Coding and Modulation for Spectral Efficient Transmission

Von der Fakultät Informatik, Elektrotechnik und Informationstechnik
der Universität Stuttgart zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von

Nabil Sven Muhammad

aus Fulda

Hauptberichter:	Prof. Dr.-Ing. J. Speidel
Mitberichter:	Prof. Dr.-Ing. B. Yang
Tag der mündlichen Prüfung:	20. Juli 2010

Institut für Nachrichtenübertragung der Universität Stuttgart

2010

Acknowledgements

This dissertation is the outcome of my activities as a research assistant at the Institute of Telecommunications (INÜ), University of Stuttgart, Germany.

I would like to express my gratitude to Professor Joachim Speidel for giving me the opportunity to work under his supervision. I owe him special thanks for having confidence in me and offering me the freedom to develop own ideas. His door was always open for fruitful discussions and advice. Furthermore, he sharpened my way of thinking towards a clear and scientific perspective. He constantly encouraged me in pursuing challenging tasks and enabled the publication of my research results.

I would also like to thank Professor Bin Yang for the assessment of this thesis and his valuable comments.

Warm thanks go to all my former colleagues at the INÜ. I had a great time working with you. Special thanks to Dr. Frank Schaich and Robert Fritsch for all technical and even more not-so-technical discussions we had (the same holds for all assistants at the INÜ — also thanks for all the fun we had playing "Kicker"). I'm indebted to Torsten Freckmann for carefully reading and correcting this thesis. Similarly, I thank Daniel Efinger for commenting the thesis. Besides Torsten and Daniel, I frequently bothered Andreas Müller and Matthias Breuninger with questions about LaTeX. Thanks for all the tricks you taught me. I also like to thank Dr. Hanns Thilo Hagmeyer for all prolific discussions I had with him. I always left his room wiser than when I've entered it. Many thanks to the secretaries Dorothee (Doro) Buchau and Jessica Driendl as well as to Agnes Schön-Abiry for daily support. Last, but far from least I like thank (in German language) the "backbone" of the institute: to Arnfried (Arnie) Eipper, Csaba Erdei, Günther Varady and Dieter Weber. Danke für Euere Unterstützung (ganz besonders an Arnie), für die vielen guten und warmherzigen Gespräche und für das gemeinsame Espresso trinken. Ich habe mich stets pudelwohl bei Euch gefühlt!

I would also like to appreciate the work of all students, who contributed to this dissertation by their study, diploma or master theses. To highlight just some names would not be fair to others, so I thank you all.

It is my pleasure also to thank my new colleagues at Sony. I am indebted to my big boss Dr. Dietmar Schill for offering me the job position and for enabling my latest publication at the VTC in Alaska. To Lothar Stadelmeier I bow down and express the warmest of thanks for being such a great mentor. Finally (as promised, Bob) thanks to the Sony poker mates.

Thanks to Lars Sonnabend for being such a good friend and for all the (in)glorious parties we had.

As my acknowledgment will exceed two pages anyway, I may as well thank Miss Ellie for being the cutest cat in the world. Thanks for bringing me mice every night.

To my family I send my deepest gratitude (again in German): Danke an meine wunderbaren Eltern, für all Euere Unterstützung. Nur durch Euch habe ich es so weit gebracht. Ihr seit stets für mich da, habt mir alles ermöglicht, und dafür liebe ich Euch! Danke auch an meinen Bruder und meine Schwester. Es ist gut zu wissen, dass Ihr immer ein offenes Ohr für mich habt.

I saved the final thank-you's for the sunshine of my life. Thank you, Andreea, my beloved, for being the person you are. For your enduring support and your patience with me, especially in those times, when writing this thesis filled out my complete week. Thanks for relieving me from most of the workload at home and even bringing me dinner to the office once. But most of all, I thank you for all the love you have given me. *You are my sunshine, my only sunshine. You make me happy when skies are gray.*

> We shall not cease from exploration,
> And the end of all our exploring,
> Will be to arrive where we started,
> And know the place for the first time.
> (T.S. Elliot)

Contents

Acronyms

APP	a posteriori probability
ASK	amplitude-shift keying
AWGN	additive white Gaussian noise
BCJR	APP decoding algorithm by Bahl, Cocke, Jelinek, and Raviv
BEC	binary erasure channel
BER	bit error rate
BICM	bit-interleaved coded modulation
BICM-ID	BICM with iterative demapping and decoding
BPSK	binary phase-shift keying
BSA	binary switching algorithm
DSL	digital subscriber line
DVB	digital video broadcasting
ECC	error control coding
EDGE	enhanced data rates for GSM evolution
EXIT	extrinsic information transfer
GF	Galois field
GMSK	Gaussian mimimum shift keying
GSM	global system for mobile communications
HSDPA	high speed downlink packet access
i.i.d.	independent and identically distributed
i.u.d.	independent and uniformly distributed
IMM	information-maintaining mapping
ISI	intersymbol interference
LDPC	low density parity check
LLR	log likelihood ratio
LSB	least significant bit
MdM	multidimensional mapping
MAP	maximum a posteriori
MI	mutual information
MIMO	multiple input multiple output
MISO	multiple input single output
ML	maximum likelihood
MLC	multilevel coding

MLSE	maximum likelihood sequence estimation
MMSE	minimum mean squared error
MSB	most significant bit
MSD	multistage decoding
NEC	non-expansion coefficient
OFDM	orthogonal frequency division multiplex
PAM	pulse-amplidude modulation
PAPR	peak-to-average power ratio
PCC	parallel concatenated code
pdf	probability density function
PDL	parallel decoding of individual levels of MLC
PEP	pairwise error probability
PLC	powerline communications
PSD	power spectral density
PSK	phase-shift keying
QAM	quadrature amplitude modulation
QoS	quality of service
QPSK	quaternary (or quadrature) phase-shift keying
RC	raised cosine
RS	Reed-Solomon
RSC	recursive systematic convolutional
SCC	serial concatenated code
SER	symbol error rate
SIMO	single input multiple output
SISO	single input single output
SNR	signal-to-noise ratio
SOVA	soft output Viterbi algorithm
SRRC	square root raised cosine
SSB	single-sideband
TCM	trellis coded modulation
UMTS	universal mobile telecommunications system
V-BLAST	vertical Bell Laboratories Layered Space-Time
WLAN	wireless local area network
WSS	wide-sense stationary
ZF	zero forcing

Symbols

$(.)^{\mathrm{H}}$	Hermitian operation (conjugate transposition)
$(.)^{\mathrm{T}}$	transposition
$(.)^{*}$	complex conjugation
B	occupied bandwidth of QAM signal
C_{BEC}	binary erasure channel capacity
$C_{\mathrm{B}}^{(m)}$	m-th bit level BICM capacity
$C_{\mathrm{S}}^{(m)}$	m-th bit level signal set capacity
$F(\omega)$	ideal bandpass filter at QAM receiver
$G_{\mathrm{RC}}(\omega)$	spectrum of RC pulse
$G_{\mathrm{rx}}(\omega)$	spectrum of matched filter at the receiver
$G_{\mathrm{tx}}(\omega)$	spectrum of SRRC from the transmitter
$H(x)$	entropy of x
H_k	MIMO channel matrix
$I(s_k, r_k)$	MI between s_k and r_k
$I_{\mathrm{E1}}^{(m)}$	m-th bit level EXIT function
I_a	MI, given perfect knowledge of a bits in $\mathbf{c}$
$L_{\mathrm{A},k,\mathrm{m}}$	a priori L-value of bit $\hat{c}_{k,m}$
$L_{\mathrm{E},k,\mathrm{m}}$	extrinsic L-value of bit $\hat{c}_{k,m}$
$L_{\mathrm{P},k,\mathrm{m}}$	a posteriori L-value of bit $\hat{c}_{k,m}$
N_0	one-sided power spectral density of AWGN
N_{d}	number of distinct distances in distance spectrum
$R_{nn}(\tau)$	autocorrelation function of process $n(t)$
$S_{nn}(\omega)$	double-sided power spectral density of process $n(t)$
T_i	transfer characteristic of i-th constituent decoder
Z_L	matrix multiplication to force zero mean mapping
$[H_k]_{i,j}$	MIMO channel coefficient from transmit antenna i to receive antenna j
C_{B}	BICM capacity
C_{C}	Shannon's channel capacity
$I_{\mathrm{A}i}$	MI between a priori L-values and transmitted bits of decoder i
$I_{\mathrm{E1,IR}}$	transfer characteristic of irregular inner code
$I_{\mathrm{E}i}$	MI between extrinsic L-values and transmitted bits of decoder i

$\mathbf{L}_{\mathrm{A},\kappa}^{(\mathrm{c})}$	vector of a priori L-values of the coded bits for the outer decoder
$\mathbf{L}_{\mathrm{A},k}$	vector of a priori L-values $L_{\mathrm{A},k,\mathrm{m}}$ for s_k
$\mathbf{L}_{\mathrm{E},\kappa}^{(\mathrm{c})}$	vector of extrinsic L-values of the coded bits from the outer decoder
$\mathbf{L}_{\mathrm{E},k}$	vector of extrinsic L-values $L_{\mathrm{E},k,\mathrm{m}}$ for s_k
$\mathbf{L}_{\mathrm{P},\kappa}^{(\mathrm{c})}$	vector of APP L-values of the coded bits from the outer decoder
$\mathbf{L}_{\mathrm{P},\kappa}^{(\mathrm{i})}$	vector of APP L-values of the uncoded (information) bits from the outer decoder
$\mathbf{L}_{\mathrm{P},k}$	vector of APP L-values $L_{\mathrm{P},k,\mathrm{m}}$ for s_k
ω_{N}	Nyquist angular frequency
f_{N}	Nyquist frequency
$\Omega(\xi)$	partial derivative of $p\left(\xi\middle\|1\right)$ with respect to x_l
$P[x]$	probability of event x
Φ	abbreviation for EXIT functions
$P_{\mathrm{A},k,\mathrm{m}}$	a priori probability of bit $\hat{c}_{k,m}$
$P_{\mathrm{P},k,m}$	a posteriori probability of $\hat{c}_{k,m}$, given r_k and $\mathbf{P}_{\mathrm{A},k[m]}$
$\mathbf{P}_{\mathrm{A},k}$	vector of a priori probabilities $P_{\mathrm{A},k,\mathrm{m}}$ for s_k
Ψ_s	multiplicity of distance $\mathrm{d}_{\mathrm{E},s}$ in distance spectrum
C_{S}	signal set capacity
Θ	normalization factor for regular 16-QAM mapping
Υ	upper bound on $I_{\mathrm{E1}}(1)$ for arbitrary M
Ξ	integrated squared error between inner and outer EXIT function
$\|.\|$	absolute value of a scalar, magnitude of vector, cardinality of a set
$\|\mathbb{X}\|$	cardinality of a set $\mathbb{X}$
$\mathbf{1}_L$	all-one row vector of length L
α_n	fraction of time, code component G_n is used
$\mathbb{X}_b^m$	set of all symbols, whose m-th bit label equals b
$\boldsymbol{\alpha}$	row vector of all α_n
$\mathscr{A}_{l_1}$	area in complex plane
$\arg(x)$	argument, angle of complex number x
$\arg\max_x f(x)$	argument x, which maximizes $f(x)$
$\arg\min_x f$	argument x, which minimizes $f(x)$
$\overset{*}{}$	convolution operator
$\overline{b}$	inversion of bit b
P_{b}	bit error rate
$c_{k,i,m}$	m-th bit from i-th bit vector out of vector $\mathbf{c}_k$ in MIMO transmission
$c_{k,m}$	m-th bit from vector $\mathbf{c}_k$
$c_{\kappa,i}$	i-th bit from code vector $\mathbf{c}_\kappa$
$\hat{\mathbf{c}}_k, \hat{\mathbf{c}}'_k$	estimate of bit vector $\mathbf{c}_k$
$\mathbf{c}_k$	encoded bit vector, input to QAM mapper
$\mathbf{c}_\kappa$	bit vector from channel encoder
$\mathbf{c}_j^{(a)}$	bit vector of length a of perfectly known bits in $\mathbf{c}$
$\mathbf{c}_{k,i}$	i-th bit vector out of vector $\mathbf{c}_k$ in MIMO transmission

χ	decision threshold for preselection in optimization algorithm
N	length of code word
N_c	number of output bits from convolutional encoder at discrete-time κ
R_c	code rate
$f(\mathbf{x})$	cost function for QAM mapping $\mathbf{x}$
$\cup$	union (of sets)
ω_c	cutoff angular frequency of SRRC
f_c	cutoff frequency of SRRC
D	delay operator
δ	shortest squared Euclidean distance between two distinct symbol vectors
$\delta(t)$	continuous-time Dirac impulse
$\delta_j^{(i)}$	coefficient of polynomial I_E1 with respect to q
δ_n	discrete-time unit impulse
M_c	number of input bits to inner block encoder
κ	discrete-time index
e	Euler's number, 2.7182818...
E_b	average energy per information bit
ε	erasure
$\mathrm{erfc}(.)$	complementary error function
$\hat{c}_{k,i}, \hat{c}'_{k,i}$	estimate of bit $c_{k,i}$
$\hat{\mathbf{x}}_l, \hat{\mathbf{x}}'_l$	estimate of MIMO transmit vector $\mathbf{x}_l$
$\mathsf{E}_\theta[x]$	expectation (average) of random variable x taken with respect to θ
$\circ\!\!-\!\!\bullet$	Fourier transform
γ_i	coefficient of polynomial I_E1 with respect to p
G	generator matrix
D_h	harmonic mean of squared Euclidean distances for MdM optimization
$\hat{\mathbf{b}}_\kappa$	estimate of source bit vector $\mathbf{b}_\kappa$
$\hat{s}_k$	hard decided symbol at demapper
$\hat{x}, \hat{x}'$	hypothetical transmit symbols, symbol estimates
I_N	identity matrix of dimension N
$\mathrm{Im}\{x_l\}$	imaginary part of x_l
K	length of information word
K_c	number of input bits to convolutional encoder at discrete-time κ
N_int	interleaver depth
$\bullet\!\!-\!\!\circ$	inverse Fourier transform
ι	integrand
j	imaginary unit, $\sqrt{-1}$
λ_0	carrier wavelength
$\ln(.)$	natural logarithm (base e)
$\log_b(.)$	logarithm to base b
M	modulation order, number of bits per QAM symbol
$\mathbb{C}$	set of complex numbers

$\mathbb{R}$	set of real numbers
$\mathbb{X}$	symbol alphabet
$\mathbb{Z}$	set of integer numbers
$\mathbf{0}$	all-zero row vector
$\mathbf{1}$	all-one column vector of length $N_\mathrm{t}M$
$\mathbf{h}_\kappa$	discrete-time impulse response of convolutional encoder
$\mathbf{x}_{[n]}$	vector $\mathbf{x}$ without the n-th entry
$\mathscr{C}$	inner encoder of SCC
$\max\limits_x f(x)$	maximum of $f(x)$ with respect to x
E_s	average symbol energy
$\mu_{l_1,m}$	mean value of Gaussian distribution of $L_{\mathrm{E},k,\mathrm{m}}$ for m-th bit of symbol x_{l_1}
d_min	minimum Euclidean distance of QAM constellation
$\min\limits_x f(x)$	minimum of $f(x)$ with respect to x
$\mu(.)$	labeling function
$\mu^{-1}(.)$	inverse labeling (demapping) function
μ_A	mean value of Gaussian distributed a priori L-values
$\mu_\mathrm{2d}(.), \mu'_\mathrm{2d}(.)$	two-dimensional reference mappings for MdM
$\nabla f(\mathbf{x}_i)$	gradient of f with respect to $\mathbf{x}_i$
Λ	non-expansion coefficient, for mapping optimization
$n!$	factorial of non-negative integer n
$n_1(t)$	in-phase component of bandpass noise $\tilde{n}(t)$
$n_2(t)$	quadrature-phase component of bandpass noise $\tilde{n}(t)$
$n_\mathrm{I}(t)$	in-phase noise component after matched filter at receiver
$n_\mathrm{Q}(t)$	quadrature-phase noise component after matched filter at receiver
$n_{k,\mathrm{I}}$	sample of in-phase noise component $n_\mathrm{I}(t)$ at receiver
$n_{k,\mathrm{Q}}$	sample of quadrature-phase noise component $n_\mathrm{Q}(t)$ at receiver
$n_{k,j}$	j-th noise component, in MIMO transmission
n_k	complex noise sample, $n_k = n_{k,\mathrm{I}} + \mathrm{j}n_{k,\mathrm{Q}}$
$\mathbf{n}_k$	MIMO noise vector
N_r	number of MIMO receive antennas
N_t	number of MIMO transmit antennas
ν	number of delay (memory) elements in convolutional encoder
L	number of QAM symbols, 2^M
ω	angular frequency
$\overline{\mathbf{e}_i}$	i-th inverted unit column vector of dimension $N_\mathrm{t}M$
P	parity check matrix
ϕ_k	phase of Rayleigh fading coefficient
π	ratio of circle's circumference to its diameter, 3.1415926...
$\mathscr{P}_x$	(dimensionless) power of complex transmit symbols
$p(x)$	probability density function of random variable x
$\mathrm{Re}\{x_l\}$	real part of x_l
α	rolloff factor of pulse shaper

$r(t)$	continuous-time received QAM signal
$r_{\mathrm{I}}(t)$	continuous-time in-phase component at QAM receiver
$r_{\mathrm{Q}}(t)$	continuous-time quadrature-phase component at QAM receiver
$r_{\mathrm{f}}(t)$	continuous-time received and passband filtered QAM signal
$r_{k,\mathrm{I}}$	sample of in-phase at receiver
$r_{k,\mathrm{Q}}$	sample of quadrature-phase at receiver
$r_{k,j}$	j-th MIMO receive symbol
r_k	complex received sample, $r_k = r_{k,\mathrm{I}} + \mathrm{j}r_{k,\mathrm{Q}}$
$\mathbf{r}_k$	MIMO receive vector
P_{s}	symbol error rate
$\mathrm{sgn}(x)$	sign of x
σ_{A}^2	variance of Gaussian distributed a priori L-values
σ_{M}^2	variance for mutation vector in optimization algorithm
σ_{h}^2	variance of Rayleigh fading coefficients h_k
$\sigma_{\mathrm{n_I}}^2$	variance (power) of in-phase noise component in received sample
$\sigma_{\mathrm{n_Q}}^2$	variance (power) of quadrature-phase noise component in received sample
σ_{n}^2	variance (power) of signal part in complex received sample
σ_{s}^2	variance (power) of complex signal part in received sample
$\sigma_{\tilde{n}}^2$	variance (power) of bandpass noise
$\sigma_{l_1,m}^2$	variance of Gaussian distribution of $L_{\mathrm{E},k,\mathrm{m}}$ for m-th bit of symbol x_{l_1}
$\mathrm{sinc}(x)$	sinc-function, $\frac{\sin(x)}{x}$
$b_{\kappa,i}$	i-th bit from source bit vector $\mathbf{b}_\kappa$
$\mathbf{b}_\kappa$	source bit vector
η	spectral efficiency (best-case, $\alpha = 0$)
γ	step size for gradient-search algorithm
$s_{k,\mathrm{I}}$	in-phase component of QAM symbol
$s_{k,\mathrm{Q}}$	quadrature-phase component of QAM symbol
x_l	QAM symbol out of set $\mathbb{X}$
τ	time
c	free space velocity of light
vec	stacking operator
$\tilde{n}(t)$	bandpass noise
$\tilde{\mathbb{X}}$	MIMO symbol alphabet
$\tilde{R}_{\mathrm{c}}$	code rate of convolutional encoder including trellis termination
$\tilde{\eta}$	spectral efficiency of QAM with arbitrary rolloff factor α
$s(t)$	continuous-time transmitted QAM signal
$s_{\mathrm{I}}(t)$	continuous-time in-phase component at QAM transmitter
$s_{\mathrm{Q}}(t)$	continuous-time quadrature-phase component at QAM transmitter
$s_{k,i}$	i-th MIMO transmit symbol
s_k	complex transmit symbol, $s_k = s_{k,\mathrm{I}} + \mathrm{j}s_{k,\mathrm{Q}}$
$\mathbf{s}_k$	MIMO transmit vector
$u_{k,i,m}$	m-th bit from i-th bit vector out of vector $\mathbf{u}_k$ for MdM

$\mathbf{u}_k$	coded bit vector for MdM
$\mathbf{u}_{k,i}$	i-th bit vector out of vector $\mathbf{u}_k$ for MdM
$\mathbf{x}$	ordered symbol alphabet, mapping vector
$\mathbf{x}_l$	MIMO transmit vector
$\mathbf{x}_{\mathrm{G}}$	Gray mapping vector
$\mathbf{x}_{\mathrm{M}}$	variation (mutation) mapping vector in optimization algorithm
ξ	dummy variable for integrations
b	bit
d_G	Hamming weight of a matrix
d_p	degree of polynomial I_{E1} with respect to p
d_q	degree of polynomial I_{E1} with respect to q
$d_{l_1,m}$	Euclidean distance between x_{l_1} and the symbol that differs in m-th bit label
f	frequency
$f_{\kappa,n}$	feedforward bits in convolutional encoder for n-th code bit
g_0	maximum value of SRRC
$g_{\mathrm{RC}}(t)$	impulse response of raised-cosine
$g_{\mathrm{rx}}(t)$	matched filter impulse response at the receiver
$g_{\mathrm{tx}}(t)$	SRRC pulse shaper at the transmitter
$g_{\mathrm{f},n,i}$	i-th coefficient of feedforward generator polynomial $g_{\mathrm{f},n}(\mathrm{D})$
$g_{\mathrm{f},n}(\mathrm{D})$	feedforward generator polynomial of n-th code bit
$g_{\mathrm{r},n,i}$	i-th coefficient of feedback generator polynomial $g_{\mathrm{r},n}(\mathrm{D})$
$g_{\mathrm{r},n}(\mathrm{D})$	feedback generator polynomial of n-th code bit
$h(\tau,t)$	time-variant impulse response of bandpass channel
h_k	Rayleigh fading coefficient
q,p	erasure probability of communication and extrinsic channel, respectively
r_i	inner radius of a mapping with symbols on concentric circles
r_o	outer radius of a mapping with symbols on concentric circles
$r_{\kappa,n}$	feedback bits in convolutional encoder for n-th code bit
t	time
$\mathrm{d}_{\mathrm{E},s}$	s-th distance in distance spectrum
ω_0	carrier angular frequency
f_0	carrier frequency
k	discrete-time index
T_{s}	symbol duration, signaling interval

Abstract

We investigate channel coding and modulation for bandwidth-efficient transmission. For the bit-interleaved coded modulation (BICM) system, we optimize mappings for quadrature amplitude modulation with respect to the resulting symbol and bit error rates at the demapper, and to the achievable BICM and signal set capacities. Further, we design mappings that are most suited for receivers, which exchange information between demapper and decoder in an iterative manner. The mappings are not restricted to certain symbol alphabets, and thus, the optimized mappings outperform similar state-of-the-art schemes. For multi-antenna systems, we examine multidimensional mappings as a means to minimize bit error rates at an iterative receiver. Finally, we derive for block codes over the binary erasure channel closed-form expressions for their transfer characteristics to be used in the extrinsic information transfer chart. Based on that, capacity approaching irregular, i.e., time-varying, codes are suggested.

Kurzfassung

Diese Arbeit untersucht Kanalkodierung und Modulation für bandbreite-effiziente Nachrichtenübertragung. Für das so genannte “bit-interleaved coded modulation”-System — einer seriellen Verkettung von Kanalkode, Bit-Verwürfelung und Quadratur-Amplituden Modulation (QAM), werden QAM-Mappings vorgestellt, die bezüglich der Symbol- und Bitfehlerhäufigkeit, die beim Detektor auftreten, optimal sind, sowie bezüglich der BICM- und der Signalvorrats-Kapazität, d.h. informationstheoretisch die Übertragungsgeschwindigkeit maximieren. Darüber hinaus werden Mappings entworfen, die für Empfänger geeignet sind, welche iterativ Informationen zwischen Detektor und Dekodierer austauschen. Die Mappings sind nicht auf herkömmliche Symbolanordnungen beschränkt, wodurch die optimierten Lösungen den Standard-Systemen überlegen sind. Bei Mehrantennen-Systemen werden multidimensionale Mappings untersucht, welche die Bitfehlerhäufigkeit bei iterativen Empfängern reduzieren. Schließlich werden Transinformations-Kennlinien für Block-Kodes hergeleitet, für den Fall, dass der Kanal durch einen binären Auslöschungskanal modelliert wird. Diese Kennlinien sind Bestandteil des so genannten EXIT-Diagramms. Darauf aufbauend werden irreguläre, also zeitvariante, Kodes vorgeschlagen, die nahe an die Kapazitäts-Grenze herankommen.

Chapter 1

Introduction

The ever growing demand for higher bit rates requires communications systems to be steadily upgraded. Wireless, nomadic, and wireline communications systems are ever trying to provide these higher data rates in order to allow for new applications like mobile video telephony and high-speed web browsing. That in turn raises the commercial popularity of these applications, bringing forth further consumers that request more and more data transmission. History has shown that this avalanche effect was still met by communications engineers, but the challenge for increased data rates is always omnipresent. Similar to Moore's law for the development of integrated circuits, Edholm's law of bandwidth claims empirically that data rates have been and will be doubled every 18 months [1].

One obvious way for increasing throughput is to allow for more bandwidth for the data transmission. However, bandwidth is a precious good when regulated, e.g., by federal network agencies as in case of the UMTS (universal mobile telecommunications system) license auctions. Even if costs are not the limiting factor, bandwidth is always bounded by the transmission system components themselves. Antennas for wireless communications are bandlimited to a degree depending mostly on their geometry, as are filters, which operate as desired only in a limited frequency range. Finally, most communication channels are also bandlimited: copper wires and coaxial cables act as lowpass filters and optical fibers transmit light only in a certain wavelength interval.

Bandwidth-efficient digital transmission techniques provide an alternative approach to increase the throughput. The most common technique is quadrature amplitude modulation (QAM), which was first proposed in [2]. Currently, many communications systems are enhanced by a simple software update to allow for higher order QAM, i.e., more bits are transmitted per channel usage. Mobile communication systems apply high order QAM to increase throughput, e.g., EDGE (enhanced data rates for GSM evolution) uses 8-phase-shift keying (PSK) compared to Gaussian mimimum shift keying (GMSK) in GSM (global system for mobile communications), so the number of bits per channel usage is tripled in this case. Similar enhancements in third generation networks use even higher spectral efficiencies, e.g., 16-QAM in HSDPA (high speed downlink packet access). The second generations of digital video broadcasting (DVB) will allow for high definition television transmission by further

extending the QAM order. Other examples of communication systems, which steadily increase their spectral efficiencies by high order QAM, are digital subscriber line (DSL) and wireless local area network (WLAN). Even optical fiber transmission systems have started to take their leave of binary intensity modulation, i.e., binary amplitude-shift keying (ASK), to further exploit their capacity [3].

In his seminal paper, Shannon determined the fundamental limits for reliable data communication [4]. For each communication channel and a given spectral efficiency, there exists a minimum signal-to-noise ratio (SNR) (sometimes referred to as capacity limit), which is necessary for error-free transmission. Bandwidth-efficient QAM schemes are more difficult designed to be also power-efficient than simple binary modulation systems. Thus, the challenge in spectral-efficient QAM transmission lies in the need for more sophisticated coding schemes. The larger the signal alphabet, the more error-prone is the transmission. Capacity approaching QAM transmission is only possible, if both the channel code and the modulation scheme are adapted well to the channel characteristics. As was already noted in [5], coding and modulation should be used as one entity. Based on this idea, trellis coded modulation (TCM) was proposed in [6] and [7]. By increasing the modulation order and partitioning the signal sets through a trellis encoder, the Euclidean free distance can be maximized, allowing for good performance on the (non-fading) additive white Gaussian noise (AWGN) channel. This target was in contrast to previous coding schemes, which rather tried to maximize the Hamming distance. For fading channels, especially if time-varying, the strict combination of channel code and modulation has the disadvantage of less flexibility. TCM schemes optimized for AWGN channels usually perform poor for fading channels [8]. Another approach, that combines channel coding and modulation, is multilevel coding (MLC) [9], which was thoroughly investigated in [10]. The idea of MLC is to provide for each bit belonging to a QAM symbol an individual channel code. Based on multiuser information theory, it was shown in [10], that multistage decoding (MSD) at the receiver side can provide capacity approaching performance. Each individual code, i.e., each level, is decoded in a well-defined order, providing a priori information for the next level. The drawbacks of MLC are its high decoding complexity and its tendency for error propagation.

To overcome the shortage of TCM over fading channels, Zehavi introduced a bitwise interleaver in between a convolutional code and a Gray [11] mapper to increase temporal diversity [12]. This pragmatic approach outperformed the best known TCM schemes over a Rayleigh fading channel. Later, this approach was denoted as bit-interleaved coded modulation (BICM) and analyzed in detail in [13]. In fact, BICM is a variant of MLC, where the receiver does not perform MSD, but rather a parallel decoding of individual levels (PDL) [10]. Close to capacity performance is possible with BICM, if Gray mapping is applied together with a strong code [13]. Candidates for the latter are large codes on graphs, like turbo [14] or low density parity check (LDPC) codes [15], which are usually decoded in an iterative manner. Since BICM is a concatenated coding scheme itself — with the mapper as the inner encoder and the channel code as the outer encoder, iterative demapping and decoding was proposed in [16, 17, 18, 19]. In a series of publications, BICM with iterative demapping and decoding (BICM-ID) was further investigated [20, 21, 22, 23, 24, 25, 26, 27]. The main

results are that a rather weak outer code is sufficient for capacity approaching performance, provided that the modulator applies anti-Gray mapping. BICM and BICM-ID have the advantage that they perform well on both fading and non-fading channels. Especially BICM-ID allows for much flexibility, since the number of iterations at the receiver may depend on the channel characteristics. Furthermore, the small but inherent gap to signal set capacity in BICM can be diminishingly small for BICM-ID. Altogether, channel coding in combination with high order modulation is a topic worth investigating in much detail.

This thesis is concerned with BICM with or without an iterative receiver. Most results will be derived for the non-fading AWGN channel model, but comparisons with the Rayleigh fading channel will show, that the results are rather robust with respect to different channels. Answers will be given to the following questions: Which mappings provide the lowest bit error rate (BER), if the receiver applies no iterations and which mappings are optimal for perfect a priori information that might arise from an iterative feedback loop. In contrast to [28] and other research, we relax the condition, that QAM symbols must lie on conventionally applied constellations such as rotationally symmetric PSK or orthogonal QAM grids. We also propose signal constellations that yield minimized symbol error rate (SER) after a hard decision. From an information theoretic perspective, we propose constellations that maximize the signal set capacity as well as mappings, which maximize the BICM capacity. Further, we investigate multidimensional mappings for multiple input multiple output (MIMO) systems. We introduce a precoder to generate these mappings, and propose schemes, which minimize the error-floor that occurs in iterative MIMO demapping schemes. Finally, we describe arbitrary QAM mappings as the concatenation of a rate 1 precoder followed by Gray mapping. Analysis of the mapping is reduced to the investigation of the precoder. Based on analytical formulas, we find sets of precoders to be used in a time-varying manner. These sets will achieve the best match to a given outer encoder.

The structure of this thesis is as follows. Chapter 2 summarizes the most important fundamentals, which are necessary for the understanding of this thesis. In Chapter 3, we propose QAM mappings, which are optimum with respect to various cost functions, such as the resulting symbol and bit error rates, the achievable capacities or performance for iterative demapping. Multidimensional MIMO mappings are discussed in Chapter 4. A precoding scheme for their generation is introduced, and based on that, precoders are derived, which minimize the error-floor for iterative MIMO receivers. Further, we derive in Chapter 5 closed-form expressions for EXIT functions of the precoder under the binary erasure channel (BEC) assumption and apply the results both to the optimization of irregular inner codes and to the EXIT functions of a QAM demapper. Chapter 6 concludes this thesis with an outlook of possible future research topics.

Chapter 2

Fundamentals

This chapter summarizes all fundamentals that are necessary for the understanding of this thesis. We first derive the BICM system model and its mathematical description, followed by the characterization of different channel models, which are examined in this thesis. Error control coding techniques are outlined and optimum decoding rules recapitulated, which will be the basis of the iterative receiver of BICM-ID. Finally, the most important fundamentals of information theory are explained and the EXIT chart is introduced.

2.1 System Model for QAM

In this section, we explain the system model of spectral-efficient (digital) QAM. This scheme is of such importance, that it shall be developed step by step. Starting with the continuous-time description of the transmitted waveforms, we examine the optimum receiver filter and derive the average powers of both signal and noise. Based on that we introduce the equivalent baseband description for digital QAM, which allows for a more convenient notation. A classification of QAM mappings concludes this section.

2.1.1 Waveform Description of QAM

Figure 2.1 depicts the transmitter of the continuous-time QAM as well as the bandpass channel. The data inputs to the QAM transmitter per channel usage are two real-valued symbols, $s_{k,\mathrm{I}}$ and $s_{k,\mathrm{Q}}$. They arrive each T_s seconds, where T_s is the symbol duration or signaling interval, and are enumerated by discrete-time index k. Subscripts “I” and “Q” denote the in-phase and quadrature-phase component, respectively. The symbols $s_{k,\mathrm{I}}$ and $s_{k,\mathrm{Q}}$ are individually independent. This assumption will be justified by the insertion of an interleaver in the BICM scheme, cf. Section 2.4. In principal, they may also be jointly independent, but many QAM schemes interrelate them, as both are typically treated as one joint symbol. This dependency is considered more closely in Subsection 2.1.3, as well as the set of discrete levels, which

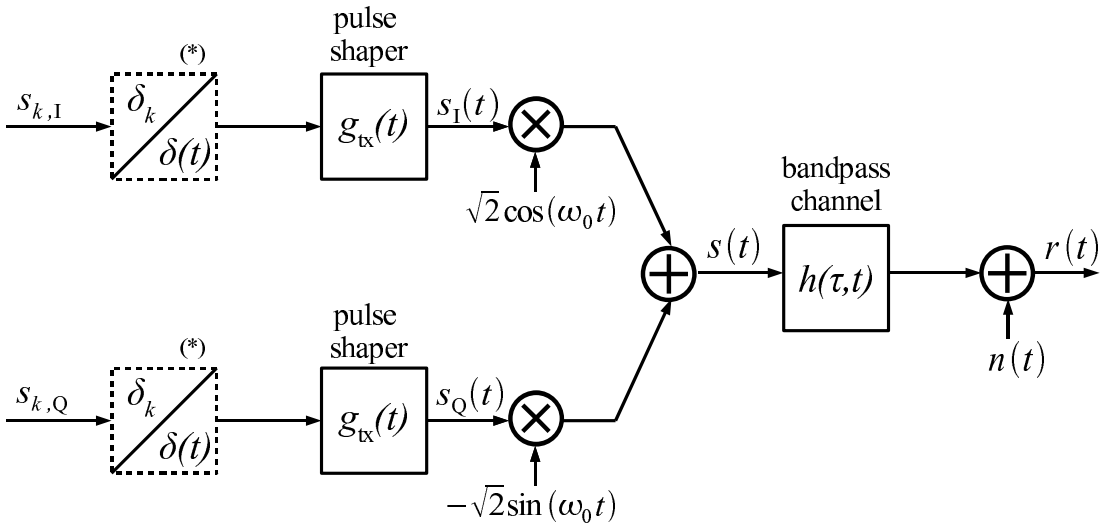


Figure 2.1: Transmitter and channel for continuous-time QAM

$s_{k,\mathrm{I}}$ and $s_{k,\mathrm{Q}}$ might take on. The dashed box (*) converts an incoming unit impulse δ_k to a Dirac-impulse $\delta(t)$. Its only significance is due to a precise system-theoretical description. The pulse shaper at the transmitter $g_{\mathrm{tx}}(t)$ shall possess the well-known square root raised cosine (SRRC) impulse response

$$g_{\mathrm{tx}}(t) = g_0 T_{\mathrm{s}} \cdot \frac{\sin\left(\pi(1-\alpha)\frac{t}{T_{\mathrm{s}}}\right) + 4\alpha\frac{t}{T_{\mathrm{s}}}\cos\left(\pi(1+\alpha)\frac{t}{T_{\mathrm{s}}}\right)}{\pi t\left(1-\left(4\alpha\frac{t}{T_{\mathrm{s}}}\right)^2\right)}. \tag{2.1}$$

$\alpha \in [0,1]$ is the rolloff factor and $g_0 = g_{\mathrm{tx}}(0)$ is the maximum value of the SRRC. It represents the dimension of the physical nature of the waveform, e.g., V for an electrical voltage or $\frac{\mathrm{V}}{\mathrm{m}}$ for the electric field of an electromagnetic wave. Note that $g_{\mathrm{tx}}(t)$ is non-causal and thus, real applications must apply a windowing function and an appropriate temporal delay to achieve $g_{\mathrm{tx}}(t) = 0,\ \forall t < 0$, for causality. The corresponding Fourier spectrum is given by

$$G_{\mathrm{tx}}(\omega) = \begin{cases} g_0 T_{\mathrm{s}} & , \ \frac{|\omega|}{\omega_{\mathrm{N}}} \leq 1-\alpha \\ g_0 T_{\mathrm{s}} \cdot \cos\left(\frac{\pi}{4}\frac{|\omega|-(1-\alpha)\omega_{\mathrm{N}}}{\alpha\omega_{\mathrm{N}}}\right) & , \ 1-\alpha \leq \frac{|\omega|}{\omega_{\mathrm{N}}} \leq 1+\alpha \\ 0 & , \ \frac{|\omega|}{\omega_{\mathrm{N}}} \geq 1+\alpha \end{cases} \tag{2.2}$$

and can be seen in Figure 2.2. The spectrum of the SRRC is strictly bandlimited with cutoff angular frequency $\omega_{\mathrm{c}} = 2\pi f_{\mathrm{c}} = (1+\alpha)\omega_{\mathrm{N}}$, where $\omega_{\mathrm{N}} = 2\pi f_{\mathrm{N}} = \frac{\pi}{T_{\mathrm{s}}}$ is called Nyquist angular frequency. The purpose for this special class of pulses will become evident soon.

The continuous-time in-phase component

$$s_{\mathrm{I}}(t) = \sum_{k=-\infty}^{\infty} s_{k,\mathrm{I}} \cdot g_{\mathrm{tx}}(t - kT_{\mathrm{s}}) \tag{2.3}$$

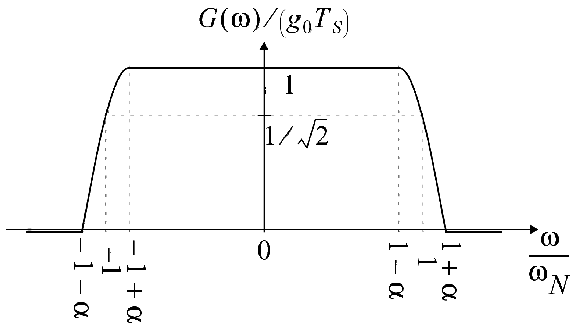


Figure 2.2: Spectrum of SRRC for arbitrary rolloff factor α

is multiplied by a cosine-carrier with angular carrier frequency $\omega_0 = 2\pi f_0$. The resulting frequency shift of the spectrum in (2.2) yields the condition that $f_0 \geq f_c$. It is worth noting, that two exceptions for this condition exist. In case of baseband transmission, $f_0 = 0$, and thus only the in-phase part is transmitted. This means, that the QAM scheme reduces to pulse-amplidude modulation (PAM). Applying in addition single-sideband (SSB) modulation [29], we may even allow for an arbitrary f_0. The factor of $\sqrt{2}$ in Figure 2.1 is just a normalization to assure the desired result, if transmitter and receiver apply the same multipliers. Similarly, the continuous-time quadrature-phase component

$$s_{\mathrm{Q}}(t) = \sum_{k=-\infty}^{\infty} s_{k,\mathrm{Q}} \cdot g_{\mathrm{tx}}(t - kT_{\mathrm{s}}) \tag{2.4}$$

is multiplied by a sine-carrier of the same carrier frequency. The fact that cosine and sine are two functions that stand in quadrature, i.e., they are orthogonal to each other, justifies the notation “quadrature amplitude modulation”. The transmitted signal after addition of upper and lower branch in Figure 2.1 becomes

$$s(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} \left[s_{k,\mathrm{I}} \cdot g_{\mathrm{tx}}(t - kT_{\mathrm{s}}) \cdot \cos(\omega_0 t) - s_{k,\mathrm{Q}} \cdot g_{\mathrm{tx}}(t - kT_{\mathrm{s}}) \cdot \sin(\omega_0 t) \right]. \tag{2.5}$$

The linear bandpass channel is described by its time-variant impulse response $h(\tau, t)$. If the channel is deterministic, $h(\tau, t)$ can be appropriately modeled as a time-variant linear filter, and the Fourier-transforms with respect to both τ and t exist. In all other cases, the channel behavior may be analyzed statistically and the sample function $h(\tau, t)$ is treated as a stochastic process. In any case, its dimension shall be $\frac{1}{s}$, so that channel input and output have the same dimension.

First, we assume, that the channel is ideal, i.e., $h(\tau, t) = \delta(\tau, t)$. This will yield the non-fading AWGN channel, described in Subsection 2.2.1. Hence, the channel does not attenuate (or amplify) the transmit signal, nor does it introduce a temporal delay of the signal. These assumptions are valid, if the actual attenuation of the channel is already incorporated into the transmitted signal and if the actual delay is considered at the receiver by appropriate synchronization methods. An instance of a fading channel will be discussed in Subsection 2.2.2.

Noise $n(t)$ is added to the channel output, so the received signal in the non-fading case is

$$r(t) = s(t) + n(t). \tag{2.6}$$

Typically, $n(t)$ is assumed to be a wide-sense stationary (WSS) process with Gaussian distribution and zero mean. Its double-sided power spectral density (PSD) is $S_{nn}(\omega) = N_0/2, \forall \omega$. This constant composition of all frequencies resembles that of white light in the visual domain. Hence, $n(t)$ is said to be additive white Gaussian noise (AWGN). At this point, it is necessary to introduce an ideal bandpass filter $F(\omega)$ at the receiver input to limit the noise power, which would be infinite otherwise. $F(\omega)$ is 1 for $|\omega| \in [\omega_0 - \omega_c, \omega_0 + \omega_c]$ and 0 else. This filter does not affect the transmit signal $s(t)$, which has only spectral components in the passband of $F(\omega)$. The filtered and thus bandlimited noise part of the received signal is denoted as $\tilde{n}(t)$ and its PSD $S_{\tilde{n}\tilde{n}}(\omega)$ can be seen in Figure 2.3. The filtered received signal is

$$r_{\mathrm{f}}(t) = s(t) + \tilde{n}(t). \tag{2.7}$$

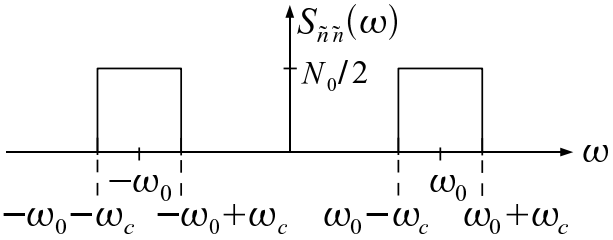


Figure 2.3: Power spectral density of filtered noise $\tilde{n}(t)$

The power of $\tilde{n}(t)$ equals its variance, since linear filters maintain zero mean, and is computed as

$$\sigma_{\tilde{n}}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\tilde{n}\tilde{n}}(\omega)\, d\omega = N_0 \cdot B, \tag{2.8}$$

where B is the bandwidth, which is occupied by the QAM signal,

$$B = (f_0 + f_{\mathrm{c}}) - (f_0 - f_{\mathrm{c}}) = 2 f_{\mathrm{c}}. \tag{2.9}$$

The (coherent) QAM receiver is depicted in Figure 2.4. We now analyze the upper in-phase branch in detail. Perfect carrier and clock synchronization are assumed. The received in-phase component after matched filtering with $g_{\mathrm{rx}}(t)$ is given as

$$\begin{aligned} r_{\mathrm{I}}(t) \quad &= \quad \left[r_{\mathrm{f}}(t) \cdot \sqrt{2} \cos(\omega_0 t) \right] * g_{\mathrm{rx}}(t) \\ &\overset{(2.5),(2.7)}{=} \quad \left\{ \sum_{k=-\infty}^{\infty} 2 \left[s_{k,\mathrm{I}} \cdot \cos^2(\omega_0 t) - s_{k,\mathrm{Q}} \cdot \sin(\omega_0 t) \cos(\omega_0 t) \right] g_{\mathrm{tx}}(t - kT_{\mathrm{s}}) \right. \\ &\qquad \left. + \sqrt{2} \left[\tilde{n}(t) \cdot \cos(\omega_0 t) \right] \right\} * g_{\mathrm{rx}}(t) \end{aligned} \tag{2.10}$$

An optimum receiver for the AWGN channel applies the well-known matched filter, which maximizes the SNR [30]. Note that the filtered noise $\tilde{n}(t)$ is white in the frequency band of

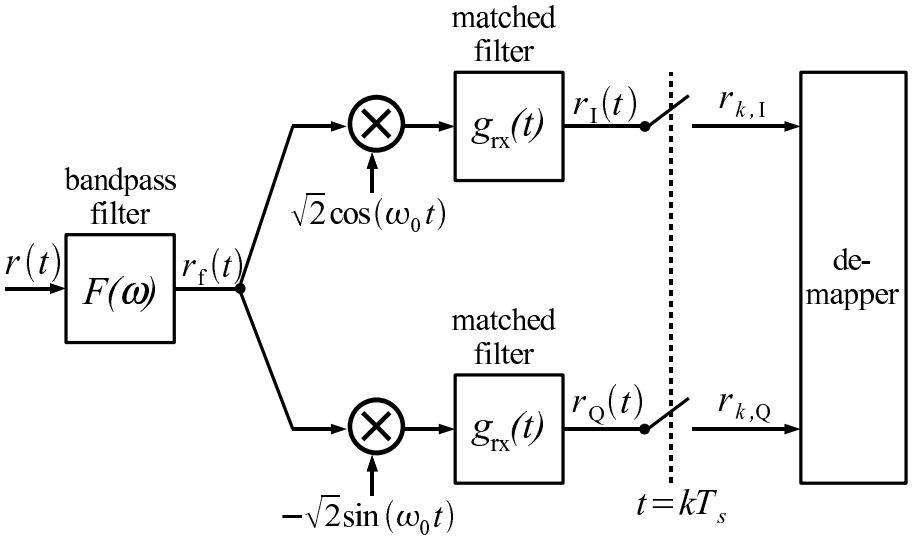


Figure 2.4: Receiver for continuous-time QAM

interest. Thus, the (real-valued) impulse response and transfer function of the matched filter are

$$g_{\mathrm{rx}}(t) = \frac{1}{g_0 T_\mathrm{s}} g_{\mathrm{tx}}(-t) \circ\!\!-\!\!\bullet\, G_{\mathrm{rx}}(\omega) = \frac{1}{g_0 T_\mathrm{s}} G_{\mathrm{tx}}^*(\omega). \tag{2.11}$$

The operator $(.)^*$ denotes complex conjugation. A dimensionless transfer function is ensured by the factor $\frac{1}{g_0 T_\mathrm{s}}$. Since $g_{\mathrm{tx}}(t)$ in (2.1) has even symmetry, its spectrum $G_{\mathrm{tx}}(\omega)$ in (2.2) is real-valued and also even. Thus, the same properties apply on $g_{\mathrm{rx}}(t)$ and $G_{\mathrm{rx}}(\omega)$, and we can write $g_{\mathrm{rx}}(t) = \frac{1}{g_0 T_\mathrm{s}} g_{\mathrm{tx}}(t)$ and $G_{\mathrm{rx}}(\omega) = \frac{1}{g_0 T_\mathrm{s}} G_{\mathrm{tx}}(\omega)$. With the relations $\cos^2(x) = \frac{1}{2}\left(1+\cos(2x)\right)$ and $\sin(x)\cos(x) = \frac{1}{2}\sin(2x)$, and the fact, that the matched filter is strictly bandlimited with cutoff frequency f_c, we rewrite (2.10) as

$$r_\mathrm{I}(t) = \sum_{k=-\infty}^{\infty} s_{k,\mathrm{I}} \cdot g_{\mathrm{tx}}(t - kT_\mathrm{s}) * g_{\mathrm{rx}}(t) + n_\mathrm{I}(t), \tag{2.12}$$

with the in-phase noise component

$$n_\mathrm{I}(t) = \sqrt{2}\left[\tilde{n}(t) \cdot \cos(\omega_0 t)\right] * g_{\mathrm{rx}}(t). \tag{2.13}$$

Note that $\left|G_{\mathrm{rx}}(\omega)\right|^2$ is the (dimensionless) transfer function of the raised cosine (RC) pulse $g_{\mathrm{RC}}(t) \circ\!\!-\!\!\bullet\, G_{\mathrm{RC}}(\omega)$,

$$G_{\mathrm{RC}}(\omega) = \left|G_{\mathrm{rx}}(\omega)\right|^2 = \begin{cases} 1 & , \ \frac{|\omega|}{\omega_\mathrm{N}} \le 1-\alpha \\ \cos^2\left(\frac{\pi}{4}\frac{|\omega|-(1-\alpha)\omega_\mathrm{N}}{\alpha\omega_\mathrm{N}}\right) & , \ 1-\alpha \le \frac{|\omega|}{\omega_\mathrm{N}} \le 1+\alpha \\ 0 & , \ \frac{|\omega|}{\omega_\mathrm{N}} \ge 1+\alpha \end{cases}, \tag{2.14}$$

$$g_{\mathrm{RC}}(t) = g_{\mathrm{rx}}(t) * g_{\mathrm{rx}}(-t) = \frac{1}{T_\mathrm{s}} \mathrm{sinc}\left(\pi \frac{t}{T_\mathrm{s}}\right) \frac{\cos\left(\pi\alpha t/T_\mathrm{s}\right)}{1-\left(2\alpha t/T_\mathrm{s}\right)^2}. \tag{2.15}$$

The sinc-function $\mathrm{sinc}(x)$ is defined as $\frac{\sin(x)}{x}$. If we insert $g_{\mathrm{tx}}(t)$ from (2.11) in (2.12) and introduce (2.15), we get after changing the summation index

$$r_{\mathrm{I}}(t) = g_0 T_{\mathrm{s}} \sum_{l=-\infty}^{\infty} s_{l,\mathrm{I}} \cdot g_{\mathrm{RC}}(t - lT_{\mathrm{s}}) + n_{\mathrm{I}}(t). \tag{2.16}$$

Sampling at discrete-time instances $t = kT_{\mathrm{s}}$, the in-phase input to the demapper becomes

$$r_{k,\mathrm{I}} = r_{\mathrm{I}}(kT_{\mathrm{s}}) = g_0 T_{\mathrm{s}} \sum_{l=-\infty}^{\infty} s_{l,\mathrm{I}} \cdot g_{\mathrm{RC}}((k-l)T_{\mathrm{s}}) + n_{\mathrm{I}}(kT_{\mathrm{s}}). \tag{2.17}$$

From (2.15), it is easily verified that $g_{\mathrm{RC}}(0) = {T_{\mathrm{s}}}^{-1}$ and $g_{\mathrm{RC}}(t) = 0$ for all $t = kT_{\mathrm{s}}, k \in \mathbb{Z} \setminus \{0\}$. At this point, the choice for the SRRC pulse becomes clear: its matched filter output fulfills the Nyquist criterion, so that discrete-time samples are free of intersymbol interference (ISI) [30]. Denoting the in-phase noise sample $n_{\mathrm{I}}(kT_{\mathrm{s}})$ as $n_{k,\mathrm{I}}$, the final result is

$$r_{k,\mathrm{I}} = g_0 \cdot s_{k,\mathrm{I}} + n_{k,\mathrm{I}}. \tag{2.18}$$

In the same manner, the quadrature-phase sample is expressed as

$$r_{k,\mathrm{Q}} = g_0 \cdot s_{k,\mathrm{Q}} + n_{k,\mathrm{Q}}, \tag{2.19}$$

where $n_{k,\mathrm{Q}}$ is the sample at $t = kT_{\mathrm{s}}$ of the quadrature-phase noise component

$$n_{\mathrm{Q}}(t) = -\sqrt{2}\left[\tilde{n}(t) \cdot \sin(\omega_0 t)\right] * g_{\mathrm{rx}}(t). \tag{2.20}$$

Based on the observations $r_{k,\mathrm{I}}$ and $r_{k,\mathrm{Q}}$ and possibly a priori knowledge, the demapper performs either a hard or soft decision. This will be discussed in Subsection 2.3.3.

Let us now examine the noise properties more closely. A sample function of the (narrowband) bandpass Gaussian noise process $\tilde{n}(t)$ can be modeled as

$$\tilde{n}(t) = n_1(t)\cos(\omega_0 t) - n_2(t)\sin(\omega_0 t), \tag{2.21}$$

where $n_1(t)$ and $n_2(t)$ are jointly WSS lowpass processes. It can be shown (e.g. [30], Chapter 2.9) that both processes are independent and identically distributed (i.i.d.), having Gaussian distribution with zero mean and variance equal to $\sigma_{\tilde{n}}^2 = N_0 \cdot B$. Thus, their PSD $S_{n_i n_i}(\omega), i = 1,2$, is N_0 for $|\omega| < \omega_{\mathrm{c}}$ and 0 elsewhere. The same considerations as in (2.12) yield for the in- and quadrature-phase noise components

$$n_{\mathrm{I}}(t) = \frac{1}{\sqrt{2}} n_1(t) * g_{\mathrm{rx}}(t) \quad \text{and} \quad n_{\mathrm{Q}}(t) = \frac{1}{\sqrt{2}} n_2(t) * g_{\mathrm{rx}}(t). \tag{2.22}$$

The PSD of both in- and quadrature-phase noise component are computed with the Wiener-Khinchin theorem [31] as

$$S_{n_{\mathrm{I}} n_{\mathrm{I}}}(\omega) = S_{n_{\mathrm{Q}} n_{\mathrm{Q}}}(\omega) = \frac{1}{2} S_{n_1 n_1}(\omega) \cdot \left|G_{\mathrm{rx}}(\omega)\right|^2 = \frac{N_0}{2} \cdot G_{\mathrm{RC}}(\omega), \tag{2.23}$$

where we have used (2.14). The corresponding autocorrelation functions are obtained from the inverse Fourier transform of (2.23) as

$$S_{n_\mathrm{I} n_\mathrm{I}}(\omega) \circledast\!\!-\!\!\circ R_{n_\mathrm{I} n_\mathrm{I}}(\tau) = R_{n_\mathrm{Q} n_\mathrm{Q}}(\tau) = \frac{N_0}{2} \cdot g_\mathrm{RC}(\tau). \tag{2.24}$$

Obviously, $R_{n_\mathrm{I} n_\mathrm{I}}(kT_\mathrm{s}) = 0, \forall k \in \mathbb{Z} \setminus \{0\}$, so that the samples $n_{k,\mathrm{I}}$ are orthogonal. Since they have zero mean, they are uncorrelated and from their Gaussian distribution, we can follow that they are independent as well. The same holds true for samples $n_{k,\mathrm{Q}}$. The variances of $n_{k,\mathrm{I}}$ and $n_{k,\mathrm{Q}}$ are

$$\sigma^2_{\mathrm{n_I}} = \sigma^2_{\mathrm{n_Q}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{n_\mathrm{I} n_\mathrm{I}}(\omega)\, d\omega = R_{n_\mathrm{I} n_\mathrm{I}}(0) \overset{(2.15)}{=} \frac{N_0}{2T_\mathrm{s}}\,, \quad \forall \alpha \in [0,1]. \tag{2.25}$$

2.1.2 Equivalent Baseband Description of Digital QAM

The lowpass (or baseband) equivalent of a bandpass signal allows for a simpler mathematical description. However, the signals become in general complex-valued. An elaborate derivation of a lowpass equivalent by means of its analytical signal can be found, e.g., in [30]. If we focus on digital QAM, it is sufficient to consider the sampled, i.e., discrete-time, output of the QAM receiver. We define $r_k = r_{k,\mathrm{I}} + \mathrm{j} r_{k,\mathrm{Q}}$ and in the same manner the complex values s_k and n_k. The relations (2.18) and (2.19) from the previous section can now be combined in compact notation as

$$r_k = g_0 \cdot s_k + n_k. \tag{2.26}$$

Figure 2.5 depicts the corresponding system model. The transmitter from Figure 2.1 is replaced by a QAM mapper that assigns every T_s seconds a bit vector $\mathbf{c}_k$ of length M to one symbol $s_k \in \mathbb{X}$ according to a labeling function μ. This will be discussed in the next subsection.

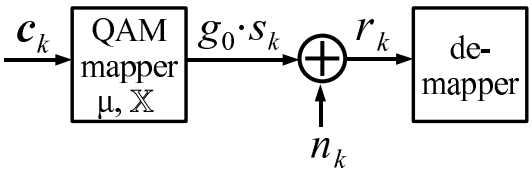


Figure 2.5: Equivalent baseband representation of discrete-time QAM

It can be shown that $s_\mathrm{I}(t)$ and $s_\mathrm{Q}(t)$ are uncorrelated, if $s(t)$ has zero mean [30]. This is due to the fact that cosine and sine in (2.5) are mathematically orthogonal functions. The same is then true for the signal part of the matched filter samples $s_{k,\mathrm{I}}$ and $s_{k,\mathrm{Q}}$, i.e., for the real and imaginary part of the transmit symbols s_k. The variance of a complex process, whose real and imaginary part are uncorrelated, is the sum of the variances of its real and imaginary part. Hence, if s_k has zero mean, the information-bearing signal part of the received sample in (2.26) has variance

$$\sigma^2_\mathrm{s} = g_0^2 \cdot \mathsf{E}_{s_k}\left[s_k \cdot s_k^*\right] = g_0^2 \cdot \mathscr{P}_x, \tag{2.27}$$

where the operator $\mathsf{E}_\theta[x]$ denotes expectation of random variable[1] x taken with respect to θ, and $\mathscr{P}_x$ is the average (dimensionless) symbol power, i.e., the second moment of the transmit symbols s_k, so $\mathscr{P}_x = \mathsf{E}_{s_k}\left[|s_k|^2\right]$. It is reasonable to assume that the signaling alphabet $\mathbb{X}$, from which the possible transmit symbols s_k are drawn, is designed in such a way that $\mathsf{E}_{s_k}[s_k] = 0$. Any non-zero mean would induce a larger average transmit power. So, in case of zero mean, the average signal power equals variance σ_s^2.

The power of the complex noise n_k is denoted as σ_n^2 and is according to (2.25)

$$\sigma_\mathrm{n}^2 = \sigma_{\mathrm{n_I}}^2 + \sigma_{\mathrm{n_Q}}^2 = \frac{N_0}{T_\mathrm{s}}, \tag{2.28}$$

i.e., twice as large as the variance per dimension, because real and imaginary part of the noise sample n_k are statistically independent and thus also uncorrelated.

In this thesis, we will make frequent use of two different notations of SNR. First, the ratio of average signal to noise *power* in logarithmic notation is

$$\left.\frac{\sigma_\mathrm{s}^2}{\sigma_\mathrm{n}^2}\right|_\mathrm{dB} = 10 \cdot \log_{10}\left(\frac{g_0^2 \cdot \mathscr{P}_x}{N_0/T_\mathrm{s}}\right). \tag{2.29}$$

As the signaling interval is T_s, the average symbol energy $E_\mathrm{s} = \sigma_\mathrm{s}^2 \cdot T_\mathrm{s}$. We will see in Section 2.4 that the input to the QAM mapper are M bits per T_s, which are in general encoded. Only a fraction of $R_\mathrm{c} \cdot M$ bits per T_s are information bits that leave the bit source, the rest is redundancy introduced by a channel encoder. The factor R_c is referred to as the code rate.

We now define the spectral efficiency $\tilde{\eta}$ as the amount of information bits per second that is transmitted in a bandwidth B, which is a function of the rolloff factor α

$$\tilde{\eta} = \frac{R_\mathrm{c} \cdot M/T_\mathrm{s}}{B} \overset{(2.9)}{=} \frac{R_\mathrm{c} \cdot M/T_\mathrm{s}}{2 \cdot f_\mathrm{c}} = \frac{R_\mathrm{c} \cdot M/T_\mathrm{s}}{(1+\alpha)/T_\mathrm{s}} = \eta \cdot \frac{1}{(1+\alpha)}, \tag{2.30}$$

where $\eta = R_\mathrm{c} \cdot M$ is the maximum spectral efficiency, which can be achieved by setting $\alpha = 0$, while $\alpha = 1$ yields the smallest spectral efficiency of $\eta/2$. However, an SRRC pulse with $\alpha = 0$ has several drawbacks. On the one hand, its impulse response has a slow degradation of the slopes, which results in larger ISI, if clock synchronization is imperfect, i.e., if sampling jitter occurs. On the other hand, it becomes more difficult to design a pulse shaper that yields almost infinitely steep transitions in frequency domain. Nevertheless, we will consider in this thesis the best case for spectral efficiency and thus set $\tilde{\eta} = \eta$. With that we can compute the average energy per information bit as $E_\mathrm{b} = E_\mathrm{s}/\eta$.

The second SNR definition relates the average *energy* per information bit to the noise power spectral density, i.e.,

$$\left.\frac{E_\mathrm{b}}{N_0}\right|_\mathrm{dB} = 10 \cdot \log_{10}\left(\frac{E_\mathrm{s}}{N_0 \cdot \eta}\right) = 10 \cdot \log_{10}\left(\frac{\sigma_\mathrm{s}^2 \cdot T_\mathrm{s}}{N_0 \cdot \eta}\right) \overset{(2.29)}{=} \left.\frac{\sigma_\mathrm{s}^2}{\sigma_\mathrm{n}^2}\right|_\mathrm{dB} - 10 \cdot \log_{10}\eta. \tag{2.31}$$

[1] In order to ease notation, we will not differentiate between a random variable and a certain realization, but rather use the same letter for both. It will become clear from the context, which case is meant.

Both SNRs, (2.29) and (2.31), carry the pseudo-dimension *decibel*, or dB in short. If no confusion is to be feared, we will omit the subindex "dB" and simply refer to the SNRs as to $\frac{\sigma_s^2}{\sigma_n^2}$ and $\frac{E_b}{N_0}$, respectively.

The discrete-time equivalent baseband notation in (2.26) allows also for simple implementation in simulation programs as well as for a more abstract description. For this purpose, the dimensions are irrelevant[2] and we will set for the rest of this thesis without loss of generality $g_0 = 1$ and $T_s = 1$. Furthermore, as only the *ratios* in (2.29) and (2.31) are of interest, we also normalize $\mathscr{P}_x = 1$.

2.1.3 QAM Mappings

The QAM mapper in Figure 2.5 assigns symbols to incoming bit vectors. Every T_s seconds, it receives M bits as the vector $\mathbf{c}_k = \left(c_{k,0}, c_{k,1}, \ldots, c_{k,M-1}\right)$ and outputs a complex symbol $s_k \in \mathbb{X}$ according to labeling function μ, i.e., $s_k = \mu(\mathbf{c}_k)$. In correspondence to most coding theory books, we let vectors be per definition *row* vectors and denote them by bold letters. The mapping is fully defined by the symbol alphabet $\mathbb{X}$ and the one-to-one (or bijective) labeling function μ. The inverse function performs demapping from symbols to bits, i.e., $\mu^{-1}(s_k) = \left(\mu_0^{-1}(s_k), \ldots, \mu_{M-1}^{-1}(s_k)\right) = \left(c_{k,0}, c_{k,1}, \ldots, c_{k,M-1}\right)$. The set $\mathbb{X}$ consists of $L = 2^M$ symbols x_l that are in general complex, but are not necessarily different. Hence, the set is $\mathbb{X} = \{x_0, \ldots, x_l, \ldots, x_{L-1}\}$ and its cardinality is $|\mathbb{X}| = L$. The plot of all $x_l \in \mathbb{X}$ in the complex plane is called constellation or signal space diagram. For the latter, however, special care is required about the scaling of the axes, as those represent an orthonormal basis [30]. So we will rather use the former notation.

The average symbol power $\mathscr{P}_x$ as previously defined considers the discrete-time averaging of $|s_k|^2$. According to the labeling function μ, each s_k corresponds to one realization $x_l \in \mathbb{X}$. Thus, the process is ergodic and we can also consider the ensemble average of all possible symbols x_l. It can be assumed that all symbols x_l occur equally likely with probability $P[x_l] = L^{-1}$. Signal shaping techniques [32] with non-equiprobable symbols constitute an exception to this assumption, but are not considered in this thesis. The symbol power constraint is then formulated as

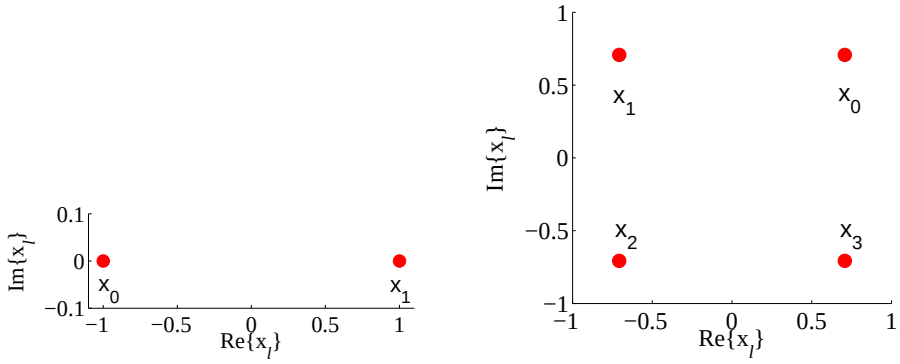
$$\mathscr{P}_x = \mathsf{E}_{s_k}\left[|s_k|^2\right] = \mathsf{E}_{x_l}\left[|x_l|^2\right] = \frac{1}{L}\sum_{l=0}^{L-1}|x_l|^2 \stackrel{!}{=} 1. \tag{2.32}$$

Examples: Let us consider the simplest case of $M = 1$, thus $\mathbb{X} = \{x_0, x_1\}$. Two variants are possible. First, for 2-ASK, we choose $\mu(0) = x_0 = 0$ and $\mu(1) = x_1 = \sqrt{2}$, such that $\mathscr{P}_x = 1$. This is the preferred modulation scheme for optical fiber communications due to its

[2] Only in light of a dimensionless description, the imprecise statement of some publications and textbooks, that the variance per noise component shall equal $N_0/2$, may be tolerated.

simplicity, as a laser has to be switched on and off only. This scheme is called binary intensity modulation, too. Note that 2-ASK has $\mathsf{E}_{s_k}[s_k] = \sqrt{2}/2 \neq 0$. However, all mappings, treated in this thesis, have zero mean. The second variant is binary phase-shift keying (BPSK), which is an antipodal signaling scheme with $\mu(0) = x_0 = -1$ and $\mu(1) = x_1 = 1$. Its constellation diagram is shown in Figure 2.6(a). It is obvious that BPSK achieves smaller SER for the same $\frac{E_b}{N_0}$ as 2-ASK, because its squared symbol distance $|x_0 - x_1|^2 = 4$ is twice as large as for 2-ASK. Both cases are one-dimensional mappings, in which only the in-phase component has to be considered.

For $M = 2$, the four possible symbols are given as $x_l = \mathrm{e}^{\mathrm{j}\left(\frac{\pi}{4} + \frac{\pi}{2} l\right)}, l \in \{0,1,2,3\}$, cf. Figure 2.6(b). This scheme is called quaternary (or quadrature) phase-shift keying (QPSK). Note that any rotation of the symbols does not affect the performance, because the noise has zero mean and its real and imaginary part are independent (also called circularly symmetric noise process). Hence, we could also choose the symbols $x'_l = \mathrm{e}^{\mathrm{j}\frac{\pi}{2} l}$. From symbols x_l, however, we can directly see that a QPSK can be decomposed into two BPSKs with appropriately scaled symbol distances, one transmitted over the in-, the other over the quadrature-phase component.



(a) $M = 1$, BPSK

(b) $M = 2$, QPSK

Figure 2.6: Constellation diagrams for $M = 1$ and $M = 2$

There are only two different labelings possible for QPSK. If adjacent symbols differ in one bit (out of bit vector $\mathbf{c}_k$) only, we refer to it as *Gray* labeling or altogether as Gray mapping [11]. This is the case if, e.g., $\mu(0,0) = x_0, \mu(1,0) = x_1, \mu(1,1) = x_2, \mu(0,1) = x_3$. If we interchange the positions of $\mu(1,0)$ and $\mu(1,1)$, we obtain an *anti-Gray* labeling (or anti-Gray mapping).

Clearly, one or more of the following *invariant transformations* do not affect the properties of a mapping:

1. rotation of all symbols by arbitrary angle φ (see discussion above),
2. inversion of m-th bit $c_m = b \in \{0,1\}$ to $c_m = \bar{b}$, where the bar indicates inversion,

3. interchanging of bit positions c_{m_1} and c_{m_2},

4. reflection on $\mathrm{Re}\{x_l\}$- and/or $\mathrm{Im}\{x_l\}$-axis.

Thus, every other QPSK mapping can be transformed to either the Gray or anti-Gray mapping as proposed above. We now define a more stringent definition of Gray labeling. This is needed especially for high order mappings (large M). First, let us introduce the minimum Euclidean distance as

$$d_{\min} = \min_{\substack{l_1,l_2=0,\ldots,L-1 \\ l_1 \neq l_2}} \left| x_{l_1} - x_{l_2} \right| \tag{2.33}$$

and further the subset of all symbols, whose m-th bit label equals b, as

$$\mathbb{X}_b^m = \left\{ x_l \,\middle|\, \mu_m^{-1}(x_l) = b \right\}. \tag{2.34}$$

It is plain to see that $\mathbb{X} = \mathbb{X}_b^m \cup \mathbb{X}_{\overline{b}}^m$.

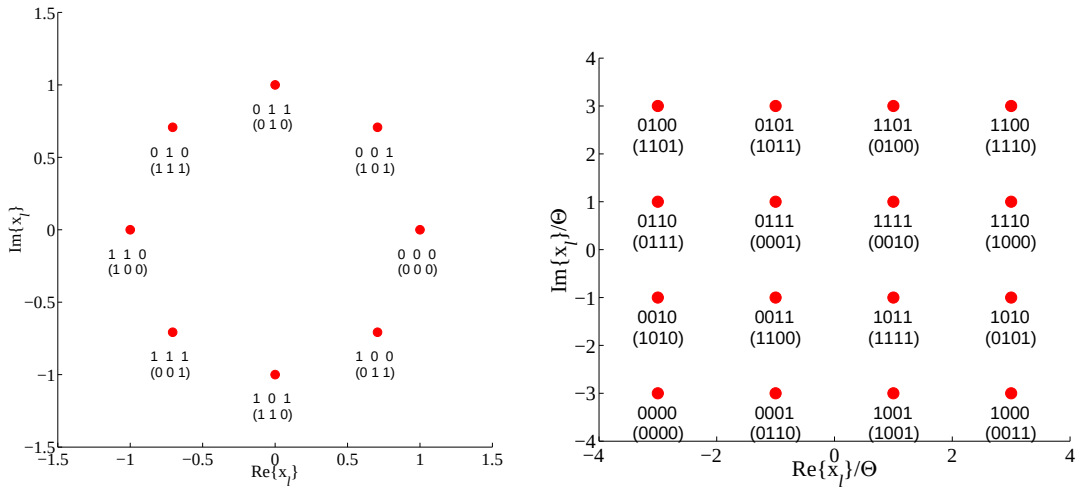
Definition 2.1 *A labeling function is called Gray labeling, if* $\forall m \in \{0,\ldots,M-1\}$ *and* $b \in \{0,1\}$, *each symbol* $x_{l_1} \in \mathbb{X}_b^m$ *has at most one symbol* $x_{l_2} \in \mathbb{X}_{\overline{b}}^m$ *at distance* d_{min} *[33].*

Even though this definition is not as intuitive as the one before, it can be shown that any mapping, for which symbols at distance $d_{\min}$ differ in more than one bit label, cannot fulfill Definition 2.1. The advantage of Gray labeling is that zero mean Gaussian noise most likely yields at the demapper erroneous symbols that are located at $d_{\min}$ from the transmitted symbols. In these cases, only one out of M bits is decided wrong. However, as we will see, this labeling strategy does hardly benefit from a priori knowledge that might be available at the demapper from a feedback loop from the channel decoder.

We denote as *anti-Gray* labeling *every* labeling that does not satisfy Definition 2.1. One main contribution of this thesis is to classify both Gray and anti-Gray mappings with respect to different figures of merit, such as SER, BER, achievable capacity, with or without a priori knowledge.

In Figure 2.7, we continue with two more examples of Gray mappings. Figure 2.7(a) depicts the symbols of an 8-PSK, i.e., $x_l = \mathrm{e}^{\mathrm{j}\frac{\pi}{4}l}, l \in \{0,\ldots,7\}$, and the corresponding bit labels $\left(c_{k,0}, c_{k,1}, c_{k,2}\right)$. Note that the random variables $\mathrm{Re}\{x_l\}$ and $\mathrm{Im}\{x_l\}$ are statistically dependent. This can easily be seen by considering $P\left[\mathrm{Re}\{x_l\} = 1 \,\middle|\, \mathrm{Im}\{x_l\} \neq 0\right] = 0 \neq P\left[\mathrm{Re}\{x_l\} = 1\right] = 1/8$. They are however uncorrelated, as shown in the previous subchapter, so that (2.27) and (2.32) are still valid. The BPSK, QPSK and 8-PSK constellations all have in common that $|x_l| = 1, \forall l$. The information is contained in the phase only. In Figure 2.7(b), the symbols of a 16-QAM, belonging to labels $\left(c_{k,0}, c_{k,1}, c_{k,2}, c_{k,3}\right)$, have different magnitudes and phases. Any QAM mapping can be regarded as combined ASK/PSK. We will thus use the notation of QAM as the overarching term. Both axes in Figure 2.7(b) are normalized by $\Theta = 10^{-1/2}$ to satisfy $\mathscr{P}_x = 1$.

The bit labelings given in brackets in Figure 2.7 correspond to anti-Gray mappings, which are optimized for the AWGN channel to exploit perfect a priori information most effectively [23], see also Section 3.5. ■



(a) $M = 3$, 8-PSK mapping

(b) $M = 4$, 16-QAM mapping, $\Theta = 10^{-1/2}$

Figure 2.7: Gray mappings for $M = 3, 4$. In brackets: anti-Gray bit labels [23]

2.2 Channel Models

We focus in this thesis only on frequency flat, i.e., memoryless, channel models. We do so to avoid the following considerations: frequency selectivity would result in ISI and an equalizer should be employed. However, the resulting noise samples would not be orthogonal as in (2.24) anymore. One has to apply an appropriate whitening filter [30]. As an alternative, one could divide the bandwidth in smaller fractions, each of which being approximately frequency flat. This is the case for the orthogonal frequency division multiplex (OFDM). Our analysis can then be applied to each subcarrier of the OFDM signal. All these details shall be neglected here. Moreover, we assume that the receiver has perfect knowledge of the channel, whereas the transmitter has no knowledge of it. We include the additive noise in our notation of a channel. The overall channel can be described by the probability density functions (pdfs) $p\left(r_k \,|s_k = x_l\right), \forall x_l \in \mathbb{X}$, which are also called likelihood or transition pdfs.

2.2.1 Additive White Gaussian Noise Channel

The non-fading[3] AWGN channel models the impairment that arises from additive noise. The noise sources may be manifold — random fluctuations of the wave in free space communications, or at the receiver side thermal noise that is omnipresent in electrical conductors or shot noise in semiconductors, such as amplifiers. The AWGN channel model is well suited for satellite-earth or deep space communications. We will make frequent use of it due to its rather simple mathematical description. The received sample is from (2.26)

$$r_k = s_k + n_k. \tag{2.35}$$

The noise samples $n_k = r_k - s_k$ are complex Gaussian distributed with zero mean and variance $\sigma_\mathrm{n}^2/2$ per dimension, so that the transition pdf is a bivariate Gaussian function

$$p\left(n_k = r_k - s_k \,|s_k = x_l\right) = p\left(r_k \,|s_k = x_l\right) = \left.\frac{\exp\left(-\frac{|r_k - s_k|^2}{\sigma_\mathrm{n}^2}\right)}{\pi\sigma_\mathrm{n}^2}\right|_{s_k = x_l}. \tag{2.36}$$

2.2.2 Rayleigh Fading Channel

In wireless transmission systems, such as mobile communications or terrestrial broadcasting, the electromagnetic wave travels typically over more than one path from transmitter to receiver. Multipath propagation occurs, where obstacles act as scatterers, reflectors or diffractors. The multiple received waves differ in amplitude and phase (due to temporal delays). At the receiver, the superposition of all incoming parts may lead to destructive interference. This effect is called (small-scale) *fading*. To integrate fading in the channel model, the transmit symbol in (2.26) is weighted by a complex factor[4], the magnitude of which corresponds to the attenuation (or possibly amplification, if larger than 1), while its phase is related to the delay. The central limit theorem [31] states that if the number or received waves goes to infinity, this complex number approximates a circular symmetric Gaussian random variable, denoted as $h_k = |h_k|\,\mathrm{e}^{\mathrm{j}\phi_k}$. This fading coefficient has zero mean and variance σ_h^2, i.e., variance $\sigma_\mathrm{h}^2/2$ per real dimension [34]. We set $\sigma_\mathrm{h}^2 = 1$ to model a passive channel that conserves transmit power on average. The SNR definitions (2.29) and (2.31) still apply in that case. Any real attenuation, such as free space path loss or attenuation through shadowing obstacles (also called large-scale fading), should be incorporated in the average transmit power, as discussed in Subsection 2.1.1. We now have the relation

$$r_k = h_k \cdot s_k + n_k, \tag{2.37}$$

with the corresponding system model depicted in Figure 2.8. We assume that h_k are i.i.d., which is the case for fast fading (or if interleavers are chosen appropriately long). Since the receiver has perfect knowledge of h_k, the Rayleigh channel can be interpreted as a scaled non-fading AWGN channel, with instantaneous noise power $\sigma_\mathrm{n}^2 \cdot \ |h_k|^{-2}$.

[3] We may drop the attribute "non-fading" to shorten notation.

[4] We assume that no Doppler effect occurs.

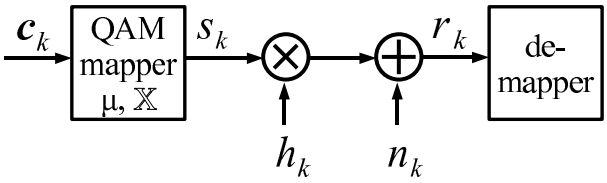


Figure 2.8: Equivalent baseband representation of discrete-time QAM over Rayleigh fading channel

The Rayleigh channel owes its name to the pdf of magnitude $|h_k|$. This is a Rayleigh distribution with variance 1, [30],

$$p\left(|h_k|\right) = \begin{cases} |h_k| \cdot \exp\left(-\frac{|h_k|^2}{2}\right) & , \quad \text{for } |h_k| > 0 \\ 0 & , \quad \text{else .} \end{cases} \tag{2.38}$$

The angle ϕ_k of the fading coefficient is uniformly distributed in $[0, 2\pi)$.

2.2.3 Multiple Input Multiple Output Channel

Up to now, we have focused on channels with one input and one output. The transmission rate can only be increased by selecting schemes with high spectral efficiency. For each channel, however, there exists a fundamental limit for the spectral efficiency, as will be discussed in Section 2.5. Fading channels in particular have lower limits than non-fading channel. Multipath propagation was long seen as a difficulty for wireless transmission, until in 1996 Foschini came up with the idea of rather exploiting the fading channel [35]. Multiple antennas at both transmitter and receiver side can be utilized in a rich-scattering wireless channel to increase the throughput [36]. This is called multiple input multiple output (MIMO) transmission. The variants single input multiple output (SIMO) [37] and multiple input single output (MISO) [38, 39] were already known to the scientific community. For MIMO, several strategies exist (see [40, 41, 42] and references therein): beamforming at transmitter and interference suppression at receiver, exploitation of spatial diversity by space-time coding or spatial multiplexing. We consider only the latter approach, where each antenna transmits an independent symbol. Thus, the spectral efficiency increases linearly with the number of transmit antennas N_t, i.e.,

$$\eta = N_\mathrm{t} \cdot R_\mathrm{c} \cdot M, \tag{2.39}$$

if all antennas transmit symbols drawn from the same alphabet $\mathbb{X}$.

The MIMO system for spatial multiplexing is shown in Figure 2.9. In order to separate the transmitted symbols at the MIMO detector with a linear filter, the number of receive antennas, N_r, must not be smaller than the number of transmit antennas. Optimum MIMO detection, which allows in principle for arbitrary N_r, is discussed in Subsection 2.3.3.

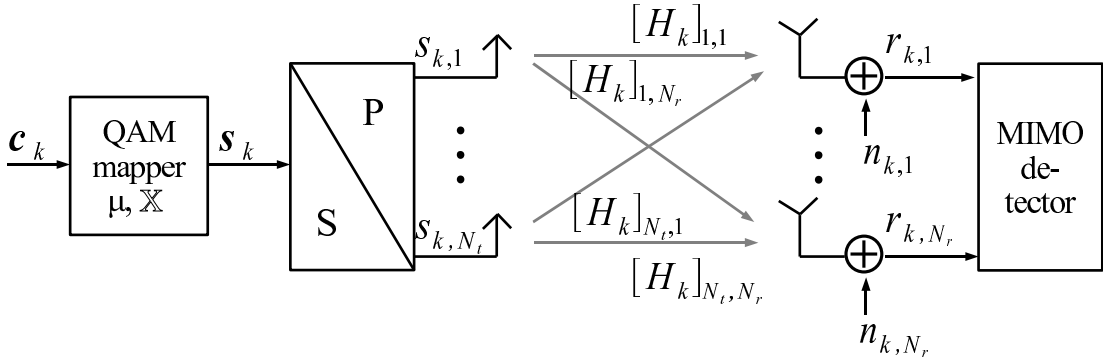


Figure 2.9: Spatial multiplexing over MIMO channel

The bit vector $\mathbf{c}_k$ now has a length of $N_\mathrm{t} \cdot M$ bits per signaling interval. It can be decomposed as $\mathbf{c}_k = \left(\mathbf{c}_{k,1}, \ldots, \mathbf{c}_{k,i}, \ldots, \mathbf{c}_{k,N_\mathrm{t}}\right)$, where each component is itself a vector, $\mathbf{c}_{k,i} = \left(c_{k,i,0}, \ldots, c_{k,i,M-1}\right)$. The QAM mapper operates on each $\mathbf{c}_{k,i}$ and yields symbol vector

$$\mathbf{s}_k = \left(s_{k,1}, \ldots, s_{k,i}, \ldots, s_{k,N_\mathrm{t}}\right) = \mu(\mathbf{c}_k) = \left(\mu(\mathbf{c}_{k,1}), \ldots, \mu(\mathbf{c}_{k,i}), \ldots, \mu(\mathbf{c}_{k,N_\mathrm{t}})\right). \tag{2.40}$$

In accordance with the majority of MIMO literature, we enumerate the symbols starting with 1. We differ, however, in so far as we maintain row vectors for consistency. After serial-to-parallel conversion, all N_t antennas transmit the vector $\mathbf{s}_k$ simultaneously at discrete-time k in the same frequency band. The i-th antenna transmits $s_{k,i}$, $i \in \{1, \ldots, N_\mathrm{t}\}$. This structure is also called vertical Bell Laboratories Layered Space-Time (V-BLAST), because the code words from $\mathbf{c}_k$ are distributed vertically over the transmit antennas in Figure 2.9, [43].

The fading coefficient from transmit antenna i to receive antenna j is denoted as $[H_k]_{i,j}$, $i \in \{1, \ldots, N_\mathrm{t}\}$, $j \in \{1, \ldots, N_\mathrm{r}\}$, and constitutes the i-th row, j-th column entry of channel matrix H_k. This matrix[5] has dimensions $N_\mathrm{t} \times N_\mathrm{r}$ and hence, we denote the antenna setup from Figure 2.9 as an $N_\mathrm{t} \times N_\mathrm{r}$ MIMO channel. We consider the case that all $[H_k]_{i,j}$ are i.i.d. with Rayleigh distribution from the previous subsection. This assumption is allowed, if the antenna arrays have a spacing of at least $\lambda_0/2$, where $\lambda_0 = \mathrm{c}/f_0$ is the wavelength of the carrier and c is the free space velocity of light. We then have $\mathsf{E}_{H_k}\left[H_k^\mathrm{H} H_k\right] = N_\mathrm{t} \cdot I_{N_\mathrm{r}}$, where I_{N_r} and $(.)^\mathrm{H}$ denote the N_r-dimensional identity matrix and Hermitian operation, respectively.

We define for discrete-time k the receive vector as $\mathbf{r}_k = \left(r_{k,1}, \ldots, r_{k,N_\mathrm{r}}\right)$ and the noise vector as $\mathbf{n}_k = \left(n_{k,1}, \ldots, n_{k,N_\mathrm{r}}\right)$. At each receive antenna $j \in \{1, \ldots, N_\mathrm{r}\}$, the superposition of all transmitted symbols, weighted by the fading coefficients, is observed. In addition, the signal is impaired by white Gaussian noise $n_{k,j}$, yielding $r_{k,j}$. All elements in $\mathbf{n}_k$ are i.i.d. as discussed in Subsection 2.2.1, so that $\mathsf{E}_{\mathbf{n}_k}\left[\mathbf{n}_k^\mathrm{H} \mathbf{n}_k\right] = \sigma_\mathrm{n}^2 \cdot I_{N_\mathrm{r}}$. In matrix notation, we now have

$$\mathbf{r}_k = \mathbf{s}_k \cdot H_k + \mathbf{n}_k. \tag{2.41}$$

[5] In this thesis, matrices are given in capital letters.

The transition pdf is the multivariate Gaussian function

$$p\left(\mathbf{r}_k \,|\mathbf{s}_k\right) = \frac{\exp\left(-\frac{\left|\mathbf{r}_k - \mathbf{s}_k \cdot H_k\right|^2}{\sigma_\mathrm{n}^2}\right)}{\left(\pi\sigma_\mathrm{n}^2\right)^{N_\mathrm{r}}}. \tag{2.42}$$

The average received power is N_r times the transmitted power, because the channel is passive. Each transmit antenna radiates on average $\frac{\mathscr{P}_x}{N_\mathrm{t}}$, such that the overall transmit power equals $\mathscr{P}_x$. The overall average receive power is $N_\mathrm{r} \cdot \mathscr{P}_x$. To adjust the SNR definitions of $\frac{E_\mathrm{b}}{N_0}$ to MIMO settings, we insert this received power and (2.39) in (2.31), such that

$$\left.\frac{E_\mathrm{b}}{N_0}\right|_\mathrm{dB} = 10 \cdot \log_{10}\left(\frac{N_\mathrm{r} \cdot \mathscr{P}_x}{N_0 \cdot N_\mathrm{t} \cdot R_\mathrm{c} \cdot M}\right). \tag{2.43}$$

This is the SNR at the total receiver as used e.g. in [44, 45], and not at one receive antenna.

Finally, we remark that MIMO principles are not restricted to wireless communications via multiple antennas. In powerline communications (PLC), the different phase wires may constitute MIMO systems up to order 2×4 [46]. Other MIMO applications are the reduction of crosstalk in cable bundles for DSL [47] or equalization in optical transmission systems by exploitation of different wave modes in a multi-mode fiber [48].

2.2.4 Binary Erasure Channel

All previous channels can be described as discrete input, continuous output channels. The transmit symbols are drawn from a finite set $\mathbb{X}$ or more general $\{\mathbb{X}\}^{N_\mathrm{t}}$, which includes spatial multiplexing MIMO systems. The output is continuous-valued due to the addition of noise. On a more abstract level, we now consider the binary erasure channel (BEC) from Figure 2.10, where both input c_k and output y_k are discrete. Transmitted bits c_k are either received correctly with probability $1-q$ or the receiver declares an erasure, if detection is too unreliable or impossible for other reasons. This event, denoted as ε, occurs with erasure probability q. Thus, the input is binary, while the output is ternary. Obviously, no transmission errors are possible, only erasures. The BEC is often used to model the channel of magnetic recording systems, and we will make use of its mathematical simplicity in Chapter 5.

2.3 Error Control Coding

All channels from the previous section distort the transmitted bit sequence, c_k. The resulting BER might be unacceptable for reliable data communications. Error control coding (ECC) techniques try to mitigate this degradation, by inserting redundancy in a controlled manner. As this new encoded sequence should be adapted to the channel impairments, this technique

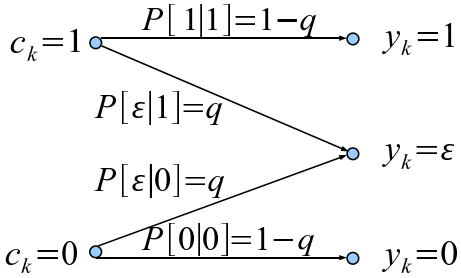


Figure 2.10: BEC model

is also called channel coding. In this section, we summarize all principles that are necessary for the further understanding. We start with (binary) block and convolutional codes, then introduce the so called log likelihood ratio (LLR) or in short L-value, and end with concatenated coding schemes. For more details on ECC, the reader is referred to textbooks as [49, 50, 51, 52].

2.3.1 Block Codes

A block encoder maps an information word $\mathbf{b}_\kappa = \left(b_{\kappa,0},\ldots,b_{\kappa,K-1}\right)$ of length K to a code word $\mathbf{c}_\kappa = \left(c_{\kappa,0},\ldots,c_{\kappa,N-1}\right)$ of length $N \geq K$. Discrete-time index κ is not necessarily synchronous with discrete-time k from QAM signaling. As the encoder operates block-wise, it is memoryless. A linear block code can be described by its generator matrix G of dimensions $K \times N$. Encoding is given by

$$\mathbf{c}_\kappa = \mathbf{b}_\kappa \cdot G, \tag{2.44}$$

where operations as additions and multiplications are performed in Galois field GF(2), i.e., modulo-2 is appended to conventional operations. We will imply GF(2) operations whenever binary en-/decoding is described. The set of code words constitutes the code $\mathscr{C}$. Thus, $\mathscr{C}$ is the K-dimensional subspace in the N-dimensional space with binary entries, $b_{\kappa,i}, c_{\kappa,j} \in \{0,1\}$, $i \in \{0,\ldots,K-1\}$, $j \in \{0,\ldots,N-1\}$. For bijective encoding, the generator matrix G has to be of rank K. The ratio $R_\mathrm{c} = \frac{K}{N} \leq 1$ is called code rate and is the fraction of information that is contained in a code word.

To each generator matrix G of rank K, there exists a parity check matrix P of dimensions $N-K \times N$, such that

$$G \cdot P^\mathrm{T} = 0_{K\times N-K}, \text{ with } P \neq 0_{N-K\times N}, \tag{2.45}$$

in which $0_{N_1\times N_2}$ is the all-zero matrix of dimensions $N_1 \times N_2$ and $(.)^\mathrm{T}$ denotes the transposition. The second condition in (2.45) excludes the trivial solution. Right-multiplying (2.44) with P^T and inserting (2.45), we obtain $N-K$ scalar parity check equations, that have to be fulfilled by any valid code word $\mathbf{c}_\kappa$. In matrix notation, we have

$$\mathbf{c}_\kappa \cdot P^\mathrm{T} = \mathbf{b}_\kappa \cdot G \cdot P^\mathrm{T} = 0_{1\times N-K}. \tag{2.46}$$

The majority of block codes belongs to the class of algebraic codes that are typically decoded by hard decisions. One example of hard decision decoding is based on the fact, that bit errors imposed on $\mathbf{c}_\kappa$ can lead to a vector that is not included in $\mathscr{C}$. Parity check (2.46) will then give a non-zero row vector, called *syndrome*. Each syndrome belongs to a most likely error pattern, which is added on the received word and will cancel out the true error pattern if both are identical.

In 1963, Gallager proposed in his Ph.D. thesis [53] the LDPC codes, which belong to the class of codes on graphs. The idea is to use very large code words length N and let only a small fraction of code bits be involved in a parity check. From (2.46), it follows that the parity check matrix P has to be of low density, i.e., the number of 1's, must be relatively small. Gallager proved that LDPC codes are asymptotically good, if N increases, provided a random distribution of (few) 1's in P, which is in accordance to Shannon's random coding theorem [4]. However, optimum decoding of LDPC codes is hardly feasible. Suboptimum decoding algorithms as iterative message passing with the sum-product algorithm [54], which is based on the Tanner graph representation [55] of the code, were proposed after the discovery of the turbo codes, cf. Subsection 2.3.4.

2.3.2 Convolutional Codes

In contrast to block codes, convolutional codes [56] can operate in principle on an infinite length of uncoded bits $b_\kappa, \kappa \in \mathbb{Z}$. Encoding is performed with a shift register with ν delay elements, i.e., convolutional encoders introduce memory. Thus, long code sequences can be generated with small encoding complexity. If each bit b_κ yields N_c code bits $\mathbf{c}_\kappa = \left(c_{\kappa,0}, \ldots, c_{\kappa,n}, \ldots, c_{\kappa,N_\mathrm{c}-1}\right)$, the code rate is $R_\mathrm{c} = N_\mathrm{c}^{-1}$. In general, any number K_c of input bits can generate any number $N_\mathrm{c} \geq K_\mathrm{c}$ of output bits, allowing R_c to be any rational number in $[0,1]$, but this will not be treated here. Encoding of the n-th output bit $c_{\kappa,n}$ is depicted in Figure 2.11.

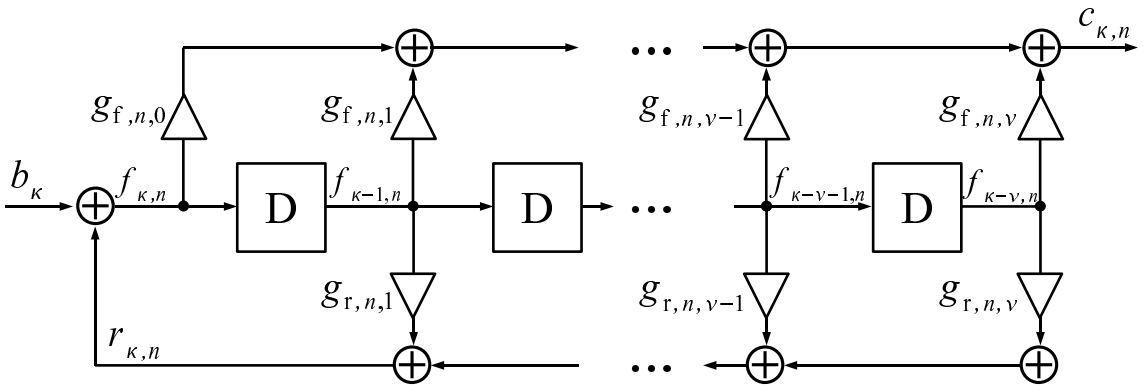


Figure 2.11: Shift register for n-th output bit of a convolutional encoder

Let us define for the n-th output the feedforward and feedback generator polynomials as $g_{\mathrm{f},n}(\mathrm{D}) = \sum_{i=0}^{\nu} g_{\mathrm{f},n,i} \cdot \mathrm{D}^i$ and $g_{\mathrm{r},n}(\mathrm{D}) = 1 + \sum_{i=1}^{\nu} g_{\mathrm{r},n,i} \cdot \mathrm{D}^i$, respectively. The indeterminate D

represents the delay operator and the coefficients $g_{\mathrm{f},n,i}, g_{\mathrm{r},n,i} \in \{0,1\}$ describe, if a connection to the modulo-2 adders exists or not, cf. Figure 2.11. The feedforward part yields code bit $c_{\kappa,n}$ as

$$c_{\kappa,n} = \sum_{i=0}^{\nu} f_{\kappa-i,n} \cdot g_{\mathrm{f},n,i}, \tag{2.47}$$

in which the intermediate feedforward bits $f_{\kappa,n}$ are computed by

$$f_{\kappa,n} = b_\kappa + r_{\kappa,n}, \tag{2.48}$$

and the intermediate feedback bits $r_{\kappa,n}$ are given by the recursive (thus the index "r") relation

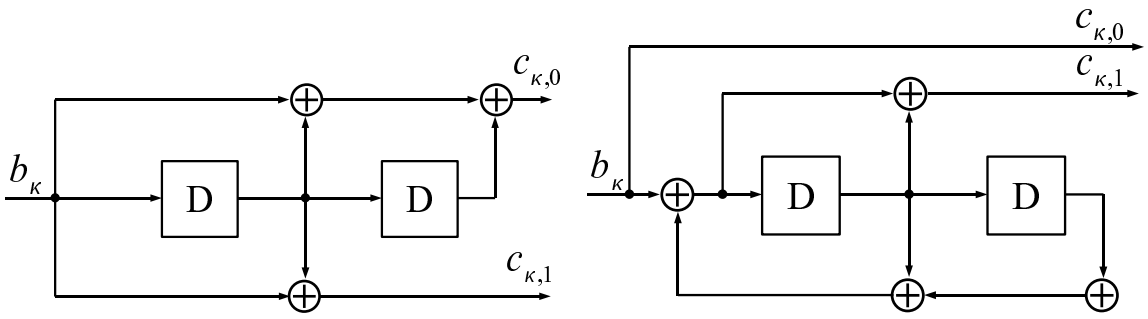
$$r_{\kappa,n} = \sum_{i=1}^{\nu} f_{\kappa-i,n} \cdot g_{\mathrm{r},n,i} \overset{(2.48)}{=} \sum_{i=1}^{\nu} \left(b_{\kappa-i} + r_{\kappa-i,n}\right) \cdot g_{\mathrm{r},n,i}. \tag{2.49}$$

Equations (2.47) and (2.49) can be interpreted as the convolution of intermediate bits $f_{\kappa-i,n}$ and $r_{\kappa-i,n}$, respectively, with a sequence of generator polynomial coefficients, justifying the name convolutional encoder. The fraction $g_{\mathrm{f},n}(\mathrm{D})/g_{\mathrm{r},n}(\mathrm{D})$ is the z-domain transfer function from input b_κ to output $c_{\kappa,n}$, if we substitute z^{-1} for D. Thus, a rate N_c^{-1} convolutional code is fully described by its transfer functions $\left(g_{\mathrm{f},0}(\mathrm{D})/g_{\mathrm{r},0}(\mathrm{D}), \ldots, g_{\mathrm{f},N_\mathrm{c}-1}(\mathrm{D})/g_{\mathrm{r},N_\mathrm{c}-1}(\mathrm{D})\right)$. Obviously, the same set of code words, i.e., the same code $\mathscr{C}$, is obtained, if we multiply each transfer function $g_{\mathrm{f},n}(\mathrm{D})/g_{\mathrm{r},n}(\mathrm{D}), n \in \{0,\ldots,N_\mathrm{c}-1\}$, with $g_{\mathrm{r},0}(\mathrm{D})/g_{\mathrm{f},0}(\mathrm{D})$, [57, 58]. The transfer functions are then $\left(1, \ldots, \left(g_{\mathrm{f},N_\mathrm{c}-1}(\mathrm{D}) \cdot g_{\mathrm{r},0}(\mathrm{D})\right) \Big/ \left(g_{\mathrm{r},N_\mathrm{c}-1}(\mathrm{D}) \cdot g_{\mathrm{f},0}(\mathrm{D})\right)\right)$. As a consequence, this code realization is systematic, i.e., the first code bit $c_{\kappa,0}$ equals the uncoded bit b_κ. However, the assignment from information words to code words differs. In analogy with QAM mappings, the code $\mathscr{C}$ defines the alphabet (set of all code words), while the code realization corresponds to the labeling of code words. From that we can conclude that if the decoder does not obtain a priori information of the information bits b_κ, then a non-systematic (block or convolutional) code performs identical as its systematic equivalent.

In this thesis, we apply convolutional codes as outer encoders in a serial concatenated coding scheme. At the receiver side, the respective decoder obtains information of coded bits $\mathbf{c}_\kappa$. Hence, it makes no difference, if a non-systematic or systematic code is employed. Without loss of generality, we restrict ourselves to systematic convolutional codes of rate $R_\mathrm{c} = \frac{1}{2}$. Extensions to other code rates are straightforward. Denoting the transfer function for the second code bit as $g_\mathrm{f}(\mathrm{D})/g_\mathrm{r}(\mathrm{D})$, the overall code is either described by $\left(1, g_\mathrm{f}(\mathrm{D})/g_\mathrm{r}(\mathrm{D})\right)$ or $\left(g_\mathrm{r}(\mathrm{D}), g_\mathrm{f}(\mathrm{D})\right)$. We use the latter notation and give the generator polynomials as octal numbers. This is achieved by multiplying the generator polynomial by 2^ν, substituting 2^{-1} for D and changing the result to the octal format.

Example: Let us consider the $(07,06)$ code. The preceding "0" indicates octal notation. It can be readily seen that $\nu = 2$, because $\max(6,7) < 2^{2+1}$, and further $g_\mathrm{r}(\mathrm{D}) = 1+\mathrm{D}+\mathrm{D}^2$, $g_\mathrm{f}(\mathrm{D}) = 1+\mathrm{D}$, because $2^2 \cdot \left(1+1\cdot 2^{-1}+1\cdot 2^{-2}\right) = 7$ and $2^2 \cdot \left(1+1\cdot 2^{-1}\right) = 6$. The non-systematic realization of this encoder is depicted in Figure 2.12(a) and its systematic equivalent $\left(1, 06/07\right)$ in Figure 2.12(b). For the latter, a recursive part is needed. Note that the

transfer function for the systematic bit is 1 and would be 04 in octal notation. ■



(a) Non-systematic $(07,06)$ realization

(b) Systematic $\left(1,06/07\right)$ realization

Figure 2.12: Example of convolutional encoder $(07,06)$

For real applications of convolutional codes, especially if employed as a constituent code in a concatenated coding scheme, the input bit sequence has to have finite lengths, say K. To ease decoding, initial and final states[6] of the encoder are usually forced to the all-zero states. This means that all memory elements store bit 0, if a new sequence of information bits is to be encoded, and that after this sequence, $\nu \cdot N_\mathrm{c}$ "tail bits" are appended, such that each of the N_c shift registers is forced to the all-zero state again (trellis termination, [58]). A convolutional code of rate $R_\mathrm{c} = N_\mathrm{c}^{-1}$ computes for K information bits $N = K \cdot N_\mathrm{c}$ code bits. Due to the tail bits, the true code rate is

$$\tilde{R_\mathrm{c}} = \frac{K}{K \cdot N_\mathrm{c} + \nu \cdot N_\mathrm{c}} \approx N_\mathrm{c}^{-1} = R_\mathrm{c}, \tag{2.50}$$

where the approximation holds for large K, which is typically the case. We will thus neglect this small variation in code rate.

It is worth to mention that a terminated convolutional code can be described as a block code. Following up the previous example of the rate $R_\mathrm{c} = \frac{1}{2}$ code, we can easily see that the non-systematic $(07,06)$ encoder in Figure 2.12(a) produces the two impulse responses, if the input is a unit impulse, $b_\kappa = \delta_\kappa$. They are given by inverse z-transform of the transfer functions $1 + z^{-1} + z^{-2}$ and $1 + z^{-1}$, respectively. The upper branch produces $c_{\kappa,0} = 1$ for $\kappa \in \{0,1,2\}$ and 0 else, while the lower branch yields $c_{\kappa,1} = 1$ for $\kappa \in \{0,1\}$ and 0 else. The overall impulse response $\mathbf{h}_\kappa = \left(c_{\kappa,0}, c_{\kappa,1}\right)$ for $\kappa \in \{0,1,2\}$ in vector notation is of the form $(11\,11\,10)$. If the initial state is the all-zero state, then any input sequence $\mathbf{b}_\kappa$ of length K yields the coded sequence (without tail bits) of length $2K$ through

$$\mathbf{c}_\kappa = \mathbf{b}_\kappa \cdot G, \tag{2.51}$$

[6] Each of the 2^ν states of a shift register is defined by the entries of its memory elements.

where the $K \times 2K$ generator matrix is of the form

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & \cdots \\ \ddots & & \ddots & & \ddots & & \ddots & & \ddots & & \ddots \end{pmatrix}. \tag{2.52}$$

For an arbitrary convolutional code, the generator matrix can be constructed by the following rule: the k-th row of G, with $k \in \{0, \ldots, K-1\}$, starts with $k \cdot N_\mathrm{c}$ zeros, followed by impulse response $\mathbf{h}_\kappa$, and the rest is filled up with zeros, if no recursive part is involved (otherwise, $\mathbf{h}_\kappa$ is of infinite length).

2.3.3 Optimum Decoding

2.3.3.1 A posteriori probabilities

The task of the channel code is to introduce both redundancy in and dependencies among a code word. Also the QAM mapper from Subsection 2.1.3 combines M bits in $\mathbf{c}_k$ and outputs a symbol s_k. This corresponds to a rate 1 code, with no additional redundancy, however with dependencies among the bits that belong to s_k. We now review the optimum decoding rule for channel codes and include in this discussion optimum demapping as an instance of channel decoding. As a matter of fact, we focus on optimum demapping, because extensions to rate $R_\mathrm{c} < 1$ codes are straightforward.

Figure 2.13 shows a QAM demapper, which receives channel observation r_k as a result of transmit symbol $s_k = \mu(\mathbf{c}_k)$. Remember that the receiver has perfect channel knowledge, thus it knows the transition pdfs $p(r_k | s_k = x_l)$. The second input to the demapper consists of *a priori* probabilities of the bits $\mathbf{c}_k$. To be precise, these probabilities concern the bit estimates of the demapper $\hat{\mathbf{c}}_k$, not the actual transmitted bits $\mathbf{c}_k$, as these are deterministic, after being sent. This information is available in iterative receivers, where a feedback loop from a channel decoder back to the demapper exists, cf. Section 2.4. Let us define the a priori probability of the m-th bit in $\hat{\mathbf{c}}_k$ as

$$P_{\mathrm{A},k,m} = P\left[\hat{c}_{k,m} = 1\right] = 1 - P\left[\hat{c}_{k,m} = 0\right], \tag{2.53}$$

and the vector of these probabilities, which belong to one symbol s_k, as

$$\mathbf{P}_{\mathrm{A},k} = \left(P_{\mathrm{A},k,0}, \ldots, P_{\mathrm{A},k,M\text{-}1}\right). \tag{2.54}$$

The decoding rule, which is optimum in the sense that it minimizes the number of bit errors in $\hat{c}_{k,m}$, is called maximum a posteriori (MAP) [30]. For that, the *a posteriori probability* (APP) is computed for $\hat{c}_{k,m} = b, b \in \{0,1\}$, based on all available information and the decision is made in favor of bit b, which maximizes the APP. The available information for decoding

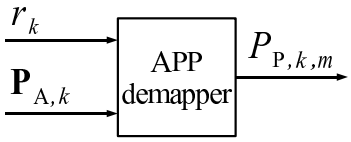


Figure 2.13: APP demapper with channel observations and a priori input

bit $\hat{c}_{k,m}$ is the received sample r_k and the a priori probabilities $\mathbf{P}_{\mathrm{A},k}$. We denote the APP for bit $\hat{c}_{k,m}$ as $P_{\mathrm{P},k,m}$ and compute it as

$$\begin{aligned}
P_{\mathrm{P},k,m} = \;& P\left[\hat{c}_{k,m}=1\,\middle|\,r_k,\mathbf{P}_{\mathrm{A},k}\right] = 1-P\left[\hat{c}_{k,m}=0\,\middle|\,r_k,\mathbf{P}_{\mathrm{A},k}\right] \\
\overset{(1)}{=}\;& \frac{p\left(\hat{c}_{k,m}=1,r_k,\mathbf{P}_{\mathrm{A},k}\right)}{p\left(r_k,\mathbf{P}_{\mathrm{A},k}\right)} \\
\overset{(2)}{=}\;& \frac{p\left(\mathbf{P}_{\mathrm{A},k}\right)\cdot p\left(\hat{c}_{k,m}=1,r_k\,\middle|\,\mathbf{P}_{\mathrm{A},k}\right)}{p\left(\mathbf{P}_{\mathrm{A},k}\right)\cdot p\left(r_k\right)} \\
\overset{(3)}{=}\;& \frac{1}{p\left(r_k\right)}\cdot\sum_{\hat{x}\in\mathbb{X}_1^m} p\left(s_k=\hat{x},r_k\,\middle|\,\mathbf{P}_{\mathrm{A},k}\right) \\
\overset{(4)}{=}\;& \frac{1}{p\left(r_k\right)}\cdot\sum_{\hat{x}\in\mathbb{X}_1^m} P\left[s_k=\hat{x}\,\middle|\,\mathbf{P}_{\mathrm{A},k}\right]\cdot p\left(r_k\,\middle|\,s_k=\hat{x},\mathbf{P}_{\mathrm{A},k}\right) \\
\overset{(5)}{=}\;& \frac{1}{p\left(r_k\right)}\cdot\sum_{\hat{x}\in\mathbb{X}_1^m}\prod_{i=0}^{M-1} P\left[\hat{c}_{k,i}=\mu_i^{-1}\left(\hat{x}\right)\,\middle|\,\mathbf{P}_{\mathrm{A},k}\right]\cdot p\left(r_k\,|s_k=\hat{x}\right) \\
\overset{(6)}{=}\;& \frac{P\left[\hat{c}_{k,m}=1\right]}{p\left(r_k\right)}\cdot\sum_{\hat{x}\in\mathbb{X}_1^m}\prod_{\substack{i=0\\ i\neq m}}^{M-1} P\left[\hat{c}_{k,i}=\mu_i^{-1}\left(\hat{x}\right)\right]\cdot p\left(r_k\,|s_k=\hat{x}\right)
\end{aligned} \tag{2.55}$$

In step (1), Bayes' rule was applied, in (2), the pdfs were factorized, and the fact was used that a priori knowledge $\mathbf{P}_{\mathrm{A},k}$ is independent[7] of the channel observation r_k. After canceling pdf $p\left(\mathbf{P}_{\mathrm{A},k}\right)$, we apply in (3) the law of total probability and sum over all hypothetical transmit symbols $\hat{x}$, for which the m-th bit label equals 1. (4) is simply a factorization of a joint pdf that includes a condition. In the second factor, we can neglect $\mathbf{P}_{\mathrm{A},k}$, because $s_k=\hat{x}$ is already a sufficient condition. The probability that $\hat{x}$ was transmitted is given by the product in (5). Note that each factor in this product is determined by a priori probability $P_{\mathrm{A},k,\mathrm{i}}$ from (2.53), i.e., $P\left[\hat{c}_{k,i}=\mu_i^{-1}\left(\hat{x}\right)\,\middle|\,\mathbf{P}_{\mathrm{A},k}\right]=P_{\mathrm{A},k,\mathrm{i}}$, if $\mu_i^{-1}\left(\hat{x}\right)=1$, else $1-P_{\mathrm{A},k,\mathrm{i}}$. We thus drop the condition of $\mathbf{P}_{\mathrm{A},k}$ in step (6), where we have also factored out $P\left[\hat{c}_{k,m}=1\right]$.

[7] Independence should be guaranteed by an interleaver between channel code and mapper, cf. Section 2.4.

The MAP rule is now to choose

$$\hat{c}_{k,m} = \arg \max_{b \in \{0,1\}} P\left[\hat{c}_{k,m} = b \,\middle|\, r_k, \mathbf{P}_{\mathrm{A},k}\right]. \tag{2.56}$$

For the maximization, the factor $p(r_k)$ is irrelevant and can be neglected. If no a priori knowledge is available, then MAP reduces to maximum likelihood (ML) decoding, which in turn reduces for an AWGN channel to the nearest Euclidean neighbor search.

Remarks on APP decoding: A generic APP decoder is presented in [59]. The APP computation as in (2.55) can be executed with manageable complexity for short block codes or as in our setting, for demappers. APP decoders for convolutional codes can exploit the trellis structure of the code. The corresponding algorithm was named BCJR by the initials of its discoverers [60, 61]. A suboptimum soft decoding for convolutional codes is the soft output Viterbi algorithm (SOVA) from [62]. For long block code words as in case of LDPC codes, an iterative approximation of the true APPs yields close to optimum performance at a much smaller decoding complexity. The best approximation of the APP is achieved by the message passing algorithm [54], which is also called sum-product algorithm due to its twofold computation in the log-domain. Wiberg revealed in his Ph.D. thesis [63] that the application of the message passing algorithm on concatenated (turbo) codes is exactly the same as the iterative exchange of extrinsic information between the component decoders, cf. Subsection 2.3.4, [64]. Shortly after, both approaches were identified as instances of Pearl's belief propagation on Bayesian networks [65].

2.3.3.2 L-value notation

Decoders, which output bit estimates $\hat{c}_{k,m} = b, b \in \{0,1\}$, as in (2.56), are called hard decision decoders. If, however, the output is a metric that includes the information of $P_{\mathrm{P},k,m}$ from (2.55), we denote this as soft decoding. To ease mathematical computations and increase numerical stability, L-values [66] are preferred as metrics.

Defining the a priori L-value of bit $\hat{c}_{k,m}$ as

$$L_{\mathrm{A},k,\mathrm{m}} = \ln\left(\frac{P\left[\hat{c}_{k,m}=1\right]}{P\left[\hat{c}_{k,m}=0\right]}\right) \overset{(2.53)}{=} \ln\left(\frac{P_{\mathrm{A},k,\mathrm{m}}}{1-P_{\mathrm{A},k,\mathrm{m}}}\right), \tag{2.57}$$

we can inversely compute the a priori probability that $\hat{c}_{k,m} = b, b \in \{0,1\}$, as

$$P\left[\hat{c}_{k,m}=b\right] = \left.\frac{\exp\left(\left(2\hat{c}_{k,m}-1\right)\cdot L_{\mathrm{A},k,\mathrm{m}}/2\right)}{\exp\left(L_{\mathrm{A},k,\mathrm{m}}/2\right)+\exp\left(-L_{\mathrm{A},k,\mathrm{m}}/2\right)}\right|_{\hat{c}_{k,m}=b}. \tag{2.58}$$

The vector of all a priori L-values that belong to one symbol s_k shall be

$$\mathbf{L}_{\mathrm{A},k} = \left(L_{\mathrm{A},k,0}, \ldots, L_{\mathrm{A},k,M\text{-}1}\right). \tag{2.59}$$

We further define in general a vector $\mathbf{x}_{[n]}$ as the vector $\mathbf{x}$ without the n-th entry. Similar to (2.57), we define the a posteriori L-value of $\hat{c}_{k,m}$ as

$$L_{\mathrm{P},k,\mathrm{m}} = \ln\left(\frac{P\left[\hat{c}_{k,m}=1\,\middle|\,r_k,\mathbf{P}_{\mathrm{A},k}\right]}{P\left[\hat{c}_{k,m}=0\,\middle|\,r_k,\mathbf{P}_{\mathrm{A},k}\right]}\right) \overset{(2.55)}{=} \ln\left(\frac{P_{\mathrm{P},k,m}}{1-P_{\mathrm{P},k,m}}.\right) \tag{2.60}$$

Inserting the APP from step (6) of (2.55), we can cancel out $p(r_k)$ and separate with (2.57) the a priori L-value of bit $\hat{c}_{k,m}$

$$L_{\mathrm{P},k,\mathrm{m}} = L_{\mathrm{A},k,\mathrm{m}} + \ln\left(\frac{\sum\limits_{\hat{x}\in\mathbb{X}_1^m}\prod\limits_{\substack{i=0\\i\neq m}}^{M-1} P\left[\hat{c}_{k,i}=\mu_i^{-1}(\hat{x})\right]\cdot p\left(r_k\,|s_k=\hat{x}\right)}{\sum\limits_{\hat{x}'\in\mathbb{X}_0^m}\prod\limits_{\substack{i=0\\i\neq m}}^{M-1} P\left[\hat{c}'_{k,i}=\mu_i^{-1}(\hat{x}')\right]\cdot p\left(r_k\,|s_k=\hat{x}'\right)}\right) \tag{2.61}$$

In the denominator, we have marked the hypothetical symbols and bits by a prime to distinguish them from the numerator. Now we substitute (2.58) for the a priori probabilities and expand the fraction by the factor

$$\prod_{\substack{i=0\\i\neq m}}^{M-1}\left(\exp\left(L_{\mathrm{A},k,\mathrm{i}}/2\right)+\exp\left(-L_{\mathrm{A},k,\mathrm{i}}/2\right)\right)\cdot\exp\left(L_{\mathrm{A},k,\mathrm{i}}/2\right) = \prod_{\substack{i=0\\i\neq m}}^{M-1}\left(\exp\left(L_{\mathrm{A},k,\mathrm{i}}\right)+1\right), \tag{2.62}$$

which is independent of the choice for $\hat{x},\hat{x}'$. This finally yields

$$\begin{aligned} L_{\mathrm{P},k,\mathrm{m}} &= L_{\mathrm{A},k,\mathrm{m}} + \ln\left(\frac{\sum\limits_{\hat{x}\in\mathbb{X}_1^m}\prod\limits_{\substack{i=0\\i\neq m}}^{M-1}\exp\left(\hat{c}_{k,i}\cdot L_{\mathrm{A},k,\mathrm{i}}\right)\cdot p\left(r_k\,|s_k=\hat{x}\right)\Big|_{\hat{\mathbf{c}}_k=\mu^{-1}(\hat{x})}}{\sum\limits_{\hat{x}'\in\mathbb{X}_0^m}\prod\limits_{\substack{i=0\\i\neq m}}^{M-1}\exp\left(\hat{c}'_{k,i}\cdot L_{\mathrm{A},k,\mathrm{i}}\right)\cdot p\left(r_k\,|s_k=\hat{x}'\right)\Big|_{\hat{\mathbf{c}}'_k=\mu^{-1}(\hat{x}')}}\right) \\ &= L_{\mathrm{A},k,\mathrm{m}} + L_{\mathrm{E},k,\mathrm{m}}, \end{aligned} \tag{2.63}$$

where the *extrinsic* L-value of bit $\hat{c}_{k,m}$ is after one more manipulation

$$L_{\mathrm{E},k,\mathrm{m}} = \ln\left(\frac{\sum\limits_{\hat{x}\in\mathbb{X}_1^m}\exp\left(\hat{\mathbf{c}}_{k[m]}\cdot\mathbf{L}_{\mathrm{A},k[m]}^{\mathrm{T}}\right)\cdot p\left(r_k\,|s_k=\hat{x}\right)\Big|_{\hat{\mathbf{c}}_k=\mu^{-1}(\hat{x})}}{\sum\limits_{\hat{x}'\in\mathbb{X}_0^m}\exp\left(\hat{\mathbf{c}}'_{k[m]}\cdot\mathbf{L}_{\mathrm{A},k[m]}^{\mathrm{T}}\right)\cdot p\left(r_k\,|s_k=\hat{x}'\right)\Big|_{\hat{\mathbf{c}}'_k=\mu^{-1}(\hat{x}')}}\right). \tag{2.64}$$

Similar to (2.59), we combine all APP and extrinsic L-values that belong to one symbol s_k in the vectors $\mathbf{L}_{\mathrm{P},k}$ and $\mathbf{L}_{\mathrm{E},k}$, respectively.

It is easily seen that the sign of any L-value allows for a hard decision, i.e., $\hat{c}_{k,m} = 1$, if the L-value is positive and 0 else. Furthermore, the magnitude of the L-value reflects the reliability of this prediction. The extrinsic information from (2.64) includes all *new* information that is obtained by the code constraints (redundancy and interdependencies). To be precise, $L_{\mathrm{E},k,\mathrm{m}}$ in our example of an optimum demapper is the extrinsic *and* channel L-value, as it includes also channel information (sometimes referred to as *intrinsic* information).

For an $N_\mathrm{t} \times N_\mathrm{r}$ MIMO system, we have to introduce some modifications in our notation. Let $\mathbf{x}_l$ be an N_t-dimensional row vector from the set $\tilde{\mathbb{X}} = \{\mathbb{X}\}^{N_\mathrm{t}}$ and $\hat{\mathbf{x}}_l, \hat{\mathbf{x}}'_l$ estimates thereof. The mapping from bit vector $\mathbf{c}_k$ of length $N_\mathrm{t}M$ to $\mathbf{x}_l$ is given by $\mu(\mathbf{c}_k) = \mathbf{x}_l$. This is similar to (2.40), in which $\mathbf{s}_k$ corresponds to a certain realization at discrete-time k. We enumerate the m-th bit of symbol $s_{k,i}$, $m \in \{0,\ldots,M-1\}$, $i \in \{1,\ldots,N_\mathrm{t}\}$, by index $p = (i-1)M + m$. The subset $\tilde{\mathbb{X}}_b^p$ shall now comprise all symbol vectors, whose p-th bit equals b, where $p \in \{0,\ldots,N_\mathrm{t}M-1\}$. The extrinsic L-value of the p-th bit $c_{k,p}$ thus becomes

$$L_{\mathrm{E},k,p} = \ln\left(\frac{\displaystyle\sum_{\hat{\mathbf{x}}_l \in \tilde{\mathbb{X}}_1^p} \exp\left(\hat{\mathbf{c}}_{k[p]} \cdot \mathbf{L}_{\mathrm{A},k[p]}^{\mathrm{T}}\right) \cdot p\left(\mathbf{r}_k \,|\mathbf{s}_k = \hat{\mathbf{x}}_l\right)\Big|_{\hat{\mathbf{c}}_k = \mu^{-1}(\hat{\mathbf{x}}_l)}}{\displaystyle\sum_{\hat{\mathbf{x}}'_l \in \tilde{\mathbb{X}}_0^p} \exp\left(\hat{\mathbf{c}}'_{k[p]} \cdot \mathbf{L}_{\mathrm{A},k[p]}^{\mathrm{T}}\right) \cdot p\left(\mathbf{r}_k \,\middle|\mathbf{s}_k = \hat{\mathbf{x}}'_l\right)\Big|_{\hat{\mathbf{c}}'_k = \mu^{-1}(\hat{\mathbf{x}}'_l)}}\right). \tag{2.65}$$

The vectors $\mathbf{L}_{\mathrm{A},k}$, $\mathbf{L}_{\mathrm{P},k}$, and $\mathbf{L}_{\mathrm{E},k}$ of those L-values that belong to one transmit vector $\mathbf{s}_k$ are now of length $N_\mathrm{t}M$. We still have the relation $\mathbf{L}_{\mathrm{P},k} = \mathbf{L}_{\mathrm{A},k} + \mathbf{L}_{\mathrm{E},k}$. As the complexity of APP decoding grows exponentially with $N_\mathrm{t}M$, suboptimal decoding strategies exist. These are discussed e.g. in [40, 41, 42] and are outside the scope of this thesis.

2.3.4 Concatenated Codes — The Turbo Principle

It is well known from information theory that good channel codes should consist of long code words and have a rather random structure [4, 67]. However, long code word lengths hinder optimum decoding as its complexity becomes unmanageable, and random-like codes are difficult to generate. Convolutional and — to an even greater extend, algebraic block codes are inherently structured. One way out of this dilemma is to generate long codes by combining several short ones, which results in concatenated [68] or product codes. Instead of decoding the overall (long) code, the receiver tackles this by successive decoding of the component codes. For a long time, a serial concatenation of a Reed-Solomon (RS) code [69], followed by an interleaver (rearrangement of bit order) and a convolutional code were considered the most powerful of channel codes [70]. The receiver first applies the Viterbi algorithm [71, 72], which is a hard decision ML *sequence estimation* (MLSE). Residual bit errors typically occur in bursts, but are distributed more evenly after deinterleaving (inverse operation of interleaver). These errors might now be fixed by the RS decoder. Note that the introduction of an interleaver randomizes the overall code structure.

In 1993, Berrou et al. proposed the *turbo code* as a parallel concatenated code (PCC), consisting of convolutional encoders, which are separated by interleavers, and suggested an iterative decoding algorithm based on soft APP decoding [14]. The idea is to update each decoder by the new extrinsic information, gained from the other decoders. In a subsequent iteration, this information is interpreted as a priori information. With each iteration, the reliability of the processed probabilities increases, until convergence is achieved. The receiver can either try to diagnose convergence behavior by means of stopping criteria, see e.g. [73, 74], or it simply stops after a certain number of iterations. In any case, a hard decision on all available information of the information bits is carried out. Therefore, the MAP rule according to (2.56) is applied on each information bit. The overall decoding performance was unmatched at that time, while the decoding complexity is still feasible. In [65], it is proven theoretically that iterative decoding approximates the overall joint decoding. The gap to an overall (or *flat*) decoder was shown empirically in [75] to be less than about 0.5 dB.

Shortly after, this turbo principle [76] was adopted on a myriad of other scenarios: turbo equalization [77], iterative source-channel decoding [78], or iterative multiuser detection [79], to name just a few. All these schemes have in common that they possess the structure of a serial concatenated code (SCC). However, the term "code" is now to be understood in a broader sense. Any abstract block that induces some dependencies among bits or symbols can be exploited in an iterative receiver as being a code. For our given examples, the dispersive channel, the source encoder, or the multiuser interference, respectively, are instances of rate 1 codes under this perception. Also the BICM-ID scheme, cf. Section 2.4, belongs to the class of SCCs, because the mapper introduces dependencies. We now summarize the most important facts about concatenated codes.

For PCCs, the information sequence is the input to *all* encoders, only that it is reshuffled by interleavers. This information sequence is also transmitted. This means that each constituent decoder at the receiver side obtains channel observations directly. In order to provide each decoder only with new information, the channel information must be removed from (2.64). This gives the "pure" extrinsic L-value, see [14, 80, 81]. Note that all decoders may work in parallel and then exchange their extrinsic information to start a new iteration.

For an SCC, we restrict ourselves on only two constituent encoders. Thus, the information sequence is encoded first by an *outer* encoder, next interleaved, and finally encoded by an *inner* encoder. The receiver operates in reversed order. As, however, the outer decoder does not receive any channel knowledge, the full extrinsic information of (2.64) is fed forward. This means that after deinterleaving, the outer APP decoder interprets the deinterleaved extrinsic L-values as if they are channel observations of the (outer) coded bits. More precisely, they are treated, as if they stem from noisy BPSK symbol observations. The outer APP decoder may also accept a priori knowledge of the uncoded bits, but typically, this is not available. In accordance with the majority of the literature, we will also denote the deinterleaved extrinsic L-values as a priori input to the outer decoder, but keep in mind, that these values are a priori knowledge of the *coded* bits.

Both SCCs and PCCs separate the constituent encoders by interleavers. The number of bits

that are interleaved in one block is called interleaver depth N_{int}. As already mentioned, one task of an interleaver is to scatter bursty error patterns to enhance the decoding capability of an outer decoder in case of SCCs. We have further seen that the interleaver creates a random-like code. This is the case, if we choose a pseudo-random interleaver, as we will do in this thesis. Other interleaver designs are discussed in [82]. Two more attributes of an interleaver should be pointed out, that depend on two different theories about turbo codes. The first point of view is that interleavers reduce the multiplicity of low weight code words, which was examined in [80, 81, 83, 84, 85, 86]. There, the analysis of the turbo code is with respect to its overall decoding error probability. The multiplicity of these low weight code words was shown to decrease inversely proportional to the interleaving depth N_{int}, provided that the inner encoder in case of an SCC is recursive. In case of a PCC, all constituent encoders have to be recursive to achieve this so called *interleaver gain*. The second interpretation of an interleaver is that it decorrelates the bit sequences that leave each component encoder. This is necessary in order to guarantee that the a priori information in the iterative receiver stems from an independent source. Moreover, the APP computation in (2.55) assumes that the a priori probabilities are independent, see step (5) of this equation. The effect of short interleavers can be seen in the extrinsic information transfer (EXIT) chart [87, 88], which describes the turbo decoder in terms of its iterative exchange of extrinsic information.

2.4 BICM with and without Iterative Demapping

Figure 2.14 depicts the transmitter of the BICM scheme. The information bits $\mathbf{b}_\kappa$ that leave the source are assumed to be independent and uniformly distributed (i.u.d.). This means that for each bit, $P[b_\kappa = 1] = P[b_\kappa = 0] = \frac{1}{2}$ holds[8]. As mentioned in Subsection 2.3.2, the outer encoder is a systematic convolutional code of rate $R_{\text{c}} = \frac{1}{2}$, which will be described by its generator polynomials $\big(g_{\text{r}}(\text{D}), g_{\text{f}}(\text{D})\big)$. It yields the coded bit sequence $\mathbf{c}_\kappa$, which is input to the pseudo-random interleaver. We set the code word length N to be equal to the interleaver depth N_{int}. For large enough N_{int}, we may assume that the interleaved bits are i.u.d., too. This is typically the case, if $N_{\text{int}} > 10^3$ [50]. The interleaved bit sequence has a different arrangement and further, the bits are grouped as described in Subsection 2.1.3 to form the vectors $\mathbf{c}_k$. The mapper acts as an inner encoder of rate 1. It outputs symbols s_k, which are transmitted over the discrete-time equivalent baseband channel.

The BICM receiver can be seen in Figure 2.15. Based on channel observation r_k and a priori information $\mathbf{L}_{\text{A},k}$, the optimum demapper computes the APP L-values $\mathbf{L}_{\text{P},k}$ according to (2.63). During the first pass through the demapper, all a priori L-values are set to zero. The extrinsic L-values $\mathbf{L}_{\text{E},k}$ are obtained by subtracting $\mathbf{L}_{\text{A},k}$ from $\mathbf{L}_{\text{P},k}$ as in (2.64). After deinterleaving, they are used as a priori L-values of the coded bits for the outer decoder and are denoted as $\mathbf{L}^{(\text{c})}_{\text{A},\kappa}$. The superscript "(c)" indicates that the information is with respect to the coded bits, while the subscript "κ" stands for the original arrangement of the bits, i.e., before interleaving.

[8] Otherwise, the bit sequence includes redundancy, which can be eliminated by appropriate source coding.

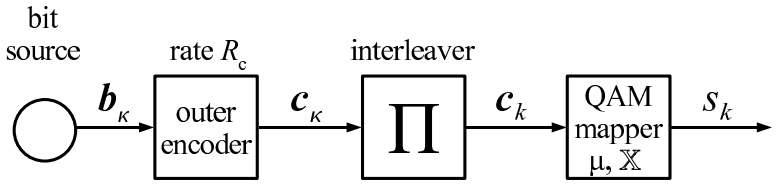


Figure 2.14: BICM transmitter

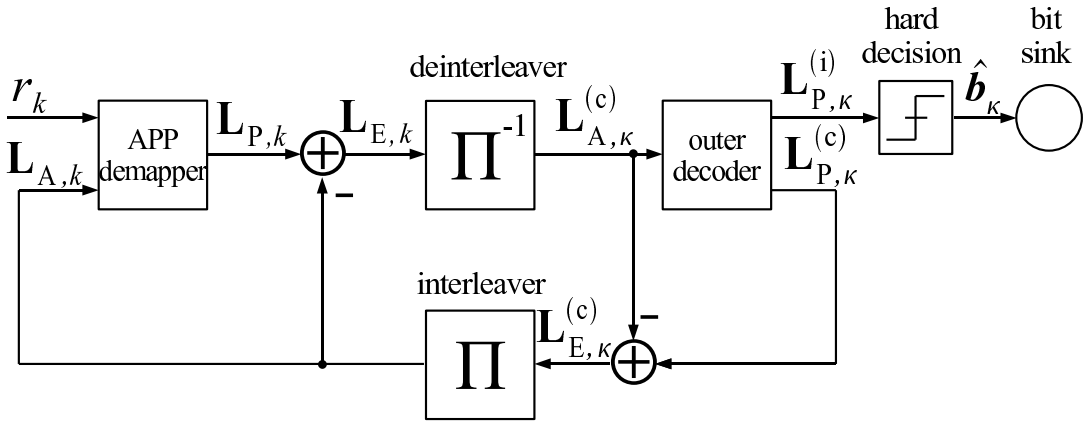


Figure 2.15: BICM receiver with iterative feedback loop

The optimum outer decoder applies the BCJR algorithm in the log-domain [89, 90] and outputs both APP L-values $\mathbf{L}^{(\mathrm{c})}_{\mathrm{P},\kappa}$ and $\mathbf{L}^{(\mathrm{i})}_{\mathrm{P},\kappa}$ for coded and uncoded (information) bits, respectively. The latter are indicated by superscript "(i)" and are input to a hard decision device, which simply considers the sign of these L-values. This yields the estimates of the information bits, $\hat{\mathbf{b}}_\kappa$, for the bit sink. A non-iterative BICM receiver stops at this point. Note that in that case, MLSE with the Viterbi algorithm is sufficient for the outer decoder [33]. This results in a small performance degradation compared to bitwise MAP decoding [91], but has a much smaller decoding complexity (about one third of the BCJR algorithm [92]).

An iterative receiver for the BICM-ID scheme requires the lower feedback branch. The extrinsic L-values of the coded bits $\mathbf{L}^{(\mathrm{c})}_{\mathrm{E},\kappa} = \mathbf{L}^{(\mathrm{c})}_{\mathrm{P},\kappa} - \mathbf{L}^{(\mathrm{c})}_{\mathrm{A},\kappa}$ are interleaved and can now be utilized in a new iteration as a priori information $\mathbf{L}_{\mathrm{A},k}$ at the demapper.

2.5 Fundamentals of Information Theory

In this section, we review some fundamentals of information theory. We start with the notion of entropy and mutual information. Then we define channel capacity and distinguish between different constrained capacities. Both mutual information and capacity will become targets for optimization of QAM mappings and are thus discussed here. For all other parts of information theory and further details, the reader is referred to [67].

2.5.1 Entropy and Mutual Information

The *entropy* of an arbitrary random variable x is defined as

$$H(x) = -\mathsf{E}_x\left[\log_2 p(x)\right]. \tag{2.66}$$

The logarithm is taken to the base 2, such that the entropy is measured in bits per symbol. Depending on the point of view, different interpretations for entropy exist. If the random variable is the discrete-time output of a source, i.e., x corresponds to a process s_k, then the entropy $H(s_k)$ reflects the average information content of this source. The second interpretation follows from Shannon's source coding theorem, which states that $H(s_k)$ is the minimum average number of bits per symbol that are necessary to encode s_k [4]. Finally, for a receiver, $H(s_k)$ is a measure for the a priori uncertainty of the transmitted symbols s_k. By means of channel observations r_k, the uncertainty can be turned into information. Due to channel impairments, however, the uncertainty does not resolve completely into information, but some amount is inevitably lost.

This loss is determined by the conditional entropy (also called a posteriori entropy or equivocation)

$$H(s_k|r_k) = -\mathsf{E}_{s_k,r_k}\left[\log_2 p\left(s_k|r_k\right)\right], \tag{2.67}$$

where the expectation operator averages with respect to the joint probability density function $p(s_k, r_k) = P[s_k] \cdot p\left(r_k|s_k\right)$ — assuming s_k is discrete-valued here.

The throughput of information is the difference

$$I(s_k, r_k) = H(s_k) - H(s_k|r_k), \tag{2.68}$$

and is denoted as the (average) *mutual information* (MI) or *transinformation* between s_k and r_k. Inserting (2.66) and (2.67) in (2.68), we get

$$\begin{aligned}
I(s_k, r_k) &= -\mathsf{E}_{s_k}\left[\log_2 P[s_k]\right] + \mathsf{E}_{s_k,r_k}\left[\log_2 p\left(s_k|r_k\right)\right] \\
&\overset{(1)}{=} \mathsf{E}_{s_k,r_k}\left[\log_2 \frac{p\left(s_k|r_k\right)}{P[s_k]}\right] \\
&\overset{(2)}{=} \mathsf{E}_{s_k,r_k}\left[\log_2 \frac{p(s_k, r_k)}{p(r_k) \cdot P[s_k]}\right] \\
&\overset{(3)}{=} \mathsf{E}_{s_k,r_k}\left[\log_2 \frac{p\left(r_k|s_k\right)}{p(r_k)}\right] = \mathsf{E}_{s_k,r_k}\left[\log_2 \frac{p\left(r_k|s_k\right)}{\sum\limits_{x'_l \in \mathbb{X}} P\left[x'_l\right] \cdot p\left(r_k \middle| s_k = x'_l\right)}\right] \\
&\overset{(4)}{=} \int\limits_{r_k \in \mathbb{C}} \sum_{x_l \in \mathbb{X}} P[x_l]\, p\left(r_k|s_k = x_l\right) \log_2 \frac{p\left(r_k|s_k = x_l\right)}{\sum\limits_{x'_l \in \mathbb{X}} P\left[x'_l\right] \cdot p\left(r_k \middle| s_k = x'_l\right)}\, dr_k.
\end{aligned} \tag{2.69}$$

In step (1), we have extended the first expectation operator also over r_k, which has no effect due to the law of total probability, and further combined both terms, because expectation is a linear operation. Expanding the fraction by $p(r_k)$ yields step (2). From this symmetry, we can easily recognize that $I(s_k, r_k) = I(r_k, s_k)$. The symmetric version is extended in (3) and in the last step, the expectation over both s_k and r_k is performed. Extensions of (2.69) to MIMO schemes are straightforward, if (2.42) is inserted for the transition pdf and scalars are replaced by vectors [44].

The MI $I(s_k, r_k)$ is a measure of the information that the observation of r_k provides about s_k or vice versa[9]. Obviously, $0 \leq I(s_k, r_k) \leq H(s_k)$. The lower bound holds, if s_k and r_k are independent, while the upper bound applies, if r_k is a function of s_k. The latter would occur for ideal communication channels.

2.5.2 Capacity Limits

2.5.2.1 Shannon's channel capacity

In Section 2.2, we described the impairments of the communication channel by its transition pdfs $p\left(r_k | s_k = x_l\right)$. With the set of transition pdfs and the distribution of transmit symbols, $p(x_l)$, the MI $I(s_k, r_k)$ can be determined according to (2.69). In general, the transmit symbols $x_l \in \mathbb{X}$ may be continuous-valued. Shannon introduced the *channel capacity* C_C as the maximum MI $I(s_k, r_k)$ between channel input s_k and output r_k among all possible distributions $p(x_l)$ of an arbitrary symbol alphabet $\mathbb{X}$.

$$C_\mathrm{C} = \max_{p(x_l)} I(s_k, r_k). \tag{2.70}$$

Shannon's famous channel coding theorem [4] states that reliable data communication is only possible, if the number of information bits per transmit symbol, $\eta = R_\mathrm{c} \cdot M$, does not exceed the channel capacity, i.e.,

$$\eta \leq C_\mathrm{C}. \tag{2.71}$$

In other words, the maximum possible bit rate over a given channel of bandwidth B in bits per seconds is $C_\mathrm{C} \cdot B$. Thus, the dimension of C_C is often given as $\mathrm{bits}/(\mathrm{s} \cdot \mathrm{Hz})$ or $\mathrm{bits}/\mathrm{channel\,usage}$. The proof of the channel coding theorem, which was later refined by Gallager [93], is not constructive in the sense that it claimes, *which* channel codes can provide reliable communication. It only considers the ensemble of all randomly chosen codes, and is therefore also called random coding theorem. It is still an ongoing challenge to achieve close to capacity communication systems.

[9] Another common, yet imprecise, interpretation of $I(s_k, r_k)$ is that it measures the mutual dependence between s_k and r_k. However, the same vague claim can be stated about crosscorrelation between s_k and r_k.

Examples: For the non-fading AWGN channel, the optimum signal alphabet can be shown to be (complex) Gaussian distributed with zero mean and variance σ_s^2, [4, 67, 93]. The corresponding channel capacity is

$$C_C = \log_2\left(1+\frac{\sigma_s^2}{\sigma_n^2}\right), \tag{2.72}$$

where the SNR $\frac{\sigma_s^2}{\sigma_n^2}$ is the non-logarithmic version of (2.29). Hence, the AWGN channel capacity is only a function of the SNR and its inverse function is

$$\frac{\sigma_s^2}{\sigma_n^2} = 2^{C_C} - 1. \tag{2.73}$$

Setting the spectral efficiency to its maximum possible value, which is according to (2.71) $\eta = C_C$, and inserting (2.73) into the non-logarithmic version of (2.31), we find

$$\frac{E_b}{N_0} = \frac{2^{C_C}-1}{C_C}. \tag{2.74}$$

The minimum necessary $\frac{E_b}{N_0}$ can be computed with l'Hôpital's rule as the limit

$$\left(\frac{E_b}{N_0}\right)_{\min} = \lim_{C_C\to 0}\frac{2^{C_C}-1}{C_C} = \lim_{C_C\to 0}\frac{2^{C_C}\cdot\ln(2)}{1} = \ln(2) \triangleq -1.59\text{dB}, \tag{2.75}$$

which means that the average energy per information bit must not fall below $-1.59\,\text{dB}$ of the noise power spectral density of the AWGN channel, to allow for any reliable (but low rate) communication at all.

As a second example, we consider the channel capacity of the BEC from Subsection 2.2.4 with erasure probability q. Clearly, the channel is doubly symmetric [93], because the transition pdfs are the same for all input symbols c_k and the same is true for the reciprocal channel, where input and output are interchanged. Thus, a uniform input distribution, i.e., $P[c_\kappa = 0] = \frac{1}{2}$, yields channel capacity [93, 52], which equals ■

$$C_C = \max I(c_k, y_k) = 1 - q. \tag{2.76}$$

2.5.2.2 Signal set capacity

In practical digital communication systems, the signaling alphabet can not be of infinite cardinality. For digital QAM schemes, we have $|\mathbb{X}| = L$, and usually $L = 2^M$ is a power of 2. For such a discrete-valued signal set, we can define the *signal set capacity* C_S in the same manner as in (2.70) as the maximum MI among all input distributions. However, in this thesis we restrict ourselves to equiprobable signal sets, i.e., $P[x_l] = L^{-1}$. Even though no maximization is involved, we will refer to the MI as signal set capacity, implying that the constraint exists

that equiprobable QAM symbols have to be transmitted over the communication channel. According to step (4) from (2.69), we have

$$C_{\mathrm{S}} = \int\limits_{r_k \in \mathbb{C}} \sum_{l=0}^{L-1} \frac{1}{L} \cdot p\left(r_k \,|s_k = x_l\right) \log_2 \frac{p\left(r_k \,|s_k = x_l\right)}{\frac{1}{L} \cdot \sum\limits_{l_1=0}^{L-1} p\left(r_k \,\middle|s_k = x_{l_1}\right)} dr_k. \tag{2.77}$$

2.5.2.3 BICM capacity

The signal set capacity C_{S} is the maximum possible signaling rate (in bits / channel usage) under the condition that equiprobable signals from a discrete set are transmitted. If we add as a further constraint on the transmission system that (optimum) demapping is performed only *once*, then we can compute the *BICM capacity* C_{B} as the MI between transmit symbols s_k and the output of the demapper. As the non-iterative demapper receives no a priori information, its APP output is equal to the extrinsic L-values. It was shown in [13, 33], that a fully[10] interleaved BICM schemes corresponds to the transmission of M parallel binary input channels, each of which carries one coded bit $b \in \{0,1\}$, which determines the symbol subset $\mathbb{X}_b^m$. Introducing this fact in the third step of (2.69), we get

$$\begin{aligned} C_{\mathrm{B}} &= \sum_{m=0}^{M-1} \mathsf{E}_{b,r_k}\left[\log_2 \frac{\sum\limits_{x_l \in \mathbb{X}_b^m} p\left(r_k \,|s_k = x_l\right)}{p\left(r_k\right)}\right] \\ &= \int\limits_{r_k \in \mathbb{C}} \sum_{m=0}^{M-1} \sum_{b=0}^{1} \sum_{x_l \in \mathbb{X}_b^m} \frac{1}{L} \cdot p\left(r_k \,|s_k = x_l\right) \log_2 \frac{\sum\limits_{x_l \in \mathbb{X}_b^m} p\left(r_k \,|s_k = x_l\right)}{\frac{1}{L} \cdot \sum\limits_{x_l' \in \mathbb{X}} p\left(r_k \,\middle|s_k = x_l'\right)} dr_k \end{aligned} \tag{2.78}$$

In the second line, expectation is taken over the joint pdf of b and r_k. The factor $\frac{1}{L} = \frac{1}{2^M}$ comes from averaging over $b \in \{0,1\}$, with $P[b] = 2^{-1}$ and from applying the law of total probability over all $x_l \in \mathbb{X}_b^m$, for which $P[x_l] = 2^{-M+1}$ holds.

Note that in contrast to the signal set capacity, which is determined by the alphabet $\mathbb{X}$, the BICM capacity depends also on the bit labeling function μ.

2.5.2.4 Comparison of capacity definitions

Figure 2.16 visualizes the previously introduced capacity definitions C_{C}, C_{S}, and C_{B}. For a particular communication channel, which is indicated in the block diagram by input ② and output ③, and which is fully specified by the set of transition pdfs $p\left(r_k \,|s_k = x_l\right)$, the ultimate signaling rate for reliable communication is given by the channel capacity C_{C}. This rate can

[10] Assuming an infinite depth random interleaver, which yields independent output bits.

only be achieved by optimized distributions of the transmit symbols, which are in general continuous-valued. Signal shaping techniques [32] try to approximate the optimum symbol distributions, even for a finite cardinality of the symbol alphabet.

A more realistic upper bound on the signaling rate is the signal set capacity C_S, which includes a certain signaling set as part of the channel. Thus, C_S describes the upper limit on the signaling rate for the discrete-input (see ①), continuous-output (see ③) channel. It is well known that achieving the signal set capacity requires the receiver to perform joint demapping and decoding. Hence, coded modulation systems as in MLC are able to achieve this capacity, and therefore C_S is sometimes also called *coded modulation capacity* [33]. Instead of the rather complex joint detection, iterative demodulation and decoding, as in BICM-ID, offers a close to optimum alternative. However, one must find mappings that can exploit a priori information most efficiently. These are certain anti-Gray mappings, as will be discussed in this thesis. As an advantage, a rather weak channel code is sufficient for capacity approaching performance.

If the output of an optimum demapper is further included in the overall "channel", then the BICM capacity C_B is the maximum rate from input ① to output ④. It was shown pragmatically in [13, 33] and theoretically in [94, 95], that Gray mappings achieve the largest C_B. Interestingly, Definition 2.1 allows for different classes of Gray mappings. The original recursive construction according to Gray [11] yields the so called binary reflected Gray labeling, which maximizes C_B for both AWGN and Rayleigh fading channel, while other classes of Gray labelings are inferior [94, 95]. In any case, a powerful channel code has to assure a performance close to the BICM capacity.

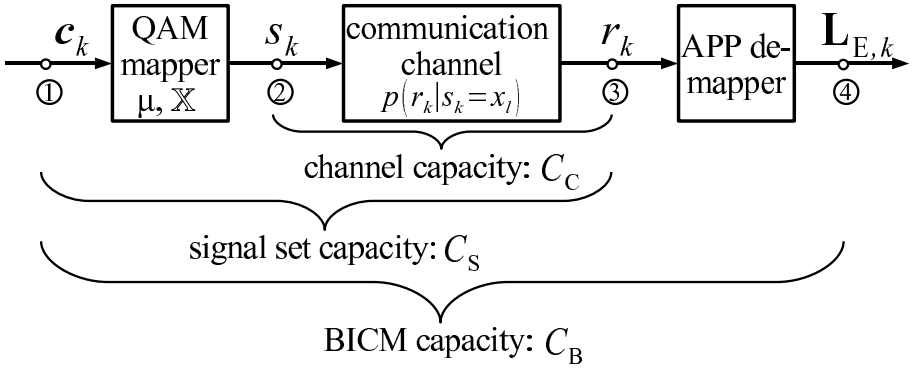


Figure 2.16: Definitions of channel, signal set, and BICM capacity

Finally, Figure 2.17 depicts the three types of capacities for an AWGN channel. The upper limit is given by Shannon's channel capacity C_C according to (2.72). The solid curves with markers belong to the signal set capacities C_S for BPSK, QPSK (cf. Figure 2.6), and 8-PSK, 16-QAM (cf. Figure 2.7), respectively. It is plain to see (and can be proven from (2.77)) that $0 \leq C_S \leq M$. The dashed curves correspond to the BICM capacities C_B. As we will discuss in Section 3.4, $0 \leq C_B \leq C_S \leq M$, where equality holds only for BPSK and QPSK with Gray labeling. Thus, C_B is not plotted for these two mappings. The depicted capacity curves belong to the Gray and anti-Gray labelings from Subsection 2.1.3. As mentioned

above, Gray labeling yields higher values for C_B than anti-Gray labelings. But even with Gray labeling, there exist a (small, but positive) difference $C_S - C_B$ for $M > 2$, which can be seen for medium SNR values in the range of about $[-5, 8]$ dB.

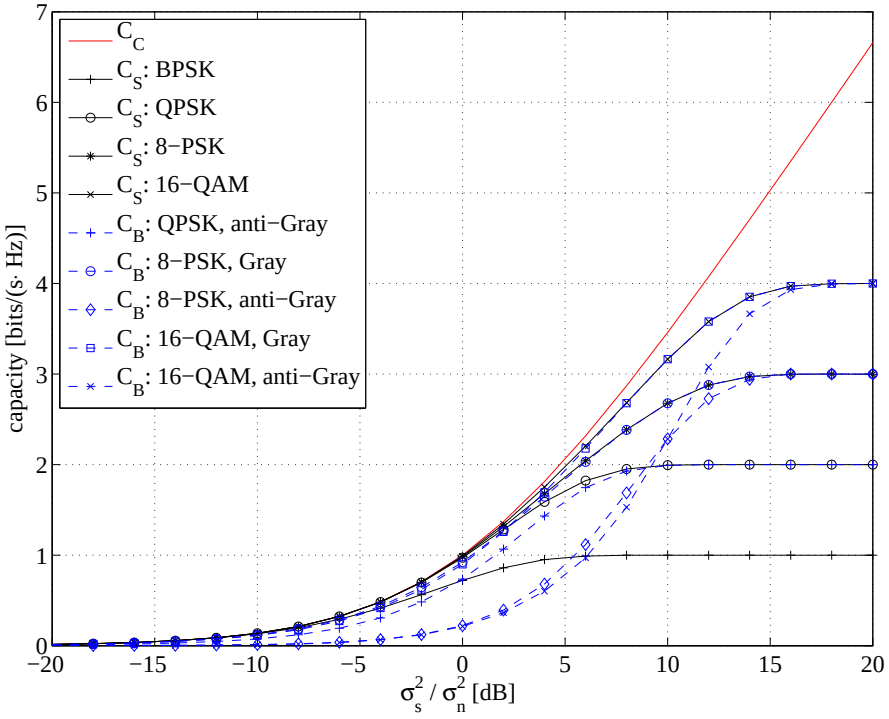


Figure 2.17: Capacities for different mappings over AWGN channel

2.6 EXIT Charts

The first turbo codes were designed by heuristic approaches, and left many open questions with respect to their decoding capabilities. Some light was shed on their inherent error-floor existence by lower bounding the BER with respect to the Hamming distance spectrum of the overall code [80, 85]. However, major questions (decrease of BER by many magnitudes after turbo-cliff, convergence of iterative decoding, interrelation of constituent codes) were not fully understood until the introduction of the EXIT chart by ten Brink [87, 88]. Since then, the EXIT chart has become an indispensable tool for designing concatenated coding schemes as well as for analyzing iterative receivers. This subsection summarizes the main characteristics of EXIT charts, while more information can be found in [96, 97].

For the EXIT chart, we consider each decoder (marked by index i) separately by its *transfer characteristic* T_i, which is also called *EXIT function*. In case of BICM-ID, the inner decoder, which corresponds to the demapper, is indexed by 1, and the outer decoder by 2. T_i describes the quality (or fidelity) of the extrinsic L-values, which are computed by the decoder, as a

function of a particular input quality, which is provided by the a priori L-values. In particular, we quantify the quality of a priori L-values of the inner demapper, $L_{\mathrm{A},k,\mathrm{m}}$, by the MI between these values and the transmitted bits $c_{k,m}$ to which they belong, and denote it as

$$I_{\mathrm{A1}} = I\left(c_{k,m}, L_{\mathrm{A},k,\mathrm{m}}\right). \tag{2.79}$$

Similarly, the output of the demapper is quantified by the MI between $L_{\mathrm{E},k,\mathrm{m}}$ and the transmitted bits $c_{k,m}$, and is called

$$I_{\mathrm{E1}} = I\left(c_{k,m}, L_{\mathrm{E},k,\mathrm{m}}\right). \tag{2.80}$$

Both quantities are bitwise ($L = 2$) MI, so $0 \leq I_{\mathrm{A1}}, I_{\mathrm{E1}} \leq 1$ holds. Further, as L-values are real-valued, equations (2.79) and (2.80) are computed according to (2.77) as

$$\int_{\xi\in\mathbb{R}} \sum_{b=0}^{1} \frac{1}{2} \cdot p\left(\xi \,\middle|\, c_{k,m} = b\right) \log_2 \frac{2 \cdot p\left(\xi \,\middle|\, c_{k,m} = b\right)}{p\left(\xi \,\middle|\, c_{k,m} = 0\right) + p\left(\xi \,\middle|\, c_{k,m} = 1\right)} d\xi, \tag{2.81}$$

where the dummy variable ξ for the pdfs $p\left(\xi | c_{k,m} = b\right)$, $b \in \{0,1\}$, is either $L_{\mathrm{A},k,\mathrm{m}}$ for computation of I_{A1} or $L_{\mathrm{E},k,\mathrm{m}}$ for I_{E1}. The transfer characteristic T_1 of the demapper is defined as I_{E1} as a function of I_{A1}, and — as the extrinsic output also depends on channel observations r_k, of the SNR $\frac{E_\mathrm{b}}{N_0}$, so $I_{\mathrm{E1}} = T_1\left(I_{\mathrm{A1}}, \frac{E_\mathrm{b}}{N_0}\right)$.

If we assume that the conditional pdfs of a priori L-values $p\left(L_{\mathrm{A},k,\mathrm{m}} | c_{k,m} = b\right)$ are Gaussian [98], then the corresponding mean value μ_A and variance σ_A^2 are interconnected as $\mu_\mathrm{A} = (2b-1) \cdot \sigma_\mathrm{A}^2/2$ [66]. Hence, the parameter σ_A is sufficient to determine I_{A1} with (2.81), which is then denoted as $J(\sigma_\mathrm{A})$ [88].

To determine T_1 for all $0 \leq I_{\mathrm{A1}} \leq 1$, where $I_{\mathrm{A1}} = 0$ and $I_{\mathrm{A1}} = 1$ correspond to no and perfect a priori information, respectively, the demapper receives channel observations r_k from Monte Carlo simulation, while the a priori L-values are created artificially as being i.i.d. with Gaussian distribution. The standard deviation is computed[11] as $\sigma_\mathrm{A} = J^{-1}(I_{\mathrm{A1}})$ and the mean value is as described above. Measuring the histograms of $L_{\mathrm{E},k,\mathrm{m}}$ for both cases, when either $c_{k,m} = 0$ or $c_{k,m} = 1$ was transmitted, allows for an approximation of the pdfs $p\left(L_{\mathrm{E},k,\mathrm{m}} | c_{k,m} = b\right)$. This in turn yields with (2.81) the MI I_{E1}.

The same procedure is carried out for the outer decoder to obtain $I_{\mathrm{E2}} = T_2\left(I_{\mathrm{A2}}\right)$. As discussed in Subsection 2.3.4, this decoder does not receive direct channel observations, so that T_2 is independent of $\frac{E_\mathrm{b}}{N_0}$. For an EXIT chart, both transfer characteristics, T_1 and T_2, are plotted in the same diagram, but the latter function with swapped axes.

Example: Figure 2.18 shows an EXIT chart for BICM-ID over the AWGN channel. For the demapper, three transfer characteristics are plotted: 16-QAM Gray mapping at $\frac{E_\mathrm{b}}{N_0} = 4\,\mathrm{dB}$

[11] $J(\sigma_\mathrm{A})$ is strictly monotonically increasing, thus its inverse function exists.

and 16-QAM anti-Gray at 3 and 4 dB. The mappings are as defined in Figure 2.7(b). The outer decoder belongs to a $(04,07)$ systematic convolutional code of rate $R_c = \frac{1}{2}$. Iterative exchange of extrinsic information (hence the name extrinsic information transfer chart) for the anti-Gray mapping at 4 dB is visualized by the zig-zag trajectory. It represents on-the-fly measurements of the points (I_{A1}, I_{E1}) and (I_{A2}, I_{E2}) during the actual iterative decoding process (10 iterations). Note that although a priori L-values are in general not Gaussian distributed, the real measurements match well with the transfer characteristics, which were generated by the Gaussian approximation. This underscores the robustness of MI against different channel assumptions [96]. It can be observed that each iteration increases MI, until the trajectory gets stuck at the intersection of T_1 and T_2. The anti-Gray mapping at 3 dB would produce an early intersection of both curves in the lower left part of the EXIT chart, resulting in high BER. This means that the turbo-cliff position is expected to occur at an SNR slightly larger than 3 dB (cf. Subsection 3.6.1). The transfer characteristic of the Gray mapping is almost flat, so the gain from iterative demapping is very small. ■

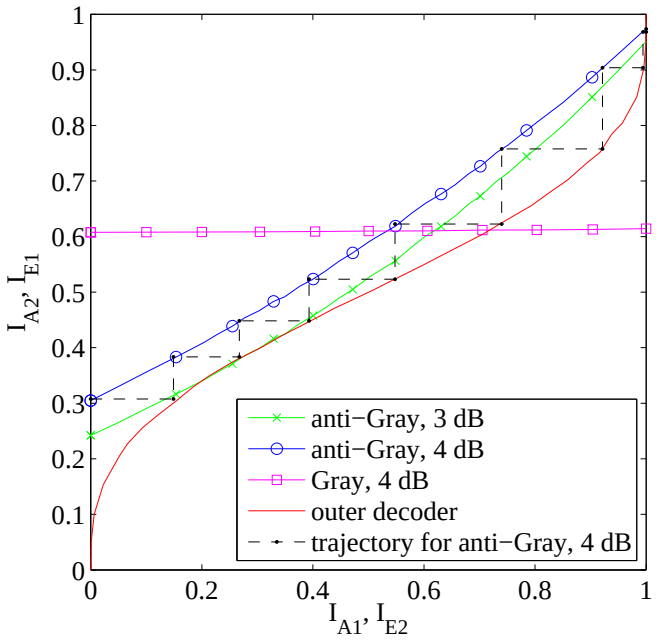


Figure 2.18: EXIT chart example for BICM-ID with 16-QAM over AWGN

We end this section with an enumeration of important properties of EXIT charts.

1. Each point $(I_{Ai}, I_{Ei}), i \in \{0,1\}$, corresponds to a BER P_{bi}, which would result, if hard decision on APP L-values (sum of a priori and extrinsic L-values) is performed. Applying the Gaussian approximation on the L-values, the BER is [88]

$$P_{bi} \approx \frac{1}{2}\operatorname{erfc}\left(\frac{\sqrt{\left[J^{-1}(I_{Ai})\right]^2 + \left[J^{-1}(I_{Ei})\right]^2}}{2\sqrt{2}}\right), \tag{2.82}$$

where erfc$(.)$ is the complementary error function [99]. Hence, it is desirable that the intersection of T_1 and T_2 reaches the upper right corner $(1,1)$ in the EXIT chart.

2. Comparing the bitwise MIs (2.80) and (2.81) for the case of no a priori information with the definition of the BICM capacity (2.78), it can be seen that [96]

$$I_{\mathrm{E}1}\left(I_{\mathrm{A}1}=0\right)=C_{\mathrm{B}}/M. \tag{2.83}$$

3. The area under the transfer characteristic T_1 of a demapper is related to the signal set capacity,

$$\int_0^1 I_{\mathrm{E}1}\,dI_{\mathrm{A}1}=C_{\mathrm{S}}/M. \tag{2.84}$$

A strict proof is only available for the scenario, that a priori values stem from a BEC [100, 101], but (2.84) holds empirically for all channels.

4. The area under the transfer characteristic T_2 of the outer decoder, where the axes are swapped, equals its code rate,

$$\int_0^1 I_{\mathrm{A}2}\,dI_{\mathrm{E}2}=1-\int_0^1 I_{\mathrm{E}2}\,dI_{\mathrm{A}2}=R_{\mathrm{c}}. \tag{2.85}$$

Again, the proof only holds for the BEC assumption of a priori input [101], while measurements on other channels back up this claim empirically.

5. Assume that the inner ($i=1$) or the outer encoder ($i=2$) is time-varying and made up of N different codes. Each of these code components is used for a fraction α_n of the total code word, such that

$$\sum_{n=1}^{N}\alpha_n=1. \tag{2.86}$$

Denote the transfer characteristic of the n-th code component as $I_{\mathrm{E}i,n}$. Then the overall transfer characteristic, which is per definition an *average* MI, can be expressed by the mixing property [101] as

$$I_{\mathrm{E}i}=\sum_{n=1}^{N}\alpha_n\cdot I_{\mathrm{E}i,n}. \tag{2.87}$$

From properties 3 and 4, Shannon's coding theorem (2.71) can be reformulated as a curve-fitting problem: reliable communication is only possible, if the demapper's transfer characteristic T_1 lies above the curve T_2 from the channel code.

Chapter 3

Optimization of QAM Mappings

In this chapter, high order mappings are investigated and optimized with respect to different cost functions. In particular, we give results for L-ary QAM, with $L \in \{4, 8, 16, 32\}$. As discussed in Subsection 2.1.3, a QAM mapping is defined by its alphabet $\mathbb{X}$ (constellation diagram), as well as the labeling function μ, which describes the assignment of $M = \log_2 L$ bits to symbols $x_l \in \mathbb{X}$. The cost functions are on the one hand the resulting SER and BER, if hard decision is performed after the first pass through the demapper, which ought to be minimized. On the other hand, we propose mappings that maximize either signal set or BICM capacity as well as mappings, which exploit perfect a priori knowledge at the demapper most efficiently. Finally, we search for mappings that allow for a good tradeoff between no and perfect a priori knowledge at the demapper. We restrict ourselves mostly to the AWGN channel. The resulting mappings were published in parts in [102, 26]. We begin this chapter by introducing some notations and with the formulation and investigation of the optimization problem. Then, the optimization algorithm is described and results are given for different cost functions in subsequent sections.

3.1 Prerequisites

3.1.1 Problem Formulation

The symbol alphabet $\mathbb{X}$ is the set of all possible transmit symbols, and per definition a set is an unordered collection of elements. In order to combine alphabet $\mathbb{X}$ and labeling function μ, we now introduce the mapping vector $\mathbf{x}$ as the *ordered* arrangement of all $x_l \in \mathbb{X}$ as

$$\mathbf{x} = \left(x_0, \ldots, x_{L-1}\right) \quad \text{with} \quad x_l = \mu\left(l_{\text{bin}}\right),\ l \in \{0, \ldots, L-1\}. \tag{3.1}$$

l_{bin} is the binary representation of integer l, where — without loss of generality due to the third type of invariant transformations in Subsection 2.1.3, the leftmost bit should correspond to the most significant bit (MSB). In this chapter, we exclude mappings for MIMO systems, hence no confusion with MIMO transmit vector $\mathbf{x}_l$ from (2.65) is to be feared.

Examples: The 8-PSK Gray mapping from Figure 2.7(a) can be described with (3.1) as

$$\mathbf{x} = \left(\mathrm{e}^{\mathrm{j}0}, \mathrm{e}^{\mathrm{j}\frac{\pi}{4}}, \mathrm{e}^{\mathrm{j}\frac{3\pi}{4}}, \mathrm{e}^{\mathrm{j}\frac{\pi}{2}}, \mathrm{e}^{-\mathrm{j}\frac{\pi}{4}}, \mathrm{e}^{-\mathrm{j}\frac{\pi}{2}}, \mathrm{e}^{\mathrm{j}\pi}, \mathrm{e}^{-\mathrm{j}\frac{3\pi}{4}}\right), \tag{3.2}$$

while the 8-PSK anti-Gray mapping in this figure is defined by

$$\mathbf{x} = \left(\mathrm{e}^{\mathrm{j}0}, \mathrm{e}^{-\mathrm{j}\frac{3\pi}{4}}, \mathrm{e}^{\mathrm{j}\frac{\pi}{2}}, \mathrm{e}^{-\mathrm{j}\frac{\pi}{4}}, \mathrm{e}^{\mathrm{j}\pi}, \mathrm{e}^{\mathrm{j}\frac{\pi}{4}}, \mathrm{e}^{-\mathrm{j}\frac{\pi}{2}}, \mathrm{e}^{\mathrm{j}\frac{3\pi}{4}}\right). \tag{3.3}$$

■

The power constraint (2.32) can thus be reformulated as

$$\mathscr{P}_x = \mathsf{E}_{x_l}\left[|x_l|^2\right] = \frac{1}{L}|\mathbf{x}|^2 \stackrel{!}{=} 1. \tag{3.4}$$

Further, let $\mathbf{1}_L$ be a row vector of length L, whose entries are all 1. Hence, the center of gravity of the QAM mapping, i.e., the mean value of all x_l, can be computed as

$$\mathsf{E}_{x_l}[x_l] = \frac{1}{L}\mathbf{x}\cdot\mathbf{1}_L^{\mathrm{T}} \tag{3.5}$$

and should be zero to minimize $\mathscr{P}_x$ [103]. Assume, $\mathbf{x}$ possesses a non-zero mean. If we subtract this mean value from all components of $\mathbf{x}$, the resulting zero mean vector $\mathbf{x}_{\mathrm{Z}}$ is

$$\mathbf{x}_{\mathrm{Z}} = \mathbf{x} - \frac{1}{L}\mathbf{x}\cdot\mathbf{1}_L^{\mathrm{T}}\cdot\mathbf{1}_L = \mathbf{x}\cdot Z_L, \tag{3.6}$$

in which the matrix $Z_L \in \mathbb{R}^L$ forces zero mean and is computed as

$$Z_L = I_L - \frac{1}{L}\mathbf{1}_L^{\mathrm{T}}\cdot\mathbf{1}_L, \tag{3.7}$$

i.e., the diagonal elements of Z_L are $1-\frac{1}{L}$, and the rest equals $-\frac{1}{L}$.

The task at hand is to find a vector $\mathbf{x}$, which maximizes a cost function $f(\mathbf{x})$, and fulfills the power constraint (3.4),

$$\mathbf{x}_{\mathrm{opt}} = \arg\max_{\mathbf{x}:\mathscr{P}_x=1} f(\mathbf{x}). \tag{3.8}$$

In Section 3.2, we set $f = -P_{\mathrm{s}}$, in order to minimize SER, in Section 3.3 and Section 3.4 BICM and signal set capacity are maximized by $f = C_{\mathrm{B}}$ and $f = C_{\mathrm{S}}$, respectively. It will turn out that the former is identical with minimization of BER (under the Gaussian assumption of extrinsic L-values). Utilization of perfect a priori information at the demapper in the most efficient way will be covered in Section 3.5 by letting $f = I_{\mathrm{E1}}(1)$, and finally, in Section 3.6, a tradeoff function will be introduced that allows for a compromise between no and perfect a priori knowledge at the demapper. Since (3.8) yields the joint optimization of symbol constellation and labeling function, we will refer to the solutions $\mathbf{x}_{\mathrm{opt}}$ as L-QAM, e.g., 8-QAM.

The following example shall illustrate the problem of solving (3.8).

Example: Let us determine the 2-QAM vector $\mathbf{x} = (x_0, x_1)$, which maximizes the signal set capacity $f(\mathbf{x}) = C_\mathrm{S}$ (which equals BICM capacity C_B, because $M = 1$) at $\frac{\sigma_\mathrm{s}^2}{\sigma_\mathrm{n}^2} = -2.82\,\mathrm{dB}$. The BPSK from Figure 2.6(a) achieves $C_\mathrm{S} = 0.5$[1]. Equating $\eta = C_\mathrm{S}$ yields with (2.31) the SNR $\frac{E_\mathrm{b}}{N_0} = 0.19\,\mathrm{dB}$, which is the well-known limit for BPSK signaling with rate $R_\mathrm{c} = \frac{1}{2}$ over the AWGN channel [104, 105]. Without loss of generality, the quadrature-phase component of the symbols shall be zero, so $x_0, x_1 \in \mathbb{R}$. Figure 3.1 depicts the unconstrained capacity C_S for all $x_0, x_1 \in [-2, 2]$, i.e., (3.4) may be violated.

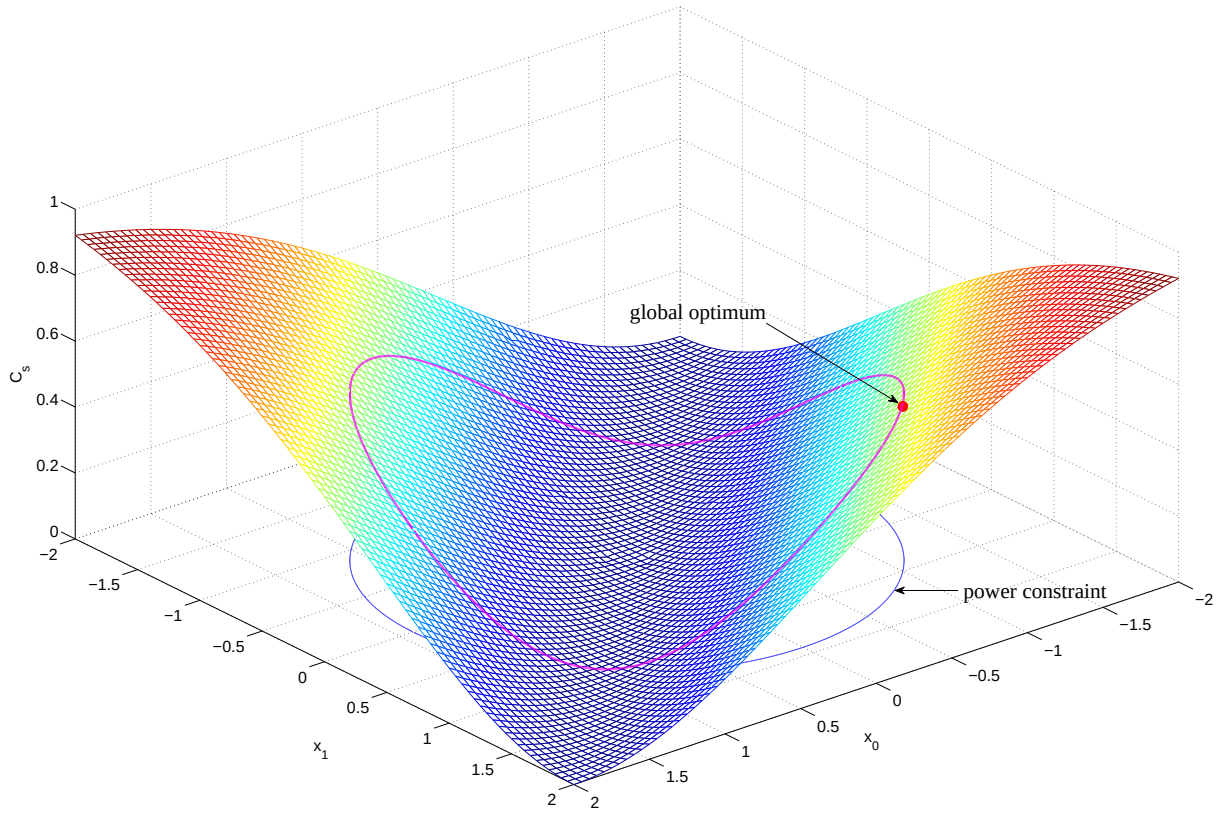


Figure 3.1: Unconstrained capacity of 2-QAM at $\frac{\sigma_\mathrm{s}^2}{\sigma_\mathrm{n}^2} = -2.82\,\mathrm{dB}$

Obviously, the capacity plane is symmetric to both the $x_0 = x_1$- and the $x_0 = -x_1$-plane. It is further plain to see, that the capacity is zero, if both symbols are identical, i.e., $C_\mathrm{S} = 0$ for $x_0 = x_1$, which can be seen in the viewing direction of Figure 3.1. Besides this minimum, the capacity plane has no other zero gradient. The gradient approaches zero (and maximizes the unconstrained capacity), if $x_0 = -x_1$ and $|x_0| \to \infty$, which violates the power constraint, $\mathscr{P}_x = \frac{1}{2}(x_0^2 + x_1^2) = 1$. All vectors $\mathbf{x}$ that fulfill this constraint are located on a circle in the x_0, x_1-plane, centered around the origin with radius $\sqrt{2}$, indicated by the blue line in the figure, while the magenta line shows the corresponding constrained capacity. The task is thus to find the global maximum of this line. One global optimum is marked by a red point, which corresponds to the normal BPSK with $\mathbf{x}_\mathrm{opt} = (-1, 1)$. However, due to the symmetry,

[1] We neglect the pseudo-dimension bits / symbol in the following.

another global optimum exists at $-\mathbf{x}_{\text{opt}}$. For higher modulation orders, $M > 1$, the number of global optima increases even more, which is plausible due to the invariant transformations from Subsection 2.1.3. Another problem that arises is the fact that in general, the cost functions are not convex [106]. Hence, Kuhn-Tucker conditions for constrained convex optimization are of no avail [107]. Also, the constrained optimization by Lagrange multipliers [107] becomes unmanageable for high order mappings, because the power constraint and the gradient are non-linear functions, the latter might include integrals. In the next subsection, we will therefore propose an algorithm, which can solve (3.8).

We end this example, by proving analytically, that $\mathbf{x}_{\text{opt}}$ maximizes C_{S} for a 2-QAM:

Clearly, x_0 and x_1 should be of opposite signs, as can be seen in Figure 3.1. Otherwise the center of gravity of $\mathbf{x}$ (cf. (3.5)) is non-zero and thus a zero mean mapping $\mathbf{x}_{\text{z}} = \mathbf{x} \cdot Z_2$ exists, in which the Euclidean distances between the symbols is increased, yielding $C_{\text{S}}(\mathbf{x}_{\text{z}}) > C_{\text{S}}(\mathbf{x})$. Hence, we set $x_0 = -\text{sgn}(x_1)\sqrt{2 - x_1^2}$ for $x_1 \in \left[-\sqrt{2}, \sqrt{2}\right] \setminus \{0\}$ and $x_0 = \pm\sqrt{2}$ for $x_1 = 0$, where $\text{sgn}(.)$ denotes the sign (signum) function. This fulfills constraint (3.4) and reduces the problem to a one-dimensional function. This constrained capacity is depicted in Figure 3.2 and corresponds to the projection of the magenta curve from Figure 3.1 onto the x_1, C_{S}-plane for the two quadrants, in which $x_0 \cdot x_1 \leq 0$ holds. Computation of the gradient and equating it to zero yields the optimum solutions $\mathbf{x}_{\text{opt}}$ (indicated again by a red point) and $-\mathbf{x}_{\text{opt}}$, and is derived in Appendix A.3.2. The four minima $(0, \pm\sqrt{2})$ and $(\pm\sqrt{2}, 0)$ correspond to 2-ASK mappings, with symbol spacing of $\sqrt{2}$. This minimum capacity is 0.3, and corresponds to a 3 dB reduction in SNR. ■

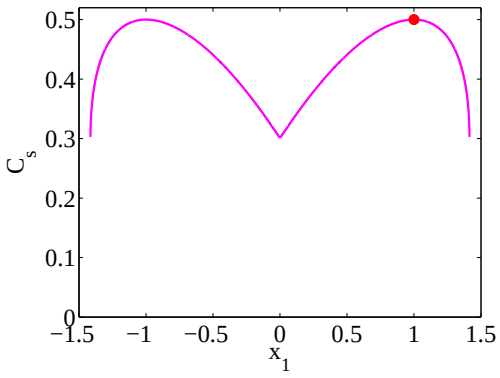


Figure 3.2: Constrained capacity of 2-QAM at $\frac{\sigma_{\text{s}}^2}{\sigma_{\text{n}}^2} = -2.82\,\text{dB}$

3.1.2 Optimization Algorithm

The previous example has shown that optimization according to (3.8) is already cumbersome for $M = 1$. In the same way, optimality of the BPSK for all cost functions can be derived. But for high order mappings, we have to resort e.g. to the following optimization algorithm,

which is a modified version of the constrained gradient-search algorithm [106, 108]. Starting with a randomly drawn mapping vector $\mathbf{x}_0$, a new vector is obtained by moving along the gradient. This gradient points in the direction of steepest ascent, yielding a larger cost function. However, the constraint (3.4) is in general violated and the vector has to be renormalized. This procedure is repeated iteratively, until the cost function stops improving. As this function is in general not convex, the algorithm converges towards a local optimum. Starting it all over again for many trials, the likelihood increases that the global optimum can be found. We have set the number of trails to 10^5. The following pseudo-code describes one trial.

1. **Initialization:**
   (a) Shuffle a random start mapping vector $\mathbf{x}'_0 \in \mathbb{C}^L$. Real and imaginary part of each component are drawn independently from a uniform distribution in $[-1,1]$.
   (b) Shift the center of gravity to the origin: $\mathbf{x}'_{0,\mathrm{z}} = \mathbf{x}'_0 \cdot Z_L$.
   (c) Normalize transmit power to satisfy (3.4): $\mathbf{x}_0 = \frac{\mathbf{x}'_{0,\mathrm{z}}}{\left|\mathbf{x}'_{0,\mathrm{z}}\right|} \cdot \sqrt{L}$.
2. **For each iteration** i **in** $0 \leq i < \infty$**:**
   (a) Compute cost function $f(\mathbf{x}_i)$ and gradient $\nabla f(\mathbf{x}_i)$.
   (b) Set step size to $\gamma = 1$.
      i. Compute intermediate vector $\mathbf{x}'_{i+1} = \mathbf{x}_i + \gamma \cdot \frac{\nabla f(\mathbf{x}_i)}{\left|\nabla f(\mathbf{x}_i)\right|}$.
      ii. Normalize transmit power to satisfy (3.4): $\mathbf{x}_{i+1} = \frac{\mathbf{x}'_{i+1}}{\left|\mathbf{x}'_{i+1}\right|} \cdot \sqrt{L}$
      iii. If the intermediate vector improves the cost function, i.e., if $f\left(\mathbf{x}_{i+1}\right) > f(\mathbf{x}_i)$, goto 2a and start next iteration.
      Otherwise, if $\gamma > 10^{-6}$, multiply step size γ by factor 10^{-1} and goto step 2(b)i, else finish with next step.
3. **Termination:** Store $\mathbf{x}_i$ and $f(\mathbf{x}_i)$ as a local maximum.

Only the initial mapping vector is forced to have zero mean in step 1b, because this is no constraint. It turns out that this property is maintained almost exactly in each iteration. Note that unlike the algorithm of [106], we use in step 2b the normalized gradient (unit vector) and further adapt the step size appropriately down to a magnitude of 10^{-6}. This avoids the problem of oscillation around a local optimum. It must be emphasized that this algorithm does not substitute a proof of optimality. We therefore define an indicator, which provides information about the tendency of $\mathbf{x}_i$ during the gradient-search algorithm:

Note that the renormalization in step 2(b)ii corresponds to a radial contraction of the mapping, i.e., the vector $\mathbf{x}_i$ is parallel to $\mathbf{x}'_i$, but has a shorter length. Now assume that $\mathbf{x}_i$ fulfills (3.4) and provides a local maximum of $f(\mathbf{x})$. In this case, the gradient points in the same direction as $\mathbf{x}_i$, which means that the only possible improvement comes from expansion of

the symbols, which would violate (3.4). Hence, in step 2(b)iii, no improvement is possible, as $\mathbf{x}_{i+1}$ will be identical with $\mathbf{x}_i$. Only if the gradient is not parallel to $\mathbf{x}_i$, an improvement after renormalization is possible. Dropping the index "i", we expand the l-th component of the mapping vector $\mathbf{x}$ in (3.1) as

$$x_l = \mathrm{Re}\{x_l\} + \mathrm{j}\mathrm{Im}\{x_l\} = x_{l,\mathrm{R}} + \mathrm{j}x_{l,\mathrm{I}} \tag{3.9}$$

and the l-th component of the gradient $\nabla f(\mathbf{x})$ as

$$\frac{\partial f(\mathbf{x})}{\partial x_l} = \mathrm{Re}\left\{\frac{\partial f(\mathbf{x})}{\partial x_l}\right\} + \mathrm{j}\mathrm{Im}\left\{\frac{\partial f(\mathbf{x})}{\partial x_l}\right\} = \frac{\partial f(\mathbf{x})}{\partial x_{l,\mathrm{R}}} + \mathrm{j}\frac{\partial f(\mathbf{x})}{\partial x_{l,\mathrm{I}}}. \tag{3.10}$$

In polar coordinates, we denote the angles of x_l and of $\frac{\partial f(\mathbf{x})}{\partial x_l}$ as

$$\varphi_l = \arg(x_l) = \arctan\left(\frac{x_{l,\mathrm{I}}}{x_{l,\mathrm{R}}}\right) + \begin{cases} 0 & , \text{ if } x_{l,\mathrm{R}} \geq 0 \\ \pi & , \text{ else} \end{cases}, \tag{3.11}$$

$$\psi_l = \arg\left(\frac{\partial f(\mathbf{x})}{\partial x_l}\right) = \arctan\left(\frac{\frac{\partial f(\mathbf{x})}{\partial x_{l,\mathrm{I}}}}{\frac{\partial f(\mathbf{x})}{\partial x_{l,\mathrm{R}}}}\right) + \begin{cases} 0 & , \text{ if } \frac{\partial f(\mathbf{x})}{\partial x_{l,\mathrm{R}}} \geq 0 \\ \pi & , \text{ else} \end{cases}. \tag{3.12}$$

With that, we can define the non-expansion coefficient (NEC) as

$$\Lambda = \sum_{l=0}^{L-1} \left|\varphi_l - \psi_l\right|. \tag{3.13}$$

This value approaches zero at a local optimum.

Example: Recalling the previous example of the capacity for two symbols. The gradient at the optimum vector $\mathbf{x}_{\mathrm{opt}} = (-1,1)$ equals $0.49 \cdot (-1,1)$, is therefore parallel to $\mathbf{x}_{\mathrm{opt}}$ and yields a NEC of $\Lambda = 0$. The suboptimum 2-ASK with $\mathbf{x} = (0, \sqrt{2})$ has the gradient $0.51 \cdot (-1,1)$, which is not parallel to $\mathbf{x}$. The non-zero NEC of $\Lambda = \pi$ indicates the tendency of $\mathbf{x}$ to move towards $\mathbf{x}_{\mathrm{opt}}$ during further iterations of the optimization algorithm. ■

Remarks on partial derivative: Strictly speaking, the left hand side from (3.10) does not exist, because $f(\mathbf{x})$ does not satisfy the Cauchy-Riemann differential equations and is thus not a holomorphic function [99]. However, we define the partial derivative as on the right hand side, in accordance with the notation from [106]. Instead of interpreting $\mathbf{x}$ as a vector from $\mathbb{C}^L$, it should rather be seen as a vector from $\mathbb{R}^{2L}$. The partial derivatives with respect to both $x_{l,\mathrm{R}}$ and $x_{l,\mathrm{I}}$ can be combined to form the complex derivative from (3.10), which is added to the l-th complex component x_l during the gradient-search algorithm.

3.2 Mappings for Minimum Symbol Error Rate

If all symbols are transmitted equally likely and if we assume that the demapper has no a priori knowledge of bits $\mathbf{c}_k$, then ML detection yields minimum SER, [30]. For the AWGN channel, the ML rule decides in favor of symbol $\hat{s}_k$, which has the smallest Euclidean distance to the received sample r_k. If we assume further that the demapper has only to decide between two symbols, x_{l_1} and x_{l_2}, and that the former was transmitted, $s_k = x_{l_1}$, the pairwise error probability (PEP) corresponds to the probability that the projection of the noise sample n_k onto the unit vector $\left(x_{l_2} - x_{l_1}\right) \cdot \left|x_{l_2} - x_{l_1}\right|^{-1}$ exceeds the decision threshold $\left|x_{l_2} - x_{l_1}\right|/2$. From the circular symmetry of n_k, we can conclude that its projection onto any unit vector is Gaussian distributed with zero mean and variance $\sigma_\mathrm{n}^2/2$. The PEP for this scenario is

$$P\left[\hat{s}_k = x_{l_2} \left| s_k = x_{l_1}\right.\right] = \int\limits_{|x_{l_2} - x_{l_1}|/2}^{\infty} \frac{\exp\left(-\frac{\xi^2}{\sigma_\mathrm{n}^2}\right)}{\sqrt{\pi}\sigma_\mathrm{n}} d\xi = \frac{1}{2}\operatorname{erfc}\left(\frac{\left|x_{l_2} - x_{l_1}\right|}{2\sqrt{2}\sigma_\mathrm{n}}\right). \tag{3.14}$$

Now, considering all transmit symbols, $x_l \in \mathbb{X}$, the (average) SER is expressed as

$$P_\mathrm{s} = \frac{1}{L}\sum_{l_1=0}^{L-1} P\left[\hat{s}_k \neq x_{l_1} \left| s_k = x_{l_1}\right.\right] = \frac{1}{L}\sum_{l_1=0}^{L-1}\left(1 - \int\limits_{\xi \in \mathscr{A}_{l_1}} \frac{\exp\left(-\frac{|\xi|^2}{\sigma_\mathrm{n}^2}\right)}{\pi\sigma_\mathrm{n}^2} d\xi\right). \tag{3.15}$$

The area $\mathscr{A}_{l_1}$ corresponds to all points ξ in the complex plane, for which $p\left(\xi \left| s_k = x_{l_1}\right.\right) > p\left(\xi \left| s_k = x_{l_2}\right.\right), \forall l_2 \neq l_1$. For the AWGN channel, this reduces to all points ξ, which are closer to x_{l_1} than to any other symbol x_{l_2}. However, these areas of closest neighbors are difficult to determine for arbitrary alphabets $\mathbb{X}$. Moreover, the integrals have in general no closed-form. Thus, we apply the well-known union bound on the SER, which considers only PEPs [30],

$$P_\mathrm{s} \lessapprox \frac{1}{L}\sum_{l_1=0}^{L-1}\sum_{\substack{l_2=0 \\ l_2 \neq l_1}}^{L-1} P\left[\hat{s}_k = x_{l_2} \left| s_k = x_{l_1}\right.\right] = \frac{1}{2L}\sum_{l_1=0}^{L-1}\sum_{\substack{l_2=0 \\ l_2 \neq l_1}}^{L-1} \operatorname{erfc}\left(\frac{\left|x_{l_2} - x_{l_1}\right|}{2\sqrt{2}\sigma_\mathrm{n}}\right). \tag{3.16}$$

Obviously, (3.16) provides an upper bound on P_s, because the summation of all PEPs repeatedly adds contributions from overlapping decision areas. Nevertheless, we will treat (3.16) as an equation to avoid yet another notation.

Example: We have simulated the SER for the QPSK, 8-PSK and 16-QAM mappings from Subsection 2.1.3 and computed the SER with (3.16). The results are depicted in Figure 3.3 and underscore the tightness of the union bound, especially for high SNR. This is because all addends from the second summation in (3.16) tend to zero, except the one that includes the nearest neighboring symbol of x_{l_1}. Thus, the overlapping contributions diminish. ■

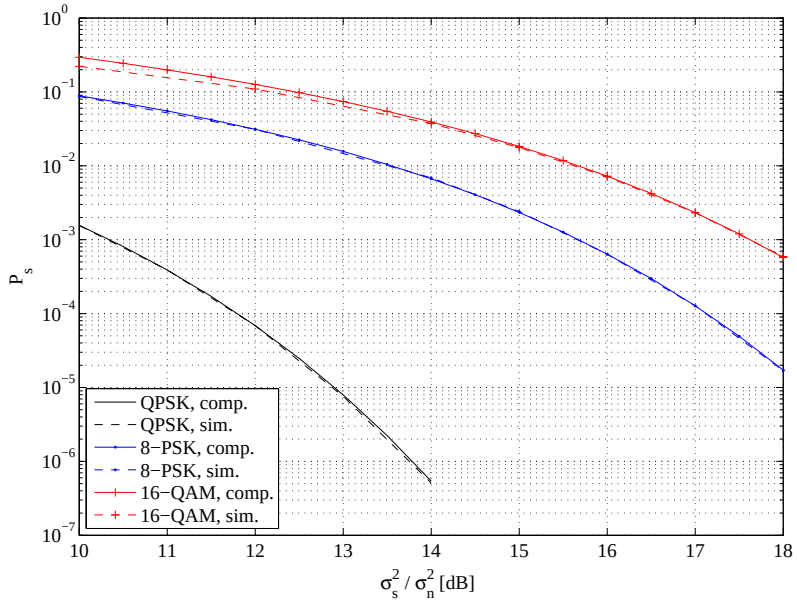


Figure 3.3: SER for different constellations: computed (—) and simulated (- - -)

The l-component of the gradient $\nabla P_\mathrm{s}(\mathbf{x})$ from the SER is derived in the Appendix A.1, and is with the notation from (3.10)

$$\frac{\partial P_\mathrm{s}(\mathbf{x})}{\partial x_l} = \frac{1}{L} \sum_{\substack{l_1=0 \\ l_1 \neq l}}^{L-1} \frac{x_{l_1} - x_l}{\left| x_{l_1} - x_l \right|} \cdot \frac{\exp\left(-\frac{\left| x_{l_1} - x_l \right|^2}{4\sigma_\mathrm{n}^2} \right)}{\sqrt{\pi}\sigma_\mathrm{n}}. \tag{3.17}$$

In the following subsections, we apply the optimization algorithm from Subsection 3.1.2 on (3.16) and (3.17) to obtain mappings $\mathbf{x}$ for minimum SER. Note that the cost function $f = -P_\mathrm{s}$ depends only on the constellation $\mathbb{X}$, which means that any reordered vector $\mathbf{x}'$ yields the same result. The results from Foschini et al. [106] serve as a baseline and will be denoted as $\mathbf{x}_{L,\mathrm{Fos}}$. However, these mappings were obtained by a slightly modified algorithm, as discussed before (fixed step size). Furthermore, an upper bound was applied on the erfc-function, which holds only for high SNR. Finally, each proposed mapping in [106] was found for a certain SNR (not given explicitly), which yields $P_\mathrm{s} \approx 10^{-6}$ for this mapping.

3.2.1 4-QAM for Minimum Symbol Error Rate

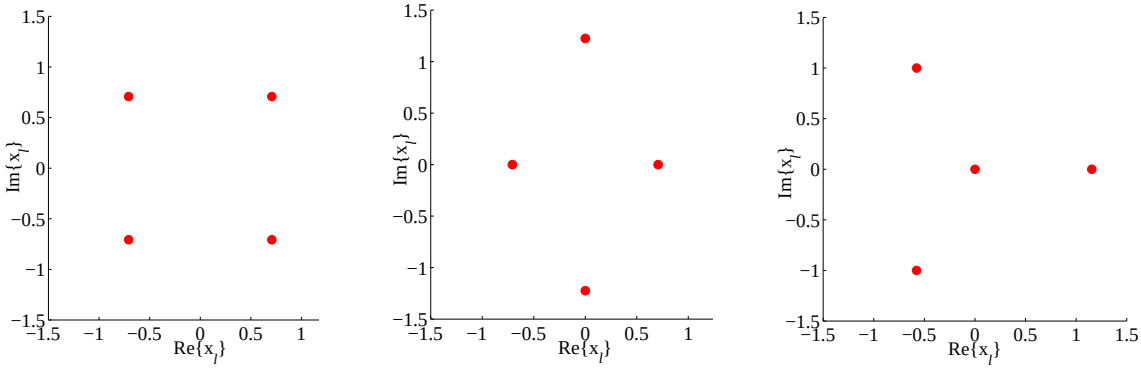
For $L = 4$, the optimum constellation for all SNRs is the QPSK mapping from Figure 2.6(b), which we denote as $\mathbf{x}_{4,\mathrm{opt}}$. For the sake of completeness, we examine two further candidates,

$$\mathbf{x}_{4,\mathrm{tri}} = \frac{1}{\sqrt{2}} \cdot \left(-1, 1, \mathrm{j}\sqrt{3}, -\mathrm{j}\sqrt{3} \right), \tag{3.18}$$

where the four symbols are located on equilateral triangles, and

$$\mathbf{x}_{4,1\_3} = \frac{2}{\sqrt{3}} \cdot \left(0, 1, \mathrm{e}^{\mathrm{j}\frac{2\pi}{3}}, \mathrm{e}^{-\mathrm{j}\frac{2\pi}{3}}\right), \tag{3.19}$$

which is an ASK/PSK combination with two amplitude levels and three phases. The mappings are depicted in Figure 3.4, and the resulting SER according to (3.16) can be seen in Figure 3.5.



(a) $\mathbf{x}_{4,\mathrm{opt}}$ mapping (b) $\mathbf{x}_{4,\mathrm{tri}}$ mapping (c) $\mathbf{x}_{4,1\_3}$ mapping

Figure 3.4: Three candidates for 4-QAM constellations

In general, an L-ary QAM mapping can be described by $L(L-1)/2$ Euclidean distances between symbol pairs x_{l_1} and x_{l_2}, $l_1 \neq l_2$, which are comprised in the *distance spectrum* [23]. Although $\mathbf{x}_{4,\mathrm{opt}}$ and $\mathbf{x}_{4,\mathrm{tri}}$ possess the same minimum Euclidean distance of $d_{\min} = 2/\sqrt{2}$, the resulting SER of the former is smaller. The reason is that $\mathbf{x}_{4,\mathrm{opt}}$ has four symbol pairs at $d_{\min}$ and two at distance 2, while $\mathbf{x}_{4,\mathrm{tri}}$ has five pairs at $d_{\min}$ and one at a larger distance of $2\sqrt{3}$. However, the SER is mostly determined by $d_{\min}$ and its multiplicity [30]. The $\mathbf{x}_{4,1\_3}$ mapping performs even worse, because here $d_{\min} = 2/\sqrt{3}$. Its NEC is $\Lambda = \pi$, so further iterations in the optimization algorithm would result in the global optimum, $\mathbf{x}_{4,\mathrm{opt}}$, which was also proposed in [106]. Both $\mathbf{x}_{4,\mathrm{opt}}$ and $\mathbf{x}_{4,\mathrm{tri}}$ are local optima, because their NEC is $\Lambda = 0$.

3.2.2 8-QAM for Minimum Symbol Error Rate

In contrast to $L \in \{2, 4\}$, the optimum mappings of 8-ary (and higher order) QAMs depend on the SNR. Applying the optimization algorithm at an SNR of $\frac{\sigma_s^2}{\sigma_n^2} = 10\,\mathrm{dB}$, the mapping with the smallest SER can be seen in Figure 3.6(a). Its mapping vector is

$$\mathbf{x}_{8,1\_7} = \sqrt{\frac{8}{7}} \cdot \left(0, 1, \mathrm{e}^{\mathrm{j}\frac{2\pi}{7}}, \mathrm{e}^{\mathrm{j}\frac{4\pi}{7}}, \mathrm{e}^{\mathrm{j}\frac{6\pi}{7}}, \mathrm{e}^{-\mathrm{j}\frac{2\pi}{7}}, \mathrm{e}^{-\mathrm{j}\frac{4\pi}{7}}, \mathrm{e}^{-\mathrm{j}\frac{6\pi}{7}}\right). \tag{3.20}$$

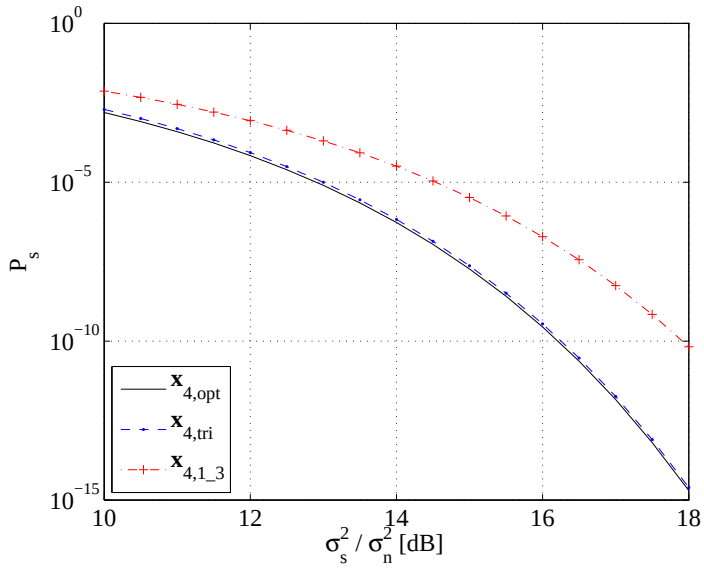
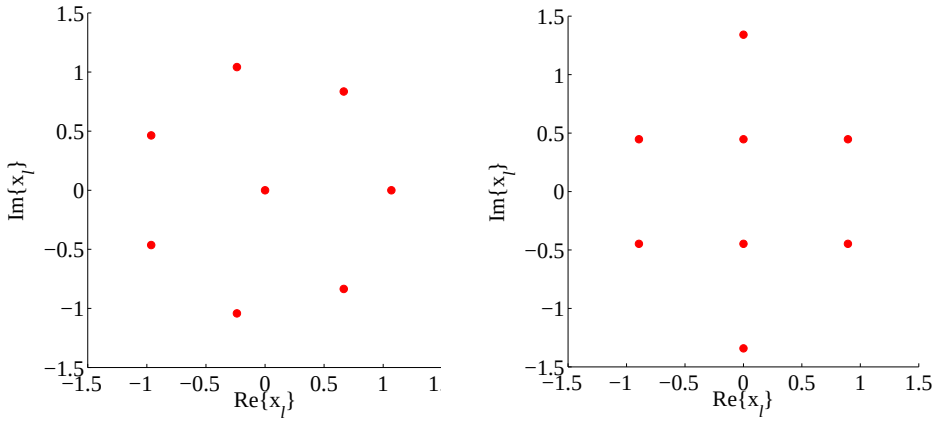


Figure 3.5: SER from union bound for 4-QAM candidates

The mapping from Figure 3.6(b) is an 8-QAM cross constellation $\mathbf{x}_{8,\text{cro}}$ with equidistant symbols, and will act as a candidate for comparison. The corresponding vector equals

$$\mathbf{x}_{8,\text{cro}} = \frac{2}{\sqrt{5}} \cdot \left(-1-0.5\text{j}, -1+0.5\text{j}, -0.5\text{j}, 0.5\text{j}, 1-0.5\text{j}, 1+0.5\text{j}, -1.5\text{j}, 1.5\text{j}\right). \tag{3.21}$$



(a) Optimum 8-QAM for $\frac{\sigma_s^2}{\sigma_n^2} = 10\,\text{dB}$ (b) $\mathbf{x}_{8,\text{cro}}$ mapping

Figure 3.6: Two candidates for 8-QAM constellations

If we increase the SNR, the rightmost symbol from $\mathbf{x}_{8,1\_7}$, i.e., $\sqrt{\frac{8}{7}}$, moves further to the right. In return, the magnitudes of all other symbols shrink to satisfy the power constraint. Moreover, the symbols move towards a lattice, which consists of equilateral triangles. For the high SNR regime and the limit of $L \to \infty$, the optimum can be proven to possess the

structure of equilateral triangles [106]. The resulting hexagonal (or honeycomb) lattice provides densest packing capabilities [109, 110]. The tendency towards this lattice can be seen in Figure 3.7, where we depict the optimum constellations for $\frac{\sigma_s^2}{\sigma_n^2} = 14$ and 20 dB, the proposed constellation from [106], and the one, which is made up of equilateral triangles. This mapping vector is calculated as follows: construct any (non-zero mean) lattice $\mathbf{x}_{8,\text{tri,nz}}$ of equilateral triangles with arbitrary edge length $d_{\min}$, e.g., let $x_0 = 0$, followed by the hexagon $x_l = d_{\min} \cdot \exp\left(\mathrm{j}\left[\frac{\pi}{3} \cdot (l-1) + \frac{\pi}{6}\right]\right)$ for $l \in \{1,\ldots,6\}$, and finally place, e.g., $x_7 = \sqrt{3} \cdot d_{\min}$. Then subtract the center of gravity, which equals $\frac{\sqrt{3}}{8} \cdot d_{\min}$, to force zero-mean. This yields the zero mean vector $\mathbf{x}_{8,\text{tri}} = \mathbf{x}_{8,\text{tri,nz}} - \frac{\sqrt{3}}{8} \cdot d_{\min} \cdot \mathbf{1}_8$. Now, apply (3.4) to compute $d_{\min}$. To do so, we note that the variances of $\mathbf{x}_{8,\text{tri,nz}}$ and $\mathbf{x}_{8,\text{tri}}$ are the same. The latter is the desired symbol power, the former is the power of $\mathbf{x}_{8,\text{tri,nz}}$ minus the squared magnitude of its center of gravity. Thus, we have $\mathscr{P}_x = \frac{d_{\min}^2}{8}(0+6\cdot 1+3) - \left|\frac{\sqrt{3}}{8} \cdot d_{\min}\right|^2 = 1$, which is fulfilled for $d_{\min} = \frac{8}{\sqrt{69}}$. This results in

$$\mathbf{x}_{8,\text{tri}} = \frac{8}{\sqrt{69}} \cdot \left[\left(0, \mathrm{e}^{\mathrm{j}\frac{\pi}{6}}, \mathrm{j}, \mathrm{e}^{\mathrm{j}\frac{5\pi}{6}}, -\mathrm{e}^{\mathrm{j}\frac{\pi}{6}}, -\mathrm{j}, \mathrm{e}^{-\mathrm{j}\frac{\pi}{6}}, \sqrt{3}\right) - \frac{\sqrt{3}}{8} \cdot \mathbf{1}_8\right]. \tag{3.22}$$

The vectors of the other mappings are summarized in the Appendix B.1. The resulting SER from (3.16) for the optimized mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 10\,\text{dB}$, which equals $\mathbf{x}_{8,1\_7}$, as well as for $\frac{\sigma_s^2}{\sigma_n^2} = 14$ and 20 dB is plotted in Figure 3.8. We also computed the SER for $\mathbf{x}_{8,\text{Fos}}$, $\mathbf{x}_{8,\text{tri}}$, $\mathbf{x}_{8,\text{cro}}$, and the 8-PSK from Figure 2.7(a). The latter performs about 2 dB worse (at $P_s = 10^{-6}$) than all optimized mappings, even though it is the most commonly applied constellation for $L = 8$ symbols. The 8-QAM cross constellation is only about 0.5 dB worse than the optimized mappings, due to its denser packing of symbols compared to 8-PSK. The proposed optimized mappings differ only slightly. However, from the zoomed sections in Figure 3.8, it can be observed that each mapping is best at its target SNR. For $\frac{\sigma_s^2}{\sigma_n^2} = 25\,\text{dB}$, the lattice constellation from $\mathbf{x}_{8,\text{tri}}$ surpasses the mapping, which is optimized at 20 dB, by a small fraction ($6.49 \cdot 10^{-27}$ vs. $6.41 \cdot 10^{-27}$, which means that the SER is already negligible).

For low SNR, the $\mathbf{x}_{8,1\_7}$ mapping offers the advantages of a small multiplicity (seven) of its minimum distance $d_{\min} = 2\sin(\pi/7) \approx 0.869$. Further, a demapper might only distinguish between two amplitude and seven phase levels, allowing for simple realizations. At higher SNR, $\mathbf{x}_{8,\text{tri}}$ benefits from its large $d_{\min} = \frac{8}{\sqrt{69}} \approx 0.963$, which outweighs its rather large multiplicity of 14. Although both $\mathbf{x}_{8,\text{tri}}$ and $\mathbf{x}_{8,\text{Fos}}$ perform close to optimum for all SNR regions, the former mapping can be constructed more conveniently, see (3.22).

3.2.3 16-QAM for Minimum Symbol Error Rate

As the proof from [106] suggests, the optimum constellations with respect to SER tend at large SNR towards lattices of equilateral triangles, if the number of symbols L increases.

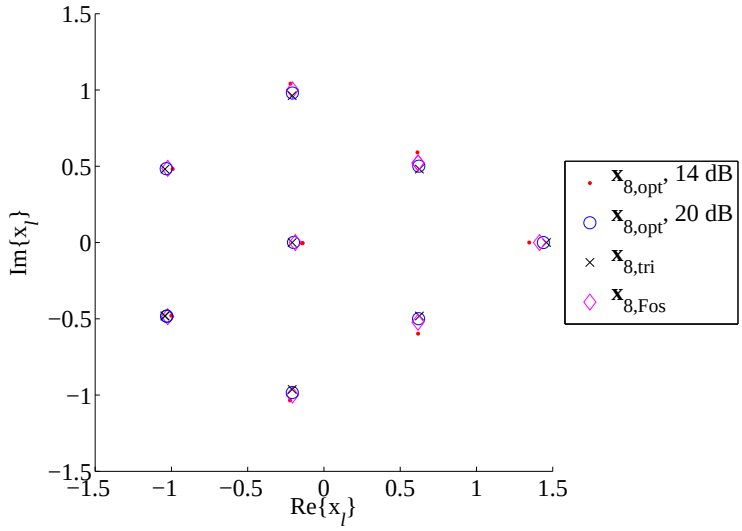


Figure 3.7: Optimum 8-QAM for $\frac{\sigma_s^2}{\sigma_n^2} = 14$ and 20 dB, and two candidates

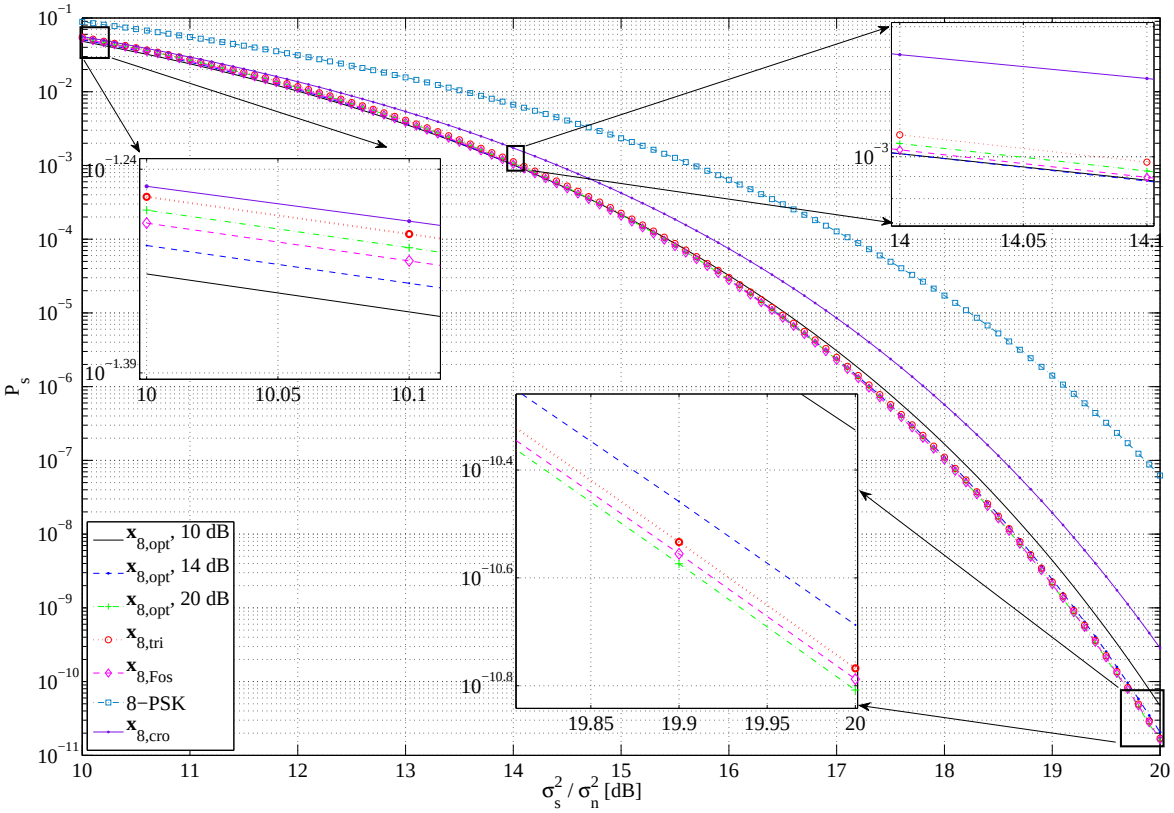


Figure 3.8: SER from union bound for 8-QAM candidates

This can already be observed for $L = 16$, also for small-to-medium SNR, as we will show in this subsection. To avoid confusion, we denote the mapping from Figure 2.7(b) as a *regular* 16-QAM constellation. Let us first construct an ideal lattice of equilateral triangles, in which

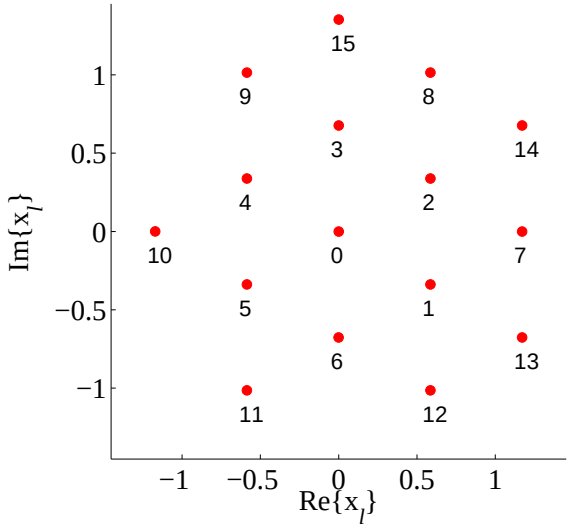


Figure 3.9: Lattice of equilateral triangles, $\mathbf{x}_{16,\text{tri,nz}}$, and symbol indices l

the mean is not yet forced to zero, denote it as $\mathbf{x}_{16,\text{tri,nz}}$ and its entries as $x_l, l \in \{0,\ldots,15\}$. From basic geometry, we find

$$x_l = d_{\min} \cdot \begin{cases} 0 & , \quad \text{for } l = 0 \\ \exp\left(\mathrm{j}\left[\frac{\pi}{3}\cdot(l-1) - \frac{\pi}{6}\right]\right) & , \quad \text{for } l \in \{1,\ldots,6\} \\ \sqrt{3}\cdot\exp\left(\mathrm{j}\frac{\pi}{3}\cdot(l-7)\right) & , \quad \text{for } l \in \{7,\ldots,12\} \\ 2\cdot\exp\left(\mathrm{j}\left[\frac{\pi}{3}\cdot(l-13) - \frac{\pi}{6}\right]\right) & , \quad \text{for } l \in \{13,\ldots,15\} \end{cases}, \tag{3.23}$$

which is depicted in Figure 3.9. Its mean value is easily seen to be $d_{\min}\cdot\left(\sqrt{3}+\mathrm{j}\right)/8$. After subtracting the mean from all components, we obtain $\mathbf{x}_{16,\text{tri}}$, see Figure 3.10,

$$\mathbf{x}_{16,\text{tri}} = \mathbf{x}_{16,\text{tri,nz}} - d_{\min}\cdot\frac{\sqrt{3}+\mathrm{j}}{8}\cdot\mathbf{1}_{16}. \tag{3.24}$$

Applying (3.4), yields the minimum Euclidean distance of this mapping as $d_{\min} = \frac{4}{\sqrt{35}} \approx 0.676$. In [111], triangular lattice mappings were proposed, but the symbol constellation is suboptimum. There, the 16 symbols are located on a hexagonal grid with four rows, consisting each of four symbols. In our constellation in Figure 3.10, three symbols are aligned on a (skew) row, beneath which are three rows, each with four symbols and one symbol is located at the bottom. In the non-zero mean lattice of Figure 3.9, the uppermost (skew) row consists of symbols x_{15}, x_8, and x_{14}. The minimum Euclidean distance in [111] is only $d_{\min} = \frac{2}{3} \approx 0.667$, which is smaller than for our dense packing lattice. Placing one of the four symbols from the uppermost row to the bottom and correcting the mean value with (3.6) would transform the mapping from [111] to our mapping.

The optimized constellations $\mathbf{x}_{16,\text{opt}}$ for $\frac{\sigma_s^2}{\sigma_n^2} = 10$, 14, and 20 dB are plotted in Figure 3.10 together with $\mathbf{x}_{16,\text{tri}}$ and the baseline mapping $\mathbf{x}_{16,\text{Fos}}$. This latter mapping was optimized

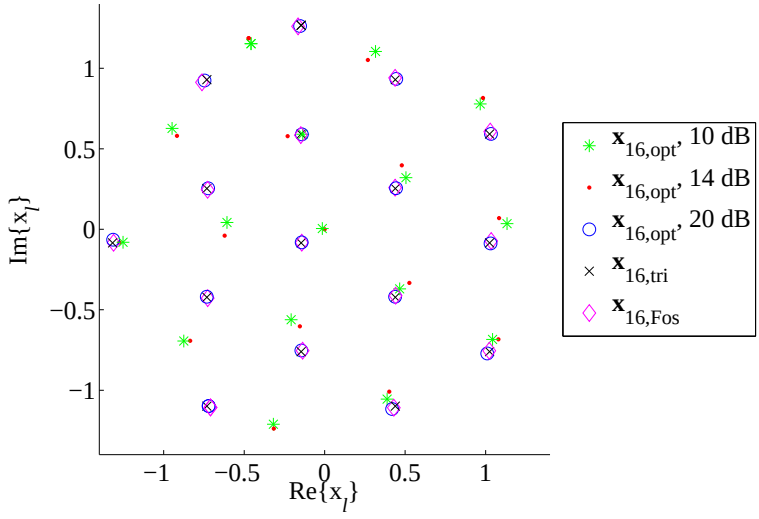


Figure 3.10: Optimum 16-QAM for $\frac{\sigma_s^2}{\sigma_n^2} = 10$, 14 and 20 dB, and two candidates

for high SNR and is thus almost identical to the ideal lattice of $\mathbf{x}_{16,\text{tri}}$. Also the optimized 16-QAM for 20 dB is very similar to $\mathbf{x}_{16,\text{tri}}$. However, the vector $\mathbf{x}_{16,\text{tri}}$ further evolves towards $\mathbf{x}_{16,\text{opt}}$ during the optimization process at 20 dB. This is signaled by the NECs of both mappings. $\mathbf{x}_{16,\text{tri}}$ has $\Lambda \approx 10.56$, while $\mathbf{x}_{16,\text{opt}}$ has a relatively small value of $\Lambda \approx 1.01$. The optimized constellations for the three SNR values are specified in the Appendix B.1.

SER over SNR performances of the 16-QAM mappings are summarized in Figure 3.11. The regular 16-QAM mapping is about 0.5 dB worse than the proposed constellations for SER below 10^{-3}. From Figure 2.7(b), we see that its minimum distance is $2\Theta = \frac{2}{\sqrt{10}} \approx 0.633$. Again, it can be seen that each mapping yields the smallest SER at its target SNR. The ideal lattice $\mathbf{x}_{16,\text{tri}}$ with $d_{\min} \approx 0.676$ performs well over a large SNR range, and offers a closed-form expression from (3.23) and (3.24). Moreover, from 24 dB on, it achieves the smallest SER among all constellations (not shown).

3.2.4 32-QAM for Minimum Symbol Error Rate

We first discuss two often applied constellations for 32-QAM. The 32-QAM cross constellation $\mathbf{x}_{32,\text{cro}}$ in Figure 3.12(a) has a denser packing of the symbols than the rectangular constellation $\mathbf{x}_{32,\text{rec}}$ from Figure 3.12(b), and is the most frequently employed 32-QAM constellation. Both candidates exhibit an orthogonal lattice structure, with equidistant symbols. The minimum distance is $d_{\min} = \frac{4}{\sqrt{80}} \approx 0.447$ for the cross constellation, but only $\frac{4}{\sqrt{104}} \approx 0.392$ for the rectangular arrangement. Both mappings are easily constructed by observing that, e.g., $x_{27} = d_{\min}/2 \cdot (1+\mathrm{j})$, and that all other symbols are equidistant in both dimensions.

For the high SNR regime, a hexagonal lattice structure $\mathbf{x}_{32,\text{tri}}$ turns out to be again the optimum constellation with respect to SER. Similar to (3.23), we construct this lattice based on

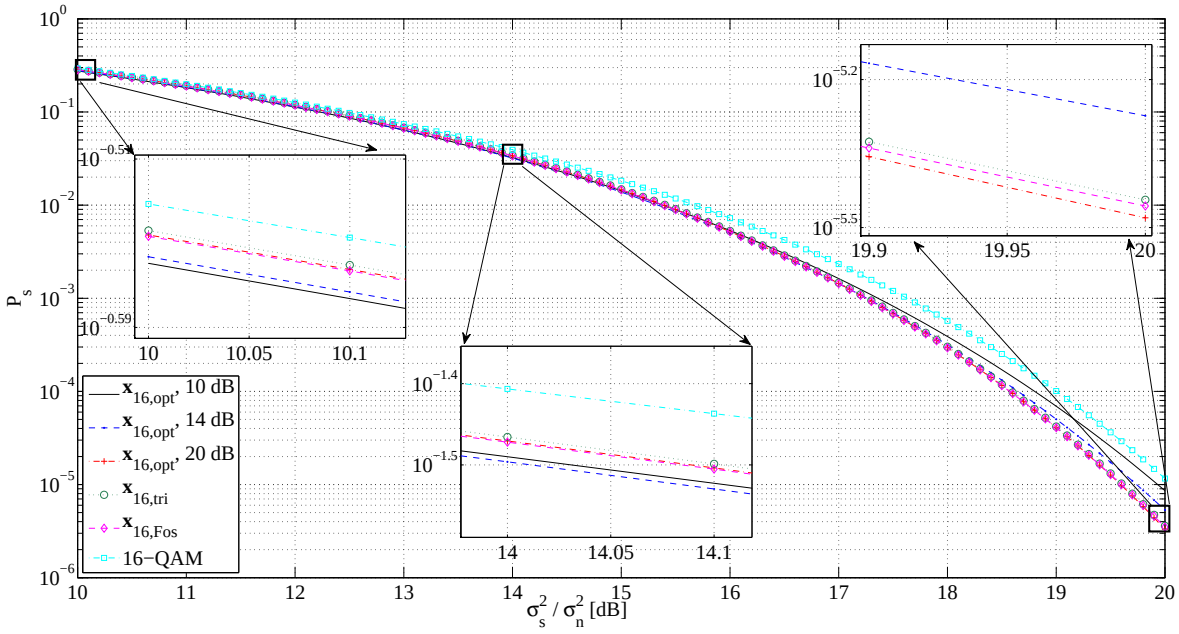
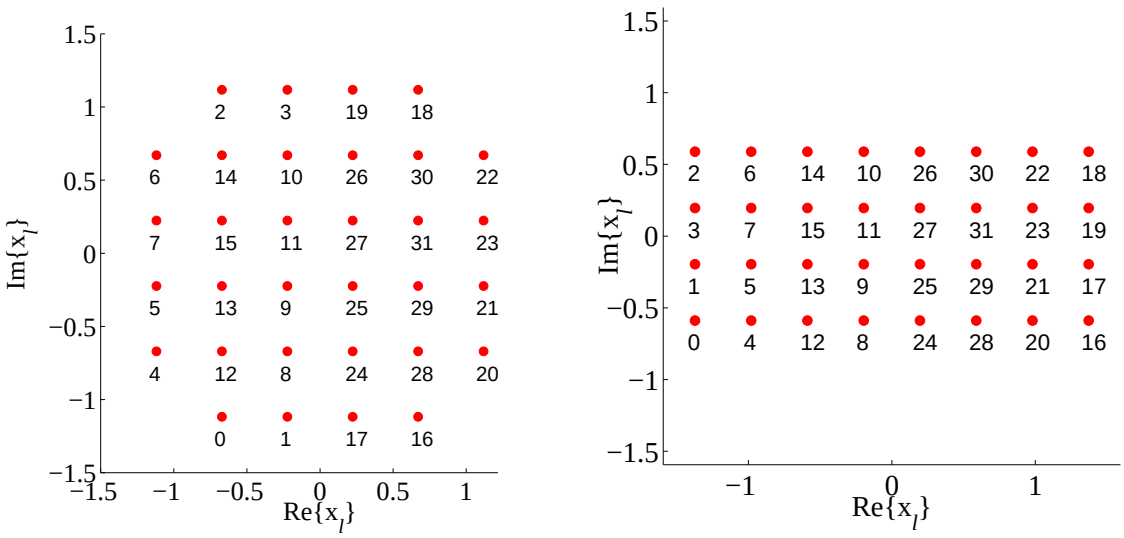


Figure 3.11: SER from union bound for 16-QAM candidates



(a) $\mathbf{x}_{32,\text{cro}}$ mapping

(b) $\mathbf{x}_{32,\text{rec}}$ mapping

Figure 3.12: Two candidates for 32-QAM constellations and symbol indices l

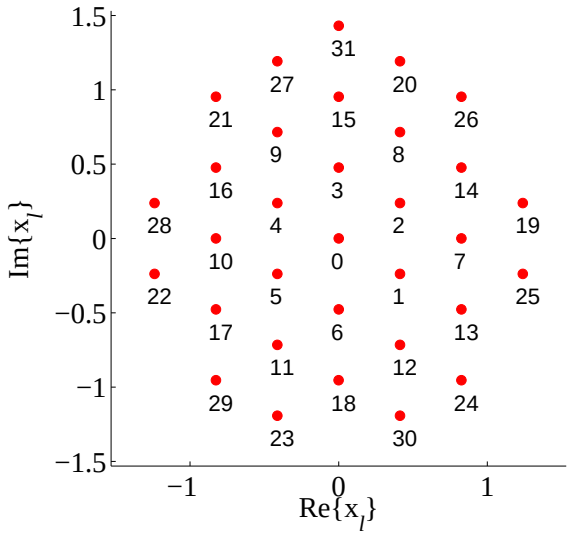


Figure 3.13: Lattice of equilateral triangles, $\mathbf{x}_{32,\text{tri,nz}}$, and symbol indices l

a lattice $\mathbf{x}_{32,\text{tri,nz}}$ of non-zero mean, with elements

$$x_l = d_{\min} \cdot \begin{cases} 0 & , \quad \text{for } l = 0 \\ \exp\left(\mathrm{j}\left[\frac{\pi}{3}\cdot(l-1)-\frac{\pi}{6}\right]\right) & , \quad \text{for } l \in \{1,\ldots,6\} \\ \sqrt{3}\cdot\exp\left(\mathrm{j}\frac{\pi}{3}\cdot(l-7)\right) & , \quad \text{for } l \in \{7,\ldots,12\} \\ 2\cdot\exp\left(\mathrm{j}\left[\frac{\pi}{3}\cdot(l-13)-\frac{\pi}{6}\right]\right) & , \quad \text{for } l \in \{13,\ldots,18\} \\ \left(\sqrt{3}+\mathrm{e}^{\mathrm{j}\frac{\pi}{6}}\right)\cdot\exp\left(\mathrm{j}\frac{\pi}{3}\cdot(l-19)\right) & , \quad \text{for } l \in \{19,\ldots,24\} \\ \left(\sqrt{3}+\mathrm{e}^{-\mathrm{j}\frac{\pi}{6}}\right)\cdot\exp\left(\mathrm{j}\frac{\pi}{3}\cdot(l-25)\right) & , \quad \text{for } l \in \{25,\ldots,30\} \\ 3\mathrm{j} & , \quad \text{for } l = 31 \end{cases}, \tag{3.25}$$

which are shown in Figure 3.13. The symbols $x_l, l \in \{0,\ldots,30\}$, add up to zero, due to symmetry. Therefore, the mean of $\mathbf{x}_{32,\text{tri,nz}}$ is $\frac{x_{31}}{32}$. This yields $\mathbf{x}_{32,\text{tri}}$ as

$$\mathbf{x}_{32,\text{tri}} = \mathbf{x}_{32,\text{tri,nz}} - d_{\min} \cdot \frac{3\mathrm{j}}{32} \cdot \mathbf{1}_{32}, \tag{3.26}$$

and from (3.4) follows the edge length $d_{\min} = \frac{32}{\sqrt{4503}} \approx 0.477$. This minimum Euclidean distance is larger than that from both $\mathbf{x}_{32,\text{cro}}$ and $\mathbf{x}_{32,\text{rec}}$.

In order to achieve a reasonable SER, we increased the target SNRs to $\frac{\sigma_s^2}{\sigma_n^2} = 14$, 20 and 24 dB. Optimum mappings for each SNR are denoted as $\mathbf{x}_{32,\text{opt}}$, itemized in the Appendix B.1, and can be seen in Figure 3.14. Only $\mathbf{x}_{32,\text{opt}}$ for $\frac{\sigma_s^2}{\sigma_n^2} = 14$ differs noticeably from the hexagonal lattice $\mathbf{x}_{32,\text{tri}}$, which is also shown. For higher SNR, the optimum constellations approach $\mathbf{x}_{32,\text{tri}}$ quite close. However, the hexagonal lattice is still slightly inferior than the

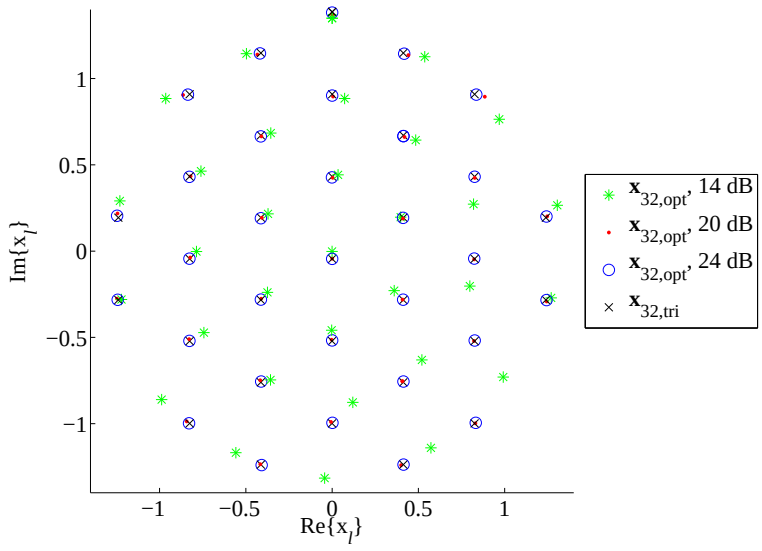


Figure 3.14: Optimum 32-QAM for different SNR and $\mathbf{x}_{32,\text{tri}}$

optimized mappings until the SNR exceeds 29 dB. In [106], no baseline mapping for $L = 32$ was given.

The resulting SER performance for all proposed constellations is plotted in Figure 3.15. Obviously, the rectangular constellation performs worst (about 1.5 dB penalty) for all SNR, while the standard 32-QAM cross constellation only yields about 0.5 dB degradation compared to optimized constellations. While the optimized mapping for 14 dB offers minimum SER at this SNR, its SER degrades at higher SNRs, but is still smaller than the SER from the cross constellation. Again, the ideal hexagonal lattice $\mathbf{x}_{32,\text{tri}}$ offers close to optimum SER for the whole SNR range.

3.2.5 Summary for Minimum Symbol Error Rate Mappings

We have proposed L-ary QAM constellations for different target SNRs, which minimize the union bound SER from (3.16). This bound was shown to be tight, especially for large SNR. From its form, it is obvious that the SER is influenced by the overall distance spectrum. However, symbol pairs x_{l_1} and x_{l_2}, which have a large Euclidean distance, contribute only to a smaller amount than closest neighbor symbols at $d_{\min}$. Especially for the high SNR regime, all addends in (3.16) tend to zero, except the symbol pairs at $d_{\min}$. Thus, for large SNR, an optimum constellation should exhibit a maximum $d_{\min}$, which is achived by the hexagonal structures of $\mathbf{x}_{L,\text{tri}}$. The drawback of these lattices at smaller SNR is the rather large multiplicity of $d_{\min}$, because all adjacent symbols are spaced $d_{\min}$ apart. The same is true for regular QAM constellations, e.g., for the 32-QAM cross or rectangular constellation, but since these are *orthogonal* lattices, they possess a kissing number (number of closest neighbors) [109] of four. On the other hand, the kissing number of a hexagonal lattice is six, yielding the large multiplicity of $d_{\min}$. After applying the Chernoff bound [30] on the

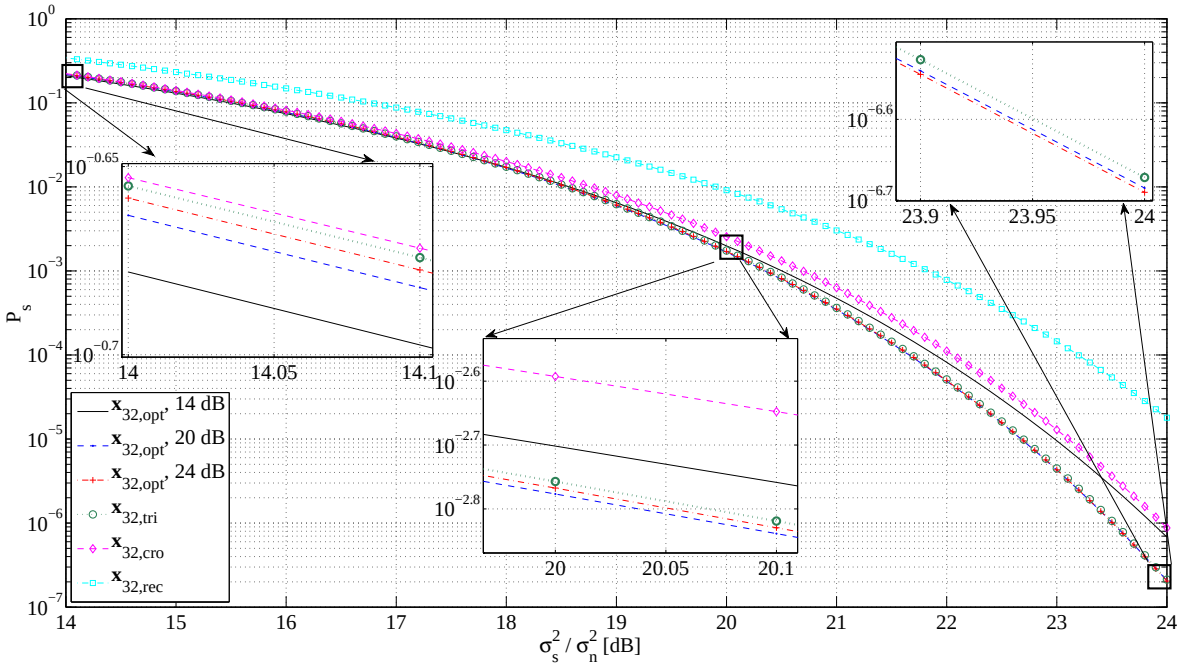


Figure 3.15: SER from union bound for 32-QAM candidates

erfc-function, it becomes clear that the SER decreases exponentially with $d_{\min}$, while its multiplicity has only a linear influence. Hence a large $d_{\min}$ outweighs its multiplicity at high SNR. For medium SNR, a tradeoff between minimum Euclidean distance and its multiplicity has to be achieved and for small SNR, the influence of all other neighbors contribute more and more to the overall SER. Yet, we have shown, that the ideal hexagonal structures $\mathbf{x}_{L,\text{tri}}$ are close to optimum for all SNR values. Another advantage of these constellations is their closed-form mapping vector, for which we have outlined the construction rules.

3.3 Mappings for Maximum BICM Capacity

3.3.1 Relation between Bit Error Rate and BICM Capacity

The BER P_b, which results from optimum ML demapping without a priori knowledge, is in general difficult to compute. The demapper applies the MAP decision rule from (2.56), or equivalently, computes the extrinsic L-value from (2.64) and performs a hard decision based on its sign. Note that the MAP rule reduces to ML detection, if no a priori information is available. A simplified demapper may employ the *max-log approximation* [89] in (2.64), i.e.,

$$\ln\left(\mathrm{e}^{a_1} + \mathrm{e}^{a_2}\right) \approx \max\ \{a_1, a_2\}\,. \tag{3.27}$$

The following hard decision reduces this max-log ML demapping to the search for the closest neighbor x_l to the received sample r_k, which is similar to the symbol detection from

Section 3.2. Hence, we can derive from (3.15) an approximation of the BER for the sub-optimum max-log demapper. Based on a particular symbol error, we need to consider the number of involved bit errors. However, we face the same problem that for arbitrary constellations $\mathbb{X}$, the integration area $\mathscr{A}_{l_1}$ is hard to determine. We resort again to the union bound and adjust (3.16) to obtain the upper bound on BER

$$P_\mathrm{b} \lessapprox \frac{1}{2L} \sum_{l_1=0}^{L-1} \sum_{\substack{l_2=0 \\ l_2 \neq l_1}}^{L-1} \operatorname{erfc}\left(\frac{\left| x_{l_2} - x_{l_1} \right|}{2\sqrt{2}\sigma_\mathrm{n}} \right) \cdot \frac{\mathrm{d_H}\left(x_{l_1}, x_{l_2}\right)}{M}. \tag{3.28}$$

The Hamming distance $\mathrm{d_H}\left(x_{l_1}, x_{l_2}\right)$ between symbols x_{l_1} and x_{l_2} is defined as the number of bit positions, in which the labels of both symbols differ, i.e., between $\mu^{-1}\left(x_{l_1}\right)$ and $\mu^{-1}\left(x_{l_2}\right)$.

Example: The Hamming distance $\mathrm{d_H}\left(x_1, x_6\right)$ is with our nomenclature from (3.1) the difference in bit positions between 001 and 110, thus $\mathrm{d_H}\left(x_1, x_6\right) = 3$. ■

In (3.28), the multiplication by $\mathrm{d_H}\left(x_{l_1}, x_{l_2}\right)/M$ weights each PEP by the number of erroneous bits per symbol. Note that the BER for a max-log demapper is larger than or at least equal to the BER from optimum ML demapping. We now discuss an alternative approach to compute BER for ML demapping.

Recall from (2.83) that the leftmost point from the demapper's transfer characteristic is related to the BICM capacity, i.e., $(I_{\mathrm{A}1}, I_{\mathrm{E}1}) = \left(0, C_\mathrm{B}/M\right)$. Inserting this point into (2.82) yields the BER after optimum ML demapping for no a priori information. Dropping the index "1", we obtain the relation

$$P_\mathrm{b} \approx \frac{1}{2} \operatorname{erfc}\left(\frac{J^{-1}\left(C_\mathrm{B}/M\right)}{2\sqrt{2}} \right). \tag{3.29}$$

The approximation comes from the assumption that the extrinsic L-values are Gaussian distributed. This is in general not the case, but — as we have already mentioned, the MI is robust against changes in the underlying distribution. It is worth mentioning that (3.29) can be computed *without* knowledge of optimum decision regions $\mathscr{A}_{l_1}$. Note that the erfc-function is monotonically decreasing, while its argument —the inverse J-function, is increasing. This means that the BER decreases monotonically with increasing BICM capacity C_B. Therefore, mapping vectors, which maximize C_B, also minimize BER P_b, and we do not have to cover the BER as an additional cost function.

Remarks on inverse J-function: In [96, 112], closed-form approximations for both $J(\sigma_\mathrm{A})$ and its inverse $J^{-1}(I)$ are given. However, if C_B/M approaches 1, these approximations are too coarse. We circumvent these numerical problems as follows: First note that from its definition, $J(\sigma_\mathrm{A})$ corresponds to the signal set (or BICM) capacity of BPSK signaling over an AWGN channel with (in-phase) noise variance $\sigma_{\mathrm{n_I}}^2 = 4/\sigma_\mathrm{A}^2$, [96, 88]. Thus, we compute $J(\sigma_\mathrm{A})$ with (2.77) for an arbitrarily fine grained scale of σ_A and store the results in a table. The inverse function $J^{-1}(I)$ is then obtained from this look-up table with arbitrary precision.

Example: We have simulated the BER for optimum ML and suboptimum max-log ML demapping for three different mappings: QPSK Gray mapping, $\mathbf{x}_{4,1\_3}$ as defined in (3.19) and the 8-QAM, which was optimized for 14 dB with respect to SER, cf. (B.1). Note that the last two mappings are anti-Gray mappings. The results are depicted in Figure 3.16, together with analytical results from (3.29), which holds for optimum ML demapping, and from the union bound (3.28) for the max-log approximation. It can be observed that the closed-form approximations for both demapper types are tight. Optimum demapping hardly differs from its suboptimum counterpart for the $\mathbf{x}_{4,1\_3}$ mapping, because only one addend in the numerator and denominator of (2.64) is neglected. The same is true for QPSK, but at higher SNR, the difference in BER becomes apparent. Obviously, the Gaussian assumption for (3.29) yields accurate BER prediction, emphasizing again the robustness of the MI.■

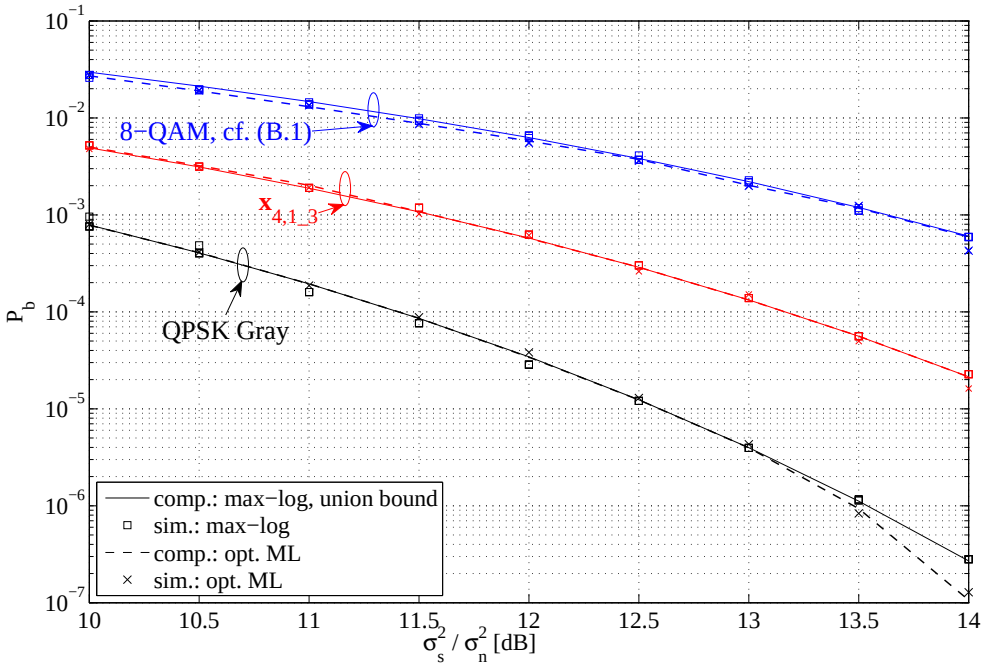


Figure 3.16: max-log ML BER from union bound, cf. (3.28) (—) and simulation (□), and optimum ML BER, analytically, cf. (3.29) (- - -) and simulation (×)

3.3.2 Enhanced Optimization Algorithm

Let us first reformulate the BICM capacity and determine its partial derivative. From (2.78), we can extract the addend $\log_2 L = M$ from the logarithm. Averaging over this constant and applying the law of total probability leaves just this factor. If we use a new index l_1 for summation, drop discrete-time index k, and introduce the short hand notation $p\left(r\left|x_{l_1}\right.\right)$ for

$p\left(r\,|s=x_{l_1}\right)$, we can write

$$C_{\mathrm{B}} = M - \frac{1}{L}\int\limits_{r\in\mathbb{C}} \sum_{m=0}^{M-1}\sum_{b=0}^{1}\sum_{x_{l_1}\in\mathbb{X}_b^m} p\left(r\,|x_{l_1}\right)\log_2\frac{L\cdot p(r)}{\sum\limits_{x_{l_1}\in\mathbb{X}_b^m} p\left(r\,|x_{l_1}\right)}\,dr. \tag{3.30}$$

Based on this expression, we derive in the Appendix A.2 the l-th component of its gradient $\nabla C_{\mathrm{B}}(\mathbf{x})$, which is expressed as

$$\begin{aligned}\frac{\partial C_{\mathrm{B}}(\mathbf{x})}{\partial x_l} &= \frac{2}{L}\int\limits_{r\in\mathbb{C}} p\left(r\,|x_l\right)\cdot\frac{x_l - r}{\sigma_{\mathrm{n}}^2}\Bigg\{M\cdot\log_2\left(L\cdot p(r)\right)\\ &\quad -\sum_{m=0}^{M-1}\log_2\left(\sum_{x_{l_1}\in\mathbb{X}_{b_{l,m}}^m} p\left(r\,|x_{l_1}\right)\right)\Bigg\}\,dr,\end{aligned} \tag{3.31}$$

where $b_{l,m}$ denotes the m-th bit label of symbol x_l, i.e., $b_{l,m} = \mu_m^{-1}(x_l)$. From the summation over subsets $\mathbb{X}_b^m$, it is evident that the cost function $f = C_{\mathrm{B}}$ depends on both alphabet $\mathbb{X}$ and bit labeling μ, such that the order of the elements in $\mathbf{x}$ is of importance. The mapping $\mathbf{x}$ should possess a "Gray-like" bit labeling [13], which is easily understood from a pragmatic point of view. During the optimization algorithm from Subsection 3.1.2, however, many of the randomly chosen start vectors $\mathbf{x}_0$ violate this condition and yield a rather low initialization cost function. Therefore, we enhance the algorithm by bootstrapping the initialization step as follows.

We take as an heuristic starting vector a regular Gray mapping, i.e., one which is typically applied in the literature, and denote it as $\mathbf{x}_{\mathrm{G}}$. For $L \in \{4,8,16\}$, $\mathbf{x}_{\mathrm{G}}$ was already introduced in Subsection 2.1.3, while for $L = 32$, we choose the 32-QAM rectangular Gray mapping from Figure 3.12(b), $\mathbf{x}_{\mathrm{G}} = \mathbf{x}_{32,\mathrm{rec}}$. Note that there exists no pure Gray labeling for the 32-QAM cross constellation [113], as can be seen in Figure 3.12(a), where the four upper- and lowermost symbols differ from their vertical neighbors in two bits. Next, we shuffle a random vector $\mathbf{x}_{\mathrm{M}}$, from which real and imaginary part of each component are drawn independently from a Gaussian distribution (as opposed to the uniform distribution in the original algorithm) with zero mean and variance σ_{M}^2. Adding this vector to the Gray mapping, we "mutate" the start vector, which yields $\mathbf{x}_0' = \mathbf{x}_{\mathrm{G}} + \mathbf{x}_{\mathrm{M}}$. As in the last two initialization steps from the original algorithm, we also shift the center of gravity to the origin and normalize the transmit power to obtain $\mathbf{x}_0$. Now, we compare its cost function $f(\mathbf{x}_0)$ with the current optimum f_{opt}, which was found during the previous trials (we set $f_{\mathrm{opt}} = f(\mathbf{x}_{\mathrm{G}})$ in the first trail). If $f(\mathbf{x}_0) < \chi\cdot f_{\mathrm{opt}}$, where $\chi \in [0,1]$ is a certain threshold, we discard this trail and repeat the procedure with a new mutation vector $\mathbf{x}_{\mathrm{M}}$. On the other hand, if a trail results in a new (local) optimum f_{opt}, the initialization vectors for the next trials will be compared with this larger cost function. Hence, the probability that $\mathbf{x}_0$ will be discarded, increases, such that only more promising starting vectors will be optimized by the gradient search algorithm. This idea comes from the *great deluge algorithm* [114], where the rising flood level

corresponds to the comparison with an increasing cost function. By varying the degree of mutation and the expanse preselection by means of σ_M^2 and χ, respectively, we can find an appropriate tradeoff between computation time and amount of optimization results.

In the next subsection, we discuss optimized mappings with respect to the BICM capacity for $L \in \{8, 16, 32\}$. For 4-QAM, the QPSK Gray mapping is obviously the best choice for all SNR values. Its capacity is twice as large as that from BPSK signaling, as can be seen in Figure 2.17. This is because the QPSK Gray mapping is made up of two BPSKs on both components, as we have discussed in Subsection 2.1.3. From the fact that BPSK is the optimum 2-QAM constellation, follows the optimality of the QPSK Gray mapping. For the other mappings, we will consider three target SNRs $\frac{\sigma_s^2}{\sigma_n^2}$ for the optimization. Assume, the L-ary QAM should be applied in a (non-iterative) BICM scheme, with an outer encoder of rate R_c, such that the spectral efficiency $\eta = R_c \cdot M$ is slightly smaller than the BICM capacity, i.e., $R_c \cdot M \lessapprox C_B$. From the EXIT chart, it is clear, that a capacity approaching outer code should possess a transfer characteristic, which is close to a step-function: for $C_B/M = I_{A2} < R_c$, the output MI I_{E2} is almost zero, while for $C_B/M = I_{A2} \gtrapprox R_c$, it jumps to almost one [96]. Further assume that we have this outer code available and want to design a capacity achieving BICM system with the standard Gray mapping $\mathbf{x}_G$, then for each R_c, we obtain the target SNR by equating

$$C_B\left(\mathbf{x}_G, \frac{\sigma_s^2}{\sigma_n^2}\right) = R_c \cdot M. \tag{3.32}$$

We focus on $R_c \in \{0.5, 0.75, 0.999\}$. From the discussion of the previous subsection, it is interesting to note that a capacity approaching code has to act as a "BER downsizer". If a hard decision is applied after the demapper, a BER P_b results, that depends on code rate R_c according to (3.29), assuming equality in (3.32). The following decoder thus must be able to decrease this input BER down to an arbitrarily small output BER. For $R_c = 0.5$, the input BER is $P_b \approx 0.153$, while for the higher code rate, $R_c = 0.75$, less errors have to be corrected, because $P_b \approx 0.071$. The last code rate of $R_c = 0.999$ corresponds a channel encoder, which adds just 0.1% redundancy. The expected BER of the demapper is $P_b \approx 2.39 \cdot 10^{-4}$. We remark, however, that in all cases, *soft* extrinsic information must be fed forward to the decoder to allow for capacity approaching performance. Hence, the BER is not an accurate quantity for describing the reliability of the soft information from the demapper.

3.3.3 8-QAM for Maximum BICM Capacity

Consider first an 8-QAM mapping that maximizes BICM capacity and which should be employed in conjunction with a rate $R_c = 0.5$ encoder. Thus, the spectral efficiency is 1.5. The reference 8-PSK Gray mapping $\mathbf{x}_G$ requires an SNR of $\frac{\sigma_s^2}{\sigma_n^2} = 3.241\,\text{dB}$ to achieve $C_B = 1.5$. In order to verify our results, Figure 3.17 shows the simulated transfer characteristic T_1 of the 8-PSK Gray mapping at this target SNR. It can be seen that the starting point of T_1

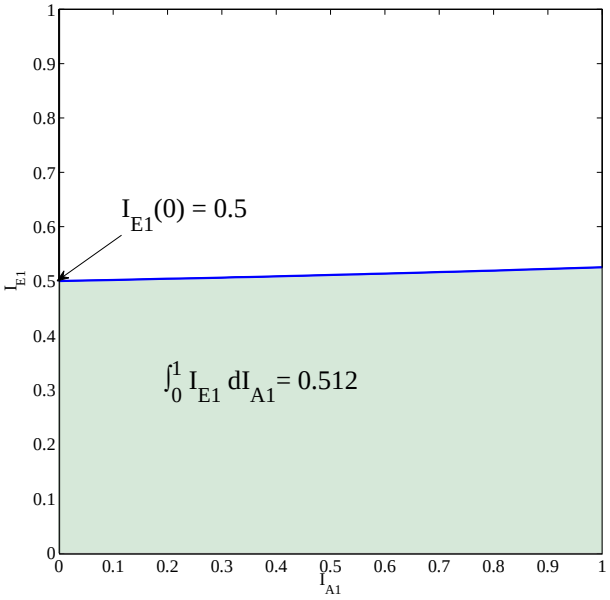


Figure 3.17: Transfer characteristic of 8-PSK Gray mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 3.241\,\text{dB}$

equals C_B/M. Further, the area under T_i is 0.512, while the signal set capacity at this SNR is $C_\text{S} = 3 \cdot 0.513$, which is a close match.

For this target SNR value, we apply the enhanced optimization algorithm from Subsection 3.3.2 and find as the global optimum the following mapping

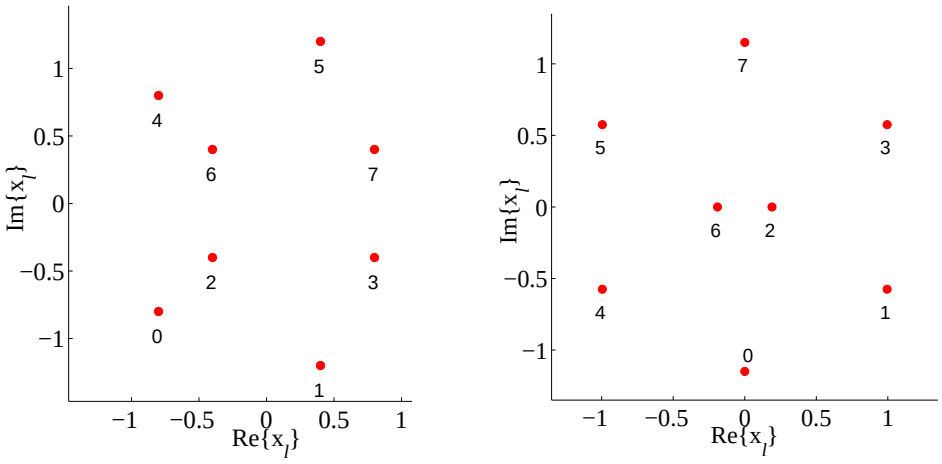
$$\mathbf{x}_{8,\text{opt}} = 0.4 \cdot \left(-2-2\text{j}, 1-3\text{j}, -1-\text{j}, 2-\text{j}, -2+2\text{j}, 1+3\text{j}, -1+\text{j}, 2+\text{j}\right), \tag{3.33}$$

which is depicted in Figure 3.18(a). It outperforms the standard 8-PSK Gray mapping to a small degree, because it achieves $C_\text{B} = 1.511$. Even though its minimum distance $d_\text{min} = 0.4 \cdot \sqrt{2} \approx 0.566$ is rather small, it has the advantage of small multiplicity of two (between x_0 and x_2 and between x_4 and x_6). Moreover, all other symbols are spaced further apart in return. Both mappings possess a small NEC — $\Lambda = 1.18$ for $\mathbf{x}_{8,\text{opt}}$, and $\Lambda = 1.01$ for $\mathbf{x}_\text{G}$, pointing out that these are local (and global) optima.

The next target SNR is $\frac{\sigma_s^2}{\sigma_n^2} = 7.216\,\text{dB}$, for which $\mathbf{x}_\text{G}$ achieves $C_\text{B} = 2.25$, i.e., we focus on BICM with a rate $R_\text{c} = 0.75$ encoder. The optimum 8-QAM is now a hexagon with a circumference of r_o and two symbols on a circle with inner radius $r_i = 0.19$. The outer radius follows from (3.4) as $r_o = \sqrt{(4-r_i^2)/3} \approx 1.15$. Thus, the mapping vector is

$$\mathbf{x}_{8,\text{opt}} = r_o \cdot \left(-\text{j}, \text{e}^{-\text{j}\frac{\pi}{6}}, r_i/r_o, \text{e}^{\text{j}\frac{\pi}{6}}, -\text{e}^{\text{j}\frac{\pi}{6}}, -\text{e}^{-\text{j}\frac{\pi}{6}}, -r_i/r_o, \text{j}\right), \tag{3.34}$$

and is shown in Figure 3.18(b). Again, the mapping possesses a small minimum distance ($2 \cdot 0.19 = 0.38$), but this time the multiplicity is only one. This mapping achieves $C_\text{B} \approx 2.300$. It is remarkable that even a 7-QAM, which results from setting $r_i = 0$, outperforms the standard 8-PSK Gray mapping, as it accomplishes $C_\text{B} = 2.291$ at this target SNR. We



(a) $\mathbf{x}_{8,\text{opt}}$ mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 3.241\,\text{dB}$ (b) $\mathbf{x}_{8,\text{opt}}$ mapping $\frac{\sigma_s^2}{\sigma_n^2} = 7.216\,\text{dB}$

Figure 3.18: 8-QAM mappings for maximum BICM capacity

can interpret this mapping either as a 7-QAM with non-equiprobable symbols or as an 8-QAM with equiprobable, but not all distinguishable symbols. Under the first perception, we observe that the six symbols at the vertices of the hexagon occur with probability $\frac{1}{8}$, while the innermost symbol (fusion of x_2 and x_6 from the original mapping in Figure 3.18(b)) appears with $\frac{2}{8}$. If the demapper interprets this mapping, however, as a degraded 8-QAM, it exhibits an error-floor of $P_\text{b} = 1/24 \approx 0.042$, as SNR approaches infinity. The reason is that two (x_2 and x_6) out of eight symbols are decided wrong with 50%, affecting one out of three bits. The symbols x_2 and x_6 have the last two bits in common, namely 10. With that we can compute the maximum BICM capacity of this 7-QAM with non-equiprobable symbols as follows: with probability $\frac{6}{8}$, three bits are transmitted per channel usage, as the equivocation diminishes to zero, while with $\frac{2}{8}$, only two bits (10) can be transmitted reliably. Thus, $C_{\text{B}\max} = \frac{6}{8} \cdot 3 + \frac{2}{8} \cdot 2 = \frac{11}{4} = 2.75$, which is still larger than the baseline capacity of 2.25.

Finally, the very high code rate $R_\text{c} = 0.999$ results in a spectral efficiency of $\eta = 2.997$. The reference Gray mapping $\mathbf{x}_\text{G}$ requires an SNR of $\frac{\sigma_s^2}{\sigma_n^2} = 15.810\,\text{dB}$ for $C_\text{B} = \eta$. Figure 3.19 shows the optimized mapping, whose vector is given explicitly in the Appendix B.2. Its constellation is quite close to the ones from Subsection 3.2.2, where the SER was minimized for medium to high SNR. However, the small difference in Figure 3.19 yields slightly better results. We also depict the ideal hexagonal lattice $\mathbf{x}_{8,\text{tri}}$ from (3.22), but with a different labeling, named quasi-Gray, because it is "Gray-like" for most of its symbols. Note that the hexagonal structure allows for pure Gray labeling only for 64-QAM or higher order QAMs, because the kissing number of the hexagonal lattice is six. Also the 8-QAM cross constellation from Figure 3.6(b) does not allow for Gray labeling and was thus excluded as a baseline mapping here.

The BICM capacities over a large range of SNR values are depicted in Figure 3.20 for various

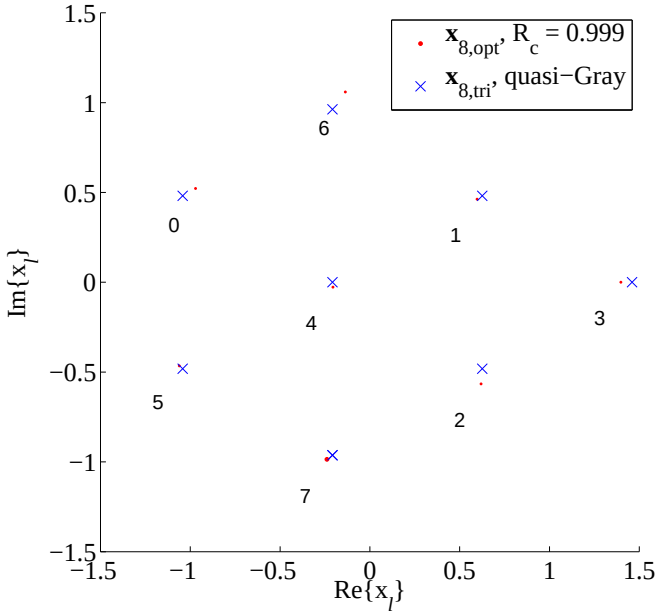


Figure 3.19: 8-QAM for maximum BICM capacity at $\frac{\sigma_s^2}{\sigma_n^2} = 15.810\,\mathrm{dB}$ and quasi-Gray $\mathbf{x}_{8,\mathrm{tri}}$. The labels l apply for both mappings.

mappings. If the mapping should achieve a fraction $R_c = 0.5$ of the maximum capacity, then 8-PSK Gray mapping $\mathbf{x}_G$ is close to optimum, as can be seen in the left zoomed section. The optimized mapping offers just a 0.05 dB gain. Note however, that we did not search for mappings that achieve $C_B = 1.5$ with *minimum* SNR, which would yield a larger gain, but is difficult to optimize. Rather we determined the SNR, which is necessary for the standard mapping $\mathbf{x}_G$ to achieve this capacity, and then optimized mappings at just this SNR with respect to maximum BICM capacity. The optimized mapping, that is designed for a rate $R_c = 0.75$ code, offers a 0.35 dB gain compared to the $\mathbf{x}_G$ mapping. As discussed above, even a 7-QAM yields a larger BICM capacity than $\mathbf{x}_G$, but approaches its smaller limit of 2.75 for increasing SNR. For a rather large target SNR value of $\frac{\sigma_s^2}{\sigma_n^2} = 15.810\,\mathrm{dB}$, the optimized mapping performs slightly better than the quasi-Gray mapping for the lattice of equilateral triangles $\mathbf{x}_{8,\mathrm{tri}}$.

We conclude by comparing the 8-QAM candidates with respect to the simulated BER after optimum ML demapping, cf. Figure 3.21. For small SNR, the baseline Gray mapping performs almost identical to the optimized mapping, but is suboptimum from about $\frac{\sigma_s^2}{\sigma_n^2} = 6\,\mathrm{dB}$ on. As already mentioned, the 7-QAM exhibits an error-floor $P_b = 1/24$, even for the noise-free case, because the receiver interprets this mapping as an 8-ary QAM with two overlapping symbols. Note that even though at $\frac{\sigma_s^2}{\sigma_n^2} = 7.216\,\mathrm{dB}$, the BICM capacity of the 7-QAM exceeds that of $\mathbf{x}_G$, its resulting BER is larger than that of the baseline mapping. This contradicts the prediction of (3.29), which holds under the Gaussian approximation. However, the overall distribution of extrinsic L-values for the 7-QAM violates this assumption to an even greater extend than other mappings do. Obviously, the first bit of symbols x_2 and x_6 of this

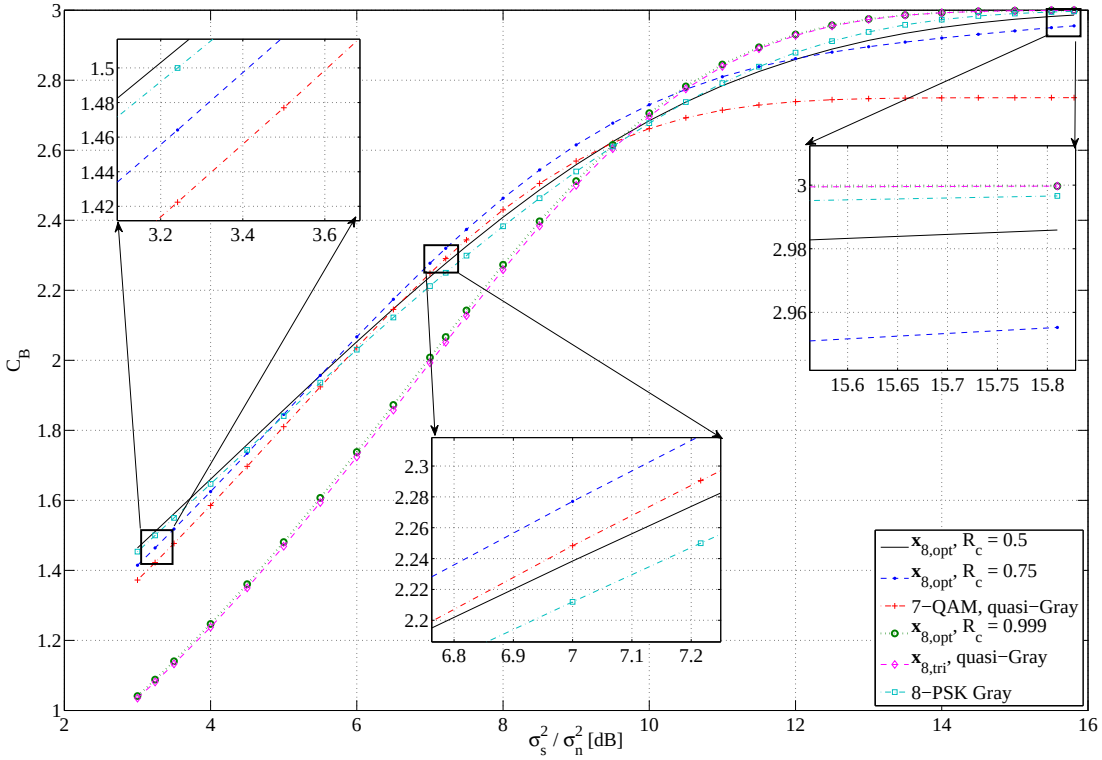


Figure 3.20: BICM capacity for 8-QAM candidates

degraded 8-QAM, which cannot be detected even for infinite SNR, yields a delta function in the distribution at the origin. One might regard the delta function as the limit of a normal distribution, whose variance approaches zero, but this Gaussian distribution is just one part of the overall L-value distribution. Thus, the pdf of extrinsic L-values does not possess Gaussian-like shape. From approximately 10 dB on, both $\mathbf{x}_{8,\text{opt}}$ for the very high code rate and the quasi-Gray lattice of equilateral triangles $\mathbf{x}_{8,\text{tri}}$ outperform all other candidates. For small BER (below 10^{-3}), the optimized mapping offers a gain of 1.3 dB compared to the 8-PSK Gray mapping.

3.3.4 16-QAM for Maximum BICM Capacity

The baseline mapping $\mathbf{x}_{\text{G}}$ for 16-QAM is depicted in Figure 2.7(b). Its symbol constellation is given by the vertices of an orthogonal lattice, and its bit labeling is Gray. For a BICM capacity of $C_{\text{B}} = R_{\text{c}} \cdot M$, the required SNR is $\frac{\sigma_{\text{s}}^2}{\sigma_{\text{n}}^2} = 5.280\,\text{dB}$, when $R_{\text{c}} = 0.5$. The optimized mapping for this target SNR can be seen in Figure 3.22(a) and is defined as

$$\mathbf{x}_{16,\text{opt}} = (r_i \cdot \mathbf{x}, r_o \cdot \mathbf{x}), \tag{3.35}$$

where $\mathbf{x}$ is the 8-PSK Gray mapping vector from (3.2). The inner radius is $r_i = 0.501$ and the outer radius is $r_o = \sqrt{2 - r_i^2} \approx 1.323$. While the reference mapping $\mathbf{x}_{\text{G}}$ offers $C_{\text{B}} = 2$, the optimized mapping yields $C_{\text{B}} = 2.0274$. Interestingly, the same constellation of concentric

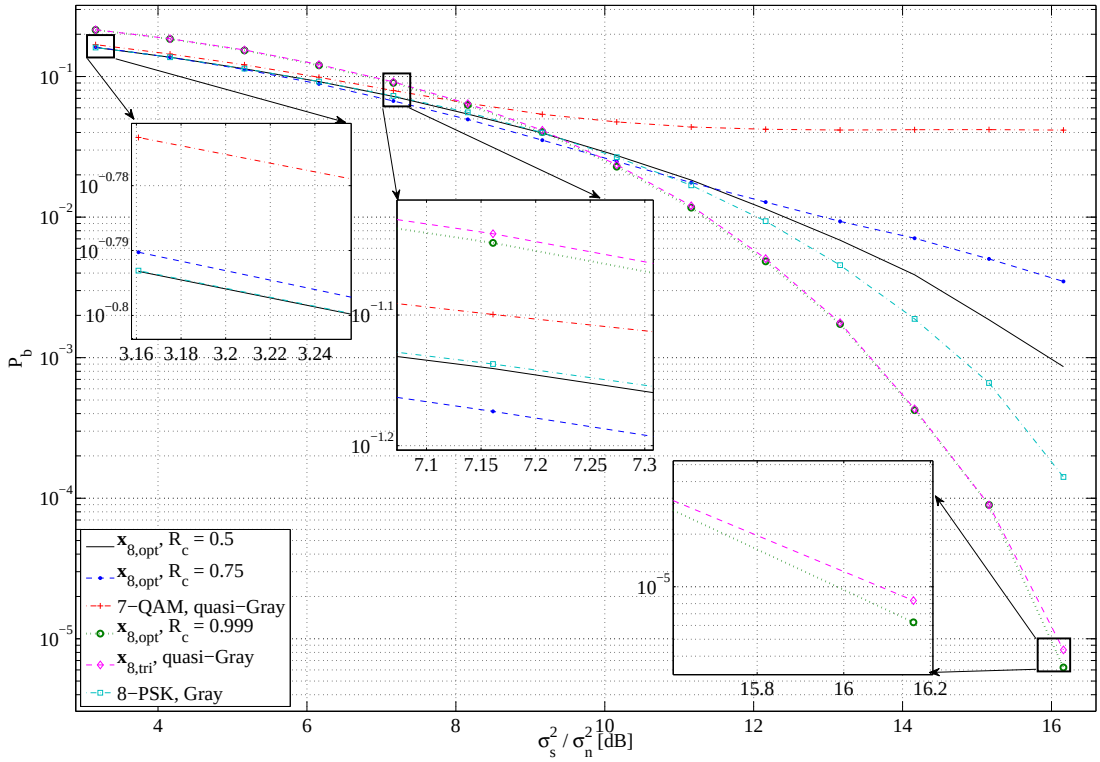


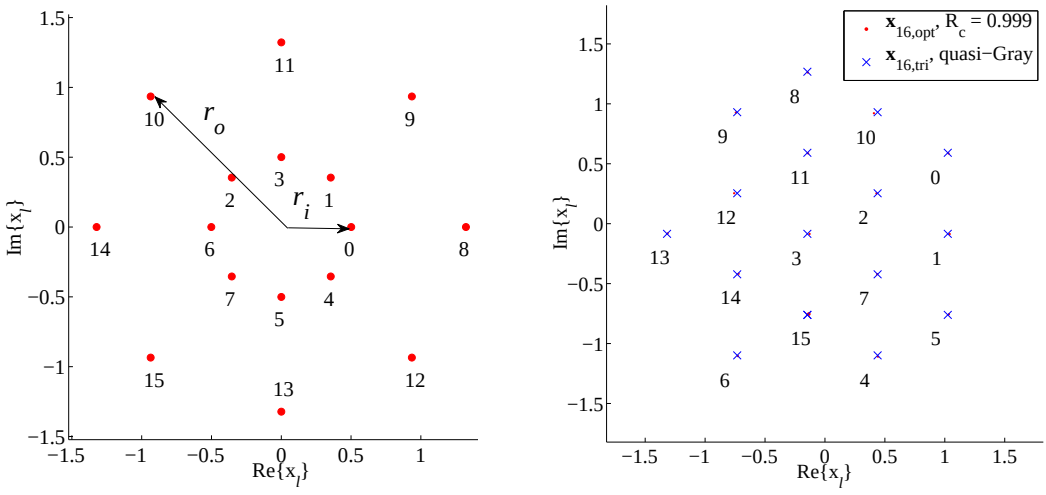
Figure 3.21: Simulated BER for 8-QAM candidates

circles maintains for the optimum mapping, if the target SNR is $\frac{\sigma_s^2}{\sigma_n^2} = 9.309\,\text{dB}$, for which $\mathbf{x}_\text{G}$ achieves 0.75 of its maximum BICM capacity. However, the inner radius expands to $r_i = 0.577$, while the outer radius shrinks in return to $r_o \approx 1.291$.[2] This mapping permits a capacity of $C_\text{B} = 3.026$. For the previous SNR of $\frac{\sigma_s^2}{\sigma_n^2} = 5.280\,\text{dB}$, it results in a BICM capacity, which is just 0.044% smaller than the global optimum. Both mappings in Figure 3.22(a) possess a minimum Euclidean distance, which is determined by adjacent symbols from the inner circle, i.e., $d_\text{min} = 2\sin\left(\pi/8\right) \cdot r_i$. For $r_i = 0.501$ and $r_i = 0.577$, we have $d_\text{min} = 0.383$ and 0.441, respectively. Both values are smaller than $d_\text{min} = \frac{2}{\sqrt{10}}$ from $\mathbf{x}_\text{G}$, however, the multiplicity is only eight, compared to 24 for $\mathbf{x}_\text{G}$.

At this point, it is worth mentioning that concentric ring constellations have recently been applied to investigate the channel capacity limits of fiber-optic transmission systems. There, the constellation was deduced from the fact that a concentric ring constellation with equal number of symbols per ring can approximate the complex Gaussian distribution, which is known to maximize channel capacity over the linear AWGN channel [115, 116].

The optimized mapping for the very high code rate $R_\text{c} = 0.999$ is very close to the ideal lattice of equilateral triangles with quasi-Gray labeling, as can be seen in in Figure 3.22(b). The explicit mapping vector $\mathbf{x}_{16,\text{opt}}$ is given in the Appendix B.2. Compared to the reference

[2] We note that this tendency does not allow for a rule of thumb, such as, e.g., a larger target SNR yields (3.35) with increasing r_i. Further optimizations at nearby SNR values resulted in different constellations.



(a) $\mathbf{x}_{16,\text{opt}}$ mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 5.28$ and 9.309 dB

(b) $\mathbf{x}_{16,\text{opt}}$ mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 17.66\,\text{dB}$ and quasi-Gray $\mathbf{x}_{16,\text{tri}}$

Figure 3.22: 16-QAM mappings for maximum BICM capacity. The labels l in each subfigure apply for both depicted mappings.

Gray mapping, which requires $\frac{\sigma_s^2}{\sigma_n^2} = 17.660\,\text{dB}$ for $C_\text{B} = 3.996$, $\mathbf{x}_{16,\text{opt}}$ yields a very small increase in BICM capacity (0.015%).

Figure 3.23 shows the BICM capacities of various 16-QAM mappings as a function of SNR. It can be observed that each mapping performs best at its target SNR. The baseline 16-QAM Gray mapping $\mathbf{x}_\text{G}$ is close to optimum for the whole range of depicted SNRs. From about 10.5 until 15 dB, it seems as if $\mathbf{x}_\text{G}$ is the optimum mapping. However, we did not give optimized mappings for SNRs in this interval. For high SNR, the optimized mapping and the quasi-Gray $\mathbf{x}_{16,\text{tri}}$ mapping achieve the highest C_B, followed by the baseline mapping.

The simulated BER for all 16-QAM candidates are plotted in Figure 3.24. At $\frac{\sigma_s^2}{\sigma_n^2} = 5.280\,\text{dB}$, the respective optimized mapping has higher BER than both the baseline mapping and the optimized mapping for $\frac{\sigma_s^2}{\sigma_n^2} = 9.309\,\text{dB}$, although it allows for higher C_B. This shows again that the BER prediction (3.29) through C_B is an approximation, which assumes Gaussian L-value distributions. For all other target SNRs, maximum C_B reflects minimum BER. In the enlarged section for large SNR values, it can be seen that the optimum mapping for the very high code rate offers a 0.1 dB gain compared to the reference 16-QAM.

3.3.5 32-QAM for Maximum BICM Capacity

Let $\mathbf{x}_\text{G}$ be the 32-QAM rectangular Gray mapping from Figure 3.12(b). Setting $R_\text{c} = 0.5$, a BICM capacity, which should equal the spectral efficiency $R_\text{c} \cdot M = 2.5$, requires an SNR of $\frac{\sigma_s^2}{\sigma_n^2} = 8.276\,\text{dB}$. At this target SNR, the optimum 32-QAM is shown in Figure 3.25(a).

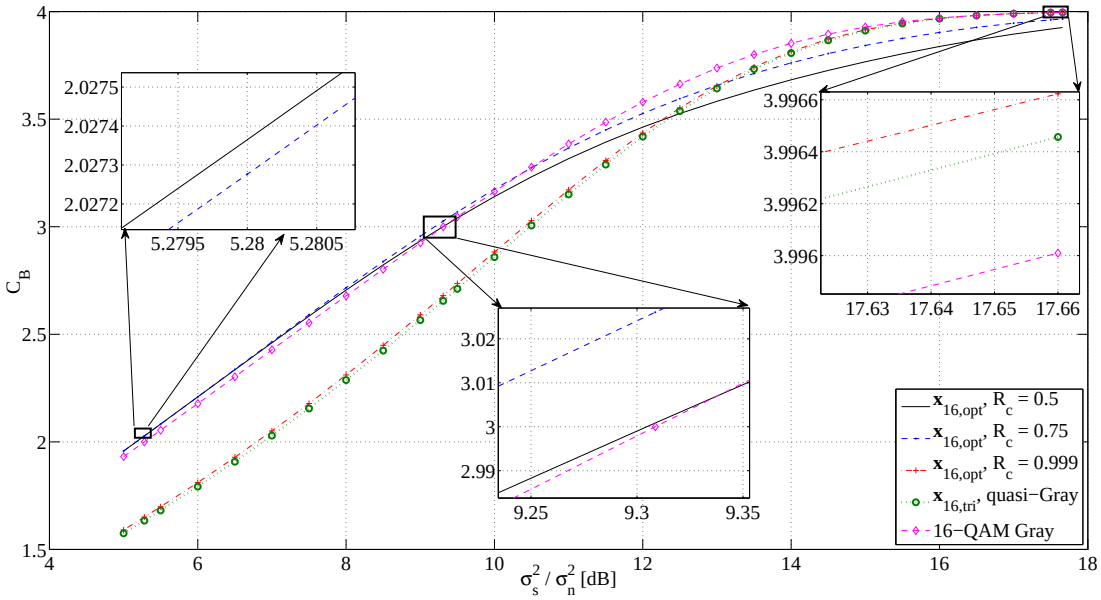


Figure 3.23: BICM capacity for 16-QAM candidates

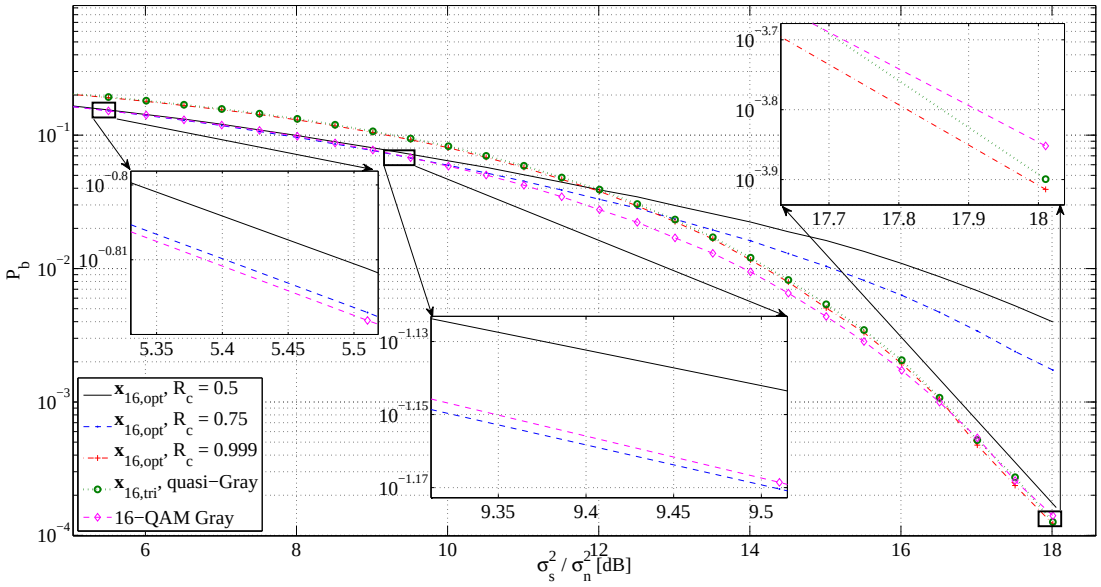


Figure 3.24: Simulated BER for 16-QAM candidates

Similar to the optimized 16-QAM for this code rate, the mapping is composed of 8-PSK Gray mappings $\mathbf{x}$ with different amplitude levels,

$$\mathbf{x}_{32,\mathrm{opt}} = \left(r_1 \cdot \mathbf{x}, r_2 \cdot \mathbf{x}, r_3 \cdot \mathbf{x}, r_4 \cdot \mathbf{x}\right), \tag{3.36}$$

in which $r_1 = 0.4791, r_2 = 0.6159, r_3 = 1.2319$ and $r_4 = 1.3688$. The difference between the two innermost and two outermost radii is almost identical[3]. This mapping yields $C_{\mathrm{B}} = 2.77$, which is about 11% more than the capacity of $\mathbf{x}_{\mathrm{G}}$. Ignoring the small difference between inner and outer radii, we find the minimum distance $d_{\min} = r_2 - r_1 \approx r_4 - r_3 \approx 0.137$, which is much smaller than $\frac{4}{\sqrt{104}} \approx 0.392$, which occurs as $d_{\min}$ in $\mathbf{x}_{\mathrm{G}}$ with a multiplicity of 52, whereas the multiplicity of $d_{\min}$ for the optimized mapping is only 16.

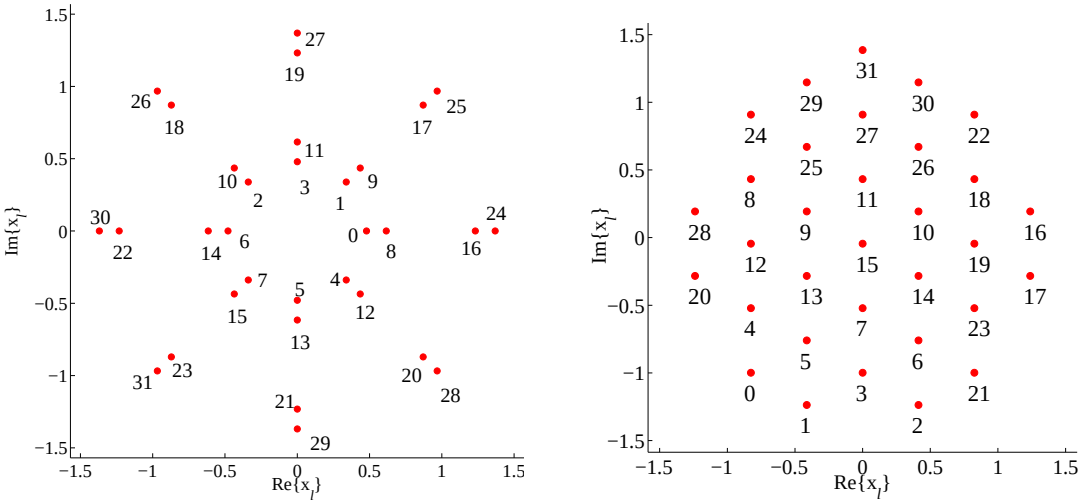
Remarks on previous mapping vector: It should be noted that this highly structured 32-QAM was found in the following way. First, the outputs of the optimization algorithm were observed. The best mappings after several 10,000 trails were examined and similarities were extracted. However, most of these mappings possessed no perfect symmetries, i.e., the mapping was already close to (3.36), but some symbols varied slightly. It turned out that the artificially "symmetrized" mapping (3.36) achieved a slightly higher cost function than the outputs of the algorithm, which were found until then. Thus, we conclude that the optimum mapping would have been found by the algorithm, if even more trails had been passed.

A target SNR of $\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2} = 12.831\,\mathrm{dB}$ yields 0.75 of the maximum BICM capacity for $\mathbf{x}_{\mathrm{G}}$. Here, the optimum mapping turns out to be the 32-QAM cross mapping with quasi-Gray labeling from Figure 3.12(a), allowing for $C_{\mathrm{B}} = 3.929$. This is about 5% more than the respective capacity of the baseline Gray mapping.

Finally, we consider the very high code rate $R_{\mathrm{c}} = 0.999$. The optimum mapping for the corresponding target SNR of $\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2} = 21.686\,\mathrm{dB}$ is exactly the ideal lattice of equilateral triangles with quasi-Gray labeling $\mathbf{x}_{32,\mathrm{tri}}$ and is depicted in Figure 3.25(b). The gain in C_{B}, compared to the reference mapping is only about 0.09%. Note that both optimized mappings for $R_{\mathrm{c}} = 0.75$ and $R_{\mathrm{c}} = 0.999$ do not fulfill Definition 2.1. Hence, at larger SNR, it is more important, that symbol errors are avoided, because $\mathbf{x}_{32,\mathrm{tri}}$ was also optimum at high SNR with respect to SER, than the property that neigboring symbols must not differ in more than one bit.

We summarize all presented mappings in Figure 3.26, where the BICM capacities over SNR are presented. Note that $\mathbf{x}_{32,\mathrm{opt}}$ for $R_{\mathrm{c}} = 0.75$ equals the 32-QAM cross mapping $\mathbf{x}_{32,\mathrm{cro}}$, and $\mathbf{x}_{32,\mathrm{opt}}$ for $R_{\mathrm{c}} = 0.999$ is the hexagonal lattice $\mathbf{x}_{32,\mathrm{tri}}$. It can be seen that $\mathbf{x}_{32,\mathrm{cro}}$ achieves higher BICM capacities than the rectangular Gray mapping $\mathbf{x}_{\mathrm{G}} = \mathbf{x}_{32,\mathrm{rec}}$ for all SNR values. Even though its labeling is not exactly Gray, the mapping offers a larger minimum distance due to a more compact symbol constellation. Notice that the relative large gains of the optimized mappings in C_{B} compared with $\mathbf{x}_{\mathrm{G}}$ can be explained by the fact that this Gray mapping does not provide a good baseline capacity. Rather, the cross constellation with quasi-Gray labeling

[3] If we force this difference to be exactly the same, a slightly smaller C_{B} results.



(a) $\mathbf{x}_{32,\text{opt}}$ mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 8.276\,\text{dB}$

(b) $\mathbf{x}_{32,\text{opt}}$ mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 21.686\,\text{dB}$

Figure 3.25: 32-QAM mappings for maximum BICM capacity

serves as a better reference mapping. For small SNR, the optimized mapping with symbols on concentric circles, cf. (3.36), gives the highest C_B, but from about 12 dB on, it suffers from its small minimum distance and approaches the maximum capacity very slowly with increasing SNR. If the SNR is larger than 18.42 dB, the hexagonal lattice with quasi-Gray labeling yields the largest BICM capacities, followed by the 32-QAM cross mapping.

Figure 3.27 shows the simulated BER, which results after ML demapping with hard decision. Again, the highest C_B does not necessarily result in minimum BER for low SNR. Here, the 32-QAM cross mapping is optimum, and also the baseline mapping has smaller BER than $\mathbf{x}_{32,\text{opt}}$ for $\frac{\sigma_s^2}{\sigma_n^2} = 8.276\,\text{dB}$. From about 18 dB on, $\mathbf{x}_{32,\text{tri}}$ results in the smallest BER. At $P_\text{b} = 10^{-2}$, the 32-QAM cross mapping offers a 1 dB gain compared to the rectangular Gray mapping $\mathbf{x}_\text{G}$. The same gap was found for suboptimum ML demapping with the max-log approximation in [113].

3.3.6 Summary for Maximum BICM Capacity Mappings

In this section, we have proposed mappings that maximize the BICM capacity for different target SNRs. These SNR values were related to the BICM capacity limits of well-known Gray mappings, such as 8-PSK. As we have pointed out, a strong outer encoder must ensure reliable transmission down to this SNR limit. Examples of these channel codes are codes on graphs, such as turbo and LDPC codes. In [117, 118], LDPC and the related class of extended irregular repeat-accumulate codes were presented for spectral-efficient modulation, which approach the BICM capacity by about 0.6 dB. Our optimized mappings, however, achieve higher BICM capacities at these SNR limits, or in other words, allow the same spec-

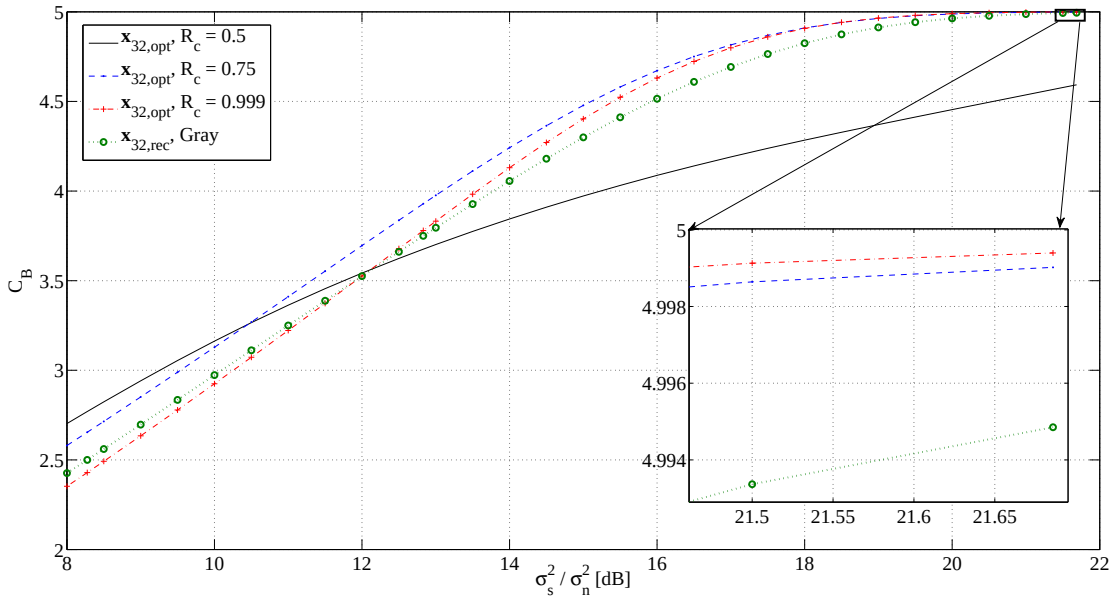


Figure 3.26: BICM capacity for 32-QAM candidates

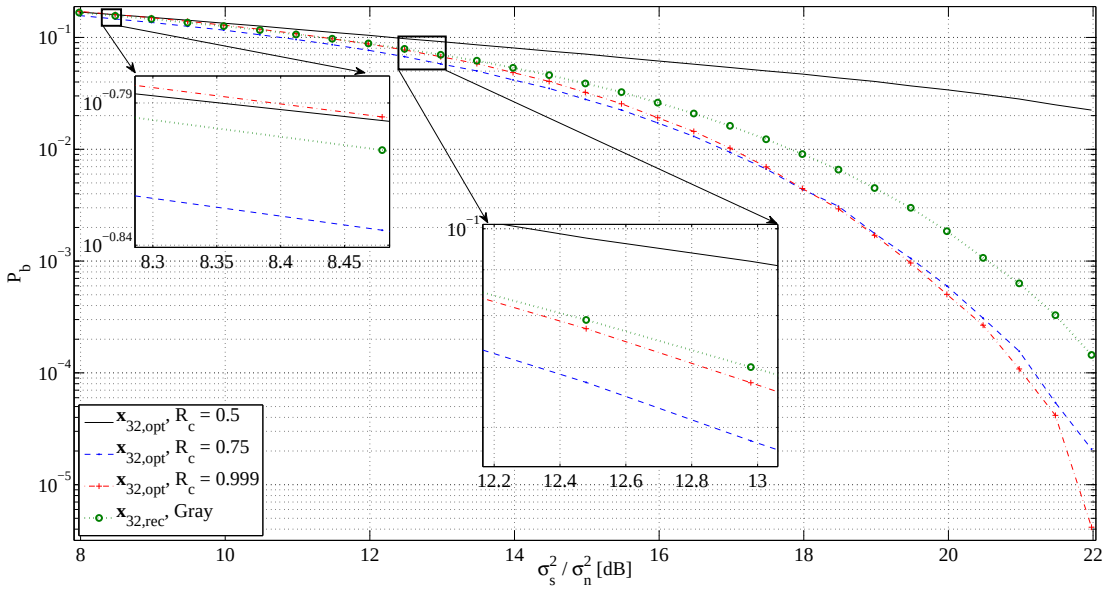


Figure 3.27: Simulated BER for 32-QAM candidates

tral efficiency at a lower SNR. We have shown that gains in the order of several tenths of a dB are possible, if both the bit labeling *and* the symbol constellation are optimized. These gains come with no additional complexity, because a capacity approaching receiver has to perform optimum ML demapping anyway. Only if suboptimum max-log ML demapping is considered, which allows for detection by decision areas $\mathscr{A}_{l_1}$, the presented constellations possess more complicated decision thresholds. A rule of thumb for the optimized constellations with respect to C_{B} is similar to the one which applies for minimum SER: For large SNR, the dominant contribution comes from closest neighbors, such that the minimum Euclidean distance has to be maximized. This is fulfilled by the hexagonal lattice mappings. In this case, it is not even required that the resulting mapping is exactly Gray labeled. For medium-to-low SNR, however, relatively small minimum distances occur in the optimized mappings, but also with a small multiplicity. In addition to the BICM capacity C_{B}, we have examined the resulting BER after ML demapping, which we have related to C_{B}. Here, the gains in SNR are even larger. For example, the optimized 8-QAM for a target SNR of $\frac{\sigma_s^2}{\sigma_n^2} = 15.81\,\mathrm{dB}$ allows $P_{\mathrm{b}} = 10^{-3}$ about 1.3 dB earlier than the typically employed 8-PSK Gray mapping.

Finally, we note that in [119], several well-known 8- and 16-QAM mappings are compared with respect to their BICM capacities for different SNR values. No optimization of the mapping vectors was performed, such that the author concludes, that, e.g., the 16-QAM Gray mapping, which we have considered as the baseline mapping, achieves the maximum capacity for all SNRs.

3.4 Mappings for Maximum Signal Set Capacity

While the BICM capacity C_{B} relates to the starting point of the demapper's transfer characteristic T_1, the signal set capacity C_{S} is proportional to the area under T_1. The capacity C_{S} assumes that the optimum receiver performs *joint* demapping and decoding. Close to optimum receivers operate iteratively, as in the BICM-ID scheme. We have seen in Section 2.6 that capacity approaching BICM-ID requires a good match between T_1 and the transfer characteristic of the outer decoder T_2. The two main strategies for this goal are either to apply a Gray mapping, for which T_1 is almost flat together with a powerful outer decoder, for which T_2 is almost a step-function, as discussed in Section 3.3, or to apply an anti-Gray mapping that exploits a priori information more effectively and an outer decoder, whose transfer characteristic matches T_1. For this case, T_1 has a steep gradient, and thus, the channel encoder can be a simple convolutional code with small constraint length, because its T_2 curve possesses the corresponding form [96]. In both cases, only the gradient of T_1 can be changed by different bit labelings, not the area under this curve. In order to lift this curve to higher MIs for the same SNR, it is necessary to change the symbol constellation, such that C_{S} increases. In this section, we will optimize mappings that achieve maximum signal set capacity C_{S} for a given target SNR. As C_{S} depends only on the symbol constellation $\mathbb{X}$, the optimized mappings are described by the elements of the mapping vector $\mathbf{x}$, but the order is irrelevant. Nevertheless, the order, i.e., the bit labeling μ, matters for the shape of T_1, but will not be discussed here.

With the notation from (3.30), the signal set capacity becomes

$$C_{\mathrm{S}} = \frac{1}{L} \int_{r \in \mathbb{C}} \sum_{l_1=0}^{L-1} p\left(r \left| x_{l_1}\right.\right) \log_2 \frac{p\left(r \left| x_{l_1}\right.\right)}{p\left(r\right)} dr. \tag{3.37}$$

Note that $C_{\mathrm{S}} \geq C_{\mathrm{B}}$, because the receiver is not constrained to perform APP demapping only once as in case of a BICM receiver. Only for BPSK and QPSK Gray mapping, the transfer characteristic T_1 is exactly flat. As we have discussed in Subsection 2.1.3, the latter mapping corresponds to two independent BPSK mappings — one on each orthogonal QAM phase. Thus, fed back a priori knowledge can not be exploited in an APP demapper, resulting in a constant T_1 curve. All other mappings, Gray labeled or otherwise, introduce dependencies among the M bits, which are mapped onto the two-dimensional constellation diagram. From this point of view, it is also clear that no more than two bits per symbol can be mapped independently onto two-dimensional symbols.

The partial derivative of (3.37) is computed in the Appendix A.3.1 as

$$\frac{\partial C_{\mathrm{S}}\left(\mathbf{x}\right)}{\partial x_l} = \frac{2}{L} \int_{r \in \mathbb{C}} p\left(r \left| x_l\right.\right) \cdot \frac{r - x_l}{\sigma_{\mathrm{n}}^2} \log_2 \frac{p\left(r \left| x_l\right.\right)}{p\left(r\right)} dr. \tag{3.38}$$

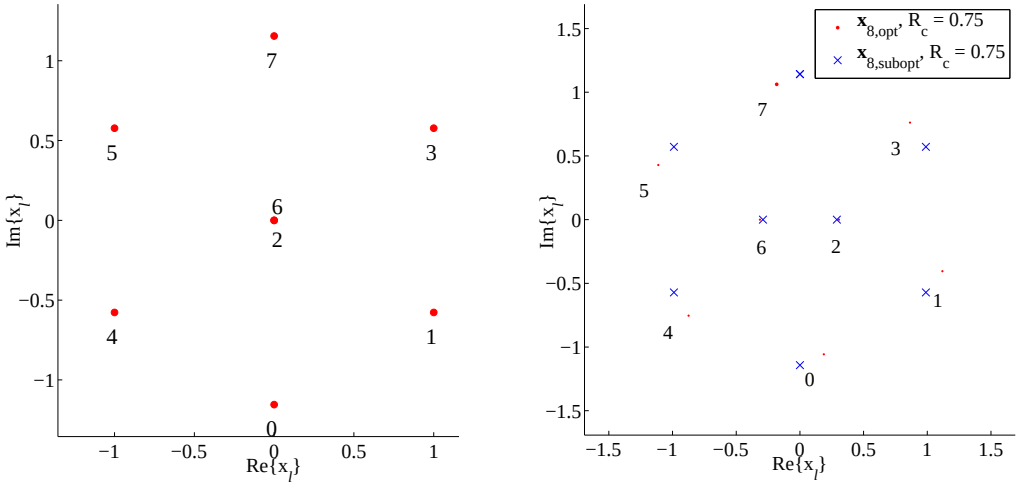
In the following subsections, we present L-ary QAM mappings that maximize C_{S} for $L \in \{8, 16, 32\}$. For each L, we examine two different target SNR values: one, for which the baseline constellation $\mathbf{x}_{\mathrm{G}}$ achieves a fraction of $R_{\mathrm{c}} = 0.5$ of its maximum signal set capacity and the other one for $R_{\mathrm{c}} = 0.75$. The BICM capacity approaches the signal set capacity closely for large SNRs, as can be seen in Figure 2.17. Hence, we do not optimize the mappings for concatenations with very high code rates, such as the rate $R_{\mathrm{c}} = 0.999$ from the previous subsection. We can rather state that the (close to) hexagonal lattice constellations for the high SNR domain are optimum also with respect to C_{S}. The baseline mappings are as previously defined the 8-PSK and the 16-QAM from Figure 2.7, but for 32-QAM, we will focus on the cross constellation from Figure 3.12(a). As we will see, the target SNRs for the same code rate are always smaller than the respective values from the previous subsection, where we equated $\eta = C_{\mathrm{B}}$, rather than $\eta = C_{\mathrm{S}}$. Stated differently, the same SNR yields $C_{\mathrm{S}} \geq C_{\mathrm{B}}$.

3.4.1 8-QAM for Maximum Signal Set Capacity

The code rate of $R_{\mathrm{c}} = 0.5$ yields a spectral efficiency of $\eta = 1.5$ bits per channel usage. Equating the signal set capacity C_{S} with η yields a target SNR of $\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2} = 3.028\,\mathrm{dB}$ for the baseline 8-PSK mapping $\mathbf{x}_{\mathrm{G}}$. The optimum 8-QAM mapping achieves $C_{\mathrm{S}} = 1.555$ for this SNR. Its constellation equals that of the 7-QAM from (3.34), when the inner radius is zero,

$$\mathbf{x}_{8,\mathrm{opt}} = \frac{2}{\sqrt{3}} \cdot \left(-\mathrm{j}, \mathrm{e}^{-\mathrm{j}\frac{\pi}{6}}, 0, \mathrm{e}^{\mathrm{j}\frac{\pi}{6}}, -\mathrm{e}^{\mathrm{j}\frac{\pi}{6}}, -\mathrm{e}^{-\mathrm{j}\frac{\pi}{6}}, 0, \mathrm{j}\right), \tag{3.39}$$

and is depicted in Figure 3.28(a). To be precise, this corresponds to an 8-QAM, in which two symbols are identical, $x_2 = x_6 = 0$. If we would discard one of these symbols and compute C_S for this true 7-QAM, in which each symbols occurs with probability $\frac{1}{7}$, the result is only $C_S = 1.536$, but still larger than the capacity of $\mathbf{x}_G$. Thus, the minimum distance of $\mathbf{x}_{8,opt}$ is zero with multiplicity one. The second smallest distance is $2\sin\left(\pi/6\right) = 1 < \frac{2}{\sqrt{3}} \approx 1.155$, which occurs between adjacent symbols from the outer circle. Hence, the multiplicity is six. This distance is much larger than d_{min} of $\mathbf{x}_G$, which is $2\sin\left(\pi/8\right) \approx 0.765$. We can interpret this optimum as follows: By sacrificing the symbol x_2, which can not be discriminated anymore from symbol x_6, a larger spacing of all other symbols results, which outweighs the loss of one symbol. It must also be considered that the signal set capacity is not coupled with a certain bit labeling, contrary to the BICM capacity. Therefore, two identical symbols do not necessarily yield a higher symbol error probability. It rather means that one symbol ($x_2 = 0$) occurs twice as often as the other six symbols, located on the outer circle. A scheme, which approaches the signal set capacity, could encode the information bits directly onto the seven symbols, where the symbol in the origin appears twice as likely. On the other hand, a true 7-QAM with equiprobable symbols would have one symbol in the origin and six on a circle with radius $r_o = \sqrt{\frac{7}{6}} \approx 1.08$, which is smaller than the radius of the degraded 8-QAM with $x_2 = x_6 = 0$.



(a) $\mathbf{x}_{8,opt}$ mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 3.028\,\text{dB}$

(b) optimum and suboptimum 8-QAM mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 7.18\,\text{dB}$

Figure 3.28: 8-QAM mappings for maximum signal set capacity

In Subsection 3.3.1, we have established a relation between C_B and the BER after ML demapping. One might conjecture that the same is true about C_S and the SER after ML demapping, which is also independent of a particular bit labeling. We disprove a possible connection between these two cost functions with the following counterexample: The optimized mapping with respect to the SER at $\frac{\sigma_s^2}{\sigma_n^2} = 3.028\,\text{dB}$ is $\mathbf{x}_{8,1\_7}$ from (3.20), i.e., the same vector, which

was already optimum at $\frac{\sigma_s^2}{\sigma_n^2} = 10\,\text{dB}$. However, this mapping achieves only $C_S = 1.535$ and is therefore not optimum with respect to the signal set capacity.

Next, we examine the reference 8-PSK $\mathbf{x}_G$ for a fraction of $R_c = 0.75$ of its maximum signal set capacity. The mapping $\mathbf{x}_G$ achieves $C_S = 2.25$ at $\frac{\sigma_s^2}{\sigma_n^2} = 7.180\,\text{dB}$. Optimization at this SNR yields the mapping $\mathbf{x}_{8,\text{opt}}$ that is shown in Figure 3.28(b). Its mapping vector is listed in the Appendix B.3. It resembles the mapping from (3.34), with inner radius $r_i \approx 0.3$, which we have also shown in this figure. We denote this as the suboptimum solution $\mathbf{x}_{8,\text{subopt}}$. Note however, that the six symbols of the optimum mapping with larger magnitude are also rotated compared to $\mathbf{x}_{8,\text{subopt}}$ (on average by about 10°). The advantage of $\mathbf{x}_{8,\text{subopt}}$ is its closed-form mapping vector. For this target SNR, $\mathbf{x}_{8,\text{opt}}$ achieves $C_S = 2.404$, and $\mathbf{x}_{8,\text{subopt}}$ gives $C_S = 2.401$.

The signal set capacities over a large scale of SNR values is depicted in Figure 3.29. In addition to the optimum and the baseline 8-PSK mappings, we give the results for the 8-QAM cross constellation $\mathbf{x}_{\text{cro}}$. From 0 until about 5 dB, the mapping that we optimized for $R_c = 0.5$ achieves the highest C_S. For $C_S = 1.5$, it requires 0.28 dB less than the 8-PSK mapping. This baseline mapping is also outperformed by $\mathbf{x}_{\text{cro}}$ for all depicted SNRs. For $C_S = 2.25$, i.e., at 0.75 of the maximum capacity, 8-PSK needs about 0.57 dB more SNR than $\mathbf{x}_{\text{cro}}$ and even 0.8 dB more than the optimized mapping for this code rate. At high SNR, the lattice of equilateral triangles $\mathbf{x}_{8,\text{tri}}$ gives the largest C_S, closely followed by $\mathbf{x}_{\text{cro}}$. As we have discussed in Subsection 3.3.3, the mapping from (3.39), which is an 8-QAM with two identical symbols, offers at most a capacity of 2.75. Thus, from about 10 dB on, it performs worst among all presented mappings.

3.4.2 16-QAM for Maximum Signal Set Capacity

The regular 16-QAM mapping achieves $C_S = 2 = 0.5 \cdot M$ at $\frac{\sigma_s^2}{\sigma_n^2} = 5.118\,\text{dB}$. The optimum mapping for this SNR is identical with the optimized mapping with respect to C_B for $R_c = 0.5$. Thus, equation (3.35) describes $\mathbf{x}_{16,\text{opt}}$, with $r_i = 0.501$ and $r_o = \sqrt{2 - r_i^2} \approx 1.323$. The respective signal set capacity is $C_S = 2.057$.

If we consider an SNR of $\frac{\sigma_s^2}{\sigma_n^2} = 9.304\,\text{dB}$, the baseline mapping reaches 0.75 of its maximum capacity. Figure 3.30 shows the optimized 16-QAM for this target SNR, which achieves $C_S = 3.075$. Its mapping vector is described by

$$\mathbf{x}_{16,\text{opt}} = (0, r_i \cdot \mathbf{x}_i, r_o \cdot \mathbf{x}_o)\,, \tag{3.40}$$

where the inner radius is $r_i = 0.575$ and the outer is $r_o = \sqrt{16 - 6r_i^2}/3 \approx 1.248$. One symbol is in the origin, six are located on $r_i \cdot \mathbf{x}_i$, where the l-th element of $\mathbf{x}_i$ is given by $\exp\left(\text{j}\frac{2\pi}{6}l\right)$, and the last nine symbols are located on the outer circle $r_o \cdot \mathbf{x}_o$. Here, the l-th element of $\mathbf{x}_o$ is $\exp\left(\text{j}\left(\frac{2\pi}{9}l + \varphi_0\right)\right)$, where the offset angle is $\varphi_0 = 0.286 \approx 0.091 \cdot \pi$. In addition to this

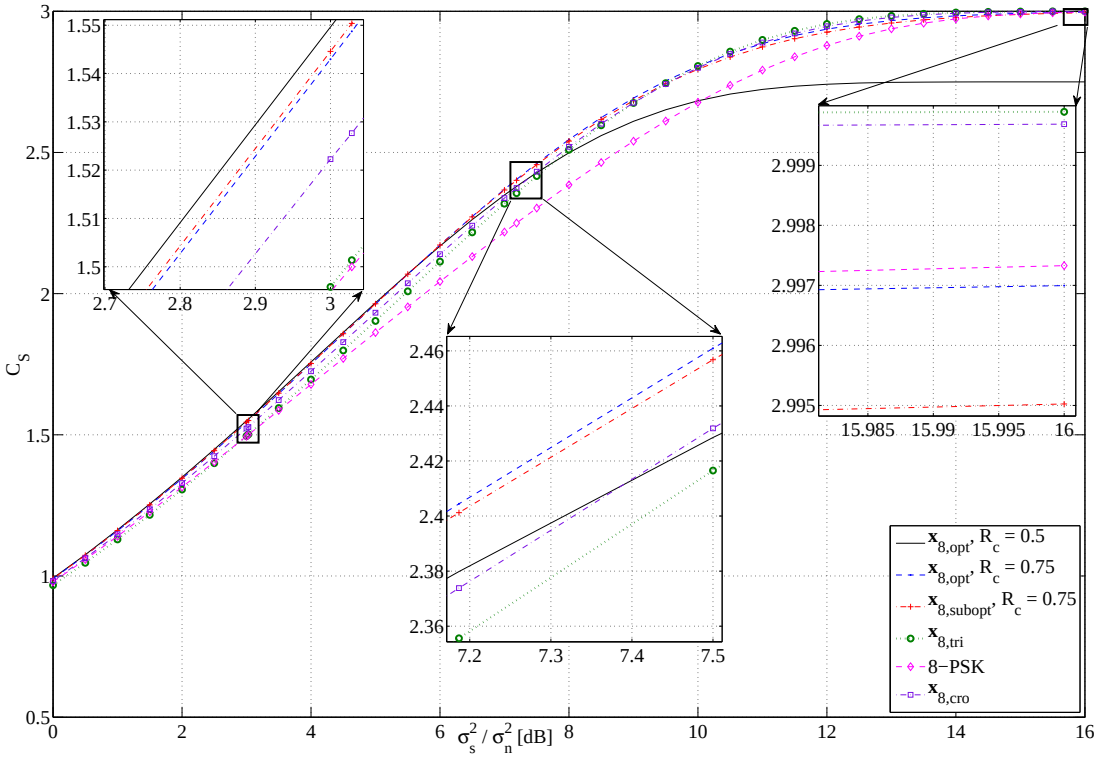


Figure 3.29: Signal set capacity for 8-QAM candidates

closed-form of $\mathbf{x}_{16,\text{opt}}$, its detailed components can be found in the Appendix B.3. It should be noted that the same remark as for the mapping in (3.36) holds for this case.

In Figure 3.31, we show the signal set capacities over SNR of all 16-QAM candidates. It can be seen that the regular 16-QAM mapping performs worst until 10 dB. For small-to-medium SNR (until about 7.3 dB), the optimized mapping for application with a rate $R_\text{c} = 0.5$ code yields the highest C_S, but the slope of its capacity curve flattens at larger SNRs. In the zoomed section in the middle, we can observe that the optimized mapping for $R_\text{c} = 0.75$ offers $C_\text{S} = 3$ about 0.3 dB earlier than the baseline mapping. Finally, the hexagonal lattice $\mathbf{x}_{16,\text{tri}}$ results in maximum C_S for large SNR values.

3.4.3 32-QAM for Maximum Signal Set Capacity

We use the 32-QAM cross constellation as the baseline mapping, $\mathbf{x}_\text{G} = \mathbf{x}_{32,\text{cro}}$. This mapping achieves half of its maximum capacity at $\frac{\sigma_\text{s}^2}{\sigma_\text{n}^2} = 7.053\,\text{dB}$. The optimized mapping, which yields $C_\text{S} = 2.574$ at this SNR, was found by "symmetrizing" the best outputs of the computer algorithm, such that the remark from (3.36) applies here, too. Its constellation is depicted in Figure 3.32(a) and can be described by

$$\mathbf{x}_{32,\text{opt}} = \left(0, r_1 \cdot \text{j}, -r_1 \cdot \text{j}, r_2 \cdot \mathbf{x}_2, r_3 \cdot \mathbf{x}_3, r_4 \cdot \mathbf{x}_4\right). \tag{3.41}$$

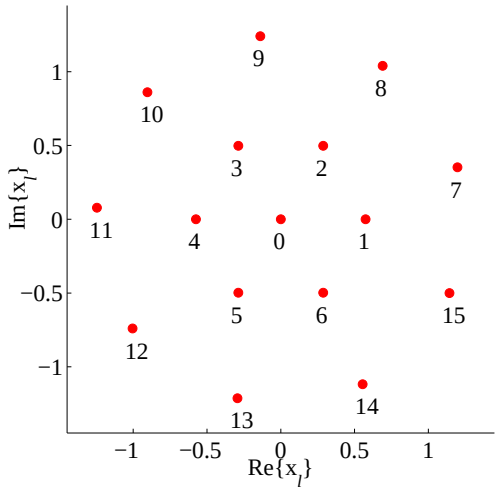


Figure 3.30: 16-QAM for maximum signal set capacity at $\frac{\sigma_s^2}{\sigma_n^2} = 9.304\,\mathrm{dB}$

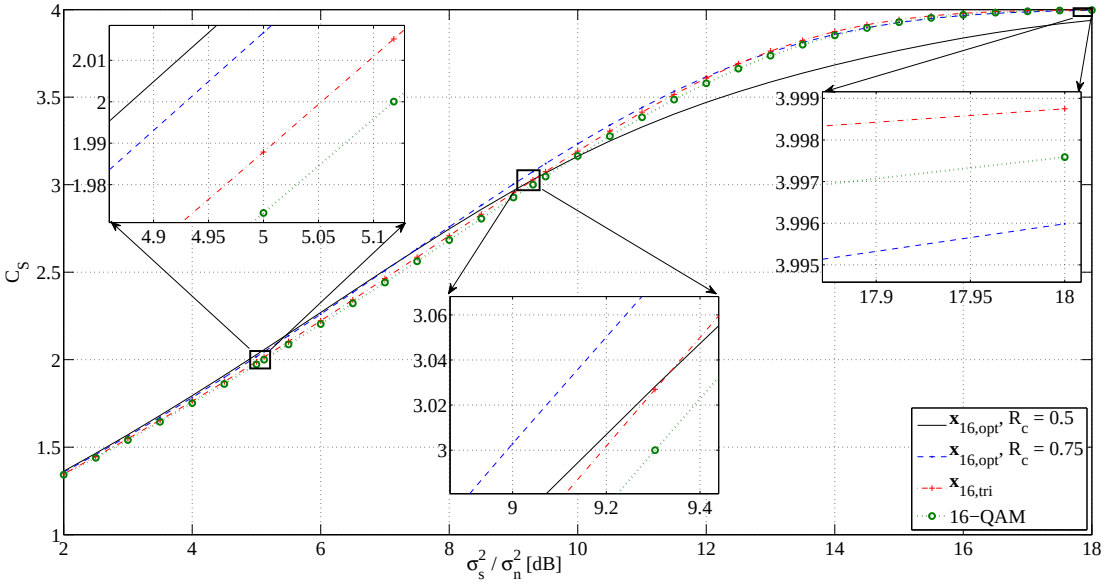
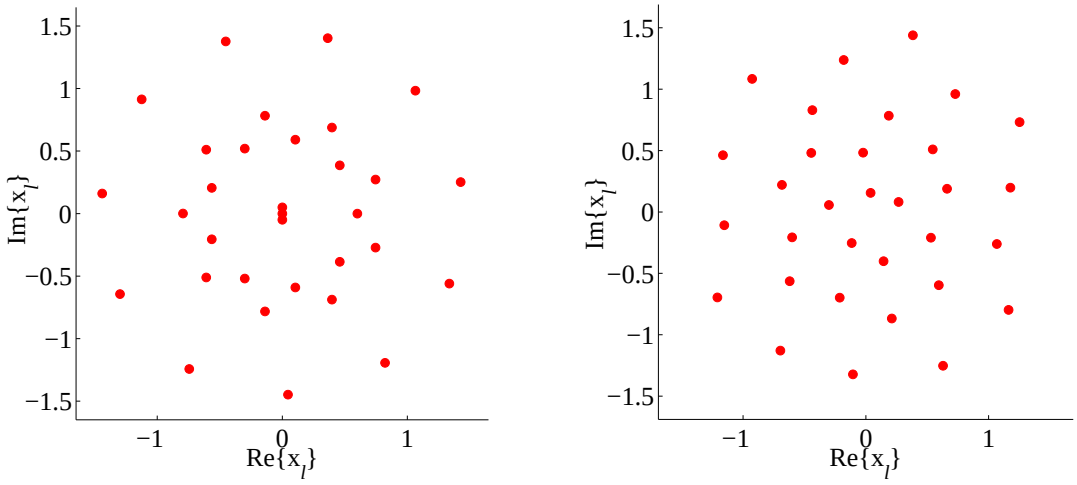


Figure 3.31: Signal set capacity for 16-QAM candidates

The radii are $r_1 = 0.05, r_2 = 0.6, r_3 = 0.794$, and $r_4 = \sqrt{\left(32 - 9\left(r_2^2 + r_3^2\right) - 2r_1^2\right)/11} \approx 1.449$. Each of the two vectors $\mathbf{x}_2$ and $\mathbf{x}_3$ comprises nine symbols, where the l-th element of $\mathbf{x}_2$ is $\exp\left(\mathrm{j}\frac{2\pi}{9}l\right)$ and for $\mathbf{x}_3$, an additional phase offset is included, $\exp\left(\mathrm{j}\left(\frac{2\pi}{9}l + \varphi_0\right)\right)$, with $\varphi_0 = \frac{\pi}{9}$. The remaining eleven symbols are located on the outer circle of $\mathbf{x}_4$, whose l-th entry is $\exp\left(\mathrm{j}\left(\frac{2\pi}{11}l + \varphi_1\right)\right)$, with $\varphi_1 = \frac{\pi}{18}$. The vector $\mathbf{x}_{32,\mathrm{opt}}$ is also specified in the Appendix B.3.



(a) $\mathbf{x}_{32,\mathrm{opt}}$ mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 7.053\,\mathrm{dB}$

(b) $\mathbf{x}_{32,\mathrm{opt}}$ mapping at $\frac{\sigma_s^2}{\sigma_n^2} = 11.740\,\mathrm{dB}$

Figure 3.32: 32-QAM mappings for maximum signal set capacity

At $\frac{\sigma_s^2}{\sigma_n^2} = 11.740\,\mathrm{dB}$, the reference mapping $\mathbf{x}_{32,\mathrm{cro}}$ has reached 75% of its maximum capacity, i.e., $C_S = 3.75$. The optimum vector for this target SNR is shown in Figure 3.32(b) and its coordinates are listed in the Appendix B.3. It allows for a higher signal set capacity of $C_S = 3.83$. Its mapping vector can not be related to a closed-form expression, even though its symbols possess some symmetric structures.

Figure 3.33 summarizes the signal set capacities of all 32-QAM candidates over a large range of SNR. As for the 8- and 16-QAM, each optimized mapping performs best at its target SNR. The optimized mapping for $R_c = 0.5$ achieves the best results for low SNR. At about 10 dB, it falls behind the optimized mapping for $R_c = 0.75$, and from 12 dB on, it performs worst among all presented mappings, mostly due to its very small minimum Euclidean distance of $d_{\min} = 0.05$. However, for $C_S = 2.5$, it offers 0.27 dB gain compared to the baseline mapping $\mathbf{x}_{32,\mathrm{cro}}$. It is worth noting that the rectangular constellation $\mathbf{x}_{32,\mathrm{rec}}$ needs even 0.925 dB more than $\mathbf{x}_{32,\mathrm{cro}}$ to achieve this capacity (not shown). For $C_S = 3.75$, the optimum mapping for $R_c = 0.75$ allows about 0.25 dB gain compared to $\mathbf{x}_{32,\mathrm{cro}}$. Finally, at high SNR, the hexagonal lattice mapping $\mathbf{x}_{32,\mathrm{tri}}$ yields maximum C_S, followed closely by $\mathbf{x}_{32,\mathrm{cro}}$.

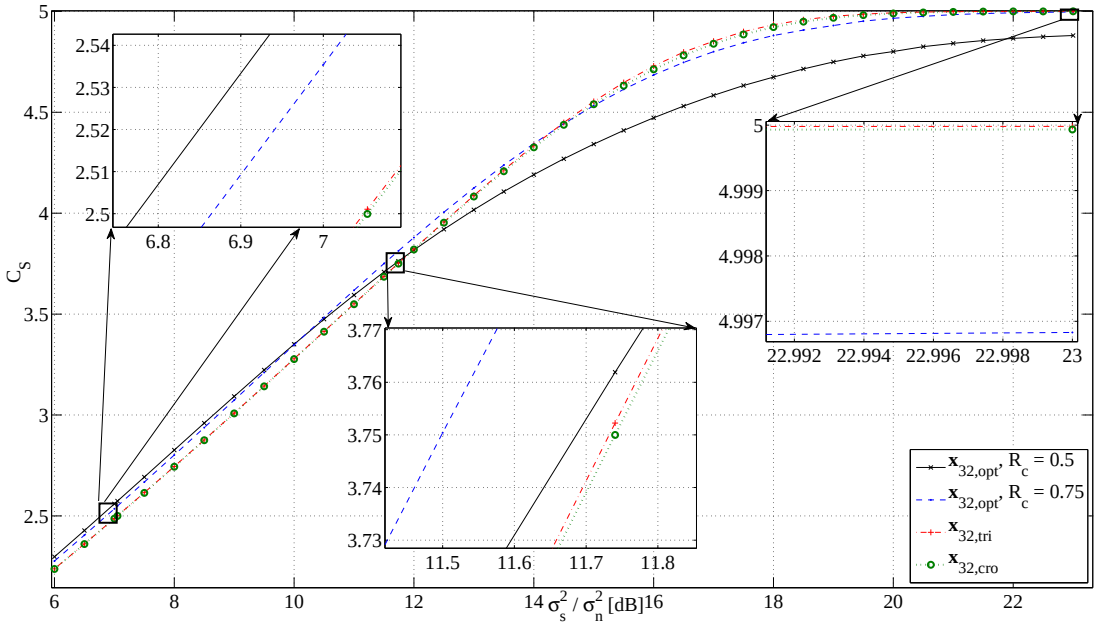


Figure 3.33: Signal set capacity for 32-QAM candidates

3.4.4 Summary for Maximum Signal Set Capacity Mappings

It is well known that the optimum symbol alphabet for an AWGN channel is continuous-valued, and the pdf of the transmit symbols is Gaussian, cf. part 2.5.2.1. For discrete-valued symbols, however, the optimum probabilities of occurrence for the symbols x_l are determined by the Maxwell-Boltzmann distribution, which can be interpreted as a "discretized" Gaussian pdf [32]. As a result, symbols with smaller magnitudes should be transmitted more frequently than symbols with larger magnitudes. This can be achieved either by non-equiprobable, but uniformly spaced symbols or by equiprobable, but non-uniformly spaced symbols. The ultimate gain that this so called *signal shaping* allows is $\frac{\pi e}{6} \triangleq 1.53\,\text{dB}$ per dimension [109]. However, this gain is somewhat unpractical to achieve, as the proof presumes infinite SNR and further it holds only for the limit, that the number of dimensions and thus the number of symbols goes to infinity. More realistic approaches for signal shaping were presented, e.g., in [120, 121, 122, 123], which aim at approximating the probabilities of occurrence of x_l by a Gaussian distribution and are restricted to L-ary ASK. As an example, the shaping gain in [120] for a 4-ASK with respect to the signal set capacity is 0.07 dB. Applying this new constellation independently on both components in order to design a 16-QAM would yield twice the gain, i.e., 0.14 dB. Other approaches to achieve a shaping gain try to minimize the transmit power $\mathscr{P}_x$ for a given number of symbols by means of dense packing of the symbols x_l [124]. In the two-dimensional plane, this reduces to the hexagonal lattice constellations, which we found to be optimal only for the high SNR regime.

In this section, we have tackled the maximization of signal set capacity for L-ary QAM mappings *directly*. We have shown that an optimized 8-QAM constellation yields a shaping gain of 0.8 dB compared to the widely employed 8-PSK and 0.23 dB compared to the 8-

QAM cross constellation at $C_{\mathrm{S}} = 2.25$. For 16-QAM, we presented gains of 0.3 dB and for optimized 32-QAM mappings, 0.25 to 0.27 dB are possible shaping gains.

3.5 Maximum Exploitation of Perfect A Priori Information

3.5.1 Closed-Form Expression of $I_{\mathrm{E1}}(1)$

In this section, we focus on BICM-ID, where an outer decoder provides the APP demapper with a priori information, cf. Section 2.4. From the discussion in Section 2.6, it is clear that if the SNR is sufficiently high, such that the trajectory can tunnel through the gap of both transfer characteristics, the error-floor is determined by the intersection of both curves. The higher the rightmost point from the demapper's transfer characteristic in the EXIT chart, i.e., $I_{\mathrm{E1}}(1)$, the closer this intersection comes to the desired upper right corner of the EXIT chart. A QAM mapping that maximizes $I_{\mathrm{E1}}(1)$ exploits perfect a priori information, i.e., $I_{\mathrm{A1}} = 1$, most efficiently, such that the resulting error-floor is minimized [125]. We now derive the computation of $I_{\mathrm{E1}}(1)$, based on a closed-form expression of the pdf of extrinsic L-values $L_{\mathrm{E},k,\mathrm{m}}$ for the case of perfect a priori knowledge at the demapper [26]. To do so, we determine for this scenario the conditional pdf of extrinsic L-values, assuming that bit b was transmitted, i.e., $p\left(\xi|c_{k,m}=b\right), b \in \{0,1\}$, with $\xi = L_{\mathrm{E},k,\mathrm{m}}$. These pdfs are sufficient to compute $I_{\mathrm{E1}}(1)$ according to (2.81). Note that the quantities in the EXIT chart are *average* MIs, such that the condition in $p\left(\xi|c_{k,m}=b\right)$ has to be interpreted as *any* event, in which the bit b is transmitted. Thus, we have to average over all events, where bit b was transmitted over any bit position $m \in \{0,\ldots,M-1\}$.

First, we derive this pdf under the condition that the respective transmitted bit was a 1. Averaging over all such possibilities yields

$$p\left(\xi|c_{k,m}=1\right) = \frac{1}{M}\sum_{m=0}^{M-1}\frac{1}{2^{M-1}}\sum_{x_{l_1}\in\mathbb{X}_1^m} p\left(\xi|x_{l_1}\right). \tag{3.42}$$

For each $x_{l_1} \in \mathbb{X}_1^m$, we define its counterpart symbol as $x_{l_{0,m}} \in \mathbb{X}_0^m$, where the labels of x_{l_1} and $x_{l_{0,m}}$ differ only in the m-th bit label.

Example: For $M = 4$ and $m = 2$, we consider $x_{l_1} \in \mathbb{X}_1^m$, with $l_1 = 7$. With our nomenclature from (3.1), the corresponding bit labels are 0111. Its counterpart symbol has bit labels 0101. Hence, $x_{l_{0,m}}$ is given by index $l_{0,m} = 5$. ■

In general, the relation $l_{0,m} = l_1 - 2^{M-1-m}$ holds. Assume that x_{l_1} was transmitted and that $n = n_{\mathrm{R}} + \mathrm{j}n_{\mathrm{I}}$ is the particular noise realization of the channel, such that the received sample is $r = x_{l_1} + n$. Now, we consider the extrinsic L-value $L_{\mathrm{E},k,\mathrm{m}}$ of the demapper, if a priori information is ideal and if x_{l_1} was transmitted. From (2.53), we notice that all a priori

probabilities $P_{\mathrm{A},k,m_1} \in \{0,1\}, \forall m_1 \neq m$, if this feedback information is perfect. The corresponding a priori L-values then take on the values $\pm\infty$. If we expand the fraction in (2.64) by $\exp\left(\hat{\mathbf{c}}_{k[m]} \cdot \mathbf{L}^{\mathrm{T}}_{\mathrm{A},k[m]}\right)\Big|_{\hat{\mathbf{c}}_k=\mu^{-1}(x_{l_1})}$, all addends in both numerator and denominator are forced to zero, except for the two candidates x_{l_1} and $x_{l_{0,m}}$. We thus have

$$L_{\mathrm{E},k,\mathrm{m}} = \ln \frac{p\left(r\middle|x_{l_1}\right)}{p\left(r\middle|x_{l_{0,m}}\right)}. \tag{3.43}$$

Inserting the Gaussian conditional pdf from (2.36) we obtain

$$L_{\mathrm{E},k,\mathrm{m}} = \frac{1}{\sigma_{\mathrm{n}}^2}\left(\left|r - x_{l_{0,m}}\right|^2 - \left|r - x_{l_1}\right|^2\right), \tag{3.44}$$

and with $r = x_{l_1} + n$, we get

$$\begin{aligned} L_{\mathrm{E},k,\mathrm{m}} &= \frac{1}{\sigma_{\mathrm{n}}^2}\left(\left|x_{l_1} - x_{l_{0,m}} + n\right|^2 - |n|^2\right) \\ &= \frac{1}{\sigma_{\mathrm{n}}^2}\left(\left|x_{l_1} - x_{l_{0,m}}\right|^2 + |n|^2 + 2\mathrm{Re}\left\{\left(x_{l_1} - x_{l_{0,m}}\right) n^*\right\} - |n|^2\right) \\ &= \frac{1}{\sigma_{\mathrm{n}}^2}\left(d_{l_1,m}^2 + 2\left(x_{l_1,\mathrm{R}} - x_{l_{0,m},\mathrm{R}}\right) n_{\mathrm{R}} + 2\left(x_{l_1,\mathrm{I}} - x_{l_{0,m},\mathrm{I}}\right) n_{\mathrm{I}}\right), \end{aligned} \tag{3.45}$$

where we have used the notation from (3.9) and introduced the Euclidean distance between the symbol x_{l_1} and its counterpart $x_{l_{0,m}}$, which differs in the m-th bit label, as $d_{l_1,m}$, i.e.,

$$d_{l_1,m} = \left|x_{l_1} - x_{l_{0,m}}\right|. \tag{3.46}$$

From the fact that n_{R} and n_{I} are Gaussian i.i.d. with zero mean and variance $\sigma_{\mathrm{n}}^2/2$, follows that $L_{\mathrm{E},k,\mathrm{m}}$ is also Gaussian distributed. Thus, the pdf of $\xi = L_{\mathrm{E},k,\mathrm{m}}$, if x_{l_1} was transmitted, is

$$p\left(\xi\middle|x_{l_1}\right) = \frac{\exp\left(-\frac{\left(\xi - \mu_{l_1,m}\right)^2}{2\sigma_{l_1,m}^2}\right)}{\sqrt{2\pi}\,\sigma_{l_1,m}}, \tag{3.47}$$

where the respective mean and variance can readily be seen from (3.45) to be

$$\mu_{l_1,m} = \frac{d_{l_1,m}^2}{\sigma_{\mathrm{n}}^2} = \frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2} \cdot d_{l_1,m}^2, \tag{3.48}$$

$$\begin{aligned} \sigma_{l_1,m}^2 &= \left(\frac{1}{\sigma_{\mathrm{n}}^2}\right)^2 \cdot \left(4\left(x_{l_1,\mathrm{R}} - x_{l_{0,m},\mathrm{R}}\right)^2 \cdot \frac{\sigma_{\mathrm{n}}^2}{2} + 4\left(x_{l_1,\mathrm{I}} - x_{l_{0,m},\mathrm{I}}\right)^2 \cdot \frac{\sigma_{\mathrm{n}}^2}{2}\right) \\ &= 2 \cdot \frac{d_{l_1,m}^2}{\sigma_{\mathrm{n}}^2} = 2 \cdot \mu_{l_1,m}. \end{aligned} \tag{3.49}$$

In (3.48), we applied equation (2.27) with the normalization $\sigma_s^2 = 1$ from Subsection 2.1.2. Again, the mean value and the variance of the L-value distribution are interconnected, as we have already seen for the a priori L-values in Section 2.6. We can interpret the result as follows: For ideal a priori information, the demapper takes for each bit $c_{k,m}$ only the two possibilities x_{l_1} and $x_{l_{0,m}}$ into account, independent of the total number of symbols L. This corresponds to a BPSK with symbol distance $d_{l_1,m}^2$.

Inserting (3.47) into (3.42) yields the closed-form expression for the distribution of $L_{\mathrm{E},k,\mathrm{m}}$ for ideal a priori knowledge. To do so, we introduce a distance spectrum similar to [23]. Consider all distinct Euclidean distances between symbol pairs that differ in one bit label only and denote them as $\mathrm{d}_{\mathrm{E},s}$, $s \in \{1,\ldots,N_\mathrm{d}\}$, where N_d is the number of distinct distances. Let Ψ_s be the multiplicity of distance $\mathrm{d}_{\mathrm{E},s}$, such that

$$\sum_{s=1}^{N_\mathrm{d}} \Psi_s = M \cdot 2^{M-1}. \tag{3.50}$$

Examples: For QPSK, perfect a priori knowledge means that one bit is exactly known, while the other is currently investigated by $L_{\mathrm{E},k,\mathrm{m}}$. Gray labeling has only $N_\mathrm{d} = 1$ distance, namely the minimum Euclidean distance $\mathrm{d}_{\mathrm{E},1} = 2/\sqrt{2}$, with multiplicity $\Psi_1 = 4$. However, anti-Gray labeling offers another distance, so $N_\mathrm{d} = 2$, $\mathrm{d}_{\mathrm{E},1} = 2/\sqrt{2}$ with $\Psi_1 = 2$ and $\mathrm{d}_{\mathrm{E},2} = 2$ with $\Psi_2 = 2$. This larger distance increases the reliability of the computed extrinsic L-value and thereby offers a higher $I_{\mathrm{E}1}(1)$. ■

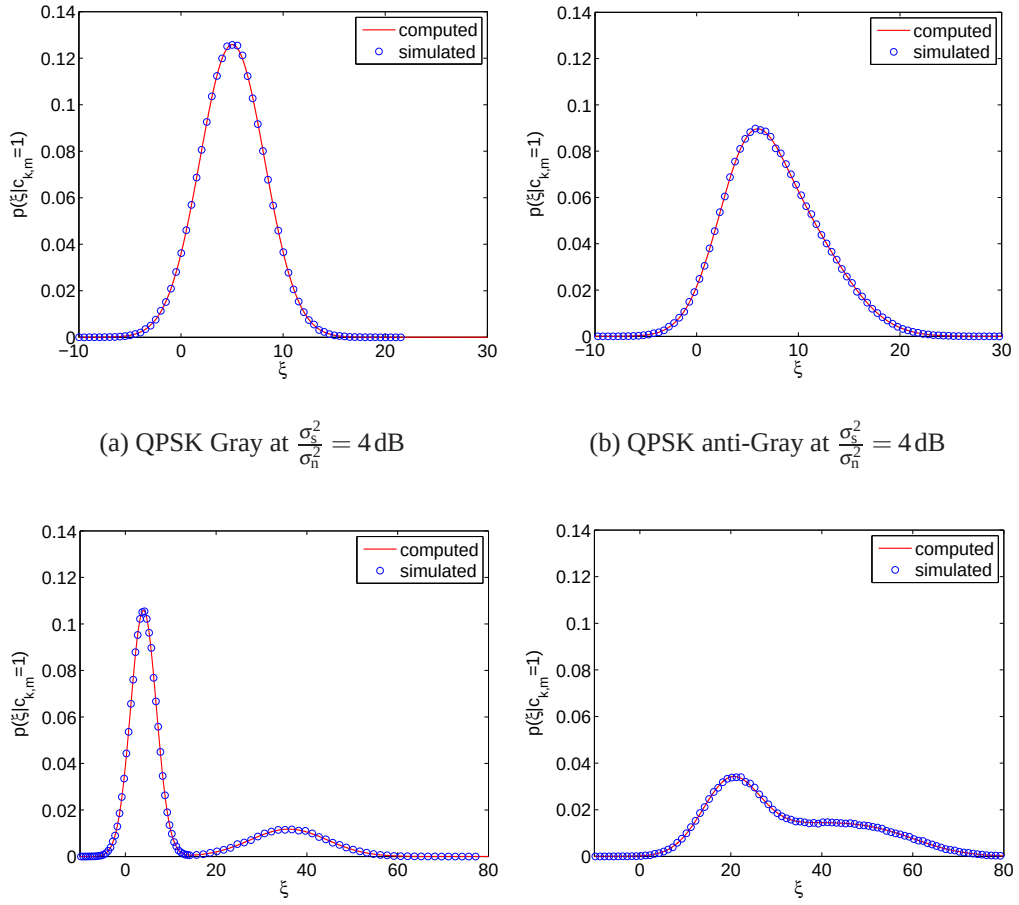
With that, we can write the distribution of $\xi = L_{\mathrm{E},k,\mathrm{m}}$ for perfect a priori knowledge as

$$p\left(\xi | c_{k,m} = 1\right) = \frac{1}{M \cdot 2^{M-1}} \sum_{s=1}^{N_\mathrm{d}} \Psi_s \cdot \frac{\exp\left(-\frac{\left(\xi - \mathrm{d}_{\mathrm{E},s}^2/\sigma_\mathrm{n}^2\right)^2}{4\mathrm{d}_{\mathrm{E},s}^2/\sigma_\mathrm{n}^2}\right)}{2\sqrt{\pi}\mathrm{d}_{\mathrm{E},s}/\sigma_\mathrm{n}}. \tag{3.51}$$

It can be seen that the distribution results as a superposition of Gauss-functions, each having mean and variance that depend on the SNR $\frac{\sigma_s^2}{\sigma_\mathrm{n}^2}$ and the Euclidean distance between symbols, differing in one bit label. Hence, only if the distance spectrum for perfect a priori information reduces to one remaining entry, $N_\mathrm{d} = 1$, the overall distribution is Gaussian. This holds for BPSK and QPSK Gray mapping. For all other cases, the true distribution is not Gaussian.

Examples: The extrinsic L-value distributions for the two QPSK mappings from the previous example are depicted in Figure 3.34 for $I_{\mathrm{A}1} = 1$ at $\frac{\sigma_s^2}{\sigma_\mathrm{n}^2} = 4\,\mathrm{dB}$. It can be seen that the simulated histogram measurements match the predicted pdfs from (3.51). Only the QPSK Gray mapping in Figure 3.34(a) yields a Gaussian pdf. In addition, we show the pdfs for the 16-QAM Gray and anti-Gray mapping from Figure 2.7(b) at $\frac{\sigma_s^2}{\sigma_\mathrm{n}^2} = 10\,\mathrm{dB}$. The corresponding distance spectra are given in [23]. Observe, that the cross-over area $\int_{-\infty}^{0} p\left(\xi | c_{k,m} = 1\right) d\xi$, which corresponds to the BER after a hard decision, is very small for the anti-Gray labeling

and that many large — and thus, reliable, L-values appear. This bit *labeling* was optimized in [23] for the given 16-QAM constellation $\mathbb{X}$. ■



(a) QPSK Gray at $\frac{\sigma_s^2}{\sigma_n^2} = 4\,\text{dB}$

(b) QPSK anti-Gray at $\frac{\sigma_s^2}{\sigma_n^2} = 4\,\text{dB}$

(c) 16-QAM Gray at $\frac{\sigma_s^2}{\sigma_n^2} = 10\,\text{dB}$

(d) 16-QAM anti-Gray at $\frac{\sigma_s^2}{\sigma_n^2} = 10\,\text{dB}$

Figure 3.34: Extrinsic L-value distribution for perfect a priori information

Let us now examine the other condition, that the transmitted bit was a 0. From symmetry considerations, we find

$$p\left(\xi\,|c_{k,m}=0\right) = p\left(-\xi\,|c_{k,m}=1\right). \tag{3.52}$$

This can easily be seen from (3.43), where inverting the fraction or negating the L-value, respectively, yields the desired expression, while all following derivations are equivalent. Inserting (3.52) into (2.81), and using the short hand notation $p\left(-\xi|1\right) = p\left(-\xi|c_{k,m}=1\right)$, yields after some manipulations

$$I_{\text{E1}}(1) \;=\; \frac{1}{2}\int_{\xi\in\mathbb{R}} \left\{ p\left(-\xi|1\right) \left\{ 1-\log_2\left(1+\frac{p\left(\xi|1\right)}{p\left(-\xi|1\right)}\right)\right\}\right.$$

$$+p\left(\xi|1\right)\left\{1-\log_2\left(1+\frac{p\left(-\xi|1\right)}{p\left(\xi|1\right)}\right)\right\}\Bigg\}\,d\xi$$

$$= \quad 1-\int\limits_{\xi\in\mathbb{R}} p\left(\xi|1\right)\log_2\left(1+\frac{p\left(-\xi|1\right)}{p\left(\xi|1\right)}\right)d\xi, \tag{3.53}$$

where the preceding 1 comes from the law of total probability. Further, we have combined both addends, because they are identical, which can easily be seen after substituting $-\xi$ with ξ'.

3.5.2 Mappings for Maximum $I_{\mathrm{E1}}(1)$

We derive the partial derivative of $I_{\mathrm{E1}}(1)$ with respect to x_l in the Appendix A.4 and find

$$\frac{\partial I_{\mathrm{E1}}(1)}{\partial x_l} = -\int\limits_{\xi\in\mathbb{R}}\Omega(\xi)\left\{\log_2\left(1+\frac{p\left(-\xi|1\right)}{p\left(\xi|1\right)}\right)+\frac{p\left(\xi|1\right)\cdot\mathrm{e}^{-\xi}-p\left(-\xi|1\right)}{\ln(2)\cdot\left(p\left(\xi|1\right)+p\left(-\xi|1\right)\right)}\right\}d\xi. \tag{3.54}$$

The function $\Omega(\xi)$ denotes the partial derivative of $p\left(\xi|1\right)$ with respect to x_l, and is

$$\Omega(\xi) = \Omega_0\sum_{m=0}^{M-1}\left(x_l - x_l^{(m)}\right)\cdot\frac{\xi^2-\mu_{l_2,m}^2-2\mu_{l_2,m}}{\mu_{l_2,m}^2\cdot\sqrt{\mu_{l_2,m}}}\cdot\exp\left(-\frac{\left(\xi-\mu_{l_2,m}\right)^2}{4\mu_{l_2,m}}\right), \tag{3.55}$$

with the constant factor

$$\Omega_0 = \left(M\cdot 2^{M+1}\cdot\sqrt{\pi}\cdot\sigma_{\mathrm{n}}^2\right)^{-1} \tag{3.56}$$

and the symbol $x_l^{(m)}$, which is obtained from x_l by inverting the m-th bit label $\mu_m^{-1}(x_l)$. Its index ζ is determined by

$$x_l^{(m)} = x_\zeta \;\text{ with }\; \zeta = l+2^{M-1-m}\cdot\left[1-2\cdot\mu_m^{-1}(x_l)\right]. \tag{3.57}$$

The mean values in (3.55) are computed as

$$\mu_{l_2,m} = \frac{\left|x_l - x_l^{(m)}\right|^2}{\sigma_{\mathrm{n}}^2}. \tag{3.58}$$

It is clear from (3.42), that the cost function $f = I_{\mathrm{E1}}(1)$ depends on both the symbol constellation and the bit labeling. Intuitively, it should be expected that a mapping, which maximizes $I_{\mathrm{E1}}(1)$, is diametrically opposed to a Gray mapping. Despite the mathematical complexity of the gradient $\nabla I_{\mathrm{E1}}(1)$, the original optimization algorithm from Subsection 3.1.2 converges rather fast to the global optima. It turns out that the optimum mapping $\mathbf{x}_{\mathrm{opt}}$ is independent of

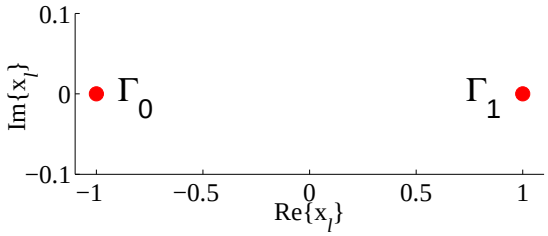


Figure 3.35: Mapping for maximum $I_{\mathrm{E1}}(1)$

the SNR and has the same structure for all orders M: one half of the symbols x_l from $\mathbf{x}_{\mathrm{opt}}$, whose bit labels exhibit even Hamming weight[4], are colocated at -1 and will be denoted as the group Γ_0. The remaining symbols, whose labels possess odd Hamming weight, are all placed at $+1$ and are denoted as Γ_1. Figure 3.35 shows such an optimum mapping. This mapping is justified, because the symbols in each group differ in at least two bits and hence, perfect a priori knowledge of all other $M-1$ bits allows the bit under consideration in (2.64) to be estimated most likely. The demapper only has to decide between the two groups, which are separated with maximum Euclidean distance, bearing in mind that the power constraint (3.4) has to be fulfilled. The maximum value for $I_{\mathrm{E1}}(1)$, which can be achieved through mapping $\mathbf{x}_{\mathrm{opt}}$, can be computed as follows: First, we note that all Euclidean distances from (3.46) are the same, namely 2. The standard deviation for the Gaussian distribution of extrinsic L-values is with (3.49) $2\sqrt{2}\frac{\sigma_{\mathrm{s}}}{\sigma_{\mathrm{n}}}$. Thus, $\mathbf{x}_{\mathrm{opt}}$ results in $I_{\mathrm{E1}}(1) = J\left(2\sqrt{2}\frac{\sigma_{\mathrm{s}}}{\sigma_{\mathrm{n}}}\right)$. From our remark in Subsection 3.3.1, we conclude that this is exactly the (BICM or signal set) capacity of BPSK signaling at an SNR of $\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2}$. Thus, we obtain for an arbitrary mapping at any given SNR $\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2}$ the upper bound Υ on $I_{\mathrm{E1}}(1)$ as

$$I_{\mathrm{E1}}(1) \leq \Upsilon = J\left(2\sqrt{2}\frac{\sigma_{\mathrm{s}}}{\sigma_{\mathrm{n}}}\right) = C_{\mathrm{S}}\left(\mathbf{x} = (-1,+1), \frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2}\right). \tag{3.59}$$

Nevertheless, the optimized mapping $\mathbf{x}_{\mathrm{opt}}$ is obviously unpractical, as it prevents the iterative exchange of extrinsic information to get started. This is demonstrated by the example in Figure 3.36. We have plotted the same transfer characteristics for the outer decoder and for the 16-QAM anti-Gray mapping from Figure 2.7(b) at $\frac{E_{\mathrm{b}}}{N_0} = 3\,\mathrm{dB}$, which we have already depicted in Figure 2.18. In addition, we show the transfer characteristic of the 16-QAM, which is optimized with respect to $I_{\mathrm{E1}}(1)$, also at 3 dB. In this degraded mapping, eight symbols with even Hamming weight labels are located at -1, the other eight at $+1$. While the regular 16-QAM anti-Gray mapping requires slightly more than 3 dB to allow for an open tunnel between inner and outer transfer characteristic, the starting point $I_{\mathrm{E1}}(0)$ of $\mathbf{x}_{\mathrm{opt}}$ is almost zero and suppresses further iterative decoding. To be precise, $I_{\mathrm{E1}}(0) = 9.7 \cdot 10^{-7}$ from simulation and $I_{\mathrm{E1}}(0) = 7.6 \cdot 10^{-7}$ from equation (2.83). The ending point is $I_{\mathrm{E1}}(1) = 0.9905$ from simulation and $I_{\mathrm{E1}}(1) = \Upsilon = 0.9904$ from equation (3.59).

[4]The Hamming weight is defined as the Hamming distance to the all-zero word.

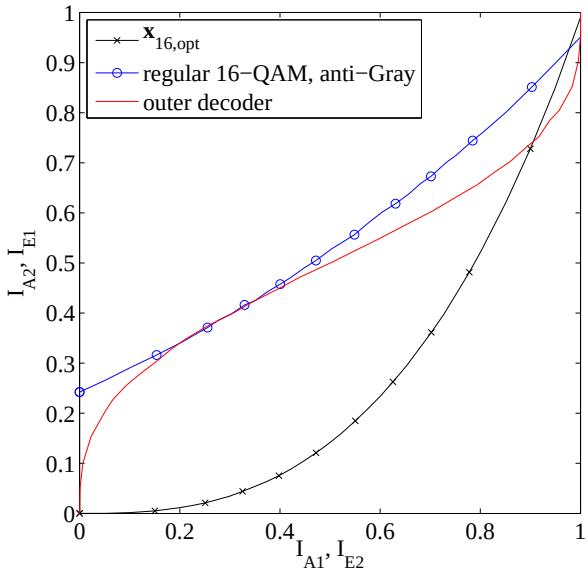


Figure 3.36: EXIT chart for $\mathbf{x}_{\text{opt}}$ and regular 16-QAM anti-Gray at $\frac{E_b}{N_0} = 3\,\text{dB}$

3.5.3 Summary for Maximum $I_{\text{E1}}(1)$ Mappings

We have derived a closed-form expression for the pdf of extrinsic L-values, if the demapper is provided with perfect a priori knowledge. This allows for fast computation of $I_{\text{E1}}(1)$ by a one-dimensional integral, cf. (3.53). Another approach would be the straightforward application of the MI formula (2.69), and considering reduced subsets of symbols, which differ in one bit label only [23]. This however includes integration over the two-dimensional plane of complex numbers. We have related the pdf of extrinsic L-values to the SNR and the distance spectrum of the mapping. No strict rule can be formalized from neither the arithmetic, geometric, nor harmonic mean of (squared) Euclidean distances. Besides some exceptions that we observed (not shown here), a rule of thumb suggests that a large average Euclidean distance yields large $I_{\text{E1}}(1)$. Our calculation of the extrinsic L-value distribution for error-free a priori information has shown that only BPSK and QPSK Gray mappings fulfill the Gaussian assumption. In [126], the extrinsic L-value distribution is examined for 16-QAM Gray mapping, if no a priori information is available, and if the ML demapper applies the max-log approximation. Also in this case, the pdf differs from a Gaussian distribution. Further, we have proposed mappings that utilize perfect a priori information at the demapper most efficiently and thereby maximize $I_{\text{E1}}(1)$. All such mappings for $M > 1$ exhibit two groups of indistinguishable symbols, in which the Hamming weight of the bit labels is either even or odd. Based on this extreme mapping, (3.59) gives an upper bound on $I_{\text{E1}}(1)$ for any mapping. However, the optimized mapping is not suited for iterative demapping, because the MI for no a priori knowledge, $I_{\text{E1}}(0)$, is close to zero. Thus, we will investigate in the next section a tradeoff between no and perfect a priori knowledge.

3.6 Tradeoff between No and Perfect A Priori Knowledge

For a BICM-ID scheme, it is necessary that the demapper can exploit a priori knowledge that it receives from the outer decoder. An anti-Gray mapping assures this condition, because it introduces dependencies among the M bits, which are mapped onto one QAM symbol. In its most pronounced form, such an anti-Gray mapping collapses to two groups of indistinguishable symbols, as we have seen in Figure 3.35. On the other hand, the demapper must provide the outer decoder with reliable extrinsic information, even for no or only unreliable a priori information at its input. Thus, a compromise between the two extreme cases of no and of perfect a priori information must be found. In this section, we examine mappings that maximize the following cost function

$$f = a_{\mathrm{W}} \cdot I_{\mathrm{E1}}(1) + (1 - a_{\mathrm{W}}) \cdot (1 - P_{\mathrm{b}}), \tag{3.60}$$

where P_{b} is the BER from the union bound of (3.28) and $a_{\mathrm{W}} \in [0,1]$ is a weighting factor to balance between the two contradicting goals: for $a_{\mathrm{W}} = 0$, the mapping is optimized, such that P_{b} is minimized, if no a priori information is available, and for $a_{\mathrm{W}} = 1$, the mapping allows for maximum exploitation of perfect a priori information. For every other weighting factor in between these two borders, a compromise is to be found. Note that for $a_{\mathrm{W}} = 0$, we have chosen to minimize the BER for a suboptimum max-log ML demapper, instead of directly maximizing the starting value $I_{\mathrm{E1}}(0) = C_{\mathrm{B}}/M$ in the EXIT chart. The reason is the tremendous mathematical complexity — and thus the overall computation time of the optimization algorithm, if the gradients of the BICM capacity (3.31) and of $I_{\mathrm{E1}}(1)$, cf. (3.54) until (3.58), have to be calculated. As we have shown in Subsection 3.3.1, the BER from max-log ML demapping differs only slightly from the BER of optimum demapping, and the latter was related to $I_{\mathrm{E1}}(0)$ in (2.82). The gradient of P_{b} from (3.28) allows much faster computation and is obtained directly from (3.17) by inclusion of the factor $\mathrm{d_H}\left(x_{l_1}, x_l\right)/M$ for each addend.

3.6.1 16-QAM Tradeoff Mappings

We will now examine optimum mappings for different weighting factors a_{W}, and will concentrate on 16-QAM. Other examples were investigated in [127, 102, 26]. Further, we restrict ourselves on an outer encoder of rate $R_{\mathrm{c}} = \frac{1}{2}$, with two memory elements and generator polynomials (04,07) and a target SNR of $\frac{E_{\mathrm{b}}}{N_0} = 6\,\mathrm{dB}$. Even though the signal set capacity of the regular 16-QAM achieves 0.5 of its maximum capacity at $\frac{E_{\mathrm{b}}}{N_0} = \left(5.118 - 10 \cdot \log_{10}(2)\right)\,\mathrm{dB} \approx 2.11\mathrm{dB}$, as we have shown in Subsection 3.4.2, we decided for the larger target SNR, because the union bound on the BER (3.28) yields unrealistic large values (> 1) otherwise.

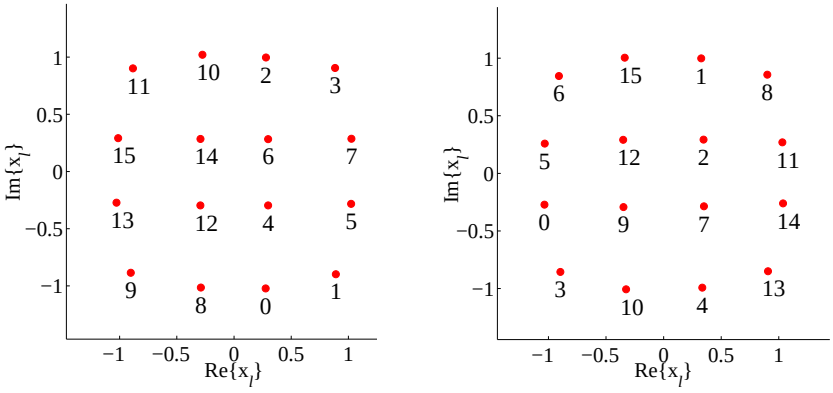
The optimum mapping for $a_{\mathrm{W}} = 0$ is shown in Figure 3.37(a) and resembles the regular 16-QAM Gray mapping from Figure 2.7(b). Adjacent symbols differ in one bit only, but are no longer located exactly on an orthogonal grid. Note that $\frac{E_{\mathrm{b}}}{N_0} = 6\,\mathrm{dB}$ corresponds to

$\frac{\sigma_s^2}{\sigma_n^2} = 9.01\,\text{dB}$, which is close to the target SNR with respect to the BICM capacity for rate $R_c = 0.75$, which was about 9.3 dB, cf. Subsection 3.4.2. There, we have shown that the mapping, which maximizes $I_{E1}(0)$, has its symbols located on two circles. However, this mapping results in a union bound BER at $\frac{\sigma_s^2}{\sigma_n^2} = 9.01\,\text{dB}$ of $P_b = 0.1398$, while the optimized mapping in Figure 3.37(a) achieves $P_b = 0.1277$.

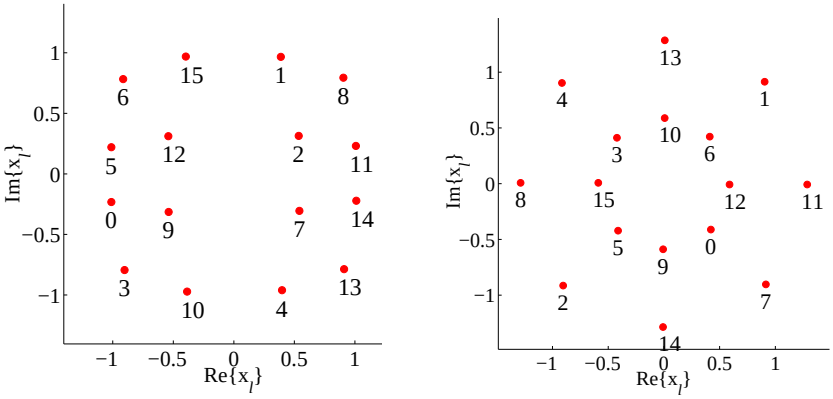
Increasing a_W until 0.4 alters the optimum mapping only on a small scale, but from $a_W = 0.5$ on, the optimum is no longer Gray labeled. As can be seen from Figure 3.37(b), the symbols that belong to groups Γ_0 and Γ_1 are already separated spatially in the complex plane for $a_W = 0.5$. If a_W is further increased, the symbols of the two groups move closer together, as shown in Figure 3.37(c) for $a_W = 0.95$. The tendency towards the extreme mapping from the previous section is visualized in Figure 3.38, where the emphasis in the cost function is put more and more on $I_{E1}(1)$. It can be observed that each half of the symbols will finally merge at ± 1 for $a_W = 1$. For $a_W \geq 0.6$, the gradient algorithm only converges towards the global optimum, if in the initial mapping $\mathbf{x}_0$, the symbol groups with even and odd Hamming weight were already located in separate areas. For most initial conditions, suboptimal mappings (local, but no global optima) were found, in which the two groups are located on two concentric ellipses. If $a_W = 0.8$, the ellipses degenerate into circles, with inner radius $r_i = 0.59$ and the outer radius is $r_o = \sqrt{2 - r_i^2} \approx 1.29$, see Figure 3.37(d). All mapping vectors, except those from Figure 3.38, can be found in the Appendix B.4. Finally, we note that optimized mappings for a different target SNR vary only slightly from the presented ones for 6 dB.

The simulated BER of various 16-QAM candidates for BICM-ID with 30 iterations is plotted in Figure 3.39. Note that only around each turbo-cliff, such a large number of iterations is necessary, while for higher SNR, only 10 to 15 iterations are needed. Further, we have applied a pseudo-random interleaver of a rather large depth $N_{int} = 9.6 \cdot 10^4$, to achieve a pronounced turbo-cliff. The optimized mapping for $a_W = 0$ performs slightly better than the regular 16-QAM Gray mapping. After the first pass through the demapper, the BER of these two mappings is smaller than the BER of mappings with larger weighting factors, but further iterations do not improve the BER. We have also depicted the BER for the 16-QAM mapping, which was optimized with respect to the BICM capacity for a code rate of $R_c = 0.5$, cf. Figure 3.22(a). The demapper in our receiver performs optimum ML demapping, without the max-log approximation. Thus, the mapping, which was optimized with respect to C_B gives a better BER than the mapping for $a_W = 0$, which aims at minimizing the BER after max-log ML demapping. All other mappings in Figure 3.39 are designed to exploit a priori information. The baseline is set by the optimized anti-Gray labeling for a regular 16-QAM constellation [23], which has its turbo-cliff at $\frac{E_b}{N_0} = 3.2\,\text{dB}$. To the best of our knowledge, this is the preeminent BICM-ID scheme for a spectral efficiency of $\eta = 2$. It outperforms a regular 16-QAM Gray mapping in combination with a classical PCC turbo code by 1 dB [128]. It can be observed that increasing a_W yields a lower error-floor, but shifts the turbo-cliff position to higher SNR. With optimized mappings for $a_W \geq 0.9$[5], the error-floor is

[5]The optimized mapping for $a_W = 0.9$ was not shown in Figure 3.37, but is specified in the Appendix B.4.



(a) Global optimum for $a_W = 0$ (b) Global optimum for $a_W = 0.5$



(c) Global optimum for $a_W = 0.95$ (d) Local optimum for $a_W = 0.8$

Figure 3.37: Optimized 16-QAM mappings for different weighting factors a_W

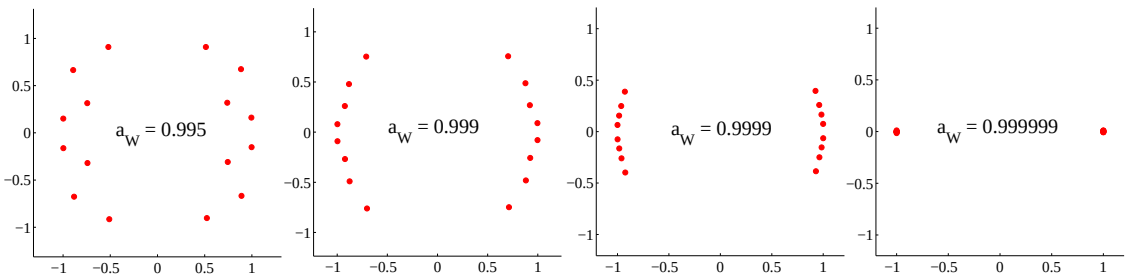


Figure 3.38: Optimized 16-QAM mappings for increasing a_W. Labels are similar as in Figure 3.37(c).

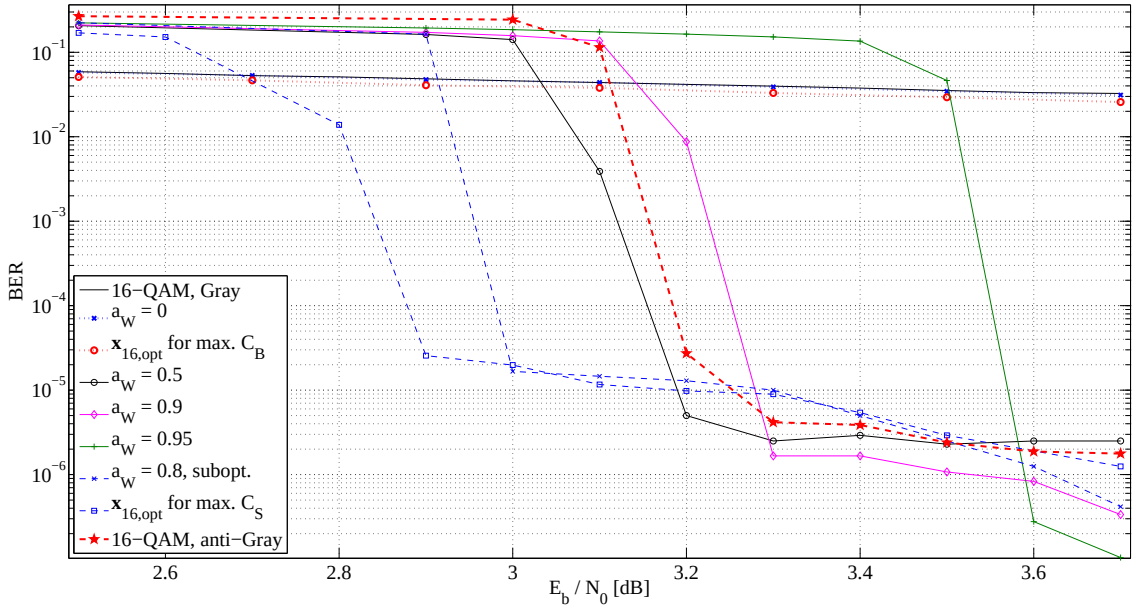


Figure 3.39: BER after iterative demapping for 16-QAM candidates

below that from the baseline mapping. The suboptimal mapping for $a_W = 0.8$ outperforms the reference mapping in both requirements, i.e., the turbo-cliff is about 0.2 dB earlier and the error-floor finally is lower. As the constellation of this mapping is similar to the optimized constellation $\mathbf{x}_{opt}$ with respect to the signal set capacity for a code rate of $R_c = 0.5$, cf. Subsection 3.4.2, which differs only in the ratio of inner to outer radii, we adopt the bit labeling from Figure 3.37(d) to $\mathbf{x}_{opt}$. This lowers the turbo-cliff position even further to $\frac{E_b}{N_0} = 2.9$ dB, which is only about 0.79 dB away from the SNR limit for $C_S = 2$. However, the error-floor is slightly lifted.

We end this chapter by examining the previous mappings, which were optimized for the non-fading AWGN channel, if the channel is substituted by frequency flat Rayleigh fading, see Subsection 2.2.2. In [23], a bit labeling function was optimized for the regular 16-QAM constellation, such that the Chernoff bound for PEP is minimized for the Rayleigh channel. However, simulation results therein have shown that the BER after iterative demapping of this mapping differs only marginally from the optimized mapping for the AWGN channel. Hence, we maintain as a baseline the same anti-Gray mapping as for the AWGN channel. Further, the same convolutional encoder is applied. The results are shown in Figure 3.40, where the number of iterations was reduced to 20. This is because the iterative process converges earlier for this setting. It can be observed that the mappings perform relative to each other in a similar manner as for the AWGN channel, i.e., increasing a_W yields lower error-floors, at the cost of a later turbo-cliff position. All turbo-cliffs are shifted to higher values (about 1.5 dB) and the error-floors are lifted (about one order of magnitude). Again, the mapping, which was optimized with respect to the signal set capacity for the AWGN channel and for $R_c = 0.5$, achieves a good tradeoff between an early turbo-cliff (about 0.2 dB gain compared to the baseline mapping) and low error-floor (slightly higher than that of

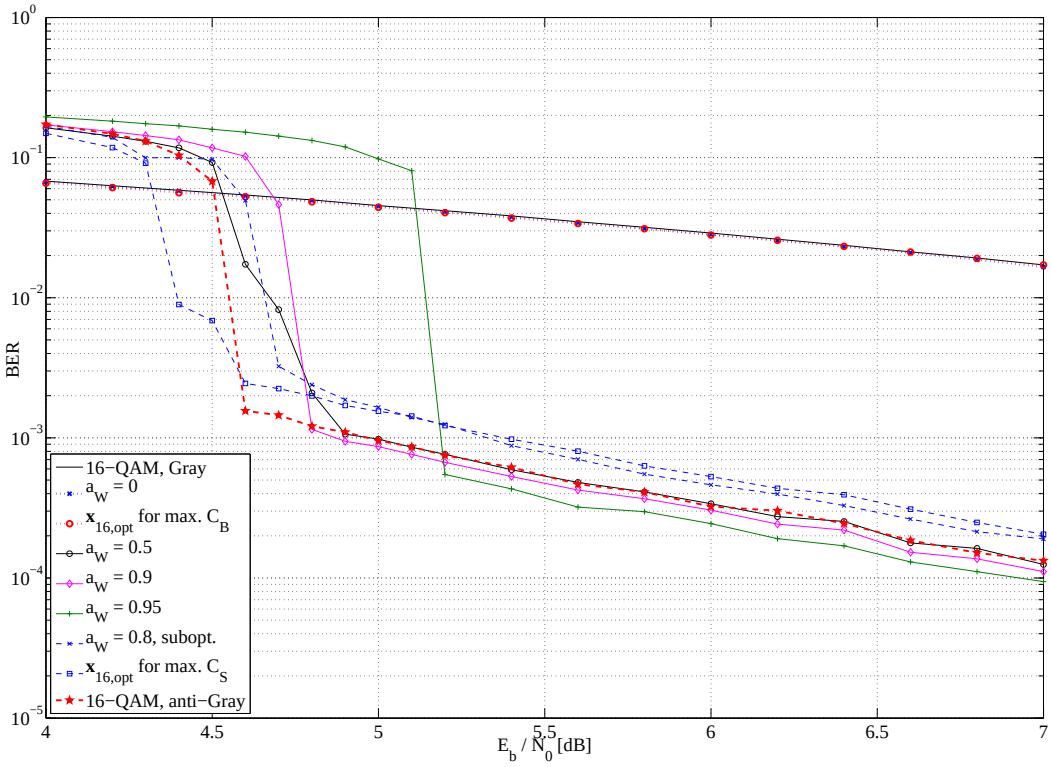


Figure 3.40: BER after iterative demapping for 16-QAM candidates over Rayleigh fading channel

the baseline mapping). It is worth noting that this mapping performs well also on a MIMO channel with iterative demapping, see Subsection 4.1.1 of [45].

3.6.2 Summary for Tradeoff Mappings

In this section, we have combined knowledge of previous optimization targets for QAM mappings. We have seen that Gray-like mappings achieve the smallest BER compared to other mappings at low SNR as well as for non-iterative receivers, but they can not exploit a priori information from an outer decoder. Hence, this decoder has to be capable to further reduce the BER down to acceptable values for a certain application. This is typically specified by the quality of service (QoS) requirement. One example would be the application of an outer turbo code together with a mapping, which is optimized with respect to the BICM capacity. This way, the demapper provides the turbo decoder with the most reliable extrinsic information. Another approach is BICM-ID, in which a rather weak outer decoder feeds back extrinsic information to the demapper, which in turn tries to exploit this additional information most efficiently. From Section 3.5, it became clear that the emphasis in mapping optimization must not be put fully on maximum exploitation of perfect a priori knowledge, but that rather a compromise has to be found. Hence, we proposed 16-QAM mappings, which allow for an early turbo-cliff *and* achieve low error-floors. Depending on a weighting factor a_W, one of these two goals is traded more against the other. Especially the local op-

timum mapping for $a_{\mathrm{W}} = 0.8$, which resembled the mapping that achieves maximum signal set capacity, allowed for capacity approaching performance with an error-floor below 10^{-5} and outperforms the best regular 16-QAM anti-Gray mapping scheme by 0.3 dB. Finally, we applied the proposed mappings on a Rayleigh fading channel, even though they were optimized for the non-fading AWGN channel. The resulting BER after iterative demapping has shown that the mappings perform well, also if the underlying channel model is changed.

Chapter 4

Multidimensional Mappings for Iterative MIMO Detection

In the last chapter, we proposed mappings, which were optimized for various cost functions over a single input single output (SISO) channel. For the BICM-ID scheme, where a receiver performs iterations over the demapper and the outer decoder, we investigated anti-Gray mappings that can lower the error-floor. Recall that a QAM mapping corresponds to a bijective assignment of M bits to a symbol x_l in the two-dimensional complex plane. MIMO schemes allow for additional degrees of freedom, namely in the spatial dimensions. So called multidimensional mapping (MdM) was first introduced in [129, 130]. The design target was to exploit perfect a priori knowledge most effectively at the MIMO demapper. However, in [129], the optimization target was restricted to the most likely PEP. As a result, MdMs were proposed that maximize the minimum Euclidean distance between symbol vectors, which differ in one bit. All other PEPs were neglected. Thus, the presented MdMs perform worse than the MdMs, which we will propose in Section 4.3. In [130], on the other hand, the design goal was the same as in our approach, but only a brute-force search optimization was executed.

We begin this chapter with the delineation of the underlying system model. In Section 4.2, we derive an effective description of MdMs based on block code generator matrices. Using the EXIT chart and a bound on PEP for BICM-ID, we propose in Section 4.3 optimized MdMs for several antenna configurations, which allow for transmission with minimal error-floor and turbo-cliff positions close to capacity limit. The results were published in [27].

4.1 System Model for Multidimensional MIMO Mappings

The transmitter in Figure 4.1 is a combination of BICM (see Section 2.4) with multiple transmit antennas (see Subsection 2.2.3). It consists of a random bit source with output sequence $\mathbf{b}_\kappa$. These information bits are first convolutionally encoded by an outer encoder of rate R_c, then randomly bit-interleaved with an interleaving depth N_int. We will maintain

the rate $R_c = \frac{1}{2}$ recursive systematic convolutional code with generator polynomials $(04, 07)$ from the previous chapter, as well as the interleaving depth of $N_{\text{int}} = 9.6 \cdot 10^4$. At each discrete-time instance k, the input to the multidimensional mapper is a row vector $\mathbf{c}_k$, which consists of $N_t M$ bits. An MdM is defined by the relation $\mathbf{s}_k = \mu(\mathbf{c}_k)$, where $\mu(.)$ is now generalized as the multidimensional mapping function. The block diagram for the iterative MIMO receiver (not shown) is similar to Figure 2.15, except that the APP demapper receives the vector $\mathbf{r}_k$ from the N_r receive antennas and operates according to (2.65). Note that this corresponds to optimal vectorwise APP demapping.

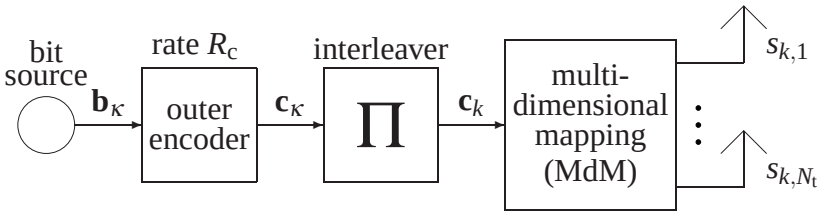


Figure 4.1: MIMO transmitter with multidimensional mapper

Let us first recall conventional mappings for spatial multiplexing MIMO systems, as, e.g., in the V-BLAST architecture. For this special case of MdMs, the concatenation of a QAM mapper with a serial-to-parallel conversion corresponds to the MdM function $\mathbf{s}_k = \mu(\mathbf{c}_k)$, as we have seen in Figure 2.9. We will call this symbolwise mapping, because each transmit symbol is obtained by assigning M bits of $\mathbf{c}_{k,i}$ to $s_{k,i}$, $i \in \{1, \ldots, N_t\}$, independently of all other symbols. The real and imaginary part of $s_{k,i}$ are obtained by the two-dimensional (2d) QAM mapping function, which we will further denote as $\mu_{2d}(.)$. We can thus write the transmit vector, similar to (2.40), as

$$\mathbf{s}_k = \left(s_{k,1}, \ldots, s_{k,i}, \ldots, s_{k,N_t}\right) = \mu(\mathbf{c}_k) = \left(\mu_{2d}(\mathbf{c}_{k,1}), \ldots, \mu_{2d}(\mathbf{c}_{k,i}), \ldots, \mu_{2d}(\mathbf{c}_{k,N_t})\right). \quad (4.1)$$

A linear MIMO detector applies zero forcing (ZF) or minimum mean squared error (MMSE) equalization, followed by symbolwise APP demapping. Vectorwise APP demapping is also applicable and is known to be the optimal detection scheme [40]. Now, let us examine an example for MdM.

Example: We define $\mu_{2d}(.)$ to be the BPSK mapping function from Figure 2.6(a). Although BPSK is a one-dimensional mapping, we keep the notation $\mu_{2d}(.)$. For each transmit antenna $i \in \{1, \ldots, N_t\}$, one bit $\mathbf{c}_{k,i} = c_{k,i,0}$ is mapped according to $\mu_{2d}(c_{k,i,0}) = 2 \cdot c_{k,i,0} - 1$, if conventional symbolwise mapping is used. An MdM now generalizes a mapping by allowing for any one-to-one correspondence from the total bit vector $\mathbf{c}_k$ to symbol vector $\mathbf{s}_k = \mu(\mathbf{c}_k)$. However, we restrict ourselves to the case, where all symbols $s_{k,i}$ are drawn from the same alphabet $\mathbb{X}$. Further, we will only consider standard constellations, such as QPSK and the regular 16-QAM, with symbols located on an orthogonal grid. By restriction to a fixed constellation, we might call the function μ a mapping, whereas its strict definition is only with respect to the bit labeling. Table 4.1 shows an MdM for $N_t = 2$ transmit antennas and BPSK symbols, $M = 1$.

Table 4.1: Example for MdM: $N_\mathrm{t} = 2$ transmit antennas and BPSK, $M = 1$

$\mathbf{c}_{k,1}$	$\mathbf{c}_{k,2}$	$s_{k,1}$	$s_{k,2}$	$\mathbf{u}_{k,1}$	$\mathbf{u}_{k,2}$
0	0	-1	-1	0	0
1	0	$+1$	$+1$	1	1
0	1	$+1$	-1	1	0
1	1	-1	$+1$	0	1

The multidimensional mapper uses the look-up table to allocate $N_\mathrm{t}M$ input bits $\mathbf{c}_k$ *combined* to a transmit symbol *vector* $\mathbf{s}_k$. For example, $\mathbf{s}_k = \mu\left((1,0)\right) = (+1,+1)$. Thus, linear ZF or MMSE detection is not possible anymore, because the transmit symbols $s_{k,i}$ can not be treated independently. Therefore, the right hand side of (4.1) does not apply anymore, i.e., the mapping can not be decoupled. ■

In general, the look-up table consists of $2^{N_\mathrm{t}M}$ rows. For small M and N_t, the table is still manageable. However, for larger M and N_t, its size gets very large, e.g., for $M = 2$ and $N_\mathrm{t} = 4$, the table would have 256 rows. We thus derive an alternative description of MdM in the following section. Note that optimization of MdMs by brute-force search is literally intractable, since there exist $(2^{N_\mathrm{t}M})!$ different possibilities to define the look-up table. With $M = 2$ and $N_\mathrm{t} = 4$, there are about 10^{505} different look-up tables. In [130], a random search was performed for this setting over $5.5 \cdot 10^5$ different MdMs, which is not enough compared to the total number of possible MdMs.

4.2 Generation of MdM

In many cases, an MdM can be described by

$$\mathbf{s}_k = \mu(\mathbf{c}_k) = \left(\mu_{2\mathrm{d}}(\mathbf{u}_{k,1}), \ldots, \mu_{2\mathrm{d}}(\mathbf{u}_{k,i}), \ldots, \mu_{2\mathrm{d}}(\mathbf{u}_{k,N_\mathrm{t}})\right), \tag{4.2}$$

in which $\mathbf{u}_{k,i}$ are M-dimensional subvectors

$$\mathbf{u}_{k,i} = \left(u_{k,i,0}, \ldots, u_{k,i,m}, \ldots, u_{k,i,M-1}\right) \tag{4.3}$$

of an encoded vector of length $N_\mathrm{t}M$, which is obtained by

$$\mathbf{u}_k = \left(\mathbf{u}_{k,1}, \ldots, \mathbf{u}_{k,i}, \ldots, \mathbf{u}_{k,N_\mathrm{t}}\right) = \mathbf{c}_k \cdot G. \tag{4.4}$$

Comparison with (2.44) shows that $\mathbf{u}_k$ is the result of linear block encoding with G. The generator matrix G is square, non-singular and of dimension $N_\mathrm{t}M$. Additions and multiplications in (4.4) have to be performed in the Galois field GF(2). From (4.2), we observe that an

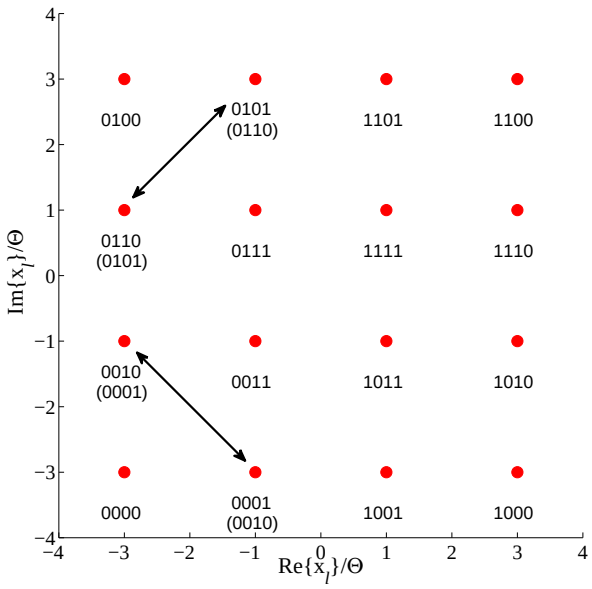


Figure 4.2: Reference mappings for 16-QAM: Gray mapping $\mu_{2\mathrm{d}}(.)$ and modified mapping $\mu'_{2\mathrm{d}}(.)$ (with changes in brackets)

MdM can be constructed by applying conventional symbolwise mapping $\mu_{2\mathrm{d}}(.)$ to the block coded bits $\mathbf{u}_k$. For the example in Table 4.1, we find

$$G = G_2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}. \tag{4.5}$$

The subindex "2" of G_2 indicates its dimensionality. For higher dimensions ($N_\mathrm{t}M > 2$), it is not always possible to describe any MdM by a linear precoder as in (4.4). However, all optimized MdMs presented in Section 4.3 allow this description. In Table 4.1, the coded bits $\mathbf{u}_k$ are also listed in the two rightmost columns.

$\mu_{2\mathrm{d}}(.)$ can thus be regarded as a two-dimensional reference mapping, which produces the MdM after precoding with G. For the rest of this chapter, we define the following reference mappings: If $M = 1$, we use the previous BPSK definition, if $M = 2$, we let $\mu_{2\mathrm{d}}(.)$ be the QPSK Gray mapping from Figure 2.6(b). Further, we will examine $M = 4$, for which we will make use of two different reference mappings shown in Figure 4.2. One is the 16-QAM Gray mapping from Figure 2.7(b) and will be denoted as $\mu_{2\mathrm{d}}(.)$. If we swap two times two symbols as indicated by the arrows in Figure 4.2, we obtain a modified mapping $\mu'_{2\mathrm{d}}(.)$, which is per Definition 2.1 not a Gray mapping anymore.

Figure 4.3 depicts an alternative transmitter compared to Figure 4.1 for MdM generation, using equations (4.2) until (4.4). Multiplication of the input bits $\mathbf{c}_k$ with generator matrix G corresponds to linear block precoding with code rate $R_\mathrm{c} = 1$. The block encoded bits $\mathbf{u}_k$ are input to a conventional V-BLAST transmitter, where M bits are assigned to complex, i.e., two-dimensional symbols. Finally, these symbols are serial-to-parallel converted before transmission over the N_t antennas, similar to Figure 2.9.

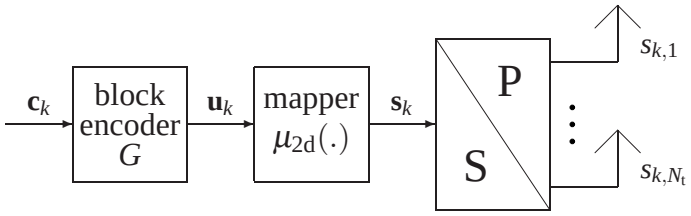


Figure 4.3: Alternative scheme for MdM transmission

At this point, it is worth mentioning that also mappings for a single-antenna transmitter ($N_\mathrm{t} = 1$) can be described by a reference mapping $\mu_{2\mathrm{d}}(.)$ and an MdM table. With properly defined $\mu_{2\mathrm{d}}(.)$, the linear precoder with generator matrix notation can be used as well. As an example, it can easily be verified that the QPSK anti-Gray mapping corresponds to block encoding with G_2 in (4.5), followed by QPSK Gray mapping. This idea will be picked up again in Chapter 5.

Finally, we introduce a short hand notation that will simplify the description of the generator matrices G. Let $\mathbf{1}$ be a column vector of dimension $N_\mathrm{t}M$, where all entries are 1. With the notation from Subsection 3.1.1, we can write $\mathbf{1} = \mathbf{1}_{N_\mathrm{t}M}^\mathrm{T}$. Further, let $\overline{\mathbf{e}_i}$, $i \in \{1,\ldots,N_\mathrm{t}M\}$, be the i-th inverted unit column vector[1] of dimension $N_\mathrm{t}M$, i.e., a vector, with all entries equal to 1, except the i-th position, where there is a 0. Then, G_2 in (4.5) can be written as $G_2 = (\mathbf{1}\ \overline{\mathbf{e}_2})$.

4.3 Design Criteria and MdM Optimization

Optimization of MdMs can be considered for two applications. First, if the receiver does not perform iterative demapping, i.e., the demapper is not supplied with a priori knowledge, it is obvious that the optimum MdM is conventional symbolwise Gray mapping. Hence, the generator matrix G is the identity matrix and the mapping functions $\mu_{2\mathrm{d}}(.)$ as defined in the previous section are used. Note that an identity matrix is a non-singular matrix with the maximum number of 0 entries, which means that the amount of introduced dependencies by the precoder is minimal (nonexistent).

Secondly, a receiver, which iterates over demapper and outer decoder, may utilize a priori information at the demapper. As in Section 3.5, we focus on the case of perfect a priori information. This means that the outer decoder provides perfect knowledge about all bits $\mathbf{c}_k$, except the bit $c_{k,p}$ under consideration in (2.65). Utilizing this knowledge most efficiently corresponds to maximizing the MI $I_{\mathrm{E}1}(1)$ in the EXIT chart. From the last chapter, it is clear that the higher $I_{\mathrm{E}1}(1)$, the closer the intersection of the transfer characteristics belonging to the demapper and the outer decoder comes to the desired upper right corner of the EXIT chart. Thereby, the error-floor after iterative demapping is lowered. Another approach to

[1] $\mathbf{1}$ and $\overline{\mathbf{e}_i}$ are the only exceptions to our conventions that vectors are treated as *row* vectors.

reduce the error-floor is to minimize the average PEP for perfect a priori knowledge at the demapper. It was shown in [33, 92] that for Rayleigh fading channel coefficients and high SNR, this PEP is monotonically decreasing with the harmonic mean D_{h} of squared Euclidean distances of symbol vectors, which differ in one bit label only. The harmonic mean is defined as

$$D_{\mathrm{h}} = \left(\frac{1}{N_{\mathrm{t}}M \cdot 2^{N_{\mathrm{t}}M-1}} \sum_{p=0}^{N_{\mathrm{t}}M-1} \sum_{\mathbf{x}_l \in \tilde{\mathbb{X}}_1^p} \left| \mathbf{x}_l - \mathbf{x}_l^{(p)} \right|^{-2} \right)^{-1}, \tag{4.6}$$

where symbol $\mathbf{x}_l^{(p)}$ is the counterpart symbol vector of $\mathbf{x}_l$, which has a 0 as the p-th bit label, but all other labels are identical to those from $\mathbf{x}_l$. Similar to (3.57), we can write

$$\mathbf{x}_l^{(p)} = \mathbf{x}_\zeta \text{ with } \zeta = l - 2^{N_{\mathrm{t}}M-1-p}. \tag{4.7}$$

The second summation in (4.6) is with respect to all $2^{N_{\mathrm{t}}M-1}$ symbol vectors from $\tilde{\mathbb{X}}$, whose p-th bit label is 1. The set $\tilde{\mathbb{X}}$ and its subsets $\tilde{\mathbb{X}}_1^p$ are as defined for equation (2.65). An optimal MdM for iterative detection should now have maximum D_{h}.

4.3.1 MdM with BPSK and QPSK

The following theorem states optimality for MdMs with BPSK and QPSK as the underlying two-dimensional mapping for the case of perfect a priori knowledge.

Theorem 4.1 *For BPSK and QPSK, i.e., $M \in \{1,2\}$, and an arbitrary number of transmit antennas N_t, D_h is maximized, if a generator matrix of the form*

$$G_{N_tM} = \left(\mathbf{1}\ \overline{\mathbf{e}_2}\ \overline{\mathbf{e}_3}\ \ldots\ \overline{\mathbf{e}_{N_tM}} \right) \tag{4.8}$$

is applied, followed by the two-dimensional mapping $\mu_{2d}(.)$, as defined in the previous section. The maximum harmonic mean for this MdM is

$$D_{h,max} = \frac{4}{M} \cdot \frac{(N_tM)^2}{N_tM+1} \overset{\mathscr{P}_x=1}{=} \frac{4E_bR_c}{N_r} \cdot \frac{(N_tM)^2}{N_tM+1}. \tag{4.9}$$

Proof: Define the shortest squared Euclidean distance between two distinct symbol vectors as δ. For $M=1$, $\delta = 2^2 = 4$, while for $M=2$, $\delta = \left(2/\sqrt{2}\right)^2 = 2$, thus for both cases, we have $\delta = 4/M$. Each bit of $\mathbf{u}_k$ is mapped to an independent dimension — be it the N_{t} spatial dimensions, if $M=1$ or in addition the two orthogonal dimensions of the in- and quadrature-phase component, if QPSK is applied. It follows that two block code words with a Hamming distance of l bits are mapped to symbol vectors with squared Euclidean distance $l \cdot \delta$. It is therefore sufficient to maximize the harmonic mean of Hamming distances of code word pairs $\mathbf{u}_k$, belonging to input vectors $\mathbf{c}_k$, which differ in one bit. Denote these Hamming distances between the corresponding code words as $d_{\mathrm{H},i}$, $i \in \{0,\ldots,N_{\mathrm{t}}M \cdot 2^{N_{\mathrm{t}}M-1}-1\}$. For

the moment, we neglect the two facts that code words have to be distinct and that the $d_{\mathrm{H},i}$ have to be discrete-valued. We only restrict $d_{\mathrm{H},i}$ to be positive and upper bounded by $N_\mathrm{t}M$. Obviously, the harmonic mean is a $\cup$-convex function of the $d_{\mathrm{H},i}$ and thus is maximal at its boundary, i.e., if all $d_{\mathrm{H},i} = N_\mathrm{t}M$. Since this is not allowed for a one-to-one correspondence between $\mathbf{c}_k$ and $\mathbf{u}_k$ (all entries of G would be 1 in this case, thus G would be a singular matrix), we decrease as many $d_{\mathrm{H},i}$ as necessary to allow for unique encoding. This would be the case, if the inversion of the p-th bit label as in (4.7) for all but one bit positions p in the input vectors $\mathbf{c}_k$ would yield $d_{\mathrm{H},i} = N_\mathrm{t}M - 1$, while one inversion, e.g., of the first bit, would result in $d_{\mathrm{H},i} = N_\mathrm{t}M$. This can be achieved by a linear block code with the generator matrix defined in (4.8).

Hence, we have $2^{N_\mathrm{t}M-1}$ times the Hamming distance $d_{\mathrm{H},i} = N_\mathrm{t}M$ and $2^{N_\mathrm{t}M-1} \cdot (N_\mathrm{t}M - 1)$ times $d_{\mathrm{H},i} = N_\mathrm{t}M - 1$. The harmonic mean of these values equals $(N_\mathrm{t}M)^2/(N_\mathrm{t}M + 1)$. Multiplication with δ yields (4.9), and the right hand side thereof follows from (2.43). □

4.3.2 MdM with 16-QAM

For the regular 16-QAM constellation in Figure 4.2, the shortest squared Euclidean distance between two distinct symbol vectors is $\delta = \left(\frac{2}{\sqrt{10}}\right)^2 = \frac{2}{5}$. However, it is not possible anymore to state that code words differing in l bits are mapped to symbol vectors with squared Euclidean distance $l \cdot \delta$. This is because four bits may be assigned to one antenna, as in the conventional symbolwise mappings, and are thus mapped to only two dimensions. Therefore, Theorem 4.1 can not be applied straightforwardly.

Let us first consider $N_\mathrm{t} = 1$. Using a generator matrix $\widetilde{G_4} = (\overline{\mathbf{e}_1}\ \mathbf{1}\ \overline{\mathbf{e}_3}\ \overline{\mathbf{e}_4})$, which is similar to (4.8), followed by Gray mapping $\mu_{2\mathrm{d}}(.)$, results in a 16-QAM anti-Gray mapping with $D_\mathrm{h} = 10.63 \cdot \frac{E_\mathrm{b} \cdot R_\mathrm{c}}{N_\mathrm{r}}$, which is already close to optimum. The following squared Euclidean distances occur: 16 times $5 \cdot \delta$, 8 times $8 \cdot \delta$ and 8 times $13 \cdot \delta$. Only if, in addition, four symbols are swapped, as indicated by the arrows in Figure 4.2, or in other words, if we apply $\widetilde{G_4}$ according to (4.8), followed by the modified mapping $\mu'_{2\mathrm{d}}(.)$, we obtain the best 16-QAM anti-Gray mapping for the AWGN channel, cf. Figure 2.7(b). We will denote this mapping as 16-QAM 2d AG. As a consequence of this symbol swapping, half of the symbol pairs, which first had squared Euclidean distances of $8 \cdot \delta$ are now $10 \cdot \delta$ apart. Computing D_h for this case yields about $10.86 \cdot \frac{E_\mathrm{b} \cdot R_\mathrm{c}}{N_\mathrm{r}}$, which is lower than $D_{\mathrm{h,max}}$, if either BPSK with $N_\mathrm{t} = 4$ or QPSK with $N_\mathrm{t} = 2$, respectively, is applied. From (4.9), we compute $D_{\mathrm{h,max}} = 12.8 \cdot \frac{E_\mathrm{b} \cdot R_\mathrm{c}}{N_\mathrm{r}}$. All three schemes transmit $\eta = 4 \cdot R_\mathrm{c}$ information bits per channel usage. As we have seen in Figure 3.40 and as was shown in [23], the 16-QAM 2d AG mapping performs over a Rayleigh channel almost identically as the mapping, which was optimized for Rayleigh fading. The latter mapping can be found in [23] and yields a slightly higher harmonic mean $D_\mathrm{h} = 10.88 \cdot \frac{E_\mathrm{b} \cdot R_\mathrm{c}}{N_\mathrm{r}}$ than 16-QAM 2d AG.

For $N_\mathrm{t} = 2$, we applied all possible 8-dimensional non-singular matrices with the maximum number of 1 entries, i.e., in which all except seven entries are 1, followed by either $\mu_{2\mathrm{d}}(.)$ or

$\mu'_{2\mathrm{d}}(.)$. The highest value for D_h was obtained with $\widetilde{G_8} = (\overline{\mathbf{e}_1}\ \mathbf{1}\ \overline{\mathbf{e}_3}\ \overline{\mathbf{e}_4}\ \overline{\mathbf{e}_5}\ \overline{\mathbf{e}_6}\ \overline{\mathbf{e}_7}\ \overline{\mathbf{e}_8})$ in combination with $\mu'_{2\mathrm{d}}(.)$ and is about $25.13 \cdot \frac{E_\mathrm{b} \cdot R_\mathrm{c}}{N_\mathrm{r}}$. This is again smaller than the optimum harmonic mean of MdMs with BPSK or QPSK symbols at the same spectral efficiency: for $M = 2$ and $N_\mathrm{t} = 4$, we get $D_{\mathrm{h,max}} = 28.44 \cdot \frac{E_\mathrm{b} \cdot R_\mathrm{c}}{N_\mathrm{r}}$. This suboptimality of high order mappings ($M > 2$) is due to the fact that the number $N_\mathrm{t}M$ of bits, which are mapped to a symbol vector $\mathbf{s}_k$, is larger than the number of independent dimensions available in $\mathbf{s}_k$, which is N_t for real-valued and $2N_\mathrm{t}$ for complex-valued symbols x_l.

Extensive investigation on MdM optimization by means of the binary switching algorithm (BSA) [131] was done in [132]. The search was not restricted to MdMs that can be generated by linear precoding. However, no MdM could be found, which has a larger cost function D_h than the MdMs designed with the proposed matrices in this chapter.

We end this section with the following theorem.

Theorem 4.2 *The harmonic mean D_h in (4.9) is constant with respect to the number of transmit antennas N_t, if conventional symbolwise mapping is applied.*

Proof: Compared with a single-antenna transmitter, the frequencies of occurrence of squared Euclidean distances in (4.9) are N_t times higher, if N_t transmit antennas and symbolwise mapping are applied. This up-scaling does not change the harmonic mean. □

Hence, if we apply symbolwise mapping with 16-QAM 2d AG for $N_\mathrm{t} = 2$, we have again just $10.86 \cdot \frac{E_\mathrm{b} \cdot R_\mathrm{c}}{N_\mathrm{r}}$.

4.4 Simulation Results of MdMs

We have simulated the BER of an iterative MIMO receiver as a function of SNR with $4.8 \cdot 10^7$ information bits per SNR value. In all cases, we have applied the optimized MdMs that provide maximum D_h, as explained in the previous section, except in some cases, where conventional symbolwise mapping is applied as a reference, which is indicated by the notation "2d" in the figures.

Figure 4.4 shows the BER after 20 iterations for systems with an information rate of $\eta = 2$ bits per channel usage. As can be concluded from Theorem 4.1, the error-floors of both MdMs, 4×2 BPSK and 2×2 QPSK, are the same. Both schemes transmit 4 encoded bits per channel usage over 4 independent dimensions: 4 spatial dimensions, provided by the i.i.d. channel coefficients for BPSK and 2 spatial dimensions subdivided into 2 orthogonal dimensions of the complex plane in case of QPSK. This orthogonality among each of the two dimensions benefits at low SNR, visible in an earlier turbo-cliff for 2×2 QPSK. The MdM for 4×4 BPSK has an additional receive diversity advantage of about 0.8 dB, as the number of receive antennas is doubled. Note that from our SNR definition (2.43), which

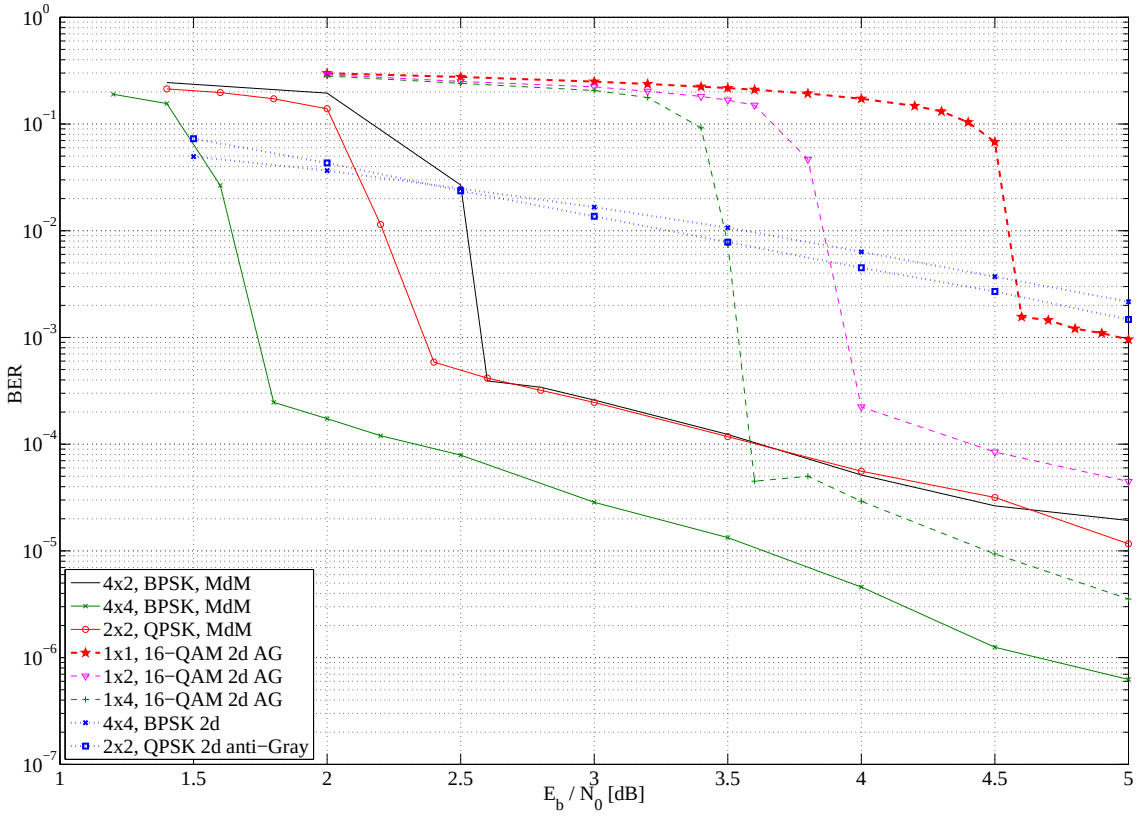


Figure 4.4: BER for MdM and symbolwise mappings for various antennas configurations and modulation orders, if $\eta = 2$

considers the total received power and not the power per receive antenna, we should bear in mind an additional 3 dB array gain [40], as twice as many receive antennas capture twice the power. The BER for conventional mapping with 16-QAM 2d AG for $N_\mathrm{r} \in \{1, 2, 4\}$ receive antennas, respectively, is also depicted. For $N_\mathrm{r} = 2$, the error-floor is higher than that of the two MdMs, which employ the same number of receive antennas, because D_h is smaller as discussed in Subsection 4.3.2. Note that the BER curve of 16-QAM 2d AG for 1×1 was already depicted in Figure 3.40. For comparison, the BER for symbolwise mappings for 4×4 BPSK and 2×2 QPSK with anti-Gray labeling is also shown. No turbo-cliff can be observed in the depicted SNR interval. In summary, we can observe that MdMs allow for lower error-floors, because they were optimized with respect to this design goal. However, it turns out that they also achieve earlier turbo-cliff positions. The reason for that is that the demapper's transfer characteristic in case of the proposed MdMs match better to the curve of the outer decoder in the EXIT chart. We will give an EXIT example for MdMs with a spectral efficiency $\eta = 4$.

If we apply the previous $N_\mathrm{t} = 4$ BPSK and the $N_\mathrm{t} = 2$ QPSK MdMs in a scenario, where the receiver employs only $N_\mathrm{r} = 1$ antenna, we can see from Figure 4.5, that their performance is still superior to a conventional 1×1 system with 16-QAM 2d AG mapping, which has the same spectral efficiency. Note that MdM utilizes multiple transmit antennas for spatial multiplexing as opposed to space-time block or trellis codes, where additional redundancy is

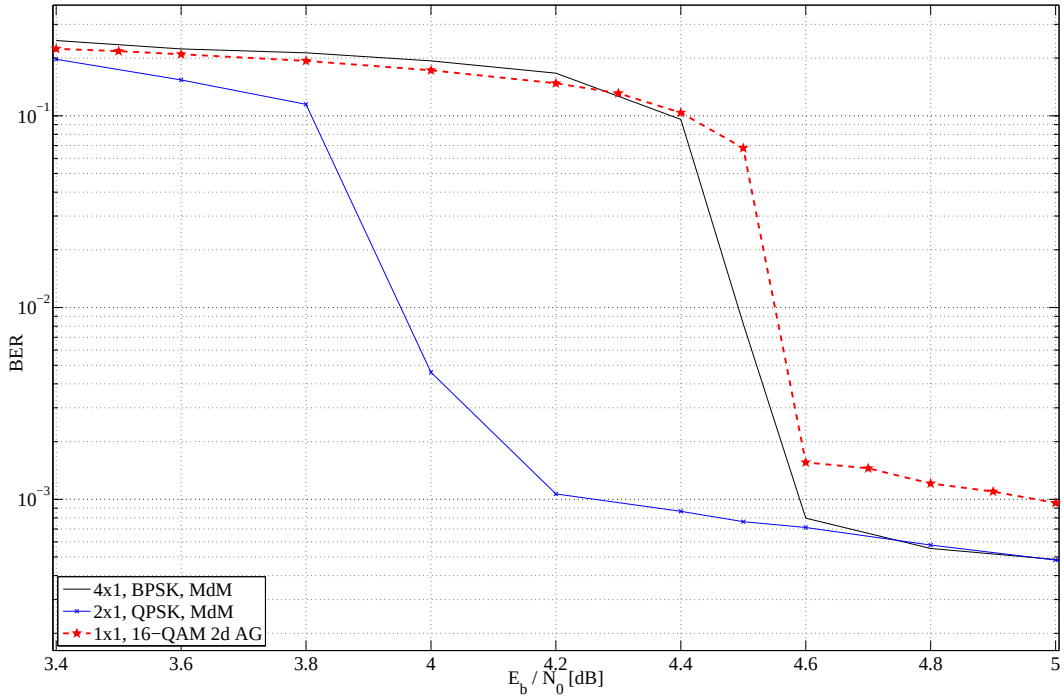


Figure 4.5: BER for MdMs and symbolwise mapping, for $N_\mathrm{r} = 1$ receive antenna and $\eta = 2$

distributed over these antennas, in order to achieve transmit diversity.

Obviously, the performance of the $N_\mathrm{t} \times 1$ MdMs will be degraded, if the channel coefficients become correlated. Let us examine as an example the 4×1 antenna configuration with BPSK symbols more closely. Assume, e.g., the extreme case, where all fading coefficients are 1. The channel matrix from (2.41) then is $H_k = (1\ 1\ 1\ 1)^\mathrm{T}$. All $2^{N_\mathrm{t}M} = 16$ possible superpositions of the 4 transmitted BPSK symbols are determined by the Hamming weight of the code word $\mathbf{u}_k$. If $\mathbf{u}_k$ is the all-zero word, then the received sample is $r_k = -4$, when the noise is neglected. We denote this group, which consists of only one element, as Γ_0. In general, if the Hamming weight of $\mathbf{u}_k$ is l, the group Γ_l results, where all superpositions are colocated at $2 \cdot l - 4$, see Figure 4.6[2]. The two groups Γ_1 and Γ_3 comprise each four indistinguishable symbols, the group Γ_2 includes even six symbols. Yet, an iterative demapper might resolve these ambiguities, when the reliability of a priori information increases. If three bits are known perfectly at the demapper, and if these bits have Hamming weight l, the demapper only has to decide between the two groups Γ_l and Γ_{l+1}, which have Euclidean distance of 2. Nevertheless, the iterative demapping process is exacerbate for this scenario.

On the other hand, assume, e.g., that $H_k = \left(\mathrm{e}^{\mathrm{j}0}, \mathrm{e}^{\mathrm{j}\frac{\pi}{4}}, \mathrm{e}^{\mathrm{j}\frac{\pi}{2}}, \mathrm{e}^{\mathrm{j}\frac{3\pi}{4}}\right)^\mathrm{T}$. Now, the resulting 16 symbols are all distinct. The superimposed constellation resembles that of the optimized 16-QAM with respect to the BICM capacity for small to medium SNR, cf. Figure 3.22(a). The ratio of outer to inner radius can be computed as $r_o/r_i = \left(2+\sqrt{2}\right)/\sqrt{2} \approx 2.414$. On

[2]This mapping has average power $4 = N_\mathrm{t}$; for the SNR definition (2.43), it is however set to unity.

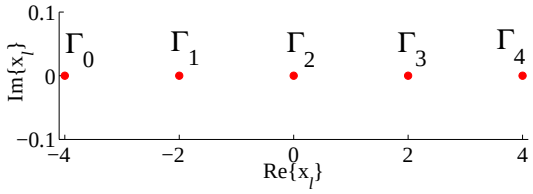


Figure 4.6: Superimposed 4×1 MdM with BPSK, all channel coefficients are 1

average, the resulting constellations in combination with the precoding by G_4, have better properties for iterative demapping than the regular 16-QAM anti-Gray mapping. The effects of different fading coefficients on the superimposed mapping were examined in [133] and the results were applied over the AWGN channel. Another approach in [134] has lead to the so called sigma-mappings, for which a suboptimal demapper with lower complexity was proposed in [135]. Its underlying principle is soft interference cancellation of each BPSK symbol by means of the fed back a priori information, similar to the techniques used for iterative MIMO detection [45].

In Figure 4.7, the BER after 40 iterations for MIMO systems with a spectral efficiency $\eta = 4$ is depicted. With the proposed MdMs, no errors could be measured after the turbo-cliff, as opposed to conventional mappings. The suboptimality of higher order modulation, as discussed in Subsection 4.3.2, can clearly be seen. Symbolwise mapping with 16-QAM 2d AG for $N_\mathrm{t} = N_\mathrm{r} = 2$ performs about 3.5 dB worse than the 4×4 QPSK MdM. An additional 3 dB array gain for the latter scheme must be considered, too. The 16-QAM MdM for $N_\mathrm{t} = N_\mathrm{r} = 2$ achieves a BER of 10^{-4} about 0.63 dB earlier than the conventional 16-QAM 2d AG for the same antenna configuration. The BER for the 1×2 16-QAM 2d AG system, which transmits only 2 bits per channel usage, is also shown. As implied in the discussion of Theorem 4.2, the error-floor is independent of the number of transmit antennas N_t for symbolwise mappings. The $N_\mathrm{t} = 4$ QPSK MdM utilizes 8 independent dimensions, 4 of which are orthogonal to each other, while the 16-QAM MdM for $N_\mathrm{t} = 2$ only provides 4 dimensions. Hence, the turbo-cliff of 4×4 QPSK MdM occurs earlier.

The optimized 4×4 QPSK MdM with the outer encoder of rate $R_\mathrm{c} = \frac{1}{2}$ in Figure 4.7 comes close to its signal set capacity limit. The turbo-cliff is at 2 dB, which is only 0.4 dB away from the capacity limit, which was computed in [44] to be 1.6 dB for this setup. In this publication, a 4×4 MIMO system with conventional QPSK Gray mapping was combined with an outer turbo code of rate $R_\mathrm{c} = \frac{1}{2}$ and a turbo-cliff at slightly less than 3 dB without measurable error-floor was achieved. A similar measurement was done in [130], but with finer granularity on the SNR axis, which revealed the turbo-cliff at 2.2 dB. In [112], the outer turbo code was replaced by an irregular LDPC code of the same code rate, which was optimized by means of the EXIT chart. As a result, a BER of 10^{-4} was achieved after 100 internal iterations in the sum-product algorithm at an SNR of only 1.95 dB. As we have mentioned in Subsection 2.3.4, iterative decoding of a concatenated coding schemes might result in a loss of up to 0.5 dB [75]. This means that optimum flat decoding of the presented MIMO systems — conventional QPSK Gray mapping with irregular LDPC or turbo code or

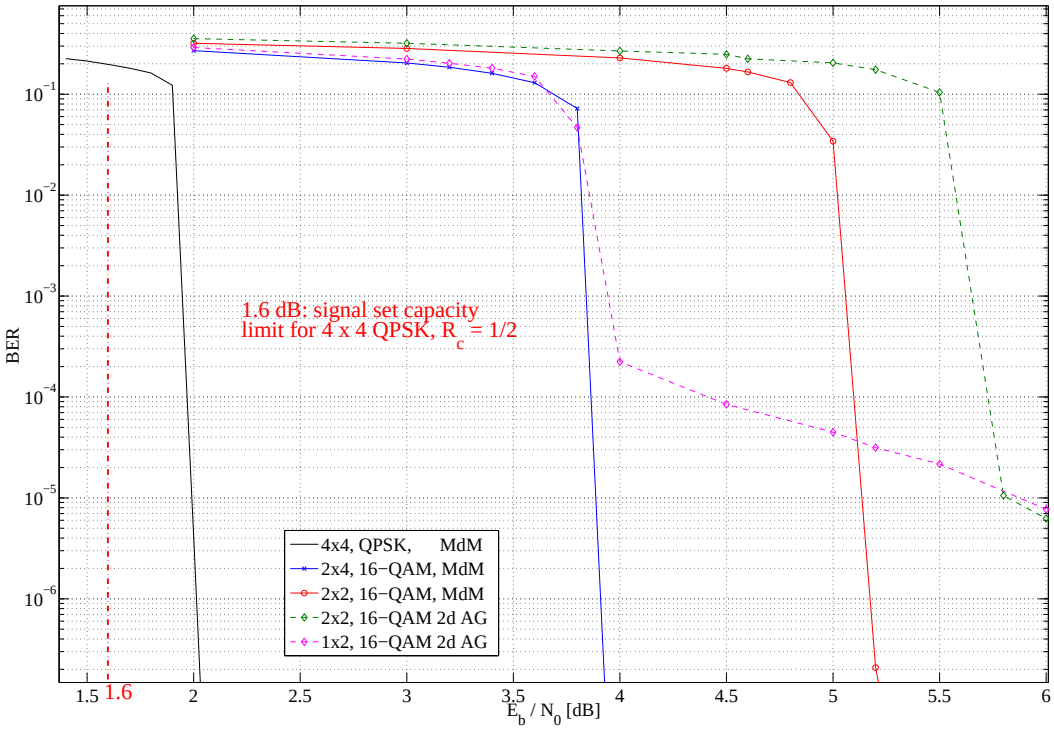


Figure 4.7: BER for MdM and symbolwise mappings for various antennas configurations and modulation orders, if $\eta = 4$, and 1×2 16-QAM 2d AG

our proposed MdM scheme with a memory two convolutional encoder, would approach the capacity limit very closely.

Another advantage of QPSK over 16-QAM MdM becomes visible in the EXIT chart of Figure 4.8. The transfer characteristic of the 4×4 QPSK MdM has an inflection point at about $I_{\mathrm{A1}} = 0.5$, which allows a good match to the transfer characteristic of the outer decoder, because this curve exhibits also an odd symmetry with respect to $(0.5, 0.5)$. This explains the early turbo-cliff position. As a drawback, the matching of transfer characteristics results in many iterations. A tunnel between both curves is already open at 2 dB, but 40 iterations are necessary for the trajectory to sneak through the narrow part of the tunnel (at medium extrinsic MI). An inflection point of $I_{\mathrm{E1}}(I_{\mathrm{A1}})$ could not be observed for any 16-QAM MdM, which were examined. Moreover, $I_{\mathrm{E1}}(1)$ is smaller in case of the 16-QAM MdM than that of the MdM with QPSK. For comparison, the transfer characteristic for symbolwise anti-Gray QPSK with $N_{\mathrm{t}} = N_{\mathrm{r}} = 4$ is also shown.

4.4.1 Summary for Multidimensional Mappings

In this chapter, we have presented a method to design MdMs by means of generator matrices. This allows a simple description of the MdMs and optimization with respect to average PEP, if perfect a priori knowledge is available at the demapper. For BPSK and QPSK as the underlying mappings, we proposed optimal generator matrices. The QPSK MdMs have an

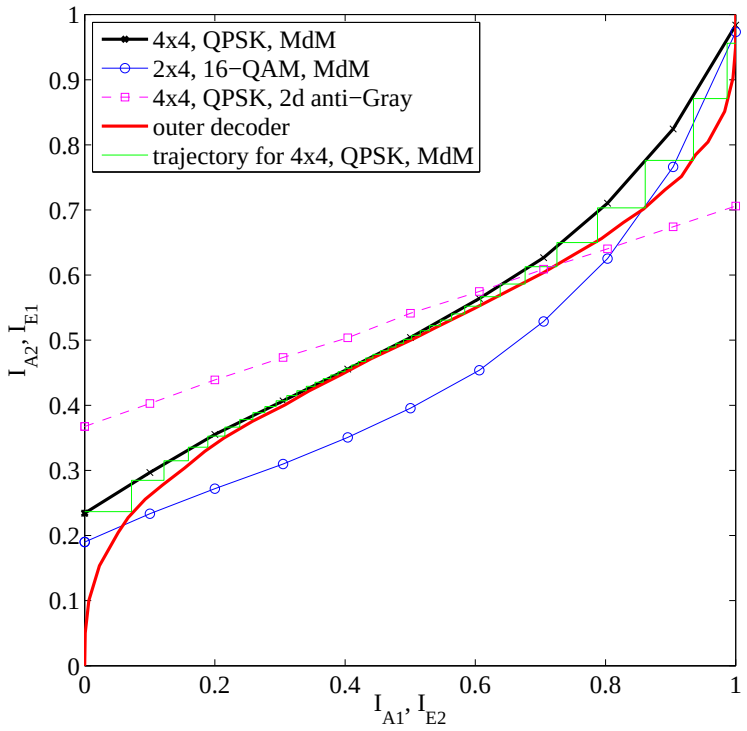


Figure 4.8: EXIT chart at $\frac{E_b}{N_0} = 2\,\text{dB}$ for MdMs and symbolwise mapping

earlier turbo-cliff than BPSK MdMs and need just half the number of transmit antennas to achieve the same PEP. Simulation results have shown that MdMs with the same PEP exhibit the same error-floor, which justifies our approach. The optimized 4×4 QPSK MdM allows transmission without measurable error-floor only 0.4 dB away from the signal set capacity limit. From a pragmatic point of view, one can explain the advantage of MdMs as follows: From our discussion in Subsection 2.3.4, it is known that the inner encoder in a SCC scheme has to be recursive in order to minimize the lower bound on decoding errors [83, 85]. A MIMO mapper operates blockwise on $N_t M$ bits, but for conventional symbolwise mapping, only M bits for each antenna can be used to introduce dependencies. This is done by an anti-Gray labeling, which we have shown to be a consequence of a block precoder followed by Gray mapping. A multidimensional mapper, on the other hand, may encode all $N_t M$ input bits, which can be exploited more efficiently at the demapper, if a priori information is fed back from an outer decoder. If we replace the block precoder by a recursive convolutional code, the error-floor can be reduced even further, see [136, 125]. The drawback of this scheme is the need for an additional BCJR decoder at the receiver side, which is rather complex. Also the MdM demapper has high complexity, as vectorwise APP decoding is required. However, reduced search sphere decoding algorithms exist also for soft APP decoding [137, 44].

Finally, it is worth mentioning that MdMs are not restricted to MIMO schemes. We can replace the spatial dimension by any other dimension, e.g., an MdM could be applied over the orthogonal subcarriers of an OFDM symbol. In [138], MdMs were examined for temporal

dimensions, where BPSK or QPSK symbols are blockwise precoded and then transmitted successively. However, as in [129], the authors optimized the system only with respect to the most likely PEP. In [139], which was published independently of [27] at the same conference, the design target was extended to minimize the average PEP. The proposed precoding is identical to our results. In the next chapter, we will investigate these temporal MdMs more closely.

Chapter 5

Inner Block Codes for Serial Concatenated Codes and Applications to BICM

5.1 Motivation

In the last chapter, we have pointed out that the application of MdMs is also possible for temporal precoding. Let us denote the number of consecutive bits $c_{k,m}$, which are influenced by an inner precoder, as M_c. For the previous spatial MdMs, we have $M_\mathrm{c} = N_\mathrm{t}M$. We restrict ourselves to rate 1 encoders, which only introduce dependencies among the bits, but no additional redundancy. After the inner code has precoded the M_c bits, they are transmitted by M_c BPSK symbols over the channel. The inner encoder may in general be non-linear. In Subsection 3.3.1 of [96], the inner encoder was denoted as a mapper, but should not be confused with the notation of a QAM mapper. There, the inner code was described by a look-up table. Also in [140], a look-up table was used to specify temporal MdMs. In this chapter, we will focus again on linear block codes, but extensions to non-linear codes are straightforward, and were outlined in [141]. While the design goal for the inner encoder in [96, 139, 140] is restricted to the maximization of $I_{\mathrm{E1}}(1)$ in order to reduce the error-floor of an iterative receiver, we will derive a closed-form expression of its transfer characteristic, if both communication and extrinsic[1] channel are modeled as a binary erasure channel (BEC). The results will be shown to be applicable also to the AWGN channel. Having available these analytical expressions for different inner codes, we can apply the mixing property of the EXIT chart and solve a curve-fitting problem to obtain a set of inner codes, which can be used as an irregular, i.e., time-varying inner code. The transfer characteristics of irregular codes were first examined in [142] and later adopted to irregular modulation, where the mapping is changing with time [143]. However, the approach was to use numerical EXIT

[1] The extrinsic channel is a virtual channel, introduced in [101] to model the quality of a priori information.

functions from Monte-Carlo simulations for the curve-fitting problem, whereas we introduce closed-form expressions.

In the next section, we present the system model for the serial concatenated code (SCC). Then, we derive the transfer characteristics for various short block codes and introduce them in the curve-fitting problem. This will yield optimized irregular inner codes, for which the presented simulation results will show that capacity approaching performance is possible. Finally, we will apply the inner precoding as a means of generating arbitrary anti-Gray mappings for high order modulation. The results of this chapter are published in [144].

5.2 System Model for SCC

5.2.1 Transmitter for SCC

Figure 5.1 depicts the transmitter model, which is similar to the BICM transmitter from Figure 2.14. The information bits $\mathbf{b}_\kappa$ from the binary source are encoded by an outer encoder of rate R_c, then randomly bit-interleaved, yielding the sequence $\mathbf{c}_k$. An inner encoder of rate $R_\mathrm{c} = 1$, which will be denoted as $\mathscr{C}$, computes from M_c consecutive bits, $c_{k,m}$, the same number of output[2] bits $u_{k,m}$, which form the output sequence $\mathbf{u}_k$. The symbol sequence $\mathbf{s}_k$ is obtained after mapping with labeling function μ and symbol alphabet $\mathbb{X}$. We distinguish between two different cases for this setup.

In case ①, the mapping applies BPSK from Figure 2.6(a), and thus we have $M_\mathrm{c} > M = 1$. This case will be used to examine the influence of the inner encoder, which combines M_c bits, while the BPSK mapper does not introduce further dependencies. Hence, the parallel input bits to the mapper in Figure 5.1 are serialized and transmitted with BPSK symbols in M_c consecutive time slots. The transmitted symbol vector can be written as $\mathbf{s}_k = 2\mathbf{u}_k - \mathbf{1}_{M_\mathrm{c}}$, where the entries are either -1 or $+1$.

Secondly, case ② will be used later for the analysis of BICM with anti-Gray mappings, which are obtained by the entity of inner encoder and Gray mapper. From our discussion in Subsection 4.3.2, it is clear that any anti-Gray mapping can be described by the concatenation of a (possibly non-linear) block code of rate 1, followed by a reference Gray mapping. In this case, $M = M_\mathrm{c}$ bits are mapped according to the Gray labeling function μ to one complex symbol $\mathbf{s}_k = s_k \in \mathbb{X}$, $|\mathbb{X}| = 2^M = L$. We restrict ourselves to short linear block codes $\mathscr{C}$, described by a square, non-singular generator matrix G, but extensions to non-linear codes or codes of rate smaller than 1 are straightforward, and independent of the code word length.

[2] Note that the notation in [144] for the input and output sequence of the inner code is vice versa.

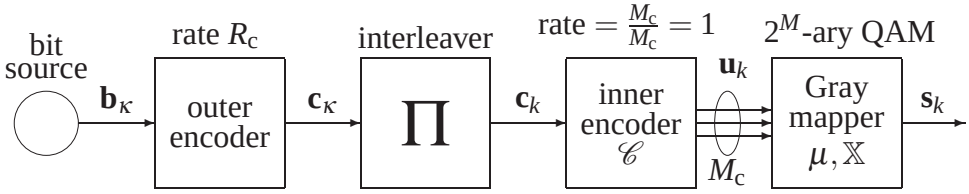


Figure 5.1: Transmitter for SCC

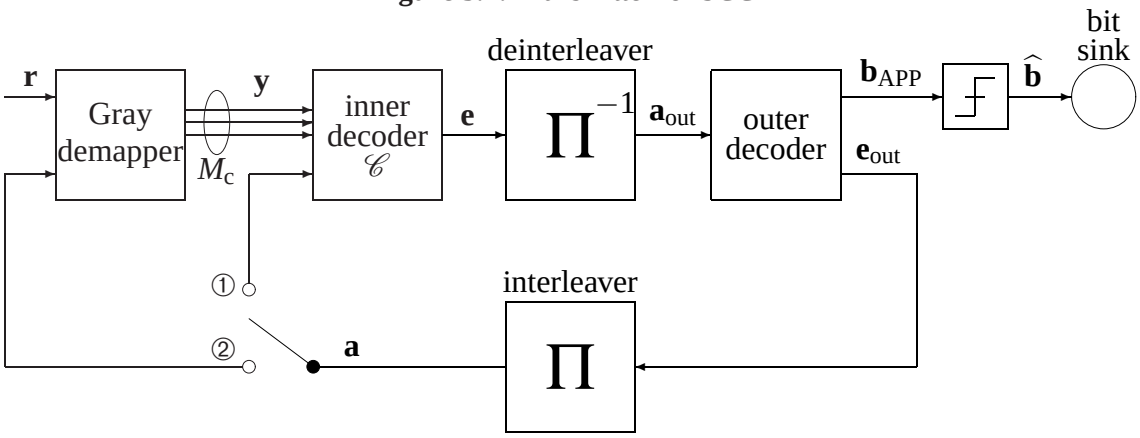


Figure 5.2: Iterative receiver for SCC

5.2.2 Receiver for SCC

The receiver in Figure 5.2 performs iterative demapping and decoding, and is a generalization of the BICM-ID receiver from Figure 2.15. To ease notation, we drop the discrete-time indices k and κ for the rest of this chapter. Thus, we focus on one inner code word only. Furthermore, we have introduced in this figure a different notation, as opposed to Figure 2.15. The reason is that this description shall also be suited for hard output processing, which is necessary for the BEC assumption.

The Gray demapper and both decoders extract extrinsic information for each bit from optimum APP processing. In case ①, M_c consecutive channel-corrupted BPSK symbols $\mathbf{r} = (r_0, \ldots, r_{M_c-1})$ are demapped one by one. All information from the received symbols is fully contained in the extrinsic vector $\mathbf{y} = (y_0, \ldots, y_{M_c-1})$, because APP decoding provides sufficient statistics [67]. The fed back extrinsic information $\mathbf{e}_{\text{out}}$ from the outer decoder is used as a priori information $\mathbf{a} = (a_0, \ldots, a_{M_c-1})$ at the inner decoder, so the switch in Figure 5.2 is in the upper position. For each bit c_m, $m \in \{0, \ldots, M_c - 1\}$, the extrinsic information e_m is computed based on the vector $\mathbf{y}$ and the a priori information of all other bits $\mathbf{a}_{[m]}$.

For case ②, the switch is in the lower position, since a priori information can already be used at the Gray demapper. This is because $C_S > C_B$ for $M > 2$, as we have seen in Chapter 3. Therefore, the transfer characteristic of a Gray demapper has a non-zero gradient. From one demapped symbol $\mathbf{r} = r$, the M_c extrinsic values of $\mathbf{y}$ are fed to the inner decoder, which computes e_m. Note that the same result is obtained, if both blocks are considered as one entity, i.e., if the corresponding anti-Gray demapping is performed in one step.

In both cases, extrinsic information is exchanged iteratively, until after a certain number of iterations, a hard decision based on APP values of the information bits $\mathbf{b}_{\mathrm{APP}}$ yields the bit estimates $\widehat{\mathbf{b}}$.

5.2.3 Channel Models and Capacity Constraints

In the next section, we will derive the transfer characteristics of the inner decoder for case ①, assuming that both communication and extrinsic channel are modeled as BECs (see Subsection 2.2.4) with erasure probability q and p, respectively. Thus, each bit y_m equals u_m with probability $1-q$ and each a_m equals c_m with probability $1-p$, while in all other cases, an erasure ε is declared. Recall from Section 2.6 that the typical assumption for EXIT chart measurements is that the a priori information is Gaussian distributed. In such a case, the extrinsic channel is an AWGN channel with zero mean and a variance, which is related via the J-function to the MI I_{A1}. For the BEC assumption, the input MI of the inner decoder is

$$I_{\mathrm{A1}} = I(c_m, a_m) = 1-p, \tag{5.1}$$

which follows directly form (2.76), if the bits c_m are assumed to be i.u.d.. This assumption is justified for sufficiently long interleavers. The average output MI is in general

$$I_{\mathrm{E1}} = \frac{1}{M_{\mathrm{c}}} \sum_{m=0}^{M_{\mathrm{c}}-1} I(c_m, e_m). \tag{5.2}$$

From property 3 of the EXIT diagram (cf. Section 2.6) and with (2.76), we can integrate (5.2), and obtain the BEC capacity,

$$\int_0^1 I_{\mathrm{E1}}\, dI_{\mathrm{A1}} = C_{\mathrm{BEC}} = 1-q. \tag{5.3}$$

When comparing a BEC with an AWGN communication channel, we will equate C_{BEC} with the signal set capacity of BPSK signaling, C_{S}, and thereby obtain a corresponding SNR $\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2}$ for the simulations.

$$C_{\mathrm{BEC}} = C_{\mathrm{S}}\left(\mathbf{x} = (-1,+1), \frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2}\right). \tag{5.4}$$

5.3 Transfer Characteristic of Inner Block Code

In [101], EXIT functions based on the Helleseth-Kløve-Levenshtein information functions [145] are derived for the BEC assumption. We now present a different method to compute similar EXIT functions of block codes for case ①, which allows for detailed insight. We will

focus on precoders of dimension $M_\mathrm{c} = 3$. Let us first rewrite (5.2) as

$$\begin{aligned} I_{\mathrm{E1}} &\overset{(1)}{=} \frac{1}{M_\mathrm{c}} \sum_{m=0}^{M_\mathrm{c}-1} I\left(c_m, \mathbf{y}\mathbf{a}_{[m]}\right) \\ &\overset{(2)}{=} \frac{1}{M_\mathrm{c}} \sum_{m=0}^{M_\mathrm{c}-1} \left(1 - H\left(c_m \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right)\right). \end{aligned} \tag{5.5}$$

The first step follows from the data processing theorem [67], because e_m provides sufficient statistics about both $\mathbf{y}$ and $\mathbf{a}_{[m]}$, and vice versa. Step (2) comes from (2.68) and the fact that the bits c_m are i.u.d.. Let us now consider for the m-th bit the uncertainty $H\left(c_m \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right)$ for a BEC more closely. From (2.67), we have the expectation

$$H\left(c_m \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right) = -\sum_{\forall \mathbf{y}, \mathbf{a}_{[m]}} P\left(c_m, \mathbf{y}\mathbf{a}_{[m]}\right) \cdot \log_2 P\left(c_m \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right). \tag{5.6}$$

Note that the possible probabilities for a BEC are either $P = \frac{1}{2}$ in case, an erasure occurred, or otherwise, the result is deterministic, such that $P \in \{0, 1\}$. Now observe that all addends vanish, in which $P\left(c_m \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right) = 1$, as the logarithm thereof yields zero. Furthermore, all addends with $P\left(c_m \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right) = 0$ disappear, because the joint probability

$$P\left(c_m, \mathbf{y}\mathbf{a}_{[m]}\right) = P\left(\mathbf{y}\mathbf{a}_{[m]}\right) \cdot P\left(c_m \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right) = 0 \tag{5.7}$$

is then also zero and from basic arithmetics, we have

$$\lim_{x \to 0} x \cdot \log_2 x = 0. \tag{5.8}$$

What remains in (5.6) are all addends, for which $P\left(c_m \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right) = \frac{1}{2}$, which means that, given $\mathbf{y}$ and $\mathbf{a}_{[m]}$, the decoder can not decode the bit c_m and has to declare an erasure. Thus, the uncertainty reduces with $\log_2\left(\frac{1}{2}\right) = -1$ to

$$H\left(c_m \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right) = \sum_{\forall \mathbf{y}, \mathbf{a}_{[m]}} P\left(c_m \text{ in erasure} \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right) \cdot P\left(c_m, \mathbf{y}\mathbf{a}_{[m]}\right). \tag{5.9}$$

The term $P\left(c_m \text{ in erasure} \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right)$ acts as an indicator, which is 1 for all realizations of $\mathbf{y}$ and $\mathbf{a}_{[m]}$, which do not allow the detection of bit c_m, and 0 otherwise. Thus, the summation in (5.9) corresponds to an expectation of all events, where c_m can not be decoded, given $\mathbf{y}$ and $\mathbf{a}_{[m]}$, and we finally have for the MI of (5.5)

$$\begin{aligned} I_{\mathrm{E1}} &= \frac{1}{M_\mathrm{c}} \sum_{m=0}^{M_\mathrm{c}-1} \mathsf{E}_{\mathbf{y}, \mathbf{a}_{[m]}} \left[P\left(c_m \text{ not in erasure} \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right)\right] \\ &= \mathsf{E}_{m, \mathbf{y}, \mathbf{a}_{[m]}} \left[P\left(c_m \text{ not in erasure} \left| \mathbf{y}\mathbf{a}_{[m]}\right.\right)\right]. \end{aligned} \tag{5.10}$$

This means that I_{E1} is the average probability that any bit c_m is not in erasure, i.e., can be decoded correctly, given all available input information, namely the observed vector $\mathbf{y}$ and the a priori vector $\mathbf{a}_{[m]}$.

Example: For $M_{\mathrm{c}} = 3$, consider the following inner encoder

$$\mathbf{u} = \mathbf{c} \cdot G_2 \text{ with } G_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.11}$$

In contrast to the notation from the previous chapter, the index of the generator matrix G simply enumerates different precoders. The first coded bit is obtained from modulo-2 addition of the first two input bits, $u_0 = c_0 + c_1$, whereas the other two bits are left uncoded, i.e., $u_1 = c_1$ and $u_2 = c_2$. The dependencies, which are induced by G_2, are illustrated in Figure 5.3, which can be regarded as the Tanner graph representation of the code [55]. Recall that the observations $\mathbf{y}$ at the receiver correspond to the upper bits $\mathbf{u}$ in this graph, if no erasure (with probability $1-q$) occurred, while the lower bits $\mathbf{c}$ are represented by $\mathbf{a}_{[m]}$, which is available for all, but the m-th bit, and can be described by erasure probability p.

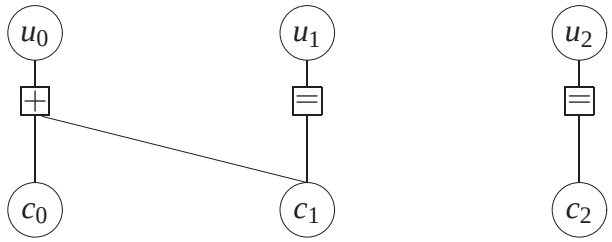


Figure 5.3: Bit dependencies induced by G_2 of equation (5.11)

Given $\mathbf{y}$ and $\mathbf{a}_{[0]}$ at the input of the inner decoder, the bit c_0 can be decoded correctly through the parity check, so u_0 (through observation y_0) *and* c_1 have to be available correctly. The bit c_1 in turn is decodable in all cases, unless both y_1 and a_1 are in erasure. Thus, we find the first addend in (5.10) to be

$$I\left(c_0, \mathbf{y}\mathbf{a}_{[0]}\right) = (1-q)(1-pq). \tag{5.12}$$

The second bit, c_1, can be decoded correctly, given $\mathbf{y}\mathbf{a}_{[1]}$, either directly through y_1 (with probability $1-q$) or — if y_1 is in erasure, indirectly through the parity check, which requires y_0 and a_0 to be correct. This happens with probability $(1-q)(1-p)$. So, together we get

$$I\left(c_1, \mathbf{y}\mathbf{a}_{[1]}\right) = (1-q) + q(1-q)(1-p). \tag{5.13}$$

Bit c_2 can only be recovered from u_2, which yields

$$I\left(c_2, \mathbf{y}\mathbf{a}_{[2]}\right) = (1-q). \tag{5.14}$$

Inserting (5.12) until (5.14) in (5.10), we obtain the EXIT function of G_2 as

$$I_{\mathrm{E1}} = \Phi\left((3+q)p^0 + (-2)p^1 + (0)p^2\right), \tag{5.15}$$

with the abbreviation $\Phi = \frac{1-q}{M_\mathrm{c}}$ and $p = 1 - I_{\mathrm{A1}}$, cf. (5.1). It can easily be verified that integration of I_{E1} yields $1-q$, which is in accordance to (5.3). ■

From this simple example, we can deduce for an arbitrary code dimension M_c the following properties of the EXIT function under the BEC assumption:

- Φ factorizes I_{E1}, because the code $\mathscr{C}$ has rate 1. This means that at least one observed bit u_l has to be correct to detect any c_m.
- I_{E1} is a polynomial in p (or I_{A1} respectively) of degree $d_p \leq M_\mathrm{c} - 1$.
- Each coefficient γ_i belonging to p^i is itself a polynomial in q of degree $d_q \leq M_\mathrm{c}$ with coefficients $\delta_j^{(i)}$.

Thus, the general form for an arbitrary rate 1 encoder can be written as

$$I_{\mathrm{E1}} = \Phi \sum_{i=0}^{M_\mathrm{c}-1} p^i \cdot \gamma_i = \Phi \sum_{i=0}^{M_\mathrm{c}-1} p^i \cdot \left(\sum_{j=0}^{M_\mathrm{c}-1} q^j \cdot \delta_j^{(i)} \right). \tag{5.16}$$

Table 5.1 summarizes the results for all non-singular generator matrices with $M_\mathrm{c} = 3$ that yield different EXIT functions. For $M_\mathrm{c} = 4$, there exist 52 different matrices, for $M_\mathrm{c} = 5$ already 965, and for $M_\mathrm{c} = 6$, more than 5000 matrices were found, which result in distinct EXIT functions [141]. The matrices in the table are written in octal notation, i.e., each column is given a number, where the upper component of the column vector is the least significant bit (LSB). For example, the identity matrix is written as $G_1 = (1,2,4)$, and is enumerated in the table by the index 1.

Table 5.1: Coefficients $\delta_j^{(i)}$ for (5.16) for matrices of dimension $M_\mathrm{c} = 3$

G (octal)	$\delta_0^{(0)}$	$\delta_1^{(0)}$	$\delta_2^{(0)}$	$\delta_0^{(1)}$	$\delta_1^{(1)}$	$\delta_2^{(1)}$	$\delta_0^{(2)}$	$\delta_1^{(2)}$	$\delta_2^{(2)}$	d_G
$G_1 = (1,2,4)$	3	0	0	0	0	0	0	0	0	3
$G_2 = (3,2,4)$	3	1	0	0	−2	0	0	0	0	4
$G_3 = (3,2,5)$	3	2	0	0	−2	−2	0	−3	3	5
$G_4 = (7,2,4)$	3	2	0	0	−4	−2	0	0	3	5
$G_5 = (7,6,4)$	3	2	1	0	−4	−4	0	0	3	6
$G_6 = (7,6,5)$	3	3	1	0	−4	−6	0	−3	6	7

Let us define the Hamming weight of a matrix as d_G, with

$$d_G = \mathrm{d_H}\left(\mathrm{vec}\,(G)^{\mathrm{T}}, \mathbf{0}\right), \tag{5.17}$$

where the vec$(.)$-operator creates a column vector from a matrix by stacking all column vectors onto each other, and $\mathbf{0}$ is the all-zero row vector. In the rightmost column of Table 5.1, the corresponding weights of the generator matrices are also included. Observe that matrix G_6 corresponds to an optimum precoding matrix for a Rayleigh fading channel, if perfect a priori knowledge is available at the inner decoder, see Theorem 4.1, i.e., it includes the maximum number of 1 elements. As we will see from the simulation results, this precoding also lowers the error-floor over the non-fading AWGN channel.

The corresponding EXIT functions for a channel erasure probability of $q = 0.5$ can be seen in Figure 5.4. Simulated EXIT functions for the BEC model are identical with analytical results from (5.16) and are thus not shown here. While I_{E1} is a constant for G_1 and a linear function for G_2, it is a polynomial in I_{A1} of degree 2 for the remaining matrices. From (5.3), we obtain the channel capacity $C_{\mathrm{BEC}} = 1 - q = 0.5$ and from equation (5.4), the corresponding SNR for the AWGN channel is $\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2} = -2.82\,\mathrm{dB}$. From this, we obtain the well-known capacity limit of a rate $R_{\mathrm{c}} = \frac{1}{2}$ scheme with BPSK signaling in terms of $\frac{E_{\mathrm{b}}}{N_0}$, which is according to (2.31) $\frac{E_{\mathrm{b}}}{N_0} = \frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2} - 10 \cdot \log_{10}(0.5) = 0.19\,\mathrm{dB}$. In Figure 5.4, we also depict the simulated EXIT functions over the AWGN channel with BPSK signaling at this SNR, and applied the Gaussian assumption for the extrinsic channel, cf. Section 2.6. It can be seen that corresponding BEC and AWGN curves have the same area of 0.5, because of constraint (5.4). The BEC curves approximate closely the AWGN curves, but each BEC curve always exhibits a slightly steeper gradient. In [101], it was observed that a BEC curve upper bounds an AWGN curve, if both curves have the same starting point $I_{\mathrm{E1}}(0)$, which corresponds to a normalization with respect to the BICM capacity C_{B}. This backs up our observation about the gradients.

We can further observe from this figure that the rightmost point of the EXIT function, i.e., $I_{\mathrm{E1}}(1)$ for perfect a priori information, increases with increasing Hamming weight of the generator matrix. To be precise, this point can be computed from (5.16) by inserting $p = 1 - I_{\mathrm{A1}} = 0$,

$$I_{\mathrm{E1}} = \Phi \sum_{j=0}^{M_{\mathrm{c}}-1} q^j \cdot \delta_j^{(0)}. \tag{5.18}$$

From Table 5.1, we find that the coefficients $\delta_j^{(0)}$ are identical for the codes G_3 and G_4, such that their value for $I_{\mathrm{E1}}(1)$ is equal. However, for the AWGN channel, we see from Figure 5.4 that this ending point is slightly smaller for G_3 than for G_4. As to the starting point $I_{\mathrm{E1}}(0)$, we can observe, that G_1 achieves the highest MI, as it introduces no dependencies, $I_{\mathrm{E1}} \equiv 0.5$, followed by G_2, G_4, G_5, G_3, and G_6. This order holds for both the BEC and the AWGN channel.

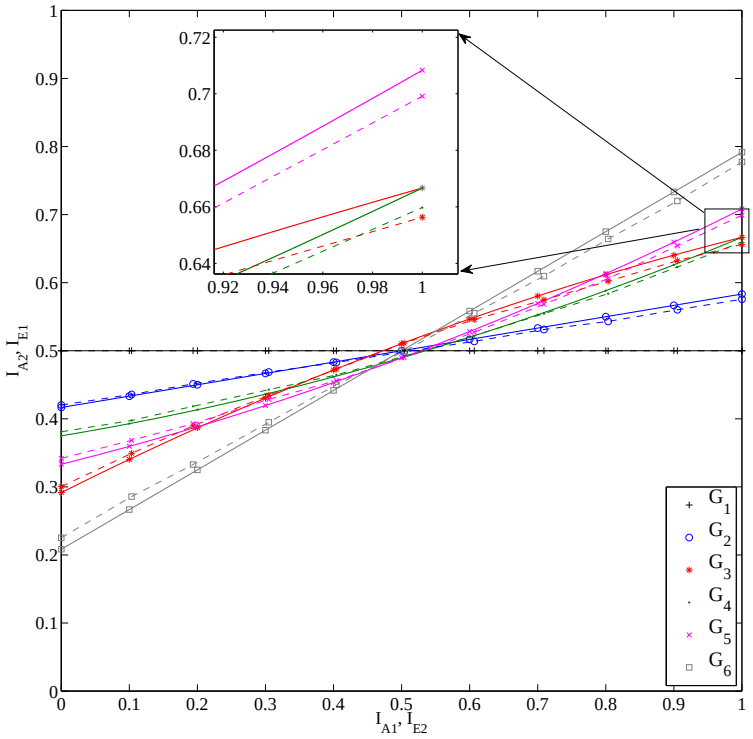


Figure 5.4: EXIT chart for inner codes of length $M_c = 3$ over BEC with $q = 0.5$ (—) and over AWGN with BPSK at $\frac{E_b}{N_0} = 0.19\,\text{dB}$ (- - -)

5.4 Optimization of Irregular Inner Codes

5.4.1 Formulation of Design Target

The inner encoder $\mathscr{C}$ in Figure 5.1 shall now be irregular, i.e., time-varying. As in [143], we allow N different codes, linear and of rate 1. Each generator matrix G_n, $n \in \{1,\ldots,N\}$, is used for a fraction α_n of the total length of the outer code word $\mathbf{c}$. If the EXIT function of code G_n is denoted by $I_{\text{E1},n}$, the EXIT function of the irregular code is expressed by the mixing property (see Section 2.6) as

$$I_{\text{E1,IR}} = \sum_{n=1}^{N} \alpha_n \cdot I_{\text{E1},n}. \tag{5.19}$$

Finding capacity approaching code components reduces to a curve-fitting problem, i.e., $I_{\text{E1,IR}}$ shall match f_{out}, which is the inverse EXIT function of the outer encoder, $f_{\text{out}} = I_{\text{A2}}(I_{\text{E2}})$. In contrast to [143], we introduce the closed-form expressions for this problem. As we ultimately want to examine the behavior over an AWGN channel, we use (5.16) as a close approximation for the AWGN transfer characteristic. This is justified by the close match between analytical formulas for the BEC assumption and the simulated AWGN curves, cf. Figure 5.4. The exact probability for decoding erasures of a convolutional decoder under the BEC assumption was derived in [146] and could be introduced in (5.10) to obtain an

analytical expression for f_{out} [141, 28]. This results again in a steeper gradient of the transfer characteristic compared to the simulated curve for the AWGN channel. Thus, we rather apply a cubic interpolation on Monte Carlo measurements, which were obtained over the AWGN channel, to approximate f_{out} in closed-form. With $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$, the design target is to minimize the integrated squared error Ξ between inner and outer EXIT function,

$$\arg\min_{\boldsymbol{\alpha}} \Xi = \arg\min_{\boldsymbol{\alpha}} \int_0^1 \left(I_{\text{E1,IR}} - f_{\text{out}}\right)^2 dI_{\text{A1}} = \arg\min_{\boldsymbol{\alpha}} \; \boldsymbol{\alpha} Q \boldsymbol{\alpha}^{\text{T}} - 2\boldsymbol{\alpha}\mathbf{v}^{\text{T}} + w,$$

$$\text{subject to: } \sum_{n=1}^{N} \alpha_n = 1 \quad \text{and} \quad 0 \leq \alpha_n \leq 1\,, \forall n \in \{1, \ldots, N\}, \tag{5.20}$$

where the matrix Q, which is used for the quadratic form, is symmetric and consists of the components

$$[Q]_{i,j} = Q_{i,j} = Q_{j,i} = \int_0^1 I_{\text{E1},i} \cdot I_{\text{E1},j}\, dI_{\text{A1}} = \int_0^1 I_{\text{E1},i} \cdot I_{\text{E1},j}\, dp \geq 0. \tag{5.21}$$

The right hand side follows from the substitution $I_{\text{A1}} = 1 - p$.

Example: The element $Q_{1,2} = Q_{2,1}$ is obtained with the coefficients from Table 5.1 as

$$Q_{1,2} = \int_0^1 \{3\Phi\} \cdot \left\{ \Phi\left[\left(3+q\right) + p\left(-2q\right) \right] \right\} dp = 9\Phi^2 \tag{5.22}$$

and is obviously the same for all $Q_{1,j}$, $j \in \{1, \ldots, N\}$. ■

In (5.20), we have further introduced the row vector $\mathbf{v}$, whose i-th entry is

$$v_i = \int_0^1 I_{\text{E1},i} \cdot f_{\text{out}}\, dI_{\text{A1}} \tag{5.23}$$

and the positive constant w, which is given by

$$w = \int_0^1 f_{\text{out}}^2\, dI_{\text{A1}}. \tag{5.24}$$

In the minimization of Ξ, the constant w can be neglected. It is easily seen from (5.21) that the matrix Q is positive semidefinite. Therefore, the function $\Xi(\boldsymbol{\alpha})$ is $\cup$-convex and equation (5.20) can be solved by constrained quadratic programming methods [147, 148]. Note that in contrast to the optimization problem in (3.8), the function Ξ has no additional local optima, but only the global solution. Finally, note that the optimization of irregular mappings in [143, 141] differs in so far from (5.20), as it relaxes the curve-fitting condition for I_{A1} close to 1.

5.4.2 Simulation Results for Irregular Inner Codes

We have solved (5.20) for $2 \leq M_{\text{c}} \leq 6$ and for all outer convolutional encoders of memory 2 and rate $R_{\text{c}} = \frac{1}{2}$. The best matches between inner and outer transfer characteristic with

respect to Ξ were always found for the $(07,05)$ RSC code. From the channel coding theorem (2.71), we set the target capacity $C_{\mathrm{BEC}} = 1-q \geq R_{\mathrm{c}}$, where for each q, a different optimum vector $\boldsymbol{\alpha}$ might arise. For lack of space, we summarize here only results for $M_{\mathrm{c}} = 3$ and $M_{\mathrm{c}} = 6$, while other examples can be found in [141].

For $M_{\mathrm{c}} = 3$, we have considered all $N = 6$ different matrices from Table 5.1. First, we set the design parameter $q = 0.476$, which corresponds to $\frac{E_{\mathrm{b}}}{N_0} = 0.5\,\mathrm{dB}$ on the AWGN channel[3]. For this case, the optimum ratios are

$$\boldsymbol{\alpha} = (0, 0, 0.25, 0.17, 0.1, 0.48) \tag{5.25}$$

and the resulting irregular EXIT function will be denoted as $I_{\mathrm{E1,IR},3,0.5}$. The first index therein equals M_{c} and the second corresponds to the design target $\frac{E_{\mathrm{b}}}{N_0}$. From (5.25), we can see that the matrix G_6 is applied most of the time, due to its steep gradient, while the first two matrices are not used at all. We have found that this optimum is quite robust against changes in q.

For $M_{\mathrm{c}} = 6$, we have chosen only $N = 10$ different matrices. In octal notation, they are summarized in Table 5.2, together with their Hamming weights. The matrix G_{10} is designed according to Theorem 4.1 and includes the maximum amount of 1 elements for $M_{\mathrm{c}} = 6$. In general, this number is ${M_{\mathrm{c}}}^2 - M_{\mathrm{c}} + 1$.

Table 5.2: $N = 10$ different matrices of dimension $M_{\mathrm{c}} = 6$ to be used in (5.20)

G (octal)	d_G	G (octal)	d_G
$G_1 = (01,02,04,10,20,40)$	6	$G_6 = (77,76,74,70,60,40)$	21
$G_2 = (07,06,04,10,20,40)$	9	$G_7 = (77,76,75,73,60,40)$	24
$G_3 = (17,16,14,10,20,40)$	12	$G_8 = (37,36,35,33,27,57)$	26
$G_4 = (77,42,44,50,60,40)$	15	$G_9 = (77,76,75,73,67,51)$	29
$G_5 = (77,76,34,30,20,40)$	18	$G_{10} = (77,76,75,73,67,57)$	31

The optimum ratios for target $\frac{E_{\mathrm{b}}}{N_0} = 0.5\,\mathrm{dB}$ are

$$\boldsymbol{\alpha} = (0.07, 0.15, 0.07, 0.29, 0.07, 0, 0, 0.25, 0.06, 0.04) \tag{5.26}$$

and the resulting irregular EXIT function will be denoted as $I_{\mathrm{E1,IR},6,0.5}$.

Figure 5.5 depicts the optimum solutions in the EXIT chart. Recall that f_{out} was measured by simulations over the AWGN channel, whereas the irregular coding curves are computed analytically based on BEC model, and a slightly smaller gradient thereof must be kept in mind for the AWGN channel. For $q = 0.476$, both solutions for $M_{\mathrm{c}} = 3$ and 6 yield exactly the same error Ξ, and both allow an open tunnel to f_{out}, but for $M_{\mathrm{c}} = 6$, the intersection

[3]The original turbo-code from [14] achieved its turbo-cliff at 0.7 dB.

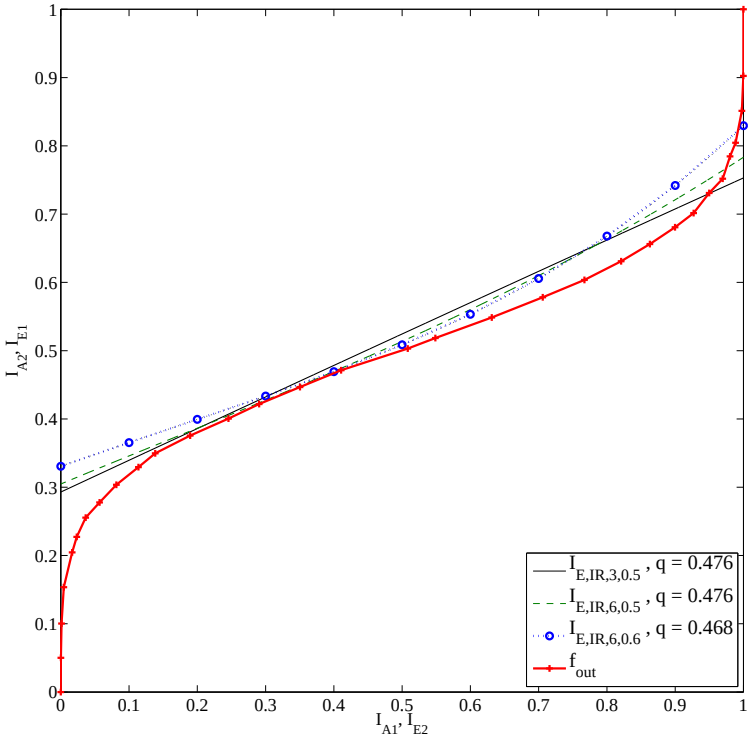


Figure 5.5: EXIT chart of optimum irregular inner codes of lengths $M_c = 3$ and $M_c = 6$, and $(07, 05)$ RSC code

with f_{out} occurs slightly later. Therefore, we expect a lower error-floor for this solution. However, its error-floor will still be rather high, as this intersection does not come close to the upper right corner of the EXIT chart. Note that a large value of $I_{\text{E1}}(1)$ for error-free a priori knowledge is not an explicit condition for the curve-fitting problem in (5.20). A reduction of this error-floor is possible by more frequent use of the matrix G_{10} for $M_c = 6$, since it contains the largest number of 1 entries. Hence, we force α_{10} to a fixed large ratio and solve (5.20) for the remaining $N-1$ ratios. For a slightly larger target SNR of $\frac{E_b}{N_0} = 0.6\,\text{dB}$ ($q = 0.468$), the largest possible value of α_{10}, which still allows an open tunnel of $I_{\text{E1,IR}}$ to f_{out}, is $\alpha_{10} = 0.6$. For this case, the optimum ratios are then

$$\boldsymbol{\alpha} = (0.24, 0.06, 0.01, 0.09, 0, 0, 0, 0, 0, 0.6), \tag{5.27}$$

yielding $I_{\text{E1,IR,6,0.6}}$, which is also shown in Figure 5.5. This transfer characteristic matches very closely with the outer decoder curve *and* intersects late with this curve.

To verify the optimized results over the AWGN channel, we plot in Figure 5.6 the simulated BER over the AWGN channel after 40 iterations for the discussed solutions, which are denoted as $I_{\text{E1,IR},M_c,x}$, where $x = \frac{E_b}{N_0}$ is the design target SNR. In addition, the BER is shown for an irregular code, which uses 10% of G_1 and 90% of G_{10}, and is the best solution for $\frac{E_b}{N_0} = 0.9\,\text{dB}$. This code has the EXIT function $I_{\text{E1,IR,6,0.9}}$ (not shown). Note that all turbo-cliffs match well with predicted target values. It can further be observed that for a given complexity, imposed by M_c, the price for an early turbo-cliff — $I_{\text{E1,IR,3,0.5}}$ has its cliff

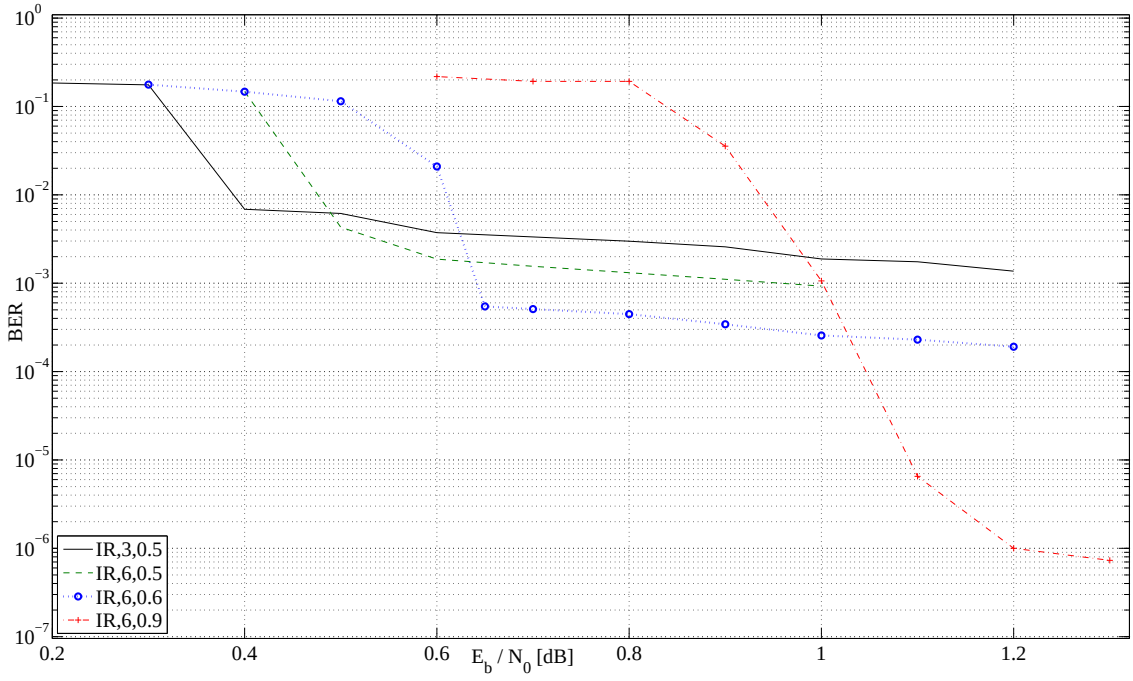


Figure 5.6: BER over AWGN channel for different irregular codes, 40 iterations

only 0.21 dB away from the capacity limit, is a rather high error-floor. This tradeoff already appeared in Section 3.6 and can always be observed for a given decoding complexity.

5.5 Applications to BICM

We now apply the previous results to a BICM scheme. We will consider high order modulation with an arbitrary mapping as the output of the inner rate 1 encoder followed by a Gray mapper as indicated in Figure 5.1. At the BICM-ID receiver, the switch in Figure 5.2 is in position ②. Up to now, the task of the inner encoder was to introduce dependencies in M_c bits, which were then transmitted over M_c time slots. If we want to adopt the closed-form expressions, which we have obtained for this scenario under the BEC assumption, to the BICM-ID scheme with high order modulation ($M = M_c$ bits per QAM symbol) over an AWGN channel, we can not simply start with the capacity constraint (5.4) to find the appropriate channel erasure probability q. In what follows, we will show, how each of the M bits contributes to the total signal set capacity C_S. Therefore, we have to introduce M different BECs for each bit and introduce the corresponding erasure probabilities in the M_c addends of equation (5.10). We begin this section with a review and thorough examination of both BICM and signal set capacities.

5.5.1 Bit Level Capacities

A non-iterative BICM receiver executes only one pass through the demapper, so the switch in Figure 5.2 is in position ①. According to (2.78), the maximum MI for equiprobable signaling is determined by the BICM capacity

$$C_{\mathrm{B}} = \sum_{m=0}^{M-1} \left\{ \int_{r\in\mathbb{C}} \sum_{b=0}^{1} \sum_{x_l\in\mathbb{X}_b^m} \frac{1}{L} \cdot p\left(r\,|x_l\right) \log_2 \frac{\sum\limits_{x_l\in\mathbb{X}_b^m} p\left(r\,|x_l\right)}{\frac{1}{L}\cdot \sum\limits_{x_l'\in\mathbb{X}} p\left(r\,\middle|x_l'\right)} dr \right\} = \sum_{m=0}^{M-1} C_{\mathrm{B}}^{(m)}, \tag{5.28}$$

where we have used the short hand notation from (3.30) and implicitly defined the bit level BICM capacities $C_{\mathrm{B}}^{(m)}$. These individual capacities are the underlying principle for designing a capacity approaching MLC [10]. They were also examined in [95] in order to construct Gray labelings for a regular constellation, which maximize C_{B}.

In a similar manner, we can now define the bit level signal set capacity $C_{\mathrm{S}}^{(m)}$ for each bit. Recall from (2.77) that for an arbitrary signal set $\mathbb{X}$, the overall signal set capacity is the average MI between the received symbols r and the equiprobable transmit symbols x_l or the corresponding bits of $\mathbf{c}$,

$$\begin{aligned} C_{\mathrm{S}} &= I(x_l,r) = I(\mathbf{c},r) = \sum_{a=0}^{M-1} I_a = \sum_{a=0}^{M-1} \overline{I\left(c_m, r\,\middle|a \text{ other bits are perfectly known}\right)} \\ &= \sum_{m=0}^{M-1} \left(\sum_{a=0}^{M-1} \sum_{j=1}^{\binom{M-1}{a}} \sum_{\forall \mathbf{c}_j^{(a)}} \frac{I(c_m, r\mid \mathbf{c}_j^{(a)})}{M\cdot 2^a \cdot \binom{M-1}{a}} \right) = \sum_{m=0}^{M-1} C_{\mathrm{S}}^{(m)}. \end{aligned} \tag{5.29}$$

The first summation over a comes from the chain rule of MI, where I_a is the average MI, given perfect knowledge of a bits in $\mathbf{c}$ [67, 100]. These known bits are comprised in the vector $\mathbf{c}_j^{(a)}$ of length a and allow 2^a different permutations. The index j enumerates the realizations, which particular a out of $M-1$ bits are known. In the next summation, the bar indicates averaging. The last line follows after carrying out the averaging in I_a and reordering the summations. Finally, the right hand side defines implicitly the bit level signal set capacities $C_{\mathrm{S}}^{(m)}$.

Let us now examine the bit level EXIT functions. The EXIT function I_{E1} of a demapper and its bit level components $I_{\mathrm{E1}}^{(m)}$ are

$$I_{\mathrm{E1}} = \frac{1}{M}\sum_{m=0}^{M-1} I\left(c_m, r\,\mathbf{a}_{[m]}\right) = \frac{1}{M}\sum_{m=0}^{M-1} I_{\mathrm{E1}}^{(m)}, \tag{5.30}$$

where the a priori feedback switch in Figure 5.2 is now in position ②. The previous equation corresponds to the definition of the average MI (5.5). Examples for bit level EXIT functions can be found in Subsection 3.3.2 of [28].

The relationship between (5.28) until (5.30) is similar to the properties from Section 2.6. Remember that the starting point of the demapper's EXIT function $I_{\mathrm{E1}}(0)$ equals the BICM capacity C_{B}, whereas the area under I_{E1} corresponds to C_{S}. The same is true for bit level components, i.e., $I_{\mathrm{E1}}^{(m)}(0) = C_{\mathrm{B}}^{(m)}$ and $\int_0^1 I_{\mathrm{E1}}^{(m)}\, dI_{\mathrm{A1}} = C_{\mathrm{S}}^{(m)}$.

If we want to apply the analytical results from (5.16) to arbitrary anti-Gray mappings, consisting of inner encoder and Gray mapper, we have to consider the bit dependencies, induced even by the Gray mapping. As discussed in Chapter 3, any mapping of M bits to a two-dimensional signal set always introduces dependencies, if $M > 2$.

As an example, consider the 8-PSK Gray mapping from Figure 2.7(a). No precoder is applied yet, i.e., G is the identity matrix of dimension $M_{\mathrm{c}} = M$.

The bit level signal set capacities of the bits c_0 and c_1 are obviously identical due to symmetry. From (5.29), we get after some manipulations

$$C_{\mathrm{S}}^{(0)} = C_{\mathrm{S}}^{(1)} = \frac{1}{2}\left(I_{\mathrm{E1}}^{(0)}(0) + I_{\mathrm{E1}}^{(0)}(1)\right), \tag{5.31}$$

where $I_{\mathrm{E1}}^{(0)}(0)$ is the bit level BICM capacity of c_0 and $I_{\mathrm{E1}}^{(0)}(1)$ is the mean value of two BPSK capacities [26]. The first BPSK corresponds to the case, when the last two bits, which are perfectly known, are either 00 or 10. The corresponding signal points are located at $\pm d_1$. The other BPSK occurs, if the known bits are either 01 or 11, and has its signal points at $\pm d_2$. Hence, the distances are $d_1 = \frac{1}{2}\left|\mu(0,0,0) - \mu(1,0,0)\right| = \sin\left(\frac{\pi}{8}\right)$ and $d_2 = \frac{1}{2}\left|\mu(0,0,1) - \mu(1,0,1)\right| = \sin\left(\frac{3\pi}{8}\right)$.

For the last bit, c_2, we compute

$$C_{\mathrm{S}}^{(2)} = \frac{1}{3}\left(I_{\mathrm{E1}}^{(2)}(0) + I_{\mathrm{E1}}^{(2)}(1) + I\left(c_2, r \mid c_0\right)\right), \tag{5.32}$$

where each term can be interpreted in a similar way: $I_{\mathrm{E1}}^{(2)}(0)$ is the bit level BICM capacity of c_2, $I_{\mathrm{E1}}^{(2)}(1)$ is the capacity of a BPSK with symbols at $\pm d_1$, and the last term corresponds to the signal set capacity of a 4-QAM, where the symbols are given by the subset $\mathbb{X}_{c_0}^0$.

The bit level capacities as a function of the SNR $\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2}$ can be seen in Figure 5.7. Clearly, $C_{\mathrm{S}}^{(m)} \geq C_{\mathrm{B}}^{(m)}, \forall m$. Equality holds for the two extremes, when the SNR is either 0 or approaches infinity, for which the bit level capacities are both 0 or 1, respectively. We have observed increasing differences for larger M (not shown). Thus, some amount of MI between the transmitted bits and the received values is lost, if no iterations are performed over the demapper, even for Gray labeling. This fact is often neglected, if high order modulation with Gray labeling is combined with a strong off-the-shelf encoder, e.g., [128]. In such cases, only the BICM capacity can be achieved, but a small non-zero gap to the signal set capacity remains. From Figure 5.7, we can further notice that the bit level capacities differ among the M bits per symbol, i.e., every BICM scheme with $M > 2$ inevitably performs unequal error protection. A preferred mapping thus should possess on the one hand a diminishing gap

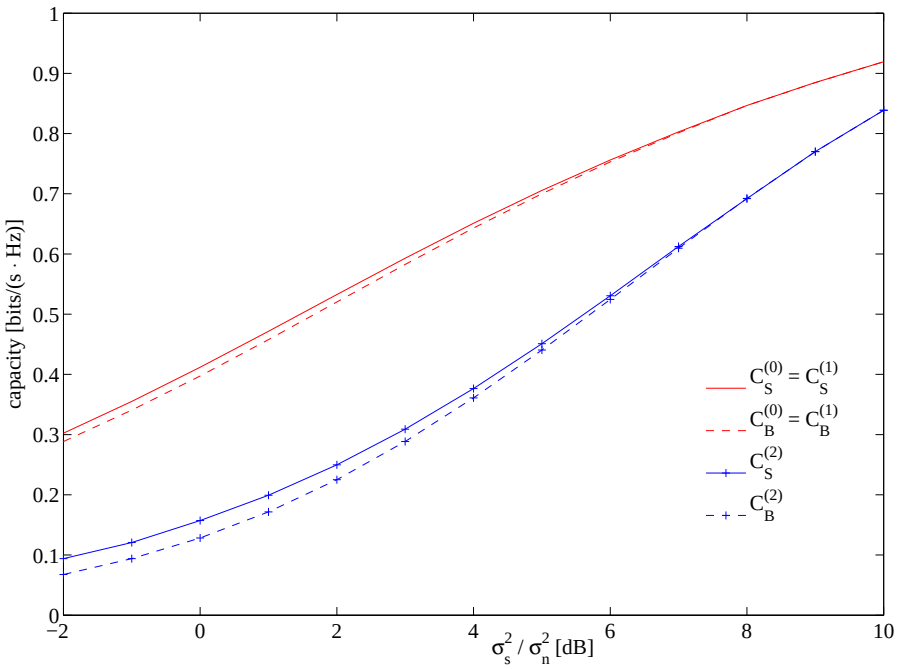


Figure 5.7: Bit level capacities for 8-PSK Gray mapping

between the signal set and the BICM capacity, and on the other hand have similar bit level capacities for all M bits. We denote this as an information-maintaining mapping (IMM). The corresponding EXIT function of its demapper should be maximally flat and its bit level EXIT functions $I_{\mathrm{E1}}^{(m)} \approx C_{\mathrm{B}}^{(m)} \approx C_{\mathrm{S}}^{(m)}, \forall m$. Typical signal set constellations however do not fulfill these conditions. The search for IMMs is still an open field of research.

5.5.2 Transfer Characteristic of Demapper

We consider the anti-Gray mapping that results from precoding with G_2 from (5.11) and 8-PSK Gray mapping. We adopt the results from Section 5.3 to obtain the transfer characteristic of the demapper. Under the assumption — which we have just shown to be not exactly true, that no MI is lost, the yet uncoded bit c_0 has the same bit level signal set capacity as the precoder output bit u_0, namely $C_{\mathrm{S}}^{(0)}$. Setting this capacity to $1-q_0$ and replacing q in (5.12) and (5.13) by q_0, we obtain BEC approximations of $I_{\mathrm{E1}}^{(0)}$ and $I_{\mathrm{E1}}^{(1)}$, respectively. For $I_{\mathrm{E1}}^{(2)}$, we replace q in (5.14) by q_2, where $q_2 = 1 - C_{\mathrm{S}}^{(2)}$. The closed-form EXIT function under the BEC assumption can be seen in Figure 5.8 together with the simulated AWGN curve, which is obtained at $\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2} = 3.5\,\mathrm{dB}$. Now the AWGN curve has a steeper gradient due to the inherent dependencies, which are already induced by the Gray mapping, and which are disregarded by our assumption that the MI is preserved. Recall from (5.11) that the BEC curve is a polynomial in I_{A1} of degree 2, whereas the AWGN curve is not a rational function. No closed-form expression for the AWGN channel has been proposed in the literature yet. However, we expect a better match between the transfer characteristics for BEC and AWGN channel, if an IMM is applied.

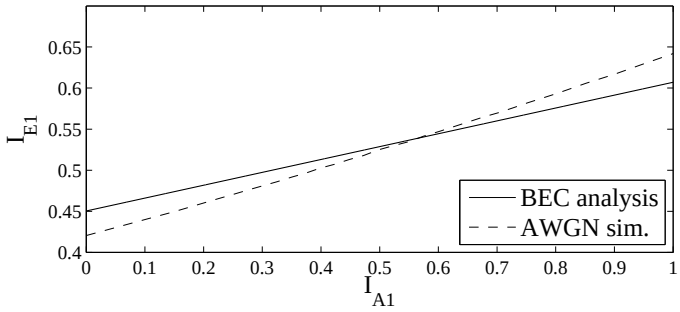


Figure 5.8: Section of EXIT chart of 8-PSK anti-Gray mapping: BEC analysis (—) and AWGN simulation (- - -)

5.6 Summary for Inner Block Codes for Serial Concatenated Codes and Applications to BICM

We have computed EXIT functions analytically for linear block codes of rate 1 under the BEC assumption and demonstrated their polynomial form. Based on these results, irregular inner code components for SCC were optimized for the AWGN channel, using our analytical expressions as close approximates. The resulting schemes approach the capacity limit by only $0.21\,\mathrm{dB}$ and have rather small decoding complexity, but as a drawback, the error-floor lies above 10^{-3}. Lower error-floors are possible at the expense of later turbo-cliff positions or with larger dimensions M_c of the inner block code, which increases the decoding complexity. In addition, we have investigated bit level capacities and bit level EXIT functions in order to apply the closed-form EXIT functions to the transfer characteristic of a BICM demapper. Detailed analysis of these bit level capacities revealed that capacity approaching BICM has to take leave of typical mappings, such as PSK or the regular QAM constellations with symbols located on an orthogonal grid, since those lose information after (optimum) demapping for $M > 2$. One approach, to retain a larger amount of MI after demapping, would be the application of those mappings, which maximize BICM capacity C_B, cf. Section 3.3. Yet, their transfer characteristic is not designed to be flat, even though its starting value $I_{\mathrm{E1}}(0)$ is maximized. This means that a non-iterative receiver still loses some MI. Thus, another approach would be the application of an IMM. We outlined as a design goal for these mappings that their EXIT function should be maximally flat.

Chapter 6

Conclusion

This thesis investigated both coding and modulation for spectral-efficient transmission. The BICM scheme plays a prominent role in practical systems, where bandwidth is limited, as well in a myriad of research topics on communications theory. However, most attempts to optimize BICM systems aim only at the bit labeling of a fixed QAM constellation. Opposed to that, we have optimized QAM mappings with respect to both bit labeling *and* symbol constellation for various cost functions. These include symbol and bit error rate that occur after the demapper, and both BICM and signal set capacities. We also examined mappings that exploit a priori knowledge, which arises in iterative BICM receivers. This target was extended to spatial multiplexing MIMO systems, where we analyzed and optimized multidimensional mappings. Simulation results have shown that all proposed schemes outperform state-of-the-art systems of the same spectral efficiencies. Finally, we have derived closed-form expressions for the EXIT functions of block codes, if the communication and extrinsic channel are appropriately modeled as binary erasure channels. The results matched well with the simulated EXIT functions of the AWGN channel. With relatively small encoding and decoding complexity, we designed irregular inner codes to be used in a serial concatenated coding system. Applied on the AWGN channel, these schemes approached the capacity limit by only 0.21 dB.

Future research could be tied in with the following topics: First, we have neglected the practically important issue of peak-to-average power ratio (PAPR). The optimized QAM mappings from Chapter 3 were only constrained to a fixed average transmit power, but not to a maximum peak power. For the multidimensional MIMO mappings, we have assumed an ideally uncorrelated fading channel. The influence of temporal and, moreover, spatial correlation has to be considered and possible application of prefilters as in [149] should be considered. The search for IMMs with flat transfer characteristics is another topic worth examining. Finally, our setup for serial concatenated codes in Section 5.2 allows further insight in the analogy of turbo and LDPC codes. Note that the concatenated encoder in Figure 5.1 can be described by a *single* generator matrix: the outer convolutional code can be reduced to a block code, cf. (2.52), the interleaver yields a square permutation matrix of dimension N_{int}, and the irregular inner encoder has an overall generator matrix, with M_{c}-dimensional

submatrices G_i, see (5.11), in its diagonal, while the rest is 0. The concatenation of the three components results in a matrix multiplication in GF(2) of the individual matrices. Thus, knowledge from both theories — turbo and LDPC, could be combined. For example, the receiver might apply the message-passing algorithm on the overall parity check matrix instead of iteratively exchanging extrinsic information from demapper to decoder [150]. Further, the interleaver could be designed in such a way that it eliminates short girths in the overall parity check matrix [55, 50]. This would lower the error-floor with no additional complexity.

Appendix A

Computation of Gradients

For a given cost function $f(\mathbf{x})$, we derive the gradient $\nabla f(\mathbf{x}_i)$ with respect to $\mathbf{x}$. Note that all cost functions will be real valued scalars. Then, the gradient is

$$\nabla f(\mathbf{x}_i) = \left(\frac{\partial f(\mathbf{x})}{\partial x_0}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_l}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_{L-1}} \right), \tag{A.1}$$

where the l-component is as given in (3.10)

$$\frac{\partial f(\mathbf{x})}{\partial x_l} = \frac{\partial f(\mathbf{x})}{\partial x_{l,\mathrm{R}}} + \mathrm{j} \frac{\partial f(\mathbf{x})}{\partial x_{l,\mathrm{I}}}. \tag{A.2}$$

This means that $f(\mathbf{x})$ has to be differentiated with respect to both $x_{l,\mathrm{R}}$ and $x_{l,\mathrm{I}}$.

A.1 Gradient of Symbol Error Probability

Let us extract the l-th component from the first summation in (3.16), such that

$$P_\mathrm{s} = \frac{1}{2L} \left[\sum_{\substack{l_1=0 \\ l_1 \neq l}}^{L-1} \sum_{\substack{l_2=0 \\ l_2 \neq l_1}}^{L-1} \operatorname{erfc}\left(\frac{\left| x_{l_2} - x_{l_1} \right|}{2\sqrt{2}\sigma_\mathrm{n}} \right) + \sum_{\substack{l_2=0 \\ l_2 \neq l_1 = l}}^{L-1} \operatorname{erfc}\left(\frac{\left| x_{l_2} - x_l \right|}{2\sqrt{2}\sigma_\mathrm{n}} \right) \right]. \tag{A.3}$$

If we differentiate P_s with respect to $x_{l,\mathrm{R}}$, all addends from the first (double) sum are zero, except those in which $l_2 = l$, while from the second (single) sum, all addends contribute to the derivative, as l_1 is already set to l. Thus,

$$\frac{\partial P_\mathrm{s}(\mathbf{x})}{\partial x_{l,\mathrm{R}}} = \frac{1}{2L} \left[\sum_{\substack{l_1=0 \\ l_1 \neq l_2 = l}}^{L-1} \frac{\partial}{x_{l,\mathrm{R}}} \operatorname{erfc}\left(\frac{\left| x_l - x_{l_1} \right|}{2\sqrt{2}\sigma_\mathrm{n}} \right) + \sum_{\substack{l_2=0 \\ l_2 \neq l_1 = l}}^{L-1} \frac{\partial}{\partial x_{l,\mathrm{R}}} \operatorname{erfc}\left(\frac{\left| x_{l_2} - x_l \right|}{2\sqrt{2}\sigma_\mathrm{n}} \right) \right]. \tag{A.4}$$

By changing the first summation index from l_1 to l_2 and observing, that the absolute value in the erfc-function is even, we can combine both sums. If we further insert the integral form of the erfc-function from (3.14) in (A.4), we obtain

$$\frac{\partial P_{\mathrm{s}}(\mathbf{x})}{\partial x_{l,\mathrm{R}}} = \frac{2}{L}\sum_{\substack{l_2=0\\ l_2\neq l}}^{L-1} \frac{\partial}{\partial x_{l,\mathrm{R}}} \int\limits_{|x_{l_2}-x_l|/2}^{\infty} \frac{\exp\left(-\frac{\xi^2}{\sigma_{\mathrm{n}}^2}\right)}{\sqrt{\pi}\sigma_{\mathrm{n}}}\, d\xi \tag{A.5}$$

The fundamental theorem of calculus [99] states that

$$\frac{\partial}{\partial x} \int\limits_{f_{\mathrm{l}}(x)}^{f_{\mathrm{u}}(x)} g(\xi)\, d\xi = \frac{\partial f_{\mathrm{u}}(x)}{\partial x} \cdot g\left(f_{\mathrm{u}}(x)\right) - \frac{\partial f_{\mathrm{l}}(x)}{\partial x} \cdot g\left(f_{\mathrm{l}}(x)\right). \tag{A.6}$$

In (A.5), the variable $x_{l,\mathrm{R}}$ is only included in the lower integration border, so

$$\begin{aligned} \frac{\partial P_{\mathrm{s}}(\mathbf{x})}{\partial x_{l,\mathrm{R}}} &= -\frac{2}{L}\sum_{\substack{l_2=0\\ l_2\neq l}}^{L-1} \frac{\partial}{\partial x_{l,\mathrm{R}}} \frac{\sqrt{\left(x_{l_2,\mathrm{R}}-x_{l,\mathrm{R}}\right)^2+\left(x_{l_2,\mathrm{I}}-x_{l,\mathrm{I}}\right)^2}}{2} \cdot \frac{\exp\left(-\frac{\left|x_{l_2}-x_l\right|^2}{4\sigma_{\mathrm{n}}^2}\right)}{\sqrt{\pi}\sigma_{\mathrm{n}}} \\ &= \frac{1}{L}\sum_{\substack{l_2=0\\ l_2\neq l}}^{L-1} \frac{x_{l_2,\mathrm{R}}-x_{l,\mathrm{R}}}{\left|x_{l_2}-x_l\right|} \cdot \frac{\exp\left(-\frac{\left|x_{l_2}-x_l\right|^2}{4\sigma_{\mathrm{n}}^2}\right)}{\sqrt{\pi}\sigma_{\mathrm{n}}} \end{aligned} \tag{A.7}$$

From the last step, it is easy to see that the partial derivative with respect to $x_{l,\mathrm{I}}$ is obtained by changing the index "R" to "I" in (A.7). Combining real and imaginary part as in (A.2) and changing index l_2 to l_1, the final result is

$$\frac{\partial P_{\mathrm{s}}(\mathbf{x})}{\partial x_l} = \frac{1}{L}\sum_{\substack{l_1=0\\ l_1\neq l}}^{L-1} \frac{x_{l_1}-x_l}{\left|x_{l_1}-x_l\right|} \cdot \frac{\exp\left(-\frac{\left|x_{l_1}-x_l\right|^2}{4\sigma_{\mathrm{n}}^2}\right)}{\sqrt{\pi}\sigma_{\mathrm{n}}}. \tag{A.8}$$

A.2 Gradient of BICM Capacity

Let us first consider the partial derivative of C_{B} with respect to the real part of the l-th component in $\mathbf{x}$, which is $x_{l,\mathrm{R}}$. For that, we define the m-th bit of the bit label from x_l, i.e.,

$\mu_m^{-1}(x_l)$ as $b_{l,m}$ and its inversion as $\overline{b_{l,m}}$. We split in (3.30) the sum over $b \in \{0,1\}$ to obtain

$$\begin{aligned} C_{\mathrm{B}} &= M - \frac{1}{L} \int\limits_{r \in \mathbb{C}} \sum_{m=0}^{M-1} \left\{ \sum_{x_{l_1} \in \mathbb{X}_{b_{l,m}}^m} p\left(r \left| x_{l_1}\right.\right) \log_2 \frac{L \cdot p(r)}{\sum\limits_{x_{l_1} \in \mathbb{X}_{b_{l,m}}^m} p\left(r \left| x_{l_1}\right.\right)} \right. \\ &\quad \left. + \sum_{x_{l_1} \in \mathbb{X}_{\overline{b_{l,m}}}^m} p\left(r \left| x_{l_1}\right.\right) \log_2 \frac{L \cdot p(r)}{\sum\limits_{x_{l_1} \in \mathbb{X}_{\overline{b_{l,m}}}^m} p\left(r \left| x_{l_1}\right.\right)} \right\} dr. \end{aligned} \tag{A.9}$$

Per definition, symbol $x_l \in \mathbb{X}_{b_{l,m}}^m$, $\forall m \in \{0,\ldots,M-1\}$. Hence

$$\begin{aligned} \frac{\partial}{\partial x_{l,\mathrm{R}}} \sum_{x_{l_1} \in \mathbb{X}_{b_{l,m}}^m} p\left(r \left| x_{l_1}\right.\right) &= \frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r \left| x_l\right.\right) \overset{(2.36)}{=} \frac{\partial}{\partial x_{l,\mathrm{R}}} \frac{\exp\left(-\frac{\left(r_{\mathrm{R}} - x_{l,\mathrm{R}}\right)^2 + \left(r_{\mathrm{I}} - x_{l,\mathrm{I}}\right)^2}{\sigma_{\mathrm{n}}^2}\right)}{\pi \sigma_{\mathrm{n}}^2} \\ &= 2 \cdot p\left(r \left| x_l\right.\right) \cdot \frac{r_{\mathrm{R}} - x_{l,\mathrm{R}}}{\sigma_{\mathrm{n}}^2}, \end{aligned} \tag{A.10}$$

in which $r = r_{\mathrm{R}} + \mathrm{j} r_{\mathrm{I}}$. On the other hand, $x_l \notin \mathbb{X}_{\overline{b_{l,m}}}^m$, $\forall m \in \{0,\ldots,M-1\}$, so

$$\frac{\partial}{\partial x_{l,\mathrm{R}}} \sum_{x_{l_1} \in \mathbb{X}_{\overline{b_{l,m}}}^m} p\left(r \left| x_{l_1}\right.\right) = 0. \tag{A.11}$$

The final prerequisite follows from $\mathbb{X} = \mathbb{X}_{b_{l,m}}^m \cup \mathbb{X}_{\overline{b_{l,m}}}^m$, $\forall m \in \{0,\ldots,M-1\}$, so that

$$\frac{\partial}{\partial x_{l,\mathrm{R}}} L \cdot p(r) = \frac{\partial}{\partial x_{l,\mathrm{R}}} \sum_{x_{l_1} \in \mathbb{X}} p\left(r \left| x_{l_1}\right.\right) = \frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r \left| x_l\right.\right). \tag{A.12}$$

Considering the product and chain rule of derivatives, and the fact that $\frac{\partial}{\partial x} \log_2 x = \left(\ln(2) \cdot x\right)^{-1}$, we write the partial derivative of C_{B} as

$$\begin{aligned} \frac{\partial C_{\mathrm{B}}(\mathbf{x})}{\partial x_{l,\mathrm{R}}} &= -\frac{1}{L} \int\limits_{r \in \mathbb{C}} \sum_{m=0}^{M-1} \left\{ \left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r \left| x_l\right.\right)\right] \log_2 \frac{L \cdot p(r)}{\sum\limits_{x_{l_1} \in \mathbb{X}_{b_{l,m}}^m} p\left(r \left| x_{l_1}\right.\right)} \right. \\ &\quad + \sum_{x_{l_1} \in \mathbb{X}_{b_{l,m}}^m} p\left(r \left| x_{l_1}\right.\right) \left\{ \frac{\left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r \left| x_l\right.\right)\right]}{\ln(2) \cdot L \cdot p(r)} - \frac{\left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r \left| x_l\right.\right)\right]}{\ln(2) \cdot \sum\limits_{x_{l_1} \in \mathbb{X}_{b_{l,m}}^m} p\left(r \left| x_{l_1}\right.\right)} \right\} \\ &\quad \left. + 0 + \sum_{x_{l_1} \in \mathbb{X}_{\overline{b_{l,m}}}^m} p\left(r \left| x_{l_1}\right.\right) \left\{ \frac{\left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r \left| x_l\right.\right)\right]}{\ln(2) \cdot L \cdot p(r)} - 0 \right\} \right\} dr. \end{aligned} \tag{A.13}$$

From that, we factor out $\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r\left|x_l\right.\right)$, combine the second with the last addend, and cancel out the sum in the third addend. This reduces the expression to

$$\begin{aligned}\frac{\partial C_{\mathrm{B}}(\mathbf{x})}{\partial x_{l,\mathrm{R}}} &= -\frac{1}{L}\int\limits_{r\in\mathbb{C}}\left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r\left|x_l\right.\right)\right]\sum_{m=0}^{M-1}\Bigg\{\log_2\frac{L\cdot p(r)}{\sum\limits_{x_{l_1}\in\mathbb{X}_{b_{l,m}}^{m}} p\left(r\left|x_{l_1}\right.\right)} \\ &\quad + \frac{\sum\limits_{x_{l_1}\in\mathbb{X}_{b_{l,m}}^{m}} p\left(r\left|x_{l_1}\right.\right) + \sum\limits_{x_{l_1}\in\mathbb{X}_{\overline{b_{l,m}}}^{m}} p\left(r\left|x_{l_1}\right.\right)}{\ln(2)\cdot L\cdot p(r)} - \frac{1}{\ln(2)}\Bigg\}\,dr. \end{aligned} \tag{A.14}$$

Now note, that the two sums in the numerator add up to $L\cdot p(r)$, such that the last addends cancel each other out. Inserting (A.10) yields

$$\begin{aligned}\frac{\partial C_{\mathrm{B}}(\mathbf{x})}{\partial x_{l,\mathrm{R}}} &= \frac{2}{L}\int\limits_{r\in\mathbb{C}} p\left(r\left|x_l\right.\right)\cdot\frac{x_{l,\mathrm{R}}-r_{\mathrm{R}}}{\sigma_{\mathrm{n}}^2}\Bigg\{M\cdot\log_2\left(L\cdot p(r)\right) \\ &\quad -\sum_{m=0}^{M-1}\log_2\left(\sum_{x_{l_1}\in\mathbb{X}_{b_{l,m}}^{m}} p\left(r\left|x_{l_1}\right.\right)\right)\Bigg\}\,dr. \end{aligned} \tag{A.15}$$

If we compute the partial derivative with respect to $x_{l,\mathrm{I}}$, we only have to change the index "R" to "I" in (A.12). Combining both partial derivatives as in (A.2), we end up with

$$\begin{aligned}\frac{\partial C_{\mathrm{B}}(\mathbf{x})}{\partial x_{l}} &= \frac{2}{L}\int\limits_{r\in\mathbb{C}} p\left(r\left|x_l\right.\right)\cdot\frac{x_{l}-r}{\sigma_{\mathrm{n}}^2}\Bigg\{M\cdot\log_2\left(L\cdot p(r)\right) \\ &\quad -\sum_{m=0}^{M-1}\log_2\left(\sum_{x_{l_1}\in\mathbb{X}_{b_{l,m}}^{m}} p\left(r\left|x_{l_1}\right.\right)\right)\Bigg\}\,dr. \end{aligned} \tag{A.16}$$

A.3 Gradient of Signal Set Capacity

A.3.1 General Expression for Gradient of Signal Set Capacity

First, we split the sum in (3.37) to obtain

$$C_{\mathrm{S}} = \frac{1}{L}\int\limits_{r\in\mathbb{C}}\left\{\sum_{\substack{l_1=0\\ l_1\neq l}}^{L-1} p\left(r\left|x_{l_1}\right.\right)\log_2\frac{p\left(r\left|x_{l_1}\right.\right)}{p(r)} + p\left(r\left|x_l\right.\right)\log_2\frac{p\left(r\left|x_l\right.\right)}{p(r)}\right\}dr. \tag{A.17}$$

Similar to (A.13), the partial derivative of C_S with respect to $x_{l,\mathrm{R}}$ can be determined with the product and chain rule of derivatives and with (A.12) as

$$\frac{\partial C_\mathrm{S}(\mathbf{x})}{\partial x_{l,\mathrm{R}}} = \frac{1}{L}\int\limits_{r\in\mathbb{C}} \left\{ \sum_{\substack{l_1=0 \\ l_1\neq l}}^{L-1} p\left(r\,|x_{l_1}\right) \left\{ 0 - \frac{\left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r\,|x_l\right)\right]}{\ln(2)\cdot L\cdot p(r)} \right\} + \left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r\,|x_l\right)\right] \cdot \log_2 \frac{p\left(r\,|x_l\right)}{p(r)} \right.$$
$$\left. + p\left(r\,|x_l\right) \cdot \left\{ \frac{\left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r\,|x_l\right)\right]}{\ln(2)\cdot p\left(r\,|x_l\right)} - \frac{\left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r\,|x_l\right)\right]}{\ln(2)\cdot L\cdot p(r)} \right\} \right\} dr. \tag{A.18}$$

Observe, that the last addend is the excluded component from the first sum, such that both parts can be combined to one sum. Extracting the common factor $\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r\,|x_l\right)$ and canceling $p\left(r\,|x_l\right)$ from the last but one addend, we get

$$\begin{aligned} \frac{\partial C_\mathrm{S}(\mathbf{x})}{\partial x_{l,\mathrm{R}}} &= \frac{1}{L}\int\limits_{r\in\mathbb{C}} \left[\frac{\partial}{\partial x_{l,\mathrm{R}}} p\left(r\,|x_l\right)\right] \left\{ \sum_{l_1=0}^{L-1} p\left(r\,|x_{l_1}\right) \frac{-1}{\ln(2)\cdot L\cdot p(r)} \right. \\ &\quad \left. + \log_2 \frac{p\left(r\,|x_l\right)}{p(r)} + \frac{1}{\ln(2)} \right\} dr. \end{aligned} \tag{A.19}$$

As the first sum (without the fraction) equals $L\cdot p(r)$, it cancels out the last addend. With (A.10) , the final result becomes

$$\frac{\partial C_\mathrm{S}(\mathbf{x})}{\partial x_{l,\mathrm{R}}} = \frac{2}{L}\int\limits_{r\in\mathbb{C}} p\left(r\,|x_l\right)\cdot \frac{r_\mathrm{R} - x_{l,\mathrm{R}}}{\sigma_\mathrm{n}^2} \log_2 \frac{p\left(r\,|x_l\right)}{p(r)}\, dr, \tag{A.20}$$

while the combined partial derivative according to (A.2) is

$$\frac{\partial C_\mathrm{S}(\mathbf{x})}{\partial x_l} = \frac{2}{L}\int\limits_{r\in\mathbb{C}} p\left(r\,|x_l\right)\cdot \frac{r - x_l}{\sigma_\mathrm{n}^2} \log_2 \frac{p\left(r\,|x_l\right)}{p(r)}\, dr. \tag{A.21}$$

A.3.2 Optimality of BPSK

We derive a formal proof, that BPSK maximizes signal set capacity for equiprobable symbols $P[x_0] = P[x_1] = \frac{1}{2}$. As in Subsection 3.1.1, we assume without loss of generality that

$$x_1 \in \left[0, \sqrt{2}\right] \quad \text{and} \quad x_0 = -\sqrt{2 - x_1^2} \in \left[-\sqrt{2}, 0\right], \tag{A.22}$$

where the condition on x_0 comes from power constraint (3.4). Thus, the signal set capacity (A.17) is a function of one real-valued variable x_1, as is its derivative. From (A.17), we get

$$C_\mathrm{S}(x_1) = \frac{1}{2}\int\limits_{r\in\mathbb{R}} \left\{ p\left(r\,|x_0\right) \log_2 \frac{p\left(r\,|x_0\right)}{p(r)} + p\left(r\,|x_1\right) \log_2 \frac{p\left(r\,|x_1\right)}{p(r)} \right\} dr, \tag{A.23}$$

with

$$p(r) = \frac{1}{2}\Big(p\left(r\,|x_0\right) + p\left(r\,|x_1\right)\Big). \tag{A.24}$$

It is sufficient to integrate over the set of real numbers, as each conditional pdf in (A.23) is a univariate Gaussian distribution

$$p\left(r\,|x_l\right) = \frac{\exp\Big(-\frac{\left(r-x_l\right)^2}{\sigma_\mathrm{n}^2}\Big)}{\sqrt{\pi}\sigma_\mathrm{n}},\ l \in \{0,1\}. \tag{A.25}$$

In order to differentiate (A.23) with respect to x_1, we need the following prerequisites:

$$\frac{\partial p\left(r\,|x_1\right)}{\partial x_1} = 2 \cdot p\left(r\,|x_1\right) \cdot \frac{r-x_1}{\sigma_\mathrm{n}^2}, \tag{A.26}$$

where we have inserted (A.25) and proceeded similar to (A.10). For the next equation, we apply the chain rule of derivatives and obtain

$$\frac{\partial p\left(r\,|x_0\right)}{\partial x_1} = \frac{\partial p\left(r\,|x_0\right)}{\partial x_0} \cdot \frac{\partial x_0}{\partial x_1} \overset{\text{(A.22)}}{=} 2 \cdot p\left(r\,|x_0\right) \cdot \frac{r-x_0}{\sigma_\mathrm{n}^2} \cdot \frac{x_1}{\sqrt{2-x_1^2}}. \tag{A.27}$$

Combining the last two equations with (A.24), we find

$$\frac{\partial p(r)}{\partial x_1} = p\left(r\,|x_1\right) \cdot \frac{r-x_1}{\sigma_\mathrm{n}^2} + p\left(r\,|x_0\right) \cdot \frac{r-x_0}{\sigma_\mathrm{n}^2} \cdot \frac{x_1}{\sqrt{2-x_1^2}}. \tag{A.28}$$

With (A.26) - (A.28) and the product rule of derivatives, we compute

$$\begin{aligned}
\frac{\partial C_\mathrm{S}\left(x_1\right)}{\partial x_1} = \frac{1}{2} \int\limits_{r\in\mathbb{R}} \Bigg\{ & 2 \cdot p\left(r\,|x_0\right) \cdot \frac{r-x_0}{\sigma_\mathrm{n}^2} \cdot \frac{x_1}{\sqrt{2-x_1^2}} \cdot \log_2 \frac{p\left(r\,|x_0\right)}{p(r)} \\
& + p\left(r\,|x_0\right) \cdot \Bigg\{ \frac{2 \cdot p\left(r\,|x_0\right) \cdot (r-x_0) \cdot x_1}{\ln(2) \cdot p\left(r\,|x_0\right) \cdot \sqrt{2-x_1^2} \cdot \sigma_\mathrm{n}^2} \\
& \qquad - \frac{p\left(r\,|x_1\right) \cdot (r-x_1) + p\left(r\,|x_0\right) \cdot (r-x_0) \cdot \frac{x_1}{\sqrt{2-x_1^2}}}{\ln(2) \cdot p(r) \cdot \sigma_\mathrm{n}^2} \Bigg\} \\
& + 2 \cdot p\left(r\,|x_1\right) \cdot \frac{r-x_1}{\sigma_\mathrm{n}^2} \cdot \log_2 \frac{p\left(r\,|x_1\right)}{p(r)} \\
& + p\left(r\,|x_1\right) \cdot \Bigg\{ \frac{2 \cdot p\left(r\,|x_1\right) \cdot (r-x_1)}{\ln(2) \cdot p\left(r\,|x_1\right) \cdot \sigma_\mathrm{n}^2} \\
& \qquad - \frac{p\left(r\,|x_1\right) \cdot (r-x_1) + p\left(r\,|x_0\right) \cdot (r-x_0) \cdot \frac{x_1}{\sqrt{2-x_1^2}}}{\ln(2) \cdot p(r) \cdot \sigma_\mathrm{n}^2} \Bigg\} \Bigg\} dr.
\end{aligned} \tag{A.29}$$

Notice that the third and the last addend both have the fraction in common, which we can factor out. What remains is $2 \cdot p(r)$ times this fraction, cf. (A.24). After canceling out

common factors, we observe that this fraction adds up to zero together with the second and the last but one addend. What is left, is

$$\frac{\partial C_{\mathrm{S}}(x_1)}{\partial x_1} = \frac{1}{\sigma_{\mathrm{n}}^2} \int_{r\in\mathbb{R}} \left\{ p\left(r\,|x_0\right) \cdot (r-x_0) \cdot \frac{x_1}{\sqrt{2-x_1^2}} \cdot \log_2 \frac{p\left(r\,|x_0\right)}{p(r)} + p\left(r\,|x_1\right) \cdot (r-x_1) \cdot \log_2 \frac{p\left(r\,|x_1\right)}{p(r)} \right\} dr. \tag{A.30}$$

Now, we insert (A.25) for the conditional pdfs in the logarithms. Some simple manipulations finally yield

$$\begin{aligned} \frac{\partial C_{\mathrm{S}}(x_1)}{\partial x_1} = \frac{1}{\sigma_{\mathrm{n}}^2} \int_{r\in\mathbb{R}} \Bigg\{ & p\left(r\,|x_0\right) \cdot (r-x_0) \\ & \cdot \left\{ 1 - \log_2 \left(1 + \exp\left(-\frac{x_1-x_0}{\sigma_{\mathrm{n}}^2} \cdot (x_0+x_1-2r) \right) \right) \right\} \\ & + p\left(r\,|x_1\right) \cdot (r-x_1) \\ & \cdot \left\{ 1 - \log_2 \left(1 + \exp\left(+\frac{x_1-x_0}{\sigma_{\mathrm{n}}^2} \cdot (x_0+x_1-2r) \right) \right) \right\} \Bigg\} dr. \end{aligned} \tag{A.31}$$

Setting this derivative to zero yields the extremum of C_{S}. One trivial solution comes from $x_1 = x_0 = \pm 1$. This is however a minimum, because $C_{\mathrm{S}} = 0$. Moreover, it violates the condition (A.22). The integral in (A.31) becomes zero, if, e.g., its integrand is an odd function in r. This is possible for $x_1 = -x_0 = 1$, as we will show now. Denote the integrand for this mapping as $\imath(r)$ and consider

$$\begin{aligned} -\imath(-r) = & -p\left(-r\,|-1\right) \cdot (-r+1) \cdot \left\{ 1 - \log_2 \left(1 + \exp\left(-\frac{2}{\sigma_{\mathrm{n}}^2} \cdot 2r \right) \right) \right\} \\ & - p\left(-r\,|+1\right) \cdot (-r-1) \cdot \left\{ 1 - \log_2 \left(1 + \exp\left(\frac{2}{\sigma_{\mathrm{n}}^2} \cdot 2r \right) \right) \right\}. \end{aligned} \tag{A.32}$$

From (A.25), it is clear that $p\left(r\,|x_l\right) = p\left(-r\,|-x_l\right)$. Thus $-\imath(-r) = \imath(r)$, and integration of an odd function over an even interval yields zero.

A.4 Gradient of $I_{\mathrm{E1}}(1)$

Recall from (3.53), that the MI for perfect a priori knowledge $I_{\mathrm{E1}}(1)$ is

$$I_{\mathrm{E1}}(1) = 1 - \int_{\xi\in\mathbb{R}} p\left(\xi|1\right) \log_2\left(1 + \frac{p\left(-\xi|1\right)}{p\left(\xi|1\right)}\right) d\xi, \tag{A.33}$$

with

$$p\left(\xi|1\right) \overset{(3.42)}{=} \frac{1}{M\cdot 2^{M-1}} \sum_{m=0}^{M-1} \sum_{x_{l_1}\in\mathbb{X}_1^m} p\left(\xi|x_{l_1}\right), \tag{A.34}$$

$$p\left(\xi|x_{l_1}\right) \overset{(3.47)}{=} \frac{\exp\left(-\frac{\left(\xi-\mu_{l_1,m}\right)^2}{4\mu_{l_1,m}}\right)}{\sqrt{4\pi\mu_{l_1,m}}}, \tag{A.35}$$

of which the mean is

$$\mu_{l_1,m} \overset{(3.48)}{=} \frac{d_{l_1,m}^2}{\sigma_{\mathrm{n}}^2} \overset{(3.46)}{=} \frac{\left|x_{l_1} - x_{l_{0,m}}\right|^2}{\sigma_{\mathrm{n}}^2}. \tag{A.36}$$

The index $l_{0,m}$ of the counterpart symbol to x_{l_1}, i.e., the symbol that has a 0 as the m-th bit label, whereas all other labels are identical to those of x_{l_1}, is determined by

$$l_{0,m} = l_1 - 2^{M-1-m}. \tag{A.37}$$

The partial derivative of $I_{\mathrm{E1}}(1)$ with respect to x_l as defined in (A.2) can be decomposed as

$$\begin{aligned}\frac{\partial I_{\mathrm{E1}}(1)}{\partial x_l} = -\int_{\xi\in\mathbb{R}} \Bigg\{ & \left[\frac{\partial}{\partial x_l} p\left(\xi|1\right)\right] \log_2\left(1 + \frac{p\left(-\xi|1\right)}{p\left(\xi|1\right)}\right) \\ & + p\left(\xi|1\right) \frac{\left[\frac{\partial}{\partial x_l}\frac{p\left(-\xi|1\right)}{p\left(\xi|1\right)}\right]}{\ln(2)\cdot\left(1+\frac{p\left(-\xi|1\right)}{p\left(\xi|1\right)}\right)} \Bigg\} d\xi.\end{aligned} \tag{A.38}$$

Let us introduce the abbreviation

$$\Omega(\xi) = \frac{\partial}{\partial x_l} p\left(\xi|1\right), \tag{A.39}$$

from which it is clear that

$$\Omega(-\xi) = \frac{\partial}{\partial x_l} p\left(-\xi|1\right). \tag{A.40}$$

With that, we can compute the derivative in the second addend as

$$\frac{\partial}{\partial x_l}\frac{p\left(-\xi|1\right)}{p\left(\xi|1\right)} = \frac{p\left(\xi|1\right)\cdot\Omega(-\xi) - p\left(-\xi|1\right)\cdot\Omega(\xi)}{\left[p\left(\xi|1\right)\right]^2}. \tag{A.41}$$

Now, let us examine the derivative $\Omega(\xi)$. With (A.34), we get

$$\Omega(\xi) = \frac{1}{M \cdot 2^{M-1}} \sum_{m=0}^{M-1} \frac{\partial}{\partial x_l} \sum_{x_{l_1} \in \mathbb{X}_1^m} p\left(\xi | x_{l_1}\right). \tag{A.42}$$

For each bit position m, the symbol x_l appears in the second summation only once, either directly as one symbol x_{l_1} or as one of its counterparts, which differs in the m-th bit label only. Let us first assume that the m-th bit label of x_l is 1. Then there exists one l_2, such that $x_{l_2} \in \mathbb{X}_1^m$ and $l_2 = l$. Therefore,

$$\frac{\partial}{\partial x_l} \sum_{x_{l_1} \in \mathbb{X}_1^m} p\left(\xi | x_{l_1}\right) = \frac{\partial}{\partial x_l} p\left(\xi | x_{l_2}\right) = \frac{\partial \mu_{l_2,m}}{\partial x_l} \cdot \frac{\partial p\left(\xi | x_{l_2}\right)}{\partial \mu_{l_2,m}}. \tag{A.43}$$

Taking definition (A.2) into account, the first derivative on the right hand side of (A.43) is

$$\frac{\partial \mu_{l_2,m}}{\partial x_l} = \frac{2}{\sigma_{\mathrm{n}}^2} \left(x_l - x_l^{(m)}\right), \tag{A.44}$$

where $x_l^{(m)}$ is the symbol, which differs only in the m-th bit label compared to x_l. Now, we examine the case, that the m-th bit label of x_l is 0. Then there exists one l_2, such that $x_{l_2} \in \mathbb{X}_1^m$ and $l_{0,m} = l_2 - 2^{M-1-m} = l$. For this case, (A.44) is still true, considering that $x_l^{(m)}$ is now x_{l_2}. Generalizing (A.37), the counterpart symbol to x_l, which differs in the m-th bit label $\mu_m^{-1}(x_l)$[1], is

$$x_l^{(m)} = x_\zeta \text{ with } \zeta = l + 2^{M-1-m} \cdot \left[1 - 2 \cdot \mu_m^{-1}(x_l)\right]. \tag{A.45}$$

In any case, the mean value of interest in equation (A.43) is labeled by l_2, which is either the index of x_l or of its counterpart symbol,

$$\mu_{l_2,m} = \frac{\left|x_l - x_l^{(m)}\right|^2}{\sigma_{\mathrm{n}}^2}. \tag{A.46}$$

Next, we compute the other derivative in (A.43) as

$$\begin{aligned} \frac{\partial p\left(\xi | x_{l_2}\right)}{\partial \mu_{l_2,m}} &\overset{(A.35)}{=} \frac{\partial}{\partial \mu_{l_2,m}} \frac{\exp\left(-\frac{\left(\xi - \mu_{l_2,m}\right)^2}{4\mu_{l_2,m}}\right)}{\sqrt{4\pi\mu_{l_2,m}}} \\ &= \frac{\sqrt{4\pi\mu_{l_2,m}} \cdot D_1 - \exp\left(-\frac{\left(\xi - \mu_{l_2,m}\right)^2}{4\mu_{l_2,m}}\right) \cdot D_2}{4\pi\mu_{l_2,m}}, \end{aligned} \tag{A.47}$$

[1] It should be self-evident, that $\mu(.)$ is the labeling function in this context, not a mean value.

in which the two abbreviations are

$$\begin{aligned} D_1 &= \frac{\partial}{\partial \mu_{l_2,m}} \exp\left(-\frac{\left(\xi - \mu_{l_2,m}\right)^2}{4\mu_{l_2,m}} \right) \\ &= \exp\left(-\frac{\left(\xi - \mu_{l_2,m}\right)^2}{4\mu_{l_2,m}} \right) \cdot \frac{2\mu_{l_2,m} \cdot \left(\xi - \mu_{l_2,m}\right) + \left(\xi - \mu_{l_2,m}\right)^2}{4\mu_{l_2,m}^2} \\ &= \exp\left(-\frac{\left(\xi - \mu_{l_2,m}\right)^2}{4\mu_{l_2,m}} \right) \cdot \frac{\xi^2 - \mu_{l_2,m}^2}{4\mu_{l_2,m}^2}, \end{aligned} \tag{A.48}$$

and

$$D_2 = \frac{\partial}{\partial \mu_{l_2,m}} \sqrt{4\pi\mu_{l_2,m}} = \frac{4\pi}{2\sqrt{4\pi\mu_{l_2,m}}} = \sqrt{\frac{\pi}{\mu_{l_2,m}}}. \tag{A.49}$$

For $\Omega(-\xi)$, we have to invert the argument ξ, which is only included in (A.47). It can be observed that we obtain this expression simply by

$$\frac{\partial p\left(-\xi|x_{l_2}\right)}{\partial \mu_{l_2,m}} = \frac{\partial p\left(\xi|x_{l_2}\right)}{\partial \mu_{l_2,m}} \cdot \frac{\exp\left(-\frac{\left(-\xi - \mu_{l_2,m}\right)^2}{4\mu_{l_2,m}} \right)}{\exp\left(-\frac{\left(\xi - \mu_{l_2,m}\right)^2}{4\mu_{l_2,m}} \right)} = \frac{\partial p\left(\xi|x_{l_2}\right)}{\partial \mu_{l_2,m}} \cdot \mathrm{e}^{-\xi}, \tag{A.50}$$

independent of the bit position m. Next, we combine (A.47) until (A.49) with (A.42) until (A.44) and find after one more manipulation

$$\Omega(\xi) = \Omega_0 \sum_{m=0}^{M-1} \left(x_l - x_l^{(m)} \right) \cdot \frac{\xi^2 - \mu_{l_2,m}^2 - 2\mu_{l_2,m}}{\mu_{l_2,m}^2 \cdot \sqrt{\mu_{l_2,m}}} \cdot \exp\left(-\frac{\left(\xi - \mu_{l_2,m}\right)^2}{4\mu_{l_2,m}} \right), \tag{A.51}$$

in which the preceding factor is

$$\Omega_0 = \left(M \cdot 2^{M+1} \cdot \sqrt{\pi} \cdot \sigma_{\mathrm{n}}^2 \right)^{-1}. \tag{A.52}$$

From (A.50), it is clear that

$$\Omega(-\xi) = \Omega(\xi) \cdot \mathrm{e}^{-\xi}. \tag{A.53}$$

Finally, we insert (A.41) with (A.51) into (A.38) and apply (A.53). This yields

$$\frac{\partial I_{\mathrm{E1}}(1)}{\partial x_l} = -\int_{\xi \in \mathbb{R}} \Omega(\xi) \left\{ \log_2\left(1 + \frac{p\left(-\xi|1\right)}{p\left(\xi|1\right)} \right) + \frac{p\left(\xi|1\right) \cdot \mathrm{e}^{-\xi} - p\left(-\xi|1\right)}{\ln(2) \cdot \left(p\left(\xi|1\right) + p\left(-\xi|1\right) \right)} \right\} d\xi. \tag{A.54}$$

Appendix B

Optimized Mapping Vectors

B.1 Mapping Vectors for Minimum Symbol Error Rate

The optimum 8-QAM constellation at $\frac{\sigma_s^2}{\sigma_n^2} = 14\,\text{dB}$ is described by

$$\begin{matrix}(0.6131+0.5911\text{j} & , & -0.2232-1.0344\text{j} & , & -0.2195+1.0421\text{j} & , & 1.3465, \\ 0.6175-0.5976\text{j} & , & -1.0010-0.4795\text{j} & , & -0.9923+0.4822\text{j} & , & -0.1412-0.0040\text{j})\end{matrix}, \tag{B.1}$$

and at $\frac{\sigma_s^2}{\sigma_n^2} = 20\,\text{dB}$ by

$$\begin{matrix}(0.6220+0.5005\text{j} & , & -0.2079-0.9835\text{j} & , & -0.2071+0.9806\text{j} & , & 1.4391, \\ 0.6203-0.4993\text{j} & , & -1.0316-0.4831\text{j} & , & -1.0345+0.4844\text{j} & , & -0.2001+0.0004\text{j})\end{matrix}. \tag{B.2}$$

The optimum 16-QAM constellation at $\frac{\sigma_s^2}{\sigma_n^2} = 10\,\text{dB}$ equals

$$\begin{matrix}(-0.6070+0.0426\text{j} & , & -0.3189-1.2112\text{j} & , & 0.3163+1.1051\text{j} & , & 0.3882-1.0551\text{j}, \\ -0.1360+0.5903\text{j} & , & 0.4658-0.3700\text{j} & , & 1.1331+0.0353\text{j} & , & 0.5053+0.3207\text{j}, \\ -0.8752-0.6940\text{j} & , & 0.9664+0.7790\text{j} & , & 1.0436-0.6843\text{j} & , & -0.0143+0.0043\text{j}, \\ -1.2531-0.0804\text{j} & , & -0.9482+0.6266\text{j} & , & -0.2084-0.5617\text{j} & , & -0.4573+1.1528\text{j})\end{matrix}, \tag{B.3}$$

at $\frac{\sigma_s^2}{\sigma_n^2} = 14\,\text{dB}$, it is

$$\begin{matrix}(-1.2755-0.0818\text{j} & , & -0.8358-0.6932\text{j} & , & 0.2687+1.0521\text{j} & , & 0.4010-1.0091\text{j}, \\ -0.6212-0.0399\text{j} & , & 0 & , & 1.0800-0.6835\text{j} & , & 0.4792+0.3974\text{j}, \\ -0.9176+0.5807\text{j} & , & 0.5260-0.3329\text{j} & , & 0.9838+0.8159\text{j} & , & -0.3163-1.2383\text{j}, \\ -0.2299+0.5785\text{j} & , & -0.1540-0.6031\text{j} & , & 1.0836+0.0695\text{j} & , & -0.4720+1.1877\text{j})\end{matrix}, \tag{B.4}$$

and at $\frac{\sigma_s^2}{\sigma_n^2} = 20\,\text{dB}$, it is

$$\begin{matrix}(-0.1449-0.7530\text{j} & , & -0.7241+0.2545\text{j} & , & 0.4377-0.4189\text{j} & , & -0.7463+0.9246\text{j}, \\ 0.4189-1.1170\text{j} & , & -0.1427-0.0816\text{j} & , & -0.7325-0.4195\text{j} & , & -1.3147-0.0651\text{j}, \\ 1.0310-0.0879\text{j} & , & 1.0332+0.5926\text{j} & , & 0.4452+0.9342\text{j} & , & -0.1412+0.5897\text{j}, \\ 0.4434+0.2549\text{j} & , & -0.1524+1.2627\text{j} & , & 1.0102-0.7714\text{j} & , & -0.7209-1.0985\text{j})\end{matrix}. \tag{B.5}$$

For $L = 32$, the optimum constellation at $\frac{\sigma_s^2}{\sigma_n^2} = 14\,\text{dB}$ is described by

$$\begin{array}{cccc}
(-0.0435-1.3155\text{j} & , & 0.3974+0.1968\text{j} & , & 0.1200-0.8767\text{j} & , & 0.4845+0.6432\text{j}, \\
-0.0002-0.0021\text{j} & , & -0.3556+0.6842\text{j} & , & 0.9685+0.7637\text{j} & , & -0.7860-0.0025\text{j}, \\
-0.9876-0.8602\text{j} & , & -0.3569-0.7465\text{j} & , & 0.9915-0.7298\text{j} & , & 0.3600-0.2293\text{j}, \\
0.0724+0.8842\text{j} & , & 0.7984-0.2027\text{j} & , & -0.9639+0.8840\text{j} & , & 1.2698-0.2710\text{j}, \\
0.5728-1.1398\text{j} & , & 1.3048+0.2655\text{j} & , & -0.7432-0.4725\text{j} & , & -1.2205-0.2806\text{j}, \\
-0.3716+0.2160\text{j} & , & -0.5577-1.1681\text{j} & , & -0.4967+1.1441\text{j} & , & -0.3746-0.2389\text{j}, \\
0.8198+0.2723\text{j} & , & -1.2300+0.2911\text{j} & , & -0.0027-0.4586\text{j} & , & 0.5366+1.1262\text{j}, \\
0.5202-0.6306\text{j} & , & 0.0343+0.4419\text{j} & , & -0.7601+0.4632\text{j} & , & 1.3494\text{j})
\end{array}, \tag{B.6}$$

at $\frac{\sigma_s^2}{\sigma_n^2} = 20\,\text{dB}$, the 32-QAM with minimum SER is

$$\begin{array}{cccc}
(\ 0.0002-0.0440\text{j} & , & -1.2462-0.2753\text{j} & , & -0.4111-0.2759\text{j} & , & -0.4333+1.1388\text{j}, \\
0.8300-0.9977\text{j} & , & -0.8203+0.4335\text{j} & , & -0.0029-0.5134\text{j} & , & 0.8227-0.0475\text{j}, \\
-0.4072+0.1932\text{j} & , & -1.2439+0.2179\text{j} & , & 1.2431-0.2921\text{j} & , & 0.4000-1.2406\text{j}, \\
0.0051+0.8955\text{j} & , & 0.8224-0.5210\text{j} & , & 0.4107+0.1888\text{j} & , & 1.2465+0.2014\text{j}, \\
0.4083-0.2808\text{j} & , & 0.4193+0.6612\text{j} & , & -0.8428-0.9865\text{j} & , & 0.8263+0.4241\text{j}, \\
-0.8637+0.9048\text{j} & , & -0.0063-0.9898\text{j} & , & 1.3700\text{j} & , & 0.8845+0.8949\text{j}, \\
0.4417+1.1356\text{j} & , & -0.4099+0.6645\text{j} & , & -0.4161-1.2351\text{j} & , & 0.0034+0.4251\text{j}, \\
-0.8226-0.0372\text{j} & , & -0.4164-0.7486\text{j} & , & -0.8286-0.5101\text{j} & , & 0.4071-0.7539\text{j})
\end{array}, \tag{B.7}$$

and at $\frac{\sigma_s^2}{\sigma_n^2} = 24\,\text{dB}$, it is

$$\begin{array}{cccc}
(\ 0.8265-0.5183\text{j} & , & -1.2457+0.2047\text{j} & , & 0.4118-0.2813\text{j} & , & 0.8236-0.0433\text{j}, \\
-0.4085-1.2399\text{j} & , & 0.4140-1.2372\text{j} & , & -0.4128-0.2812\text{j} & , & -0.4113-0.7556\text{j}, \\
0.0017-0.9951\text{j} & , & 0.4111+0.1927\text{j} & , & 0.8354+0.9067\text{j} & , & -0.8267-0.0441\text{j}, \\
-0.8266+0.4306\text{j} & , & -0.8268-0.5201\text{j} & , & -0.0011+0.4268\text{j} & , & -0.4133+0.6648\text{j}, \\
1.2420+0.2002\text{j} & , & -0.4127+0.1905\text{j} & , & 0.8322-0.9951\text{j} & , & 1.3821\text{j}, \\
-0.8356+0.9068\text{j} & , & 0.4177+1.1439\text{j} & , & 0.0001+0.9013\text{j} & , & 0.0004-0.0452\text{j}, \\
-0.5177\text{j} & , & 0.4133-0.7559\text{j} & , & 0.8256+0.4310\text{j} & , & -1.2427-0.2823\text{j}, \\
1.2424-0.2837\text{j} & , & -0.8288-0.9979\text{j} & , & -0.4185+1.1454\text{j} & , & 0.4133+0.6665\text{j})
\end{array}. \tag{B.8}$$

B.2 Mapping Vectors for Maximum BICM Capacity

For a code rate of $R_c = 0.999$, the optimum 8-QAM mapping is

$$\begin{array}{cccc}
(-0.9708+0.5223\text{j} & , & 0.5972+0.4628\text{j} & , & 0.6192-0.5660\text{j} & , & 1.3978, \\
-0.2058-0.0269\text{j} & , & -1.0613-0.4654\text{j} & , & -0.1370+1.0593\text{j} & , & -0.2392-0.9859\text{j})
\end{array}, \tag{B.9}$$

while the optimum 16-QAM mapping for this purpose is

$$\begin{array}{cccc}
(\ 1.0311+0.5864\text{j} & , & 1.0412-0.0864\text{j} & , & 0.4297+0.2532\text{j} & , & -0.1278-0.0854\text{j}, \\
0.4422-1.1093\text{j} & , & 1.0342-0.7552\text{j} & , & -0.7257-1.1047\text{j} & , & 0.4353-0.4211\text{j}, \\
-0.1417+1.2611\text{j} & , & -0.7435+0.9270\text{j} & , & 0.4086+0.9216\text{j} & , & -0.1405+0.5866\text{j}, \\
-0.7577+0.2536\text{j} & , & -1.3125-0.0819\text{j} & , & -0.7202-0.4174\text{j} & , & -0.1348-0.7536\text{j})
\end{array}. \tag{B.10}$$

B.3 Mapping Vectors for Maximum Signal Set Capacity

The optimum 8-QAM with respect to C_S for $R_c = 0.75$ is

$$\begin{array}{cccc} (\ 0.1880-1.0569\mathrm{j} & , & 1.1189-0.4053\mathrm{j} & , & 0.3070 & , & 0.8642+0.7615\mathrm{j} \\ -0.8738-0.7537\mathrm{j} & , & -1.1113+0.4285\mathrm{j} & , & -0.3135-0.0003\mathrm{j} & , & -0.1823+1.0623\mathrm{j}) \end{array}, \quad \text{(B.11)}$$

the optimum 16-QAM mapping for this purpose is

$$\begin{array}{cccc} (\ 0 & , & 0.5750 & , & 0.2875+0.4980\mathrm{j} & , & -0.2875+0.4980\mathrm{j}, \\ -0.5750+0.0000\mathrm{j} & , & -0.2875-0.4980\mathrm{j} & , & 0.2875-0.4980\mathrm{j} & , & 1.1973+0.3521\mathrm{j}, \\ 0.6908+1.0393\mathrm{j} & , & -0.1388+1.2402\mathrm{j} & , & -0.9035+0.8608\mathrm{j} & , & -1.2455+0.0787\mathrm{j}, \\ -1.0046-0.7403\mathrm{j} & , & -0.2937-1.2129\mathrm{j} & , & 0.5546-1.1179\mathrm{j} & , & 1.1435-0.4999\mathrm{j}) \end{array}. \quad \text{(B.12)}$$

The optimum 32-QAM mapping for a code of rate $R_c = 0.5$ is

$$\begin{array}{cccc} (\ 0 & , & 0.0500\mathrm{j} & , & 0.0500\mathrm{j} & , & 0.6000, \\ 0.7461+0.2716\mathrm{j} & , & 0.4596+0.3857\mathrm{j} & , & 0.3970+0.6876\mathrm{j} & , & 0.1042+0.5909\mathrm{j}, \\ -0.1379+0.7819\mathrm{j} & , & -0.3000+0.5196\mathrm{j} & , & -0.6082+0.5104\mathrm{j} & , & -0.5638+0.2052\mathrm{j}, \\ -0.7940 & , & -0.5638-0.2052\mathrm{j} & , & -0.6082-0.5104\mathrm{j} & , & -0.3000-0.5196\mathrm{j}, \\ -0.1379-0.7819\mathrm{j} & , & 0.1042-0.5909\mathrm{j} & , & 0.3970-0.6876\mathrm{j} & , & 0.4596-0.3857\mathrm{j}, \\ 0.7461-0.2716\mathrm{j} & , & 1.4265+0.2515\mathrm{j} & , & 1.0641+0.9829\mathrm{j} & , & 0.3638+1.4021\mathrm{j}, \\ -0.4520+1.3762\mathrm{j} & , & -1.1243+0.9134\mathrm{j} & , & -1.4396+0.1606\mathrm{j} & , & -1.2979-0.6433\mathrm{j}, \\ -0.7441-1.2428\mathrm{j} & , & 0.0460-1.4478\mathrm{j} & , & 0.8214-1.1931\mathrm{j} & , & 1.3361-0.5596\mathrm{j}) \end{array}, \quad \text{(B.13)}$$

and for $R_c = 0.75$, it is described by

$$\begin{array}{cccc} (-0.3313+0.7406\mathrm{j} & , & 1.0186+0.9610\mathrm{j} & , & -0.8445+0.0014\mathrm{j} & , & 0.3783-0.1946\mathrm{j}, \\ 0.1451-0.1112\mathrm{j} & , & -0.8501+1.0492\mathrm{j} & , & -0.6110-0.2580\mathrm{j} & , & -1.2445+0.4809\mathrm{j}, \\ -0.5748-0.7132\mathrm{j} & , & 0.0400+0.4334\mathrm{j} & , & -0.4813-1.3221\mathrm{j} & , & -0.2632+1.3580\mathrm{j}, \\ 1.3775+0.1239\mathrm{j} & , & -0.0397+0.1790\mathrm{j} & , & 0.2375-1.3720\mathrm{j} & , & 0.4721+0.2230\mathrm{j}, \\ 1.0834-0.7646\mathrm{j} & , & -0.1198-0.4451\mathrm{j} & , & 0.0863+0.8693\mathrm{j} & , & 0.4326+1.2898\mathrm{j}, \\ -0.3892+0.2260\mathrm{j} & , & 0.5509-0.5382\mathrm{j} & , & 0.2642-0.7816\mathrm{j} & , & 0.4603+0.6394\mathrm{j}, \\ 1.1953-0.5264\mathrm{j} & , & -0.1304-0.8471\mathrm{j} & , & -0.2390-0.1905 & , & 0.9000+0.4117\mathrm{j}, \\ 0.8099-0.1200\mathrm{j} & , & -0.6815+0.4759\mathrm{j} & , & -1.3832-0.1781\mathrm{j} & , & 0.8464-1.0546\mathrm{j}) \end{array}. \quad \text{(B.14)}$$

B.4 Mapping Vectors for Tradeoffs

The optimum 16-QAM mapping for a weighting factor of $a_W = 0$ is

$$\begin{array}{cccc} (1.0216+0.2772\mathrm{j} & , & 0.8977+0.8906\mathrm{j} & , & -0.9965+0.2794\mathrm{j} & , & -0.9049+0.8837\mathrm{j}, \\ 0.2968+0.2982\mathrm{j} & , & 0.2829+1.0239\mathrm{j} & , & -0.2824+0.2978\mathrm{j} & , & -0.2865+1.0264\mathrm{j}, \\ 1.0145-0.2899\mathrm{j} & , & 0.8855-0.9020\mathrm{j} & , & -1.0206-0.2757\mathrm{j} & , & -0.9011-0.8832\mathrm{j}, \\ 0.2962-0.2939\mathrm{j} & , & 0.2727-1.0269\mathrm{j} & , & -0.2845-0.2933\mathrm{j} & , & -0.2917-1.0123\mathrm{j}) \end{array}, \quad \text{(B.15)}$$

for $a_W = 0.5$, we obtain

$$\begin{array}{cccc} (-1.0352-0.2715\mathrm{j} & , & 0.3265+0.9990\mathrm{j} & , & 0.3469+0.2926\mathrm{j} & , & -0.8964-0.8560\mathrm{j}, \\ 0.3359-0.9924\mathrm{j} & , & -1.0326+0.2581\mathrm{j} & , & -0.9091+0.8454\mathrm{j} & , & 0.3505-0.2866\mathrm{j}, \\ 0.9006+0.8573\mathrm{j} & , & -0.3485-0.2930\mathrm{j} & , & -0.3255-1.0075\mathrm{j} & , & 1.0314+0.2695\mathrm{j}, \\ -0.3510+0.2913\mathrm{j} & , & 0.9070-0.8500\mathrm{j} & , & 1.0370-0.2611\mathrm{j} & , & -0.3372+1.0048\mathrm{j}) \end{array}, \quad \text{(B.16)}$$

for $a_{\mathrm{W}} = 0.9$

$$\begin{array}{cccc} (-1.0266-0.2601\mathrm{j} & , & 0.3445+0.9885\mathrm{j} & , & 0.4141+0.2997\mathrm{j} & , & -0.9014-0.8384\mathrm{j}, \\ 0.3540-0.9821\mathrm{j} & , & -1.0250+0.2471\mathrm{j} & , & -0.9135+0.8276\mathrm{j} & , & 0.4187-0.2931\mathrm{j}, \\ 0.9047+0.8391\mathrm{j} & , & -0.4161-0.3007\mathrm{j} & , & -0.3427-0.9957\mathrm{j} & , & 1.0228+0.2583\mathrm{j}, \\ -0.4183+0.2985\mathrm{j} & , & 0.9110-0.8318\mathrm{j} & , & 1.0277-0.2498\mathrm{j} & , & -0.3542+0.9928\mathrm{j}) \end{array} , \quad \text{(B.17)}$$

and for $a_{\mathrm{W}} = 0.95$

$$\begin{array}{cccc} (-1.0120-0.2324\mathrm{j} & , & 0.3885+0.9671\mathrm{j} & , & 0.5362+0.3136\mathrm{j} & , & -0.9028-0.7943\mathrm{j}, \\ 0.3983-0.9609\mathrm{j} & , & -1.0117+0.2201\mathrm{j} & , & -0.9140+0.7832\mathrm{j} & , & 0.5416-0.3057\mathrm{j}, \\ 0.9054+0.7941\mathrm{j} & , & -0.5385-0.3148\mathrm{j} & , & -0.3857-0.9728\mathrm{j} & , & 1.0085+0.2310\mathrm{j}, \\ -0.5407+0.3116\mathrm{j} & , & 0.9112-0.7868\mathrm{j} & , & 1.0122-0.2221\mathrm{j} & , & -0.3966+0.9691\mathrm{j}) \end{array} . \quad \text{(B.18)}$$

The suboptimum mapping for $a_{\mathrm{W}} = 0.8$ is described by

$$\begin{array}{cccc} (\ 0.4214-0.4114\mathrm{j} & , & 0.9036+0.9147\mathrm{j} & , & -0.9036-0.9147\mathrm{j} & , & -0.4214+0.4114\mathrm{j}, \\ -0.9147+0.9036\mathrm{j} & , & -0.4114-0.4214\mathrm{j} & , & 0.4114+0.4214\mathrm{j} & , & 0.9147-0.9036\mathrm{j}, \\ -1.2857+0.0078\mathrm{j} & , & -0.0071-0.5889\mathrm{j} & , & 0.0071+0.5889\mathrm{j} & , & 1.2857-0.0078\mathrm{j}, \\ 0.5889-0.0071\mathrm{j} & , & 0.0078+1.2857\mathrm{j} & , & -0.0078-1.2857\mathrm{j} & , & -0.5889+0.0071\mathrm{j}) \end{array} . \quad \text{(B.19)}$$

Bibliography

[1] S. Cherry, “Edholm’s law of bandwidth,” *IEEE Spectrum*, vol. 41, no. 7, pp. 58–60, July 2004.

[2] C. Campopiano and B. Glazer, “A coherent digital amplitude and phase modulation scheme,” *IRE Transactions on Communications Systems*, vol. 10, no. 1, pp. 90–95, Mar. 1962.

[3] M. Ohm, “Multilevel optical modulation formats with direct detection,” Ph.D. dissertation, Universität Stuttgart, Institut für Nachrichtenübertragung, Pfaffenwaldring 47, 70569 Stuttgart, Germany, June 2006, ISBN 3-8322-5357-2.

[4] C. E. Shannon, “A mathematical theory of communication,” *Bell Labs System Journal*, vol. 27, pp. 379–423, 623–656, July, Oct. 1948.

[5] J. L. Massey, “Coding and modulation in digital communications,” in *Int. Zürich Sem. Digital Communications*, Zürich, Switzerland, Mar. 1974.

[6] G. Ungerböck and I. Csajka, “On improving data-link performance by increasing the channel alphabet and introducing sequence coding,” in *Int. Symp. Inf. Theory (ISIT)*, Ronneby, Schweden, June 1976.

[7] G. Ungerböck, “Channel coding with multilevel/phase signals,” *IEEE Trans. Inform. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[8] D. Divsalar and M. K. Simon, “The design of trellis coded MPSK for fading channels: performance criteria,” *IEEE Trans. Commun.*, vol. 36, no. 9, pp. 1004–1012, Sept. 1988.

[9] H. Imai and S. Hirakawa, “A new multilevel coding method using error-correcting codes,” *IEEE Trans. Inform. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.

[10] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, “Multilevel codes: theoretical concepts and practical design rules,” *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.

[11] F. Gray, “Pulse code communications,” *U.S. Patent No. 2632058*, Mar. 1953.

[12] E. Zehavi, “8-PSK trellis codes on Rayleigh channel,” *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.

[13] G. Caire, G. Taricco, and E. Biglieri, “Capacity of bit-interleaved channels,” *IEE Electronics Letters*, vol. 32, no. 12, pp. 1060–1061, June 1996.

[14] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: Turbo-codes,” in *IEEE International Conference on Communications (ICC)*, vol. 2, May 1993, pp. 1064–1070.

[15] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, “Design of capacity-approaching irregular low-density parity-check codes,” *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[16] X. Li and J. A. Ritcey, “Bit-interleaved coded modulation with iterative decoding,” *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 169–171, Nov. 1997.

[17] ——, “Bit-interleaved coded modulation with iterative decoding using soft feedback,” *Electronics Letters*, vol. 34, no. 10, pp. 942–943, May 1998.

[18] S. ten Brink, J. Speidel, and R. H. Yan, “Iterative demapping and decoding for multilevel modulation,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 1, Nov. 1998, pp. 579–584.

[19] ——, “Iterative demapping for QPSK modulation,” *Electronics Letters*, vol. 34, no. 15, pp. 1459–1460, July 1998.

[20] A. Chindapol and J. Ritcey, “Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels,” *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 944–957, May 2001.

[21] X. Li, A. Chindapol, and J. Ritcey, “Bit-interleaved coded modulation with iterative decoding and 8PSK signaling,” *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1250–1257, Aug. 2002.

[22] X. Li and J. Ritcey, “Trellis-coded modulation with bit interleaving and iterative decoding,” *IEEE J. Select. Areas Commun.*, vol. 17, no. 4, pp. 715–724, Apr. 1999.

[23] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch, “Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding,” *IEEE Commun. Lett.*, vol. 7, no. 12, pp. 593–595, Dec. 2003.

[24] ——, “Optimized symbol mappings for bit-interleaved coded modulation with iterative decoding,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 6, San Francisco, CA, USA, Dec. 2003, pp. 3316–3320.

[25] F. Schreckenbach and G. Bauch, “EXIT charts for iteratively decoded multilevel modulation,” in *12th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, Sept. 2004.

[26] N. S. Muhammad and J. Speidel, “Joint optimization of signal constellation and bit labeling for bit-interleaved coded modulation with iterative decoding,” *IEEE Commun. Lett.*, vol. 9, no. 9, pp. 775–777, Sept. 2005.

[27] ——, “Design of multidimensional mappings for iterative MIMO detection with minimized bit error floor,” in *4th International Symposium on Turbo Codes & Related Topics in connection with the 6th International ITG-Conference on Source and Channel Coding (SCC)*, Apr. 2006.

[28] F. Schreckenbach, “Iterative decoding of bit-interleaved coded modulation,” Ph.D. dissertation, Technische Universität München (TUM), Lehrstuhl für Nachrichtentechnik (LNT), Theresienstraße 90, 80333 München, Germany, May 2007.

[29] E. M. Deloraine, “Single sideband short wave telephone,” *Société Française Electricité*, pp. 940–1009, Sept. 1932, Bull. 2.

[30] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill Book Company, 2008, ISBN 978-0-07-295716-7.

[31] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill Book Company, 1965, ISBN 0-704-8448-1.

[32] R. F. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: John Wiley & Sons, Inc., 2002, ISBN 0471-22410-3.

[33] G. Caire, G. Taricci, and E. Biglieri, “Bit-interleaved coded modulation,” *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 927–945, May 1998.

[34] W. Davenport, *Probability and Random Processes, an Introduction for Applied Scientists and Engineers*. New York: McGraw-Hill College, 1970, ISBN 0-070-15440-6.

[35] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

[36] İ. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Europ. Trans. Telecom.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[37] W. C. Jakes, *Microwave Mobile Communications*. New York: John Wiley & Sons, Inc., 1974, ISBN 0-471-43720-4.

[38] A. Wittneben, “A new bandwidth efficient transmit antenna mdulation diversity scheme for linear digital modulations,” in *IEEE International Conference on Communications (ICC)*, vol. 2, May 1993, pp. 1630–1634.

[39] N. Seshadri and J. H. Winters, “Two signaling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity,” in *IEEE Vehicular Technology Conference (VTC)*, vol. 43, Secaucus, NJ, USA, May 1993, pp. 508–511.

[40] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003, ISBN 0-521-82615-2.

[41] V. Branka and Y. Jinhong, *Space-time coding*. Chichester: Wiley, 2003, ISBN 0-470-84757-3.

[42] J. Speidel, “Multiple-Input Multiple-Output (MIMO) - Drahtlose Nachrichtenübertragung hoher Bitrate und Qualität mit Mehrfachantennen,” *TeleKommunikation Aktuell*, vol. 59, no. 07-10, July - Oct. 2005.

[43] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, “V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel,” in *URSI Int. Symp. on Signals, Systems, and Electronics (ISSSE)*, Sept./Oct. 1998, pp. 295–300.

[44] B. M. Hochwald and S. ten Brink, “Achieving near-capacity on a multiple-antenna channel,” *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[45] A. Boronka, “Verfahren mit adaptiver Symbolauslöschung zur iterativen Detektion codierter MIMO-Signale,” Ph.D. dissertation, Universität Stuttgart, Institut für Nachrichtenübertragung, Pfaffenwaldring 47, 70569 Stuttgart, Germany, Nov. 2004, ISBN 3-832-23759-3.

[46] D. Schneider, J. Speidel, L. Stadelmeier, D. Schill, and A. Schwager, “MIMO for inhome power line communications,” in *International ITG Conference on Source and Channel Coding (SCC)*, Jan. 2008.

[47] G. Tauböck and W. Henkel, “MIMO systems in the subscriber-line network,” in *5th Int. OFDM Workshop*, Sept. 2000, pp. 18.1–18.3.

[48] W. Schöllmann, S. ; Rosenkranz, “Experimental equalization of crosstalk in a 2×2 MIMO system based on mode group diversity multiplexing in MMF systems @ 10.7 Gb/s,” in *33rd European Conference and Exhibition on Optical Communication ECOC*, Sept. 2007.

[49] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. New Jersey: Prentice Hall, 1982, ISBN 0-132- 83796-X.

[50] R. H. Morelos-Zaragoza, *The art of error correcting coding*. Chichester: Wiley, 2005, ISBN 0-471-49581-6.

[51] W. W. Peterson and E. J. Weldon, *Error-correcting codes*, 2nd ed. Cambridge: MIT Press, 1994, ISBN 0-262-16039-0.

[52] A. M. Michelson and A. H. Levesque, *Error-control techniques for digital communication*. New York: Wiley, 1985, ISBN 0-471-88074-4.

[53] R. G. Gallagher, *Low Density Parity-Check Codes*. Cambridge: MIT Press, 1963, ISBN 0-262-57177-3.

[54] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[55] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, no. 5, pp. 533–547, Sept. 1981.

[56] P. Elias, "Coding for noisy channels," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 29–37, Sept. 1954.

[57] G. Jr. Forney, "Convolutional codes I: Algebraic structure," *IEEE Trans. Inform. Theory*, vol. 16, no. 6, pp. 720–738, Nov. 1970.

[58] R. Johannesson and K. S. Zigangirov, *Fundamentals of convolutional coding*. New York: IEEE Press, 1999, ISBN 0-780-33483-3.

[59] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Lett.*, vol. 1, no. 1, pp. 22–24, Jan. 1997.

[60] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, no. 6, pp. 284–287, Mar. 1974.

[61] A. J. Viterbi, "An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 260–264, Feb. 1998.

[62] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," in *IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 3, Dallas, TX, USA, Nov. 1989, pp. 1680–1686.

[63] N. Wiberg, "Codes and decoding on general graphs," Ph.D. dissertation, Linköping University, Linköping, Sweden, Oct. 1996.

[64] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Mar. 2001.

[65] R. McEliece, D. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's "belief propagation"algorithm," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 140–152, Feb. 1998.

[66] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

[67] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 1991, ISBN 0-471-06259-6.

[68] G. D. Forney, *Concatenated Codes*. Cambridge: MIT Press, 1966, ISBN 0-262-06015-9.

[69] I. S. Reed and G. Solomon, “Polynomial codes over certain finite fields,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, June 1960.

[70] Consultative Committee for Space Data System, “Telemetry channel coding,” *Recommendation for Space Data System Standards*, no. 2, Jan. 1987, CCSDS 101, 0-B-2, Blue Book.

[71] A. J. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE Trans. Inform. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

[72] G. Forney, “The Viterbi algorithm,” *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[73] R. Shao, S. Lin, and M. Fossorier, “Two simple stopping criteria for turbo decoding,” *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1117–1120, Aug. 1999.

[74] Y. Wu, B. Woerner, and W. Ebel, “A simple stopping criterion for turbo decoding,” *IEEE Commun. Lett.*, vol. 4, no. 8, pp. 258–260, Aug. 2000.

[75] M. Breiling and L. Hanzo, “Optimum non-iterative turbo-decoding,” in *Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 2, Helsinki, Finland, Sept. 2002, pp. 714–718.

[76] J. Hagenauer, “The turbo principle: Tutorial introduction and state of the art,” in *1st Int. Symp. Turbo Codes*, vol. 2, Brest, France, Sept. 1997, pp. 1–12.

[77] R. Kötter, A. Singer, and M. Tüchler, “Turbo equalization,” *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 67–80, Jan. 2004.

[78] R. Bauer and J. Hagenauer, “Iterative source/channel-decoding using reversible variable length codes,” in *Data Compression Conference (DCC)*, Snowbird, UT, USA, Mar. 2000, pp. 93–102.

[79] H. Poor, “Iterative multiuser detection,” *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 81–88, Jan. 2004.

[80] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 409–428, Mar. 1996.

[81] ——, "Design of parallel concatenated convolutional codes," *IEEE Trans. Commun.*, vol. 44, no. 5, pp. 591–600, May 1996.

[82] H. Sadjadpour, N. Sloane, M. Salehi, and G. Nebe, "Interleaver design for turbo codes," *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 831–837, May 2001.

[83] S. Benedetto and G. Montorsi, "Serial concatenation of block and convolutional codes," *IEE Electronics Letters*, vol. 32, no. 10, pp. 887–888, May 1996.

[84] ——, "Iterative decoding of serially concatenated convolutional codes," *IEE Electronics Letters*, vol. 32, no. 13, pp. 1186–1188, June 1996.

[85] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

[86] L. Perez, J. Seghers, and D. Costello, "A distance spectrum interpretation of turbo codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1698–1709, Nov. 1996.

[87] S. ten Brink, "Convergence of iterative decoding," *Electronics Letters*, vol. 35, no. 10, pp. 806–808, May 1999.

[88] ——, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[89] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal map decoding algorithms operating in the log domain," in *IEEE International Conference on Communications (ICC)*, vol. 2, June 1995, pp. 1009–1013.

[90] H. Wang, H. Yang, and D. Yang, "Improved log-MAP decoding algorithm for turbo-like codes," *IEEE Commun. Lett.*, vol. 10, no. 3, pp. 186–188, Mar. 2006.

[91] D.-F. Yuan and X.-H. Shan, "Research on the Viterbi and BCJR decoding schemes of convolutional codes under different sources," in *IEEE Vehicular Technology Conference (VTC)*, vol. 2, Rhodes, Greece, May 2001, pp. 1377–1381.

[92] S. Benedetto and E. Biglieri, *Principles of digital transmission: with wireless applications*. New York: Kluwer Academic / Plenum Publ., 1999, ISBN 0-306-45753-9.

[93] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968, ISBN 0-471-29048-3.

[94] E. Agrell, J. Lassing, E. G. Strom, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.

[95] C. Stierstorfer and R. F. H. Fischer, "(Gray) mappings for bit-interleaved coded modulation," in *IEEE Vehicular Technology Conference (VTC)*, Dublin, Irleand, Apr. 2007, pp. 1703–1707.

[96] S. ten Brink, "Design of concatenated coding schemes based on iterative decoding convergence," Ph.D. dissertation, Universität Stuttgart, Institut für Nachrichtenübertragung, Pfaffenwaldring 47, 70569 Stuttgart, Germany, Dec. 2001, ISBN 3-8322-0684-1.

[97] ——, "Designing iterative decoding schemes with the extrinsic information transfer chart," *AEÜ Arch. Elektron. Übertragung*, vol. 54, no. 6, pp. 389–398, Dec. 2000.

[98] H. El Gamal and A. Hammons Jr., "Analyzing the turbo decoder using the Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 671–686, Feb. 2001.

[99] I. N. Bronstein, K. A. Semendyayev, G. Musiol, and H. Muehlig, *Handbook of Mathematics*, 5th ed. Berlin: Springer, 2007, ISBN 3-540-72121-5.

[100] S. ten Brink, "Exploiting the chain rule of mutual information for the design of iterative decoding scheme," in *IEEE Allerton Conference on Communication, Control and Computing*, Oct. 2001, pp. 293–300.

[101] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[102] N. S. Muhammad, "Analytische Berechnung von Kenngrößen des EXIT-Charts und Optimierung von QAM-Signalpunkten," in *ITG-Fachtagung Angewandte Inf.theorie*, Apr. 2004, in German.

[103] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Waveland Press, 1990, ISBN 0881-33554-1.

[104] S. ten Brink, "A rate one-half code for approaching the Shannon limit by 0.1dB," *IEE Electronics Letters*, vol. 36, pp. 1293–1294, July 2000.

[105] J. Boutros, G. Caire, E. Viterbo, H. Sawaya, and S. Vialle, "Turbo code at 0.03 dB from capacity limit," in *IEEE International Symposium on Information Theory (ISIT)*, Lausanne, Switzerland, July 2002, p. 56.

[106] G. Foschini, R. Gitlin, and S. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Trans. Commun.*, vol. 22, no. 1, pp. 28–38, Jan. 1974.

[107] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press Inc, 1996, ISBN 1886-52904-3.

[108] D. G. Lueneberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969, ISBN 0471-55359-X.

[109] G. Jr. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Select. Areas Commun.*, vol. 2, no. 5, pp. 632–647, Sept. 1984.

[110] M. Simon and J. Smith, "Hexagonal multiple phase-and-amplitude-shift-keyed signal sets," *IEEE Trans. Commun.*, vol. 21, no. 10, pp. 1108–1115, Oct. 1973.

[111] S.-J. Park, "Triangular quadrature amplitude modulation," *IEEE Commun. Lett.*, vol. 11, no. 4, pp. 292–294, Apr. 2007.

[112] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.

[113] P. Vitthaladevuni and M.-S. Alouini, "Exact BER computation for the cross 32-QAM constellation," in *First International Symposium on Control, Communications and Signal Processing*, Mar. 2004, pp. 643–646.

[114] G. Dueck, "New optimization heuristics: The great deluge algorithm and the record-to-record travel," *Journal of Computational Physics*, vol. 104, no. 1, pp. 86–92, Jan. 1993.

[115] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity limits of information transport in fiber-optic networks," *Physical Review Letters*, vol. 101, Oct. 2008, paper 163901.

[116] T. Freckmann, R.-J. Essiambre, P. J. Winzer, G. J. Foschini, and G. Kramer, "Fiber capacity limits with optimized ring constellations," *IEEE Photon. Technol. Lett.*, submitted for publication.

[117] Y. Li and W. E. Ryan, "Design of LDPC-coded modulation schemes," in *3rd International Symposium on Turbo Codes*, Sept. 200, pp. 551–554.

[118] Y. Li and W. Ryan, "Bit-reliability mapping in LDPC-coded modulation systems," *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 1–3, Jan. 2005.

[119] S. Y. L. Goff, "Signal constellations for bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 49, no. 1, pp. 307–313, Jan. 2003.

[120] D. Sommer and G. P. Fettweis, "Approaching capacity by equiprobable signaling on the Gaussian channel," in *International ITG Conference on Source and Channel Coding (SCC)*, Jan. 2000.

[121] C. Fragouli, R. D. Wesel, D. Sommer, and G. P. Fettweis, "Turbo codes with non-uniform constellations," in *IEEE International Conference on Communications (ICC)*, June 2001.

[122] S. Y. L. Goff, B. S. Sharif, and S. A. Jimaa, "Bit-interleaved turbo-coded modulation using shaping coding," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 246–248, Mar. 2005.

[123] W. Betts, A. R. Calderbank, and R. Laroia, "Performance of nonuniform constellations on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1633–1638, Sept. 1994.

[124] G. Jr. Forney and L.-F. Wei, "Multidimensional constellations - Part I: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.

[125] A. Boronka, N. Muhammad, and J. Speidel, "Removing error floor for bit interleaved coded modulation MIMO transmission with iterative detection," in *IEEE International Conference on Communications (ICC)*, vol. 4, May 2005, pp. 2392–2396.

[126] A. Alvarado, L. Szczecinski, and R. Feick, "On the distribution of extrinsic L-values in Gray-mapped 16-QAM," in *AMC International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug. 2007, pp. 329–335.

[127] E. B. Planas, "Optimization of the signal points and their encoding of a quadrature-amplitude modulation," Master's thesis, Institut für Nachrichtenübertragung, Universität Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, Jan. 2004, Number 2115D.

[128] S. Y. L. Goff, A. Glavieux, and C. Berrou, "Turbo-codes and high spectral efficiency modulation," in *IEEE International Conference on Communications (ICC)*, vol. 2, May 1994, pp. 645–649.

[129] F. Simoens, H. Wymeersch, and M. Moeneclaey, "Spatial mapping for MIMO systems," in *IEEE Information Theory Workshop*, Oct. 2004, pp. 187–192.

[130] S. Bäro, "Turbo detection for MIMO systems: Bit labeling and pre-coding," in *4th International ITG Conference on Source and Channel Coding (SCC)*, Jan. 2004.

[131] K. Zeger and A. Gersho, "Pseudo-Gray coding," *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2147–2158, Dec. 1990.

[132] F. Breyer, "Untersuchung von multidimensionalen Mappings für MIMO-Systeme," Master's thesis, Institut für Nachrichtenübertragung, Universität Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, May 2004, Number 2115D.

[133] C. Deigendesch, "Optimierung quasi-multidimensionaler Mappings für iterative Demodulation," Master's thesis, Institut für Nachrichtenübertragung, Universität Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, June 2007, Number 2192D.

[134] X. Ma and L. Ping, "Coded modulation using superimposed binary codes," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3331–3343, Dec. 2004.

[135] J. Tong and L. Ping, "Iterative decoding of superposition coding," in *4th International Symposium on Turbo Codes & Related Topics in connection with the 6th International ITG-Conference on Source and Channel Coding (SCC)*, Apr. 2006.

[136] S. Pfletschinger and M. Navarro, "A low complexity MIMO system based on BICM with iterative decoding," in *The 13th IST Mobile & Wireless Communications Summit*, June 2004.

[137] C. Studer and H. Bolcskei, "Soft-input soft-output sphere decoding," in *IEEE International Symposium on Information Theory (ISIT)*, July 2008, pp. 2007–2011.

[138] F. Simoens, H. Wymeersch, H. Bruneel, and M. Moeneclaey, "Multidimensional mapping for bit-interleaved coded modulation with BPSK/QPSK signaling," *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 453–455, May 2005.

[139] F. Simoens, H. Wymeersch, and M. Moeneclaey, "Design and analysis of linear precoders for bit-interleaved coded modulation with iterative decoding," in *4th International Symposium on Turbo Codes & Related Topics in connection with the 6th International ITG-Conference on Source and Channel Coding (SCC)*, Apr. 2006.

[140] X. Qi, M. Zhao, S. Zhou, and J. Wang, "Multidimensional modulation used in BICM-ID," *IEE Electronics Letters*, vol. 41, no. 3, pp. 140–142, Feb. 2005.

[141] V. A. Alonso, "Closed-form expressions for the transfer characteristic of a QAM demapper," Master's thesis, Institut für Nachrichtenübertragung, Universität Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, June 2008, Number 2260M.

[142] M. Tüchler and J. Hagenauer, "EXIT charts of irregular codes," in *Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, USA, Mar. 2002, pp. 748–753.

[143] F. Schreckenbach and G. Bauch, "Bit-interleaved coded irregular modulation," *Europ. Trans. Telecom.*, vol. 17, no. 2, pp. 269–282, Mar./Apr. 2006.

[144] N. S. Muhammad, "Short inner block codes for serial concatenated codes and bit-interleaved coded modulation," in *IEEE Vehicular Technology Conference (VTC)*, vol. 70, Sept. 2009.

[145] T. Helleseth, T. Kløve, and V. I. Levenshtein, "On the information function of an error-correcting code," *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 549–557, Mar. 1997.

[146] B. Kurkoski, P. Siegel, and J. Wolf, "Exact probability of erasure and a decoding algorithm for convolutional codes on the binary erasure channel," in *Global Telecommunications Conference GLOBECOM*, vol. 3, Dec. 2003, pp. 1741–1745.

[147] E. Çela, *The Quadratic Assignment Problem: Theory And Algorithms*. Kluwer Academic Publishers, 1998, ISBN 0-792-34878-8.

[148] P. Gill, W. Murray, and M. Wright, *Practical Optimization*. London: Academic Press, 1981, ISBN 0-122-83952-8.

[149] M. Kießling, "Statistical analysis and transmit prefiltering for MIMO wireless systems in correlated fading environments," Ph.D. dissertation, Universität Stuttgart, Institut für Nachrichtenübertragung, Pfaffenwaldring 47, 70569 Stuttgart, Germany, Oct. 2004, ISBN 3-8322-3444-6.

[150] J. Sun, O. Y. Takeshita, and M. Fitz, "A highly parallel decoder for turbo codes," in *IEEE International Symposium on Information Theory*, July 2002, p. 228.