EINDHOVEN UNIVERSITY OF TECHNOLOGY

Information-theoretic analysis of a family of additive energy channels

Citation for published version (APA):
Martinez Vicente, A. (2008). *Information-theoretic analysis of a family of additive energy channels*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Electrical Engineering]. Technische Universiteit Eindhoven. https://doi.org/10.6100/IR632385

DOI:
10.6100/IR632385

Document status and date:
Published: 01/01/2008

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 16. Nov. 2023

Information-theoretic Analysis of a Family of Additive Energy Channels

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
Rector Magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op maandag 28 januari 2008 om 16.00 uur

door

Alfonso Martinez Vicente

geboren te Zaragoza, Spanje

Dit proefschrift is goedgekeurd door de promotor:

prof.dr.ir. J.W.M. Bergmans

Copromotor:
dr.ir. F.M.J. Willems

The work described in this thesis was financially supported by the Freeband Impulse Program of the Technology Foundation STW.

© 2008 by Alfonso Martinez. All rights reserved.

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

Martinez, Alfonso

Information-theoretic analysis of a family of additive energy channels / by Alfonso Martinez. - Eindhoven : Technische Universiteit Eindhoven, 2008.
Proefschrift. - ISBN 978-90-386-1754-1
NUR 959
Trefw.: informatietheorie / digitale modulatie / optische telecommunicatie.
Subject headings: information theory / digital communication / optical communication.

Ever tried. Ever failed. No matter.
Try Again. Fail again. Fail better.

Samuel Beckett

Summary

Information-theoretic analysis of a family of additive energy channels

This dissertation studies a new family of channel models for non-coherent communications, the additive energy channels. By construction, the additive energy channels occupy an intermediate region between two widely used channel models: the discrete-time Gaussian channel, used to represent coherent communication systems operating at radio and microwave frequencies, and the discrete-time Poisson channel, which often appears in the analysis of intensity-modulated systems working at optical frequencies. The additive energy channels share with the Gaussian channel the additivity between a useful signal and a noise component. However, the signal and noise components are not complex-valued quadrature amplitudes but, as in the Poisson channel, non-negative real numbers, the energy or squared modulus of the complex amplitude.

The additive energy channels come in two variants, depending on whether the channel output is discrete or continuous. In the former case, the energy is a multiple of a fundamental unit, the quantum of energy, whereas in the second the value of the energy can take on any non-negative real number. For continuous output the additive noise has an exponential density, as for the energy of a sample of complex Gaussian noise. For discrete, or quantized, energy the signal component is randomly distributed according to a Poisson distribution whose mean is the signal energy of the corresponding Gaussian channel; part of the total noise at the channel output is thus a signal-dependent, Poisson noise component. Moreover, the additive noise has a geometric distribution, the discrete counterpart of the exponential density.

Contrary to the common engineering wisdom that not using the quadrature amplitude incurs in a significant performance penalty, it is shown in this dissertation that the capacity of the additive energy channels essentially coincides

with that of a coherent Gaussian model under a broad set of circumstances. Moreover, common modulation and coding techniques for the Gaussian channel often admit a natural extension to the additive energy channels, and their performance frequently parallels those of the Gaussian channel methods.

Four information-theoretic quantities, covering both theoretical and practical aspects of the reliable transmission of information, are studied: the channel capacity, the minimum energy per bit, the constrained capacity when a given digital modulation format is used, and the pairwise error probability. Of these quantities, the channel capacity sets a fundamental limit on the transmission capabilities of the channel but is sometimes difficult to determine. The minimum energy per bit (or its inverse, the capacity per unit cost), on the other hand, turns out to be easier to determine, and may be used to analyze the performance of systems operating at low levels of signal energy. Closer to a practical figure of merit is the constrained capacity, which estimates the largest amount of information which can be transmitted by using a specific digital modulation format. Its study is complemented by the computation of the pairwise error probability, an effective tool to estimate the performance of practical coded communication systems.

Regarding the channel capacity, the capacity of the continuous additive energy channel is found to coincide with that of a Gaussian channel with identical signal-to-noise ratio. Also, an upper bound —the tightest known— to the capacity of the discrete-time Poisson channel is derived. The capacity of the quantized additive energy channel is shown to have two distinct functional forms: if additive noise is dominant, the capacity is close to that of the continuous channel with the same energy and noise levels; when Poisson noise prevails, the capacity is similar to that of a discrete-time Poisson channel, with no additive noise. An analogy with radiation channels of an arbitrary frequency, for which the quanta of energy are photons, is presented. Additive noise is found to be dominant when frequency is low and, simultaneously, the signal-to-noise ratio lies below a threshold; the value of this threshold is well approximated by the expected number of quanta of additive noise.

As for the minimum energy per nat (1 nat is $\log_2 e$ bits, or about 1.4427 bits), it equals the average energy of the additive noise component for all the studied channel models. A similar result was previously known to hold for two particular cases, namely the discrete-time Gaussian and Poisson channels.

An extension of digital modulation methods from the Gaussian channels to the additive energy channel is presented, and their constrained capacity determined. Special attention is paid to their asymptotic form of the capacity at low and high levels of signal energy. In contrast to the behaviour in the

Gaussian channel, arbitrary modulation formats do not achieve the minimum energy per bit at low signal energy. Analytic expressions for the constrained capacity at low signal energy levels are provided. In the high-energy limit simple pulse-energy modulations, which achieve a larger constrained capacity than their counterparts for the Gaussian channel, are presented.

As a final element, the error probability of binary channel codes in the additive energy channels is studied by analyzing the pairwise error probability, the probability of wrong decision between two alternative binary codewords. Saddlepoint approximations to the pairwise error probability are given, both for binary modulation and for bit-interleaved coded modulation, a simple and efficient method to use binary codes with non-binary modulations. The methods yield new simple approximations to the error probability in the fading Gaussian channel. The error rates in the continuous additive energy channel are close to those of coherent transmission at identical signal-to-noise ratio. Constellations minimizing the pairwise error probability in the additive energy channels are presented, and their form compared to that of the constellations which maximize the constrained capacity at high signal energy levels.

Samenvatting

In dit proefschrift wordt een nieuwe familie van kanaalmodellen voor niet-coherente communicatie onderzocht. Deze kanalen, aangeduid als additieve-energie kanalen, ontlenen kenmerken aan twee veelvuldig toegepaste modellen. Dit is enerzijds het discrete-tijd Gaussische kanaal, dat als model dient voor systemen met coherente communicatie op radio- en microgolf-frequenties, en anderzijds het discrete-tijd Poisson kanaal, dat doorgaans wordt gebruikt in de analyse van optische communicatiesystemen gebaseerd op intensiteit-modulatie. Zoals in het Gaussische kanaal, is de uitgang van een additieve-energie kanaal de som van het ingangs-signaal en additieve ruis. De waarden die deze uitgang kan aannemen zijn echter niet complex zoals de phasor van een electromagnetisch veld, maar niet-negatief reëel overeenkomstig de veldenergie.

Bij additieve-energie kanalen kan onderscheid worden gemaakt tussen kanalen met continue en kanalen met discrete energie. Als de energie continu is, heeft de additieve ruis een exponentiële verdeling, zoals de amplitude van circulair symmetrische complexe Gaussische ruis. Bij discrete energie is de kanaaluitgang een aantal energiekwanta. In dit geval is de signaalterm een stochast, verdeeld als een Poisson variabele waarvan de gemiddelde waarde gelijk is aan het aantal energiekwanta in het equivalente continue-energie model. Als gevolg hiervan komt een deel van de ruis (Poisson ruis) uit het signaal zelf. Verder heeft de ruisterm een geometrische verdeling, de discrete versie van een exponentiële verdeling.

In dit proefschrift wordt aangetoond dat additieve-energie kanalen vaak even goed presteren als het coherente Gaussische kanaal. In tegenstelling tot wat vaak wordt verondersteld, leidt niet-coherente communicatie niet tot een substantiëel verlies. Deze conclusie wordt gestaafd door bestudering van vier informatie-theoretische grootheden: de kanaalcapaciteit, de minimum energie per bit, de capaciteit voor algemene digitale modulaties en de paarsgewijze

foutenkans. De twee eerstgenoemde grootheden zijn overwegend theoretisch van aard en bepalen de grenzen voor de informatieoverdracht in het kanaal, terwijl de twee overige een meer praktisch karakter hebben.

Een eerste bevinding van het onderzoek is dat de capaciteit van het continue additieve-energie kanaal gelijk is aan de capaciteit van een Gaussisch kanaal met identieke signaal-ruis verhouding. Daarnaast wordt een nieuwe bovengrens afgeleid voor de capaciteit van het discrete-tijd Poisson kanaal. Voor de capaciteit van het additieve-energie kanaal met discrete energie bestaan twee limiet-uitdrukkingen. De capaciteit kan benaderd worden door de capaciteit van een kanaal met exponentiële ruis bij lage signaal-ruis verhouding, m. a. w. als geometrische ruis groter is in verwachting dan de Poisson ruis. Vanaf een bepaalde waarde van de Poisson ruis verwachting is daarentegen de capaciteit van een Poisson kanaal zonder geometrische ruis een goede benadering. Toepassing van het bovenstaande model op elektromagnetische straling, waarbij de kwanta fotonen zijn, leidt tot een formule voor de drempel in de signaal-ruis verhouding als functie van de temperatuur en de frequentie. Voor de gebruikelijke radio- en microgolf-frequenties ligt deze drempel ruimschoots boven de signaal-ruis verhouding van bestaande communicatiesystemen.

De minimum energie per nat is gelijk aan de gemiddelde waarde van de additieve ruis. Bij afwezigheid van additieve ruis is de minimum energie per bit oneindig, net als bij het Poisson kanaal.

De analyse van digitale puls-energie modulaties is gebaseerd op de "constrained capacity", de hoogste informatie rate die gerealiseerd kan worden met deze modulaties. Anders dan in het Gaussische kanaal, halen puls-energie modulaties in het algemeen niet de minimum energie per nat. Voor hoge energie zijn deze modulaties echter potentiëel efficiënter dan vergelijkbare kwadratuur amplitude modulaties voor het Gaussische kanaal.

Tot slot wordt de foutenkans van binaire codes geanalyseerd met behulp van een zadelpunt-benadering voor de paarsgewijze foutenkans, de kans op een foutieve beslissing tussen twee codewoorden. Onze analyse introduceert nieuwe en effectieve benaderingen voor deze foutenkans voor het Gaussische kanaal met fading. Zoals eerder ook met de capaciteit het geval was, is de foutenkans van binaire codes voor additieve-energie kanalen vergelijkbaar met die van dezelfde codes voor het Gaussische kanaal. Tenslotte is ook bit-interleaved coded-modulation voor additieve-energie kanalen bestudeerd. Deze modulatiemethode maakt op een eenvoudige en effectieve wijze gebruik van binaire codes in combinatie met niet-binaire modulaties. Het blijkt dat modulatie en coderingen voor het Gaussische kanaal vaak vergelijkbare prestaties leveren als soortgelijke methoden voor additieve-energie kanalen.

Acknowledgements

First of all, I would like to express my sincere gratitude to Jan Bergmans and Frans Willems for their advice, teaching, help, encouragement, and patience over the past few years in Eindhoven. From them I have received the supreme forms of support: freedom to do research and confidence on my judgement.

This thesis would have never been possible without some of my former colleagues at the European Space Agency, especially Riccardo de Gaudenzi, who directly and indirectly taught me so many things, and Albert Guillén i Fàbregas, who has kept in touch throughout the years. Through them, I have also had the pleasure of working with Giuseppe Caire. Their contribution to the analysis of the Gaussian channel has been invaluable.

I am indebted to the members of the thesis defense committee, Profs. Ton Koonen, Emre Telatar, Edward van der Meulen, and Sergio Verdú, for their presence in the defense and for their comments on the draft thesis.

I am grateful to the many people with whom I have worked the past few years in Eindhoven. In the framework of the mode-group diversity multiplexing project, I have enjoyed working with Ton Koonen and Henrie van den Boom, and also with Marcel, Helmi, and Pablo. In our signal processing group, I have found it enlightening to discuss our respective research topics with Emanuël, Jakob, Hongming, Chin Keong, and Jamal. I particularly appreciate the help from Dr. Hennie ter Morsche in the asymptotic analysis reported in section 3.C.

A special acknowledgement goes to Christos Tsekrekos and María García Larrodé, who have very often listened with attention to my ideas and given their constructive comments. Far from Eindhoven, Josep Maria Perdigués Armengol has always been willing to answer my questions on quantum communications.

And last, but by no means least, I wish to thank my parents and Tomas for the precious gift of their unconditional love. I am glad they are there.

Glossary

A_d	Number of codewords at Hamming distance d
AE-Q	Quantized additive energy channel
AEN	Additive exponential noise channel
APSK	Amplitude and phase-shift keying modulation
AWGN	Additive white Gaussian noise channel
$\mathbf{b}$	Binary codeword
b	Bit (in a codeword)
BICM	Bit-interleaved coded modulation
BNR	Bit-energy-to-noise ratio
$\mathrm{BNR}_{\min}$	Minimum bit-energy-to-noise ratio
BNR_0	Bit-energy-to-noise ratio at zero capacity
BPSK	Binary phase-shift keying modulation
C	Channel capacity, in bits(nats)/channel use
$\mathrm{C}_{\mathcal{X},\mu}$	Bit-interleaved coded modulation capacity
$\mathrm{C}_{\mathcal{X}}$	Coded modulation ("constrained") capacity
$\mathrm{C}_{\mathcal{X}}^{\mathrm{u}}$	Coded modulation uniform capacity
$\mathrm{C}(E)$	Channel capacity at energy E
$\mathrm{C}_{\mathrm{G}}(\varepsilon_s, \varepsilon_n)$	AE-Q channel capacity bound in G regime
$\mathrm{C}_{\mathrm{P}}(\varepsilon_s)$	AE-Q channel capacity bound in P regime
c_1	First-order Taylor coefficient in capacity expansion
c_2	Second-order Taylor coefficient in capacity expansion
C_1	Capacity per unit energy
$D(\cdot\|\|\cdot)$	Divergence between two probability distributions
d	Hamming distance between two binary codewords
ΔP	Power expansion ratio
ΔW	Bandwidth expansion ratio

DTP	Discrete-time Poisson channel
$\mathrm{E}[\cdot]$	Expectation of a random variable
ε_b	Bit energy (DTP, AE-Q channels)
$E_{b,\min}$	Minimum energy per bit
$\varepsilon_{b,\min}$	Minimum bit energy (DTP, AE-Q channels)
ε_{b0}	Bit-energy at zero capacity (DTP, AE-Q channels)
$\varepsilon(\cdot)$	Energy of a symbol or sequence
E_n	Average noise energy (AEN)
ε_n	Average noise (AE-Q)
$\mathcal{E}(\varepsilon)$	Exponential random variable of mean ε
E_s	Energy constraint (AWGN, AEN)
ε_s	Energy constraint (DTP, AE-Q)
η	Spectral efficiency, in bits/sec/Hz
ε_0	Energy of a quantum
$\mathcal{G}(\varepsilon_s, \nu)$	Gamma distribution with parameters ε_s and ν
γ_{e}	Euler's constant, $0.5772\ldots$
$\mathcal{G}(\varepsilon)$	Geometric random variable of mean ε
h	Planck's constant
$H_{\mathrm{Exp}}(\varepsilon)$	Differential entropy of $\mathcal{E}(\varepsilon)$
$H_{\mathrm{Geom}}(\varepsilon)$	Differential entropy of $\mathcal{G}(\varepsilon)$
$H_{\mathcal{N}_{\mathbf{C}}}(\sigma_0^2)$	Differential entropy of $\mathcal{N}_{\mathbf{C}}(\mu, \sigma^2)$
$H_{\mathrm{Pois}}(x)$	Entropy of a Poisson distribution with mean x
$H(X)$	Entropy (or differential entropy) of X
$H(Y\|X)$	Conditional (differential) entropy of X given Y
$I(X;Y)$	Mutual information between variables X and Y
$\kappa_1(r)$	Cumulant transform of bit score
$\kappa_{\mathrm{pw}}(r)$	Cumulant transform of pairwise score
k_B	Boltzmann's constant
λ_i	Log-likelihood ratio
λ	PEM constellation parameter
$\mathrm{mgf}(r)$	Moment gen. function of X, $\mathrm{E}[e^{rX}]$
m	$\log_2 \|\mathcal{X}\|$, for modulation set $\mathcal{X}$
m_f	Nakagami/gamma fading factor
$\mu_1(\mathcal{X})$	First-order moment of constellation $\mathcal{X}$
$\mu_2(\mathcal{X})$	Second-order moment of constellation $\mathcal{X}$
$\mu_{2'}(\mathcal{X})$	Pseudo second-order moment of constellation $\mathcal{X}$
n	Length of transmitted/received sequence
N_0	Noise spectral density

$\mathcal{N}_{\mathbf{C}}(0, \sigma^2)$	Circularly-symmetric complex Gaussian variable
ν	Frequency
P	Average received power
PAM	Pulse-amplitude modulation
$P_{S\|X}(s\|x)$	Signal channel output conditional probability
$P_{Y\|X}(y\|x)$	Channel output conditional probability
$p_{S\|X}(s\|x)$	Signal channel output conditional density
$p_{Y\|X}(y\|x)$	Channel output conditional density
P_b	Bit error rate
PEM	Pulse-energy modulation
$\mathrm{pep}(d)$	Pairwise error probability (Hamming distance d)
P_w	Word error rate
$\mathrm{pgf}(u)$	Probability gen. function of discrete X, $\mathrm{E}[u^X]$
$\Pr(\cdot)$	Probability of an event
$\mathcal{P}(\varepsilon)$	Poisson random variable of mean ε
PSK	Phase-shift keying modulation
$P_X(\cdot)$	Input distribution
QAM	Quadrature-amplitude modulation
$q(x, y)$	Symbol decoding metric
$q_i(b, y)$	Bit decoding metric at position i
$Q(y\|x)$	Channel transition matrix
$Q_i(y\|b)$	Channel transition prob. at bit position i
QPSK	Quaternary phase-shift keying modulation
R	Transmission data rate, in bits(nats)/channel use
$\hat{r}$	Saddlepoint for tail probability
$\sigma^2(\mathcal{X})$	Variance of constellation $\mathcal{X}$
$\hat{\sigma}^2(\mathcal{X})$	Pseudo variance of constellation $\mathcal{X}$
σ^2	Average noise energy (AWGN)
s_k	Discrete-time received signal
SNR	Signal-to-noise ratio
T_0	Ambient temperature
$u(t)$	Step function
$\mathrm{Var}(\cdot)$	Variance of a random variable
W	Bandwidth (in Hz)
W_{eff}	Total effective bandwidth (in Hz)
W_s	Spatial bandwidth; number of degrees of freedom
W_t	Temporal bandwidth (in Hz)
w	Index of transmitted message

$\lvert\mathcal{W}\rvert$	Cardinality of the set of messages w
$\hat{w}$	Index of estimated message at receiver
$\mathbf{x}$	Sequence of transmitted symbols
$\mathcal{X}$	Alphabet of transmitted symbols
$\mathcal{X}_i^b$	Set of symbols with bit b in i-th label
$\mathcal{X}_\lambda$	PEM constellation of parameter λ
$\mathcal{X}_\lambda^\infty$	Continuous PEM of parameter λ
$X_{\mathcal{E}(\varepsilon_s)}$	Input distributed as $\mathcal{E}(\varepsilon_s)$
$X_{\mathcal{G}(\varepsilon_s,\nu)}$	Input distributed as $\mathcal{G}(\varepsilon_s,\nu)$
Ξ_b	Decoder decision bit score
Ξ_{pw}	Decoder decision pairwise score
x_k	Discrete-time received signal
$\mathbf{y}$	Sequence of received symbols
$\mathcal{Y}$	Alphabet of received symbols
y_k	Discrete-time received signal
z_k	Discrete-time (additive) noise

Contents

1

Introduction

1.1 Coherent Transmission in Wireless Communications

One of the most remarkable social developments in the past century has been the enormous growth in the use of telecommunications. In a process sparked by the telegraph in the 19th century, followed by Marconi's invention of the radio, and proceeding through the telephone system and the communication satellites, towards the modern cellular networks and the Internet, the possibility of communication at a distance, for that is what telecommunication means, has changed the ways people live and work. Fuelling these changes, electrical engineers have spent large amounts of time and resources in better understanding the communication capabilities of their systems and in devising new alternatives with improved performance. Among the possible names, let us just mention three pioneers: Nyquist, Kotelnikov, and Shannon.

Harry Nyquist, as an engineer working at the Bell Labs in the early 20th century, identified bandwidth and noise as two key parameters that affect the efficiency of communications. He then went on to provide simple, yet accurate, tools to represent both of them. In the case of bandwidth, his name is associated with the sampling theorem, specifically with the statement that the number of independent pulses that may be sent per unit time through a telegraph or radio channel is limited to twice the bandwidth of the channel. As for noise, he studied thermal noise, present in all radio receivers, and derived the celebrated formula giving the noise spectral density N_0 as a function of the ambient temperature T_0 and the radio frequency ν,

$$N_0 = \frac{h\nu}{e^{\frac{h\nu}{k_B T_0}} - 1}, \tag{1.1}$$

where h and k_B are respectively Planck's and Boltzmann's constants. At radio

frequencies, $h\nu \ll k_B T_0$, and one recovers the well-known formula $N_0 \simeq k_B T_0$.

Vladimir Kotelnikov, working in the Soviet Union in the 1930's and 1940's, independently formulated the sampling theorem, complementing Nyquist's result with an interpolation formula that yields the original signal from the sample amplitudes. In addition, he extensively analysed the performance of communication systems in the presence of noise, in particular of Gaussian noise; in this context, he provided heuristic reasons to justify the Gaussianity of noise in the communication receiver, essentially by an invocation of the central limit theorem of probability theory.

Kotelnikov also pioneered the use of a geometric, or vector space, approach to model communication systems. More formally, consider a signal $y(t)$ at the input of a radio receiver, say one polarization of the electromagnetic field impinging in the receiving antenna. Often, the signal $y(t)$ is given by the sum of a useful signal component, $x(t)$, and an additive noise component, $z(t)$. In the geometric approach, the signal $y(t)$ is replaced by a vector of numbers y_k, each of whom is the projection of $y(t)$ onto the k-th coordinate of an underlying vector space. Since projection onto a basis is a linear operation, we have that

$$y_k = x_k + z_k, \tag{1.2}$$

where x_k and z_k respectively denote the useful signal and the noise components along the k-th coordinate. The resulting discrete-time model is the standard additive white Gaussian noise (AWGN) channel, where z_k are independent Gaussian random variables with identical variance. When the complex-valued quantities y_k are determined at the receiver, we talk of coherent signal detection. In physical terms, coherent detection corresponds to accurately estimating the frequency and the phase of the electromagnetic field.

Claude Shannon, another engineer employed at the Bell Labs, is possibly the most important figure in the field of communication theory. Among the many fundamental results in his well-known paper "A Mathematical Theory of Communication" [1], of special importance is his discovery of the existence of a quantity, the channel capacity, which determines the highest data rate at which reliable transmission of information over a channel is possible. In this context, reliably means with vanishing probability of wrong message detection at the receiving end of the communication link. For a radio channel of bandwidth W (in Hz) in additive white Gaussian noise of spectral density N_0 and with average received power P, the capacity C (in bits/second, or bps) equals

$$C = W \log_2\left(1 + \frac{P}{W N_0}\right). \tag{1.3}$$

In [1], Shannon expressed the channel capacity in terms of entropies of random variables and, using the fact that the Gaussian distribution has the largest entropy of all random variables with a given variance, he went on to prove that Gaussian noise is the worst additive noise, in the sense that other noise distributions with the same variance allow for a larger channel capacity. More recently, Lapidoth proved [2] that a system designed for the worst-case noise, namely maximum-entropy Gaussian noise, is likely to operate well under other noise distributions, thus providing a further engineering argument to the use of a Gaussian noise model.

As the capacity C is the maximum data rate at which reliable communication is possible, Eq. (1.3) provides guidance on the way to attain ever higher rates. Indeed, the evolution of telecommunications in the second half of the 20th century can be loosely described as a form of "conversation" with Eq. (1.3). Several landmarks in this historical evolution are shown in Fig. 1.1, along with the spectral efficiency η, given by $\eta = C/W$ (in bps/Hz), as a function of the signal-to-noise ratio SNR, defined as $\mathrm{SNR} = P/(WN_0)$.

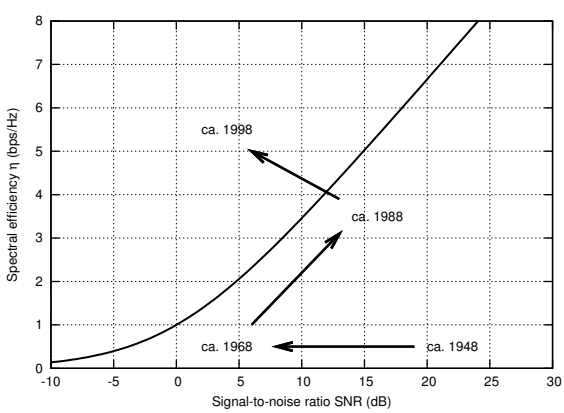


Figure 1.1: Trade-off between signal-to-noise ratio and spectral efficiency.

In the first radio systems, represented by the label 'ca. 1948' in Fig. 1.1, signal-to-noise ratios SNR of the order of 20-30 dB were typical, together with low spectral efficiencies, say $\eta \leq 0.5$ bps/Hz. After Shannon's analysis, channel codes were devised in the 1950's and 1960's for use in various communication systems, e. g. satellite transmission. These channel codes allow for a reduction of the required signal-to-noise ratio at no cost in spectral efficiency, as indicated

by the label 'ca. 1968' in Fig. 1.1. Later, around the 1980's, requirements for higher data rates, e. g. for telephone modems, led to the use of multi-level modulations, which trade increased power for higher data rates; a label 'ca. 1988' is placed at a typical operating point of such systems.

When it seemed that the whole space of feasible communications was covered, a new way forward was found. It was discovered that the total "bandwidth" W_{eff} available for communication, i. e. the total number of independent pulses that may be sent per unit time through a radio channel, ought to have two components, one spatial and one temporal [3]. Loosely speaking, $W_{\text{eff}} = W_t W_s$, where W_t, measured in Hz, is the quantity previously referred to as bandwidth, and W_s is the number of spatial degrees of freedom, typically related to the number of available antennas. In order to account for this effect, W should be replaced in Eq. (1.3) by W_{eff} and, consequently, the spectral efficiency η becomes

$$\eta = \frac{C}{W_t} = W_s \log_2\left(1 + \frac{P}{W_{\text{eff}} N_0}\right). \tag{1.4}$$

Exploitation of this "spatial bandwidth" is the underlying principle behind the use of multiple-antenna (MIMO) systems [4, 5] for wireless communications, where spectral efficiencies exceeding 5 bps/Hz are possible for a fixed signal-to-noise ratio, now defined as $\text{SNR} = P/(W_{\text{eff}} N_0)$; these values of the spectral efficiency are represented by the label 'ca. 1998' in Fig. 1.1.

Having sketched how communication engineers have exploited tools derived from the information-theoretic analysis of coherent detection to design efficient communication systems for radio and microwave frequencies, we next shift our attention to optical communications.

1.2 Intensity Modulation in Optical Communications

In parallel to the exploitation of the radio and microwave frequencies, much higher frequencies have also been put into use. One reason for this move is the fact that the available bandwidth becomes larger as the frequency increases. At optical frequencies, in particular, bandwidth is effectively unlimited. Moreover, at optical frequencies efficient "antennas" are available for the transmitting and receiving ends of the communication link, in the form of lasers and photodiodes, and an essentially lossless transmission medium, the optical fibre, exists. As a consequence, optical fibres, massively deployed in the past few decades, carry very high data rates, easily reaching hundreds of gigabits/second.

Most optical communication systems do not modulate the quadrature amplitude $x(t)$ of the electromagnetic field, but the instantaneous field intensity, defined as $|x(t)|^2$. The underlying reason for this choice is the difficulty of building oscillators at high frequencies with satisfactory phase stability properties. At the receiver side, coherent detection is often unfeasible, due to similar problems with the oscillator phase stability. Direct detection, based on the photoelectric effect, is frequently used as an alternative.

The photoelectric effect manifests itself as a random process with discrete output. More precisely, the measurement of a direct detection receiver over an interval $(0, T)$ is a random integer number (of photons, or quanta of light), distributed according to a Poisson distribution with mean υ. The mean υ, given by $\upsilon = \int_0^T |y(t)|^2 \, dt$, depends on the instantaneous squared modulus of the field at the receiver, $|y(t)|^2$, and is independent of the phase of the complex-valued amplitude $y(t)$. Since the variance of a Poisson random variable coincides with its mean υ, it is in general non-zero and there is a noise contribution arising from the signal itself. This noise contribution is called shot noise.

In information theory, a common model for optical communications systems with intensity modulation and direct detection is the Poisson channel, originally proposed by Bar-David [6]. A short, yet thorough, historical review by Verdú [7] lists the main contributions to the information-theoretic analysis of the Poisson channel. The input to the Poisson channel is a continuous-time signal, subject to a constraint on the peak and average value. The input is usually denoted by $\lambda(t)$, which corresponds to an instantaneous field intensity, i. e. $\lambda(t) = |y(t)|^2$, in our previous notation. In an arbitrary interval $(0, T)$, the output is a random variable distributed according to a Poisson distribution with mean $\upsilon = \int_0^T \lambda(t) \, dt$. As found by Wyner [8], the capacity of the Poisson channel is approached by functions $\lambda(t)$ whose variation rate grows unbounded. Practical constraints on the variation rate of $\lambda(t)$ may be included by assuming that the input signal is piecewise constant [9], in which case the Poisson channel is naturally represented by a discrete-time channel model, whose output y_k has a Poisson distribution of the appropriate mean.

Next to the Poisson channel models, physicists have also independently studied various channel models for communication at optical frequencies; for a relatively recent review, see the paper by Caves [10]. In particular, the design and performance of receivers for optical coherent detection has been considered. In this case, it is worthwhile remarking that direct application of Eq. (1.3) with N_0 given by the corresponding value of Eq. (1.1) at optical frequencies, proves problematic since $N_0 \simeq 0$, which would indicate an infinite

capacity. Phenomena absent in the models for radio communication, must be taken into account. A good overview of such phenomena is given by Oliver [11].

Under different assumptions on signal, noise, and/or detection method, different models and therefore different values for the channel capacity are obtained [10, 12–15]. To any extent, and regardless of the precise value of the channel capacity at optical frequencies, it is safe to state that deployed optical-fibre communications systems are, qualitatively, somehow still around their equivalent of 'ca. 1948' in Fig. 1.1. Designers have not yet pushed towards the ultimate capacity limit, in contrast to the situation in wireless communications which was sketched during the discussion on Fig. 1.1. Both modulation (binary on-off keying) and multiplexing methods (wavelength division multiplexing) have remained remarkably constant along the years. Nevertheless, and in anticipation of future needs for increased spectral efficiency, research has been conducted on channel codes [16] —corresponding roughly to 'ca. 1968'—, modulation techniques [17], multi-level modulations [18] —for 'ca. 1988'—, or multiple-laser methods [19, 20] —as the techniques in 'ca. 1998'—.

A common thread of the lines of research listed in the previous paragraph is the extension of techniques common to radio frequencies to optical frequencies. In a similar vein, it may also prove fruitful to extend to optical frequencies some key features of the models used for radio frequencies. One such key feature is the presence of additive maximum-entropy Gaussian noise at the channel output; as we previously mentioned, Shannon proved that this noise distribution allows for the lowest possible channel capacity when the signal is power constrained. For other channel models, a similar role could be played by the corresponding maximum-entropy distribution; this is indeed the case for non-negative output and exponential additive noise, as found by Verdú [21]. This observation suggests an extension of the discrete-time Poisson channel so as to include maximum-entropy additive noise. Such a channel model includes two key traits of the Poisson channel, namely non-negativity of the input signal and the quantized nature of the channel output and adds the new feature of a maximum-entropy additive noise. In the next section we incorporate these elements into the definition of the additive energy channels.

1.3 The Additive Energy Channels

The family of additive energy channels occupies an intermediate region between the discrete-time Poisson channel and the discrete-time Gaussian channel. As in the Poisson channel, communication in the additive energy channels is non-

coherent, in the sense that the signal and noise components are not represented by a complex-valued quadrature amplitude, but rather by a non-negative number, which can be identified with the squared modulus of the quadrature amplitude. From the Gaussian channel, the additive energy channels inherit the properties of discreteness in time and of additivity between a useful signal and a noise component, drawn according to a maximum-entropy distribution.

In analogy with Eq. (1.2), the k-th channel output, denoted by y'_k, is given by the sum of a useful signal x'_k and a noise component z'_k, that is

$$y'_k = x'_k + z'_k. \tag{1.5}$$

In general, we refer to additive energy channels, in plural, since there are two distinct variants, depending on whether the output is continuous or discrete.

When the output is discrete, the energy is a multiple of a quantum of energy of value ε_0. The useful signal component x'_k is now a random variable with a Poisson distribution of mean $|x_k|^2$, say the signal energy in the k-th component of an AWGN channel. The additive noise component z'_k is distributed according to a geometric (also called Bose-Einstein) distribution, which has the largest entropy of all distributions for discrete, non-negative random variables subject to a fixed mean value [22].

For continuous output, the value of the signal (resp. additive noise) component x'_k (resp. z'_k) coincides with the signal energy in the k-th coordinate of an AWGN channel, i. e. $x'_k = |x_k|^2$ (resp. $z'_k = |z_k|^2$), a non-negative number. The noise z'_k follows an exponential distribution, since z_k is Gaussian distributed. The exponential density is the natural continuous counterpart of the geometric distribution and also has the largest entropy among the densities for continuous, non-negative random variables with a constraint on the mean [22]. The channel model with continuous output is an additive exponential noise channel, studied in a different context by Verdú [21]. The continuous-output channel model may also be derived from the discrete-output model by letting the number of quanta grow unbounded, simultaneously keeping fixed the total energy. Equivalently, the energy of a single quantum ε_0 may be let go to zero while the total average energy is kept constant.

The additive energy channel models are different from the most common information-theoretic model for non-coherent detection, obtained by replacing the AWGN output signal y_k by its squared modulus (see e. g. the recent study [23] and references therein). The channel output, now denoted by y''_k, is then

$$y''_k = |y_k|^2 = |x_k + z_k|^2. \tag{1.6}$$

By construction, the output y_k'' conditional on x_k follows a non-central chi-square distribution and is therefore not the sum of the energies of x_k and z_k, i. e. $y_k'' \neq |x_k|^2 + |z_k|^2$.

The main contribution of this dissertation is the information-theoretic analysis of the additive energy channels. We will see that, under a broad set of circumstances, the information rates and error probabilities in the additive energy channels are very close to those attained in the Gaussian channel with the same signal-to-noise ratio. Somewhat surprisingly, the performance of direct detection turns out to be close to that of coherent detection, where we have borrowed terminology from optical communications. In Section 1.4, we outline the main elements of this analysis, as a preview of the dissertation itself.

1.4 Outline of the Dissertation

In this dissertation, we analyze a family of additive energy channels from the point of view of information theory. Since these channels are mathematically similar to the Gaussian channel, we find it convenient to apply tools and techniques originally devised for Gaussian channels, with the necessary adaptations wherever appropriate. In some cases, the adaptations shed some new light on the results for the Gaussian channel, in which case we also discuss at some length the corresponding results.

In Chapter 2, we formally describe the additive energy channels, both for continuous output (additive exponential noise channel) and for discrete output (quantized additive energy channel). In the latter, the analysis is carried out in terms of quanta, with a brief application at the end of the chapter of the results to the case where the quanta of energy are photons of an arbitrary frequency.

Four information-theoretic quantities, covering both theoretical and practical aspects of the reliable transmission of information, are studied: the channel capacity, the constrained capacity when a given digital modulation format is used, the minimum energy per bit, and the pairwise error probability.

As we stated before Eq. (1.3), the channel capacity gives the fundamental limit on the transmission capabilities of a channel. More precisely, the capacity is the highest data rate at which reliable transmission of information over a channel is possible. In Chapter 3, the channel capacity of the additive energy channels is determined. The capacity of the continuous additive energy channel is shown to coincide with that of a Gaussian channel with identical signal-to-noise ratio. Then, an upper bound —the tightest known to date— to the capacity of the discrete-time Poisson channel is obtained by applying a

method recently used by Lapidoth [24] to derive upper bounds to the capacity of arbitrary channels. The capacity of the quantized additive energy channel is shown to have two distinct functional forms: if additive noise is dominant, the capacity approaches that of the continuous channel with the same energy and noise levels; when Poisson noise prevails, the capacity is similar to that of a discrete-time Poisson channel with no additive noise.

An analogy with radiation channels of an arbitrary frequency, for which the quanta of energy are photons, is presented. Additive noise is found to be dominant when frequency is low and, simultaneously, the signal-to-noise ratio lies below a threshold; the value of this threshold is well approximated by the expected number of quanta of additive noise.

Unfortunately, the capacity is often difficult to compute and knowing its value does not necessarily lead to practical, workable methods to approach it. On the other hand, the minimum energy per bit (or its inverse, the capacity per unit cost) turns out to be easier to determine and further proves useful in the performance analysis of systems working at low levels of signal energy, a common operating condition. Even closer to a practical figure of merit is the constrained capacity, which estimates the largest amount of information which can be transmitted by using a specific digital modulation format. In Chapter 4, we cover coded modulation methods for the Gaussian channel, with particular emphasis laid on the performance at low signal-to-noise ratios, the so-called wideband regime, of renewed interest in the past few years after an important paper by Verdú [25]. Some new results on the characterization of the wideband regime are presented. The discussion is complemented by an analysis of bit-interleaved coded modulation, a simple and efficient method proposed by Caire [26] to use binary codes with non-binary modulations.

In Chapter 5, an extension of digital modulation methods from the Gaussian channels to the additive energy channel is presented, and their corresponding constrained capacity when used at the channel input determined. Special attention is paid to the asymptotic form of the capacity at low and high levels of signal energy. In the low-energy region, our work complements previous work by Prelov and van der Meulen [27, 28], who considered a general discrete-time additive channel model, and determined the asymptotic Taylor expansion at zero signal-to-noise ratio, in that the additive energy channels are constrained on the mean value of the input, rather than the variance, and similarly the noise is described by its mean, not its variance; the models considered by Prelov and van der Meulen rather deal with channels where the second-order moments, both for signal energy and noise level, are of importance. Our work extends their analysis to the family of additive energy channels, where the first-order

moments are constrained. In the high-energy limit, simple pulse-energy modulations are presented which achieve a larger constrained capacity than their counterparts for the Gaussian channel.

In addition, techniques devised by Verdú [29] to compute the capacity per unit cost are exploited to determine the minimum energy per nat (recall that $1\,\text{nat} = \log_2 e\,\text{bits}$, or about $1.4427\,\text{bits}$), which is found to equal the average energy of the additive noise component for all the channel models we study. We note here that this result was known to hold in two particular cases, namely the discrete-time Gaussian and Poisson channels [29, 30].

We complement our study of the constrained capacity by the computation of the pairwise error probability, an important tool to estimate the performance of practical binary codes used in conjunction with digital modulations. In Chapter 6, the error probability of binary channel codes in the additive energy channels is studied. Saddlepoint approximations to the pairwise error probability are given, both for binary modulation and for bit-interleaved coded modulation. The methods yield new simple approximations to the error probability in the fading Gaussian channel. It is proved that the error rates in the continuous additive energy channel are close to those of the coherent transmission at identical signal-to-noise ratio. Finally, constellations minimizing the pairwise error probability in the additive energy channels are presented, and their form compared to that of the constellations which maximize the constrained capacity at high signal energy levels.

Concluding the dissertation, Chapter 7 contains a critical discussion of the main findings presented in the preceding chapters and sketches possible extensions and future lines of work.

2

The Additive Energy Channels

2.1 Introduction: The Communication Channel

In this dissertation we study the transmission of information across a communication channel from the point of view of information theory. As schematically depicted in Fig. 2.1, very similar to the diagram in Shannon's classical paper [1], information is transmitted by sending a message w, generated at the source of the communication link, to the receiving end. The meaning, form, or content of the message are not relevant for the communication problem, and only the number of different messages generated by the source is relevant. For convenience, we model the message w as an integer number.

The encoder transforms the message into an array of n symbols, which we denote by $\mathbf{x}$. The symbols in $\mathbf{x}$ are drawn from an alphabet $\mathcal{X}$, or set, that depends on the underlying channel. In this dissertation, symbols are either complex or non-negative real numbers, as is common practice for the modelling of, respectively, wireless radio and optical-fibre channels.

The symbol for encoder output $\mathbf{x}$ also stands for the communication channel input. The channel maps the array $\mathbf{x}$ onto another array $\mathbf{y}$ of n symbols, an array which is detected at the receiver. The channel is noisy, in the sense that $\mathbf{x}$ (and therefore w) may not be univocally recoverable from $\mathbf{y}$.

The decoder block generates a message estimate, $\hat{w}$, from $\mathbf{y}$ and delivers it

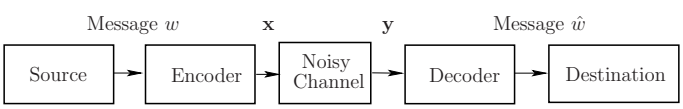


Figure 2.1: Generic communication link.

to the destination. The noisy nature of the communication channel causes the estimate $\hat{w}$ to possibly differ from the original message w. A natural problem is to make the probability of the estimate being wrong low enough, where the precise meaning of low enough depends on the circumstances and applications.

Information theory studies both theoretical and practical aspects of how to generate an estimate $\hat{w}$ very likely to coincide with the source message w. First, and through the concept of channel capacity, information theory gives an answer to the fundamental problem of how many messages can be reliably distinguished at the receiver side in the limit $n \to \infty$. Here reliably means that the probability that the receiver's estimate of the message, $\hat{w}$, differs from the original message at the source, w, is vanishingly small. In Chapter 3, we review the concept of channel capacity, and determine its value for the channel models described in this chapter.

Pairs of encoder and decoder which allow for the reliable transmission of the largest possible number of messages are said to achieve the channel capacity. In practice, simple yet suboptimal encoder and decoders are used. Information theory also provides tools to analyze the performance of these specific encoders and decoders. The performance of some encoder and decoder pairs for the models described in this chapter are covered in Chapters 4, 5, and 6.

Models with arrays as channel input and output naturally appear in the analysis of so-called waveform channels, for which functions of a continuous time variable t are transformed into a discrete array of numbers via an application of the sampling theorem or a Fourier decomposition. Details of this discretization can be found, for instance, in Chapter 8 of Gallager's book [31]. Since the time variable is discretized, these models often receive the name discrete-time, a naming convention we adopt.

In the remainder of this chapter, we present and discuss the channel models used in the dissertation. The various models are defined by the alphabet, or set, of possible channel inputs; the alphabet of possible channel outputs; and a probability density function $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ (for continuous output, if the output is discrete, a probability mass function $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ is used) on the set of outputs $\mathbf{y}$ for each input $\mathbf{x}$. We consider memoryless and stationary channels, for which $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ (resp. $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$) admits a decomposition

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^{n} p_{Y|X}(y_k|x_k), \tag{2.1}$$

where the symbols x_k and y_k are the k-th component of their respective arrays. The conditional density $p_{Y|X}(\cdot|\cdot)$ (resp. $P_{Y|X}(\cdot|\cdot)$) does not depend on the

value of k. An alternative name for the output conditional density is channel transition matrix, denoted by $Q(\cdot|\cdot)$. This name is common when the channel input and output alphabets are discrete.

We assume that one output symbol is produced for every input symbol, as depicted in Fig. 2.2 for the k-th component, or time k. The output y_k is the sum of two terms, the signal component s_k and the additive noise z_k, both of them taking values in the same alphabet as y_k (or possibly in a subset of the output alphabet). The probability law of z_k is independent of x_k. The channel output is

$$y_k = s_k(x_k) + z_k. \tag{2.2}$$

The signal component s_k is a function of the channel input x_k. The mapping $s_k(x_k)$ need not be deterministic, in which case it is described by a probability density function $p_{S|X}(s_k|x_k)$ ($P_{S|X}(s_k|x_k)$ for discrete output), common for all time indices k.

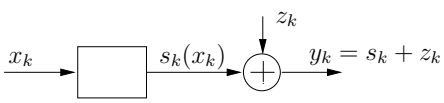


Figure 2.2: Channel operation at time k.

On Notation We agree that a symbol in small caps, u, refers to the numerical realization of the associated random variable, denoted by the capital letter U. Its probability density function, of total unit probability, is denoted by $p_U(u)$. Here U may stand for the encoder output X_k, the channel output Y_k, the noise realization Z_k, or a vector thereof. The input density may be a mixture of continuous and discrete components, in which case the density may include a number of Dirac delta functions. When the variable U takes values in a discrete set, we denote its probability mass function by $P_U(u)$.

Throughout the dissertation, integrals, Taylor series expansions, or series sums without explicit bibliographic reference can be found listed in [49] or may otherwise be computed by using Mathematica.

2.2 Complex-Valued Additive Gaussian Noise Channel

Arguably, the most widely analyzed, physically motivated, channel model is the discrete-time Additive White Gaussian Noise (AWGN) channel [31–33]. In

this case, the time components arise naturally from an application of the sampling theorem to a waveform channel, as well as in the context of a frequency decomposition of the channel into narrowband parallel subchannels.

We consider the complex-valued AWGN channel, whose channel input x_k and output y_k at time index $k = 1, \ldots, n$, are complex numbers related by

$$y_k = x_k + z_k. \tag{2.3}$$

In this case, the channel input x_k and its contribution to the channel output $s_k(x_k)$ coincide; they will differ for other channel models. The noise component z_k is drawn according to a circularly-symmetric complex Gaussian density of variance σ^2, a fact shorthanded to $Z_k \sim \mathcal{N}_{\mathbf{C}}(0, \sigma^2)$. The noise density is

$$p_Z(z) = \frac{1}{\pi\sigma^2} e^{-\frac{|z|^2}{\sigma^2}}. \tag{2.4}$$

The channel transition matrix is $Q(y|x) = p_Z(y - x)$.

We define the instantaneous (at time k) signal energy, denoted by $\varepsilon(x_k)$, as $|x_k|^2$. Similarly, the noise instantaneous energy $\varepsilon(z_k)$ is $\varepsilon(z_k) = |z_k|^2$. The average noise energy, where the averaging is performed over the possible realizations of z_k, is $\mathrm{E}[|Z_k|^2] = \mathrm{Var}(Z_k) = \sigma^2$.

The channel is used under a constraint on the total energy $\varepsilon(\mathbf{x})$, of the form

$$\varepsilon(\mathbf{x}) = \sum_{k=1}^{n} \varepsilon(x_k) = \sum_{k=1}^{n} |x_k|^2 \leq nE_s, \tag{2.5}$$

where E_s denotes the maximum permitted energy per channel use.

The average signal-to-noise ratio, denoted by SNR, is given by

$$\mathrm{SNR} = \frac{E_s}{\sigma^2}. \tag{2.6}$$

Here, and with no loss of generality, we assume that the constraint in Eq. (2.5) holds with equality. This step will be justified in Chapter 3, when we state how the channel capacity links the constraint on the total energy per message, as in Eq. (2.5), with a constraint on the average energy per channel use, or equivalently on the average signal-to-noise ratio.

It is common practice to replace the AWGN model given in Eqs. (2.3) by an alternative model whose input signal x'_k and channel output y'_k are respectively given by $x_k = \sqrt{E_s} x'_k$ and

$$y'_k = \frac{1}{\sigma} y_k = \sqrt{\mathrm{SNR}} x'_k + z'_k, \tag{2.7}$$

so that both signal x'_k and additive noise z'_k have unit average energy, i. e. $\mathrm{E}[|X'_k|^2] = 1$ and $Z'_k \sim \mathcal{N}_{\mathbf{C}}(0,1)$. The channel transition matrix is then

$$Q(y|x) = \frac{1}{\pi} e^{-|y-\sqrt{\mathrm{SNR}}x|^2}. \tag{2.8}$$

Both forms of the AWGN channel model are equivalent since they describe the same physical realization.

2.3 Additive Exponential Noise Channel

We next introduce the additive exponential noise (AEN) channel as a variation of an underlying AWGN channel. Since we use the symbols for the variables x_k, z_k, and y_k for both channels, we distinguish the AWGN variables by appending a prime.

In the AEN channel, the channel output is an energy, rather than a complex number as it was in the AWGN case. In the previous section, we defined the instantaneous energy of the signal and noise variables as the squared magnitude of the complex number x'_k or z'_k, and avoided referring to the energy of the output y'_k. The reason for this avoidance is that there are two natural definitions for the output energy.

First, in the AEN channel, the channel output y_k is defined to be the sum of the energies in x'_k and z'_k, that is, $y_k = x_k + z_k$, where the signal and noise components x_k and z_k are related to their Gaussian counterparts by

$$x_k = \varepsilon(x'_k) = |x'_k|^2, \quad z_k = \varepsilon(z'_k) = |z'_k|^2. \tag{2.9}$$

Figure 2.3 shows the relationship between the AWGN and AEN channels. We hasten to remark that this postulate is of a mathematical nature, possibly independent of any underlying physical model, since the quantity $\varepsilon(x'_k) + \varepsilon(z'_k)$ cannot be directly derived from y'_k only.

The inputs x_k and outputs y_k are non-negative real numbers, as befits an energy. The correspondence with the AWGN channel, $x_k = |x'_k|^2$, leads to a natural constraint on the average energy per channel use E_s, $\sum_{k=1}^{n} x_k \leq nE_s$, where E_s is the constraint in the AWGN channel.

The noise energy $z_k = \varepsilon(z'_k) = |z'_k|^2$, that is the squared amplitude of a circularly-symmetric complex Gaussian noise, has an exponential density [32] of mean $E_n = \sigma^2$,

$$p_Z(z) = \tfrac{1}{E_n} e^{-\frac{z}{E_n}} u(z), \tag{2.10}$$

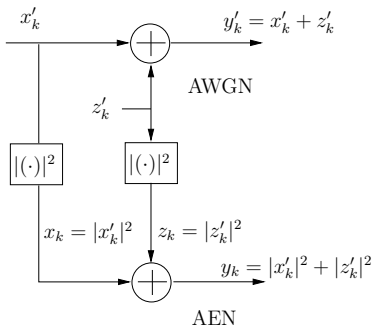


Figure 2.3: AEN model and its AWGN counterpart.

where $u(z)$ is a step function, $u(z) = 1$ for $z \geq 0$, and $u(z) = 0$ for $z < 0$. We use the notation $Z_k \sim \mathcal{E}(E_n)$, where the symbol $\mathcal{E}$ is used to denote an exponential distribution. The channel transition matrix has the form $Q(y|x) = p_Z(y - x)$.

The average signal-to-noise ratio, denoted by SNR, is given by

$$\mathrm{SNR} = \frac{\mathrm{E}\big[\varepsilon(X_k)\big]}{\mathrm{Var}^{\frac{1}{2}}(Z_k)} = \frac{E_s}{E_n}, \tag{2.11}$$

again assuming that the average energy constraint holds with equality. Here we used that the variance of an exponential random variable is E_n^2.

Even though the simplicity of this channel matches that of the AWGN channel, the AEN channel seems to have received limited attention in the literature. Exceptions are Verdú's work [21, 34] in the context of queueing theory, where exponential distributions often appear.

A second model derived from the AWGN channel is the non-coherent AWGN channel, see [23] and references therein. Its operation is depicted in Fig. 2.4. In this model, the output energy and channel output is $|y'_k|^2$, that is

$$|y'_k|^2 = |x'_k|^2 + |z'_k|^2 + 2\,\mathrm{Re}(x'^*_k z'_k). \tag{2.12}$$

The square root of the output, $|y'_k|$, is distributed according to the well-known Rician distribution [32]. In general, the value of $|y'_k|^2$ does not coincide with the sum of the energies $\varepsilon(x'_k) + \varepsilon(z'_k)$. By construction, the output energy in the AEN channel is the sum of the signal and noise energies. We say that the channel is an additive energy channel.

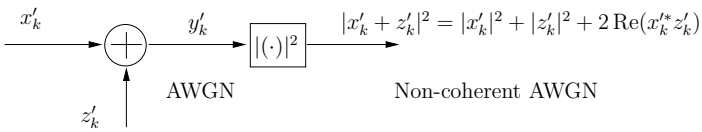


Figure 2.4: Non-coherent AWGN model.

In the AEN channel, multiplication of the output y_k by a real, non-zero factor α, leaves the model unchanged. The choice $\alpha = E_n^{-1}$, together with a new input x'_k, whose average energy is 1, leads to a new pair of signals x'_k and z'_k, with $\mathrm{E}[X'_k] = 1$ and $Z'_k \sim \mathcal{E}(1)$, such that the output becomes

$$y'_k = \mathrm{SNR}\, x'_k + z'_k. \tag{2.13}$$

Here the prime refers to the equivalent AEN channel, not to the Gaussian counterpart. As it happened for the AWGN channel, the channel capacity and the error rate performance achieved by specific encoder and decoder blocks coincide for either description of the AEN channel. We shall often use the model described in Eq. (2.13), whose channel transition matrix is given by

$$Q(y|x) = e^{-(y-\mathrm{SNR}\,x)} u(y - \mathrm{SNR}\,x). \tag{2.14}$$

We next consider a different additive energy channel model, whose channel output is a discrete number of quanta of energy. First, in the next section, we discuss the discrete-time Poisson channel, for which there is no additive noise component z_k, and then move on to the quantized additive energy channel.

2.4 Discrete-Time Poisson Channel

The discrete-time Poisson (DTP) channel appears naturally in the analysis of optical systems using the so-called pulse-amplitude modulation. Examples in the literature are papers by Shamai [9] and Brady and Verdú [35]. When ambient noise is negligible [12], it also models the bandlimited optical channel.

In the DTP model, the channel output is an energy, which comes in a discrete multiple of a quantum of energy ε_0. At time k the input is a non-negative real number $x_k \geq 0$, and we agree that the input energy is $x_k \varepsilon_0$; in optical communications the number x_k is the energy carried by an electromagnetic wave of

the appropriate frequency. The channel is used under a constraint on the (maximum) average number of quanta per channel use, ε_s, $\sum_{k=1}^{n} x_k \leq n\varepsilon_s$, with the understanding that the average energy per channel use E_s is $\varepsilon_s\varepsilon_0 = E_s$.

The channel output depends on the input x_k via a Poisson distribution with parameter x_k, that is, $Y_k = S_k \sim \mathcal{P}(x_k)$, where the symbol $\mathcal{P}$ is used to denote a Poisson distribution. Hence, the conditional output distribution is given by

$$P_{S|X}(s|x) = e^{-x}\frac{x^s}{s!}, \tag{2.15}$$

which also gives the channel transmission matrix $Q(y|x)$, with s replaced by y.

Since the channel output Y_k is a Poisson random variable, its variance is equal to x_k [36], $\mathrm{Var}(Y_k) = x_k$. Differently from AWGN or AEN channels, noise is now signal-dependent, coming from the signal term s_k itself. We refer to this noise as shot noise or Poisson noise.

As the number of quanta s_k becomes arbitrarily large for a fixed value of input energy $x_k\varepsilon_0$, the standard deviation of the channel output, of value $\sqrt{s_k}$, becomes negligible compared to its mean, s_k, and the density of the output energy $s_k\varepsilon_0$, viewed as a continuous random variable, approaches

$$p_{S|X}(s_k\varepsilon_0|x_k\varepsilon_0) = \lim_{\substack{\Delta x\to 0, x_k\to\infty \\ x_k\varepsilon_0 \text{fixed}}} \frac{1}{\Delta x}\Pr\left(x_k - \frac{\Delta x}{2} \leq s_k \leq x_k + \frac{\Delta x}{2}\right) \tag{2.16}$$

$$= \delta\big((s_k - x_k)\varepsilon_0\big), \tag{2.17}$$

i. e. a delta function, as for the signal energy in the AWGN and AEN channels, models for which there is no Poisson noise at the input.

2.5 Quantized Additive Energy Channel

The quantized additive energy channel (AE-Q) appears as a natural generalization of the discrete-time Poisson and the additive exponential channels.

First, it shares with the DTP channel the characteristic that the output energy is discrete, an integer number of quanta of energy ε_0 each.

In parallel, it generalizes the DTP channel in the sense that an additive noise component is present at the channel output. The correspondence with the AEN channel is established by assuming that the noise component has a geometric distribution, the natural discrete counterpart of the exponential density in Eq. (2.10). Also, the geometric distribution has the highest entropy

among all discrete, non-negative random variables of a given mean, a property shared by the exponential density among the continuous random variables [22].

The output y_k is an integer number of quanta of energy, each of energy ε_0. As in the additive exponential noise channel, the output y_k at time k is the sum of a signal and an additive noise components, that is

$$y_k = s_k + z_k, \qquad k = 1, \dots, n, \tag{2.18}$$

where the numbers y_k, s_k and z_k are now non-negative integers, i. e. are in $\{0, 1, 2, \dots\}$.

The input is a non-negative real number $x_k \geq 0$, related to its AWGN (and AEN) equivalent by $x_k \varepsilon_0 = |x'_k|^2$, where x'_k is the AWGN value. There is a constraint on the total energy, expressed in terms of the average number of quanta per channel use, ε_s, by $\sum_{k=1}^{n} x_k \leq n\varepsilon_s$. As in the discrete-time Poisson case, the signal component at the output s_k has a Poisson distribution of parameter x_k, whose formula is given in Eq. (2.15).

The additive noise z_k has a geometric distribution of mean ε_n, that is

$$P_Z(z) = \frac{1}{1+\varepsilon_n}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^z. \tag{2.19}$$

We agree on the shorthand $Z_k \sim \mathcal{G}(\varepsilon_n)$, where the symbol $\mathcal{G}$ denotes a geometric distribution. Its variance is $\varepsilon_n(1+\varepsilon_n)$.

In order to establish a correspondence between the various channel models, we choose $\varepsilon_n \varepsilon_0 = \sigma^2 = E_n$, the average noise energies in the AWGN and AEN channels. From the discussion at the end of Section 2.4, the AEN model is recovered in the limiting case where the number of quanta becomes very large, and consequently the Poisson noise becomes negligible.

In the AE-Q channel, we identify two limiting situations, the G and P regimes, distinguished by which noise source, either additive geometric noise or Poisson noise, is predominant.

In the G regime, additive geometric noise dominates over signal-dependent noise. This happens when $\varepsilon_n \gg 1$ and $x_k \ll \varepsilon_n^2$. Note that, in addition to large number of quanta, a second condition relating the noise and signal levels is of importance. Then, the (in)equalities $\mathrm{Var}(Y_k) \simeq \mathrm{Var}(Z_k) \simeq \varepsilon_n^2 \gg x_k$ hold.

In the P regime, Poisson noise is prevalent. In terms of variances, $\mathrm{Var}(Y_k) \simeq \mathrm{Var}(S_k) \simeq x_k \gg \mathrm{Var}(Z_k)$. Since the additive geometric noise is negligible, the signal-to-noise ratio for the AEN or AWGN channel models would become infinite in this case.

It is obvious that the transition between the G and P regimes does not take place in the AWGN and AEN models, where increasing the signal energy

makes the additive noise component ever smaller. Moreover, since the models remain unchanged when both signal and noise components are multiplied by a constant value, these alternative forms are characterized by the ratio of the noise and signal energy levels, and not by their absolute values taken separately. A consequence is that signal-to-noise ratio can be freely apportioned to signal or noise components, and the choices leading to Eqs. (2.7) and (2.13) often prove convenient.

On the other hand, the AE-Q channel is sensitive to the absolute values of signal and noise, in addition to their ratio, since application of a scaling factor may easily change operation from the G to the P regime, or vice versa. The presence of the quantum as a fundamental unit of energy and the discreteness of the output fundamentally alter the behaviour of the channel under changes of scale. For the DTP or AE-Q channel models it is the absolute energies, not only their ratio, what determines the channel performance.

To the best of our knowledge the AE-Q channel model is new. A discrete counterpart of the chi-square (squared Rician) distribution, which describes the output of the non-coherent AWGN channel in Eq. (2.12), is the Laguerre distribution with parameters x and ε_n [37], and is studied in the literature on optical communications. In Section 2.6, we particularize the AE-Q channel for the case when the quanta of energy are photons of frequency ν.

The main characteristics of each channel considered so far are listed in Table 2.1: the parameters include the signal input, its energy, the total output, the signal output, the additive noise distribution and its parameters; the expressions for the channel transition matrices $Q(y|x)$ are

$$Q(y|x) = \frac{1}{\pi\sigma^2} e^{-\frac{|y-x|^2}{\sigma^2}} \qquad \text{for AWGN,} \tag{2.20}$$

$$Q(y|x) = \frac{1}{E_n} e^{-\frac{y-x}{E_n}} u(y-x) \qquad \text{for AEN,} \tag{2.21}$$

$$Q(y|x) = e^{-x} \frac{x^y}{y!} \qquad \text{for DTP,} \tag{2.22}$$

$$Q(y|x) = \sum_{l=0}^{y} \frac{e^{-x}}{1+\varepsilon_n} \left(\frac{\varepsilon_n}{1+\varepsilon_n} \right)^y \frac{\left(x\left(1+\frac{1}{\varepsilon_n}\right) \right)^l}{l!} \qquad \text{for AE-Q.} \tag{2.23}$$

	AWGN	AEN	DTP	AE-Q
Input x_k Alphabet	$\mathbf{C}$	$[0,\infty)$	$[0,\infty)$	$[0,\infty)$
Signal Energy	$\lvert x_k\rvert^2$	x_k	$x_k\varepsilon_0$	$x_k\varepsilon_0$
Output y_k Alphabet	$\mathbf{C}$	$[0,\infty)$	$\{0,1,\dots\}$	$\{0,1,\dots\}$
Signal Output s_k	x_k	x_k	$\sim\mathcal{P}(x_k)$	$\sim\mathcal{P}(x_k)$
Signal Output Mean	x_k	x_k	x_k	x_k
Signal Output Variance	0	0	x_k	x_k
Additive Noise z_k	$\sim\mathcal{N}_{\mathbf{C}}(0,\sigma^2)$	$\sim\mathcal{E}(E_n)$	0	$\sim\mathcal{G}(\varepsilon_n)$
Average Noise Energy	σ^2	$E_n=\sigma^2$	0	$\varepsilon_n\varepsilon_0=E_n$
Noise Variance $\mathrm{Var}(Z_k)$	σ^2	E_n^2	0	$\varepsilon_n(1+\varepsilon_n)$

Table 2.1: Input and output descriptions for the various channel models.

2.6 Photons as Quanta of Energy

For electromagnetic radiation of frequency ν, Einstein linked the energy ε_0 of the quantum of radiation [38], the photon, to its frequency as $\varepsilon_0 = h\nu$, where h is Planck's constant, $h = 6.626\cdot 10^{-34}$ Js. Moreover, since the additive component z_k stems from the environment at a physical temperature T_0, a natural choice for ε_n is the blackbody radiation formula [38],

$$\varepsilon_n = \left(e^{\frac{h\nu}{k_B T_0}} - 1\right)^{-1}, \tag{2.24}$$

where T_0 is the ambient temperature and k_B is Boltzmann's constant, $k_B = 1.381\cdot 10^{-23}$ J/K. Table 2.2 shows ε_n for several frequencies at $T_0 = 290$ K.

Even though we need not have an integer number of quanta at the input, it is nevertheless possible to define an equivalent number of photons in the signal input, x_k, from the AWGN value, x'_k, as

$$x_k = \frac{|x'_k|^2}{h\nu}. \tag{2.25}$$

The constraint on the average signal energy becomes $\varepsilon_s h\nu = E_s$, where ε_s is the average number of photons.

In the low-frequency (or high-temperature) limit, i. e. when $h\nu \ll k_B T_0$ Eq. (2.24) for the noise energy ε_n becomes $\varepsilon_n \simeq \frac{k_B T_0}{h\nu}$ and we recover the well-known relationship between the noise variance and the physical temperature for Gaussian channels, $\varepsilon_n h\nu \simeq k_B T_0 = \sigma^2$.

Frequency Band	Wavelength (c/ν)	$\frac{h\nu}{k_B T_0}$	ε_n
HF - 6 MHz	50 m	10^{-6}	10^6
VHF - 60 MHz	5 m	10^{-5}	10^5
UHF - 2 GHz	15 cm	$3 \cdot 10^{-4}$	3000
Microwave - 6 GHz	5 cm	10^{-3}	1000
EHF - 60 GHz	0.5 cm	0.01	100
sub-mm - 4 THz	75 μm	0.66	1
Optical - 500 THz	600 nm	83	10^{-36}

Table 2.2: Typical values of ε_n at $T_0 = 290\,\mathrm{K}$; $c = 3 \cdot 10^8\,\mathrm{m/s}$.

Background noise does not include any source originating at the receiver itself, for instance at amplifier or other electronic devices. Accepting this caveat, at optical frequencies, there is no background noise, as is approximately true in optical communication systems. Most of the radio communication systems operate in the region below 30 GHz, for which ε_n is at least in the order of thousands of photons, a value which increases to millions of photons when low frequencies are used.

In Chapter 3, we compare the channel capacities for the AWGN, AEN and AE-Q channels for several of the cases listed in Table 2.2, and discuss how the AWGN and AEN channels appear as a limiting case of the AE-Q model when the number of quanta becomes large.

2.7 Summary

In this chapter we have introduced and described the channel models we will use throughout the dissertation. All channels are memoryless and time-discrete. At time k, the output y_k is given by $y_k = s_k(x_k) + z_k$, where x_k is the channel input, $s_k(x_k)$ the input contribution to the output, and z_k an additive noise. Four forms of this model have been presented:

1. The additive white Gaussian noise (AWGN) channel, in Section 2.2.
2. The additive exponential noise (AEN) channel, described in Section 2.3.
3. The discrete-time Poisson (DTP) channel, described in Section 2.4.
4. The additive energy (AE-Q) channel, introduced in Section 2.5.

Using the AWGN as a basis for the comparison, relationships between the various channels have been derived:

- The AEN is derived by postulating that the channel output is given by the sum of the energies of the signal and noise components in the AWGN channel.
- The AE-Q is derived from the AEN model by postulating that energy is discrete and the channel output is a non-negative integer, the number of energy quanta.
- The DTP is an AE-Q channel whose additive noise component is zero.

In the DTP and AE-Q channels, energy is discrete. In Section 2.6 we have used radiation as a guide to obtain the orders of magnitude of the signal and noise components. We use these numerical values in Chapter 3 to compare the communication capabilities of the AE-Q channel with its AWGN counterpart.

A key feature of the AWGN channel is the presence of additive maximum-entropy Gaussian noise at the channel output. For other channel models, a similar role is played by the corresponding maximum-entropy distribution. Using the DTP channel as the baseline,

- The AE-Q is derived by postulating the presence of an additive, maximum-entropy noise component. This additive noise component has a geometric (or Bose-Einstein) distribution. Note that this differs from the usual practice in the analysis of the DTP channel, where an additive noise component with Poisson distribution is often considered.
- The AEN is derived from the AE-Q channel by letting the number of quanta become infinite, keeping fixed the energy value. Equivalently, the energy of the quantum ε_0 goes to zero, having kept fixed the energy. In the limit, additive noise has an exponential density, and the Poisson noise effectively vanishes.

Table 2.1, on page 21, summarizes the characteristics of the various models. All channels, bar the AWGN, operate with energies rather than with the complex amplitude in the Gaussian channel. As the AWGN admits a physical motivation, it might prove convenient to briefly discuss the physical meaning of the additive energy channels.

In terms of electromagnetic theory, the complex amplitude in the Gaussian model corresponds to the amplitude and phase of electromagnetic fields: the

component s_k is generated by a far-away antenna, and the noise z_k is produced by the environment. Discreteness in time appears by considering, say, narrow parallel frequency sub-bands. An AWGN model assumes that the fields are added at the receiver, and then detected as a complex number. In the additive energy channels, noise and signal are postulated to be incoherent, so that their energies can be added. Since the discrete-time additive energy channels cannot be derived from the discrete-time AWGN channel, possible links with physical channels would require analyzing the continuous-time AWGN channel. Some elements of this possible analysis are mentioned in Chapter 7.

3

Capacity of the Additive Energy Channels

3.1 Outline of the Chapter

In this chapter we determine the capacity of the additive energy channels. As an introduction, we review in Section 3.2 the concept of channel capacity and discuss its relevance for the problem of reliable transmission of information. We also provide some tools necessary to compute its value. In addition, we define the concept of minimum energy per bit, which is related to the capacity per unit energy. The presentation borrows elements from standard textbooks on information theory, namely Gallager [31], Blahut [39], Cover and Thomas [22].

The first channel we consider, in Section 3.3, is the standard additive Gaussian noise channel. Then, in Section 3.4 we determine the capacity of the additive exponential noise (AEN) channel, presented in Section 2.3.

In Section 3.5, we provide good upper and lower bounds to the capacity of the discrete-time Poisson channel (DTP), introduced in Section 2.4. The bounds we provide are tighter than previous results in the literature, such as those by Brady and Verdú [35] or Lapidoth and Moser [24].

Finally, we bound the capacity of the quantized additive energy channel (AE-Q) in Section 3.6. We construct the bounds from knowledge of the behaviour in the two cases discussed previously, namely the AEN and the DTP channels. These two cases respectively correspond to the G and P regimes of the AE-Q channel (see Section 2.5, page 19), and show markedly different behaviour. Since upper and lower bounds are close to each other in each of these two regimes, the channel capacity of the AE-Q is determined to a high degree of accuracy.

3.2 Mutual Information, Entropy, and Channel Capacity

Shannon's well-known block diagram of the communication channel, depicted in Chapter 2, is reproduced in Fig. 3.1. In few words, a message w is to be transmitted from the source of the communication link to the receiving end. With no loss of generality, the message may be modelled as an integer number. This message w is encoded onto a sequence $\mathbf{x}(w)$, the encoder output, possibly under a constraint on its energy $\varepsilon(\mathbf{x})$, and sent over the channel, which generates a noisy output $\mathbf{y}(\mathbf{x})$. Both sequences $\mathbf{x}$ and $\mathbf{y}$ have length n. The receiver produces a message estimate, $\hat{w}$, from the noisy output $\mathbf{y}$.

The channels we study are discrete-time, memoryless and stationary, such that an output symbol y_k, $k = 1, \ldots, n$ is generated for each symbol x_k. The conditional output density $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ (resp. $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ for discrete output) satisfies

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^{n} p_{Y|X}(y_k|x_k) = \prod_{k=1}^{n} Q(y_k|x_k), \tag{3.1}$$

where both $p_{Y|X}(\cdot|\cdot)$, the conditional density of the output (resp. $P_{Y|X}(\cdot|\cdot)$), and $Q(\cdot|\cdot)$, the channel transition matrix, are equivalent ways of characterizing the channel.

The output at time k, y_k, satisfies a generic equation of the form

$$y_k = s_k(x_k) + z_k, \tag{3.2}$$

where $s_k(x_k)$ is a (possibly random) variable, whose realization depends on the input x_k, and the component z_k is random. Input and output alphabets, as well as measures for the different channel models, were defined in Chapter 2, in Eqs. (2.20)–(2.23).

In a sense which will be made precise later, the channel input x_k may be seen as the realization of a random variable X with density $p_X(\cdot)$. In the following we drop the sub-index k unless doing so creates ambiguity. One defines the mutual information between the channel input X and output Y, in

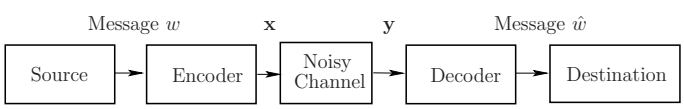


Figure 3.1: Generic communication channel.

Definition 3.1. The mutual information $I(X;Y)$ between the channel input X, with density $p_X(x)$, and the channel output Y is the quantity

$$I(X;Y) = \iint p_X(x)Q(y|x) \log \frac{Q(y|x)}{p_Y(y)} \, dy \, dx, \tag{3.3}$$

where $p_Y(y)$, the output density, is given by $p_Y(y) = \int p_X(x)Q(y|x) \, dx$.

Throughout the dissertation, logarithms are taken in base e, and the mutual information is measured in nats. The same holds for other information-theoretic quantities, such as entropies or channel capacities. Further, we agree that $0 \log 0 = 0$. In general, though, plots of the channel capacity will be in bits. In some of the cases we consider the output is discrete and/or the input density may contain Dirac delta's.

In our analysis of digital modulation in Chapters 4 and 5, x is drawn from a set $\mathcal{X}$. In that case, we denote the corresponding mutual information by $\mathrm{C}_{\mathcal{X}}$ and refer to it as constrained capacity or coded modulation capacity, where the constraint comes from using a specific modulation format.

When the random variables are purely discrete or continuous, one can define their (differential) entropy $H(U)$ (we use the same symbol for both cases).

Definition 3.2. The differential entropy $H(X)$ of a continuous random variable X with density $p_X(x)$ is given by

$$H(X) = -\int p_X(x) \log p_X(x) \, dx. \tag{3.4}$$

Definition 3.3. The entropy $H(X)$ of a discrete random variable X with distribution $P_X(x)$ is given by

$$H(X) = -\sum_x P_X(x) \log P_X(x). \tag{3.5}$$

Similar definitions hold for the conditional differential entropy, as well as the conditional entropy $H(Y|X)$. For instance, in the discrete case, the conditional entropy between two random variables Y and X is given by

$$H(Y|X) = -\sum_x P_X(x) \sum_y P_{Y|X}(y|x) \log P_{Y|X}(y|x). \tag{3.6}$$

The conditional differential entropies have the analogous definition.

We are most interested in cases where the variable Y is a (possibly stochastic) function of another variable X, viz. $Y = S(X) + Z$. Whenever the (differential) entropies $H(X)$ and $H(X|Y)$, or $H(Y)$ and $H(Y|X)$, are finite the mutual information respectively admits the well-known decompositions

$$I(X;Y) = H(Y) - H(Y|X) = H(X) - H(X|Y). \tag{3.7}$$

In the channels we study, inputs are used under an energy constraint, namely E_s for the AWGN and AEN channels and ε_s for the AE-Q and DTP channels. Recall that for each time index, the instantaneous signal energy (or cost, in general) $\varepsilon(x_k)$ is given by $\varepsilon(x_k) = |x_k|^2$ for the AWGN channel and by $\varepsilon(x_k) = x_k$ for the AEN, AE-Q and DTP channels. For an energy constraint E, one defines the channel capacity $\mathrm{C}(E)$ as

Definition 3.4. The channel capacity with energy constraint E is the supremum among all possible input distributions of the mutual information between channel input and output,

$$\mathrm{C}(E) = \sup_{p_X(\cdot)} I(X;Y), \tag{3.8}$$

where the optimization is subject to the constraint, $\mathrm{E}\big[\varepsilon(X)\big] \leq E$.

Note that the constraint on the total energy has been replaced by a constraint on the average energy of the random variable X. Additionally, in the channels we study the largest mutual information is attained when the inequality constraint on the average energy is satisfied with equality. The original constraint on the maximum energy is therefore equivalent to a constraint on the average energy.

Associated to the definition of channel capacity, a coding theorem sharply determines the limits on the reliability of communication. Shannon's coding theorem states [1,22,31,39] that, as the length of the encoder output sequence $\mathbf{x}$, n, becomes large, the probability that the receiver's message estimate, $\hat{w}$, differs from the original message at the source, w, is vanishingly small. Let the set of messages w have cardinality $|\mathcal{W}|$ and define the rate R as $R = \frac{1}{n}\log|\mathcal{W}|$. Denote the message estimate at the receiver by $\hat{w}$:

Theorem 3.5 (Noisy-Channel Coding Theorem)**.** *For each $\epsilon > 0$ for all n large enough, there exist an encoder and decoder such that the probability of error, $\Pr(\hat{W} \neq W)$ is at most ϵ, if $R < C$.*

Moreover, when $R > C$ the bit error probability at the receiver cannot be made small.

The exact computation of the channel capacity is possible only in some specific cases. The AWGN case is a well-known example, and the AEN channel will turn out to share this trait. In general, one resorts to computing bounds to the capacity. A straightforward method to derive lower bounds is

Proposition 3.6 (Lower Bound to Capacity)**.** *A fixed input density $p_X(x)$ which satisfies the energy constraint achieves a mutual information $I(X;Y)$ lower than the channel capacity, $I(X;Y) \leq C(E)$.*

There is also a simple procedure to derive upper bounds from a fixed output density, which need not correspond to a valid channel output.

Proposition 3.7 (Upper Bound to Capacity)**.** *Let $p_Y(y)$ be an arbitrary channel output density. Then, for every $\gamma \geq 0$ and every energy E the following expression gives an upper bound to the channel capacity at energy E, $C(E)$,*

$$C\big(E\big) \leq \max_x \left(\int Q(y|x) \log \frac{Q(y|x)}{p_Y(y)}\, dy - \gamma(\varepsilon(x) - E) \right). \tag{3.9}$$

We prove the result in Appendix 3.A; the discrete case is essentially identical and is omitted. The proof builds on a result by Blahut (in the proof of the convergence of the algorithm to compute the channel capacity, Section 5.4 of his book [39]). Other appearances of similar results are in Gallager's book [31], in Section 4.5 and Problem 4.17; in Section 3.3 of Csiszár and Körner's book [40]; the analysis of the capacity of the non-coherent discrete-time Rayleigh-fading channel [41]; and the derivation of upper bounds to the capacity by Lapidoth and Moser [24]. Our formulation includes the effect of the constraint.

The capacity per unit energy (or unit cost in general), which we denote by C_1 is a close relative to the channel capacity. It was defined by Verdú [29] as the largest information which can be reliably sent over the channel per unit cost. In our channel models the capacity per unit cost is closely related to the smallest energy per bit required to transmit information reliably.

Definition 3.8. In a channel with capacity $\mathrm{C}\big(E\big)$ at energy E, the capacity per unit energy C_1 is given by

$$\mathrm{C}_1 = \sup_E \frac{\mathrm{C}\big(E\big)}{E}. \tag{3.10}$$

Similarly, the minimum energy per bit is defined as $E_{b,\min} = \inf_E \frac{E}{\mathrm{C}(E)}$, where the capacity is measured in bits.

If the capacity is in nats, it is clear from the definitions that

$$E_{b,\min} = \frac{\log 2}{\mathrm{C}_1}. \tag{3.11}$$

Since the dependence on the cost constraint E is removed, the capacity per unit energy often proves easier to determine than the channel capacity itself. In particular, we have [29]

Theorem 3.9. *If the channel has an input x_0 with zero energy, i. e. $\varepsilon(x_0) = 0$, then the capacity per unit energy is given by*

$$C_1 = \sup_x \frac{D\big(Q(y|x)||Q(y|x_0)\big)}{\varepsilon(x)}, \tag{3.12}$$

where $D\big(Q(y|x)||Q(y|x_0)\big)$ is the divergence between the output distributions for inputs x and x_0, given by

$$D\big(Q(y|x)||Q(y|x_0)\big) = \sum_y Q(y|x) \log \frac{Q(y|x)}{Q(y|x_0)}. \tag{3.13}$$

For channels with continuous output, the divergence in Eq. (3.13) is defined in the natural manner.

3.3 Capacity of the Gaussian Noise Channel

We described the additive Gaussian noise (AWGN) channel in Section 2.2. Under an signal energy constraint E_s, and for circularly-symmetric complex Gaussian noise with variance σ^2, the capacity is given in

Theorem 3.10. *The capacity of the complex-valued AWGN channel is*

$$C(\mathrm{SNR}) = \log(1 + \mathrm{SNR}), \tag{3.14}$$

where SNR *is the signal-to-noise ratio,* $\mathrm{SNR} = E_s/\sigma^2$.

The proof [1] depends on the fact that a Gaussian density maximizes the differential entropy among all the densities with the same variance [1, 33].

Proposition 3.11. *The density which maximizes the differential entropy of a complex-valued random variable, subject to a constraint on the variance σ_0^2, is the Gaussian density $\mathcal{N}_{\mathbf{C}}(\mu, \sigma_0^2)$, where μ is arbitrary. Its entropy $H_{\mathcal{N}_{\mathbf{C}}}(\sigma_0^2)$ is*

$$H_{\mathcal{N}_{\mathbf{C}}}(\sigma_0^2) = \log(\pi e \sigma_0^2). \tag{3.15}$$

3.4 Capacity of the Additive Exponential Noise Channel

We described the additive exponential noise (AEN) channel in Section 2.3. It is a discrete-time model as in Eq. (3.2), with non-negative, real-valued input x_k and output y_k. The input is used under a constraint, $\sum_{k=1}^{n} x_k \leq nE_s$, where E_s is the energy per channel use. The additive noise component samples are independently drawn from an exponential distribution, $Z_k \sim \mathcal{E}(E_n)$, with E_n the average noise energy.

The capacity of the AEN channel coincides with the capacity of the equivalent AWGN channel,

Theorem 3.12. *The capacity of the additive-exponential noise channel, C, is*

$$C(\mathrm{SNR}) = \log(1 + \mathrm{SNR}), \tag{3.16}$$

where SNR *is the signal-to-noise ratio,* $\mathrm{SNR} = E_s/E_n$.

The capacity of this channel was determined by Verdú [21], in a different context. His proof exploits that an exponential density maximizes the differential entropy among the distributions of non-negative random variables [1, 21]:

Proposition 3.13. *The density which maximizes the differential entropy of a non-negative random variable U, subject to a constraint on the mean ε, is the exponential density. Its entropy $H_{Exp}(\varepsilon)$ is*

$$H_{Exp}(\varepsilon) = 1 + \log \varepsilon = \log e\varepsilon. \tag{3.17}$$

We also need the following result,

Proposition 3.14. *Let Y be a continuous non-negative real-valued random variable of the form $Y = X + Z$; X has mean E_s and Z is additive exponential noise of mean E_n. The output Y has exponential distribution of mean $E_s + E_n$ (and thus maximum differential entropy) when the input density $p_X(x)$ is*

$$p_X(x) = \frac{E_s}{(E_s + E_n)^2} \exp\left(-\frac{x}{E_s + E_n}\right) + \frac{E_n}{E_s + E_n}\delta(x), \quad x \geq 0. \tag{3.18}$$

We shall later see that a similar density maximizes the output entropy in the quantized additive energy channel.

We now reproduce the proof of Theorem 3.12.

Proof. For a given input density $p_X(\cdot)$, the mutual information $I(X;Y)$ satisfies

$$I(X;Y) = H(Y) - H(Y|X) \tag{3.19}$$
$$= H(Y) - H(X+Z|X) \tag{3.20}$$
$$= H(Y) - H(Z), \tag{3.21}$$

as the noise is additive. A maximum is achieved when $H(Y)$ is maximum, that is when Y itself it exponentially distributed. This condition is satisfied for the density in Proposition 3.14. As the output mean is $E_s + E_n$, we have

$$\mathrm{C} = \max_{p_X(\cdot)} I(X;Y) = \log(E_s + E_n) - \log E_n. \tag{3.22}$$

□

In the Gaussian case, the capacity-achieving input is Gaussian, the same distribution that the noise has. However, as found by Verdú, with additive exponential noise the input is not purely exponential, but has a delta at $x = 0$,

3.5 Capacity of the Discrete-Time Poisson Channel

3.5.1 Introduction and Main Results

In this section, and before embarking on the analysis of the quantized additive energy channel, we study the channel capacity of the discrete-time Poisson (DTP) channel. The channel was described in Section 2.5. Briefly said, it has a non-negative, real-valued input x_k, and an integer output y_k drawn from a Poisson distribution with parameter x_k, that is $Y_k \sim \mathcal{P}(x_k)$. The input is used under an energy constraint, $\sum_{k=1}^{n} x_k \leq n\varepsilon_s$, ε_s being the average number of quanta of energy per channel use.

In the context of optical communications, where this channel model sometimes appears, the capacity $\mathrm{C}(\varepsilon_s)$ was estimated by Gordon [12] as,

$$\mathrm{C}(\varepsilon_s) \simeq \frac{1}{2}\big(H_{\mathrm{Geom}}(\varepsilon_s) - \log_2(2\pi e^{\gamma_{\mathrm{e}}})\big), \tag{3.23}$$

where $\gamma_{\mathrm{e}} \simeq 0.5772\ldots$ is Euler's constant, and $H_{\mathrm{Geom}}(t)$ is the entropy of a geometric distribution with mean t,

$$H_{\mathrm{Geom}}(t) = (1+t)\log_2(1+t) - t\log_2(t). \tag{3.24}$$

This mutual information is obtained by an exponential density at the channel input, a density sometimes conjectured to be optimal [12, 15]. As ε_s becomes large, the mutual information can be further simplified, and approximated by

$$\mathrm{C}(\varepsilon_s) \simeq \frac{1}{2}\log_2 \varepsilon_s - 1. \tag{3.25}$$

From an information-theoretic perspective, Brady and Verdú [35] computed some asymptotic bounds for the channel capacity. Their analysis was recently refined by Lapidoth and Moser [42], who derived the (tighter) bounds

$$\mathrm{C}(\varepsilon_s) \geq \frac{1}{2}\log_2 \varepsilon_s + (1+\varepsilon_s)\log_2\left(1+\frac{1}{\varepsilon_s}\right) - \left(1+\sqrt{\frac{\pi}{24\varepsilon_s}}\right)\log_2 e, \tag{3.26}$$

$$\mathrm{C}(\varepsilon_s) \leq \frac{1}{2}\log_2 \varepsilon_s + o_{\varepsilon_s}(1), \tag{3.27}$$

where the error term $o_{\varepsilon_s}(1)$ tends to zero as $\varepsilon_s \to \infty$. The upper bound is of asymptotic nature.

A critical examination of the proofs for these results [12, 35] shows that the blocking elements in obtaining simple formulas are mainly two. Firstly, there is no simple formula for the entropy of a random variable with a Poisson distribution, especially compared to the neat formula for the differential entropy of a continuous variable with Gaussian density (see Eq. (3.15) on page 31). The major difficulty here is the evaluation of the term $\log m!$, where m is an integer. A common solution makes use of the Stirling approximation, which is accurate for moderate and large values of m, but leads to non exact expressions. We attack this problem by using an integral representation of the function $\log m!$ [43, 44], as we will see in Section 3.5.2. As a by-product of the analysis, we shall also determine the entropy of discrete random variables with a negative binomial distribution.

The second major difficulty is the determination of the optimum input density. In Section 3.5.3 we study in detail the mutual information of the family of gamma densities. We shall see how some gamma densities achieve a larger mutual information for low quanta counts. An exact integral representation of the mutual information will be given. As an example, for the exponential input density, the exact mutual information $I(X_{\mathcal{E}(\varepsilon_s)};Y)$ (in nats) is

$$I(X_{\mathcal{E}(\varepsilon_s)};Y) = (\varepsilon_s+1)\log(\varepsilon_s+1) + \int_0^1 \frac{\varepsilon_s^2(1-u)}{1+\varepsilon_s(1-u)}\frac{du}{\log u} - \varepsilon_s\gamma_{\mathrm{e}}. \tag{3.28}$$

Asymptotically, as $\varepsilon_s \to \infty$, we find that its behaviour is

$$I(X_{\mathcal{E}(\varepsilon_s)}; Y) \simeq \frac{1}{2}\log_2 \varepsilon_s - 0.188, \tag{3.29}$$

that is, higher than Gordon's approximation in Eq. (3.25). As a by-product of the analysis, tight lower and upper bounds to the channel capacity can be computed. A lower bound to the capacity is the mutual information achievable with a gamma input, $S \sim \mathcal{G}(1/2, \varepsilon_s)$, in Eq. (3.46). We have numerically checked, but not proved, the additional inequality

$$\mathrm{C}(\varepsilon_s) \geq I(X_{\mathcal{G}(1/2,\varepsilon_s)}; Y) \geq \frac{1}{2}\log(1+\varepsilon_s). \tag{3.30}$$

Our best upper bound admits the simple expression

$$\mathrm{C}(\varepsilon_s) \leq \log\left(\left(1 + \frac{\sqrt{2e}-1}{\sqrt{1+2\varepsilon_s}}\right)\frac{\left(\varepsilon_s + \frac{1}{2}\right)^{\varepsilon_s+\frac{1}{2}}}{\sqrt{e}\varepsilon_s^{\varepsilon_s}}\right). \tag{3.31}$$

As ε_s becomes large, this upper bound asymptotically tends to

$$\mathrm{C}(\varepsilon_s) \leq \frac{1}{2}\log \varepsilon_s + \frac{\sqrt{2e}-1}{\sqrt{2}\sqrt{\varepsilon_s}} + \mathrm{O}\left(\tfrac{1}{\varepsilon_s}\right), \tag{3.32}$$

in line with the bound in Eq. (3.27).

The remainder of this section is structured as follows. First, in Section 3.5.2 we give closed-form expressions for the entropy of random variables with Poisson or negative binomial distributions. In Section 3.5.3 similar closed-form expressions are given for the mutual information of a DTP channel when the input has a gamma density; these expressions give lower bounds to the channel capacity. In Section 3.5.4 we provide an upper bound to the channel capacity. Finally, numerical results are presented in Section 3.5.5.

3.5.2 Entropy of Poisson (and Derived) Random Variables

In our computations we shall make extensive use of the following exact identity for the logarithm of Euler's gamma function [44], $\log\Gamma(x)$,

$$\log\Gamma(x) = \int_0^1 \left(\frac{1-u^{x-1}}{1-u} - (x-1)\right)\frac{du}{\log u}. \tag{3.33}$$

Euler's gamma function [44] is defined as

$$\Gamma(x) = a^x \int_0^\infty u^{x-1} e^{-au}\, du, \tag{3.34}$$

where a is positive real number, $a > 0$. We often take $a = 1$, the common definition of the gamma function. Recall that Euler derived his gamma function as a generalization the factorial $m!$ and that for integer m we have the identity $m! = \Gamma(m+1)$. Therefore

$$\log m! = \log \Gamma(m+1) = \int_0^1 \left(\frac{1-u^m}{1-u} - m \right) \frac{du}{\log u}. \tag{3.35}$$

These expressions do not seem to admit an evaluation in terms of more fundamental functions but are nevertheless easily computed and, more importantly, they may be manipulated algebraically to simplify some expressions involving $\log \Gamma(x)$ or $\log m!$, as we now see. The technique was used by Boersma [43] to determine the entropy of a Poisson distribution.

Proposition 3.15. *The entropy $H_{Pois}(x)$ of a random variable distributed according to a Poisson law with parameter x is*

$$H_{Pois}(x) = x - x \log x + \int_0^1 \left(\frac{1-e^{-x(1-u)}}{1-u} - x \right) \frac{du}{\log u}. \tag{3.36}$$

The formula is exact and easy to evaluate with the aid of a computer. The derivation is straightforward; details can be found in Appendix 3.B.

The integral in Eq. (3.36) is proper and converges: at $u = 0$, the integrand tends to 0; at $u = 1$, an expansion of the exponential term around $u = 1$ yields

$$\lim_{u \to 1} \left(\frac{1-e^{-x(1-u)}}{1-u} - x \right) \frac{1}{\log u} = \frac{x^2}{2}. \tag{3.37}$$

Here we have removed a common term $(u-1)$ from numerator and denominator and then applied the L'Hôpital's rule.

It is worthwhile determining the asymptotic form, as $x \to \infty$, of the entropy of a Poisson distribution. To do so, we consider a slightly more general function, which will appear later in this chapter:

Proposition 3.16. *As $x \to \infty$, the function $f(x)$ given by*

$$f(x) = x - x \log x + \int_0^1 \left(u^{\nu-1} \frac{1-e^{-x(1-u)}}{1-u} - x \right) \frac{du}{\log u}, \tag{3.38}$$

where $\nu > 0$, behaves as

$$f(x) = \left(\nu - \tfrac{1}{2}\right)\gamma_e + 1 + \left(\nu - \tfrac{1}{2}\right)\log x - \frac{1}{x}\left(\frac{1}{12} - \frac{(\nu-1)^2}{2}\right) + \int_0^1 \left(\frac{(1-u)^{\nu-1}}{u\log(1-u)} + \frac{1}{u^2} - \frac{1}{u}\left(\nu - \tfrac{1}{2}\right)\right) du + O\left(\tfrac{1}{x^2}\right). \tag{3.39}$$

In particular, the entropy of the Poisson distribution ($\nu = 1$), grows as

$$H_{Pois}(x) = \frac{1}{2}\log(2\pi e x) - \frac{1}{12x} + O\left(\tfrac{1}{x^2}\right). \tag{3.40}$$

The asymptotic formula for the Poisson distribution coincides with that derived in [43].

The proof is the result of a collaboration with Dr. Hennie Ter Morsche, from the Department of Mathematics at Technische Universiteit Eindhoven.

Proof. The derivation of Eq. (3.39) can be found in Appendix 3.C. The particularization for the Poisson distribution uses the identity

$$\frac{1}{2}\gamma_e + 1 + \int_0^1 \left(\frac{1}{u\log(1-u)} + \frac{1}{u^2} - \frac{1}{2u}\right) du = \frac{1}{2}\log(2\pi e). \tag{3.41}$$

□

If the parameter s of a Poisson variable changes a new discrete random variable is generated. Let us consider the general family of gamma densities, which may be seen as a generalization of the exponential density. Each member of the family is characterized by a parameter ν and the density $p_X(x)$ is

$$p_X(x) = \frac{\nu^\nu}{\Gamma(\nu){\varepsilon_s}^\nu} x^{\nu-1} e^{-\frac{\nu x}{\varepsilon_s}}. \tag{3.42}$$

The number ε_s is the mean of the variable X. The exponential density is obtained by setting $\nu = 1$. The output induced by a gamma density is a negative binomial distribution [36], with distribution $P_Y(y)$

$$P_Y(y) = \frac{\Gamma(y+\nu)}{y!\Gamma(\nu)} \frac{\nu^\nu {\varepsilon_s}^y}{(\varepsilon_s + \nu)^{y+\nu}}. \tag{3.43}$$

For $\nu = 1$ we obtain the geometric distribution of mean ε_s, as we should.

We have the following result,

Proposition 3.17. *The entropy $H_{NegBin}(\nu, \varepsilon_s)$ of a random variable distributed according to a negative binomial law with mean ε_s and parameter ν is*

$$H_{NegBin}(\nu, \varepsilon_s) = (\varepsilon_s + \nu)\log(\varepsilon_s + \nu) - \nu \log \nu - \varepsilon_s \log \varepsilon_s + \int_0^1 \big(1 - \mathrm{pgf}(u)\big) \frac{1 - u^{\nu-1}}{1 - u} \frac{du}{\log u}, \tag{3.44}$$

where $\mathrm{pgf}(u) = \big(1 + \frac{\varepsilon_s}{\nu}(1 - u)\big)^{-\nu}$.

The derivation of Eq. (3.44) exploits the identity Eq. (3.33) to obtain an alternative form of $\Gamma(y + \nu)$ and $y!$. It can be found in Appendix 3.D.

For the geometric random variable we recover the entropy in Eq. (3.24).

As in happened for the entropy of a Poisson distribution, the integral in Eq. (3.44) converges for $u \to 1$. Applying L'Hôpital's rule twice yields

$$\lim_{u \to 1} \frac{\big(1 - \mathrm{pgf}(u)\big)(1 - u^{\nu-1})}{(1 - u)\log u} = \varepsilon_s(1 - \nu). \tag{3.45}$$

3.5.3 Lower Bound: Mutual Information with Gamma Input

As we mentioned in Section 3.2 lower bounds to the channel capacity are straightforward to generate. A fixed input density achieves a mutual information $I(X; Y)$, in general lower than the capacity. We now exploit the representation derived in the previous section to compute the mutual information in the family of gamma densities. Each of them gives thus a lower bound.

We denote the input by $X_{\mathcal{G}(\nu,\varepsilon_s)}$, to make the gamma density explicit.

Theorem 3.18. *A lower bound to the capacity of the DTP channel with energy constraint ε_s is given by the mutual information achieved by an input with gamma density with parameter ν, denoted by $I(X_{\mathcal{G}(\nu,\varepsilon_s)}; Y)$, and given by*

$$I(X_{\mathcal{G}(\nu,\varepsilon_s)}; Y) = (\varepsilon_s + \nu)\log\frac{\varepsilon_s + \nu}{\nu} + \varepsilon_s\big(\psi(\nu + 1) - 1\big) - \int_0^1 \left(\left(1 - \frac{\nu^\nu}{(\nu + \varepsilon_s(1 - u))^\nu}\right)\frac{u^{\nu-1}}{1 - u} - \varepsilon_s\right)\frac{du}{\log u}, \tag{3.46}$$

where $\psi(y)$ is Euler's digamma function, $\psi(y) = \Gamma'(y)/\Gamma(y)$.

The proof is included in Appendix 3.E.

A particularization to $\nu = 1$, the exponential density, gives the achievable mutual information $I(X_{\mathcal{E}(\varepsilon_s)}; Y)$ as

$$I(X_{\mathcal{E}(\varepsilon_s)}; Y) = (\varepsilon_s + 1)\log(\varepsilon_s + 1) + \int_0^1 \frac{\varepsilon_s^2(1-u)}{1+\varepsilon_s(1-u)} \frac{du}{\log u} - \varepsilon_s \gamma_{\mathrm{e}}. \quad (3.47)$$

Here we used that $\psi(2) = 1 - \gamma_e$, γ_e being Euler's constant.

These exact formulas, even though they do not seem to admit an expression in terms of more fundamental functions, are relatively easy to compute.

3.5.4 Duality-based Upper Bound

We consider now the family of output distributions obtained in a Poisson channel when the input density $p_\Delta(x)$ is used,

$$p_\Delta(x) = \Delta\, \delta(x) + (1-\Delta)\, p_{\mathcal{G}(\nu,\varepsilon_s)}(x), \quad (3.48)$$

that is a mixture of a gamma density with parameter ν and mean ε_s, denoted by $p_{\mathcal{G}(\nu,\varepsilon_s)}(x)$, chosen with probability $1-\Delta$, and a delta at $x = 0$ whose probability is Δ. Accordingly, the output distribution is a mixture of a negative binomial distribution with mean ε_s and parameter ν, and the zero-output variable. The output distribution, $P_\Delta(y)$ is

$$P_\Delta(y) = \begin{cases} \Delta + (1-\Delta)\dfrac{\nu^\nu}{(\varepsilon_s+\nu)^\nu}, & y = 0, \\ (1-\Delta)\dfrac{\Gamma(y+\nu)}{y!\Gamma(\nu)} \dfrac{\nu^\nu {\varepsilon_s}^y}{(\varepsilon_s+\nu)^{y+\nu}}, & y > 0. \end{cases} \quad (3.49)$$

Each distribution gives an upper bound via Eq. (3.9), with $Q(y|x) = e^{-x}\frac{y^x}{y!}$. The flexibility in the choice of ν and Δ allows us to derive a tight bound.

For this family, let us define a function $\kappa(x)$ as

$$\kappa(x) = \sum_y Q(y|x) \log \frac{Q(y|x)}{P_\Delta(m)} - \gamma(x - \varepsilon_s). \quad (3.50)$$

The function is computed in Appendix 3.F, and is found to have the form

$\kappa(x) = \kappa_0(x) + \kappa_1(x,\gamma) + \kappa_2(\gamma)$, with the summands in turn given by

$$\begin{aligned}\kappa_0(x) &= -x + x\log x + \int_0^1 \left(x - u^{\nu-1}\frac{1-e^{-x(1-u)}}{1-u}\right)\frac{du}{\log u} \\ &\quad - e^{-x}\log\left(1 + \frac{\Delta}{(1-\Delta)}\left(1+\frac{\varepsilon_s}{\nu}\right)^{\nu}\right) \qquad (3.51)\\ \kappa_1(x,\gamma) &= -x\left(\gamma - \log\frac{\varepsilon_s+\nu}{\varepsilon_s}\right) \qquad (3.52)\\ \kappa_2(\gamma) &= \nu\log\frac{\varepsilon_s+\nu}{\nu} + \gamma\varepsilon_s - \log(1-\Delta). \qquad (3.53)\end{aligned}$$

From Proposition 3.7, an upper bound to the capacity is given by $\max_x \kappa(x)$. We now separately examine the three summands to locate the maximum.

First, $\kappa_2(\gamma)$ is constant and has therefore no impact in the localization of the maximum over x. Then, choosing $\gamma = \log\frac{\varepsilon_s+\nu}{\varepsilon_s}$ makes $\kappa_1(x,\gamma) = 0$ and therefore constant, with no impact on the localization of the maximum over x. We consider only the variation of $\kappa_0(x)$ with x.

Initially, consider $\Delta = 0$, later we will optimize Δ in order to tighten the bound. From Proposition 3.16, the behaviour as $x \to \infty$ of $\kappa_0(x)$ is given by

$$\begin{aligned}\kappa_0(x) = &-\left(\nu - \tfrac{1}{2}\right)\gamma_e - 1 - \left(\nu - \tfrac{1}{2}\right)\log x + \frac{1}{x}\left(\frac{1}{12} - \frac{(\nu-1)^2}{2}\right) \\ &- \int_0^1 \left(\frac{(1-u)^{\nu-1}}{u\log(1-u)} + \frac{1}{u^2} - \frac{1}{u}\left(\nu - \tfrac{1}{2}\right)\right) du + \mathrm{O}\left(\tfrac{1}{x^2}\right), \qquad (3.54)\end{aligned}$$

and therefore diverges to $+\infty$ if $\nu < \frac{1}{2}$. This fact is corroborated by Fig. 3.2, which shows the values of $\kappa_0(x)$ in the range $0 \le x \le 10$, for some values of ν. For $\nu \ge \frac{1}{2}$, the maximum over x appears at $x = 0$, and $\max_x \kappa_0(x) = \kappa_0(x = 0) = 0$. Therefore, using the values for γ and $\kappa_1(x,\gamma)$, an upper bound to the capacity is given by $\kappa_2(\gamma)$, whose value is

$$\kappa_2(\gamma) = \nu\log\frac{\varepsilon_s+\nu}{\nu} + \gamma\varepsilon_s = (\varepsilon_s+\nu)\log(\nu+\varepsilon_s) - \varepsilon_s\log\varepsilon_s - \nu\log\nu. \quad (3.55)$$

For $\nu = 1$ we obtain $\mathrm{C}(\varepsilon_s) \le H_{\mathrm{Geom}}(\varepsilon_s)$, a trivial statement as the geometric distribution has maximum entropy.

As $\kappa_2(\gamma)$ decreases as ν becomes smaller, the best upper bound is obtained by setting $\nu = \frac{1}{2}$. Moreover, for the choice $\nu = \frac{1}{2}$ a finite asymptotic limit is

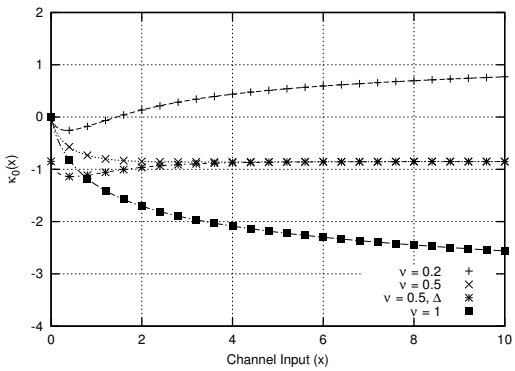


Figure 3.2: Computation of $\kappa_0(x)$ for several values of ν.

reached for $\kappa_0(x)$ as $x \to \infty$. Indeed, applying Proposition 3.16, we have that

$$\lim_{x\to\infty} \kappa_0(x) = -1 - \int_0^1 \left(\frac{(1-u)^{-1/2}}{u \log(1-u)} + \frac{1}{u^2} \right) du \tag{3.56}$$

$$= -\frac{1}{2} \log(2e). \tag{3.57}$$

In Appendix 3.G we list the numerical evaluation of $\kappa_0(x)$ for large values of x, where it is verified that the limit is approached from below, i. e. an upper bound to $\kappa_0(x)$ is given by $-\frac{1}{2}\log(2e)$. This is in accordance with Proposition 3.16, which gives the asymptotic expression for $\kappa_0(x)$,

$$\kappa_0(x) = -\frac{1}{2} \log(2e) - \frac{1}{24x} + \mathrm{O}\big(\tfrac{1}{x^2}\big), \tag{3.58}$$

and the limit is indeed approached from below for sufficiently large x. The approximation $-\frac{1}{2}\log(2e) - \frac{1}{24x}$ is also included in Appendix 3.G, and shows very good match with the function $\kappa_0(x)$ for large x.

Finally, we choose Δ so as to reduce the value at $x = 0$. Taking

$$-\log\left(1 + \frac{\Delta^*}{1-\Delta^*}\big(1+2\varepsilon_s\big)^{1/2}\right) = -\frac{1}{2}\log(2e), \tag{3.59}$$

or in other words,

$$\Delta^* = \frac{\sqrt{2e}-1}{\sqrt{2e}-1+\big(1+2\varepsilon_s\big)^{1/2}}, \tag{3.60}$$

then the value at $x = 0$ coincides with the value as $x \to \infty$, and $\kappa_0(0) = \lim_{x\to\infty} \kappa_0(x)$. This effect is visible in Fig. 3.2, where the function $\kappa_0(x)$ decreases from its value at $x = 0$ to a local minimum, and increases afterwards. We have also verified numerically that $\kappa_0'(x)$ is positive after the local minimum, that is the function is monotonic increasing and never reaches the value $\kappa_0(0)$.

Summarizing all the considerations so far, we choose an input density of the form Eq. (3.48), with $\nu = 1/2$ and the optimum Δ^* in Eq. (3.60). This yields an output distribution $P_Y(y)$, which is used together with the parameter $\gamma = \log \frac{\varepsilon_s+\nu}{\varepsilon_s}$ to derive an upper bound to the capacity, using Proposition 3.7. The maximum over x is located at $x = 0$ and $x \to \infty$ and is such that the following upper bound is obtained

$$\mathrm{C}(\varepsilon_s) \leq -\frac{1}{2}\log(2e) + \nu \log\frac{\varepsilon_s+\nu}{\nu} + \varepsilon_s \log\frac{\varepsilon_s+\nu}{\varepsilon_s} - \log(1-\Delta^*) \tag{3.61}$$

$$= \log\left(1 + \frac{\sqrt{2e}-1}{\sqrt{1+2\varepsilon_s}}\right) + \log\frac{\left(\varepsilon_s + \frac{1}{2}\right)^{\varepsilon_s+\frac{1}{2}}}{\sqrt{e}\varepsilon_s^{\varepsilon_s}}. \tag{3.62}$$

As it could be expected from Lapidoth's result in Eq. (3.27), the upper bound grows asymptotically as $\frac{1}{2}\log\varepsilon_s$ for large ε_s,

$$\frac{1}{2}\log\varepsilon_s + \frac{\sqrt{2e}-1}{\sqrt{2}\sqrt{\varepsilon_s}} + \mathrm{O}\big(\varepsilon_s^{-1}\big), \tag{3.63}$$

where we used that $\log(1+t) = t + \mathrm{O}\big(t^{-1}\big)$, as $t \to 0$.

3.5.5 Some Numerical Results

Figure 3.3 shows several bounds and approximations to the channel capacity,

1. Gordon's approximate formula Eq. (3.23).
2. Lapidoth's lower bound Eq. (3.26).
3. The exact mutual information Eq. (3.46) with two densities of the gamma family, namely $\nu = \frac{1}{2}$ and $\nu = 1$ (exponential).
4. The upper bound derived from the gamma density with $\nu = \frac{1}{2}$, Eq. (3.62).
5. An upper bound assuming a geometric distribution, $H_{\mathrm{Geom}}(\varepsilon_s)$.

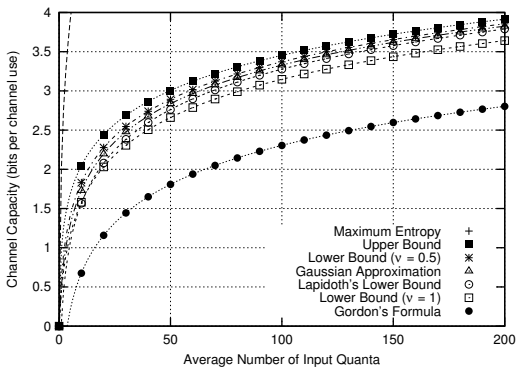


Figure 3.3: Upper and lower bounds to the channel capacity.

6. A Gaussian approximation to the capacity, $\mathrm{C}_{\mathrm{Gaus}}(\varepsilon_s) = \frac{1}{2}\log_2(1+\varepsilon_s)$.

In general our results follow the asymptotic expression for the capacity, $\frac{1}{2}\log_2 \varepsilon_s$, but are tighter than the previously known bounds, by Gordon and Lapidoth. The gamma density with $\nu = \frac{1}{2}$ is uniformly better than the exponential input, beating it by about 0.188 bits/channel use, the difference between the two asymptotic expansions of the capacity, namely

$$I(X_{\mathcal{G}(1/2,\varepsilon_s)};Y) = \frac{1}{2}\log_2 \varepsilon_s + \mathrm{o}(1), \tag{3.64}$$

$$I(X_{\mathcal{G}(1,\varepsilon_s)};Y) = \frac{1}{2}\log_2 \varepsilon_s - 0.188 + \mathrm{o}(1). \tag{3.65}$$

We have determined both expansions numerically.

In the range depicted in Fig. 3.3 the corresponding mutual information is very closely given by the Gaussian approximation $\mathrm{C}_{\mathrm{Gaus}}(\varepsilon_s)$. Figure 3.4 depicts some further curves to better compare the spectral efficiencies with $\mathrm{C}_{\mathrm{Gaus}}$. In Fig. 3.4a we show the difference

$$I(X_{\mathcal{G}(\nu,\varepsilon_s)};Y) - \mathrm{C}_{\mathrm{Gaus}}(\varepsilon_s), \tag{3.66}$$

for the mutual information for two gamma inputs, with $\nu = 1$ and $\nu = \frac{1}{2}$. The difference of $\mathrm{C}(\varepsilon_s)$ with respect to the upper bound in Eq. (3.62) is also depicted. Remarkably, the gamma density with $\nu = \frac{1}{2}$ seems to be above $\mathrm{C}_{\mathrm{Gaus}}$. This is further supported by the results plotted in Fig. 3.4b, a zoom

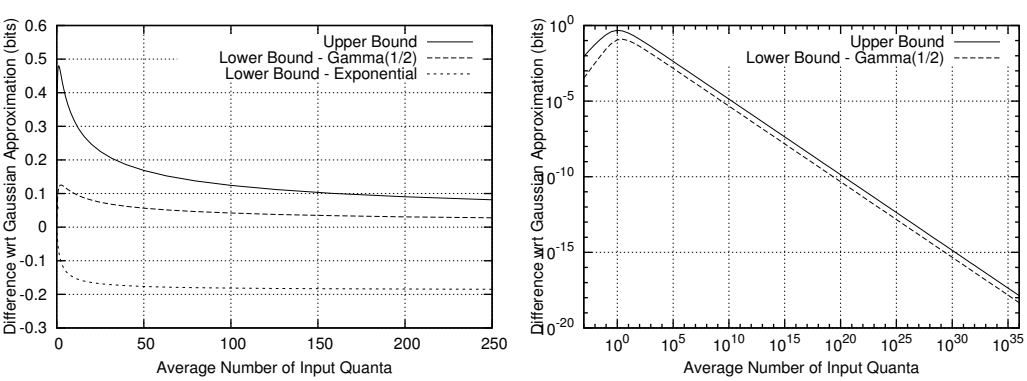
in logarithmic scale of the previous plot, where the depicted range is increased up to $\varepsilon_s = 10^{36}$ quanta.

These computations allow us to numerically determine an additional term in the asymptotic expansion to the mutual information for $\nu = \frac{1}{2}$,

$$I(X_{\mathcal{G}(1/2,\varepsilon_s)}; Y) \simeq \frac{1}{2}\log\varepsilon_s + \frac{0.324}{\sqrt{\varepsilon_s}} + \mathrm{o}(\varepsilon_s^{-1/2}). \tag{3.67}$$

The plot of the upper bound in Fig. 3.4b matches with its asymptotic form,

$$\frac{1}{2}\log\varepsilon_s + \frac{\sqrt{2e}-1}{\sqrt{2}\sqrt{\varepsilon_s}} + \mathrm{O}(\varepsilon_s^{-1}). \tag{3.68}$$



(a) Difference for low quanta counts.

(b) Difference for large quanta counts.

Figure 3.4: Comparison of channel capacity with Gaussian approximation.

From these numbers we conjecture that $\mathrm{C}_{\mathrm{Gaus}}$ is always a lower bound to the channel capacity, even though we have not been able to prove it. It is significant that the upper bound in Eq. (3.62), also represented in Fig. 3.4b, approaches $\mathrm{C}_{\mathrm{Gaus}}$ rather rapidly; the difference between the two is below 0.1 bits for 150 quanta/mode, falling to 0.01 bits at about 15000 quanta/mode. The narrow gap between the best lower and upper bounds constitutes the tightest pair of bounds known so far for this channel.

3.6 Capacity of the Quantized Additive Energy Channel

3.6.1 Introduction and Main Results

We now move on to study the capacity of the quantized additive energy (AE-Q) channel, which was described in Section 2.5. Briefly said, in the AE-Q channel, the non-negative real-valued input x_k and the integer-valued output y_k are related as

$$y_k = s_k(x_k) + z_k, \tag{3.69}$$

where the output variables are distributed as $S_k \sim \mathcal{P}(x_k)$, and $Z_k \sim \mathcal{G}(\varepsilon_n)$, i. e. with Poisson and geometric distributions respectively. The input is used under an energy constraint, $\sum_{k=1}^{n} x_k \leq n\varepsilon_s$, where ε_s is the average energy per channel use.

As with the discrete-time Poisson channel, we determine the channel capacity $\mathrm{C}(\varepsilon_s, \varepsilon_n)$ by providing tight upper and lower bounds. The main results are presented in subsequent sections and summarized here for convenience.

Upper bounds are given in Theorem 3.19 and Theorem 3.21,

$$\mathrm{C}(\varepsilon_s, \varepsilon_n) \leq \min(\mathrm{C_G}(\varepsilon_s, \varepsilon_n), \mathrm{C_P}(\varepsilon_s)), \tag{3.70}$$

where $\mathrm{C_G}(\varepsilon_s, \varepsilon_n)$ and $\mathrm{C_P}(\varepsilon_s)$ are respectively given by

$$\begin{aligned} \mathrm{C_G}(\varepsilon_s, \varepsilon_n) &= H_{\mathrm{Geom}}(\varepsilon_s + \varepsilon_n) - H_{\mathrm{Geom}}(\varepsilon_n), \\ \mathrm{C_P}(\varepsilon_s) &= \log\left(\left(1 + \frac{\sqrt{2e} - 1}{\sqrt{1 + 2\varepsilon_s}}\right) \frac{\left(\varepsilon_s + \frac{1}{2}\right)^{\varepsilon_s + \frac{1}{2}}}{\sqrt{e}\varepsilon_s^{\varepsilon_s}}\right). \end{aligned} \tag{3.71}$$

Here $H_{\mathrm{Geom}}(t)$ is the entropy of a geometric distribution with mean t, given by Eq. (3.24). Both functions $\mathrm{C_G}(\varepsilon_s, \varepsilon_n)$ and $\mathrm{C_P}(\varepsilon_s)$ are monotonically increasing functions of ε_s. Roughly speaking, $\mathrm{C_G}$ is closer to the true capacity for low values of ε_s, whereas $\mathrm{C_P}$ is more accurate for higher values of ε_s. Note that $\mathrm{C_P}(\varepsilon_s)$ does not depend on ε_n.

This threshold behaviour can be seen in Fig. 3.5, which depicts the upper and lower bounds to the capacity as a function of the input number of quanta ε_s and for several values of ε_n. The upper bound in solid lines does not depend on ε_n, as is due to the intrinsic noise in the signal itself s_k. When the additive noise z_k can be neglected, the total noise at the output does not disappear, but is dominated by the signal-dependent noise. Neglecting the signal noise, the second upper bound (in dashed lines) essentially takes the opposite path.

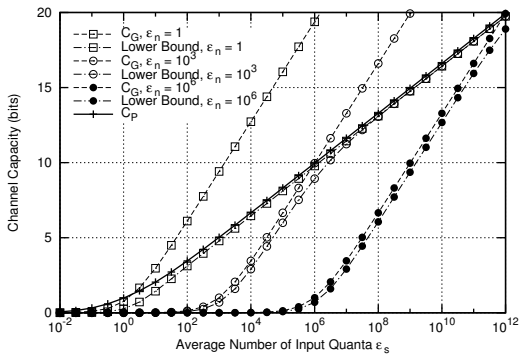


Figure 3.5: Upper bounds to the capacity for several values of ε_n.

There exists a threshold $\varepsilon_s^* = \varepsilon_n^2$ such that the behaviour of the AE-Q channel is dominated by additive geometric noise below it (the G regime), and by signal-dependent Poisson noise above it (P regime). When ε_n is large enough, a good approximation to the location of the threshold is $\varepsilon_s^* = \varepsilon_n^2$ or,

$$\varepsilon_n = \frac{\varepsilon_s^*}{\varepsilon_n} = \frac{\varepsilon_s^* \varepsilon_0}{\varepsilon_n \varepsilon_0} = \frac{E_s^*}{E_n} = \mathrm{SNR}^* \tag{3.72}$$

that is the signal-to-noise ratio of the equivalent AWGN and AEN channels. In terms of the signal-to-noise ratio of the underlying AWGN channel, below $\mathrm{SNR}^* = \varepsilon_n$ the capacities of the AE-Q and AWGN channels are very similar, whereas sufficiently above SNR^* the capacity of the AE-Q channel is half the capacity of the AWGN channel.

For each value of ε_n the lower bound $\mathrm{C}_{\mathrm{LB}}(\varepsilon_s, \varepsilon_n)$ depicted in Fig. 3.5 (in dashed-dotted lines) is the achievable mutual information with an exponential input identical to that of the additive exponential noise channel,

$$\begin{aligned}\mathrm{C}_{\mathrm{LB}}(\varepsilon_s, \varepsilon_n) = {} & H_{\mathrm{Geom}}(\varepsilon_s + \varepsilon_n) - \frac{\varepsilon_n}{\varepsilon_s + \varepsilon_n} H_{\mathrm{Geom}}(\varepsilon_n) \\ & - \frac{\varepsilon_s}{\varepsilon_s + \varepsilon_n} \left(\frac{1}{2} \log 2\pi e + \frac{1}{2} \log \left(\varepsilon_n (1 + \varepsilon_n) \frac{1}{12} \right) \times \right. \\ & \left. \times\, e^{\frac{\varepsilon_n (1+\varepsilon_n) + \frac{1}{12}}{\varepsilon_s + \varepsilon_n}} \Gamma \left(0, \frac{\varepsilon_n (1 + \varepsilon_n) + \frac{1}{12}}{\varepsilon_s + \varepsilon_n} \right) \right),\end{aligned} \tag{3.73}$$

where $\Gamma(0,t)$ is an incomplete gamma function, $\Gamma(0,t) = \int_t^\infty u^{-1}e^{-u}\,du$. Despite its unwieldy appearance, the formula is simple to evaluate numerically.

Upper and lower bounds are respectively derived in Sections 3.6.2 and 3.6.3. Numerical results, particularized for the case when quanta are identified with photons, are presented and discussed in Section 3.6.4. In this case, the threshold signal-to-noise ratio of the equivalent AWGN channel is given by

$$\mathrm{SNR}^* = \varepsilon_n \simeq \frac{6\cdot 10^{12}}{\nu}, \tag{3.74}$$

where ν is the frequency (in Hertz). In decibels, $\mathrm{SNR}^*\,(\mathrm{dB}) \simeq 37.8 - 10\log_{10}\nu$, where the frequency is in GHz.

3.6.2 Upper Bounds to the Capacity

In this section we shall give two upper bounds to the capacity C of the quantized additive energy channel as a function of the average signal energy ε_s and the average geometric noise energy ε_n. The first one will be directly related to the additive exponential noise channel, and the second derived from the discrete-time Poisson channel. Obviously, the smallest of them also constitutes an upper bound, tighter than each of them separately.

Our first result is the following

Theorem 3.19. *The capacity of the AE-Q channel is upper bounded by*

$$C(\varepsilon_s,\varepsilon_n) \le C_G(\varepsilon_s,\varepsilon_n) = H_{Geom}(\varepsilon_s+\varepsilon_n) - H_{Geom}(\varepsilon_n), \tag{3.75}$$

where $H_{Geom}(t)$ is the entropy of a geometric distribution of mean t.

Since $H_{\mathrm{Geom}}(t)$, is given by Eq. (3.24), $H_{\mathrm{Geom}}(t) = (1+t)\log(1+t) - t\log t$, an equivalent expression for the upper bound is

$$\begin{aligned}\mathrm{C_G}(\varepsilon_s,\varepsilon_n) = (1+\varepsilon_s+\varepsilon_n)\log(1+\varepsilon_s+\varepsilon_n) - (\varepsilon_s+\varepsilon_n)\log(\varepsilon_s+\varepsilon_n)\\ - (1+\varepsilon_n)\log(1+\varepsilon_n) + \varepsilon_n\log\varepsilon_n.\end{aligned} \tag{3.76}$$

We have also the following corollary,

Corollary 3.20. *For finite values of ε_s, $C_G(\varepsilon_s,\varepsilon_n)$ is bounded by*

$$\log\left(1+\frac{\varepsilon_s}{\varepsilon_n+1}\right) < C_G(\varepsilon_s,\varepsilon_n) < \log\left(1+\frac{\varepsilon_s}{\varepsilon_n}\right). \tag{3.77}$$

The capacity of the AE-Q channel is therefore strictly upper bounded by the capacity of an AEN channel of signal-to-noise ratio $\mathrm{SNR} = \varepsilon_s/\varepsilon_n$.

We now prove Theorem 3.19.

Proof. For any input $p_X(x)$ the mutual information clearly satisfies

$$I(X;Y) = H(Y) - H(Y|X) \tag{3.78}$$
$$\leq H_{\mathrm{Geom}}(\varepsilon_s + \varepsilon_n) - H\big(S(X) + Z|X\big), \tag{3.79}$$

as the geometric distribution has the highest entropy under the given constraints. Then, as conditioning reduces entropy,

$$H\big(S(X) + Z|X\big) \geq H\big(S(X) + Z|X, S\big) \tag{3.80}$$
$$= H(Z|X, S) = H(Z), \tag{3.81}$$

and Z is independent of the input X. Therefore,

$$I(X;Y) \leq H_{\mathrm{Geom}}(\varepsilon_s + \varepsilon_n) - H_{\mathrm{Geom}}(\varepsilon_n). \tag{3.82}$$

As this holds for all inputs the theorem follows. □

We now move on to prove the formulas comparing the upper bound with the capacity of the additive exponential noise channel.

Proof. First, we prove the strict inequality

$$\log(\varepsilon_s + \varepsilon_n) - \log \varepsilon_n > \mathrm{C_G}(\varepsilon_s, \varepsilon_n), \tag{3.83}$$

for all values of $\varepsilon_s > 0$, $\varepsilon_n \geq 0$. Using the definition of $\mathrm{C_G}(\varepsilon_s, \varepsilon_n)$, we have

$$\begin{aligned}(1 + \varepsilon_s + \varepsilon_n)\log(\varepsilon_s + \varepsilon_n) - (1 + \varepsilon_s + \varepsilon_n)\log(1 + \varepsilon_s + \varepsilon_n)\\ - (1 + \varepsilon_n)\log(\varepsilon_n) + (1 + \varepsilon_n)\log(1 + \varepsilon_n) > 0,\end{aligned} \tag{3.84}$$

that is,

$$(1 + \varepsilon_s + \varepsilon_n)\log\frac{\varepsilon_s + \varepsilon_n}{1 + \varepsilon_s + \varepsilon_n} > (1 + \varepsilon_n)\log\frac{\varepsilon_n}{1 + \varepsilon_n}. \tag{3.85}$$

Proving this is equivalent to proving that the function

$$f(t) = (1 + t)\log\frac{t}{1 + t} = (1 + t)\log t - (1 + t)\log(1 + t) \tag{3.86}$$

is monotonically increasing for $t > 0$. Its first derivative $f'(t)$ is

$$f'(t) = \log t + \frac{1+t}{t} - \log(1+t) - 1 \tag{3.87}$$

$$= \frac{1}{t} - \log\left(1 + \frac{1}{t}\right), \tag{3.88}$$

which is positive since $\log(1+t') < t'$ for positive t'.

We now move on to prove the second strict inequality. For $\varepsilon_s > 0$, $\varepsilon_n \geq 0$.

$$\mathrm{C_G}(\varepsilon_s, \varepsilon_n) > \log(\varepsilon_s + \varepsilon_n + 1) - \log(\varepsilon_n + 1). \tag{3.89}$$

Using the definition of $\mathrm{C_G}(\varepsilon_s, \varepsilon_n)$ and cancelling common terms,

$$\begin{aligned}(\varepsilon_s + \varepsilon_n)\log(1 + \varepsilon_s + \varepsilon_n) - (\varepsilon_s + \varepsilon_n)\log(\varepsilon_s + \varepsilon_n)\\ - \varepsilon_n \log(1 + \varepsilon_n) + \varepsilon_n \log \varepsilon_n > 0,\end{aligned} \tag{3.90}$$

a condition equivalent to

$$(\varepsilon_s + \varepsilon_n)\log\frac{1 + \varepsilon_s + \varepsilon_n}{\varepsilon_s + \varepsilon_n} > \varepsilon_n \log\frac{1 + \varepsilon_n}{\varepsilon_n}. \tag{3.91}$$

This equation is true whenever the function $f(t)$

$$f(t) = t\log\left(1 + \frac{1}{t}\right) \tag{3.92}$$

is monotonically increasing for $t > 0$. Equivalently, its first derivative $f'(t)$,

$$f'(t) = \log\left(1 + \frac{1}{t}\right) + t\frac{-\frac{1}{t^2}}{1 + \frac{1}{t}} = \log\left(1 + \frac{1}{t}\right) - \frac{1}{t+1}, \tag{3.93}$$

must be positive for $t > 0$. This condition may be rewritten as

$$\log\left(1 + \frac{1}{t}\right) > \frac{1}{t+1}, \tag{3.94}$$

or, exponentiating in both sides,

$$1 + \frac{1}{t} > e^{\frac{1}{t+1}} > 1 + \frac{1}{t+1}, \tag{3.95}$$

where the last inequality is due to the fact that $e^t > 1 + t$. Accordingly, $f'(t) > 0$ since $t + 1 > t$.

□

Our second upper bound is in

Theorem 3.21. *The capacity of the AE-Q channel is upper bounded by*

$$C(\varepsilon_s, \varepsilon_n) \leq C_P(\varepsilon_s) = \log\left(\left(1 + \frac{\sqrt{2e} - 1}{\sqrt{1 + 2\varepsilon_s}}\right) \frac{\left(\varepsilon_s + \frac{1}{2}\right)^{\varepsilon_s + \frac{1}{2}}}{\sqrt{e}\varepsilon_s^{\varepsilon_s}}\right). \tag{3.96}$$

Note that this formula does not depend on the additive noise ε_n, as will become apparent from it being derived from the discrete-time Poisson channel. The proof of Theorem 3.21 is similar to that of Theorem 3.19. It exploits that a genie has knowledge of the exact value of the additive noise component.

Proof. The variables X, $S(X)$, and $Y(S)$ form a Markov chain in this order,

$$X \to S(X) \to Y = S(X) + Z. \tag{3.97}$$

Hence, by the data processing inequality (Theorem 2.8.1 in [22]),

$$I(X;Y) \leq I\big(X; S(X)\big), \tag{3.98}$$

that is the mutual information achievable in the discrete-time Poisson channel. The bound follows then from Eq. (3.31), as an upper bound to the latter is also an upper bound to the AE-Q channel. □

3.6.3 A Lower Bound to the Capacity

Our lower bound is derived from the mutual information achievable by the input which achieves the largest output entropy. This input is characterized in

Proposition 3.22. *In the AE-Q channel the output entropy $H(Y)$ is maximized when the input density $p_X(x)$ is given by*

$$p_X(x) = \frac{\varepsilon_s}{(\varepsilon_s + \varepsilon_n)^2} \exp\left(-\frac{x}{\varepsilon_s + \varepsilon_n}\right) + \frac{\varepsilon_n}{\varepsilon_s + \varepsilon_n}\delta(x), \quad x \geq 0. \tag{3.99}$$

As a particular case we recover the exponential input, which maximizes the output entropy of a DTP channel [12]. Also, the form of the density is analogous to the result in the additive exponential noise channel, in Proposition 3.13.

The proof needs the probability generating function (pgf) $\mathrm{pgf}(u)$ of z. For a geometric random variable of mean x, the pgf is given in

Proof. Since Y is a discrete random variable, its entropy is maximum when its distribution is geometric. In that case the pgf $\mathrm{E}[u^Y]$ is given by

$$\mathrm{pgf}_{\mathrm{Geom}(\varepsilon_n)}(u) = \big(1 + \varepsilon_n(1-u)\big)^{-1}. \tag{3.100}$$

Similarly, for a Poisson distribution with mean x, its pgf is $\mathrm{pgf}_{\mathrm{Pois}(x)}(u) = e^{-x(1-u)}$. As for the signal component at the output, its pgf is a mixture of Poisson distributions, of value

$$\begin{aligned}
\mathrm{pgf}(u) &= \int_0^\infty p_X(x)\,\mathrm{pgf}_{\mathrm{Pois}(x)}(u)dx && (3.101)\\
&= \int_0^\infty \left(\frac{\varepsilon_s}{(\varepsilon_s+\varepsilon_n)^2} e^{-\frac{x}{\varepsilon_s+\varepsilon_n}} + \frac{\varepsilon_n}{\varepsilon_s+\varepsilon_n}\delta(x)\right) e^{-x(1-u)}dx && (3.102)\\
&= \frac{\varepsilon_s}{\varepsilon_s+\varepsilon_n}\frac{1}{\big((\varepsilon_s+\varepsilon_n)(1-u)+1\big)} + \frac{\varepsilon_n}{\varepsilon_s+\varepsilon_n} && (3.103)\\
&= \frac{1+\varepsilon_n(1-u)}{(\varepsilon_s+\varepsilon_n)(1-u)+1}. && (3.104)
\end{aligned}$$

As Y is the sum of two independent random variables, its pgf is the product of the corresponding pgf's, Eqs. (3.100) and (3.104), which gives

$$\big((\varepsilon_s+\varepsilon_n)(1-u)+1\big)^{-1}, \tag{3.105}$$

the pgf of a geometric distribution with mean $\varepsilon_s + \varepsilon_n$, as desired. □

We shall also need the following upper bound to the entropy of a random variable Y with a given variance $\mathrm{Var}\,Y$ (Theorem 9.7.1 of [22]),

$$H(Y) \le \frac{1}{2}\log 2\pi e\left(\mathrm{Var}\,Y + \frac{1}{12}\right). \tag{3.106}$$

We have then the following

Theorem 3.23. *A lower bound $C_{LB}(\varepsilon_s, \varepsilon_n)$ to the capacity of the AE-Q channel is*

$$\begin{aligned}
C_{LB}(\varepsilon_s,\varepsilon_n) = H_{Geom}(\varepsilon_s+\varepsilon_n) &- \frac{\varepsilon_n}{\varepsilon_s+\varepsilon_n}H_{Geom}(\varepsilon_n)\\
&- \frac{\varepsilon_s}{\varepsilon_s+\varepsilon_n}\left(\frac{1}{2}\log 2\pi e + \frac{1}{2}\log\left(\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}\right)\times\right.\\
&\left.\times\, e^{\frac{\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}}{\varepsilon_s+\varepsilon_n}}\Gamma\left(0, \frac{\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}}{\varepsilon_s+\varepsilon_n}\right)\right), && (3.107)
\end{aligned}$$

where $\Gamma(0,t)$ *is an incomplete gamma function,* $\Gamma(0,t) = \int_t^\infty u^{-1}e^{-u}\,du$.

An alternative integral expression is

$$C_{LB}(\varepsilon_s,\varepsilon_n) = -\frac{\varepsilon_s}{(\varepsilon_s+\varepsilon_n)^2}\int_0^\infty \frac{1}{2}\log 2\pi e\left(x+\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}\right)e^{-\frac{x}{\varepsilon_s+\varepsilon_n}}\,dx + H_{Geom}(\varepsilon_s+\varepsilon_n) - \frac{\varepsilon_n}{\varepsilon_s+\varepsilon_n}H_{Geom}(\varepsilon_n). \quad (3.108)$$

Proof. We choose as an input the density in Eq. (3.99) in Proposition 3.22. By construction, the output is then geometric with mean $\varepsilon_s+\varepsilon_n$ and the output entropy $H(Y)$ is therefore given by $H(Y) = H_{\text{Geom}}(\varepsilon_s+\varepsilon_n)$. We compute the mutual information with this input as $H(Y) - H(Y|X)$.

We now move on to estimate the second term, the conditional entropy, as

$$H(Y|X) = \int_0^\infty H(Y|x)\left(\frac{\varepsilon_s}{(\varepsilon_s+\varepsilon_n)^2}e^{-\frac{x}{\varepsilon_s+\varepsilon_n}} + \frac{\varepsilon_n}{\varepsilon_s+\varepsilon_n}\delta(x)\right)dx \quad (3.109)$$

$$= \frac{\varepsilon_n}{\varepsilon_s+\varepsilon_n}H(Y|x=0) + \frac{\varepsilon_s}{(\varepsilon_s+\varepsilon_n)^2}\int_0^\infty H(Y|x)e^{-\frac{x}{\varepsilon_s+\varepsilon_n}}\,dx. \quad (3.110)$$

The first term is the entropy of a geometric distribution, that is $H(Y|x=0) = H_{\text{Geom}}(\varepsilon_n)$. And the second is upper bounded by Eq. (3.106),

$$H(Y|x) \le \frac{1}{2}\log 2\pi e\left(\text{Var}(Y|x)+\frac{1}{12}\right) \quad (3.111)$$

$$= \frac{1}{2}\log 2\pi e + \frac{1}{2}\log\left(x+\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}\right). \quad (3.112)$$

Hence,

$$\int_0^\infty H(Y|x)e^{-\frac{x}{\varepsilon_s+\varepsilon_n}}\,dx \le \frac{1}{2}\log 2\pi e\int_0^\infty e^{-\frac{x}{\varepsilon_s+\varepsilon_n}}\,dx + \int_0^\infty \frac{1}{2}\log\left(x+\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}\right)e^{-\frac{x}{\varepsilon_s+\varepsilon_n}}\,dx \quad (3.113)$$

$$= (\varepsilon_s+\varepsilon_n)\frac{1}{2}\log 2\pi e + (\varepsilon_s+\varepsilon_n)\frac{1}{2}\log\left(\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}\right)\times \times e^{\frac{\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}}{\varepsilon_s+\varepsilon_n}}\Gamma\left(0,\frac{\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}}{\varepsilon_s+\varepsilon_n}\right). \quad (3.114)$$

We determined the integral in Eq. (3.113) by using Mathematica; here $\Gamma(0,t)$ is an incomplete gamma function, defined as $\Gamma(0,t) = \int_t^\infty u^{-1}e^{-u}\,du$. Hence,

$$H(Y|X) \leq \frac{\varepsilon_n}{\varepsilon_s+\varepsilon_n} H_{\mathrm{Geom}}(\varepsilon_n) + \frac{\varepsilon_s}{\varepsilon_s+\varepsilon_n}\left(\frac{1}{2}\log 2\pi e + \frac{1}{2}\log\left(\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}\right)\times\right.$$
$$\left.\times\, e^{\frac{\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}}{\varepsilon_s+\varepsilon_n}}\Gamma\left(0, \frac{\varepsilon_n(1+\varepsilon_n)+\frac{1}{12}}{\varepsilon_s+\varepsilon_n}\right)\right). \tag{3.115}$$

□

3.6.4 Numerical Results and Discussion

In this section, we compute the upper and lower bounds which we have derived in previous pages. We start by evaluating the difference between the upper and lower bounds from Theorems 3.19, 3.21, and 3.23. The quantity

$$\min\big(\mathrm{C_G}(\varepsilon_s,\varepsilon_n), \mathrm{C_P}(\varepsilon_s)\big) - \mathrm{C_{LB}}(\varepsilon_s,\varepsilon_n) \tag{3.116}$$

is plotted in Fig. 3.5 as a function of ε_s for several values of ε_n, viz. 1, 100, 1000, and 10^6 quanta. In all cases, the difference between upper and lower bounds is at most 1.1 bits, a relatively small value. At low ε_s, the gap converges to about 0.6 bits (depicted in Fig. 3.6 as a horizontal line), which is the difference between the Gaussian bound to the output differential entropy $H(Y|X=x)$, assuming a variance $\varepsilon_s + \varepsilon_n(1+\varepsilon_n) \simeq \varepsilon_n^2$, and the differential entropy of exponential noise, that is,

$$\frac{1}{2}\log(2\pi e\,\varepsilon_n^2) - \log(e\varepsilon_n) = \frac{1}{2}\log\frac{2\pi}{e} \simeq 0.60441\,\text{bits}. \tag{3.117}$$

For high values of ε_n the gap gets close to this asymptotic value (see especially the curve for $\varepsilon_n = 10^6$). Most of the loss is likely caused by a pessimistic estimate of the conditional output entropy $H(Y|X)$, which may be closer to the additive entropy $H(Y)$ than we have been able to prove. If this were the case the true channel capacity $\mathrm{C}(\varepsilon_s,\varepsilon_n)$ would be closely given by the upper bound in Theorem 3.19.

At high ε_s the gap seems to converge to about 0.18 bits (shown in Fig. 3.6 as a horizontal line), the same gap between the capacity of the discrete-time Poisson channel and the mutual information with exponential input. A different

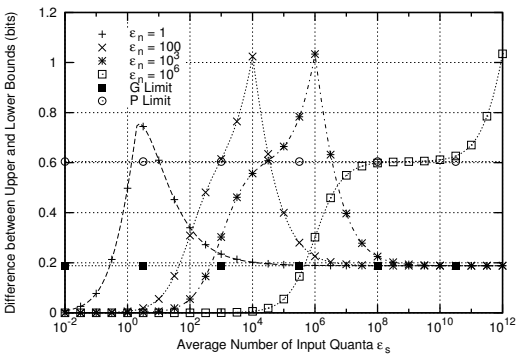


Figure 3.6: Difference between upper and lower bounds to the channel capacity.

input, such as a gamma density with $\nu = 1/2$, the density used in Section 3.5 on the capacity of the DTP channel, should close this gap.

Figure 3.7 depicts the various upper bounds for several values of ε_n: 1, 100, 10000, and 10^6 quanta. The AEN approximation is mentioned in Corollary 3.20, and is given by $\log_2\big(1 + \frac{\varepsilon_s}{\varepsilon_n}\big)$. It is indistinguishable from the upper bound $\mathrm{C_G}(\varepsilon_s, \varepsilon_n)$ for large values of ε_n, and is very close to it for $\varepsilon_n = 1$.

In line with the previous discussion, the upper and lower bounds are close for all cases, differing by at most 1.1 bits, as mentioned previously. We next estimate the crossing point between the two forms of the upper bound. For the highest values of ε_s depicted in Fig. 3.7 the double approximation $\varepsilon_n \gg 1$ and $\varepsilon_s \gg 1$ holds true and we can use the asymptotic forms of the upper bounds to the capacity, respectively given by Eq. (3.77) and Eq. (3.68),

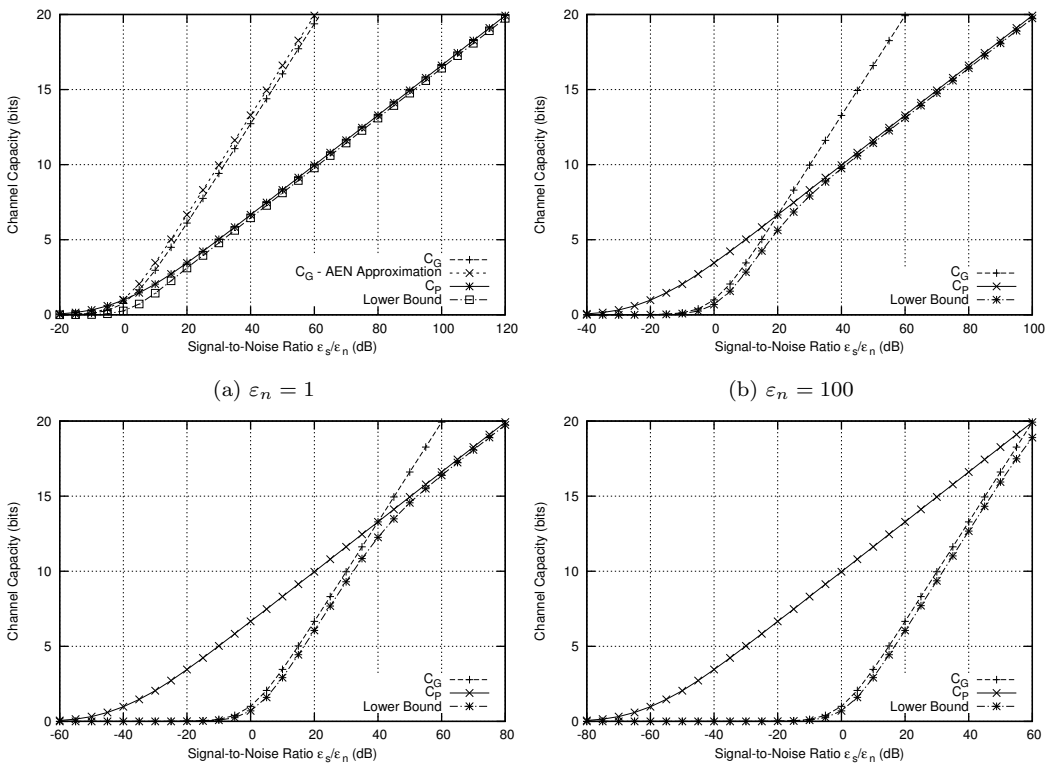
$$\mathrm{C_G}(\varepsilon_s, \varepsilon_n) \simeq \log\left(1 + \frac{\varepsilon_s}{\varepsilon_n}\right), \quad \mathrm{C_P}(\varepsilon_s) \simeq \frac{1}{2}\log(\varepsilon_s). \tag{3.118}$$

With the additional assumption that $\varepsilon_s \gg \varepsilon_n$, at the crossing point

$$\frac{\varepsilon_s^2}{\varepsilon_n^2} \simeq \varepsilon_s. \tag{3.119}$$

Equivalently, there is a threshold signal-to-noise ratio of the underlying AWGN and AEN channels,

$$\mathrm{SNR}^* = \frac{E_s}{E_n} = \frac{\varepsilon_0 \varepsilon_s}{\varepsilon_0 \varepsilon_n} \simeq \varepsilon_n, \tag{3.120}$$



(a) $\varepsilon_n = 1$

(b) $\varepsilon_n = 100$

(c) $\varepsilon_n = 10^4$

(d) $\varepsilon_n = 10^6$

Figure 3.7: Upper bounds to the capacity for several values of ε_n.

such that the behaviour is limited by additive geometric noise below it (i. e. it falls in the G region), and by the Poisson, signal-dependent noise above it (i. e. the P region). The position of the threshold roughly corresponds to input energy value for which the Poisson noise variance $\mathrm{Var}(S) = \varepsilon_s$ coincides with the additive noise variance $\mathrm{Var}(Z)$.

For the upper plots in Fig. 3.7 the position of the threshold is also given by the expression (3.119), even though the amount of additive noise is not large.

Photons as Quanta of Energy We conclude our discussion by applying the results of the AE-Q channel capacity to a channel where photons are the unit of energy. The details of this correspondence were given in Section 2.6, on page 21, and summarized in Table 2.2 therein. For a fixed frequency ν, one can identify the variables ε_s and ε_n with respectively the energy of the useful signal and of thermal noise in a physical channel which makes use of electromagnetic radiation to convey information. Moreover, ε_n is given by Eq. (2.24), the average number of thermal photons at a given temperature.

More concretely, the four cases presented in Fig. 3.7 correspond to frequencies 4 THz ($\varepsilon_n = 1$), 60 GHz ($\varepsilon_n = 100$), 600 MHz ($\varepsilon_n = 10^4$), and 6 MHz ($\varepsilon_n = 10^6$). For each of these cases the threshold signal-to-noise ratio SNR^* has the respective values 0 dB, 20 dB, 40 dB, and 60 dB, as corresponds to using Eq. (3.120),

$$\mathrm{SNR}^* = \varepsilon_n = \left(e^{\frac{h\nu}{k_B T_0}} - 1\right)^{-1} \simeq \frac{k_B T_0}{h\nu} \simeq \frac{6 \cdot 10^{12}}{\nu}. \tag{3.121}$$

In decibels, $\mathrm{SNR}^*\,(\mathrm{dB}) \simeq 37.8 - 10\log_{10}\nu$, where the frequency is in GHz.

Below threshold, the capacity of the AE-Q channel is closely given by

$$\mathrm{C}(\varepsilon_s, \varepsilon_n) \simeq \log\left(1 + \frac{\varepsilon_s}{\varepsilon_n}\right) \simeq \log\left(1 + \frac{E_s}{k_B T_0}\right), \tag{3.122}$$

where E_s is the average signal energy and $k_B T_0$ the one-sided noise spectral density of an equivalent AWGN channel.

Above the threshold SNR^* the slope of the capacity as a function of $\log(\varepsilon_s)$ changes to the value 1/2. An intriguing connection, which is worthwhile mentioning, can be made with non-coherent communications in Gaussian channels [23], where one of the two signal quadratures is not used, and a similar change in slope in the capacity takes place. A similar effect appears in phase-noise limited channels [45].

3.7 Conclusions

In this chapter we have determined the channel capacity of several additive energy channels. First, we have seen that the capacity of the complex-valued additive Gaussian noise channel coincides with that of the additive exponential noise channel with identical levels of signal and additive noise.

Second, we have obtained some new results on the capacity of the discrete-time Poisson (DTP) channel. We have derived what seems to be the tightest

known upper bound. An important tool in this analysis has been a simple method to upper bound the capacity as function of an input distribution. In addition, we have also derived closed-form proper integral expressions for the entropy of Poisson and negative binomial distributions, as well as for the average mutual information of a gamma input in the discrete-time Poisson channel. This analysis has been published in [46].

As we discussed in Chapter 2, the AE-Q channel is an intermediate model between the DTP and AEN channels. At one extreme, the additive exponential noise channel may be seen as a continuous version of the AE-Q. Next to the difference in the output alphabet (that is integer, instead of real-valued output), a more important effect is the absence or presence of Poisson (signal-dependent) noise, besides the additive noise component. In the AE-Q channel, as in the DTP channel, noise is partly signal-dependent. When this Poisson noise is taken into account, we have seen that two regimes appear,

- the G regime, where capacity is limited by additive geometric noise, and
- the P regime, where the limiting factor is Poisson noise.

We have computed the capacity by deriving upper and lower bounds. They differ by (at most) about 1.1 bits, and we expect that the channel capacity is close to the upper bound. In that case, in the G regime, the channel capacity is almost the capacity of the equivalent AWGN or AEN channels,

$$\log\left(1+\frac{\varepsilon_s}{\varepsilon_n}\right), \tag{3.123}$$

whereas in the P regime the capacity is close to the DTP value, $\frac{1}{2}\log\varepsilon_s$.

Finally, we have briefly discussed the analogy with radiation channels, for which the G regime corresponds to the thermal regime, in which additive Gaussian noise is predominant. An unexpected trait of the model is that the condition to lie in the G regime is not only to be at low frequency, as would be expected, but also to have signal-noise ratio below a threshold. Poisson noise is macroscopically visible when it prevails over additive noise, an effect which can arise in one of two ways: by reducing the additive noise level, i. e. by increasing the frequency or lowering the background temperature, or by increasing the signal level and so indirectly the variance of the signal itself, which is proportional to the mean (for a Poisson distribution).

For the frequencies commonly used in radio or wireless communications, the threshold is located at above 30 dB, and therefore the capacity of the AE-Q and

AWGN channels would coincide for all practical effects. However, for higher frequencies or larger values of signal-to-noise ratio, a discrepancy between the two models would appear. An interesting question is to establish the extent to which the AE-Q represents a real limit to the capacity of a radio or wireless communication link. Needless to say, at optical frequencies the AE-Q model collapses into a discrete-time Poisson channel, widely considered as a valid model for communication under the so-called direct detection.

Preliminary results on the analysis of the AE-Q channel were presented at the International Symposium on Information Theory (ISIT) 2006 [47], and at ISIT 2007 [48]. A paper is to be submitted to the IEEE Transactions on Information Theory.

3.A An Upper Bound to the Channel Capacity

Let $p_{\text{in}}(x)$ be a density on the input alphabet, satisfying the input constraint,

$$\int \varepsilon(x) p_{\text{in}}(x)\, dx \leq E. \tag{3.124}$$

The density $p_{\text{in}}(x)$ induces a corresponding output density, $p_1(y)$, $p_1(y) = \int p_{\text{in}}(x) Q(y|x)\, dx$, and achieves a mutual information

$$I(X_{\text{in}}; Y_1) = \iint p_{\text{in}}(x) Q(y|x) \log \frac{Q(y|x)}{p_1(y)}\, dx\, dy, \tag{3.125}$$

which in general will be smaller than the capacity $\mathrm{C}(E)$.

The right-hand side in Eq. (3.9) satisfies

$$\max_x \left(\int Q(y|x) \log \frac{Q(y|x)}{p(y)}\, dy - \gamma(\varepsilon(x) - E) \right) = \tag{3.126}$$

$$= \int p_{\text{in}}(x') \left(\max_x \left(\int Q(y|x) \log \frac{Q(y|x)}{p(y)}\, dy - \gamma(\varepsilon(x) - E) \right) \right) dx' \tag{3.127}$$

$$\geq \int p_{\text{in}}(x') \left(\int Q(y|x') \log \frac{Q(y|x')}{p(y)}\, dy - \gamma(\varepsilon(x') - E) \right) dx' \tag{3.128}$$

$$\geq \int p_{\text{in}}(x') \int Q(y|x') \log \frac{Q(y|x')}{p(y)}\, dy\, dx', \tag{3.129}$$

where Eq. (3.127) uses that $\int p_{\text{in}}(x')\, dx' = 1$ and x' is a dummy variable. When obtaining Eq. (3.128) we used that the integral becomes smaller by replacing

$\max_x f(x)$ by $f(x')$ at each x', as $f(x') \leq \max_x f(x)$. Finally Eq. (3.129) follows since $\gamma \geq 0$ and the input density $p_{\text{in}}(x)$ satisfies the constraint Eq. (3.124).

We continue by proving that the right-hand side in Eq. (3.129) is an upper bound to $I(X_{\text{in}}; Y_1)$. Indeed, cancelling common terms,

$$\int p_{\text{in}}(x') \int Q(y|x') \log \frac{Q(y|x')}{p(y)}\, dy\, dx' - \iint p_{\text{in}}(x) Q(y|x) \log \frac{Q(y|x)}{p_1(y)}\, dx\, dy$$

$$= \int p_{\text{in}}(x') \int Q(y|x') \log \frac{p_1(y)}{p(y)}\, dy\, dx' \tag{3.130}$$

$$= \int p_1(y) \log \frac{p_1(y)}{p(y)}\, dy \geq 0. \tag{3.131}$$

Here we used the definition of $p_1(y)$ and the fact that the divergence between two densities is non-negative.

We have therefore proved that

$$\max_x \left(\int Q(y|x) \log \frac{Q(y|x)}{p(y)}\, dy - \gamma(\varepsilon(x) - E) \right) \geq I(X_{\text{in}}; Y_1) \tag{3.132}$$

for all input densities which satisfy the input constraint. Since this holds for all densities, it must also hold for the supremum, the channel capacity $\mathrm{C}(E)$.

3.B Entropy of a Poisson Distribution

Using the value of $P_S(s|x) = e^{-x} \frac{x^s}{s!}$ and $\log P_S(s|x) = -x + s \log x - \log s!$, the entropy $H_{\text{Pois}}(x)$ is written as

$$H_{\text{Pois}}(x) = x - x \log x + \sum_{s=0}^{\infty} P_S(s|x) \log s!, \tag{3.133}$$

where we substituted the mean value. Using Eq (3.35) the last summand is

$$\sum_{s=0}^{\infty} P_S(s|x) \log s! = \sum_{s=0}^{\infty} e^{-x} \frac{x^s}{s!} \int_0^1 \left(\frac{1 - u^s}{1 - u} - s \right) \frac{du}{\log u} \tag{3.134}$$

$$= \int_0^1 \left(\frac{1 - e^{-x} e^{xu}}{1 - u} - x \right) \frac{du}{\log u}, \tag{3.135}$$

where we have interchanged the order of summation and integration and added up the various terms in the infinite sum.

3.C Asymptotic Form of $f(x)$

This analysis was carried out in collaboration with Dr. Hennie Ter Morsche, from the Department of Mathematics at Technische Universiteit Eindhoven.

Let I denote the integral

$$I = \int_0^1 \left(x - (1-u)^{\nu-1} \frac{1-e^{-xu}}{u} \right) \frac{du}{\log(1-u)}. \tag{3.136}$$

This integral is derived from the one in Eq. (3.38) by the change of variables $u' = 1-u$. For convenience, let us express I as the sum $I = I_1 + I_2 - I_3 + I_4$, where

$$I_1 = \int_0^{1/x} \left(x - (1-u)^{\nu-1} \frac{1-e^{-xu}}{u} \right) \frac{du}{\log(1-u)}, \tag{3.137}$$

$$I_2 = x \int_{1/x}^1 \frac{du}{\log(1-u)}, \tag{3.138}$$

$$I_3 = \int_{1/x}^1 (1-u)^{\nu-1} \frac{1}{u} \frac{du}{\log(1-u)}, \tag{3.139}$$

$$I_4 = \int_{1/x}^1 (1-u)^{\nu-1} \frac{e^{-xu}}{u} \frac{du}{\log(1-u)}. \tag{3.140}$$

We approximate each of these integrals for large x, determining an expansion up to order x^{-1}.

In I_1, let us change the variable to $t = xu$. Then

$$I_1 = \int_0^1 \left(1 - \left(1 - \frac{t}{x}\right)^{\nu-1} \frac{1-e^{-t}}{t} \right) \frac{dt}{\log\left(1 - \frac{t}{x}\right)}. \tag{3.141}$$

Using the Taylor expansions of $(1-z)^{\nu-1}$ and $\log^{-1}(1-z)$, given by

$$(1-z)^{\nu-1} = 1 - (\nu-1)z + az^2 + \mathrm{O}(z^3) \tag{3.142}$$

$$\log^{-1}(1-z) = -\frac{1}{z} + \frac{1}{2} + \frac{z}{12} + \mathrm{O}(z^2), \tag{3.143}$$

with $a = \frac{1}{2}(\nu-1)(\nu-2)$, back in Eq. (3.141), we get for I_1 the expression

$$I_1 = \int_0^1 \left(1 - \left(1 - (\nu-1)\frac{t}{x} + a\frac{t^2}{x^2} + \mathrm{O}\big(\tfrac{t^3}{x^3}\big)\right)\frac{1-e^{-t}}{t}\right)\times$$
$$\times\left(-\frac{x}{t} + \frac{1}{2} + \frac{t}{12x} + \mathrm{O}\big(\tfrac{t^2}{x^2}\big)\right)dt \tag{3.144}$$
$$= \int_0^1 \left(-\frac{x}{t} + \frac{1}{2} + \frac{t}{12x}\right)\left(1 - \frac{1-e^{-t}}{t}\right)dt + \mathrm{O}\big(\tfrac{1}{x^2}\big) +$$
$$+ \int_0^1 \left((\nu-1)\frac{t}{x}\frac{1-e^{-t}}{t}\right)\left(-\frac{x}{t} + \frac{1}{2}\right)dt + \mathrm{O}\big(\tfrac{1}{x^2}\big) +$$
$$- \int_0^1 \left(a\frac{t^2}{x^2}\frac{1-e^{-t}}{t}\right)\left(-\frac{x}{t}\right)dt + \mathrm{O}\big(\tfrac{1}{x^2}\big) \tag{3.145}$$
$$= -x\int_0^1 \frac{1}{t}\left(1 - \frac{1-e^{-t}}{t}\right)dt + \frac{1}{2} + \big(\nu - \tfrac{1}{2}\big)\int_0^1 \left(\frac{1-e^{-t}}{t}\right)dt +$$
$$+ \frac{1}{x}\left(\tfrac{1}{24} + \tfrac{1}{e}\left(-\tfrac{1}{12} + \tfrac{\nu-1}{2} + a\right)\right) + \mathrm{O}\big(\tfrac{1}{x^2}\big). \tag{3.146}$$

Here we used that $\int_0^1 (1-e^{-t})\,dt = e^{-1}$.

We now move on to I_2. By adding and subtracting a term u^{-1} to the integrand, the integral is computed as

$$I_2 = x\int_{1/x}^1 \left(\frac{1}{\log(1-u)} + \frac{1}{u} - \frac{1}{u}\right)du \tag{3.147}$$
$$= x\int_{1/x}^1 \left(\frac{1}{\log(1-u)} + \frac{1}{u}\right)du - x\int_{1/x}^1 \frac{1}{u}\,du \tag{3.148}$$
$$= x\int_{1/x}^1 \left(\frac{1}{\log(1-u)} + \frac{1}{u}\right)du - x\log x. \tag{3.149}$$

Now, we extend the lower integration limit to 0, by writing

$$\int_{1/x}^1 \left(\frac{1}{\log(1-u)} + \frac{1}{u}\right)du = \int_0^1 \left(\frac{1}{\log(1-u)} + \frac{1}{u}\right)du$$
$$- \int_0^{1/x} \left(\frac{1}{\log(1-u)} + \frac{1}{u}\right)du, \tag{3.150}$$

a valid operation since, as $u \to 0$, we deduce from Eq. (3.143) that

$$\left(\frac{1}{\log(1-u)} + \frac{1}{u}\right) = \frac{1}{2} + \frac{u}{12} + \mathrm{O}(u^2). \tag{3.151}$$

Hence, the contribution of the second integral can be evaluated asymptotically,

$$x\int_{1/x}^{1}\left(\frac{1}{\log(1-u)} + \frac{1}{u}\right)du = x\int_{0}^{1}\left(\frac{1}{\log(1-u)} + \frac{1}{u}\right)du - \frac{1}{2} - \frac{1}{24x} + \mathrm{O}\big(\tfrac{1}{x^2}\big), \tag{3.152}$$

and we obtain the expression for I_2,

$$I_2 = x\int_{0}^{1}\left(\frac{1}{\log(1-u)} + \frac{1}{u}\right)du - \frac{1}{2} - \frac{1}{24x} - x\log x + \mathrm{O}\big(\tfrac{1}{x^2}\big). \tag{3.153}$$

We move on to I_3. We add and subtract a term $\frac{1}{u} - \frac{1}{2}$ to the function $\frac{1}{\log(1-u)}$ in the integrand, and write

$$\begin{aligned} I_3 = &\int_{1/x}^{1}(1-u)^{\nu-1}\left(\frac{1}{u\log(1-u)} + \frac{1}{u^2} - \frac{1}{2u}\right)du \\ &- \int_{1/x}^{1}(1-u)^{\nu-1}\left(\frac{1}{u^2} - \frac{1}{2u}\right)du. \end{aligned} \tag{3.154}$$

Since

$$\lim_{u\to 0}\left((1-u)^{\nu-1}\cdot\left(\frac{1}{u\log(1-u)} + \frac{1}{u^2} - \frac{1}{2u}\right)\right) = \frac{1}{12} + \mathrm{O}(u), \tag{3.155}$$

we can shift the lower integration limit to 0 in the first integral. This is compensated by subtracting the integral

$$\int_{0}^{1/x}(1-u)^{\nu-1}\left(\frac{1}{u\log(1-u)} + \frac{1}{u^2} - \frac{1}{2u}\right)du = \frac{1}{12x} + \mathrm{O}\big(\tfrac{1}{x^2}\big), \tag{3.156}$$

where the error term is $\mathrm{O}\big(\frac{1}{x^2}\big)$. The resulting integral between 0 and 1 does not depend on x. The second term, which we denote by $I_{3'}$, is given by

$$I_{3'} = -\int_{1/x}^{1}(1-u)^{\nu-1}\left(\frac{1}{u^2} - \frac{1}{2u}\right)du. \tag{3.157}$$

By adding and subtracting a function $1-(\nu-1)u$, we rewrite $I_{3'}$ as

$$I_{3'} = -\int_{1/x}^{1}\Big((1-u)^{\nu-1}-1+(\nu-1)u\Big)\left(\frac{1}{u^2}-\frac{1}{2u}\right)du$$
$$-\int_{1/x}^{1}\Big(1-(\nu-1)u\Big)\left(\frac{1}{u^2}-\frac{1}{2u}\right)du. \tag{3.158}$$

Again, since

$$\Big((1-u)^{\nu-1}-1+(\nu-1)u\Big)\left(\frac{1}{u^2}-\frac{1}{2u}\right) = a+\mathrm{O}(u) \tag{3.159}$$

around $u=0$, we can shift the lower integration order to 0. In doing so, we must include the summand

$$\int_0^{1/x}\Big((1-u)^{\nu-1}-1+(\nu-1)u\Big)\left(\frac{1}{u^2}-\frac{1}{2u}\right)du = \frac{a}{x}+\mathrm{O}\big(\tfrac{1}{x^2}\big). \tag{3.160}$$

The resulting integral between 0 and 1 does not depend on x.

The remaining integral in the evaluation of I_3 is evaluated as

$$-\int_{1/x}^{1}\left(\frac{1}{u^2}-\frac{1}{u}\Big(\nu-\frac{1}{2}\Big)+\frac{\nu-1}{2}\right)du = \frac{1}{u}+\left(\nu-\frac{1}{2}\right)\log u-\frac{\nu-1}{2}u\Big|_{1/x}^{1} \tag{3.161}$$

$$= 1-\frac{\nu-1}{2}-x+\left(\nu-\frac{1}{2}\right)\log x+\frac{\nu-1}{2x}. \tag{3.162}$$

Hence, collecting the previous expressions, I_3 admits the asymptotic expression

$$I_3 = \int_0^1 (1-u)^{\nu-1}\left(\frac{1}{u\log(1-u)}+\frac{1}{u^2}-\frac{1}{2u}\right)du+$$
$$-\int_0^1\Big((1-u)^{\nu-1}-1+(\nu-1)u\Big)\left(\frac{1}{u^2}-\frac{1}{2u}\right)du+\frac{3-\nu}{2}+$$
$$-x+\left(\nu-\frac{1}{2}\right)\log x+\frac{1}{x}\left(-\frac{1}{12}+a+\frac{\nu-1}{2}\right)+\mathrm{O}\big(\tfrac{1}{x^2}\big). \tag{3.163}$$

Finally, we split I_4 into two parts, $I_{4'}$ and $I_{4''}$, respectively given by

$$I_{4'} = \int_{1/x}^{1/2}(1-u)^{\nu-1}\frac{e^{-xu}}{u}\frac{du}{\log(1-u)} \tag{3.164}$$

$$I_{4''} = \int_{1/2}^{1}(1-u)^{\nu-1}\frac{e^{-xu}}{u}\frac{du}{\log(1-u)}. \tag{3.165}$$

In the first one, $I_{4'}$ let us add and subtract a term $\frac{1}{u} - \frac{1}{2}$, to obtain

$$I_{4'} = \int_{1/x}^{1/2} (1-u)^{\nu-1} \frac{e^{-xu}}{u} \left(\frac{1}{\log(1-u)} + \frac{1}{u} - \frac{1}{2} \right) du$$
$$- \int_{1/x}^{1/2} (1-u)^{\nu-1} \frac{e^{-xu}}{u} \left(\frac{1}{u} - \frac{1}{2} \right) du. \tag{3.166}$$

The first summand of $I_{4'}$, of value

$$\int_{1/x}^{1/2} (1-u)^{\nu-1} \frac{e^{-xu}}{u} \left(\frac{1}{\log(1-u)} + \frac{1}{u} - \frac{1}{2} \right) du, \tag{3.167}$$

has the form $\int_{1/x}^{1/2} g(u) e^{-xu}\, du$, where, using Eq. (3.143),

$$g(u) = (1-u)^{\nu-1} \left(\frac{1}{u \log(1-u)} + \frac{1}{u^2} - \frac{1}{2u} \right) = \frac{1}{12} + \mathrm{O}(u). \tag{3.168}$$

Replacing this expansion back in the integral, we obtain

$$\int_{1/x}^{1/2} e^{-xu} \left(\frac{1}{12} + \mathrm{O}(u) \right) du = \frac{1}{12ex} + \mathrm{O}(e^{-\frac{x}{2}}) + \int_{1/x}^{1/2} e^{-xu}\, \mathrm{O}(u)\, du \tag{3.169}$$
$$= \frac{1}{12ex} + \mathrm{O}\big(\tfrac{1}{x^2}\big), \tag{3.170}$$

where we used that

$$\int_{1/x}^{1/2} e^{-xu} u^n\, du = \mathrm{O}\big(\tfrac{1}{x^{n+1}}\big), \tag{3.171}$$

for large x and n integer.

Changing variables, to $t = xu$, we get for the second contribution to $I_{4'}$

$$-x \int_1^{x/2} \left(1 - \frac{t}{x}\right)^{\nu-1} \frac{e^{-t}}{t^2} dt + \frac{1}{2} \int_1^{x/2} \left(1 - \frac{t}{x}\right)^{\nu-1} \frac{e^{-t}}{t} dt. \tag{3.172}$$

The Taylor expansion in Eq. (3.142) gives for the first of these summands

$$
\begin{aligned}
&-x\int_1^{x/2}\left(1-(\nu-1)\frac{t}{x}+a\frac{t^2}{x^2}+\mathrm{O}\big(\tfrac{t^3}{x^3}\big)\right)\frac{e^{-t}}{t^2}dt = \\
&\quad = -x\int_1^{x/2}\frac{e^{-t}}{t^2}dt+(\nu-1)\int_1^{x/2}\frac{e^{-t}}{t}dt-\frac{a}{x}\int_1^{x/2}e^{-t}dt-\int_1^{x/2}\mathrm{O}\big(\tfrac{t}{x^2}\big)e^{-t}dt \qquad (3.173)\\
&\quad = -x\int_1^{\infty}\frac{e^{-t}}{t^2}dt+\int_1^{\infty}(\nu-1)\frac{e^{-t}}{t}dt-\frac{a}{ex}+\mathrm{O}\big(\tfrac{1}{x^2}\big), \qquad (3.174)
\end{aligned}
$$

where we replaced the upper integration order by ∞, which introduced an exponentially small error term $\mathrm{O}(e^{-x})$, and used Eq. (3.171). Similarly, the second summand in Eq. (3.172) can be expressed as

$$
\begin{aligned}
&\frac{1}{2}\int_1^{x/2}\left(1-(\nu-1)\frac{t}{x}+\mathrm{O}\big(\tfrac{t^2}{x^2}\big)\right)\frac{e^{-t}}{t}dt = \\
&\qquad = \frac{1}{2}\int_1^{x/2}\frac{e^{-t}}{t}dt+\frac{1}{2x}\int_1^{x/2}(\nu-1)e^{-t}dt+\int_1^{x/2}\mathrm{O}\big(\tfrac{t}{x^2}\big)e^{-t}dt \qquad (3.175)\\
&\qquad = \frac{1}{2}\int_1^{\infty}\frac{e^{-t}}{t}dt-\frac{1}{2ex}(\nu-1)+\mathrm{O}\big(\tfrac{1}{x^2}\big), \qquad (3.176)
\end{aligned}
$$

where we also replaced the upper integration order by ∞, with the corresponding error term $\mathrm{O}(e^{-x})$.

The last remaining term is $I_{4''}$,

$$
I_{4''} = \int_{1/2}^{1}(1-u)^{\nu-1}\frac{e^{-xu}}{u}\frac{du}{\log(1-u)}. \qquad (3.177)
$$

The magnitude of its contribution $|I_{4''}|$ is upper bounded by

$$
|I_{4''}| = e^{-\frac{x}{2}}\left|\int_{1/2}^{1}(1-u)^{\nu-1}\frac{e^{-xu}}{u}\frac{du}{\log(1-u)}\right|, \qquad (3.178)
$$

since the function is the integrand is regular, and the exponential function decays with increasing u. Since the integral exists, we conclude that

$$
I_{4''} = \mathrm{O}\big(e^{-\frac{x}{2}}\big). \qquad (3.179)
$$

Finally, combining Eqs. (3.146), (3.153), (3.163), (3.170) (3.174), (3.176), and Eq. (3.179), the value of I is given by

$$\begin{aligned}
I = &-x\int_0^1\left(\frac{1}{t}-\frac{1}{t^2}+\frac{e^{-t}}{t^2}\right)dt-\frac{3-\nu}{2}+\frac{1}{2}-\left(\nu-\tfrac{1}{2}\right)\int_0^1\left(\frac{1-e^{-t}}{t}\right)dt\\
&+x\int_0^1\left(\frac{1}{\log(1-u)}+\frac{1}{u}\right)du-\frac{1}{2}-x\log x-\left(\nu-\frac{1}{2}\right)\log x\\
&-\int_0^1(1-u)^{\nu-1}\left(\frac{1}{u\log(1-u)}+\frac{1}{u^2}-\frac{1}{2u}\right)du\\
&+\int_0^1\Big((1-u)^{\nu-1}-1+(\nu-1)u\Big)\left(\frac{1}{u^2}-\frac{1}{2u}\right)du+x-x\int_1^\infty\frac{e^{-t}}{t^2}dt\\
&+\int_1^\infty\left(\nu-\tfrac{1}{2}\right)\frac{e^{-t}}{t}dt+\frac{1}{x}\left(\frac{1}{12}-a-\frac{\nu-1}{2}\right)+\mathrm{O}\big(\tfrac{1}{x^2}\big). \qquad (3.180)
\end{aligned}$$

Here we cancelled the common terms in the coefficient of x^{-1}. Further, recall that $a=\frac{1}{2}(\nu-1)(\nu-2)$.

Further simplification is brought about by the facts that

$$\int_0^1\left(-\frac{1}{\log(1-t)}-\frac{1}{t^2}+\frac{e^{-t}}{t^2}\right)dt+\int_1^\infty\frac{e^{-t}}{t^2}dt= \qquad (3.181)$$

$$=\int_0^1\left(-\frac{1}{\log(1-t)}-\frac{1}{t^2}+\frac{e^{-t}}{t^2}+e^{-1/t}\right)dt=0 \qquad (3.182)$$

$$\int_0^1\frac{1-e^{-t}}{t}dt-\int_1^\infty\frac{e^{-t}}{t}dt=\int_0^1\frac{1-e^{-t}-e^{-1/t}}{t}dt=\gamma_e, \qquad (3.183)$$

which, together with some straightforward combinations, implies that

$$\begin{aligned}
I &= -\big(\nu-\tfrac{1}{2}\big)\gamma_e-x\log x-\frac{3-\nu}{2}+x-\big(\nu-\tfrac{1}{2}\big)\log x+\frac{1}{x}\left(\frac{1}{12}-\frac{(\nu-1)^2}{2}\right)\\
&\quad-\int_0^1\left(\frac{(1-u)^{\nu-1}}{u\log(1-u)}+\Big(1-(\nu-1)u\Big)\left(\frac{1}{u^2}-\frac{1}{2u}\right)\right)du+\mathrm{O}\big(\tfrac{1}{x^2}\big) \qquad (3.184)\\
&= -\big(\nu-\tfrac{1}{2}\big)\gamma_e-x\log x-1+x-\big(\nu-\tfrac{1}{2}\big)\log x+\frac{1}{x}\left(\frac{1}{12}-\frac{(\nu-1)^2}{2}\right)\\
&\quad-\int_0^1\left(\frac{(1-u)^{\nu-1}}{u\log(1-u)}+\frac{1}{u^2}-\frac{1}{u}\big(\nu-\tfrac{1}{2}\big)\right)du+\mathrm{O}\big(\tfrac{1}{x^2}\big). \qquad (3.185)
\end{aligned}$$

3.D Entropy of a Negative Binomial Distribution

The moment generating function $\mathrm{E}[e^{tX}]$ of a variable X with gamma distribution, denoted by $\mathrm{mgf}_{\mathcal{G}(\nu,\varepsilon_s)}(t)$, and the probability generating function $\mathrm{E}[u^X]$ of a random variable X with negative binomial distribution, denoted by $\mathrm{pgf}_{\mathrm{NegBin}(\nu,\varepsilon_s)}(u)$, are respectively given by

$$\mathrm{mgf}_{\mathcal{G}(\nu,\varepsilon_s)}(t) = \frac{\nu^\nu}{(\nu - \varepsilon_s t)^\nu} \tag{3.186}$$

$$\mathrm{pgf}_{\mathrm{NegBin}(\nu,\varepsilon_s)}(u) = \frac{\nu^\nu}{(\nu + \varepsilon_s(1-u))^\nu} = \mathrm{mgf}_{\mathcal{G}(\nu,\varepsilon_s)}(u-1). \tag{3.187}$$

By definition the entropy is the sum

$$H_{\mathrm{NegBin}}(\nu,\varepsilon_s) = -\sum_{y=0}^{\infty} P_Y(y)\log P_Y(y) \tag{3.188}$$

$$= -\sum_{y=0}^{\infty} P_Y(y)\left(\log\frac{\Gamma(y+\nu)}{y!\Gamma(\nu)} + \log\frac{\nu^\nu}{(\varepsilon_s+\nu)^\nu} + y\log\frac{\varepsilon_s}{\varepsilon_s+\nu}\right) \tag{3.189}$$

$$= -\log\frac{\nu^\nu}{(\varepsilon_s+\nu)^\nu} - \varepsilon_s\log\frac{\varepsilon_s}{\varepsilon_s+\nu} - \sum_{y=0}^{\infty} P_Y(y)\left(\log\frac{\Gamma(y+\nu)}{y!\Gamma(\nu)}\right) \tag{3.190}$$

$$= (\varepsilon_s+\nu)\log(\varepsilon_s+\nu) - \nu\log\nu - \varepsilon_s\log\varepsilon_s - \sum_{y=0}^{\infty} P_Y(y)\left(\log\frac{\Gamma(y+\nu)}{y!\Gamma(\nu)}\right). \tag{3.191}$$

We first expand the logarithm using Eq. (3.33),

$$\log\frac{\Gamma(y+\nu)}{y!\Gamma(\nu)} = \int_0^1 \left(\frac{1-u^{y+\nu-1}-1+u^y-1+u^{\nu-1}}{1-u}\right)\frac{du}{\log u} \tag{3.192}$$

$$= -\int_0^1 \left(\frac{(1-u^{\nu-1})(1-u^y)}{1-u}\right)\frac{du}{\log u}. \tag{3.193}$$

The last summand in Eq. (3.191) then equals

$$\sum_{y=0}^{\infty} P_Y(y)\left(\int_0^1 \left(\frac{(1-u^{\nu-1})(1-u^y)}{1-u}\right)\frac{du}{\log u}\right) \tag{3.194}$$

$$= \int_0^1 \left(\frac{(1-u^{\nu-1})(1-\mathrm{pgf}(u))}{1-u}\right)\frac{du}{\log u}. \tag{3.195}$$

3.E Mutual Information for a Gamma Density Input

The mutual information $I(X_{\mathcal{G}(\nu,\varepsilon_s)};Y)$ is the sum $I(X_{\mathcal{G}(\nu,\varepsilon_s)};Y) = H(Y) - H(Y|X)$. The output entropy $H(Y)$ is directly given by Eq. (3.44), since the output is distributed according to a negative binomial distribution. As for the conditional entropy $H(Y|X)$, from its very definition we note that it is a function of the entropy of a Poisson random variable, given by Eq. (3.36),

$$H(Y|X) = \int_0^\infty p_X(x) H_{\text{Pois}}(x)\, dx \tag{3.196}$$

$$= \int_0^\infty p_X(x)\left(x - x\log x + \int_0^1 \left(\frac{1-e^{-x(1-u)}}{1-u} - x\right)\frac{du}{\log u}\right) dx \tag{3.197}$$

$$= \varepsilon_s + \int_0^1 \left(\frac{1-\text{mgf}(u-1)}{1-u} - \varepsilon_s\right)\frac{du}{\log u} - \int_0^\infty p_X(x) x \log x\, dx. \tag{3.198}$$

The remaining integral can be evaluated in terms of the digamma function [49],

$$\int_0^\infty p_X(x) x \log x\, dx = \varepsilon_s \psi(\nu+1) - \varepsilon_s \log\frac{\nu}{\varepsilon_s}. \tag{3.199}$$

Putting all elements together and grouping the obvious terms gives the desired expression.

3.F Computation of the Function $\kappa(x)$

By construction, the function $\kappa(x)$ is given by

$$\kappa(x) = -H_{\text{Pois}}(x) - \sum_{y=0}^{\infty} Q(y|x)\log P_\Delta(y) - \gamma(x-\varepsilon_s). \tag{3.200}$$

The second summand includes the expression $\log P_\Delta(y)$, where $P_\Delta(y)$ given by Eq. (3.49), the channel output under the density Eq. (3.48). The second summand is given by

$$\sum_{y=0}^{\infty} Q(y|x)\log P_\Delta(y) = Q(0|x)\log P_\Delta(0) + \sum_{y=1}^{\infty} Q(y|x)\log P_\Delta(y), \tag{3.201}$$

written so as to isolate a term related to the entropy of the negative binomial.

Let us denote the negative binomial distribution as $P(y) = \Pr_{\mathrm{NegBin}(\nu,\varepsilon_s)}(y)$. Adding and subtracting a term $Q(0|x)\log\big((1-\Delta)P(0)\big)$ we have

$$\sum_{y=0}^{\infty} Q(y|x)\log P_\Delta(y) = Q(0|x)\log\frac{P_\Delta(0)}{(1-\Delta)P(0)} + \sum_{y=0}^{\infty} Q(y|x)\log\big((1-\Delta)P(y)\big) \tag{3.202}$$

$$= e^{-x}\log\left(1+\frac{\Delta}{(1-\Delta)}\left(1+\frac{\varepsilon_s}{\nu}\right)^{\nu}\right) + \log(1-\Delta) + \sum_{y=0}^{\infty} Q(y|x)\log\big(P(y)\big), \tag{3.203}$$

where we used that $Q(0|x) = e^{-x}$. The last summand was evaluated in the derivation of Eq. (3.44), included in Appendix 3.D, in Eqs. (3.189) and (3.193). Using these equations we have

$$\sum_{y=0}^{\infty} Q(y|x)\log P(y) = \nu\log\frac{\nu}{\varepsilon_s+\nu} + \sum_{y=0}^{\infty} Q(y|x)y\log\frac{\varepsilon_s}{\varepsilon_s+\nu} - \sum_{y=0}^{\infty} Q(y|x)\int_0^1\left(\frac{(1-u^{\nu-1})(1-u^y)}{1-u}\right)\frac{du}{\log u} \tag{3.204}$$

$$= \nu\log\nu + x\log\varepsilon_s - (x+\nu)\log(\varepsilon_s+\nu) - \int_0^1\left(\frac{(1-u^{\nu-1})(1-e^{-x(1-u)})}{1-u}\right)\frac{du}{\log u}. \tag{3.205}$$

Combining this term with the entropy $H_{\mathrm{Pois}}(x)$ given by Eq. (3.36), cancelling some common terms and rearranging the final form, we have

$$\kappa(x) = -x + x\log x + \int_0^1\left(x - u^{\nu-1}\frac{1-e^{-x(1-u)}}{1-u}\right)\frac{du}{\log u} - e^{-x}\log\left(1+\frac{\Delta}{(1-\Delta)}\left(1+\frac{\varepsilon_s}{\nu}\right)^{\nu}\right) - x\left(\gamma - \log\frac{\varepsilon_s+\nu}{\varepsilon_s}\right) + \nu\log\frac{\varepsilon_s+\nu}{\nu} + \gamma\varepsilon_s - \log(1-\Delta). \tag{3.206}$$

3.G Numerical Evaluation of $\kappa_0(x)$ for Large x

Table 3.1 shows the computed values of $\kappa_0(x)$ for a wide range of input variable x. Since a large precision goal is required for large x, we have set Mathematica to a high precision goal in the internal computations by using the command

```
N[x + x*Log[x]  -
NIntegrate[((1 - Exp[-x*(1-u)])/Sqrt[u]/(1-u) - x)/Log[u],
    {u, 0, 1}, WorkingPrecision -> 240, MinRecursion -> 10,
    PrecisionGoal -> 60, MaxRecursion -> 400], 30]
```

to generate the results reported in the table. As $x \to \infty$, $\kappa_0(x)$ strongly seems to approach $-\frac{1}{2}\log(2e) \simeq -0.846573590279972654708616060 72\ldots$ from below.

x	$\kappa_0(x)$	$-\frac{1}{2}\log(2e) - \frac{1}{24x}$
10^{-3}	-0.007215157306619305332096007285	-42.513240256946639321375282727 40
0.01	-0.049175062876264757290432766019	-5.013240256946639321375282727396
0.1	-0.266340584231272974982946453848	-1.263240256946639321375282727396
1	-0.773051830993302198340618735354	-0.888240256946639321375282727396
10	-0.851269835471555230853156933655	-0.850740256946639321375282727396
100	-0.846994503273299681423249654818	-0.846990256946639321375282727396
10^3	-0.846615298690951927273996819862	-0.846615256946639321375282727396
10^4	-0.846577757363383440063808026741	-0.846577756946639321375282727396
10^6	-0.846573631946680988119380164202	-0.846573631946639321375282727396
10^8	-0.846573590696639325541949471493	-0.846573590696639321375282727396
10^{10}	-0.846573590284139321375699394062	-0.846573590284139321375282727396
10^{12}	-0.846573590280014321375282769062	-0.846573590280014321375282727396
10^{14}	-0.846573590279973071375282727400	-0.846573590279973071375282727396
10^{20}	-0.846573590279972654709032727396	-0.846573590279972654709032727396
10^{25}	-0.846573590279972654708616064896	-0.846573590279972654708616064896

Table 3.1: Evaluation of $\kappa_0(x)$ for large values of x.

4

Digital Modulation in the Gaussian Channel

4.1 Introduction

In the introduction to Chapter 2, where Shannon's model for the communication channel was introduced, we mentioned the existence of an encoder, which transforms the source message into a form appropriate for reliable transmission across a channel. From a systems point of view, the encoder performs two tasks, that can be broadly described as coding and modulation:

1. The encoder adds redundancy to the source message, so that the deleterious effect of channel noise can be overcome. This is the domain of channel coding, which will be treated in some detail in Chapter 6.

2. The encoder transforms the source message into a set of modulation symbols. This set is judiciously chosen by taking into proper account the nature of the channel characteristics and impairments.

When the coding and modulation tasks are jointly performed at both the encoder and the decoder, we talk of coded modulation. Alternatively, for bit-interleaved coded modulation, the two tasks may be carried out sequentially: first, the source message is encoded with a binary code; the encoder output is then mapped onto a set of modulation symbols. At the decoder, the reverse operations are carried out. If we do not wish to distinguish among the two options, we rather use the words digital modulation.

In this chapter, we study digital modulation for the Gaussian channel. Since Gaussian channels are often used to model communication systems operating at radio and microwave frequencies, much previous work has been done in the field, see e. g. the recent reviews [50, 51]. Part of the content of this chapter is

a review of this body of work, as a prelude to deriving similar results on digital modulation for the additive energy channels, in Chapter 5.

Much attention has been paid recently to the performance of coded modulation in the so-called "wideband regime", where signal-to-noise ratio is very low, see for instance [25, 52]. The reasons for this increased attention are twofold: from a practical point of view, many wireless systems operate at low signal-to-noise ratio and theoretical results in this field are of directly applicable. In addition, mathematical analysis is somewhat simpler than in the general case, leading to neat closed-form expressions for some information-theoretic quantities. In this chapter, we shall also study the performance in this regime and derive some new results, described in more detail below. In Chapter 5, we extend our analysis to the additive energy channels.

In Chapter 3, the channel capacity was defined as the largest rate at which messages can be reliably sent over the channel. For some channels, we found the form of the distribution at the channel input which achieves the capacity. Other input distributions also allow for vanishingly small error probabilities, but at a rate, the constrained capacity, that is in general lower than the channel capacity. In practical communication systems, this reduction in rate may be compensated by the ease of design allowed by simple modulation sets, as opposed to using the capacity-achieving input. In Section 4.2 we consider the constrained capacity for some common digital modulation formats, such as phase-shift keying (PSK) or quadrature-amplitude modulation (QAM).

Since the constrained capacity does not admit a simple expression, it often proves convenient to study alternative figures of merit, such as the capacity per unit energy or the minimum energy-to-noise ratio [25, 52]; they approximate the constrained capacity by its second-order Taylor series around zero signal-to-noise ratio. In Sections 4.2.2 and 4.2.3, we review these concepts and provide an independent derivation of the Taylor series, which does not depend on Prelov's previous analysis [52]. The method used for the derivation will prove directly applicable to the additive energy channels in Chapter 5. Another figure of merit, relevant at large signal-to-noise ratios, is the shaping gain, which determines the asymptotic loss incurred by not using the optimum channel input. The shaping gain is considered in Section 4.2.6.

In Section 4.3 we describe bit-interleaved coded modulation, in general terms and particularized to the Gaussian channel, and analytically determine its behaviour at low signal-to-noise ratio.

The results presented in this chapter have been partly published in [53] and a paper has been submitted to the IEEE Transactions on Information Theory [54].

4.2 Coded Modulation in the Gaussian Channel

4.2.1 Constrained Capacity

In Chapters 2 and 3, we considered the discrete-time additive Gaussian noise channel model. Its output is a complex-valued vector $(y_1, \ldots, y_n)$, whose k-th component y_k is given by the sum

$$y_k = \sqrt{\mathrm{SNR}}\, x_k + z_k. \tag{4.1}$$

Here the z_k are independent samples of circularly symmetric complex-valued Gaussian noise of variance 1, $z_k \sim \mathcal{N}_{\mathbf{C}}(0,1)$, and each x_k is k-th the complex-valued channel input, namely a modulation symbol drawn from a set with unit average energy. SNR is the average signal-to-noise ratio at the receiver.

Input symbols are drawn according to a common distribution $P_X(\cdot)$ from a set $\mathcal{X}$ with $|\mathcal{X}| = 2^m$ elements. When the symbols used with equal probabilities, m bits are required to choose a symbol. We define the first and second constellation moments, respectively denoted by $\mu_1(\mathcal{X})$ and $\mu_2(\mathcal{X})$, as

$$\mu_1(\mathcal{X}) = \sum_{x \in \mathcal{X}} x P_X(x), \quad \mu_2(\mathcal{X}) = \sum_{x \in \mathcal{X}} |x|^2 P_X(x). \tag{4.2}$$

Often, constellations have zero mean, i. e. $\mu_1(\mathcal{X}) = 0$, and unit energy, that is $\mu_2(\mathcal{X}) = 1$. Similarly, we define a variance as $\sigma^2(\mathcal{X}) = \mu_2(\mathcal{X}) - \big|\mu_1(\mathcal{X})\big|^2$.

The capacity C of the Gaussian channel was given by Theorem 3.10, $\mathrm{C}(\mathrm{SNR}) = \log(1 + \mathrm{SNR})$. We define the bit-energy-to-noise ratio as

$$\mathrm{BNR} = \frac{\mathrm{SNR}}{\mathrm{C}(\mathrm{SNR})} \log 2. \tag{4.3}$$

namely the average received energy normalized to a bit of information. It should be noted that a bit need not be effectively transmitted, since C may be smaller than 1 bit. This quantity is commonly used to compare systems operating at low signal-to-noise ratio in the Gaussian channel, in part because, as we shall see, most modulation formats asymptotically attain the same BNR at low capacities, easing the performance comparison among different formats.

Even though the common notation for BNR is E_b/N_0, we prefer the alternative symbol BNR because it can be used for the additive energy channels, for which a noise spectral density N_0 is not necessarily defined. Moreover, the symbol BNR maintains the symmetry with respect to the acronym SNR and leads to formulas which are, possibly, easier to read.

In order to achieve the capacity C, the channel input must be a properly chosen complex Gaussian distribution. In practice, however, it is common to fix a finite constellation set for the channel input. Well-known examples [32, 50] are phase-shift keying (PSK), quadrature-amplitude modulation (QAM), amplitude-phase shift keying (APSK), or pulse-amplitude modulation (PAM).

When the symbols in $\mathcal{X}$ are used with probabilities $P_X(x)$, the mutual information $I(X;Y)$ between channel input X and output Y, or constrained coded modulation capacity $\mathrm{C}_{\mathcal{X}}$ for short, is given by Eq. (3.3) on page 27. The name constrained capacity is justified by the fact that the mutual information gives the maximum rate at which information can be reliably transmitted over the channel by using the input the distribution $P_X(x)$, in an analogous manner as the channel capacity gives the maximum rate at which information can be reliably transmitted over the channel for all possible input distributions.

For the Gaussian channel, the channel transition matrix $Q(y|x)$ is given by

$$Q(y|x) = \frac{1}{\pi} \exp\bigl(-|y - \sqrt{\mathrm{SNR}}x|^2\bigr), \tag{4.4}$$

as in Chapter 2, Eq. (2.8) on page 15. We have then

Definition 4.1. For the additive Gaussian noise channel with average signal-to-noise ratio SNR, the constrained capacity $\mathrm{C}_{\mathcal{X}}$ with a modulation set $\mathcal{X}$ used with probabilities $P_X(x)$ is

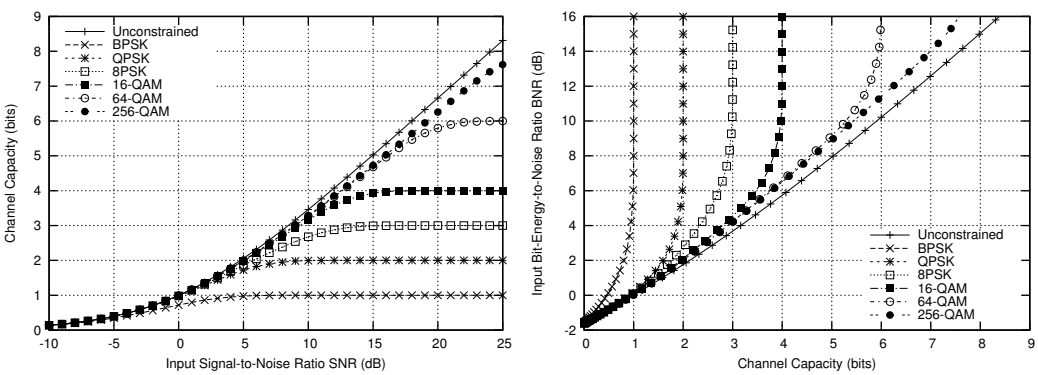
$$\mathrm{C}_{\mathcal{X}}(\mathrm{SNR}) = -\,\mathrm{E}\left[\log\left(\sum_{x'\in\mathcal{X}} P_X(x') \exp\bigl(-|\sqrt{\mathrm{SNR}}(X - x') + Z|^2 + |Z|^2\bigr)\right)\right]. \tag{4.5}$$

The expectation is performed over all input symbols X and all possible noise realizations Z, $Z \sim \mathcal{N}_{\mathbf{C}}(0,1)$.

If the symbols are used with equal probabilities, i. e. $P_X(x) = M^{-1}$, we refer to the constrained capacity as uniform capacity, and denote it by $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}}$.

Figure 4.1 shows the constrained coded modulation capacity for several typical modulations as a function of SNR and of BNR, defined as $\mathrm{BNR} = \frac{\mathrm{SNR}}{\mathrm{C}_{\mathcal{X}}(\mathrm{SNR})} \log 2$. Note that plot 4.1b has the x and y axes reversed with respect to the common practice, as in [50], and so shows BNR as a function of the capacity $\mathrm{C}_{\mathcal{X}}$, rather than the capacity as a function of BNR. Behind this choice lies the convention in plotting a function $f(x)$ which reserves the x axis to present the free variable x and uses the y axis to depict the function $f(x)$. In

Fig. 4.1b the free variable is the capacity $C_{\mathcal{X}}$ (or the signal-to-noise ratio SNR), from which the bit energy-to-noise ratio BNR is derived. If, as it happens for some channels, the inverse function $C_{\mathcal{X}}(\text{BNR})$ can take two values, interpreting the standard depiction is slightly confusing, whereas a plot $\text{BNR}(C_{\mathcal{X}})$ leads to no conceptual complication. Apart from this, the plots are well-known and can be found, for instance, in [50].



(a) Capacity as a function of SNR. (b) Capacity as a function of BNR.

Figure 4.1: Channel capacity (in bits per channel use).

Two traits in Fig. 4.1b deserving special attention are the asymptotic behaviour at low and high signal-to-noise ratio. First, the well-known minimum $\text{BNR} = -1.59\,\text{dB}$ is approached at low BNR for all modulations depicted. Second, before each of the curves attains the asymptotic value $H(X) = m \log 2$ at high BNR, all lines define a gap with respect to the unconstrained capacity C, the so-called shaping gain. In the following pages, we characterize analytically the behaviour in each of these two regimes.

4.2.2 Capacity per Unit Energy

In this section, we rederive Shannon's limit for the minimum bit-energy-to-noise ratio, namely $\text{BNR}_{\min} = -1.59\,\text{dB}$, in terms of the capacity per unit energy. The capacity per unit energy C_1 is the largest number of bits per symbol which can be reliably sent over the channel per unit energy. As found by Verdú [29], the capacity per unit energy in the Gaussian channel and the minimum energy

per bit, obtained from Eq. (3.11), are respectively given by

$$\mathrm{C}_1 = \frac{1}{\sigma^2}, \quad E_{b,\min} = \sigma^2 \log 2. \tag{4.6}$$

Alternatively, C_1 can be determined by using Theorem 3.9 on page 30.

Another equivalent form is the minimum bit-energy-to-noise ratio $\mathrm{BNR}_{\min}$, $\mathrm{BNR}_{\min} = \frac{E_{b,\min}}{\sigma^2} = \log 2$, or -1.59 dB. This quantity is the well-known Shannon limit. As previously mentioned, this result is apparent from Fig. 4.1b.

4.2.3 Asymptotic Behaviour at Low SNR

In this section, we examine the asymptotic behaviour of the constrained capacity $\mathrm{C}_{\mathcal{X}}$ as $\mathrm{SNR} \to 0$. We provide some analytic results which complement the recent study by Verdú of the "wideband, power-limited regime" [25].

Before considering the effect of modulation, let us first discuss the unconstrained case. A Taylor expansion of the capacity C around $\mathrm{SNR} = 0$ gives

$$\mathrm{C} = \log(1 + \mathrm{SNR}) = \mathrm{SNR} - \tfrac{1}{2}\,\mathrm{SNR}^2 + \mathrm{O}(\mathrm{SNR}^3). \tag{4.7}$$

The inverse function, which gives SNR as a function of C, is

$$\mathrm{SNR} = e^{\mathrm{C}} - 1 = \mathrm{C} + \tfrac{1}{2}\mathrm{C}^2 + \mathrm{O}(\mathrm{C}^3). \tag{4.8}$$

In terms of the bit-energy-to-noise ratio BNR, we have

$$\mathrm{BNR} = \log 2 + \tfrac{1}{2}\mathrm{C}\log 2 + \mathrm{O}(\mathrm{C}^2), \tag{4.9}$$

or, keeping only the linear term in BNR,

$$\mathrm{C} \simeq \big(\mathrm{BNR} - \log 2\big)\frac{2}{\log 2}. \tag{4.10}$$

The capacity is positive if the bit-energy-to-noise ratio exceeds $\log 2$, or about -1.59 dB, in line with the analysis of the capacity per unit energy. Additionally, the capacity can be approximated by an affine function when BNR is small. Recently, Verdú suggested approximating the constrained capacity $\mathrm{C}_{\mathcal{X}}$ by an affine function of BNR (in decibels, $\mathrm{BNR}^{\mathrm{dB}} = 10\log_{10}\mathrm{BNR}$), that is,

$$\mathrm{C}_{\mathcal{X}} \simeq \alpha_0\big(\mathrm{BNR}^{\mathrm{dB}} - \mathrm{BNR}_0^{\mathrm{dB}}\big), \tag{4.11}$$

where α_0 and $\mathrm{BNR}_0^{\mathrm{dB}}$ depend on the constellation set through the first terms of the Taylor series of $\mathrm{C}_{\mathcal{X}}(\mathrm{SNR})$ at $\mathrm{SNR} = 0$. In general, we have

Definition 4.2 (Taylor coefficients)**.** We denote by c_1 and c_2 the first two coefficients of a Taylor expansion of the capacity around SNR $= 0$, i. e. $\mathrm{C}(\mathrm{SNR}) = c_1\,\mathrm{SNR} + c_2\,\mathrm{SNR}^2 + \mathrm{o}(\mathrm{SNR}^2)$.

We will use the same notation for constrained capacity and for the additive energy channels, where the expansion is around zero energy.

Then we have (Theorem 9 of [25]),

Theorem 4.3. *As* SNR $\to 0$*, the channel capacity admits an expression in terms of BNR of the form*

$$C_{\mathcal{X}} \simeq \gamma_0\big(BNR^{dB} - BNR_0^{dB}\big) + \mathrm{o}\big((\Delta BNR^{dB})\big), \tag{4.12}$$

where $\Delta BNR^{dB} = BNR^{dB} - BNR_0^{dB}$*,* γ_0 *is a wideband slope, and* BNR_0 *the BNR at* SNR $\to 0$*.* γ_0 *and* BNR_0 *are respectively given by*

$$\gamma_0 = -\frac{c_1^2}{c_2 10\log_{10} 2}, \quad BNR_0 = \frac{\log 2}{c_1}. \tag{4.13}$$

For the unconstrained case $c_1 = 1$ and $c_2 = -\frac{1}{2}$. In [52], Prelov and Verdú determined c_1 and c_2 for proper-complex constellations, introduced by Neeser and Massey [55]. These constellations satisfy $\mu_{2'}(\mathcal{X}) = 0$, where

$$\mu_{2'}(\mathcal{X}) = \sum_{x\in\mathcal{X}} x^2 P_X(x), \tag{4.14}$$

is a second-order pseudo-moment, borrowing notation from the paper [55]. We define a pseudo-variance $\hat{\sigma}^2(\mathcal{X})$ as $\hat{\sigma}^2(\mathcal{X}) = \mu_{2'}(\mathcal{X}) - \mu_1^2(\mathcal{X})$. We compute the coefficients for slightly more general modulation formats in

Theorem 4.4. *The first two Taylor-series coefficients of the coded modulation capacity over a signal set* $\mathcal{X}$ *used with probabilities* $P_X(x)$ *are given by*

$$c_1 = \sigma^2(\mathcal{X}) = \mu_2(\mathcal{X}) - \big|\mu_1(\mathcal{X})\big|^2 \tag{4.15}$$

$$c_2 = -\frac{1}{2}\Big(\sigma^4(\mathcal{X}) + \big|\hat{\sigma}^2(\mathcal{X})\big|^2\Big). \tag{4.16}$$

In particular, when the average symbol is zero, and the mean energy is one,

$$c_1 = 1, \quad c_2 = -\frac{1}{2}\Big(1 + \big|\mu_{2'}(\mathcal{X})\big|^2\Big), \tag{4.17}$$

and the bit-energy-to-noise ratio at zero SNR *is* $BNR_0 = \log 2$.

The second-order coefficient is bounded by $-1 \leq c_2 \leq -\frac{1}{2}$*, the maximum* ($c_2 = -1/2$) *being attained when the constellation has uncorrelated real and imaginary parts and the energy is equally distributed among the real and imaginary parts.*

Note that constellation sets with unit energy and zero mean achieve the minimum energy per bit at low SNR.

Proof. The formulas for c_1 and c_2 can be derived from Theorem 5 of [52]. The details are given in Appendix 4.A. An alternative direct proof, which does not require knowledge of Theorem 5 in [52], may be found in Appendix 4.B. The minimum BNR is directly given by Theorem 4.3.

It is obvious that $c_2 \leq -\frac{1}{2}$. From the Cauchy-Schwartz inequality we deduce that $|\mu_{2'}(\mathcal{X})| = \left|\mathrm{E}[X^2]\right| \leq \mathrm{E}\left[|X|^2\right] = \mu_2(\mathcal{X})$, and therefore $c_2 \geq -1$. The maximum $c_2 = -\frac{1}{2}$ is attained when $\mu_{2'}(\mathcal{X}) = 0$, that is when $\mathrm{E}[X^2] = 0$. This condition is equivalent to

$$\mathrm{E}[X_r^2] - \mathrm{E}[X_i^2] = 0, \qquad \text{and } \mathrm{E}[X_r X_i] = 0. \tag{4.18}$$

Therefore, the real and imaginary parts must be uncorrelated ($\mathrm{E}[X_r X_i] = 0 = \mathrm{E}[X_r]\,\mathrm{E}[X_i]$) and with the same energy in the real and imaginary parts; this is indeed satisfied by proper-complex constellations [55]. □

The first part, giving c_1, can be found as Theorem 4 of [25]. The formula for c_2 seems to be new; note however that it can be easily derived from Theorem 5 in [52]. Application of Theorem 4.4 to some signal constellations of practical interest (used with equal probabilities) yields the following corollary.

Corollary 4.5. *Pulse amplitude modulation (PAM) with* 2^m *levels has* $c_2 = -1$.

A mixture of n_r $|\mathcal{X}_n|$*−PSK constellations for* $n = 1, \ldots, n_r$ *has* $c_2 = -\frac{1}{2}$ *if* $|\mathcal{X}_n| > 2$ *for all rings/sub-constellations* $n = 1, \ldots, n_r$.

Proof. In PAM, all symbols lie on a line in the complex plane, up to an irrelevant constant phase factor, we have $|\mu_{2'}(\mathcal{X})| = \mu_2(\mathcal{X}) = 1$.

Assume a PSK modulation with $|\mathcal{X}| > 2$. Up to a constant, irrelevant phase, the sum over the constellation symbols in $\mu_{2'}(\mathcal{X})$ gives

$$\mu_{2'}(\mathcal{X}) = \frac{1}{|\mathcal{X}|} \sum_{i=0}^{|\mathcal{X}|-1} e^{j2\frac{2\pi i}{|\mathcal{X}|}} = \frac{1}{|\mathcal{X}|} \frac{1 - e^{j2\frac{2\pi|\mathcal{X}|}{|\mathcal{X}|}}}{1 - e^{j2\frac{2\pi}{|\mathcal{X}|}}} = 0, \tag{4.19}$$

as $e^{j4\pi} = 1$. This is true only if the denominator is nonzero, which is satisfied for $|\mathcal{X}| > 2$. When $|\mathcal{X}| = 2$, we have $\left|\mu_{2'}(\mathcal{X})\right|^2 = 1$.

Let ϵ_n^2 be the squared modulus of the symbols at ring n. For simplicity, assume all rings have the same phase origin. We can split the sum over $\mathcal{X}$ as a sum over rings indexed by n,

$$\mu_{2'}(\mathcal{X}) = \frac{1}{|\mathcal{X}|} \sum_{n=1}^{n_r} \epsilon_n^2 \frac{1 - e^{j2\frac{2\pi|\mathcal{X}_n|}{|\mathcal{X}_n|}}}{1 - e^{j2\frac{2\pi}{|\mathcal{X}_n|}}}. \tag{4.20}$$

This number is zero if $|\mathcal{X}_n| > 2$ for all n. □

This result is a generalization of Theorems 11.1 and 11.2 of [25], where the results held for QPSK or mixtures of QPSK constellations. This result covers QAM and APSK modulations, for instance.

The expansion of the mutual information can be transformed into an expansion of the output channel entropy, by using the decomposition $I(X;Y) = H(Y) - H(Y|X)$, and $H(Y|X) = H(Z) = \log \pi e$,

Corollary 4.6. *The entropy of the AWGN channel output $H(Y)$ has the expansion at* $\mathrm{SNR} = 0$

$$H(Y) = \log \pi e + c_1 \,\mathrm{SNR} + c_2 \,\mathrm{SNR}^2 + O(\mathrm{SNR}^{5/2}). \tag{4.21}$$

In this section we have given closed-form approximations to the channel capacity as a function of the signal-to-noise ratio for low values of the signal-to-noise ratio. An interesting byline, recently analyzed by Verdú [25], links the coefficients c_1 and c_2 with two physical parameters in a radio communication channel, power and bandwidth. We next discuss this relationship.

4.2.4 Power and Bandwidth Trade-Off

In previous pages we have computed the first terms of the Taylor expansion of the constrained coded modulation capacity around $\mathrm{SNR} = 0$. Next to the intrinsic theoretical value of the results, the analysis also possesses some practical value, since many communication systems operate in Gaussian noise at low signal-to-noise ratio SNR. In this section we explore this link.

For a communication systems engineer interested in maximizing the data rate over a Gaussian channel, two physical variables are critical, namely the power P, or energy per unit time, and the bandwidth W, or number of channel

uses per unit time. Assuming additive Gaussian noise with spectral density N_0, the signal-to-noise ratio SNR is then given by $\mathrm{SNR} = P/(N_0 W)$.

The condition of low SNR applies to a wide variety of practical communication systems, whenever the condition that the available energy at each channel use is very low, that is, $\mathrm{SNR} \ll 1$. For fixed P and N_0, this is attained by letting the bandwidth increase unbounded, $W \to \infty$.

The capacity measured in bits per unit time is a natural figure of merit for the communications system. This capacity is given by $\mathrm{C}W = W \log(1 + \mathrm{SNR})$, where C the capacity considered in previous sections or in Chapter 3, has units of bits per unit time per Hertz. When W is very large, we have that

$$\mathrm{C}W = W \log\left(1 + \frac{P}{N_0 W}\right) = \frac{P}{N_0} - \frac{P^2}{2N_0^2 W} + \mathrm{O}\left(\frac{P^3}{N_0^3 W^2}\right). \tag{4.22}$$

For coded modulation systems whose Taylor expansion around $\mathrm{SNR} = 0$ has coefficients c_1 and c_2,

$$\mathrm{C}_{\mathcal{X}} W = c_1 \frac{P}{N_0} + c_2 \frac{P^2}{N_0^2 W} + \mathrm{O}\left(\frac{P^{5/2}}{N_0^{5/2} W^{3/2}}\right). \tag{4.23}$$

To first-order, the capacity increases linearly with the ratio P/N_0, that is linearly with the power P for fixed N_0. Following Verdú, we consider the following general situation. Two alternative transmission systems, respectively represented by an index $i = 1, 2$, with given power P_i and bandwidth W_i, achieve respective capacities per channel use C_i with Taylor coefficients c_{1i} and c_{2i}. In general, the capacities per unit time $\mathrm{C}_i W_i$ will differ.

A natural comparison between the two methods is to hold the power fixed, that is $P_1 = P_2$, and derive, as a function of W_1, the bandwidth W_2 required to have the same capacity per unit time, $\mathrm{C}_1 W_1 = \mathrm{C}_2 W_2$. Verdú carried out this analysis [25] and found that the bandwidth expansion is given by the ratio of the wideband slopes of alternatives 1 and 2, where the wideband slope is given by Theorem 4.3, and

$$\frac{W_2}{W_1} = \frac{c_{21}}{c_{22}}. \tag{4.24}$$

For instance, BPSK ($c_{22} = -1$), needs a bandwidth two times as big as QPSK ($c_{21} = -1/2$) to transmit the same rate. However, the form of Eq. (4.23) shows that the dependence of the capacity per unit time $\mathrm{C}W$ on the bandwidth W is rather weak. Since the capacity varies linearly with the available power P, a better way of achieving the same capacity is by varying P.

A natural trade-off between the two alternatives starts by fixing the power P_1 and bandwidth W_1 of system 1, and therefore signal-to-noise ratio SNR_1 and capacity per unit time $\mathrm{C}_1 W_1$. Then, one can fix the power P_2 (or the bandwidth W_2), and then determine how much bandwidth W_2 (resp. power P_2) is required to achieve the same capacity, that is $\mathrm{C}_1 W_1 = \mathrm{C}_2 W_2$. The following result determines the form of the trade-off between the power and the bandwidth for two alternative systems transmitting at the same rate,

Theorem 4.7. *Let two modulations attain respective capacities per channel use C_i, $i = 1, 2$, with respective Taylor coefficients c_{1i} and c_{2i} for* $\mathrm{SNR} \to 0$. *Consider system 1 as baseline, with power P_1, bandwidth W_1, signal-to-noise ratio* SNR_1 *and capacity $C_1 W_1$. Define $\Delta P = P_2/P_1$ and $\Delta W = W_2/W_1$ as the power and bandwidth expansion ratios between alternatives 1 and 2.*

In a neighbourhood of $\mathrm{SNR}_1 = 0$ *the capacities in bits per second, $C_1 W_1$ and $C_2 W_2$ are equal when the expansion factors ΔP and ΔW are related as*

$$\Delta W = \frac{\big(c_{22}\,\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1)\big)(\Delta P)^2}{c_{11} + c_{21}\,\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1) - c_{12}\Delta P}, \tag{4.25}$$

for ΔW as a function of ΔP and, if $c_{12} \neq 0$,

$$\Delta P = \frac{c_{11}}{c_{12}} + \left(\frac{c_{21}}{c_{12}} - \frac{c_{22}c_{11}^2}{c_{12}^3 \Delta W}\right)\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1), \tag{4.26}$$

for ΔP as a function of ΔW.

Proof. The proof can be found in Appendix 4.C. □

Remark that we use the approximation $\mathrm{SNR}_1 \to 0$. As a consequence, replacing the value of ΔP from Eq. (4.26) into Eq. (4.25) gives

$$\Delta W = \frac{1 + \mathrm{O}(\mathrm{SNR}_1)}{\frac{1}{\Delta W} + \mathrm{o}(1)}, \tag{4.27}$$

which is not exact, but consistent with the approximation that SNR_1 is small.

The previous theorem leads to the following derived results. For simplicity, we drop the terms $\mathrm{o}(\mathrm{SNR}_1)$ and replace the equality signs by approximate equalities.

Corollary 4.8. *To a first approximation, $c_{11}P_1 \simeq c_{12}P_2$, and a difference in coefficient c_1 immediately translates into a power loss.*

Corollary 4.9. *For $\Delta P = 1$, we obtain*

$$\Delta W \simeq \frac{c_{22}\,\mathrm{SNR}_1}{c_{11} + c_{21}\,\mathrm{SNR}_1 - c_{12}}, \tag{4.28}$$

and for the specific case $c_{11} = c_{12}$, $\Delta W \simeq c_{22}/c_{21}$.

The latter formula was previously obtained by Verdú [25] as a ratio of wideband slopes. As noticed in [25], and previously mentioned in Eq. (4.24), the loss in bandwidth may be significant when $\Delta P = 1$. But this point is just one of a curve relating ΔP and ΔW. For instance, with no bandwidth expansion we have

Corollary 4.10. *For $c_{11} = c_{12} = 1$, and choosing $\Delta W = 1$, $\Delta P \simeq 1 + \big(c_{21} - c_{22}\big)\,\mathrm{SNR}_1$.*

In decibels, and since SNR_1 is small, we may use a Taylor expansion of the logarithm function, and have $\Delta P \simeq 4.34\big(c_{21} - c_{22}\big)\,\mathrm{SNR}_1$ dB, where SNR_1 is in linear scale. In fact, there is no strong reason to limit ourselves to $\Delta W = 1$. For instance, for $\Delta W = \frac{1}{10}$, or a bandwidth compression of a factor 10, and for the QPSK/BPSK comparison, one gets $\Delta P \simeq 1 + 9.5\,\mathrm{SNR}_1 \simeq 41.26\,\mathrm{SNR}_1$ dB, which is indeed negligible for vanishing SNR.

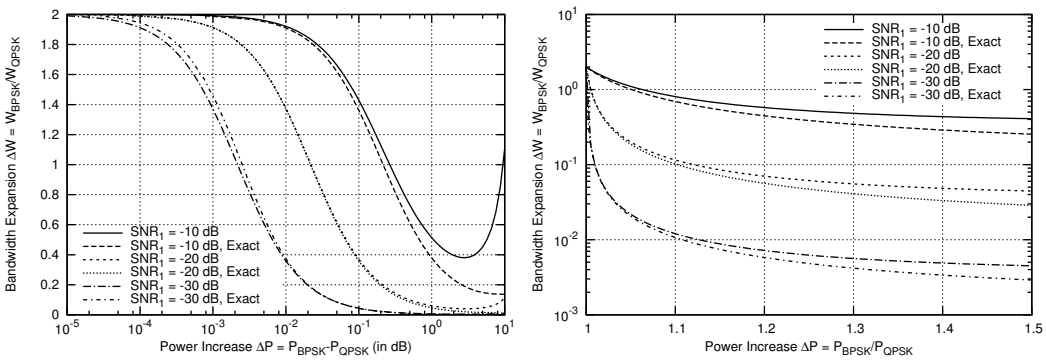
Figure 4.2 shows the trade-off curves between QPSK (baseline) and BPSK. For $\Delta P = 1$, the bandwidth expansion is indeed 2 times. However, for $\mathrm{SNR}_1 = -20\,\mathrm{dB}$, if a power loss of $1 + \frac{1}{2}\,\mathrm{SNR}_1 \simeq 0.02\,\mathrm{dB}$ is accepted, there is no bandwidth increase. Increasing power further, bandwidth may even be reduced. In either case, as $\mathrm{SNR}_1 \to 0$, the additional power loss turns negligible.

For signal-to-noise ratios below -10 dB, the approximation in Theorem 4.7 seems to be very accurate for "reasonable" power or bandwidth expansion ratios. A quantitative definition would lead to the problem of the extent to which the second order approximation to the capacity is correct, a question on which we do not dwell further.

Note that the curve derived from Theorem 4.7 diverges to $\Delta W \to \infty$ if

$$\Delta P = \frac{c_{11} + c_{21}\,\mathrm{SNR}_1}{c_{12}}, \tag{4.29}$$

which is typically close to $\Delta P = 1$ for low SNR. As we get away from $\Delta P = 1$, the bandwidth expansion quickly becomes smaller. The anomalous behaviour at large ΔP, namely the non-monotonicity of the curve, is due to the expansion at low SNR breaking down, more concretely from the minus sign of square root



(a) Power Expansion in logarithmic scale (in dB).

(b) Power Expansion in linear scale.

Figure 4.2: Trade-off ΔP vs. ΔW between QPSK (baseline) and BPSK.

in Eq. (4.142) in the proof of the theorem. In this region, the assumption of low SNR ceases being valid and the results become meaningless.

This analysis complements Verdú's approach to determine the form of trade-off between these two system variables, power and bandwidth. Next, we briefly consider the effect of fully-interleaved fading.

4.2.5 Effect of Fading on the Low SNR Regime

We now briefly study the effect of fading on the expansion at the low power regime. The fading model we consider is a fully-interleaved model, where the output y_k at time k has a form similar to Eq. (4.1),

$$y_k = \sqrt{\mathrm{SNR}}\, h_k\, x_k + z_k, \tag{4.30}$$

where z_k is a sample of Gaussian noise, h_k is a complex-valued fading coefficient, and x_k is the input. At each time instant, a new value h_k is used; its phase assumed known at the receiver. For general fading distributions, we have the following (see Theorem 12 of [25]),

Theorem 4.11. *For a fading distribution with finite moments satisfying* $\mathrm{E}[\chi] = 1$*, the Taylor expansion of the constrained capacity at low* SNR *is*

$$C_{\mathcal{X}}(\mathrm{SNR}) = c_1\,\mathrm{SNR} + c_2\,\mathrm{E}[\chi^2]\,\mathrm{SNR}^2 + \mathrm{o}(\mathrm{SNR}^2), \tag{4.31}$$

where c_1 and c_2 denote the first two terms in the expansion of the capacity in the absence of fading, Eqs. (4.15) *and* (4.16).

In Chapter 6 we shall use a Nakagami fading model to study the error performance of channel codes over Gaussian channels. In this case the squared fading coefficient $\chi_k = |h_k|^2$ follows a gamma distribution

$$p_\chi(\chi_k) = \frac{m_f^{m_f} \chi_k^{m_f - 1}}{\Gamma(m_f)} e^{-m_f \chi_k}, \tag{4.32}$$

where the parameter m_f is a real positive number, $0 < m_f < \infty$. Since $\mathrm{E}[\chi_k] = 1$, and signal and noise have average unit energy, SNR represents the average SNR at the receiver. This fading model is slightly more general than the standard Nakagami-m_f fading, since we lift the usual restriction to $m_f \geq 0.5$ [32, 51]. As the distribution is well-defined for $0 < m_f < 0.5$, reliable transmission is possible for $m_f > 0$. We recover the unfaded AWGN channel by letting $m_f \to +\infty$, a Rayleigh fading model by setting $m_f = 1$ and (an approximation to) Rician fading with parameter K_{Ric} by fixing $m_f = (K_{\mathrm{Ric}} + 1)^2/(2K_{\mathrm{Ric}} + 1)$.

Using the values of the moments of Nakagami-m_f fading, $\mathrm{E}[\chi] = 1$, and $\mathrm{E}[\chi^2] = 1 + 1/m_f$ (see Eq. (3.34)), we characterize the effect of Nakagami fading in the power and bandwidth requirements for a given modulation in the Gaussian channel,

Theorem 4.12. *Consider a modulation set $\mathcal{X}$ with average unit energy and used with power P, bandwidth W, and signal-to-noise ratio* SNR*. Let the capacity in absence of fading be characterized at low* SNR *by the coefficients $c_1 = 1$ and c_2. When used over a Nakagami-m_f channel, then $c_2(m_f) = \big(1 + \frac{1}{m_f}\big)c_2$.*

When used in the Nakagami-m_f channel with power $P(m_f)$ and bandwidth $W(m_f)$, if $P(m_f) = P$, $W(m_f) = W\big(1 + \frac{1}{m_f}\big)$, and if $W(m_f) = W$, $P(m_f) = P\big(1 - \frac{c_2}{m_f}\,\mathrm{SNR}\big)$.

As expected, for unfaded AWGN, when $m_f \to \infty$, we have $\mathrm{E}[\chi^2] = 1$. Rayleigh fading ($m_f = 1$) incurs in a bandwidth expansion of a factor 2 if the power remains fixed. On the other hand, if bandwidth is kept unchanged, there is a power penalty in dB of about $10 \log_{10}(1 - c_2\,\mathrm{SNR}) \simeq -10 c_2\,\mathrm{SNR} / \log 10 \simeq -4.343 c_2\,\mathrm{SNR}$ dB, a negligible amount for all practical effects since $\mathrm{SNR} \to 0$. The worst possible fading is $m_f \to 0$, which requires an unbounded bandwidth expansion or incurs an unlimited power penalty.

4.2.6 Asymptotic Behaviour at High SNR

In Figs. 4.1a and 4.1b, on page 75, we saw that there seemed to be only small differences among the various modulations in the low and high BNR extremes. In the last sections, we characterized analytically the behaviour at low BNR and showed that there are indeed no large differences in the rates achievable for the different modulations.

For fixed SNR and SNR $\to \infty$, the family of squared QAM constellations approaches an asymptotic mutual information, which differs from the mutual information by a constant. In this section we relate this asymptotic behaviour with the so-called shaping gain. We define the shaping gain of a constellation $\mathcal{X}$ as the ratio between the energy required to achieve a given input entropy $H(X_{\mathcal{X}})$ and the energy required by a Gaussian input to attain the same entropy.

For a fixed energy E_s, the largest entropy is achieved by a complex Gaussian distribution of variance E_s, $H(X) = \log(\pi e E_s)$, as we saw in Chapter 3, on page 31. For a fixed entropy an arbitrary constellation requires a larger energy than a Gaussian input. The differential entropies of two distributions, respectively uniformly distributed in a square and a circle, are given by [50, 56]

Proposition 4.13. *A constellation with average energy E_s uniformly distributed in a square $(-\Delta, \Delta) \times (-\Delta, \Delta)$, $\Delta = \sqrt{3E_s/2}$, has differential entropy $H(X) = \log(6E_s)$.*

A constellation with average energy E_s uniformly distributed in a circle has radius $\rho = \sqrt{2E_s}$ and differential entropy $H(X) = \log(2\pi E_s)$.

Proof. Within a square of size 2Δ, the uniform input density $p(s_r, s_i)$ satisfies $(2\Delta)^2 p(s_r, s_i) = 1$. The value of Δ is adjusted to the right average energy,

$$E_s = \int_{-\Delta}^{\Delta} \int_{-\Delta}^{\Delta} \frac{1}{4\Delta^2} (s_r^2 + s_i^2)\, ds_r\, ds_i = \frac{2\Delta^2}{3}. \tag{4.33}$$

This relates Δ and the average energy, $\Delta^2 = 3E_s/2$. The differential entropy of the uniform distribution is readily computed to be $\log(4\Delta^2) = \log(6E_s)$.

In a circle of radius ρ, the uniform input density $p(r) = (\pi\rho^2)^{-1}$ satisfies $2\pi \int_0^\rho r^3 p(r)\, dr = E_s$, and therefore $\rho^2 = 2E_s$. The differential entropy is then

$$-2\pi \int_0^\rho \frac{1}{\pi\rho^2} r \log \frac{1}{\pi\rho^2}\, dr = \log(\pi\rho^2) = \log(2\pi E_s). \tag{4.34}$$

□

Then, for a square constellation the shaping gain is $\frac{\pi e}{6}$, or 1.53 dB, whereas a circular constellation requires the marginally smaller shaping gain $\frac{e}{2}$, or 1.33 dB.

In Chapter 5, we shall carry out a similar analysis for the additive energy channels and determine their shaping gains.

This section concludes the analysis of coded modulation in the Gaussian channel. We next move to describe and analyze the technique of bit-interleaved coded modulation.

4.3 Bit-Interleaved Coded Modulation

4.3.1 On Bit-Interleaved Coded Modulation

As we saw in Section 4.2.3, for low signal-to-noise ratio the capacity of the Gaussian channel can be closely approached by using binary modulation, such as BPSK. For a fixed bit-energy-to-noise ratio BNR more efficient use of the channel is obtained by using QPSK modulation. Since the real and imaginary parts of the channel output y_k are separable into independent components for either BPSK or QPSK, the channel is as a binary-input, real-valued output channel with additive Gaussian noise. For this channel there exist good binary linear codes, viz. turbo-like codes [50], which get close to the capacity C.

For larger signal-to-noise ratios, constellations with more than two symbols are required. Since the shaping gain is relatively small, as we saw in Section 4.2.6, 2^m-ary modulations with equiprobable symbols are frequently used. When m is an integer, and for 2^m-QAM modulation, or 2^m-PSK modulation modulations, Ungerboeck proposed [57] the use of convolutional codes, combined with set-partitioning and trellis coding. These codes are efficiently decoded with the Viterbi algorithm.

A natural question is how well binary linear codes mapped with a binary mapping rule onto 2^m-ary modulations do perform, without set-partitioning and the trellis coding construction. A practical answer was given by pragmatic trellis-coded modulation [50], which replaced Ungerboeck's channel codes by a standard 64-state convolutional code. From the theoretical point of view, a more radical break with Ungerboeck's paradigm was the work by Caire *et al.* [26] on bit-interleaved coded modulation (BICM).

The operation of a bit-interleaved coded modulation scheme is depicted in Fig. 4.3. The source generates a message, which the encoder maps onto a binary codeword $\mathbf{b} = (b_{11}, \cdots, b_{m1}, \ldots, b_{1n}, \cdots, b_{mn})$, of length mn, where $|\mathcal{X}| = 2^m$, and n is the number of channel uses. Code selection is independent of the constellation set $\mathcal{X}$; more precisely, binary inputs and binary codes are used. Next, a mapper μ uses m consecutive binary digits to compute a (complex-valued) channel-input symbol $x_k = \mu(b_{1k}, \ldots, b_{mk})$, for $k = 1, \ldots, n$. A typical

choice for μ is natural reflected Gray mapping, or Gray for short. The Gray mapping for m bits may be generated recursively from the mapping for $m-1$ bits by prefixing a binary 0 to the mapping for $m-1$ bits, then prefixing a binary 1 to the reflected (i. e. listed in reverse order) mapping for $m-1$ bits.

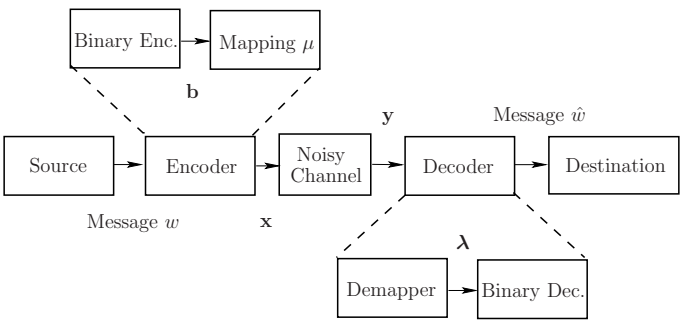


Figure 4.3: Operation of bit-interleaved coded modulation.

As in our analysis of coded modulation over the AWGN channel, the channel output is still given by Eqs. (4.1), namely $y_k = \sqrt{\mathrm{SNR}}x_k + z_k$, where z_k is a complex-valued Gaussian noise, $z_k \sim \mathcal{N}_{\mathbf{C}}(0,1)$, and SNR is the average signal-to-noise ratio. The conditional output density $Q(y|x)$ is given by Eq. (4.4).

At the receiver side, the first step is demapping, i. e. the computation from the channel output y of the metrics used by the decoder. In coded modulation, the demapper determines the symbol a posteriori probabilities $P_{X|Y}(x|y)$, which are proportional to the channel transition matrix $Q(y|x)$; the factor of proportionality is irrelevant for the decoding. With bit-interleaved coded modulation, the metrics $q(x,y)$ are given by

$$q(x,y) = \prod_{i=1}^{m} q_i(b,y), \quad b \in \{0,1\}, \tag{4.35}$$

namely the product of per-bit metrics $q_i(b,y)$. These metrics are in turn proportional to the marginal bit a posteriori probabilities, denoted by $P_{B_i|Y}(b|y)$, and given by

$$q_i(b,y) \propto P_{B_i|Y}(b|y) \propto p_{Y|B}(y|b) = \sum_{x \in \mathcal{X}_i^b} \frac{1}{|\mathcal{X}_i^b|} Q(y|x), \tag{4.36}$$

where $\mathcal{X}_i^b$ is the set of constellation symbols with bit b in the i-th position of the binary label. An alternative form is the log-likelihood ratio form,

$$\lambda_i = \log \frac{P_{B_i|Y}(1|y)}{P_{B_i|Y}(0|y)} = \log \frac{\sum_{x \in \mathcal{X}_i^1} Q(y|x)}{\sum_{x \in \mathcal{X}_i^0} Q(y|x)}. \tag{4.37}$$

It is clear that the marginal a posteriori probabilities for the bit B_i may be recovered from λ_i. The log-likelihood ratios constitute a set of sufficient statistics for the decoding.

Finally, an estimate of the source message is generated. In practice, if turbo-like codes are used, a variation of iterative decoding is used. We finally note that very good error performance is possible with simple linear codes, such as convolutional codes or turbo-like codes [26].

4.3.2 Bit-Interleaved Coded Modulation Capacity

The presentation in the previous section naturally leads to a BICM channel capacity, which in general is smaller than the coded modulation capacity we considered in previous sections. For a given constellation set $\mathcal{X}$, we denoted the constrained capacity by $\mathrm{C}_{\mathcal{X}}$. Analogously, we denote the BICM capacity for a given constellation set $\mathcal{X}$ and mapping rule μ by $\mathrm{C}_{\mathcal{X},\mu}$.

From the decoder's point of view, and following Caire's analysis [26], the channel is separated into a set of parallel subchannels, such that the input to subchannel i is bit b_i; the channel output has the form of a log-likelihood ratio λ_i. As the log-likelihood ratios are a sufficient statistics for the decoding,

$$I(B_i; \Lambda_i) = I(B_i; Y). \tag{4.38}$$

For infinite interleaving, and assuming the subchannels are independent, this would give a total rate $\mathrm{C}_{\mathcal{X},\mu}$ equal to the mutual informations in each subchannel, $I(B_i; Y)$, that is $\mathrm{C}_{\mathcal{X},\mu} = \sum_{i=1}^m I(B_i; Y)$. This total rate is the so-called BICM capacity. The ratios λ_i are generated simultaneously, by assuming a uniform prior over the remaining bits in the symbol. The correlation among the m ratios, for $i = 1, \ldots, n$, in the same symbol is ignored. We shall later see that this loss is small.

We next define the BICM capacity for a generic channel, not necessarily Gaussian. One such example of interest is provided by the additive energy channels, which we will study in Chapter 5.

Definition 4.14. For a channel with transition matrix $Q(y|x)$ the capacity BICM $\mathrm{C}_{\mathcal{X},\mu}$ with a fixed modulation set $\mathcal{X}$ and mapping rule μ is given by

$$\mathrm{C}_{\mathcal{X},\mu} = \sum_{i=1}^{m} I(B_i;Y), \tag{4.39}$$

where

$$I(B_i;Y) = \mathrm{E}\left[\log \frac{\sum_{x'\in\mathcal{X}_i^B} Q(Y|x')}{\frac{1}{2}\sum_{x'\in\mathcal{X}} Q(Y|x')}\right]. \tag{4.40}$$

The expectation is performed over the bit values b, the input symbols x in $\mathcal{X}_i^b$, and the channel realizations Y.

Remark 4.15. In recent joint work with Albert Guillén i Fàbregas, we have analyzed the BICM decoder from the point of view of mismatched decoding [58]. For the decoding metric in Eq. (4.35), we have proved that the corresponding generalized mutual information [58] is indeed given by Caire's BICM capacity, even though the interleaver is finite. This proves the achievability of the BICM capacity. A similar result had been presented in [59].

This alternative definition proves useful for the analysis at low SNR,

Proposition 4.16. *For a channel with transition matrix $Q(y|x)$, modulation set $\mathcal{X}$ and mapping rule μ, the BICM capacity $C_{\mathcal{X},\mu}$ is*

$$C_{\mathcal{X},\mu} = \sum_{i=1}^{m} \frac{1}{2} \sum_{b=0,1} (C_{\mathcal{X}}^u - C_{\mathcal{X}_i^b}^u), \tag{4.41}$$

where $C_{\mathcal{X}}^u$ and $C_{\mathcal{X}_i^b}^u$ are, respectively, the constrained capacities for equiprobable signalling in $\mathcal{X}$ and $\mathcal{X}_i^b$.

Proof. By definition, the BICM capacity is the sum over $i = 1, \ldots, m$ of the mutual informations $I(B_i;Y)$. We rewrite this mutual information as

$$I(B_i;Y) = \frac{1}{2} \sum_{b\in\{0,1\}} \mathrm{E}\left[\log \frac{\sum_{x'\in\mathcal{X}_i^b} Q(Y|x')}{\frac{1}{2}\sum_{x'\in\mathcal{X}} Q(Y|x')}\right] \tag{4.42}$$

$$= \frac{1}{2} \sum_{b\in\{0,1\}} \mathrm{E}\left[\log\left(\frac{\sum_{x'\in\mathcal{X}_i^b} \frac{2}{|\mathcal{X}|} Q(Y|x')}{Q(Y|X)} \frac{Q(Y|X)}{\frac{1}{2}\sum_{x'\in\mathcal{X}} \frac{2}{|\mathcal{X}|} Q(Y|x')}\right)\right], \tag{4.43}$$

where we have modified the variable in the logarithm by including a factor $\frac{2}{|\mathcal{X}|}Q(Y|X)$ in both numerator and denominator. Splitting the logarithm,

$$I(B_i;Y) = \frac{1}{2}\sum_{b\in\{0,1\}} \mathrm{E}\left[\log \frac{\sum_{x'\in\mathcal{X}_i^b} \frac{2}{|\mathcal{X}|}Q(Y|x')}{Q(Y|X)}\right] + \frac{1}{2}\sum_{b\in\{0,1\}} \mathrm{E}\left[\log \frac{Q(Y|X)}{\frac{1}{|\mathcal{X}|}\sum_{x'\in\mathcal{X}} Q(Y|x')}\right]. \tag{4.44}$$

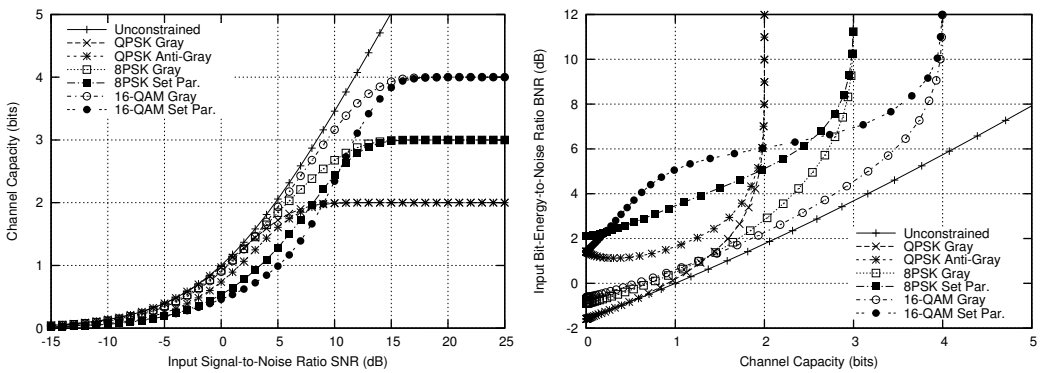
For fixed b, the expectations are respectively recognized as (minus) the mutual information achievable by using equiprobable signalling in $\mathcal{X}_i^b$, $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}_i^b}$, and the mutual information achieved by equiprobable signalling in $\mathcal{X}$, $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}}$. □

Figures 4.4a and 4.4b depict the BICM channel capacity (in bits per channel use) $\mathrm{C}_{\mathcal{X},\mu}$ for several modulations and mapping rules in the Gaussian channel. The cases depicted correspond to commonly used modulations and mapping rules: QPSK with Gray and anti-Gray mappings, 8PSK with Gray and set partitioning mappings, and 16-QAM with Gray and set-partitioning mappings (see for instance [26] for the exact mapping rules). As found in [26], the BICM capacity is close to that of coded modulation when Gray mapping is used. Use of set-partitioning leads to a significant loss in capacity.

The good behaviour of Gray mapping is somewhat limited at high SNR, as can be seen in Fig. 4.4b where BNR is plotted as a function of the capacity. In the following section, we consider the asymptotic behaviour at low SNR, and characterize it analytically. Among other things, we shall be able to explain behaviour such as that for QPSK and anti-Gray mapping, which has a negative slope at BNR_0. This plot shows why the notation BNR_0 is preferable to $\mathrm{BNR}_{\min}$, as the minimum BNR is not necessarily attained at zero capacity.

4.3.3 Asymptotic Behaviour at Low SNR

In the previous section, we derived the BICM capacity $\mathrm{C}_{\mathcal{X},\mu}$, and plotted it as a function of SNR in Fig. 4.4a. In the same plot, the CM channel capacity $\mathrm{C}_{\mathcal{X}}$, which assumes equiprobable signalling over the input constellation set, is also shown. As a function of the signal-to-noise ratio SNR, in Fig. 4.4, the BICM capacity is close to the CM value when Gray mapping is used. However, the plot as a function of the BNR (see Fig. 4.4b) reveals the suboptimality of BICM for low rates, or equivalently, in the power-limited regime.



(a) Capacity as a function of SNR. (b) Capacity as a function of BNR.

Figure 4.4: Channel capacity (in bits per channel use).

In this section, we apply the techniques developed for the analysis of coded modulation in the AWGN channel to the low-SNR regime for BICM. As we did in the coded modulation case, our goal is to compute expansions of $\mathrm{C}_{\mathcal{X},\mu}$ as a function of SNR and BNR, when $\mathrm{C}_{\mathcal{X},\mu}$ is small.

First, for fixed label index, i, and bit value b, let us define the quantities $\mu_1(\mathcal{X}_i^b)$, $\mu_2(\mathcal{X}_i^b)$, and $\mu_{2'}(\mathcal{X}_i^b)$, $\sigma^2(\mathcal{X}_i^b)$ and $\hat{\sigma}^2(\mathcal{X}_i^b)$ as the (pseudo-)moments and (pseudo-)variances of the set $\mathcal{X}_i^b$. Then, we have

Theorem 4.17. *In the AWGN channel with average signal-to-noise ratio* SNR*, the Taylor coefficients of the BICM capacity $C_{\mathcal{X},\mu}$ used over the set $\mathcal{X}$ (of zero mean and unit average energy) with mapping μ are given by*

$$c_1 = \sum_{i=1}^{m} \frac{1}{2} \sum_{b} |\mu_1(\mathcal{X}_i^b)|^2, \tag{4.45}$$

$$c_2 = -\frac{1}{2}\left\{\sum_{i=1}^{m}\frac{1}{2}\sum_{b=0,1}\left(\left(\sigma^4(\mathcal{X}) - \sigma^4(\mathcal{X}_i^b) + \left|\hat{\sigma}^2(\mathcal{X})\right|^2 - \left|\hat{\sigma}^2(\mathcal{X}_i^b)\right|^2\right)\right)\right\}. \tag{4.46}$$

Proof. From Proposition 4.16, the BICM capacity can be written as

$$\mathrm{C}_{\mathcal{X},\mu} = \sum_{i=1}^{m} \frac{1}{2} \sum_{b=0,1} (\mathrm{C}^{\mathrm{u}}_{\mathcal{X}} - \mathrm{C}^{\mathrm{u}}_{\mathcal{X}_i^b}). \tag{4.47}$$

Since the summands $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}}$ and $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}_i^b}$ admit a Taylor expansion given in Theorem 4.1,

$$c_1 = \sum_{i=1}^{m} \frac{1}{2} \sum_{b=0,1} \Big(1 - \big(\mu_2(\mathcal{X}_i^b) - |\mu_1(\mathcal{X}_i^b)|^2\big)\Big) \tag{4.48}$$

$$= \sum_{i=1}^{m} \left(\left(1 - \frac{1}{2} \sum_{b=0,1} \mu_2(\mathcal{X}_i^b)\right) + \frac{1}{2} \sum_{b=0,1} |\mu_1(\mathcal{X}_i^b)|^2 \right) \tag{4.49}$$

$$= \sum_{i=1}^{m} \frac{1}{2} \sum_{b=0,1} |\mu_1(\mathcal{X}_i^b)|^2, \tag{4.50}$$

since $\frac{1}{2}\sum_{b=0,1} \mu_2(\mathcal{X}_i^b) = \mu_2(\mathcal{X}) = 1$ by construction.

The quadratic coefficient c_2 follows from a similar application of Theorem 4.1. □

Table 4.1 shows the values of the coefficients c_1 and c_2, as well as the minimum bit signal-to-noise ratio BNR_0 for various cases, namely QPSK with Gray (Q–Gr) and anti-Gray mapping (Q–A-Gr), 8PSK and 16-QAM modulations and Gray and set partitioning mappings (respectively 8–Gr, 8-SP, 16-Gr, and 16–SP). Note that for QPSK with anti-Gray mapping the slope at BNR_0 is negative. From Theorem 4.3 this should correspond with a positive coefficient c_2, as it indeed does. A similar effect takes place for 8PSK and set partitioning, even though it is barely noticeable in the plot.

	Modulation and Mapping					
	Q–Gr	Q–A-Gr	8–Gr	8–SP	16–Gr	16–SP
c_1	1.0000	0.5000	0.8536	0.4268	0.8000	0.5000
BNR_0	0.6931	1.3863	0.8121	1.6241	0.8664	1.3863
BNR_0 (dB)	-1.5917	1.4186	-0.9041	2.1062	-0.6226	1.4186
c_2	-0.5000	0.2500	-0.2393	0.0054	-0.1600	-0.3100

Table 4.1: Bit signal-to-noise ratio and coefficients c_1, c_2 for BICM in AWGN.

In general, it seems to be difficult to draw general conclusions for arbitrary mappings from Theorem 4.17. A notable exception, however, is the analysis under natural reflected Gray mapping.

Theorem 4.18. *For 2^m-PAM and 2^{2m}-QAM (m a positive integer) and natural, binary-reflected Gray mapping, the coefficient c_1 in the Taylor expansion of the BICM capacity $C_{\mathcal{X},\mu}$ at low* SNR *is*

$$c_1 = \frac{3 \cdot 2^{2m}}{4(2^{2m}-1)}, \tag{4.51}$$

and the minimum BNR_0 is

$$BNR_0 = \frac{4(2^{2m}-1)}{3 \cdot 2^{2m}} \log 2. \tag{4.52}$$

Proof. For 2^m-PAM, the Gray mapping construction makes $\mu_1(\mathcal{X}_i^b) = 0$, for $b = 0, 1$ and all bit positions except one, which we take with no loss of generality to be $i = 1$. Therefore,

$$c_1 = \frac{1}{2}\left|\mu_1(\mathcal{X}_1^0)\right|^2 + \frac{1}{2}\left|\mu_1(\mathcal{X}_1^1)\right|^2 = \left|\mu_1(\mathcal{X}_1^0)\right|^2 = \left|\mu_1(\mathcal{X}_1^1)\right|^2. \tag{4.53}$$

The last equalities follow from the symmetry between 0 and 1.

Symbols lie on a line in the complex plane with values $\pm\beta\big(1, 3, 5, \ldots, 2^m-1\big)$, with $\beta^2 = 3/(2^{2m}-1)$; this follows from setting $2n = 2^m$ in the formula $\frac{1}{n}\sum_{i=1}^{n}(2i-1)^2 = \frac{1}{3}((2n)^2-1)$. The average symbol has modulus $|\mu_1(\mathcal{X}_1^0)| = \beta 2^{m-1}$, and therefore

$$c_1 = \left|\mu_1(\mathcal{X}_1^0)\right|^2 = \frac{3 \cdot 2^{2m}}{4(2^{2m}-1)}. \tag{4.54}$$

Extension to 2^{2m}-QAM is clear, by taking the Cartesian product along real and imaginary parts. Now, two indices i contribute, each with an identical form to that of PAM. As the energy along each axis of half that of PAM, the normalization factor β^2_{QAM} also halves and overall c_1 does not change. □

The results for BPSK, QPSK (2-PAM×2-PAM), and 16-QAM (4-PAM×4-PAM), as presented in Table 4.1, match with the Theorem, as they should.

It is somewhat surprising that the loss with respect to coded modulation at low SNR is bounded,

Corollary 4.19. *As $m \to \infty$, and under the conditions of Theorem 4.18, BNR_0 approaches $\frac{4}{3}\log 2 \simeq -0.3424$ dB from below.*

Using our analysis of the low SNR regime, Theorem 4.7 and its corollary 4.8, we deduce that the loss represents about 1.25 dB with respect to the classical CM limit, namely $\mathrm{BNR}_0 = -1.59\,\mathrm{dB}$. Using a fixed modulation for a large range of signal-to-noise ratio values, with adjustment of the transmission rate by changing the code rate, needs not result in a large loss with respect to more optimal transmission schemes, where both the rate and modulation change. Applying the trade-off between power and bandwidth of Theorem 4.7, bandwidth may be compressed at some cost in power. Figure 4.5 depicts the trade-off for between QPSK and 16-QAM (with Gray mapping) for two values of signal-to-noise ratio. The exact result, obtained by using the exact formulas for the mutual information is plotted along the result by using Theorem 4.7.

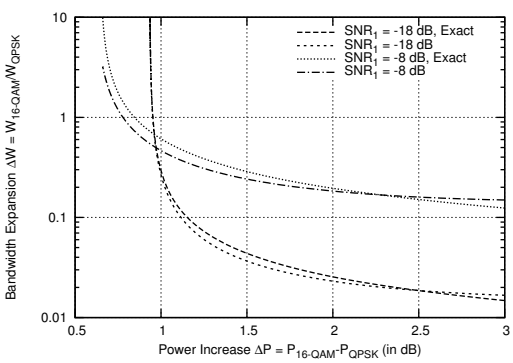


Figure 4.5: Trade-off ΔP vs. ΔW between 16-QAM (Gray map.) and QPSK.

As expected from the values of c_1 and c_2, use of 16-QAM incurs in a non-negligible power loss, given to a first approximation by Theorem 4.18. This loss may however be accompanied by a significant reduction in bandwidth, which might be of interest in some applications. For signal-to-noise ratios larger than those reported in the figure, the assumption of low SNR loses its validity and the results derived from the Taylor expansion are no longer accurate.

4.4 Conclusions

In this chapter, we have reviewed the capacity of digital modulation systems in Gaussian channels. Whenever the results presented are known, we have strived to present them in such a way that generalization to the additive energy

channel is straightforward. In addition, there are a few novel contributions in the analysis:

1. The first two derivatives of the constrained capacity at zero SNR have been computed for general modulation sets.

2. The trade-off between power penalty and bandwidth expansion between two alternative systems at low SNR has been determined. It generalizes Verdú's analysis of the wideband regime, which estimated only the bandwidth expansion. We have shown that no bandwidth expansion may often be achieved at a negligible (but non-zero) cost in power.

3. A similar trade-off between power penalty and bandwidth expansion for general Nakagami-m_f fading has been computed, with similar conclusions as in the point above: bandwidth expansion may be large at no power cost, but absent at a tiny power penalty.

4. The capacity at low BNR for BICM has been characterized by simple expressions. For binary reflected Gray mapping, the capacity loss at low SNR with respect to coded modulation is bounded by about 1.25 dB.

In the next chapter, we build on the analysis presented here and show that these results admit a natural extension to the additive energy channels.

4.A CM Capacity Expansion at Low SNR

Since the constellation moments are finite, we have that $\mathrm{E}\big[|X|^{2+\alpha}\big] < \infty$ for $\alpha > 0$. Therefore, as $\mathrm{SNR} \to 0$, for $\mu > 0$ the technical condition

$$\mathrm{SNR}^{2+\alpha}\,\mathrm{E}\big[|X|^{2+\alpha}\big] \leq (-\log\sqrt{\mathrm{SNR}})^{\mu}, \tag{4.55}$$

necessary to apply Theorem 5 of [52] holds.

Let us define a 2×1 vector $x^{(r)} = (x_r\ x_i)^T$, with components the real and imaginary parts of symbol s, respectively denoted by x_r and x_i. The covariance matrix of $x^{(r)}$, denoted by $\mathrm{cov}(X)$, is given by

$$\mathrm{cov}(X) = \begin{pmatrix} \mathrm{E}[(X_r - \hat{x}_r)^2] & \mathrm{E}\big[(X_r - \hat{x}_r)(X_i - \hat{x}_i)\big] \\ \mathrm{E}\big[(X_r - \hat{x}_r)(X_i - \hat{x}_i)\big] & \mathrm{E}[(X_i - \hat{x}_i)^2] \end{pmatrix}, \tag{4.56}$$

where $\hat{x}_r$ and $\hat{x}_i$ are the mean values of the real and imaginary parts of the constellation.

Theorem 5 of [52] gives $c_1 = \mathrm{Tr}(\mathrm{cov}(X))$ and $c_2 = -\mathrm{Tr}(\mathrm{cov}^2(X))$, or

$$c_1 = \mathrm{E}[(X_r - \hat{x}_r)^2] + \mathrm{E}[(X_i - \hat{x}_i)^2] \tag{4.57}$$

$$c_2 = -\Big(\mathrm{E}^2[(X_r - \hat{x}_r)^2] + \mathrm{E}^2[(X_i - \hat{x}_i)^2] + 2\,\mathrm{E}^2\big[(X_r - \hat{x}_r)(X_i - \hat{x}_i)\big]\Big). \tag{4.58}$$

The coefficient c_1 coincides with that in Eq. (4.15).

As for c_2, let us add a subtract a term $\mathrm{E}[(X_r - \hat{x}_r)^2]\,\mathrm{E}[(X_i - \hat{x}_i)^2]$ to Eq. (4.58). Then,

$$\begin{aligned} c_2 = -\Big(&\tfrac{1}{2}\,\mathrm{E}^2[(X_r - \hat{x}_r)^2] + \tfrac{1}{2}\,\mathrm{E}^2[(X_i - \hat{x}_i)^2] + \mathrm{E}[(X_r - \hat{x}_r)^2]\,\mathrm{E}[(X_i - \hat{x}_i)^2] \\ &+ \tfrac{1}{2}\,\mathrm{E}^2[(X_r - \hat{x}_r)^2] + \tfrac{1}{2}\,\mathrm{E}^2[(X_i - \hat{x}_i)^2] \\ &- \mathrm{E}[(X_r - \hat{x}_r)^2]\,\mathrm{E}[(X_i - \hat{x}_i)^2] + 2\,\mathrm{E}^2\big[(X_r - \hat{x}_r)(X_i - \hat{x}_i)\big]\Big), \end{aligned} \tag{4.59}$$

which in turn can be written as

$$c_2 = -\frac{1}{2}\Big(\mathrm{E}^2\big[|X - \hat{x}|^2\big] + \big|\mathrm{E}[(X - \hat{x})^2]\big|^2\Big), \tag{4.60}$$

a form which coincides with Eq. (4.16), by noting that

$$\mathrm{E}\big[|X - \hat{x}|^2\big] = \mathrm{E}\big[|X|^2\big] - |\hat{x}|^2 = \mu_2(\mathcal{X}) - \big|\mu_1(\mathcal{X})\big|^2 \tag{4.61}$$

$$\mathrm{E}[(X - \hat{x})^2] = \mathrm{E}[X^2] - \hat{x}^2 = \mu_{2'}(\mathcal{X}) - \mu_1^2(\mathcal{X}). \tag{4.62}$$

4.B CM Capacity Expansion at Low SNR – AWGN

The constrained capacity $\mathrm{C}_{\mathcal{X}}$ is given by Eq. (4.5),

$$\mathrm{C}_{\mathcal{X}} = -\sum_x P(x) \int_Y Q(y|x) \log\left(\sum_{x' \in \mathcal{X}} P(x') e^{-|\gamma(x-x')+z|^2+|z|^2}\right) dy, \tag{4.63}$$

where, for the sake of brevity, we have set $\gamma = \sqrt{\mathrm{SNR}}$.

We first rewrite each of the exponents in the $\log(\cdot)$ in Eq. (4.63),

$$\exp\big(-|\gamma(x - x') + z|^2 + |z|^2\big) = \exp\big(-\gamma^2|x - x'|^2 - 2\gamma r(x - x')\big), \tag{4.64}$$

where we have used the function $r(y)$, defined as

$$r(y) = \mathrm{Re}(yz^*). \tag{4.65}$$

In general, we will take $y = x - x'$.

Now we use a Taylor expansion of e^t around $t = 0$, $e^t = 1 + t + \frac{1}{2}t^2 + \frac{1}{6}t^3 + \frac{1}{24}t^4 + \mathrm{O}\big(t^5\big)$ to transform Eq. (4.64) into

$$\begin{aligned}
&1 - \gamma^2|x-x'|^2 - 2\gamma r(x-x') + \frac{1}{2}\Big(-\gamma^2|x-x'|^2 - 2\gamma r(x-x')\Big)^2 \\
&\qquad + \frac{1}{6}\Big(-\gamma^2|x-x'|^2 - 2\gamma r(x-x')\Big)^3 \\
&\qquad + \frac{1}{24}\Big(-\gamma^2|x-x'|^2 - 2\gamma r(x-x')\Big)^4 + \mathrm{O}(\gamma^5) \qquad (4.66)\\
&= 1 - \gamma^2|x-x'|^2 - 2\gamma r(x-x') + \frac{1}{2}\gamma^4|x-x'|^4 + \gamma^2 2r^2(x-x') \\
&\qquad + \gamma^3 2|x-x'|^2 r(x-x') - \gamma^4 2|x-x'|^2 r^2(x-x') \\
&\qquad - \gamma^3\frac{4}{3}r^3(x-x') + \gamma^4\frac{2}{3}r^4(x-x') + \mathrm{O}(\gamma^5). \qquad (4.67)
\end{aligned}$$

With the substitutions,

$$a_1' = -2r(x-x') \qquad (4.68)$$

$$a_2' = -|x-x'|^2 + 2r^2(x-x') \qquad (4.69)$$

$$a_3' = 2|x-x'|^2 r(x-x') - \frac{4}{3}r^3(x-x') \qquad (4.70)$$

$$a_4' = \frac{1}{2}|x-x'|^4 - 2|x-x'|^2 r^2(x-x') + \frac{2}{3}r^4(x-x'). \qquad (4.71)$$

the sum over x' in Eq. (4.63) becomes

$$\sum_{x'\in\mathcal{X}} P(x')\big(1 + \gamma a_1' + \gamma^2 a_2' + \gamma^3 a_3' + \gamma^4 a_4' + \mathrm{O}(\gamma^5)\big) \qquad (4.72)$$

$$= \Big(1 + \gamma a_1 + \gamma^2 a_2 + \gamma^3 a_3 + \gamma^4 a_4 + \mathrm{O}(\gamma^5)\Big). \qquad (4.73)$$

where the coefficients a_l, with $l = 1, 2, 3, 4$, are given by $a_l = \sum_{x'\in\mathcal{X}} a_l' P(x')$. Of special interest is a_1,

$$a_1 = -2\sum_{x'\in\mathcal{X}} r(x-x')P(x') = -2\big(r(x-\hat{x})\big), \qquad (4.74)$$

since $r(y)$ is linear in its argument and $\hat{x}$ is the mean value of the constellation.

As next step we expand the logarithm in Eq. (4.73) using $\log(1+t) = t - \frac{1}{2}t^2 + \frac{1}{3}t^3 - \frac{1}{4}t^4 + \mathrm{O}\big(t^5\big)$, where y is small. The term in t^4 is needed to catch all factors in γ^4. Then

$$\log\Big(1+\gamma a_1 + \gamma^2 a_2 + \gamma^3 a_3 + \gamma^4 a_4 + \mathrm{O}(\gamma^5)\Big) \tag{4.75}$$

$$\begin{aligned}
&= \gamma a_1 + \gamma^2 a_2 + \gamma^3 a_3 + \gamma^4 a_4 - \frac{1}{2}\big(\gamma a_1 + \gamma^2 a_2 + \gamma^3 a_3 + \gamma^4 a_4\big)^2 \\
&\quad + \frac{1}{3}\big(\gamma a_1 + \gamma^2 a_2 + \gamma^3 a_3 + \gamma^4 a_4\big)^3 \\
&\quad - \frac{1}{4}\big(\gamma a_1 + \gamma^2 a_2 + \gamma^3 a_3 + \gamma^4 a_4\big)^4 + \mathrm{O}(\gamma^5) \qquad (4.76)\\
&= \gamma a_1 + \gamma^2 a_2 + \gamma^3 a_3 + \gamma^4 a_4 - \frac{1}{2}\big(\gamma^2 a_1^2 + 2\gamma^3 a_1 a_2 + \gamma^4 a_2^2 + 2\gamma^4 a_1 a_3\big) \\
&\quad + \frac{1}{3}\big(\gamma^3 a_1^3 + 3\gamma^4 a_1^2 a_2\big) - \gamma^4 \frac{1}{4} a_1^4 + \mathrm{O}(\gamma^5) \qquad (4.77)\\
&= \gamma a_1 + \gamma^2\left(a_2 - \frac{1}{2}a_1^2\right) + \gamma^3\left(a_3 - a_1 a_2 + \frac{1}{3}a_1^3\right) \\
&\quad + \gamma^4\left(a_4 - \frac{1}{2}a_2^2 - a_1 a_3 + a_1^2 a_2 - \frac{1}{4}a_1^4\right) + \mathrm{O}(\gamma^5). \qquad (4.78)
\end{aligned}$$

The remaining steps are the averaging over the input symbol x and the noise realization z. To save space, we sometimes use the symbol $\mathrm{E}_{X,Z}$ to make explicit the variable (X, Z) over which the averaging is performed.

Throughout, we have factors depending on

$$\mathrm{E}_Z\, r^{k_1}(\chi_1) r^{k_2}(\chi_2), \tag{4.79}$$

χ_1 and χ_2 complex numbers, and k_1, k_2 are 0, 1, 2, 3, or 4. We have then

Lemma 4.20. *Let the function $r(y)$ be $r(y) = \mathrm{Re}(yz^*)$, where z is a complex Gaussian random variable, $z \sim \mathcal{N}_{\mathbf{C}}(0,1)$, and y is a complex number. Then*

$$\mathrm{E}_Z\, r^{k_1}(\chi_1) r^{k_2}(\chi_2) = 0, \quad \textit{if } k_1 + k_2 \textit{ is odd.} \tag{4.80}$$

and

$$\mathrm{E}_Z\, r(\chi_1) r(\chi_2) = \frac{1}{2}|\chi_1||\chi_2|\cos(\varphi_1 - \varphi_2) \tag{4.81}$$

$$\mathrm{E}_Z\, r(\chi_1) r^3(\chi_2) = \frac{3}{4}|\chi_1||\chi_2|^3\cos(\varphi_1 - \varphi_2) \tag{4.82}$$

$$\mathrm{E}_Z\, r^2(\chi_1) r^2(\chi_2) = \frac{1}{4}|\chi_1|^2|\chi_2|^2\big(2 + \cos 2(\varphi_1 - \varphi_2)\big). \tag{4.83}$$

Here φ_1 and φ_2 are the phases of the complex numbers χ_1 and χ_2 respectively. In particular,

$$\mathrm{E}_Z\, r^2(\chi_1) = \frac{1}{2}|\chi|^2 \tag{4.84}$$

$$\mathrm{E}_Z\, r^4(\chi_1) = \frac{3}{4}|\chi|^4. \tag{4.85}$$

Proof. Using a polar decomposition of the numbers χ_1, χ_2, and z, we have with the obvious definitions,

$$\chi_1 = |\chi_1|e^{j\varphi_1}, \quad \chi_2 = |\chi_2|e^{j\varphi_2}, \quad z = |z|e^{j\varphi_z}. \tag{4.86}$$

The realization of the variable whose expectation is to be computed is then

$$\Big(\mathrm{Re}\big(|\chi_1||z|e^{j\varphi_1}e^{-j\varphi_z}\big)\Big)^{k_1}\Big(\mathrm{Re}\big(|\chi_2||z|e^{j\varphi_2}e^{-j\varphi_z}\big)\Big)^{k_2} \tag{4.87}$$

$$|\chi_1|^{k_1}|\chi_2|^{k_2}|z|^{k_1+k_2}\cos^{k_1}\big(\varphi_1-\varphi_z\big)\cos^{k_2}\big(\varphi_2-\varphi_z\big). \tag{4.88}$$

We carry out the expectation separately for the modulus and phase,

$$\mathrm{E}_Z\, |z|^{k_1+k_2}\cos^{k_1}\big(\varphi_1-\varphi_z\big)\cos^{k_2}\big(\varphi_2-\varphi_z\big) \tag{4.89}$$

$$= \mathrm{E}_{|Z|}\, |z|^{k_1+k_2}\, \mathrm{E}_{\Phi_z} \cos^{k_1}\big(\varphi_1-\varphi_z\big)\cos^{k_2}\big(\varphi_2-\varphi_z\big). \tag{4.90}$$

For the phase, Mathematica gives

$$\mathrm{E}_{\Phi_z} \cos^{k_1}\big(\varphi_1-\varphi_z\big)\cos^{k_2}\big(\varphi_2-\varphi_z\big) = \begin{cases} \frac{1}{2}\cos(\varphi_1-\varphi_2), & k_1=1, k_2=1 \\ \frac{3}{8}\cos(\varphi_1-\varphi_2), & k_1=1, k_2=3 \\ \frac{1}{8}\big(2+\cos 2(\varphi_1-\varphi_2)\big), & k_1=2, k_2=2 \\ 0, & k_1+k_2 \text{ odd.} \end{cases} \tag{4.91}$$

and for the modulus,

$$\mathrm{E}_{|Z|}\, |Z|^{k_1+k_2} = \begin{cases} 1, & k_1+k_2=2 \\ 2, & k_1+k_2=4. \end{cases} \tag{4.92}$$

□

We now put all terms together in Eq. (4.78), proceeding from γ up to γ^4. We use the appropriate equation in the previous lemma at each step.

First, using Eq. (4.80) in the term with γ, for which $k_1 = 1$ and $k_2 = 0$,

$$\mathrm{E}_Z[A_1] = \mathrm{E}_Z\big[-2\big(r(x) - r(\hat{x})\big)\big] = 0. \tag{4.93}$$

Then, for γ^2, Eq. (4.84) gives

$$\mathrm{E}_Z[A_2] = \mathrm{E}_Z\left[\sum_{x'\in\mathcal{X}} P(x')\Big(-|x-x'|^2 + 2r^2(x-x')\Big)\right] \tag{4.94}$$

$$= \sum_{x'\in\mathcal{X}} P(x')\Big(-|x-x'|^2 + |x-x'|^2\Big) = 0, \tag{4.95}$$

and

$$\mathrm{E}_Z\left[-\frac{1}{2}A_1^2\right] = -2\,\mathrm{E}_Z\Big[r^2(x-\hat{x})\Big] = -|x-\hat{x}|^2. \tag{4.96}$$

Now, for γ^3, Eq. (4.80) with $k_1 = 1$ or $k_1 = 3$ and $k_2 = 0$ yields

$$\mathrm{E}_Z[A_3] = \mathrm{E}_Z\left[\sum_{x'\in\mathcal{X}} P(x')\left(2|x-x'|^2 r(x-x') - \frac{4}{3}r^3(x-x')\right)\right] = 0, \tag{4.97}$$

and

$$\mathrm{E}_Z\left[\frac{1}{3}A_1^3\right] = -\frac{8}{3}\,\mathrm{E}_Z\Big[r^3(x-\hat{x})\Big] = 0, \tag{4.98}$$

and also

$$\mathrm{E}_Z[-A_1A_2] = 2\,\mathrm{E}_Z\left[r(x-\hat{x})\sum_{x''\in\mathcal{X}} P(x')\Big(-|x-x''|^2 + 2r^2(x-x'')\Big)\right] = 0. \tag{4.99}$$

And finally, for γ^4, Eqs. (4.84) and (4.85) give

$$\mathrm{E}_Z[A_4] = \sum_{x'\in\mathcal{X}} P(x')\,\mathrm{E}_Z\left[\frac{1}{2}|x-x'|^4 - 2|x-x'|^2 r^2(x-x') + \frac{2}{3}r^4(x-x')\right] \tag{4.100}$$

$$= \sum_{x'\in\mathcal{X}} P(x')\left(\frac{1}{2}|x-x'|^4 - |x-x'|^4 + \frac{1}{2}|x-x'|^4\right) = 0. \tag{4.101}$$

Next,

$$\mathrm{E}_Z\left[-\frac{1}{2}A_2^2\right] = -\frac{1}{2}\,\mathrm{E}_Z\left[\left(\sum_{x'\in\mathcal{X}} P(x')\Big(-|x-x'|^2 + 2r^2(x-x')\Big)\right)^2\right] \tag{4.102}$$

$$= -\frac{1}{2}\,\mathrm{E}_Z\left[\left(\sum_{x'\in\mathcal{X}} P(x')\Big(-|x-x'|^2 + 2r^2(x-x')\Big)\right)\times \right.$$
$$\left.\times\left(\sum_{x''\in\mathcal{X}} P(x'')\Big(-|x-x''|^2 + 2r^2(x-x'')\Big)\right)\right], \tag{4.103}$$

since x' and x'' are dummy variables. Multiplying the terms in brackets factor by factor and using Eqs. (4.84) and (4.85) in the Lemma, the summand corresponding to x' and x'' becomes

$$P(x')P(x'')|x-x'|^2|x-x''|^2\Big(1+\cos 2\Big(\varphi(x-x')-\varphi(x-x'')\Big)\Big) \tag{4.104}$$

$$= P(x')P(x'')|x-x'|^2|x-x''|^2\cos^2\Big(\varphi(x-x')-\varphi(x-x'')\Big) \tag{4.105}$$

$$= P(x')P(x'')\Big(\mathrm{Re}\big((x-x')(x-x'')^*\big)\Big)^2 \tag{4.106}$$

$$= P(x')P(x'')\Big(\mathrm{Re}\big(|x|^2 - xx''^* - x'x^* + x'x''^*\big)\Big)^2 \tag{4.107}$$

$$\begin{aligned} = P(x')P(x'')\Big(&|x|^4 + \big(\mathrm{Re}(xx''^*)\big)^2 + \big(\mathrm{Re}(x'x^*)\big)^2 + \big(\mathrm{Re}(x'x''^*)\big)^2 \\ &- 2|x|^2\,\mathrm{Re}(xx''^*) - 2|x|^2\,\mathrm{Re}(x'x^*) + 2|x|^2\,\mathrm{Re}(x'x''^*) \\ &+ 2\,\mathrm{Re}(xx''^*)\,\mathrm{Re}(x'x^*) - 2\,\mathrm{Re}(xx''^*)\,\mathrm{Re}(x'x''^*) \\ &- 2\,\mathrm{Re}(x'x^*)\,\mathrm{Re}(x'x''^*)\Big). \end{aligned} \tag{4.108}$$

Here we used that $1+\cos 2\alpha = 2\cos^2\alpha$. Then, we average over x, x' and x'' and combine some expectations since x, x' and x'' are dummy variables taking values in the same set. Using the definition of $\hat{x}$, we get

$$\begin{aligned} \mathrm{E}_{X,Z}\left[-\frac{1}{2}A_2^2\right] = -\,\mathrm{E}_{X,X'}\Big[&|X|^4 + 3\big(\mathrm{Re}(XX'^*)\big)^2 - 4|X|^2\,\mathrm{Re}(X\hat{x}^*) \\ &+ 2|X|^2|\hat{x}|^2 - 2\,\mathrm{Re}^2(X\hat{x}^*)\Big], \end{aligned} \tag{4.109}$$

Then, with Eqs. (4.82) and (4.83), the next term of γ^4 is

$$\mathrm{E}_Z\Big[-A_1A_3\Big] = 2\,\mathrm{E}_Z\left[r(x-\hat{x})\Big(\sum_{x'\in\mathcal{X}} P(x')\Big(2|x-x'|^2 r(x-x') - \frac{4}{3}r^3(x-x')\Big)\Big)\right] \tag{4.110}$$

$$= \sum_{x'\in\mathcal{X}} 2P(x')\Big(|x-\hat{x}||x-x'|^3\cos\big(\varphi(x)-\varphi(x')\big) - |x-\hat{x}||x-x'|^3\cos\big(\varphi(x)-\varphi(x')\big)\Big) = 0. \tag{4.111}$$

The following term in γ^4 becomes

$$\mathrm{E}_Z\Big[A_1^2A_2\Big] = \mathrm{E}_Z\left[4r^2(x-\hat{x})\Big(\sum_{x'\in\mathcal{X}} P(x')\Big(-|x-x'|^2 + 2r^2(x-x')\Big)\Big)\right] \tag{4.112}$$

$$= 2\sum_{x'} P(x')|x-\hat{x}|^2|x-x'|^2\Big(1+\cos 2\big(\varphi(x)-\varphi(x-x')\big)\Big) \tag{4.113}$$

$$= 4\sum_{x'} P(x')\Big(|x-\hat{x}|^2|x-x'|^2\cos^2\big(\varphi(x)-\varphi(x-x')\big)\Big) \tag{4.114}$$

$$= 4\sum_{x'} P(x')\Big(\mathrm{Re}\big((x-\hat{x})(x-x')^*\big)\Big)^2. \tag{4.115}$$

We now expand the factor $\Big(\mathrm{Re}\big((x-\hat{x})(x-x')^*\big)\Big)^2$,

$$\begin{aligned}\Big(\mathrm{Re}\big((x-\hat{x})(x-x')^*\big)\Big)^2 = {}& |x|^4 + \big(\mathrm{Re}(xx'^*)\big)^2 + \big(\mathrm{Re}(\hat{x}x^*)\big)^2 + \big(\mathrm{Re}(\hat{x}x'^*)\big)^2 \\ & - 2|x|^2\,\mathrm{Re}(xx'^*) - 2|x|^2\,\mathrm{Re}(\hat{x}x^*) + 2|x|^2\,\mathrm{Re}(\hat{x}x'^*) \\ & + 2\,\mathrm{Re}(xx'^*)\,\mathrm{Re}(\hat{x}x^*) - 2\,\mathrm{Re}(xx'^*)\,\mathrm{Re}(\hat{x}x'^*) \\ & - 2\,\mathrm{Re}(\hat{x}x^*)\,\mathrm{Re}(\hat{x}x'^*).\end{aligned} \tag{4.116}$$

Grouping terms, carrying out the averaging over X' and X, and using that the

variables are dummy, we get

$$\mathrm{E}_{X,Z}\Big[A_1^2 A_2\Big] = 4\,\mathrm{E}_{X,X'}\Big[|X|^4 + \big(\mathrm{Re}(XX'^*)\big)^2 + 2\big(\mathrm{Re}(\hat{x}X^*)\big)^2 - 4|X|^2\,\mathrm{Re}(X^*\hat{x}) + 2|X|^2|\hat{x}|^2 - 2|\hat{x}|^4\Big]. \tag{4.117}$$

Finally,

$$\mathrm{E}_Z\left[-\frac{1}{4}a_1^4\right] = -4\,\mathrm{E}_Z\Big[r^4(x)\Big] = -3|x-\hat{x}|^4. \tag{4.118}$$

This coefficient can also be expanded,

$$|x-\hat{x}|^4 = \big(|x|^2 + |\hat{x}|^2 - 2\,\mathrm{Re}(x^*\hat{x})\big)^2 \tag{4.119}$$

$$= |x|^4 + |\hat{x}|^4 + 4\big(\mathrm{Re}(x^*\hat{x})\big)^2 + 2|x|^2|\hat{x}|^2 - 4|x|^2\,\mathrm{Re}(x^*\hat{x}) - 4|\hat{x}|^2\,\mathrm{Re}(x^*\hat{x}), \tag{4.120}$$

and summed over x,

$$\mathrm{E}_{X,Z}\left[-\frac{1}{4}a_1^4\right] = -3\,\mathrm{E}_X\Big[|X|^4 + |\hat{x}|^4 + 4\big(\mathrm{Re}(X^*\hat{x})\big)^2 + 2|X|^2|\hat{x}|^2 - 4|X|^2\,\mathrm{Re}(X^*\hat{x}) - 4|\hat{x}|^4\Big] \tag{4.121}$$

$$= -3\,\mathrm{E}_X\Big[|X|^4 - 3|\hat{x}|^4 + 4\big(\mathrm{Re}(X^*\hat{x})\big)^2 + 2|X|^2|\hat{x}|^2 - 4|X|^2\,\mathrm{Re}(X^*\hat{x})\Big]. \tag{4.122}$$

We finally collect all the coefficients of powers of γ, up to γ^4. The non-zero contributions stem from Eqs. (4.96), (4.109), (4.117), and (4.122). Therefore,

$$\mathrm{C}_{\mathcal{X}} = c_1\gamma^2 + c_2\gamma^4 + \mathrm{O}(\gamma^5) \tag{4.123}$$

$$= c_1\,\mathrm{SNR} + c_2\,\mathrm{SNR}^2 + \mathrm{O}(\mathrm{SNR}^{5/2}). \tag{4.124}$$

The first-order coefficient (in γ^2) is recovered from Eq. (4.96),

$$c_1 = \mathrm{E}_X[|X-\hat{x}|^2]. \tag{4.125}$$

The second-order coefficient of the constrained capacity (in γ^4) is derived by grouping Eqs. (4.109), (4.117), and (4.122),

$$\begin{aligned} c_2 = \mathrm{E}_{X,X'}\Big[&|X|^4 + 3\big(\mathrm{Re}(XX'^*)\big)^2 - 4|X|^2\,\mathrm{Re}(X\hat{x}^*) + 2|X|^2|\hat{x}|^2 - 2\,\mathrm{Re}^2(X\hat{x}^*) \\ &- 4|X|^4 - 4\big(\mathrm{Re}(XX'^*)\big)^2 - 8\big(\mathrm{Re}(\hat{x}X^*)\big)^2 + 16|X|^2\,\mathrm{Re}(X^*\hat{x}) \\ &- 8|X|^2|\hat{x}|^2 + 8|\hat{x}|^4 + 3|X|^4 - 9|\hat{x}|^4 + 12\big(\mathrm{Re}(X^*\hat{x})\big)^2 \\ &+ 6|X|^2|\hat{x}|^2 - 12|X|^2\,\mathrm{Re}(X^*\hat{x})\Big] \qquad (4.126) \\ = \mathrm{E}_{X,X'}\Big[&2\big(\mathrm{Re}(X^*\hat{x})\big)^2 - \big(\mathrm{Re}(XX'^*)\big)^2 - |\hat{x}|^4\Big]. \qquad (4.127) \end{aligned}$$

Finally, we verify that c_2 coincides with

$$c_2 = -\frac{1}{2}\Big(\mathrm{E}_X^2\big[|X-\hat{x}|^2\big] + \big|\mathrm{E}_X[(X-\hat{x})^2]\big|^2\Big). \qquad (4.128)$$

To do so, let us expand the summands in Eq. (4.127). First,

$$\begin{aligned} \mathrm{E}_{X,X'}\big(\mathrm{Re}(XX'^*)\big)^2 &= \mathrm{E}_{X,X'}\left[\frac{(XX'^*)^2 + (X^*X')^2 + 2|X|^2|X'|^2}{4}\right] \qquad (4.129) \\ &= \frac{1}{2}\Big(\mathrm{Re}\big(\mathrm{E}_X[X^2]\,\mathrm{E}_{X'}[X'^2]^*\big) + \mathrm{E}_X\,|X|^2\;\mathrm{E}_{X'}\,|X'|^2\Big) \qquad (4.130) \\ &= \frac{1}{2}\Big(\big|\mathrm{E}_X[X^2]\big|^2 + \mathrm{E}_X^2\,|X|^2\Big). \qquad (4.131) \end{aligned}$$

Similarly, we obtain

$$\mathrm{E}_X\big(\mathrm{Re}(X^*\hat{x})^2 = \frac{1}{2}\Big(\mathrm{Re}\big(\hat{x}^2\,\mathrm{E}_X^*[X^2]\big) + |\hat{x}|^2\,\mathrm{E}_X\,|X|^2\Big). \qquad (4.132)$$

Therefore,

$$c_2 = \mathrm{Re}\big(\hat{x}^2\,\mathrm{E}_X^*[X^2]\big) + |\hat{x}|^2\,\mathrm{E}_X\,|X|^2 - \frac{1}{2}\big|\mathrm{E}_X[X^2]\big|^2 - \frac{1}{2}\,\mathrm{E}_X^2\,|X|^2 - |\hat{x}|^4, \qquad (4.133)$$

which indeed coincides with Eq. (4.128), since

$$c_2 = -\frac{1}{2}\left(\mathrm{E}_X^2\left[|X-\hat{x}|^2\right] + \left|\mathrm{E}_X[(X-\hat{x})^2]\right|^2\right) \tag{4.134}$$

$$= -\frac{1}{2}\left(\left(\mathrm{E}_X\,|X|^2 - |\hat{x}|^2\right)^2 + \left|\mathrm{E}_X\,X^2 - \hat{x}^2\right|^2\right) \tag{4.135}$$

$$= -\frac{1}{2}\Bigg(\mathrm{E}_X^2\,|X|^2 + |\hat{x}|^4 - 2|\hat{x}|^2\,\mathrm{E}_X\,|X|^2 + |\,\mathrm{E}_X\,X^2|^2 + |\hat{x}^4| - 2\,\mathrm{Re}(\hat{x}^2\,\mathrm{E}_X^*\,X^2)\Bigg). \tag{4.136}$$

4.C Determination of the Power and Bandwidth Trade-Off

In order to have the same capacities bandwidth and/or power must change to account for the difference in capacity, so that

$$c_{11}\frac{P_1}{N_0} + c_{21}\frac{P_1^2}{W_1 N_0^2} + \mathrm{o}(W_1\,\mathrm{SNR}_1^2) = c_{12}\frac{P_2}{N_0} + c_{22}\frac{P_2^2}{W_2 N_0^2} + \mathrm{o}(W_2\,\mathrm{SNR}_2^2). \tag{4.137}$$

Simplifying common factors, we obtain

$$c_{11} + c_{21}\,\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1) = c_{12}\frac{P_2}{P_1} + \big(c_{22} + \mathrm{o}(\mathrm{SNR}_1)\big)\frac{P_2^2}{P_1^2}\frac{W_1}{W_2}\,\mathrm{SNR}_1\,. \tag{4.138}$$

Or, with the definitions $\Delta P = P_2/P_1$, and $\Delta W = W_2/W_1$,

$$c_{11} + c_{21}\,\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1) = c_{12}\Delta P + \big(c_{22}\,\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1)\big)\frac{(\Delta P)^2}{\Delta W}, \tag{4.139}$$

and

$$\Delta W = \frac{\big(c_{22}\,\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1)\big)(\Delta P)^2}{c_{11} + c_{21}\,\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1) - c_{12}\Delta P}. \tag{4.140}$$

This equation gives the trade-off between ΔP and ΔW, for a fixed (small) SNR_1, so that the capacities of scenarios 1 and 2 coincide.

Next we solve for the inverse, i. e. for ΔP as a function of ΔP. First, let us define the quantities $a = c_{22}\,\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1)$ and $b = c_{11} + c_{21}\,\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1)$.

Rearranging Eq. (4.140) we have $a(\Delta P)^2 + c_{12}\Delta W \Delta P - b\Delta W = 0$ and therefore

$$\Delta P = \frac{-c_{12}\Delta W \pm \sqrt{(c_{12}\Delta W)^2 + 4ab\Delta W}}{2a} \tag{4.141}$$

$$= \frac{c_{12}\Delta W}{2a}\left(-1 \pm \sqrt{1 + \frac{4ab}{c_{12}^2 \Delta W}}\right). \tag{4.142}$$

Often we have $c_{22} < 0$, and then the negative root is a spurious solution. We choose then the positive root. Since ab is of order SNR_1, we can use the Taylor expansion $(1+4t)^{1/2} = 1 + 2t - 2t^2 + \mathrm{o}(t^2)$, to write

$$\Delta P = \frac{c_{12}\Delta W}{2a}\left(\frac{2ab}{c_{12}^2 \Delta W} - \frac{2a^2 b^2}{c_{12}^4 (\Delta W)^2}\right) \tag{4.143}$$

$$= \frac{b}{c_{12}} - \frac{ab^2}{c_{12}^3 \Delta W}. \tag{4.144}$$

Since $\mathrm{SNR}_1 \to 0$, we group the non-linear terms in SNR_1 and so get

$$\Delta P = \frac{c_{11} + c_{21}\,\mathrm{SNR}_1}{c_{12}} - \frac{c_{22}c_{11}^2\,\mathrm{SNR}_1}{c_{12}^3 \Delta W} + \mathrm{o}(\mathrm{SNR}_1) \tag{4.145}$$

$$= \frac{c_{11}}{c_{12}} + \left(\frac{c_{21}}{c_{12}} - \frac{c_{22}c_{11}^2}{c_{12}^3 \Delta W}\right)\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1). \tag{4.146}$$

5

Digital Modulation in the Additive Energy Channels

5.1 Introduction

In this chapter, we extend the analysis of digital modulation of Chapter 4 to the family of additive energy channels. As in Chapter 4, the presentation is built around the concept of mutual information achievable with a given constellation set $\mathcal{X}$; we call this mutual information *constrained coded modulation capacity* and denote it by the symbol $\mathrm{C}_{\mathcal{X}}$. The constellation we consider is pulse energy modulation (PEM), a set of non-negative real numbers x, used with probabilities $P(x)$. The name PEM is in agreement with the standard terminology used in Gaussian channels or in optical communications, e. g. pulse-amplitude modulation (PAM). We reserve the word amplitude to refer to the quadrature amplitude, a complex-valued quantity, meanwhile the word energy refers to the squared modulus of the (quadrature) amplitude.

As we saw in Chapter 4, in amplitude modulation the constellation points have the form

$$\pm\beta_{\mathrm{PAM}}\left(\frac{1}{2}+(i-1)\right), \quad i=1,\ldots,2^{m-1}, \tag{5.1}$$

where β_{PAM} is a normalization factor fixed to ensure average unit energy. All points are used with the same probability, namely $1/2^m$. By construction, consecutive points are separated by a distance β_{PAM}. A straightforward extension to the additive energy channels would be to consider a set of the form,

$$\beta(i-1), \quad i=1,\ldots,2^m, \tag{5.2}$$

where β is another normalization factor which also ensures average unit energy. Instead, we consider a slightly more general constellation set constructed by

choosing a positive real number $\lambda > 0$ and taking the constellation points

$$\beta_{\mathrm{PEM}}\big((i-1)^{\lambda}\big), \quad i = 1, \ldots, 2^m, \tag{5.3}$$

where β_{PEM} is a third normalization factor. Further details on the constellation are given in Section 5.2. Note that uniform PEM is recovered by setting $\lambda = 1$. Having at our disposition a family of constellations dependent on a parameter λ allows us to determine the optimum value of λ for a given channel property. As we shall see in Chapter 6, the pairwise error probability is minimized for $\lambda = 1$ in the additive exponential noise channel and for $\lambda = 2$ in the discrete-time Poisson channel. In this chapter, we determine the values of λ which maximize the constrained coded modulation capacity at high and low signal energy.

We carry out the analysis for the three additive energy channels of Chapter 2, namely

1. The additive exponential noise channel (AEN).
2. The discrete-time Poisson channel (DTP).
3. The (quantized) additive energy channel (AE-Q).

Of these, the DTP channel in particular possesses some practical interest, since it is often used to model optical communication systems. Even though there are some results in the literature on the capacity of such channels, as we saw in Chapter 3, there seems to be no extensive study of the capabilities of non-binary modulation in this channel (see the recent review [15]). The results presented here are a contribution to fill this gap.

The text is sequentially organized, respectively covering the AEN, DTP, and AE-Q channels in Sections 5.3, 5.4 and 5.5. For each channel, the presentation follows closely the structure of the Gaussian case, the study being further split into a part on coded modulation, including the capacity per unit energy and the shaping gain, and a second part on bit-interleaved coded modulation (BICM). Specifically, we compute the BICM capacity and study its closeness to $\mathrm{C}_{\mathcal{X}}$ and conclude that BICM is likely to constitute a good alternative for the code design in these channels.

Particular attention will be paid to the performance in the low energy regime, as was done in the Gaussian channel. Prelov and van der Meulen, in a series of papers [27, 28], considered a general discrete-time additive channel model and determined the asymptotic Taylor expansion at zero signal-to-noise ratio. Our work differs from theirs in that the additive energy channels are constrained on the mean value of the input, rather than the variance, and similarly

the noise is described by its mean, not its variance; the models considered by Prelov and van der Meulen rather deal with channels where the second-order moments, both for signal energy and noise level, are of importance. Our work extends their analysis to the family of additive energy channels, where the first-order moments are constrained.

Another relevant body of work is that of Guo, Shamai, and Verdú [60] and Palomar and Verdú, [61], who determined the form of the first derivative of the constrained capacity $\mathrm{C}_{\mathcal{X}}$ for arbitrary signal-to-noise ratio by exploiting a link with estimation theory, e. g. with the minimum mean square error (MMSE) estimator in the Gaussian channel. Their tools, applied to the additive energy channels, could be used to compute the first derivative at zero signal energy, one of the quantities we determine. However, the extension to compute the second-order derivative does not seem to be straightforward, so we have chosen to apply the proof method developed in Chapter 4 to find the value of the first derivatives of $\mathrm{C}_{\mathcal{X}}$ at zero signal energy.

Finally, we show that the minimum energy per bit of the Gaussian channel, namely $\sigma^2 \log 2$, coincides with the minimum energy per bit of the additive energy channels, when the average level of the additive noise in the latter models is set to σ^2. Unlike the Gaussian channel, where the number $\log 2$, or $-1.59\,\mathrm{dB}$ was a universal constant at low SNR for general modulation formats, the performance at low energy strongly depends on the size of the constellation and the channel model.

5.2 Constellations for Pulse Energy Modulation

By construction, the input symbols in the additive energy channels are real non-negative numbers. The optimum input distribution for the AEN channel was determined by Verdú [21], a result recalled in Eq. (3.18), and has the form

$$p_X(x) = \frac{E_s}{(E_s+E_n)^2} e^{-\frac{x}{E_s+E_n}} + \frac{E_n}{E_s+E_n}\delta(x), \quad x \geq 0. \tag{5.4}$$

Here E_n is the noise average energy and E_s the constraint on the average signal energy. In the DTP channel, the optimum input is not known, but the density in Eq. (3.48), from Chapter 3, was found to be a good choice, especially for large quanta counts. This density is given by

$$p_X(x) = \frac{(1+2\varepsilon_s)^{1/2}}{\sqrt{2e}-1+(1+2\varepsilon_s)^{1/2}} \frac{1}{\sqrt{2\pi x \varepsilon_s}} e^{-\frac{x}{2\varepsilon_s}} + \frac{\sqrt{2e}-1}{\sqrt{2e}-1+(1+2\varepsilon_s)^{1/2}}\delta(x), \tag{5.5}$$

where ε_s is the average number of quanta per channel use. Qualitatively, the optimum density for the AE-Q channel should lie between these two densities,

its functional form depending on the precise values of ε_s and on the average number of quanta of additive noise, ε_n.

In principle, the optimum densities for the AEN and AE-Q channels depend on the value of the signal (E_s or ε_s) and noise (E_n or ε_n) levels. In this chapter, we consider an alternative input, namely scaled constellations $\alpha\mathcal{X}$, where $\mathcal{X}$ has unit energy and α is the average signal energy (E_s or ε_s). This pulse-energy modulation (PEM) follows the common practice in the Gaussian channels with amplitude or phase modulations. Additionally, we let the discrete constellation $\mathcal{X}$ depend on a free parameter λ, which will be optimized in later sections to approximate, in a sense to be made precise later, the optimum input densities.

Let the parameter λ be a positive real number, $\lambda > 0$. We define the set of constellation symbols $\mathcal{X}_\lambda$ as the 2^m points

$$\mathcal{X}_\lambda = \left\{\beta(i-1)^\lambda\right\}, \quad i = 1, \ldots, 2^m, \tag{5.6}$$

where β is a normalization factor, $\beta^{-1} = \frac{1}{2^m}\sum_{i=1}^{2^m}(i-1)^\lambda$. We assume that the points are used with identical probabilities, namely $1/2^m$. When $\lambda = 1$, we recover a uniform PEM constellation with equispaced, equiprobable symbols.

As the number of points 2^m increases, the discrete constellation $\mathcal{X}_\lambda$ approaches a continuous distribution, which we denote by $\mathcal{X}_\lambda^\infty$, and whose form can be seen in Fig. 5.1 for several values of λ. For $\lambda > 1$ the constellation density, being less flat than the uniform, is somewhat closer to the optimum densities for the AEN and DTP channels above. We shall exploit this observation to determine the optimum λ for these channel models. To the best of our knowledge, this study is new, the problem of coded modulation in the AEN channel not having been studied in the past, and the study of the DTP channel having focused on the analysis of the uncoded error probability.

The main features of the limiting continuous constellation $\mathcal{X}_\lambda^\infty$ are listed in

Proposition 5.1. *The constellation $\mathcal{X}_\lambda^\infty$ has bounded support in the interval $[0, \lambda+1]$ and a non-uniform density given by*

$$p_X(x) = \frac{x^{\frac{1}{\lambda}-1}}{\lambda(1+\lambda)^{\frac{1}{\lambda}}} \tag{5.7}$$

within the interval $[0, \lambda+1]$ and zero outside. Its mean (or first-order moment) $\mu_1(\mathcal{X}_\lambda^\infty)$ is one, and its second-order moment is given by $\mu_2(\mathcal{X}_\lambda^\infty) = \frac{(\lambda+1)^2}{2\lambda+1}$.

A scaled constellation $\alpha\mathcal{X}_\lambda$, with $\alpha > 0$, has density

$$f(x) = \frac{x^{\frac{1}{\lambda}-1}}{\alpha^{\frac{1}{\lambda}}\lambda(1+\lambda)^{\frac{1}{\lambda}}} \tag{5.8}$$

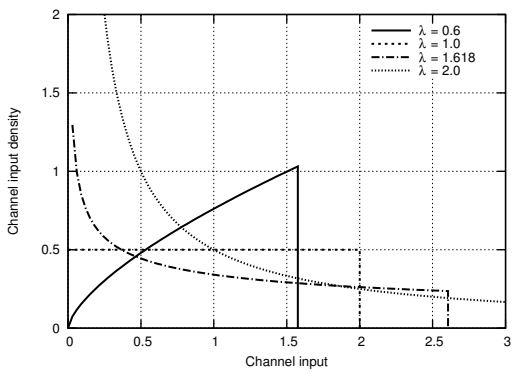


Figure 5.1: Density of the channel input $\mathcal{X}_\lambda^\infty$ for PEM with parameter λ.

in the interval $[0, \alpha(\lambda+1)]$ *and zero outside it.*

The differential entropy of the continuous constellation $\alpha\mathcal{X}_\lambda$ *is given by*

$$H(\mathcal{X}_\lambda^\infty) = 1 - \lambda + \log\bigl(\lambda(\lambda+1)\bigr) + \log\alpha. \tag{5.9}$$

As it should, the uniform distribution $\lambda = 1$ is distributed in the interval (0,2), with density 1/2, and has a differential entropy $\log 2$.

Proof. We begin by computing the support. It is obvious that the interval starts at $x = 0$. As for the upper limit $x_{\max}$, its inverse is clearly given by

$$\frac{1}{x_{\max}} = \frac{1}{2^m}\sum_{i=1}^{2^m}\left(\frac{i-1}{2^m-1}\right)^\lambda. \tag{5.10}$$

As $2^m \to \infty$, the sum can be approximated by a Riemann integral. Hence,

$$\lim_{2^m\to\infty}\frac{1}{x_{\max}} = \int_0^1 x^\lambda\,dx = \frac{1}{\lambda+1}. \tag{5.11}$$

This number is finite, and the upper limit $x_{\max}$ approaches $\lambda+1$ as $2^m \to \infty$.

The support is bounded to $[0, \lambda+1]$. We have just seen that points are selected uniformly in the interval $[0,1]$. This uniform choice induces a non-uniform measure in the interval $[0, \lambda+1]$; this measure is the density we seek. Let t denote a variable in $[0,1]$ and $x = f(t) = (1+\lambda)t^\lambda$ be the function

changing the measure. Then, since the measure dx of an infinitesimal interval around x, must coincide with the measure dt of the corresponding interval around t, that is $dx = (1+\lambda)\lambda t^{\lambda-1}\,dt$, we have that

$$dt = \frac{1}{(1+\lambda)\lambda t^{\lambda-1}}dx = \frac{x^{\frac{1}{\lambda}-1}}{(1+\lambda)^{\frac{1}{\lambda}}\lambda}dx. \tag{5.12}$$

The formulas for the scaled constellation follow from a new change of measure, now $x' = \alpha x$. Since the probability must remain unchanged, the density of x' therefore satisfies $f(x')\,dx' = f(x)\,dx$. Since $dx' = \alpha dx$, we have

$$f(x') = f(x)\frac{dx}{dx'} = \frac{x'^{\frac{1}{\lambda}-1}}{\alpha^{\frac{1}{\lambda}}\lambda(1+\lambda)^{\frac{1}{\lambda}}}. \tag{5.13}$$

The entropy of the limiting scaled constellation may be directly computed from its density, or obtained from Theorem 9.6.4 of [22]. □

5.3 Coded Modulation in the Exponential Noise Channel

5.3.1 Constrained Capacity

The channel model for the exponential noise channel was presented in Chapter 2, and we now briefly review it. The channel output consists of a real-valued vector $\mathbf{y} = (y_1, \ldots, y_n)$. For each index k, the output y_k is given by the sum

$$y_k = \mathrm{SNR}\, x_k + z_k, \tag{5.14}$$

where the z_k are independent samples of additive noise, exponentially distributed as $Z_k \sim \mathcal{E}(1)$, SNR is the average signal-to-noise ratio, and x_k is the channel input. The channel capacity, of value $\mathrm{C(SNR)} = \log(1+\mathrm{SNR})$, is achieved when the channel input has the density given by Eq. (5.4). Since the capacity cost function has the same form as in the Gaussian channel, the capacity per unit energy also coincides with its value in the Gaussian channel, an assertion which will be verified in Section 5.3.2.

In this section, we study the constrained capacity for pulse energy modulation (PEM), where the input symbols are drawn from a constellation set $\mathcal{X}$. Symbols in $\mathcal{X}$ are used with probabilities $P(x)$, and have arbitrary first- and second-order moments, $\mu_1(\mathcal{X})$ and $\mu_2(\mathcal{X})$.

The constellation is assumed to have $|\mathcal{X}| = 2^m$ elements. For later use, we add a point at infinity, defined as $x_{|\mathcal{X}|+1} = \infty$. It is useful to sort the symbols in increasing order, i. e. $x_1 \leq x_2 \leq \ldots \leq x_{|\mathcal{X}|}$.

As a final notational convention, the conditional output density $Q(y|x)$ is

$$Q(y|x) = e^{-(y-\mathrm{SNR}\,x)} u(y - \mathrm{SNR}\,x), \tag{5.15}$$

where $u(t)$ is the step function, so that the density is zero for $y < \mathrm{SNR}\,x$.

In contrast with the Gaussian case, there exists a closed-form expression for the constrained capacity,

Proposition 5.2. *In the AEN channel, the constrained capacity $C_{\mathcal{X}}(\mathrm{SNR})$ for signalling over a fixed modulation set $\mathcal{X}$ with probabilities $P(x)$ with an average signal-to-noise ratio* SNR *is given by*

$$C_{\mathcal{X}}(\mathrm{SNR}) = -\sum_{x\in\mathcal{X}} P(x) \sum_{x_j \geq x} \left(e^{\mathrm{SNR}(x-x_j)} - e^{\mathrm{SNR}(x-x_{j+1})}\right) \log\left(\sum_{x' \leq x_j} P(x') e^{-\mathrm{SNR}(x-x')}\right). \tag{5.16}$$

Proof. Eq. (3.3) on page 27, gives the mutual information across a communications channel. Writing down the expectations over all input symbols x and noise realizations z, we obtain

$$\mathrm{C}_{\mathcal{X}} = -\sum_{l=1}^{|\mathcal{X}|} P(x_l) \int_0^\infty e^{-z} \log\left(\sum_{x'\in\mathcal{X}} P(x') e^{-\gamma(x_l - x')} u\big(\gamma(x_l - x') + z\big)\right) dz. \tag{5.17}$$

We split the integral into consecutive sections of length $\gamma(x_{j+1} - x_j)$, starting at $j = l$ (the sent symbol) and extending the sum to infinity with the convention $x_{|\mathcal{X}|+1} = \infty$. Then

$$\int_0^\infty e^{-z} f(x_l, z)\, dz = \sum_{j=l}^{|\mathcal{X}|} \int_{\gamma(x_j - x_l)}^{\gamma(x_{j+1} - x_l)} e^{-z} f(x_l, z)\, dz, \tag{5.18}$$

where $f(x_l, z)$ is the logarithm function in Eq. (5.17). Within each interval the function $f(x_l, z)$ is constant, since

$$u\big(\gamma(x_l - x') + z\big) = 1, \quad \text{for } x' \leq x_j, \tag{5.19}$$

$$u\big(\gamma(x_l - x') + z\big) = 0, \quad \text{for } x' > x_j, \tag{5.20}$$

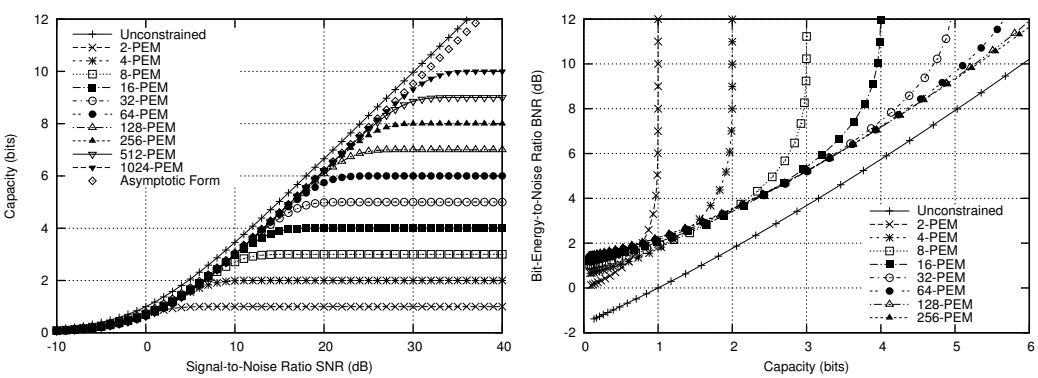
which allows us to use the integral $\int_{z_1}^{z_2} e^{-z}\,dz = e^{-z_1} - e^{-z_2}$ and derive

$$\mathrm{E}_Z\big[f(x_l, z)\big] = \sum_{j=l}^{|\mathcal{X}|}\big(e^{-\gamma(x_j - x_l)} - e^{-\gamma(x_{j+1} - x_l)}\big) f\big(x_l, \gamma(x_j - x_l)\big). \tag{5.21}$$

Substituting this expression in Eq. (5.17) gives the desired formula. □

The proof can be easily extended to apply to constellations with continuous, rather than discrete, support.

Figure 5.2 shows the constrained capacity for the uniform 2^m-PEM ($\lambda = 1$ in the constellations of Section 5.2). In Fig. 5.2a, capacity is plotted as a function of SNR, whereas in Fig. 5.2b, the bit-energy-to-noise ratio BNR, defined as $\mathrm{BNR} = \frac{\mathrm{SNR}}{\mathrm{C}_{\mathcal{X}}} \log 2$, is given as a function of the capacity. Figure 5.3 depicts the capacity $\mathrm{C}_{\mathcal{X}}$ for the same values of SNR using a constellation with $\lambda = \frac{1}{2}(1 + \sqrt{5})$. The reason for this specific choice will become apparent later.

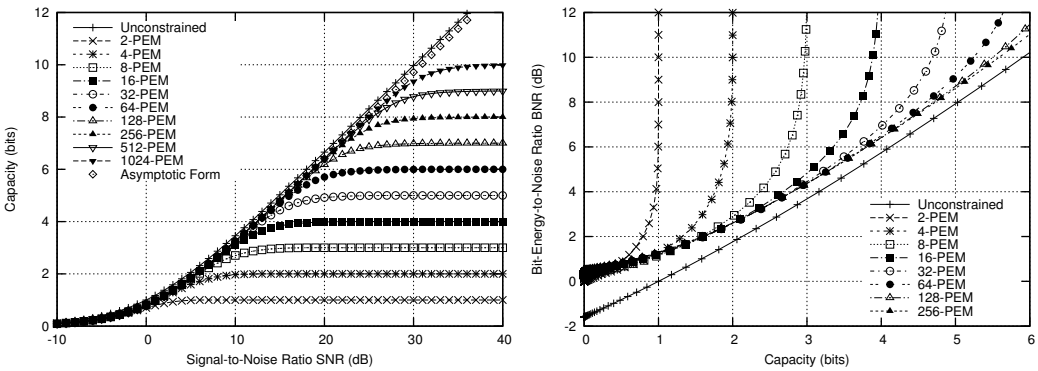


(a) Capacity as a function of SNR.

(b) BNR as a function of the capacity.

Figure 5.2: CM capacity for uniform 2^m-PEM.

As we did for the Gaussian channel in Chapter 4, we characterize analytically the behaviour of $\mathrm{C}_{\mathcal{X}}$ at low and high SNR. We start by computing the capacity per unit energy in Section 5.3.2 and relating it to the minimum energy per bit $\mathrm{BNR}_{\min}$. In Section 5.3.3, we compute the first two terms of the Taylor expansion of the capacity around SNR = 0, and relate them to the quantity BNR_0, attained as SNR tends to zero. As clearly seen in Figs. 5.2a and 5.3a,



(a) Capacity as a function of SNR.

(b) BNR as a function of the capacity.

Figure 5.3: CM capacity for equiprobable 2^m-PEM with $\lambda = \frac{1}{2}(1+\sqrt{5})$.

the capacity per unit energy is not universally achieved by arbitrary constellation sets: 2-PEM has a minimum BNR_0 of 0 dB, and an asymptotic value for BNR_0 is attained as m increases for both values of the parameter λ; this asymptotic value will also be computed. Finally, we provide constellation sets for which BNR_0 is arbitrarily close to $\mathrm{BNR}_{\min}$.

At high SNR, the curves for the capacity seem to approach a common envelope as the number of constellation points increases. To emphasize this fact, two dotted lines are shown; their form will be related in Section 5.3.4 to the differential entropy of the input. Both functions are very close to the respective envelope of the capacity, and are rather close to the channel capacity, especially in Fig. 5.3a.

Lastly, we extend the analysis of bit-interleaved coded modulation to this channel in Section 5.3.5. Curves for the BICM capacity, similar to those depicted in Figs. 5.2a and 5.3a will be obtained.

5.3.2 Capacity per Unit Energy

The capacity per unit energy is the largest number of bits per symbol which can be reliably sent over the channel per unit energy. Since the capacity equals that of the complex-valued AWGN channel, it follows that the capacity per unit cost is also equal to that of the AWGN channel, namely $\mathrm{C}_1 = E_n^{-1}$. The minimum energy per bit $E_{b,\min}$ is obtained by computing $E_{b,\min} = \frac{\log 2}{\mathrm{C}_1} = E_n \log 2$.

It takes the same energy to transmit a bit in the complex-valued AWGN and the AEN channels when the noise variance of the AWGN channel, σ^2, equals the noise mean of the AEN channel, E_n. This fact was to be expected since the channel capacity is $\log(1+\mathrm{SNR})$ in both cases. The minimum bit-energy-to-noise ratio $\mathrm{BNR}_{\min}$, given by $\mathrm{BNR}_{\min} = \frac{E_{b,\min}}{E_n} = \log 2$, equals $-1.59\,\mathrm{dB}$, in agreement with the results depicted in Figs. 5.2b and 5.3b.

Alternatively, we can use Theorem 3.9 on page 30, to determine C_1,

$$\mathrm{C}_1 = \sup_x \frac{D\big(Q(y|x)||Q(y|x=0)\big)}{E_s x} \tag{5.22}$$

$$= \sup_x \frac{1}{E_s x} \int_{\mathrm{SNR}\,x}^{\infty} e^{-(y-\mathrm{SNR}\,x)} \log \frac{e^{-(y-\mathrm{SNR}\,x)}}{e^{-y}}\, dy \tag{5.23}$$

$$= \sup_x \frac{1}{E_n} \int_{\mathrm{SNR}\,x}^{\infty} e^{-(y-\mathrm{SNR}\,x)}\, dy = \frac{1}{E_n} \sup_x 1 = \frac{1}{E_n}. \tag{5.24}$$

To write Eq. (5.23), we used the definition of the divergence between two probability distributions and the form of the channel transition probability $Q(y|x)$ in Eq. (5.15). The remaining steps are straightforward.

5.3.3 Asymptotic Behaviour at Low SNR

We start by characterizing the behaviour at low SNR of the constrained capacity $\mathrm{C}_{\mathcal{X}}$ in

Proposition 5.3. *In the AEN channel, the constrained capacity $C_{\mathcal{X}}(\mathrm{SNR})$ using a signal set $\mathcal{X}$ with average energy constraint* SNR*, admits a Taylor series expansion of the form $C_{\mathcal{X}}(\mathrm{SNR}) = c_1\,\mathrm{SNR} + c_2\,\mathrm{SNR}^2 + O(\mathrm{SNR}^3)$, as* $\mathrm{SNR} \to 0$*, with c_1 and c_2 given by*

$$c_1 = -\sum_{j=1}^{|\mathcal{X}|-1} (x_{j+1} - x_j) q_j \log q_j \tag{5.25}$$

$$c_2 = \frac{1}{2}\sigma^2(\mathcal{X}) + \sum_{j=1}^{|\mathcal{X}|-1} (x_{j+1} - x_j)\Big(\tfrac{1}{2}(x_j + x_{j+1}) q_j - q'_j\Big) \log q_j, \tag{5.26}$$

where $\sigma^2(\mathcal{X}) = \mu_2(\mathcal{X}) - \mu_1^2(\mathcal{X})$, $q_j = \sum_{x' \le x_j} P(x')$ and $q'_j = \sum_{x' \le x_j} x' P(x')$.

Proof. The detailed derivation can be found in Appendix 5.A. □

We verify next that this characterization indeed describes the results in Fig. 5.2. For the special case of equiprobable 2^m-PEM, the proposition becomes

Corollary 5.4. *In the AEN channel, 2^m-PEM modulation with parameter λ has*

$$c_1 = -\frac{2^m}{\sum_{j=1}^{2^m}(j-1)^\lambda} \sum_{j=1}^{2^m-1} \left(j^\lambda - (j-1)^\lambda\right) \frac{j}{2^m} \log \frac{j}{2^m}. \tag{5.27}$$

For 2-PEM, $c_1 = \log 2$ and $BNR_0 = 1$, or 0 dB. As $m \to \infty$, an asymptotic limit is reached,

$$\lim_{m\to\infty} c_1 = \frac{\lambda}{\lambda+1}, \tag{5.28}$$

$$\lim_{m\to\infty} c_2 = -\frac{\lambda^2}{2(1+2\lambda)^2}, \tag{5.29}$$

and therefore

$$\lim_{m\to\infty} BNR_0 = \frac{\lambda+1}{\lambda} \log 2. \tag{5.30}$$

As $\lambda \to \infty$ and $m \to \infty$, BNR_0 approaches $\log 2$, the value derived from the capacity per unit energy; under the same conditions c_2 approaches $-\frac{1}{8}$.

Proof. Eq. (5.27) immediately follows from the constellation definition.

As $m \to \infty$, the probability $q_j = \sum_{x' \le x_j} P(x')$ of the limiting continuous distribution is given by the Riemann integral

$$q_j = \int_0^{x_j} \frac{x^{\frac{1}{\lambda}-1}}{\lambda(1+\lambda)^{\frac{1}{\lambda}}}\, dx = \left(\frac{x_j}{\lambda+1}\right)^{\frac{1}{\lambda}}. \tag{5.31}$$

Therefore, the first-order coefficient is given by

$$c_1 = -\int_0^{\lambda+1} \left(\frac{x_j}{\lambda+1}\right)^{\frac{1}{\lambda}} \log\left(\frac{x_j}{\lambda+1}\right)^{\frac{1}{\lambda}} dx_j = \frac{\lambda}{\lambda+1}. \tag{5.32}$$

Next, the parameter q'_j is given by the integral

$$q'_j = \int_0^{x_j} x \frac{x^{\frac{1}{\lambda}-1}}{\lambda(1+\lambda)^{\frac{1}{\lambda}}}\, dx = \left(\frac{x_j}{\lambda+1}\right)^{1+\frac{1}{\lambda}}. \tag{5.33}$$

Approximating the sum in the formula for c_2 by an integral, and setting $\frac{1}{2}(x_j + x_{j+1}) = x_j$, we get

$$x_j q_j - q'_j = x_j \left(\frac{x_j}{\lambda+1}\right)^{\frac{1}{\lambda}} - \left(\frac{x_j}{\lambda+1}\right)^{1+\frac{1}{\lambda}} = \lambda \left(\frac{x_j}{\lambda+1}\right)^{1+\frac{1}{\lambda}}, \tag{5.34}$$

and, since the constellation variance is $\sigma^2(\mathcal{X}) = \frac{\lambda^2}{1+2\lambda}$, we get

$$c_2 = \frac{1}{2}\frac{\lambda^2}{1+2\lambda} + \int_0^{\lambda+1} \lambda \left(\frac{x_j}{\lambda+1}\right)^{1+\frac{1}{\lambda}} \log\left(\frac{x_j}{\lambda+1}\right)^{\frac{1}{\lambda}} dx_j \tag{5.35}$$

$$= \frac{1}{2}\frac{\lambda^2}{1+2\lambda} - \frac{\lambda^2(\lambda+1)}{(1+2\lambda)^2} = -\frac{\lambda^2}{2(1+2\lambda)^2}. \tag{5.36}$$

□

As seen in Fig. 5.2b, the value of 2-PEM indeed approaches zero capacity at 0 dB. Uniform PEM requires a bit energy of $2\log 2$ as the number of points increases, a loss of 3 dB compared to the asymptotic value in the AWGN case. Use of the optimum constellation, with $\lambda = \frac{1}{2}(1+\sqrt{5})$ reduces this loss by about 1 dB, since $\mathrm{BNR}_0 = \frac{1}{2}(1+\sqrt{5})\log 2$, or 0.50 dB. Further reductions are possible by increasing λ, but $\mathrm{BNR}_{\min}$ is approached relatively slowly.

A more efficient method of attaining the minimum $\mathrm{BNR}_{\min}$ is by using flash signalling [25], a generalized form of binary modulation with two points at positions 0, with probability p, and $1/(1-p)$, with probability $1-p$. Indeed

Corollary 5.5. *The coefficients c_1 and c_2 of binary modulation with one symbol at $x = 0$ used with probability p are*

$$c_1 = -\frac{p}{1-p}\log p, \quad c_2 = \frac{p}{2(1-p)}\left(1 + \frac{\log p}{1-p}\right). \tag{5.37}$$

As $p \to 1$, the coefficient $c_1 \to 1$, and the minimum bit-energy-to-noise ratio approaches $BNR \to \log 2$. As for c_2, it approaches $-1/4$.

Proof. There is just one term in the summation over j in Eq. (5.16), $q_1 = p$ and $q'_1 = 0$ and $x_2 = (1-p)^{-1}$. Also $\mu_2(\mathcal{X}) = (1-p)^2$. Then

$$c_1 = -P(x_1)(x_2 - x_1)\log P(x_1) = -\frac{p}{1-p}\log p \tag{5.38}$$

$$c_2 = \tfrac{1}{2}\big((1-p)^{-1} - 1\big) + (x_2 - x_1)\Big(\tfrac{1}{2}(x_1 + x_2)q_1 - q'_1\Big)\log q_1 \tag{5.39}$$

$$= \frac{1}{2}\frac{p}{1-p} + \frac{1}{2}\frac{p}{(1-p)^2}\log p = \frac{1}{2}\frac{p}{1-p}\left(1 + \frac{\log p}{1-p}\right). \tag{5.40}$$

The limits as $p \to 1$ are readily computed by using the Taylor expansion $\log p = p - 1 - \frac{1}{2}(p-1)^2 + \mathrm{O}(p-1)^3$. $\square$

For instance, the choice $p \simeq 0.77$ already gives $\mathrm{BNR}_0 \simeq -1\,\mathrm{dB}$. For this choice the symbols are located at positions 0 and $(1-p)^{-1} \simeq 4.27$, a relatively compact constellation. In order to achieve a similar value of BNR_0, a value of $\lambda \simeq 6.85$ would be required, a much less effective method than flash signalling, since the constellation is more peaky and needs more points (recall that BNR_0 is attained as $m \to \infty$, and that for 2-PEM BNR_0 is $0\,\mathrm{dB}$).

5.3.4 Asymptotic Behaviour at High SNR

We estimate the asymptotic behaviour of $\mathrm{C}_{\mathcal{X}}$ at high SNR of the constellations described in Section 5.2 by following a method very close to that of the Gaussian channel in Section 4.2.6. We first assume that the input is described by the limiting continuous density in Eq. (5.7), and then approximate the output (differential) entropy $H(Y)$ by the differential entropy of the input $H(X)$. As we shall shortly see, the PEM constellations described in Section 5.2 allow for rather neat closed-form expressions.

As we saw in Proposition 5.1, the differential entropy of the input $\mathcal{X}_\lambda$ is

$$H(\mathcal{X}_\lambda) = 1 - \lambda + \log\bigl(\lambda(\lambda+1)\bigr). \tag{5.41}$$

Scaled by SNR, its entropy is $H(\mathcal{X}_\lambda) + \log \mathrm{SNR}$. As for the noise entropy, as we determined in Proposition 3.13 on page 31, is given by $H(Y|X) = H(Z) = \log(e)$. The asymptotic expansion of the capacity at high SNR is

$$\mathrm{C}_{\mathcal{X}_\lambda}(\mathrm{SNR}) \simeq \log \mathrm{SNR} - \lambda + \log\bigl(\lambda(\lambda+1)\bigr). \tag{5.42}$$

This formula begets the natural question of determining the value of λ which maximizes its value. Setting the first derivative to zero, we derive the optimum λ as the solution of

$$-1 + \frac{1}{\lambda} + \frac{1}{\lambda+1} = 0 \Longrightarrow -\lambda(\lambda+1) + \lambda + \lambda + 1 = 0, \tag{5.43}$$

that is $\lambda^2 - \lambda - 1 = 0$. The positive solution of this equation is $\lambda = \frac{1+\sqrt{5}}{2} \simeq 1.618\ldots$, the golden number. Hence, choosing λ as the golden number maximizes the capacity at large SNR in the constellation family described in Section 5.2.

The capacity of the AEN channel, of value $\log(1 + \mathrm{SNR})$, asymptotically behaves as $\log \mathrm{SNR}$ for large SNR, slightly larger than Eq. (5.42). For a fixed SNR, let SNR' the signal-to-noise ratio necessary to achieve the same capacity by using PEM. Then, we can compute the energy loss incurred by using a non-optimal constellation by setting

$$\log \mathrm{SNR} = \log \mathrm{SNR}' - \lambda + \log\big(\lambda(\lambda + 1)\big), \tag{5.44}$$

or equivalently

$$\frac{\mathrm{SNR}'}{\mathrm{SNR}} = \frac{e^{\lambda}}{\lambda(\lambda + 1)}. \tag{5.45}$$

This quantity is larger than one. Figure 5.4 depicts the loss, in decibels, incurred by the PEM constellations we consider. The lowest loss, achieved for $\lambda = \frac{1}{2}(1 + \sqrt{5})$ is approximately 0.76 dB, lower than the shaping loss incurred by uniform square QAM distribution in the Gaussian channel, namely 1.53 dB. Uniform PEM, with $\lambda = 1$, suffers from an energy loss $2/e$, or approximately 1.33 dB, the same value as a uniform distribution in a circle for the Gaussian channel (see Proposition 4.13 in Section 4.2.6). For the sake of comparison, the losses for the square and circle constellations in the Gaussian channel are shown in Fig. 5.4 as horizontal lines.

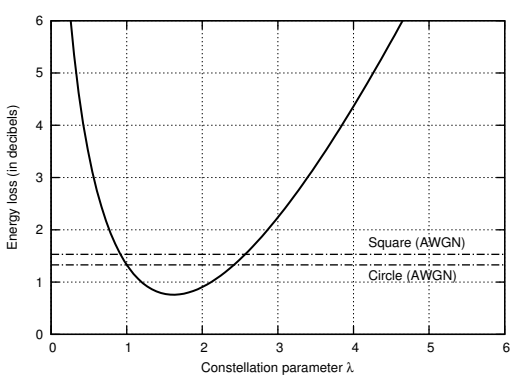


Figure 5.4: Asymptotic power loss in decibels for PEM constellations.

We conclude that PEM constellations in the AEN channel may be more energy-efficient at high SNR than QAM or PAM modulations in the AWGN. This reverts the situation at low SNR, where typical AWGN constellations require a lower SNR to achieve the same capacity, as we saw in Section 5.3.3.

5.3.5 Bit-Interleaved Coded Modulation

The principles for bit-interleaved coded modulation, presented in Section 4.3, apply essentially unchanged to the AEN channel. In particular, we can define the BICM capacity $\mathrm{C}_{\mathcal{X},\mu}$ for a given constellation set $\mathcal{X}$ and mapping rule μ. Using Proposition 4.16, we have the following convenient form,

Definition 5.6. In the AEN channel, the BICM capacity $\mathrm{C}_{\mathcal{X},\mu}$ for a modulation set $\mathcal{X}$ used under the binary mapping rule μ is given by

$$\mathrm{C}_{\mathcal{X},\mu} = \sum_{i=1}^{m} \frac{1}{2} \sum_{b=0,1} (\mathrm{C}^{\mathrm{u}}_{\mathcal{X}} - \mathrm{C}^{\mathrm{u}}_{\mathcal{X}_i^b}), \tag{5.46}$$

where $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}}$ and $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}_i^b}$ are, respectively, the constrained capacity of equiprobable signalling in the sets $\mathcal{X}$ and $\mathcal{X}_i^b$. The set $\mathcal{X}_i^b$ is the collection of all constellation symbols with bit b in the i-th position of the binary label.

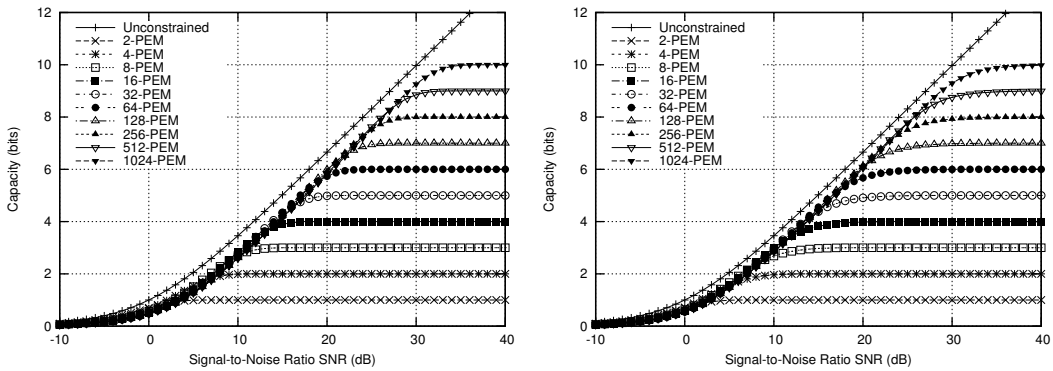
Efficient computation of $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}}$ and $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}_i^b}$ can be performed by using Eq. (5.16) with respective probabilities $1/|\mathcal{X}|$ and $2/|\mathcal{X}|$.

Figure 5.5 depicts the BICM capacity for several cases of equiprobable 2^m-PEM constellations under binary reflected Gray mapping. In Fig. 5.5a, the constellation points follow a uniform distribution, $\lambda = 1$, whereas $\lambda = \frac{1}{2}(1 + \sqrt{5})$, the optimum value for coded modulation, in Fig. 5.5b. In both cases, the curves are rather close to the respective constrained modulation capacities $\mathrm{C}_{\mathcal{X}}$, depicted in Figs. 5.2a and 5.3a.

Proposition 5.3 can be applied to derive the asymptotic expansion of the BICM capacity at low SNR. However, the resulting formula does not seem to be particularly informative and we have chosen not to write it down explicitly.

In contrast with the Gaussian case, inspection of Fig. 5.5 at high SNR suggests that the asymptotic form of the BICM capacity $\mathrm{C}_{\mathcal{X},\mu}$ with Gray mapping is appreciably smaller than the corresponding constrained capacity $\mathrm{C}_{\mathcal{X}}$. This impression is supported by Fig. 5.6, which shows the difference (in bits) between the channel capacity $\log(1 + \mathrm{SNR})$ and the capacities $\mathrm{C}_{\mathcal{X}}$ and $\mathrm{C}_{\mathcal{X},\mu}$ for several values of λ. For fixed λ and SNR, the largest capacity for all values of $m \leq 12$ was determined; the difference between this number and the channel capacity is depicted in the figure.

The divergence at large SNR is explained by the fact that the largest size of the constellation sets was 4096, limiting the capacity to 12 bits. If the channel capacity is appreciably larger than this number, a gap appears. At the other



(a) Uniform 2^m-PEM. (b) 2^m-PEM with $\lambda = \frac{1}{2}(1+\sqrt{5})$.

Figure 5.5: BICM capacity as a function of SNR for 2^m-PEM, Gray mapping.

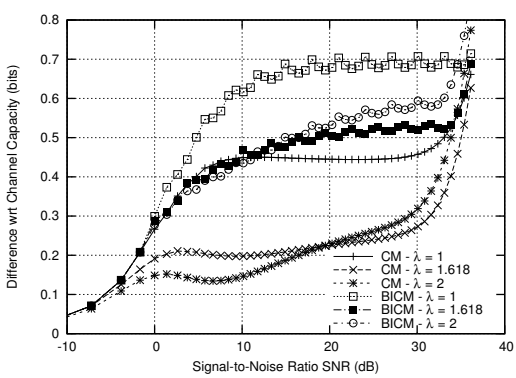


Figure 5.6: Difference with respect to the channel capacity for 2^m-PEM.

extreme, the difference among the various curves is small, since the absolute difference with the capacity is bounded by the capacity itself, a small number. Of more interest is the behaviour at moderate-to-large SNR. First, the capacity $\mathrm{C}_{\mathcal{X}}$ for $\lambda = 1$ is about 0.45 bits smaller than C, in line with Eq. (5.42) from Section 5.3.4. The values for $\lambda = 1.618$ and $\lambda = 2$ also approach their asymptotic form, Eq. (5.42), but do so more slowly. We have not found a convincing explanation for this fact. More intriguing is that the BICM ca-

pacities are around 0.25–0.30 bits lower than the CM capacities at high SNR. This trait, which was not appreciable in the Gaussian channel, should relate to the decoding metric for BICM, which ignores the correlation among the log-likelihood ratios. Moreover, the use of a mapping other than natural reflected Gray might yield some additional improvement.

5.4 Coded Modulation in the Discrete-Time Poisson Channel

5.4.1 Introduction

We now move on to the analysis of the discrete-time Poisson (DTP) channel. As we saw in Chapter 2, in this channel information is modulated onto the input symbol energy, as it was in the AEN channel. Energy at the channel output is quantized, and the input energy constraint is given in terms of the average number of quanta, denoted by ε_s.

The input at time k, a non-negative real-valued number x_k, induces a non-negative integer-valued output y_k distributed according to a Poisson distribution of parameter $\varepsilon_s x_k$, $Y_k \sim \mathcal{P}(\varepsilon_s x_k)$, there is no additive noise. We assume that x_k is drawn with probability $P(x_k)$ from a set $\mathcal{X}$; the set $\mathcal{X}$ is normalized to unit energy. The cardinality of $\mathcal{X}$ is $|\mathcal{X}| = 2^m$, that is m bits are required to index a symbol. The conditional output density $Q(y|x)$ is

$$Q(y|x) = e^{-\varepsilon_s x} \frac{(\varepsilon_s x)^y}{y!}. \tag{5.47}$$

The analysis in this section examines the transmission capabilities of the channel under pulse-energy modulations, with special attention being paid to the family described in Section 5.2. As such, our study complements existing work, such as [15], by deriving some new results:

- The performance at low quanta-count ε_s is described with techniques imported from the analysis of the Gaussian channel at low signal-to-noise ratio. In particular, the capacity per unit energy, the first two derivatives of the mutual information at zero quanta count, and the minimum number of quanta per bit for generic PEM formats are determined.

- The asymptotic form of the constrained coded modulation capacity at high quanta counts is determined by using a Gaussian approximation. A constellation with small loss with respect to the capacity is provided.

- Bit-interleaved coded modulation is shown to be a good method to design practical channel codes, since its capacity is close to that of coded modulation. In addition, the performance loss with uniform PEM modulation and Gray mapping at low quanta counts is given by a simple expression.

These items are respectively covered in Sections 5.4.3-5.4.4, 5.4.5, and 5.4.6.

5.4.2 Constrained Capacity

From the definition of mutual information in Eq. (3.3) on page 27, we define the constrained capacity $\mathrm{C}_{\mathcal{X}}$ for a fixed modulation set $\mathcal{X}$ as

Definition 5.7. In the DTP channel with average quanta count ε_s, the capacity $\mathrm{C}_{\mathcal{X}}$ for signalling over a modulation set $\mathcal{X}$ with probabilities $P(x)$ is

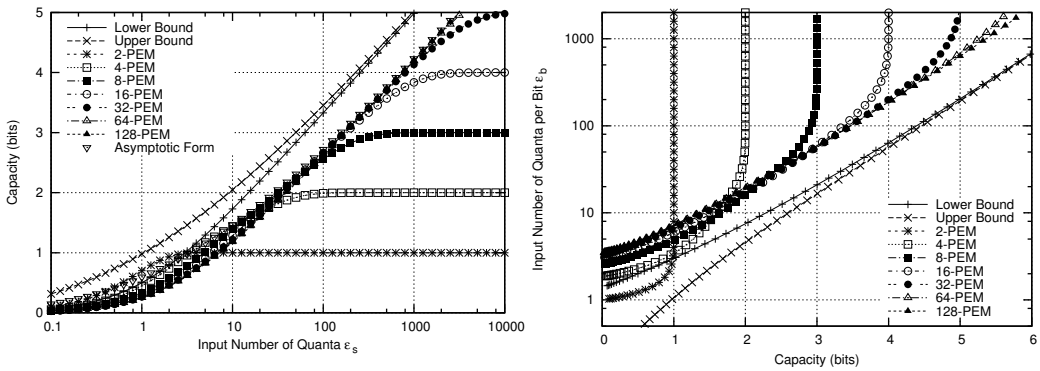
$$\mathrm{C}_{\mathcal{X}}(\varepsilon_s) = -\sum_{x} P(x) \sum_{y=0}^{\infty} e^{-\varepsilon_s x} \frac{(\varepsilon_s x)^y}{y!} \log\left(\sum_{x' \in \mathcal{X}} P(x') e^{\varepsilon_s (x-x')} \left(\frac{x'}{x}\right)^y \right). \tag{5.48}$$

If necessary, we also use the convention $x \log x = 0$ for $x = 0$. Further, for $x = 0$, the only possible output is $y = 0$, and Eq. (5.48) is well-defined by the natural limiting procedure replacing the summation over y by

$$\log\left(\sum_{x' \in \mathcal{X}} P(x') e^{-\varepsilon_s x'} \right). \tag{5.49}$$

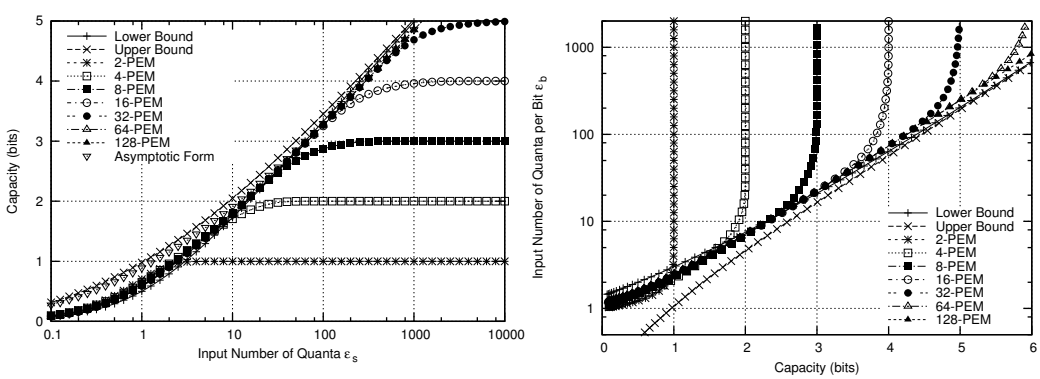
Figures 5.7 and 5.8 show the capacity for equiprobable 2^m-PEM, respectively with $\lambda = 1$ and $\lambda = 1 + \sqrt{3}$. In both cases, the upper plot depicts the capacity as a function of ε_s, whereas the lower one shows the number of quanta per bit ε_b, defined in analogy to the AWGN and AEN channels as $\varepsilon_b = \frac{\varepsilon_s}{\mathrm{C}_{\mathcal{X}}} \log 2$, as a function of the capacity $\mathrm{C}_{\mathcal{X}}$. Since for this channel the capacity is not known, we also depict the upper and lower bounds from Chapter 3.

We will see in Section 5.4.5 how the choice $\lambda = 1 + \sqrt{3}$ maximizes the asymptotic expression of the capacity $\mathrm{C}_{\mathcal{X}}$ at high ε_s under a Gaussian approximation. These asymptotic expressions match well with the envelope of the capacity curves as the number of constellation points increases. To any extent, the constrained capacity for this value of λ is markedly larger than for uniform PEM, $\lambda = 1$; inspection of the plots shows an energy gain factor above 3 for a fixed capacity or about 0.65 bits for fixed ε_s. Further, constellations with $\lambda = 1 + \sqrt{3}$ are better than uniform PEM also for low ε_s.



(a) Capacity as a function of ε_s. (b) ε_b as a function of the capacity.

Figure 5.7: CM capacity for uniform 2^m-PEM.



(a) Capacity as a function of ε_s. (b) ε_b as a function of the capacity.

Figure 5.8: CM capacity for equiprobable 2^m-PEM with $\lambda = 1 + \sqrt{3}$.

As it happened in the AEN case, the modulations do not converge to a single line for low ε_s. This behaviour is analytically characterized in Section 5.4.4, where we shall also see that the coded modulation capacity for flash signalling closely approaches the upper bound to the true channel capacity. First, in Section 5.4.3, we determine the capacity per unit energy of the DTP channel.

5.4.3 Capacity per Unit Energy

The capacity per unit energy C_1 is defined as the largest number of bits per symbol which can be reliably sent over the channel per unit energy. Using Theorem 3.9 on page 30, C_1 is determined as

$$\mathrm{C}_1 = \sup_x \frac{D\big(Q(y|x)||Q(y|x=0)\big)}{\varepsilon_s x} = \infty, \tag{5.50}$$

since the support of the conditional output $Q(y|x=0)$ is limited to $y=0$, and therefore the divergence is infinite. This value was already computed by Verdú [29] and before him by Pierce [30] in the context of optical communications.

The minimum bit-energy-to-noise ratio $\varepsilon_{b,\min}$ is $\varepsilon_{b,\min} = \frac{\log 2}{\mathrm{C}_1} = 0$. In other words, to transmit an asymptotically vanishing amount of information, the required number of quanta approaches zero faster than linearly in the amount of information.

As a final remark, we note that for the AEN (resp. AWGN) channels, the bit energy $E_{b,\min} = E_n \log 2$ (resp. $E_{b,\min} = \sigma^2 \log 2$) also vanishes as the average noise E_n (resp. σ^2) goes to zero. Since there is no additive noise in the DTP channel, the value of $\varepsilon_{b,\min}$ in the DTP channel is in agreement with this behaviour in the continuous channels. In Section 5.5.2, we shall see that $\varepsilon_{b,\min}$ in the quantized additive energy channel is $\varepsilon_n \log 2$, where ε_n is the expected additive noise, making complete the analogy with the continuous channels.

5.4.4 Asymptotic Behaviour at Low ε_s

At low ε_s the behaviour of the coded modulation capacity is characterized in

Proposition 5.8. *In the DTP channel, the constrained capacity $C_{\mathcal{X}}(\varepsilon_s)$ using a signal set $\mathcal{X}$ with probabilities $P(x)$ and average energy constraint ε_s, admits a Taylor series expansion of the form $C_{\mathcal{X}}(\varepsilon_s) = c_1\varepsilon_s + c_2\varepsilon_s^2 + O(\varepsilon_s^3)$, as $\varepsilon_s \to 0$, with c_1 and c_2 given by*

$$c_1 = \sum_{x\in\mathcal{X}} xP(x) \log \frac{x}{\mu_1(\mathcal{X})} \tag{5.51}$$

$$c_2 = \frac{1}{2}\sigma^2(\mathcal{X}) - \frac{1}{2}\mu_2(\mathcal{X}) \log \frac{\mu_2(\mathcal{X})}{\mu_1^2(\mathcal{X})}. \tag{5.52}$$

Proof. The proof can be found in Appendix 5.B. □

For constellations with unit mean energy, the bit energy at zero capacity ε_b, ε_{b0}, is given by Theorem 4.3,

$$\varepsilon_{b0} = \log 2 \left(\sum_{x \in \mathcal{X}} x P(x) \log x \right)^{-1}. \tag{5.53}$$

Since the function $t \log t$ is convex, this quantity is in general larger than zero, the value of $\varepsilon_{b,\min}$ derived from the capacity per unit energy.

For the PEM constellations described in Section 5.2, it is possible to compute the asymptotic limit as the number of points becomes infinite,

Corollary 5.9. *For 2-PEM, $c_1 = \log 2$ and $BNR_0 = 1$, or 0 dB.*

In the DTP channel, as $m \to \infty$, 2^m-PEM with parameter λ has

$$\lim_{m \to \infty} c_1 = \log(\lambda + 1) - \frac{\lambda}{\lambda + 1} \tag{5.54}$$

$$\lim_{m \to \infty} c_2 = \frac{1}{2} \left(\frac{\lambda^2}{2\lambda + 1} - \frac{(\lambda + 1)^2}{2\lambda + 1} \log \frac{(\lambda + 1)^2}{2\lambda + 1} \right), \tag{5.55}$$

and therefore

$$\lim_{m \to \infty} \varepsilon_{b0} = \frac{\log 2}{\log(\lambda + 1) - \frac{\lambda}{\lambda + 1}}. \tag{5.56}$$

As $\lambda \to \infty$ and $m \to \infty$, BNR_0 approaches 0, the value derived from the capacity per unit energy; under the same conditions c_2 approaches $-\infty$.

For instance, uniform 2^m-PEM modulation in the discrete-time Poisson channel requires about $\varepsilon_{b,\min} = \frac{\log 2}{\log 2 - \frac{1}{2}} \simeq 3.59$ quanta/bit, as the number of quanta vanishes, $\varepsilon_s \to 0$.

Proof. For the PEM constellations under consideration, average energy is one and $\mu_1(\mathcal{X}) = 1$. For large m, the sum over the modulation symbols is replaced by a Riemann integral, and we have

$$\lim_{m \to \infty} c_1 = \int_0^{\lambda+1} \frac{x^{\frac{1}{\lambda} - 1}}{\lambda (1 + \lambda)^{\frac{1}{\lambda}}} x \log x \, dx = \log(\lambda + 1) - \frac{\lambda}{\lambda + 1}. \tag{5.57}$$

$\varepsilon_{b,\min}$ follows immediately.

The limit for the second-order coefficient c_2 is obtained by using the formula for $\mu_2(\mathcal{X})$ in Proposition 5.1. □

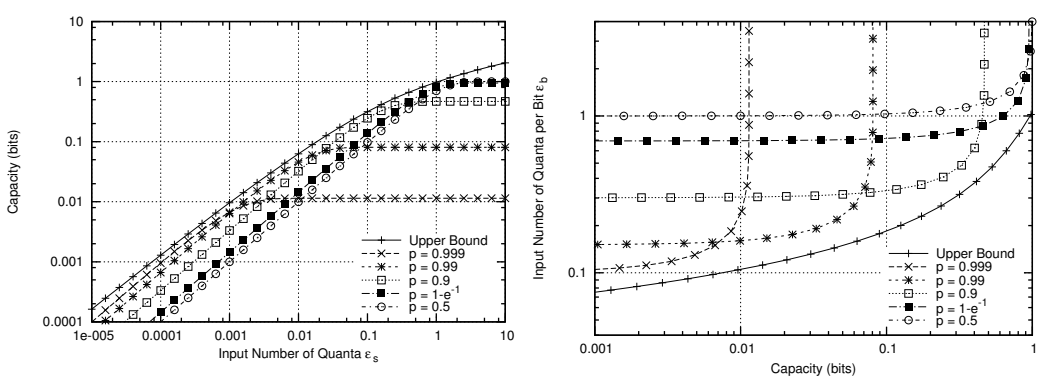
We next consider a generalized form of on-off keying, flash signalling [25]. With our definitions, the classical on-off keying is defined by two points at positions 0 and 2, randomly selected each time instant with probability 1/2. In flash signalling, the two possible inputs are $x_1 = 0$, with probability p, and $x_1 = 1/(1-p)$, with probability $1-p$, and the value of p is taken close to 1. Then, it is straightforward to obtain

Corollary 5.10. *Flash signalling in the DTP channel has coefficients*

$$c_1 = -\log(1-p), \quad c_2 = \frac{p + \log(1-p)}{2(1-p)}. \tag{5.58}$$

In the limit $p \to 1$, $c_1 \to \infty$ and the number of quanta per bit approaches $\varepsilon_{b0} = 0$. In addition, c_2 approaches $-\infty$. This behaviour is consistent with a functional form $\mathrm{C}(\varepsilon_s) \simeq \varepsilon_s \log \varepsilon_s^{-1}$ as $\varepsilon_s \to 0$, as opposed to linear in ε_s, i. e. $\mathrm{C}(\varepsilon_s) \propto \varepsilon_s$, as was the case in the AWGN and AEN channels.

Figure 5.9 depicts the capacity of flash signalling for several values of p. In Fig. 5.9a, capacity is presented as function of ε_s. The difference between the highest capacity $\mathrm{C}_{\mathcal{X}}$ and the upper bound to the true channel capacity is quite small. Figure 5.9b shows ε_b as a function of the capacity. Even though ε_{b0} indeed approaches zero, it does so very slowly. Curiously, the choice $p = 1-e^{-1}$ gives $c_1 = 1$, as in the Gaussian channel. This input appears in the information-theoretic analysis of the Z-channel at high noise levels, see [62].



(a) Capacity as a function of ε_s.

(b) ε_b as a function of the capacity.

Figure 5.9: CM capacity for flash signalling.

5.4.5 Asymptotic Behaviour at High ε_s

We now move on to the estimate of the asymptotic behaviour of the constrained capacity at high ε_s. The constrained capacity may be written as the difference between two terms, the output entropy $H(Y)$ and the conditional entropy $H(Y|X)$. We approximate the output entropy the same way we did in the AWGN and AEN channels, namely by approximating $H(Y)$ by the differential entropy of the input $H(X)$, when the input is a continuous density. As we saw in Proposition 5.1, the differential entropy of the scaled input $\varepsilon_s \mathcal{X}_\lambda$ is

$$H(\mathcal{X}_\lambda) = 1 - \lambda + \log\big(\lambda(\lambda+1)\big) + \log \varepsilon_s. \tag{5.59}$$

As for the conditional entropy, differently from the AWGN/AEN channels, noise is not additive in the DTP channel, a fact which must be properly taken into account. We use a Gaussian approximation to the conditional entropy, which is also valid as an asymptotic formula for the entropy of a Poisson distribution, $H(Y|X=x) \simeq \frac{1}{2}\log(2\pi e x)$ (see Proposition 3.40 in Section 3.5.2). Integrated over the input x, this gives

$$H(Y|X) \simeq \int_0^{\varepsilon_s(\lambda+1)} \frac{x^{\frac{1}{\lambda}-1}}{\varepsilon_s^{\frac{1}{\lambda}}\lambda(1+\lambda)^{\frac{1}{\lambda}}} \frac{1}{2}\log(2\pi e x)\, dx \tag{5.60}$$

$$= \frac{1}{2}\Big(\log\big(2\pi e \varepsilon_s(\lambda+1)\big) - \lambda\Big). \tag{5.61}$$

Therefore, assuming the Gaussian approximation is valid, the asymptotic behaviour of C at high ε_s is given by

$$\mathrm{C}(\varepsilon_s) \simeq 1 - \lambda + \log\big(\lambda(\lambda+1)\big) + \log \varepsilon_s - \frac{1}{2}\Big(\log\big(2\pi e \varepsilon_s(\lambda+1)\big) - \lambda\Big) \tag{5.62}$$

$$= \frac{1}{2}\log \varepsilon_s + \frac{1}{2}\big(1-\lambda\big) + \frac{1}{2}\log\big(\lambda^2(\lambda+1)\big) - \frac{1}{2}\log(2\pi). \tag{5.63}$$

As we did in the AEN case, it is interesting to determine the value of λ with the highest asymptotic capacity. It is derived by computing the first derivative with respect to λ and setting it to zero,

$$-\frac{1}{2} + \frac{1}{\lambda} + \frac{1}{2(\lambda+1)} = 0 \Longrightarrow -\lambda(\lambda+1) + 2(\lambda+1) + \lambda = 0, \tag{5.64}$$

that is $\lambda^2 - 2\lambda - 2 = 0$. Its positive root is $\lambda = 1 + \sqrt{3} \simeq 2.732\ldots$.

Since the capacity of the DTP channel behaves as $\frac{1}{2}\log\varepsilon_s$ for large ε_s (see Section 3.5.4), using PEM requires a larger energy ε_s' to attain the same data rate. We compute the energy loss incurred by PEM modulation by setting

$$\frac{1}{2}\log\varepsilon_s = \frac{1}{2}\log\varepsilon_s' + \frac{1}{2}\big(1-\lambda\big) + \frac{1}{2}\log\big(\lambda^2(\lambda+1)\big) - \frac{1}{2}\log(2\pi). \tag{5.65}$$

Solving for the ratio between the energies,

$$\frac{\varepsilon_s'}{\varepsilon_s} = \frac{2\pi e^{\lambda-1}}{\lambda^2(\lambda+1)}. \tag{5.66}$$

This quantity is larger than 1 for $\lambda > 0$ and represents a relative loss in energy. Figure 5.10 depicts the loss, in decibels, incurred by the PEM constellations. The lowest loss, achieved for $\lambda = 1+\sqrt{3}$, is slightly above 1 dB, better than the shaping loss of circular or square QAM constellations in the Gaussian channel or a uniform density in the AEN channel, also drawn in the figure as horizontal lines for the sake of comparison. Uniform PEM, with $\lambda = 1$, suffers from a power loss of π, or about 4.97 dB. The improvement brought about by a good choice of λ is significant. As we shall review in Chapter 6, the value $\lambda = 2$ minimizes the pairwise error probability in the DTP channel.

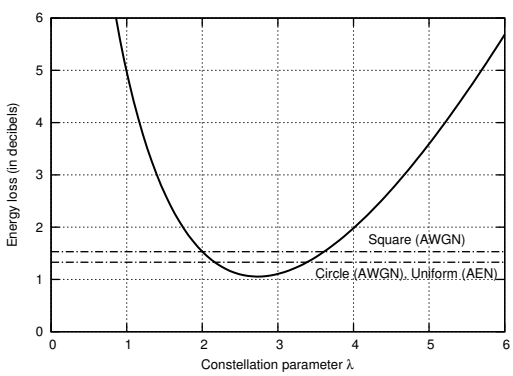


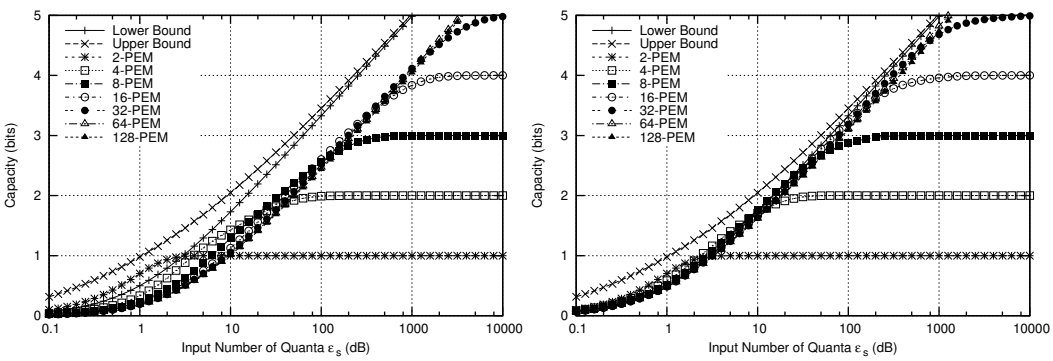
Figure 5.10: Asymptotic power loss in decibels for PEM constellations.

5.4.6 Bit-Interleaved Coded Modulation

It is probably not very surprising that a bit-interleaved coded modulation channel can be defined for the DTP channel, as was done for the Gaussian and AEN

channels. In particular, we have a BICM capacity $\mathrm{C}_{\mathcal{X},\mu}$, given by the expression (4.39) or alternatively by a definition similar to that of Eq. (5.46).

Figure 5.11 depicts the BICM capacity for several equiprobable 2^m-PEM modulation and Gray mapping. The upper and lower bounds to the channel capacity from Chapter 2 are also depicted. The most remarkable fact is that the curves are very close to those of the coded modulation capacity.



(a) Uniform 2^m-PEM.

(b) 2^m-PEM with $\lambda = 1 + \sqrt{3}$.

Figure 5.11: BICM capacity as a function of ε_s for 2^m-PEM and Gray mapping.

The loss incurred by BICM with respect to the upper bound to the channel capacity is depicted in Fig. 5.12. More precisely, for each value of ε_s and λ, the highest BICM capacity for all $m \leq 7$ is determined, and subtracted from the upper bound to the channel capacity; the result is plotted as a function of ε_s. The same steps are repeated for the constrained capacity.

The BICM capacity is very close to that of coded modulation, as was the case in the Gaussian channel. Qualitatively, the behaviour of the various curves is very similar to that of the AEN channel. For low ε_s the difference among the various curves is small, since the absolute difference with the capacity is bounded by the capacity itself, a small number; all curves coincide because the highest capacity is achieved by 2-PEM, and in that case $\mathrm{C}_{\mathcal{X}}$ and $\mathrm{C}_{\mathcal{X},\mu}$ coincide. For larger values of ε_s, the capacities $\mathrm{C}_{\mathcal{X}}$ and $\mathrm{C}_{\mathcal{X},\mu}$ approach an asymptotic form, whose value for $\mathrm{C}_{\mathcal{X}}$ is very closely given by the expressions in Section 5.4.5.

For the DTP channel, a neat formula exists for the linear terms of the Taylor expansion of $\mathrm{C}_{\mathcal{X},\mu}$ around ε_s. Then, we have

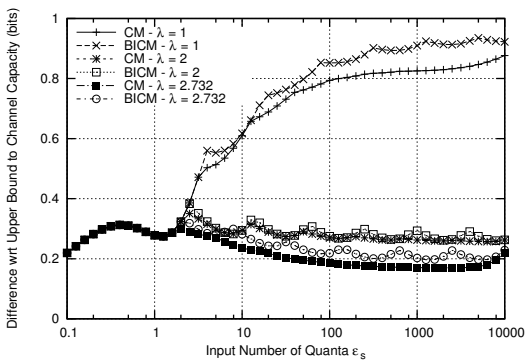


Figure 5.12: Difference with respect to the channel capacity for 2^m-PEM.

Proposition 5.11. *The BICM capacity in the DTP channel for a constellation $\mathcal{X}$ with average unit energy admits an expansion in Taylor series around $\varepsilon_s = 0$ whose first term c_1 is given by*

$$c_1 = \sum_{i=1}^{m} \frac{1}{2} \sum_{b} \mu_1(\mathcal{X}_i^b) \log \mu_1(\mathcal{X}_i^b) \tag{5.67}$$

where $\mu_1(\mathcal{X}_i^b)$ is the average symbol for a fixed label index i, and bit value b.

Proof. From Proposition 4.16, the BICM capacity can be written as

$$\mathrm{C}_{\mathcal{X},\mu} = \sum_{i=1}^{m} \frac{1}{2} \sum_{b=0,1} (\mathrm{C}^{\mathrm{u}}_{\mathcal{X}} - \mathrm{C}^{\mathrm{u}}_{\mathcal{X}_i^b}). \tag{5.68}$$

The summands $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}}$ and $\mathrm{C}^{\mathrm{u}}_{\mathcal{X}_i^b}$ admit a Taylor expansion given in Proposition 5.8,

$$c_1 = \sum_{i=1}^{m} \frac{1}{2} \sum_{b=0,1} \left(\sum_{x \in \mathcal{X}} x \frac{1}{|\mathcal{X}|} \log \frac{x}{\mu_1(\mathcal{X})} - \sum_{x \in \mathcal{X}_i^b} x \frac{2}{|\mathcal{X}|} \log \frac{x}{\mu_1(\mathcal{X}_i^b)} \right) \tag{5.69}$$

$$= \sum_{i=1}^{m} \frac{1}{2} \sum_{b=0,1} \mu_1(\mathcal{X}_i^b) \log \mu_1(\mathcal{X}_i^b), \tag{5.70}$$

since $\mu_1(\mathcal{X}) = 1$, and the summations over $\mathcal{X}_i^b$ for $b = 0$ and $b = 1$ can be either combined to yield a summation over $\mathcal{X}$, hence cancelling the contribution from the first summand, or carried out to give $\mu_1(\mathcal{X}_i^b)$. □

As in the Gaussian case, we can compute the result in Theorem 5.11 for natural Gray mapping (see the definition on page 87). Then

Corollary 5.12. *For uniform 2^m-PEM and Gray mapping, c_1 is given by*

$$c_1 = \frac{1}{2}\frac{2^{m-1}-1}{2^m-1}\log\frac{2^{m-1}-1}{2^m-1} + \frac{1}{2}\frac{3\,2^{m-1}-1}{2^m-1}\log\frac{3\,2^{m-1}-1}{2^m-1}. \tag{5.71}$$

As $m \to \infty$, the coefficient c_1 approaches $c_1 \to \log\frac{3^{3/4}}{2} \simeq 0.13$, whereas the minimum bit energy approaches $\varepsilon_{b0} \to 7.24$ dB.

The loss represents about 1.69 dB with respect to the CM limit as $m \to \infty$, which was found in Corollary 5.9 to be $\varepsilon_{b0} = 5.55$ dB.

Proof. For 2^m-PEM, the Gray mapping construction above makes $\mu_1(\mathcal{X}_i^b) = 1$ for $b = 0, 1$ and all bits but one, say $i = 1$. Therefore,

$$c_1 = \frac{1}{2}\mu_1(\mathcal{X}_1^0)\log\mu_1(\mathcal{X}_1^0) + \frac{1}{2}\mu_1(\mathcal{X}_1^1)\log\mu_1(\mathcal{X}_1^1). \tag{5.72}$$

Symbols x_i, $i = 0, \ldots, 2^m - 1$ take the value $x_i = \frac{2i}{2^m-1}$, and therefore

$$\mu_1(\mathcal{X}_1^0) = \frac{2}{2^m-1}\frac{1}{2^{m-1}}\sum_{i=0}^{2^{m-1}-1} i = \frac{2^{m-1}-1}{2^m-1}, \tag{5.73}$$

$$\mu_1(\mathcal{X}_1^1) = \frac{2}{2^m-1}\frac{1}{2^{m-1}}\sum_{i=2^{m-1}}^{2^m-1} i = \frac{3\,2^{m-1}-1}{2^m-1}. \tag{5.74}$$

The coefficient c_1 then follows. The limit $m \to \infty$ is obvious. □

5.5 Coded Modulation in the Quantized Additive Energy Channel

5.5.1 Constrained Capacity

The quantized additive energy (AE-Q) channel was introduced in Chapter 2 as an intermediate between the additive exponential (AEN) and discrete-time Poisson (DTP) channels. From the former, it inherits the trait that the channel output at time k y_k is the sum of a signal and an additive noise components, respectively denoted by s_k and z_k, i. e. $y_k = s_k + z_k$. It shares with the DTP channel the properties that the output y_k is a non-negative integer and that

the signal component s_k follows a Poisson distribution. The output suffers from Poisson noise (from s_k) and additive noise (from z_k). The noise z_k has a geometric distribution of mean ε_n.

In this section, we assume that the channel input at time k is a non-negative real number, of value $\varepsilon_s x_k$, where ε_s is the average signal energy count, and x_k are drawn from a set $\mathcal{X}$ with probability $P(x_k)$. The elements of $\mathcal{X}$ are the modulation symbols, and are normalized to have unit energy.

The DTP channel is recovered from the AE-Q channel by setting $\varepsilon_n = 0$. At the other extreme, as $\varepsilon_n \to \infty$, we saw in Chapter 3 that the AE-Q channel becomes ever closer to an equivalent AEN channel with signal-to-noise ratio SNR if also ε_s goes to infinity as $\varepsilon_s = \mathrm{SNR}\,\varepsilon_n$. In this section, we study the way the AE-Q channel is an intermediate between the AEN and DTP channels when coded modulation is considered. Our main contributions are the computations of the capacity per unit energy, in Section 5.5.2, and of the Taylor expansion of the constrained capacity for low values of ε_s, in Section 5.5.3.

Recall that the constrained capacity $\mathrm{C}_{\mathcal{X}}(\varepsilon_s, \varepsilon_n)$ is given by

$$\mathrm{C}_{\mathcal{X}}(\varepsilon_s, \varepsilon_n) = -\sum_x P(x) \sum_{y=0}^{\infty} Q(y|x) \log\left(\sum_{x' \in \mathcal{X}} P(x') \frac{Q(y|x')}{Q(y|x)} \right), \tag{5.75}$$

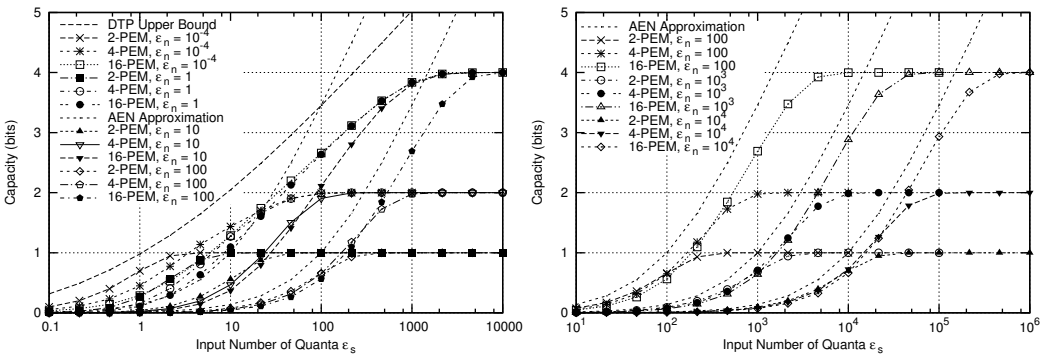
with $Q(y|x)$ depends on ε_s and ε_n, and is given by Eq. (2.23),

$$Q(y|x) = \frac{e^{\frac{\varepsilon_s x}{\varepsilon_n}}}{1+\varepsilon_n} \left(\frac{\varepsilon_n}{1+\varepsilon_n} \right)^y \sum_{l=0}^{y} e^{-\varepsilon_s x \left(1+\frac{1}{\varepsilon_n}\right)} \frac{\left(\varepsilon_s x \left(1+\frac{1}{\varepsilon_n}\right)\right)^l}{l!}. \tag{5.76}$$

Figure 5.13 shows the capacity $\mathrm{C}_{\mathcal{X}}(\varepsilon_s, \varepsilon_n)$ for uniform 2^m-PEM modulation and for several values of ε_n. For the sake of comparison, the channel capacities for the AEN (with signal-to-noise ratio $\varepsilon_s/\varepsilon_n$) and DTP channels (with signal energy ε_s) are also depicted.

The first and expected conclusion drawn from the plots is that the required signal energy to achieve a given rate migrates to higher values as the noise level ε_n increases. In addition, most of the heuristic results for the channel capacity derived in Chapter 3 extend to coded modulation. For low values of the additive noise ε_n, say from $\varepsilon_n = 0$ up to $\varepsilon_n = 1$, the constrained capacity is closely given by that of a discrete-time Poisson channel with $\varepsilon_n = 0$.

For values of additive noise $\varepsilon_n \gg 1$, a good approximation to the capacity is the capacity of the AEN channel with signal-to-noise ratio given by $\mathrm{SNR} = \varepsilon_s/\varepsilon_n$. This fact is especially visible in Fig. 5.13b. Nevertheless, for each value



(a) Low ε_s.

(b) Moderate-to-high ε_s.

Figure 5.13: CM capacity as a function of ε_s for uniform 2^m-PEM.

of ε_n, there exists a threshold ε_s^* above which the constrained capacity diverges from the AEN value and approaches that of a DTP channel with count ε_s. As we saw in Eq. (3.119), on page 52, this threshold is approximately given by $\varepsilon_s^* = \varepsilon_n^2$, or equivalently by a threshold signal-to-noise ratio $\mathrm{SNR}^* \simeq \varepsilon_n$.

In Section 5.3.4 we studied the asymptotic expression of $\mathrm{C}_{\mathcal{X}}$ for uniform 2^m-PEM in the AEN channel, and found it was $\log \mathrm{SNR} - \log \frac{e}{2}$. In the analysis of the DTP channel with the Gaussian approximation, in Section 5.4.5, the asymptotic expansion in Eq. (5.63) is given by $\frac{1}{2}\log \varepsilon_s - \frac{1}{2}\log \pi$. The plots in Fig. 5.13 suggest that the shaping gain in the AE-Q channel experiences a smooth transition when moving from the DTP-like to the AEN-like regimes.

Identical qualitative results also hold for other non-uniform 2^m-PEM constellations. An interesting and rather surprising phenomenon concerns the behaviour for low quanta counts, which we discuss next.

5.5.2 Capacity per Unit Energy

We start by determining the capacity per unit energy C_1, the largest number of bits per symbol which can be reliably sent over the channel per unit energy.

Theorem 5.13. *In the AE-Q channel with additive geometric noise of mean ε_n, the capacity per unit energy C_1 and the minimum energy per bit $\varepsilon_{b,min}$ are*

respectively given by

$$C_1 = \frac{1}{\varepsilon_n}, \quad \varepsilon_{b,min} = \varepsilon_n \log 2. \tag{5.77}$$

The energy per bit has its natural meaning. Its form is reminiscent of the results for the Gaussian channel, $E_{b,\min} = \sigma^2 \log 2$, and for the exponential channel $E_{b,\min} = E_n \log 2$.

Proof. The proof can be found in Appendix 5.C. □

This is the natural counterpart of the result in the AEN channel, where the minimum energy per bit was found to be E_n. The presence of Poisson noise does not reduce the capacity per unit energy, or equivalently increase the minimum energy per bit.

5.5.3 Asymptotic Behaviour at Low ε_s

For a capacity $\mathrm{C}_{\mathcal{X}}$, it is natural to define a number of quanta per bit, ε_b, as $\varepsilon_b = \frac{\varepsilon_s}{\mathrm{C}_{\mathcal{X}}} \log 2$. As we saw in the DTP channel, this quantity plays an analogous role to that of BNR in Gaussian channels. The capacity curves in Fig. 5.13 may be turned around and represented as ε_b versus the channel capacity. The resulting plot is depicted in Fig. 5.14. One immediately notices an unexpected behaviour in the region $\mathrm{C}_{\mathcal{X}} \to 0$, namely the seeming divergence to infinity of number of quanta per bit ε_b. This impression is further corroborated by Fig. 5.14b, essentially a logarithmic unfolding of the previous plot, in which the x scale is plotted in logarithmic scale, so as to better examine the shape of the curve.

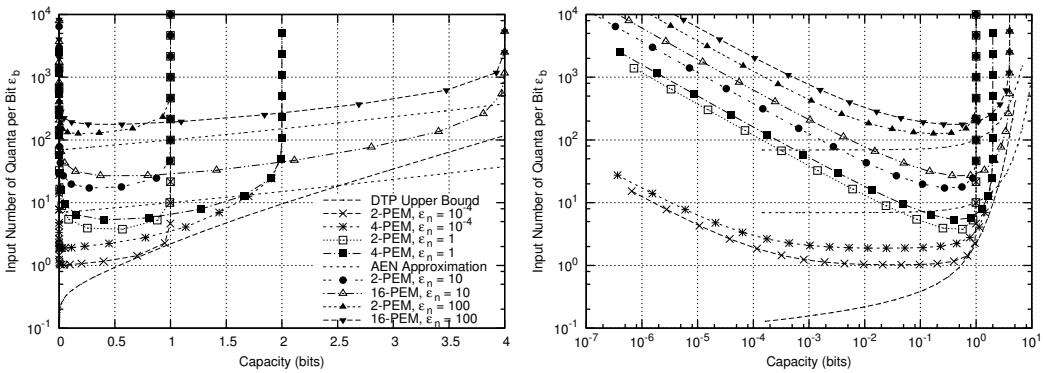
This odd behaviour is no artifact of the computations, as stated in

Proposition 5.14. *In the AE-Q channel with geometric noise of mean $\varepsilon_n > 0$ and average signal count ε_s, the Taylor expansion of the constrained capacity $C_{\mathcal{X}}(\varepsilon_s, \varepsilon_n)$ around $\varepsilon_s = 0$ is given by*

$$C_{\mathcal{X}}(\varepsilon_s, \varepsilon_n) = \frac{1}{2}\sigma^2(\mathcal{X})\frac{\varepsilon_s^2}{\varepsilon_n} + O(\varepsilon_s^3). \tag{5.78}$$

Since the linear term is zero, the energy per bit at vanishing ε_s is $\varepsilon_{b0} = \infty$.

Proof. The proof can be found in Appendix 5.D. □



(a) Linear scale. (b) Logarithmic scale; legend as left.

Figure 5.14: Bit energy versus CM capacity $\mathrm{C}_{\mathcal{X}}$ for uniform 2^m-PEM.

Since the result holds for arbitrary PEM modulations, the use of flash signalling (binary modulation, with one symbol at zero used with probability p) does not achieve a finite ε_{b0}. It would however, achieve a large c_2 coefficient, since $c_2 = \frac{p}{2(1-p)} \to \infty$ for flash signalling with probability $p \to 1$, as we saw in the proof of Corollary 5.5.

This is a most surprising fact, for which we have not found an intuitive explanation. In Sections 5.3.3 and 5.4.4, on the respective analysis of the limiting cases AEN and DTP, we saw that the required energy per bit at zero SNR had a finite, well-defined value. It would have been natural that the intermediate case, the AE-Q channel, changes these well-defined values in a smooth transition between the two extremes. The transition seems to be however different, as implied by the curves in Fig. 5.14b and Proposition 5.14. For the DTP channel, ε_{b0} is finite, of value

$$\varepsilon_{b0} = \log 2 \left(\sum_{x \in \mathcal{X}} x P(x) \log x \right)^{-1}, \tag{5.79}$$

given in Theorem 5.8. From this point, the addition of some additive noise ε_n, however small, makes the curve ${\varepsilon_b}^{\text{AE-Q}}(\mathrm{C}_{\mathcal{X}})$ diverge from the DTP value. The dual behaviour takes place in the transition to the AEN limit.

This seeming contradiction is solved by noting that this happens at a very small capacity, below 10^{-3} bits for $\varepsilon_n = 10^{-4}$, as seen in Fig. 5.14b. For even

lower values of ε_n, the divergence takes place at even lower capacities. A similar transition takes place when approaching the AEN limit. Fortunately, the divergence is not visible in the limit, unless one wishes to work at tiny capacities. Further, if a system operated at such low rates, it could be argued whether ε_b is an appropriate figure of merit, as opposed to ε_s. An interesting question is what impact this behaviour may have for situations when the channel is not close to either limit, such as ε_n in the order of hundreds.

We note that we have not been able to analytically determine the position of the inflection point in the curves for $\varepsilon_b(\mathrm{C}_{\mathcal{X}})$, which would correspond to the minimum energy per bit for a given modulation set. To any extent, an absolute lower bound to the energy per bit is given by $\varepsilon_{b,\min} = \varepsilon_n \log 2$.

5.5.4 Bit-Interleaved Coded Modulation

We conclude our analysis of the AE-Q channel by briefly considering bit-interleaved coded modulation. A BICM capacity can be defined as in the previously studied channels, and its value computed. When using Gray mapping, the results turn out to be very close to the CM case, and identical conclusions to the ones reached previously can be derived for BICM. In particular, there is a transition between the AEN and the DTP channels, with the AE-Q channel model bridging the gap between the two of them. We conclude that bit-interleaved coded modulation is likely to be a good method to general channel codes for the AE-Q channel.

In addition, thanks to the decomposition of the BICM capacity in Proposition 4.16, it is clear that the first term c_1 of the Taylor series of the capacity around $\varepsilon_s = 0$ is zero, in line with the result for coded modulation.

5.6 Conclusions

The additive energy channel models were constructed in Chapter 2 in such a way that the AE-Q channel has the AEN and DTP channels as limiting forms. This idea was further supported in Chapter 3, where we saw that the capacity of the AE-Q channel indeed converges to the capacity of these limiting cases in a natural manner. In this chapter, we have examined this transition when coded modulation is considered and we have in the process extended the analysis of coded modulation, carried out for the Gaussian channel in Chapter 4, to the additive energy channels.

More specifically, we have determined the capacity achievable by a family of pulse-energy modulations (PEM) whose constellation points have the form

$\beta(i-1)^\lambda$, for $i = 1, \ldots, 2^m$, where $\lambda > 0$ is a parameter and β a normalization factor; points are used with the same probabilities. As m becomes very large, the constellation can be described by a continuous density with support limited to the interval $[0, \lambda + 1]$. In addition, we have considered flash signalling, a generalized form of binary modulation with two points at positions 0, with probability p, and $1/(1-p)$, with probability $1-p$.

For the various additive energy channels, we have computed the constrained capacity $\mathrm{C}_{\mathcal{X}}$ achievable by using a set $\mathcal{X}$ and the bit-interleaved coded modulation (BICM) capacity $\mathrm{C}_{\mathcal{X},\mu}$ for the set $\mathcal{X}$ under mapping rule μ. As for the Gaussian channel, both capacities are rather similar when binary reflected Gray mapping is used for BICM. BICM is therefore likely to be a good method to perform channel coding in the energy channels, by using simple binary codes, such as convolutional or turbo-like codes.

The behaviour of the constrained capacity at low energy levels, when the capacity becomes vanishingly small, has been determined. The main results are summarized in Table 5.1, where the capacity per unit energy, the minimum energy per nat[1], and the energy at zero-capacity for PEM modulation and flash signalling are presented. It is remarkable that the energy per nat (or the capacity per unit energy) shows a smooth transition between the AWGN and DTP channels, going through the AEN and AE-Q models: for all models, it is equal to the average energy of the additive noise component.

Even though the capacity per unit energy of the Gaussian channel is attained by any zero-mean unit-energy constellation, this result does not extend to the additive energy channels. The performance of a specific modulation format at low energy levels strongly depends on the modulation format itself. As seen in Table 5.1, the minimum energy per nat is attained by the appropriate limits of PEM(λ) of flash signalling in the AEN and DTP channels. Intriguingly, it is not the case for the AE-Q channel, whose energy per nat at zero energy is infinite. These seemingly contradictory extremes have been reconciled in Section 5.5.3 by looking at somewhat higher channel capacities, for which the AE-Q model represents a smooth transition between the AEN and DTP channel models. Care must be exercised when analyzing the asymptotic behaviour, as it may not be representative of a more practical scenario.

For all the channel models, we have computed the first two coefficients of

[1] In the chapter, we have rather considered the energy per bit. Since $\log 2$ nats=1 bit, giving the energy per nat removes the ubiquitous factor $\log 2$ in the formulas for the energy per bit and makes for more compact results. Needless to say, this change of units has no effect on the conclusions.

Channel model	AWGN	AEN	DTP	AE-Q
Cap./unit energy	$1/\sigma^2$	$1/E_n$	∞	$1/\varepsilon_n$
Min. energy/nat	σ^2	E_n	0	ε_n
c_1 – 2^m-PEM (λ)	1	$\frac{\lambda}{\lambda+1}$	$\log(\lambda+1) - \frac{\lambda}{\lambda+1}$	0
Energy at zero	σ^2	$\geq E_n$	> 0	∞
c_1 – flash (p)	1	$-\frac{p}{1-p}\log p$	$-\log(1-p)$	0
Energy at zero	σ^2	$\geq E_n$	> 0	∞

Table 5.1: Minimum energies in the additive energy channels.

the Taylor expansion of the capacity at zero energy, namely

$$\mathrm{C}(\gamma) = c_1\gamma + c_2\gamma^2 + \mathrm{o}(\gamma^2), \tag{5.80}$$

where γ is SNR for the AEN (and AWGN) channels, and ε_s for the DTP and AE-Q channels. The coefficients c_1 are given by Propositions 5.3, 5.8, and 5.14,

$$c_1^{\mathrm{AEN}} = -\sum_{j=1}^{|\mathcal{X}|-1} (x_{j+1} - x_j) q_j \log q_j \tag{5.81}$$

$$c_1^{\mathrm{DTP}} = \sum_{x\in\mathcal{X}} x P(x) \log \frac{x}{\mu_1(\mathcal{X})} \tag{5.82}$$

$$c_1^{\text{AE-Q}} = 0, \tag{5.83}$$

where $q_j = \sum_{s'\leq s_j} P(x')$. The linear coefficients c_1 for 2^m-PEM and flash signalling are given in Table 5.1. As for the coefficients c_2, they are given by

$$c_2^{\mathrm{AEN}} = \frac{1}{2}\sigma^2(\mathcal{X}) + \sum_{j=1}^{|\mathcal{X}|-1} (s_{j+1} - s_j)\Big(\tfrac{1}{2}(s_j + s_{j+1})q_j - q'_j\Big) \log q_j, \tag{5.84}$$

$$c_2^{\mathrm{DTP}} = \frac{1}{2}\sigma^2(\mathcal{X}) - \frac{1}{2}\mu_2(\mathcal{X}) \log \frac{\mu_2(\mathcal{X})}{\mu_1^2(\mathcal{X})}, \tag{5.85}$$

$$c_2^{\text{AE-Q}} = \frac{1}{2\varepsilon_n}\sigma^2(\mathcal{X}), \tag{5.86}$$

where $\sigma^2(\mathcal{X}) = \mu_2(\mathcal{X}) - \mu_1^2(\mathcal{X})$, and $q'_j = \sum_{s'\leq s_j} s' P(x')$. This analysis complements the results obtained for the Gaussian channel, in Proposition 4.4.

Next to the low-energy region, we have studied the asymptotic expression of the capacity for large energy levels. From Eqs. (5.42) and (5.63) (under a Gaussian approximation), we have

$$\mathrm{C}_{\mathcal{X}}^{\mathrm{AEN}} \simeq \log \mathrm{SNR} - \lambda + \log\big(\lambda(\lambda+1)\big), \tag{5.87}$$

$$\mathrm{C}_{\mathcal{X}}^{\mathrm{DTP}} \simeq \tfrac{1}{2}\log \varepsilon_s + \tfrac{1}{2}\big(1-\lambda\big) + \tfrac{1}{2}\log\big(\lambda^2(\lambda+1)\big) - \tfrac{1}{2}\log(2\pi), \tag{5.88}$$

for 2^m-PEM constellations with parameter λ. The value of λ which maximizes these expressions was determined, and found to be $\lambda^{\mathrm{AEN}} = \frac{1}{2}(1+\sqrt{5})$ and $\lambda^{\mathrm{DTP}} = 1+\sqrt{3}$. The AE-Q channel shows a transition between the two as the number of quanta grows and Poisson noise becomes more important, compared to the additive noise. Uniform PEM was found to suffer from an energy loss factor of $\frac{e}{2}$ (AEN) and π (DTP), to be compared with the value $\frac{\pi e}{6}$ for square constellations in the Gaussian channel. Optimizing λ reduced these values to 1.19 (AEN), saving 0.58 dB, and 1.27 (DTP), a reduction of 3.92 dB.

Even though some of the results seem of a rather technical nature, it is hoped that the analysis presented in this section may prove useful in the design and performance analysis of, at least, some optical communication systems. In the following chapter we complement the analysis presented so far by studying the pairwise error probability for binary codes, a fundamental tool to study the effective performance of channel codes.

5.A CM Capacity Expansion at Low SNR – AEN

We follow the same steps as for AWGN, with the necessary adaptations. For the sake of compactness, we define $\gamma = \mathrm{SNR}$ and respectively denote the first- and second-order moments of the constellation by μ_1 and μ_2. Recall that the CM capacity is

$$\mathrm{C}_{\mathcal{X}} = -\sum_{l=1}^{|\mathcal{X}|} P(x_l) \sum_{j=l}^{|\mathcal{X}|} \big(e^{-\gamma(x_j - x_l)} - e^{-\gamma(x_{j+1}-x_l)}\big) \log\left(\sum_{x' \le x_j} P(x') e^{-\gamma(x_l - x')}\right). \tag{5.89}$$

For each of the summands in the $\log(\cdot)$ in Eq. (5.89) we use $e^t = 1 + t + \frac{1}{2}t^2 + \mathrm{O}(t^3)$ to obtain

$$e^{-\gamma(x_l - x')} = 1 + \gamma(x' - x_l) + \tfrac{1}{2}\gamma^2\big(x'^2 + x_l^2 - 2x'x_l\big) + \mathrm{O}(\gamma^3). \tag{5.90}$$

Let us now define the variables q_j, q'_j, and q''_j respectively as

$$q_j = \sum_{x' \le x_j} P(x'), \quad q'_j = \sum_{x' \le x_j} x' P(x'), \quad q''_j = \sum_{x' \le x_j} x'^2 P(x'). \tag{5.91}$$

Rearranging, the sum in the logarithm over x' such that $x' \le x_j$ gives

$$\sum_{x' \le x_j} e^{\gamma(x'-x_l)} P(x') = q_j \left(1 + \gamma \frac{q'_j}{q_j} - \gamma x_l + \tfrac{1}{2}\gamma^2 \frac{q''_j}{q_j} - \gamma^2 x_l \frac{q'_j}{q_j} + \tfrac{1}{2}\gamma^2 x_l^2 + \mathrm{O}(\gamma^3) \right). \tag{5.92}$$

Taking logarithms, and using the expansion $\log(1+t) = t - \frac{1}{2}t^2 + \mathrm{O}(t^3)$ around $\gamma = 0$, we obtain

$$\log q_j + \gamma \frac{q'_j}{q_j} - \gamma x_l + \tfrac{1}{2}\gamma^2 \frac{q''_j}{q_j} - \gamma^2 x_l \frac{q'_j}{q_j} + \tfrac{1}{2}\gamma^2 x_l^2 - \tfrac{1}{2}\gamma^2 \left(\frac{q'_j}{q_j} - x_l \right)^2 + \mathrm{O}(\gamma^3) \tag{5.93}$$

$$= \log q_j + \gamma \frac{q'_j}{q_j} - \gamma x_l + \tfrac{1}{2}\gamma^2 \frac{q''_j}{q_j} - \tfrac{1}{2}\gamma^2 \left(\frac{q'_j}{q_j} \right)^2 + \mathrm{O}(\gamma^3). \tag{5.94}$$

We now move on to the summation over j in Eq. (5.89). We use the Taylor expansion of the exponential function in the summation over j, separating the last term as special. Starting at it, $j = |\mathcal{X}|$, we note that the sum over $x' \le x_{|\mathcal{X}|}$ includes all symbols, and its contribution to the sum is

$$\Big(1 - \gamma(x_{|\mathcal{X}|} - x_l) + \tfrac{1}{2}\gamma^2 (x_{|\mathcal{X}|} - x_l)^2\Big) \times$$
$$\times \left(\log q_j + \gamma \frac{q'_j}{q_j} - \gamma x_l + \tfrac{1}{2}\gamma^2 \frac{q''_j}{q_j} - \tfrac{1}{2}\gamma^2 \left(\frac{q'_j}{q_j} \right)^2 \right) + \mathrm{O}(\gamma^3) \tag{5.95}$$

$$= \Big(1 - \gamma(x_{|\mathcal{X}|} - x_l) + \tfrac{1}{2}\gamma^2 (x_{|\mathcal{X}|} - x_l)^2\Big) \Big(\gamma(\mu_1 - x_l) + \tfrac{1}{2}\gamma^2 \big(\mu_2 - \mu_1^2\big) \Big) + \mathrm{O}(\gamma^3), \tag{5.96}$$

since $\sum_{x'} P(x') = 1$. Carrying out the expectation over x_l, and discarding terms of order $\mathrm{O}(\gamma^3)$, this term contributes with

$$\sum_l P(x_l) \Big(\gamma(\mu_1 - x_l) + \tfrac{1}{2}\gamma^2 \big(\mu_2 - \mu_1^2\big) - \gamma^2 (x_{|\mathcal{X}|} - x_l)(\mu_1 - x_l) \Big) = \tag{5.97}$$

$$= -\tfrac{1}{2}\gamma^2 \big(\mu_2 - \mu_1^2\big). \tag{5.98}$$

As for the terms $j < |\mathcal{X}|$, the following terms contribute

$$\sum_{j=l}^{|\mathcal{X}|-1} \Big(\gamma(x_{j+1}-x_j) + \tfrac{1}{2}\gamma^2(x_j^2 - x_{j+1}^2 - 2(x_j - x_{j+1})x_l)\Big)\left(\log q_j + \gamma\frac{q_j'}{q_j} - \gamma x_l\right) \tag{5.99}$$

$$= \sum_{j=l}^{|\mathcal{X}|-1} \Bigg\{\gamma(x_{j+1}-x_j)\log q_j + \tfrac{1}{2}\gamma^2(x_j^2 - x_{j+1}^2)\log q_j - \gamma^2(x_j - x_{j+1})x_l \log q_j + \gamma^2(x_{j+1}-x_j)\left(\frac{q_j'}{q_j} - x_l\right)\Bigg\}. \tag{5.100}$$

The only remaining step is the averaging over x_l, which yields

$$\mathrm{C}_{\mathcal{X}} = -\sum_l P(x_l) \sum_{j=l}^{|\mathcal{X}|-1} \Bigg\{\gamma(x_{j+1}-x_j)\log q_j + \tfrac{1}{2}\gamma^2(x_j^2 - x_{j+1}^2)\log q_j - \gamma^2(x_j - x_{j+1})x_l \log q_j + \gamma^2(x_{j+1}-x_j)\left(\frac{q_j'}{q_j} - x_l\right)\Bigg\} + \tfrac{1}{2}\gamma^2\big(\mu_2 - \mu_1^2\big) + \mathrm{O}(\gamma^3). \tag{5.101}$$

The order of the double summation over l and j can be reversed, with the summation limits becoming

$$\sum_{j=1}^{|\mathcal{X}|-1} \sum_{l\le j} P(x_l)\Bigg\{\gamma(x_{j+1}-x_j)\log q_j + \gamma^2(x_j - x_{j+1})\big(\tfrac{1}{2}(x_j + x_{j+1}) - x_l\big)\log q_j + \gamma^2(x_{j+1}-x_j)\left(\frac{\sum_{x'\le x_j} x' P(x')}{q_j} - x_l\right)\Bigg\} \tag{5.102}$$

$$= \sum_{j=1}^{|\mathcal{X}|-1} \Big(\gamma(x_{j+1}-x_j)q_j \log q_j + \gamma^2(x_j - x_{j+1})\big(\tfrac{1}{2}(x_j + x_{j+1})q_j - q_j'\big)\log q_j\Big). \tag{5.103}$$

Therefore the desired expression for the CM capacity is

$$\mathrm{C}_{\mathcal{X}} = -\sum_{j=1}^{|\mathcal{X}|-1} \Big(\gamma(x_{j+1}-x_j)q_j \log q_j - \gamma^2(x_{j+1}-x_j)\big(\tfrac{1}{2}(x_j + x_{j+1})q_j - q_j'\big)\log q_j\Big) + \tfrac{1}{2}\gamma^2\big(\mu_2 - \mu_1^2\big) + \mathrm{O}(\gamma^3). \tag{5.104}$$

5.B CM Capacity Expansion at Low ε_s – DTP

The CM capacity is

$$\mathrm{C}_{\mathcal{X}} = -\sum_{x} P(x) \sum_{y=0}^{\infty} Q(y|x) \log\left(\sum_{x' \in \mathcal{X}} P(x') e^{\varepsilon_s (x-x')} \left(\frac{x'}{x} \right)^{y} \right). \tag{5.105}$$

For the sake of compactness, we respectively denote the first- and second-order moments of the constellation by μ_1 and μ_2.

Using the Taylor expansion of the exponential, $e^t = 1 + t + \frac{1}{2}t^2 + \mathrm{O}(t^3)$, we notice that, in the low ε_s region, there are only three possible channel outputs to order ε_s^3, namely

$$y = 0, \quad Q(y|x) = 1 - \varepsilon_s x + \tfrac{1}{2}\varepsilon_s^2 x^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.106}$$

$$y = 1, \quad Q(y|x) = \varepsilon_s x - \varepsilon_s^2 x^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.107}$$

$$y = 2, \quad Q(y|x) = \tfrac{1}{2}\varepsilon_s^2 x^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.108}$$

$$y \geq 2, \quad Q(y|x) = \mathrm{O}(\varepsilon_s^3). \tag{5.109}$$

Since each of these cases behaves differently, we examine them separately.

We rewrite the variable in the $\log(\cdot)$ in Eq. (5.105) with the appropriate approximation. When the output is $y = 0$, the variable is

$$\sum_{x' \in \mathcal{X}} P(x') e^{\varepsilon_s (x-x')} = \sum_{x' \in \mathcal{X}} P(x') \Big(1 + \varepsilon_s (x - x') + \tfrac{1}{2}\varepsilon_s^2 (x - x')^2 + \mathrm{O}(\varepsilon_s^3) \Big) \tag{5.110}$$

$$= 1 + \varepsilon_s (x - \mu_1) + \tfrac{1}{2}\varepsilon_s^2 (x^2 + \mu_2 - 2x\mu_1) + \mathrm{O}(\varepsilon_s^3). \tag{5.111}$$

Taking logarithms, and using the formula $\log(1+t) = t - \frac{1}{2}t^2 + \mathrm{O}(t^3)$, we obtain

$$\varepsilon_s (x - \mu_1) + \tfrac{1}{2}\varepsilon_s^2 (x^2 + \mu_2 - 2x\mu_1) - \tfrac{1}{2}\varepsilon_s^2 (x^2 + \mu_1^2 - 2x\mu_1) + \mathrm{O}(\varepsilon_s^3) \tag{5.112}$$

$$= \varepsilon_s (x - \mu_1) + \tfrac{1}{2}\varepsilon_s^2 \big(\mu_2 - \mu_1^2\big) + \mathrm{O}(\varepsilon_s^3). \tag{5.113}$$

When the output is $y = 1$, the variable in the logarithm in Eq. (5.105) is

$$\sum_{x' \in \mathcal{X}} P(x') e^{\varepsilon_s (x-x')} \frac{x'}{x} = \frac{1}{x} \sum_{x' \in \mathcal{X}} P(x') x' \Big(1 + \varepsilon_s (x - x') + \mathrm{O}(\varepsilon_s^2) \Big) \tag{5.114}$$

$$= \frac{1}{x} \Big(\mu_1 + \varepsilon_s \big(x\mu_1 - \mu_2 \big) + \mathrm{O}(\varepsilon_s^2) \Big) \tag{5.115}$$

$$= \frac{\mu_1}{x} \Big(1 + \frac{\varepsilon_s}{\mu_1} \big(x\mu_1 - \mu_2 \big) + \mathrm{O}(\varepsilon_s^2) \Big). \tag{5.116}$$

Taking logarithms, and using the Taylor expansion of the logarithm, we get

$$\log \frac{\mu_1}{x} + \frac{\varepsilon_s}{\mu_1}\big(x\mu_1 - \mu_2\big) + \mathrm{O}(\varepsilon_s^2). \tag{5.117}$$

We will later verify that no higher-order terms are required.

At last, for $y = 2$, the variable in the logarithm in Eq. (5.105) is

$$\sum_{x' \in \mathcal{X}} P(x') e^{\varepsilon_s (x - x')} \frac{x'^2}{x^2} = \frac{1}{x^2} \sum_{x' \in \mathcal{X}} P(x') x'^2 \Big(1 + \mathrm{O}(\varepsilon_s)\Big) \tag{5.118}$$

$$= \frac{1}{x^2}\Big(\mu_2 + \mathrm{O}(\varepsilon_s)\Big) \tag{5.119}$$

$$= \frac{\mu_2}{x^2}\big(1 + \mathrm{O}(\varepsilon_s)\big). \tag{5.120}$$

Taking logarithms, and using the Taylor expansion of the logarithm, we get

$$\log \frac{\mu_2}{x^2} + \mathrm{O}(\varepsilon_s). \tag{5.121}$$

Later, we will verify that no higher-order terms are required.

After carrying out the averaging over y, we first combine Eqs. (5.113), (5.117) and (5.121) with the probabilities in Eqs. (5.106)–(5.108) and then group all terms up to $\mathrm{O}(\varepsilon_s^3)$ to derive

$$\begin{aligned}
&\Big(1 - \varepsilon_s x + \tfrac{1}{2}\varepsilon_s^2 x^2\Big)\Big(\varepsilon_s(x - \mu_1) + \tfrac{1}{2}\varepsilon_s^2\big(\mu_2 - \mu_1^2\big)\Big) + \\
&\quad + \big(\varepsilon_s x - \varepsilon_s^2 x^2\big)\left(\log \frac{\mu_1}{x} + \frac{\varepsilon_s}{\mu_1}\big(x\mu_1 - \mu_2\big)\right) + \tfrac{1}{2}\varepsilon_s^2 x^2 \log \frac{\mu_2}{x^2} + \mathrm{O}(\varepsilon_s^3) \qquad (5.122) \\
&= \varepsilon_s(x - \mu_1) + \tfrac{1}{2}\varepsilon_s^2\big(\mu_2 - \mu_1^2\big) - \varepsilon_s^2(x^2 - x\mu_1) + \varepsilon_s x \log \frac{\mu_1}{x} \\
&\quad - \varepsilon_s^2 x^2 \log \frac{\mu_1}{x} + \varepsilon_s^2 \frac{1}{\mu_1}\big(x^2\mu_1 - x\mu_2\big) + \tfrac{1}{2}\varepsilon_s^2 x^2 \log \frac{\mu_2}{x^2} + \mathrm{O}(\varepsilon_s^3). \qquad (5.123)
\end{aligned}$$

The expectation over x is straightforward, and gives

$$\mathrm{C}_{\mathcal{X}} = \varepsilon_s \sum_{x \in \mathcal{X}} P(x) x \log \frac{x}{\mu_1} + \frac{1}{2}\varepsilon_s^2 \left(\mu_2 - \mu_1^2 - \mu_2 \log \frac{\mu_2}{\mu_1^2}\right) + \mathrm{O}(\varepsilon_s^3). \tag{5.124}$$

5.C Capacity per Unit Energy in the AE-Q Channel

From Theorem 3.9, the capacity per unit energy can computed as

$$\mathrm{C}_1 = \sup_x \frac{D\big(Q(y|x) || Q(y|x = 0)\big)}{\varepsilon_s x}. \tag{5.125}$$

Using Eq. (5.76) for $Q(\cdot|\cdot)$ and the definition of divergence, we have

$$D\big(Q(y|x)||Q(y|x=0)\big) = \sum_y Q(y|x)\log\left(e^{\frac{\varepsilon_s x}{\varepsilon_n}}\sum_{l=0}^{y} e^{-\alpha}\frac{\alpha^l}{l!}\right), \tag{5.126}$$

where $\alpha = \varepsilon_s x\big(1+\frac{1}{\varepsilon_n}\big)$. Let us define $P(l) = e^{-\alpha}\frac{\alpha^l}{l!}$ and the quantity $q(y) = \sum_{l=0}^{y} P(l)$, i. e. the cumulative distribution function of a Poisson random variable with mean α.

Moving the exponential out of the logarithm, we obtain

$$D\big(Q(y|x)||Q(y|x=0)\big) = \frac{\varepsilon_s x}{\varepsilon_n} + \sum_y Q(y|x)\log\big(q(y)\big). \tag{5.127}$$

Hence, the capacity per unit energy is given by

$$\mathrm{C}_1 = \frac{1}{\varepsilon_n} + \sup_x \frac{\sum_y Q(y|x)\log\big(q(y)\big)}{\varepsilon_s x}. \tag{5.128}$$

Since $q(y) \leq 1$, its logarithm is always non-positive, and

$$\mathrm{C}_1 \leq \frac{1}{\varepsilon_n}. \tag{5.129}$$

The proof is completed by proving that

$$\lim_{x\to\infty} \frac{e^{\frac{\varepsilon_s x}{\varepsilon_n}}}{\varepsilon_s x(1+\varepsilon_n)}\left(\sum_y \left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y)\log\big(q(y)\big)\right) = 0, \tag{5.130}$$

where we expressed $Q(y|x)$ as a function of $q(y)$. If this condition holds true, then Eq. (5.129) becomes an equality.

In Eq. (5.130) we split the summation over y into two parts, from 0 to $y^* = \lfloor\alpha\rfloor$, and from y^*+1 to infinity. In the first part, $e^{-\alpha}\frac{\alpha^y}{y!}$ is an increasing function in y, and therefore

$$q(y) = \sum_{l=0}^{y} P(l) \geq \sum_{l=0}^{y} P(0) = (y+1)P(0) = (y+1)e^{-\alpha}. \tag{5.131}$$

Hence, the summation for $y \leq y^*$ is bounded as

$$\sum_{y=0}^{y^*}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y)\log\big(q(y)\big) \geq \sum_{y=0}^{y^*}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1)e^{-\alpha}\big(\log(y+1)-\alpha\big). \tag{5.132}$$

And, multiplying by the exponential factor $e^{\frac{\varepsilon_s x}{\varepsilon_n}}$, we have

$$\sum_{y=0}^{y^*} e^{-\varepsilon_s x}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1)\log(y+1) - \sum_{y=0}^{y^*} e^{-\varepsilon_s x}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1)\alpha. \tag{5.133}$$

As $y \to \infty$, each of these two summands goes to zero. The second has the form

$$e^{-\varepsilon_s x}\sum_{y=0}^{y^*}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1)\alpha, \tag{5.134}$$

which decays exponentially in x, since the sum satisfies

$$\sum_{y=0}^{y^*}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1) \le \sum_{y=0}^{\infty}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1) = (1+\varepsilon_n)^2, \tag{5.135}$$

and $e^{-\varepsilon_s x}\alpha(1+\varepsilon_n)^2$ vanishes for large x. Similarly, the summation

$$\sum_{y=0}^{y^*}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1)\log(y+1) \tag{5.136}$$

remains bounded, since it is the partial sum of a convergent series, with n-th coefficient $\beta^n(n+1)\log(n+1)$ and $\beta = \varepsilon_n/(1+\varepsilon_n) < 1$. This is verified by the checking the ratio test, as

$$\lim_{n\to\infty} \frac{\beta^n(n+1)\log(n+1)}{\beta^{n-1}n\log n} = \beta < 1. \tag{5.137}$$

Boundedness of the partial sum implies that, after multiplying times an exponential factor $e^{-\varepsilon_s x}$, the first summand vanishes as $x \to \infty$.

Next, we consider the remainder of the summation in Eq. (5.130),

$$\sum_{y=y^*+1}^{\infty}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y)\log\big(q(y)\big). \tag{5.138}$$

Clearly, $q(0) = e^{-\alpha} \le q(y) \le 1$ and therefore $-\alpha \le \log q(y) \le 0$, so each summand is negative and bounded by

$$\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y)\log\big(q(y)\big) \ge -\alpha\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y) \ge -\alpha\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y. \tag{5.139}$$

Summing over y,

$$\sum_{y=y^*+1}^{\infty}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y)\log\big(q(y)\big) \geq -\alpha(1+\varepsilon_n)\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^{y^*+1}. \tag{5.140}$$

Using the definition of $\alpha = \varepsilon_s x\big(1+\frac{1}{\varepsilon_n}\big)$ and taking into account the denominator $\varepsilon_s x(1+\varepsilon_n)$ in Eq. (5.130), we must study the behaviour of

$$-\frac{\varepsilon_n+1}{\varepsilon_n}e^{\frac{\varepsilon_s x}{\varepsilon_n}}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^{y^*+1} = -\frac{\varepsilon_n+1}{\varepsilon_n}e^{\frac{\varepsilon_s x}{\varepsilon_n}-(y^*+1)\log\left(1+\frac{1}{\varepsilon_n}\right)} \tag{5.141}$$

as $x \to \infty$. By construction, $y^*+1 > \alpha$, and therefore

$$\frac{\varepsilon_s x}{\varepsilon_n} - (y^*+1)\log\left(1+\frac{1}{\varepsilon_n}\right) < \frac{\varepsilon_s x}{\varepsilon_n} - \varepsilon_s x\left(1+\frac{1}{\varepsilon_n}\right)\log\left(1+\frac{1}{\varepsilon_n}\right) \tag{5.142}$$

$$= \varepsilon_s x\left(\frac{1}{\varepsilon_n} - \left(1+\frac{1}{\varepsilon_n}\right)\log\left(1+\frac{1}{\varepsilon_n}\right)\right). \tag{5.143}$$

Since $t \leq (1+t)\log(1+t)$ for $t > 0$, a fact which follows from the inequality $\log(1+t) \leq t$, this left-hand side of Eq. (5.142) is strictly upper bounded by a function ax, where a is negative. Hence, the function in Eq. (5.141) vanishes exponentially as $x \to \infty$, and so does the term

$$\frac{e^{\frac{\varepsilon_s x}{\varepsilon_n}}}{\varepsilon_s x(1+\varepsilon_n)}\sum_{y=y^*+1}^{\infty}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y)\log\big(q(y)\big). \tag{5.144}$$

This proves the limit in Eq. (5.130) and shows that Eq. (5.129) holds with equality.

5.D CM Capacity Expansion at Low ε_s – AE-Q

The constrained capacity for coded modulation is

$$\mathrm{C}_{\mathcal{X}} = -\sum_x P(x)\sum_{y=0}^{\infty} Q(y|x)\log\left(\frac{\sum_{x'\in\mathcal{X}} P(x')Q(y|x')}{Q(y|x)}\right), \tag{5.145}$$

where $Q(y|x)$ is given by Eq. (2.23), namely

$$Q(y|x) = \frac{1}{1+\varepsilon_n}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y\left(\sum_{l=0}^{y}\left(\frac{1+\varepsilon_n}{\varepsilon_n}\right)^l e^{-\varepsilon_s x}\frac{(\varepsilon_s x)^l}{l!}\right). \tag{5.146}$$

As it happened in the discrete-time Poisson channel, the Taylor expansion of the exponential, $e^t = 1 + t + \frac{1}{2}t^2 + \mathrm{O}(t^3)$, implies that there are only three possible channel outputs s to order ε_s^3, that is,

$$s = 0, \quad P(s|x) = 1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.147}$$

$$s = 1, \quad P(s|x) = \varepsilon_s x - (\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.148}$$

$$s = 2, \quad P(s|x) = \frac{1}{2}(\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.149}$$

$$s > 2, \quad P(s|x) = \mathrm{O}(\varepsilon_s^3). \tag{5.150}$$

Hence the channel output $y = s + z$ only includes these contributions. We distinguish three cases, viz. $y = 0$, $y = 1$, and $y \geq 2$.

In the first case, $y = s = z = 0$, and $Q(y|x)$ becomes

$$Q(y|x) = \frac{1}{1+\varepsilon_n}\left(1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3)\right). \tag{5.151}$$

For $y = 1$, we combine the outputs $s = 0$ and $s = 1$,

$$Q(y|x) = \frac{\varepsilon_n}{(1+\varepsilon_n)^2}\left(1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2 + \frac{1+\varepsilon_n}{\varepsilon_n}\big(\varepsilon_s x - (\varepsilon_s x)^2\big) + \mathrm{O}(\varepsilon_s^3)\right) \tag{5.152}$$

$$= \frac{\varepsilon_n}{(1+\varepsilon_n)^2}\left(1 + \frac{\varepsilon_s}{\varepsilon_n}x - \varepsilon_s^2 x^2\left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right) + \mathrm{O}(\varepsilon_s^3)\right). \tag{5.153}$$

For $y \geq 2$, we combine the outputs $s = 0$, $s = 1$, and $s = 2$,

$$Q(y|x) = \frac{\varepsilon_n^y}{(1+\varepsilon_n)^{y+1}}\left(1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2 + \big(\varepsilon_s x - (\varepsilon_s x)^2\big)\left(\frac{1+\varepsilon_n}{\varepsilon_n}\right) + \frac{1}{2}(\varepsilon_s x)^2\left(\frac{1+\varepsilon_n}{\varepsilon_n}\right)^2 + \mathrm{O}(\varepsilon_s^3)\right) \tag{5.154}$$

$$= \frac{\varepsilon_n^y}{(1+\varepsilon_n)^{y+1}}\left(1 + \frac{\varepsilon_s}{\varepsilon_n}x + \frac{\varepsilon_s^2}{2\varepsilon_n^2}x^2 + \mathrm{O}(\varepsilon_s^3)\right), \tag{5.155}$$

after combining some terms together.

We next rewrite the numerator and denominator in the $\log(\cdot)$ in Eq. (5.145) with the appropriate approximation. For $y = 0$, the common term $(1+\varepsilon_n)^{-1}$ cancels, and the numerator is

$$\sum_{x' \in \mathcal{X}} Q(y|x')P(x') = \sum_{x' \in \mathcal{X}} \big(1 - \varepsilon_s x' + \frac{1}{2}(\varepsilon_s x')^2 + \mathrm{O}(\varepsilon_s^3)\big)P(x') \tag{5.156}$$

$$= \left(1 - \varepsilon_s \mu_1 + \frac{1}{2}\varepsilon_s^2 \mu_2\right) + \mathrm{O}(\varepsilon_s^3). \tag{5.157}$$

In the denominator, we keep the expansion

$$1 - \varepsilon_s x + \tfrac{1}{2}(\varepsilon_s x)^2\, \mathrm{O}(\varepsilon_s^3). \tag{5.158}$$

Taking logarithms of Eqs. (5.157) and (5.158), using the formula $\log(1+t) = t - \frac{1}{2}t^2 + \mathrm{O}(t^3)$, and combining numerator and denominator, we obtain

$$\log\left(1 - \varepsilon_s \mu_1 + \frac{1}{2}\varepsilon_s^2 \mu_2 + \mathrm{O}(\varepsilon_s^3)\right) - \log\left(1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3)\right) \tag{5.159}$$

$$= -\varepsilon_s \mu_1 + \frac{1}{2}\varepsilon_s^2 \mu_2 - \frac{1}{2}\varepsilon_s^2 \mu_1^2 + \varepsilon_s x - \frac{1}{2}(\varepsilon_s x)^2 + \frac{1}{2}(\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.160}$$

$$= -\varepsilon_s(\mu_1 - x) + \frac{1}{2}\varepsilon_s^2(\mu_2 - \mu_1^2) + \mathrm{O}(\varepsilon_s^3). \tag{5.161}$$

If the output is $y = 1$, we use Eq. (5.153). Summing over x' in the numerator, we get

$$\sum_{x' \in \mathcal{X}} Q(y|x')P(x') = \left(1 + \frac{\varepsilon_s}{\varepsilon_n}\mu_1 - \left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\mu_2 \varepsilon_s^2 + \mathrm{O}(\varepsilon_s^3)\right), \tag{5.162}$$

with the agreement that a common term $\varepsilon_n/(1+\varepsilon_n)^2$ has been cancelled.

Combining numerator and denominator, taking logarithms, and using the Taylor expansion of the logarithm, we obtain

$$\frac{\varepsilon_s}{\varepsilon_n}\mu_1 - \left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\mu_2 \varepsilon_s^2 - \frac{\varepsilon_s^2}{2\varepsilon_n^2}\mu_1^2 - \frac{\varepsilon_s}{\varepsilon_n}x + \left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)x^2 \varepsilon_s^2 + \frac{\varepsilon_s^2}{2\varepsilon_n^2}x^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.163}$$

$$= \frac{(\mu_1 - x)\varepsilon_s}{\varepsilon_n} - \left(\left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\mu_2 + \frac{\mu_1^2}{2\varepsilon_n^2} - \frac{x^2(1+\varepsilon_n)^2}{2\varepsilon_n^2}\right)\varepsilon_s^2 + \mathrm{O}(\varepsilon_s^3). \tag{5.164}$$

If the output is $y \geq 2$, in an analogous way we use Eq. (5.155) to rewrite the logarithm of the ratio of numerator and denominator as

$$\log\left(1 + \frac{\varepsilon_s}{\varepsilon_n}\mu_1 + \frac{\varepsilon_s^2}{2\varepsilon_n^2}\mu_2 + \mathrm{O}(\varepsilon_s^3)\right) - \log\left(1 + \frac{\varepsilon_s}{\varepsilon_n}x + \frac{\varepsilon_s^2}{2\varepsilon_n^2}x^2 + \mathrm{O}(\varepsilon_s^3)\right). \tag{5.165}$$

Using now the Taylor expansion of the logarithm, we obtain

$$\frac{\varepsilon_s}{\varepsilon_n}\mu_1 + \frac{\varepsilon_s^2}{2\varepsilon_n^2}\mu_2 - \frac{\varepsilon_s^2}{2\varepsilon_n^2}\mu_1^2 - \frac{\varepsilon_s}{\varepsilon_n}x - \frac{\varepsilon_s^2}{2\varepsilon_n^2}x^2 + \frac{\varepsilon_s^2}{2\varepsilon_n^2}x^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.166}$$

$$= \frac{\varepsilon_s}{\varepsilon_n}(\mu_1 - x) + \frac{\varepsilon_s^2}{2\varepsilon_n^2}(\mu_2 - \mu_1^2) + \mathrm{O}(\varepsilon_s^3). \tag{5.167}$$

The remaining steps are the averaging over x and y. We first carry out the expectation over x. From Eq. (5.161), the averaging over x yields

$$\sum_x \frac{1}{1+\varepsilon_n}\left(1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2\right)\left((x-\mu_1)\varepsilon_s + \frac{1}{2}\varepsilon_s^2(\mu_2 - \mu_1^2)\right) + \mathrm{O}(\varepsilon_s^3) \tag{5.168}$$

$$= \frac{1}{1+\varepsilon_n}\left(\frac{1}{2}\varepsilon_s^2(\mu_2 - \mu_1^2) - \varepsilon_s^2(\mu_2 - \mu_1^2)\right) + \mathrm{O}(\varepsilon_s^3) \tag{5.169}$$

$$= \frac{1}{1+\varepsilon_n}\frac{1}{2}\varepsilon_s^2(\mu_1^2 - \mu_2) + \mathrm{O}(\varepsilon_s^3). \tag{5.170}$$

Similarly, from Eq. (5.164) we obtain (bar for a constant factor $\frac{\varepsilon_n}{(1+\varepsilon_n)^2}$)

$$\sum_x \left(1 + \frac{\varepsilon_s}{\varepsilon_n}x - \varepsilon_s^2 x^2\left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\right) \times$$
$$\times \left(\frac{(\mu_1 - x)\varepsilon_s}{\varepsilon_n} - \left(\left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\mu_2 + \frac{\mu_1^2}{2\varepsilon_n^2} - \frac{x^2(1+\varepsilon_n)^2}{2\varepsilon_n^2}\right)\varepsilon_s^2\right) + \mathrm{O}(\varepsilon_s^3) \tag{5.171}$$

$$= \left(-\left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\mu_2 - \frac{\mu_1^2}{2\varepsilon_n^2} + \frac{\mu_2(1+\varepsilon_n)^2}{2\varepsilon_n^2} + \frac{\varepsilon_s^2}{\varepsilon_n^2}(\mu_1^2 - \mu_2)\right)\varepsilon_s^2 + \mathrm{O}(\varepsilon_s^3) \tag{5.172}$$

$$= (\mu_1^2 - \mu_2)\frac{\varepsilon_s^2}{2\varepsilon_n^2} + \mathrm{O}(\varepsilon_s^3). \tag{5.173}$$

And finally, from Eq. (5.167), for $y \geq 2$, we get

$$\sum_x \frac{\varepsilon_n^y}{(1+\varepsilon_n)^{y+1}} \left(1 + \frac{\varepsilon_s}{\varepsilon_n} x + \frac{\varepsilon_s^2}{2\varepsilon_n^2} x^2\right) \left(\frac{\varepsilon_s}{\varepsilon_n}(\mu_1 - x) + \frac{\varepsilon_s^2}{2\varepsilon_n^2}(\mu_2 - \mu_1^2)\right) + \mathrm{O}(\varepsilon_s^3) \tag{5.174}$$

$$= \frac{\varepsilon_n^y}{(1+\varepsilon_n)^{y+1}} \left(\frac{\varepsilon_s^2}{2\varepsilon_n^2}(\mu_2 - \mu_1^2) + \frac{\varepsilon_s^2}{\varepsilon_n^2}(\mu_1^2 - \mu_2)\right) + \mathrm{O}(\varepsilon_s^3) \tag{5.175}$$

$$= \frac{\varepsilon_n^y}{(1+\varepsilon_n)^{y+1}} \frac{\varepsilon_s^2}{2\varepsilon_n^2}(\mu_1^2 - \mu_2) + \mathrm{O}(\varepsilon_s^3). \tag{5.176}$$

The summation over $y \geq 2$ can be carried out and yields

$$\sum_{y=2}^{\infty} \frac{\varepsilon_n^y}{(1+\varepsilon_n)^y} = \frac{\varepsilon_n^2}{1+\varepsilon_n}. \tag{5.177}$$

Then, combining Eq. (5.177) into Eq. (5.176), and summing with Eqs. (5.170) and (5.173) (including the factor $\frac{\varepsilon_n}{(1+\varepsilon_n)^2}$), we obtain

$$\frac{1}{1+\varepsilon_n} \frac{1}{2} \varepsilon_s^2 \big(\mu_1^2 - \mu_2\big) \left(1 + \frac{1}{\varepsilon_n(1+\varepsilon_n)} + \frac{1}{\varepsilon_n^2} \frac{\varepsilon_n^2}{1+\varepsilon_n}\right) + \mathrm{O}(\varepsilon_s^3) \tag{5.178}$$

$$= \frac{1}{2} \varepsilon_s^2 \big(\mu_1^2 - \mu_2\big) \frac{1}{\varepsilon_n} + \mathrm{O}(\varepsilon_s^3). \tag{5.179}$$

And, finally, the expansion for $\mathrm{C}_{\mathcal{X}}$ follows,

$$\mathrm{C}_{\mathcal{X}} = \frac{1}{2}(\mu_2 - \mu_1^2) \frac{\varepsilon_s^2}{\varepsilon_n} + \mathrm{O}(\varepsilon_s^3). \tag{5.180}$$

6

Pairwise Error Probability for Coded Transmission

6.1 Introduction

As we saw in Chapter 3, the channel capacity gives the largest rate at which information can be sent reliably, i. e. at low error rates, over a communication channel. Reliability is obtained by the operation of a channel code, or more practically of an encoder and a decoder, respectively located at the transmitting and receiving ends of the channel. We discussed in Chapters 4 and 5 some aspects of the operation of the channel code, related to the use of various modulation formats and to the use of bit-interleaved coded modulation (BICM) as a method to combine good binary codes with these modulations.

A possible continuation of our study would be the design of such binary codes. Instead, we have opted for a simpler alternative, in which we estimate the so-called pairwise error probability, rather than the performance of specific channel codes. Since the pairwise error probability is a key element in the analysis of the code performance, the tools developed here may prove to be useful in future code design for the additive energy channels.

In this chapter we follow a common practice in the optical communications literature, as represented by Helstrom [63] and Einarsson [64], and provide approximations to the error probability, rather than true upper and lower bounds, the usual approach in an information-theoretic analysis. As we shall see, the saddlepoint approximation yields an accurate estimate of the pairwise error probability. The methodology is presented in Sections 6.2 and 6.3.

In Sections 6.4 and 6.5, we analyze binary transmission and BICM in Gaussian fading channels and derive accurate and easy to compute approximations to the pairwise error probability. We next consider the two limiting cases of the additive energy channel, namely the additive exponential noise channel, in Section 6.6, and the discrete-time Poisson channel, in Section 6.7. The results

presented in these sections seem to be new, except for some special cases of the binary DTP channel. One of the most intriguing results, to be presented in Section 6.6, is the closeness of the error rates for binary transmission in the AWGN channel and its AEN equivalent.

6.2 Error Probability and the Union Bound

Two fundamental figures of merit measure the probability of error: the word error rate and the bit error rate. We start our exposition with their definition.

Let a specific message, say message w, be selected at the transmitter; all messages have the same probability. We assume that the message is represented by a sequence of l_0 input (information) bits. Then, we encode the message with a binary code and identify it with a binary codeword $\mathbf{b}$, an array of l coded bits. The codeword is mapped onto an array of modulation symbols of length n by using a mapping rule $\mu : \{0,1\}^m \to \mathcal{X}$, which groups m consecutive bits and outputs a corresponding modulation symbol x_k,

$$x_k = \mu(b_{m(k-1)+1}, b_{m(k-1)+2}, \ldots, b_{m(k-1)+m}), \quad \text{for } k = 1, \ldots, n. \tag{6.1}$$

Since a modulation symbol x is uniquely selected by m bits, we denote the sequence of bits corresponding to the symbol x by $\big(b_1(x), \ldots, b_m(x)\big)$.

The corresponding k-th channel output is denoted by y_k. At the receiver, the decoder uses a metric function $q : \mathcal{X} \times \mathcal{Y} \to \mathbf{R}$ to make a decision on which codeword was sent; here $\mathcal{Y}$ denotes the output alphabet. For a received sequence $\mathbf{y} = (y_1, \ldots, y_n)$, the codeword with largest metric

$$\prod_{k=1}^{n} q\big(\mu(b_{m(k-1)+1}, \ldots, b_{m(k-1)+m}), y_k\big) \tag{6.2}$$

is selected as estimate of the transmitted codeword. In case of ties one of the candidate codewords is randomly chosen.

The word error rate, which we denote by P_w, is the probability of selecting a codeword different from the one transmitted. Similarly, we define the bit error rate, denoted by P_b, as the average number of input bits in error.

Exact expressions for the error probabilities P_w or P_b are difficult to obtain, and one often resorts to some form of bounding, the simplest of which is probably the union bound. The union bound is built on the realization that an error event is the union of many pairwise error events. Even though these pairwise error events are in general not statistically independent, an accurate estimate

of the error probability in a region above the cut-off rate [65] is obtained by summing their probabilities. Since many existing channel codes operate in this region, viz. convolutional codes or turbo-like codes in the floor region, the union bound has wide applicability.

The word error rate P_w is upper bounded by

$$P_w \leq \sum_d A_d \ \mathrm{pep}(d), \tag{6.3}$$

where d is the Hamming distance between two codewords, namely the number of bits in which they differ, and A_d denotes the number of codewords at Hamming distance d. As for the bit error rate, P_b is given by the right-hand side of Eq. (6.3) with A_d replaced by $A'_d = \sum_{j'} \frac{j'}{l_0} A_{j',d}$, $A_{j',d}$ being the number of codewords at output Hamming distance d and input Hamming distance j'. In either case, the pairwise error probability $\mathrm{pep}(d)$ is of fundamental importance, and we devote the remainder of this chapter to its study.

6.3 Approximations to the Pairwise Error Probability

In this section, we discuss the estimation of the pairwise error probability by using a method similar to the analysis in Sections 5.3 and 5.4 of Gallager's book [31], or to the presentation of Chapters 2 and 3 of Viterbi's book [65].

We assume that the decoding metric function is of the form

$$q\big(\mu(b_1, \ldots, b_m), y\big) = \prod_{i=1}^{m} q_i(b_i, y), \tag{6.4}$$

namely the product of per-bit metric functions $q_i : \{0,1\} \times \mathcal{Y} \to \mathbf{R}$. Inspired by the analysis of BICM in Section 4.3, we let the function q_i be given by

$$q_i(b, y) = \sum_{x' \in \mathcal{X}_i^b} Q(y|x'), \tag{6.5}$$

where $\mathcal{X}_i^b$ is the set of symbols with bit b at label index i. One simple example of a different metric is $q_i(b, y) = \max_{x' \in \mathcal{X}_i^b} Q(y|x')$. Since our analysis can be easily extended to general metrics, we concentrate on the metric in Eq. (6.5).

Later, we will make use of the fact that the metric q_i in Eq. (6.5) is proportional to the transition probability $Q_i(y|b)$ of the i-th BICM parallel channel,

$$Q_i(y|b) = \frac{1}{|\mathcal{X}_i^b|} \sum_{x' \in \mathcal{X}_i^b} Q(y|x'). \tag{6.6}$$

Let us denote by b_j and b'_j the j-th bits of the transmitted and alternative codewords respectively. Taking logarithms in Eq. (6.2), the decoder makes a pairwise error if the pairwise score Ξ_{pw}, or score for short, given by

$$\Xi_{\mathrm{pw}} = \sum_{k=1}^{n} \sum_{i=1}^{m} \log \frac{q_i(b'_{m(k-1)+i}, y_k)}{q_i(b_{m(k-1)+i}, y_k)}, \tag{6.7}$$

is positive, i. e. if $\Xi_{\mathrm{pw}} > 0$. When $\Xi_{\mathrm{pw}} = 0$ an error is made with probability $\frac{1}{2}$. The pairwise score is the sum of n symbol scores Ξ_s, in turn given by the sum of m bit scores[1] Ξ_b, with their obvious definitions.

It proves fruitful to view the bit score Ξ_b as a random variable, whose distribution depends on all the random elements in the channel, including the transmitted bit b and the index i. For each value of k, a choice is made at the transmitter between the mapping rule μ in Eq. (6.1) and its binary complement. This choice is known at the receiver. Denoting the cumulant transform [67] of the bit score by $\kappa_1(r) = \log \mathrm{E}[e^{r\Xi_b}]$, we have

Definition 6.1. The cumulant transform of the bit score Ξ_b is given by

$$\kappa_1(r) = \log \mathrm{E}[e^{r\Xi_b}] = \log \left(\frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} \sum_{b \in \{0,1\}} \mathrm{E}_Y \left[\frac{q_i(\bar{b}, Y)^r}{q_i(b, Y)^r} \right] \right), \tag{6.8}$$

where the bit $\bar{b}$ is the binary complement of b and Y is the set of possible channel outputs with bit b is transmitted at position i; the expectation over Y is done according to the transition probability $Q_i(y|b)$, in Eq. (6.6).

Clearly, only the bit indices for which $b_j \neq b'_j$ have a (possibly) non-zero bit score; the number of such positions is the Hamming distance d. Since bits in the same symbol are affected by the same noise, the bit scores are generally statistically dependent. This dependence can be analyzed in a number of ways:

1. Assume the codeword length is infinite, so that the bits b_j belong to different symbols with probability 1.

2. Carry out an averaging over all possible interleavers of a given size n, and compute the average pairwise error probability, in a situation reminiscent of the uniform interleaving appearing in the analysis of turbo-like codes.

[1] In [66] we gave the somewhat clumsy name "a posteriori log-likelihood ratio" to the bit score. We use here the shorter name bit score, which has the additional advantage of being well-defined for metrics other than the one in Eq. (6.5).

3. Include in the analysis information of the position where bits are mapped into. Here, the Hamming distance is not the only relevant variable, as the position of the bits at which the codewords differ is also important.

For the binary-input Gaussian channel, all three alternatives coincide. For BICM, we choose the first option, as in [26]. We further discuss the validity of this choice in Section 6.5.3, where the analysis of Yeh *et al.* [68] for the second alternative is exploited to study QPSK modulation over fading channels.

We can compute the pairwise error probability as the tail probability of a random variable, the pairwise score, in a form that we shall use in our analysis,

Proposition 6.2. *The pairwise error probability* $\mathrm{pep}(d)$ *is given by*

$$\mathrm{pep}(d) = \Pr\left(\sum_{j=1}^{d} \Xi_{b,j} > 0\right) + \frac{1}{2}\Pr\left(\sum_{j=1}^{d} \Xi_{b,j} = 0\right), \tag{6.9}$$

where the variables $\Xi_{b,j}$ *are bit scores.*

In estimates of tail probabilities, such as in Proposition 6.2, use of the cumulant transform proves convenient [67]. Let us denote the cumulant transform of the pairwise score by $\kappa_{\mathrm{pw}}(r)$. For independent, identically distributed bit scores, we have $\kappa_{\mathrm{pw}}(r) = d\kappa_1(r)$ [67]. The main advantage brought about by using the cumulant transform is that the behaviour of the random variable is to great extent determined by its behaviour around the saddlepoint $\hat{r}$, defined as the value r for which $\kappa'_{\mathrm{pw}}(\hat{r}) = 0$ and therefore $\kappa'_1(\hat{r}) = 0$.

Proposition 6.3. *The saddlepoint is located at* $\hat{r} = \frac{1}{2}$*, that is* $\kappa'_1\left(\frac{1}{2}\right) = 0$*. In addition, the next derivatives of the cumulant transform verify*

$$\kappa''_1(\hat{r}) = \frac{\mathrm{E}\left[\Xi_b^2 e^{\hat{r}\Xi_b}\right]}{\mathrm{E}\left[e^{\hat{r}\Xi_b}\right]} \geq 0, \quad \kappa'''_1(\hat{r}) = \frac{\mathrm{E}\left[\Xi_b^3 e^{\hat{r}\Xi_b}\right]}{\mathrm{E}\left[e^{\hat{r}\Xi_b}\right]} = 0. \tag{6.10}$$

Remark that we are explicitly using the metric in Eq. (6.5).

Proof. The computation is included in Appendix 6.A. □

The cumulant transform, evaluated at the saddlepoint, appears in the Chernoff (or Bhatthacharyya) bound to the pairwise error probability,

Proposition 6.4 (Chernoff Bound)**.** *The Chernoff bound to the pairwise error probability is*

$$\mathrm{pep}(d) \leq e^{d\kappa_1(\hat{r})}. \tag{6.11}$$

This result gives a true bound and very easy to compute. It is further known to correctly give the asymptotic exponential decay [39] of the error probability for large d. However, it is not exact and in some cases, such as Gaussian fading channels, is not tight. As we next see, the saddlepoint approximation complements the Chernoff bound by including a multiplicative coefficient to obtain an expression of the form $\alpha \cdot e^{d\kappa_1(\hat{r})}$.

In Appendix 6.B we derive the saddlepoint approximation to the tail probability of a continuous random variable, including an estimate of the approximation error. The derivation was published in [66]. Even though the derivation is valid for small values of the saddlepoint $\hat{r}$, we do not exploit this property as $\hat{r} = \frac{1}{2}$. Unlike other presentations, such as Gallager's [31], we are mainly interested in finite values of d. Using that $\kappa_1'''(\hat{r}) = 0$ and denoting the fourth derivative of $\kappa_1(r)$ by $\kappa_1^{IV}(\hat{r})$, we have

Theorem 6.5. *In channels with continuous bit score Ξ_b, the pairwise error probability can be approximated to first-order by*

$$\mathrm{pep}(d) \simeq \frac{1}{\sqrt{2\pi d\kappa_1''(\hat{r})}\hat{r}}\, e^{d\kappa_1(\hat{r})}, \tag{6.12}$$

where $\hat{r} = \frac{1}{2}$. The second-order approximation is

$$\mathrm{pep}(d) \simeq \left(1 + \frac{1}{d\kappa_1''(\hat{r})}\left(-\frac{1}{\hat{r}^2} + \frac{\kappa_1^{IV}(\hat{r})}{8\kappa_1''(\hat{r})}\right)\right)\frac{1}{\sqrt{2\pi d\kappa_1''(\hat{r})}\hat{r}}\, e^{d\kappa_1(\hat{r})}. \tag{6.13}$$

Higher-order expansions are characterized by having a correction factor α polynomial in inverse powers of $\left(d\kappa_1''(\hat{r})\right)^{-1}$. In the next section we compute the pairwise error probability in the binary-input AWGN channel by explicitly using the bit score. The effect of the correction α is found to be negligible.

The approximation is different when the bit score takes values on the points of a lattice, i. e. it has the form $\alpha + i\beta$, where β is the span of the lattice, α an offset, and i indexes the random variable. In that case, the probability distribution is well approximated (see section 5.4 of [31]) by

Theorem 6.6. *For a discrete random variable Z inscribed in a lattice of span β, with cumulant transform $\kappa(r)$, let $\hat{r}$ be the root of $\kappa'(\hat{r}) = 0$. Then, the saddlepoint approximation to the probability mass function is given by*

$$P_Z(z) \simeq \frac{\beta e^{\kappa(\hat{r}) - \hat{r}z}}{\sqrt{2\pi\kappa''(\hat{r})}}. \tag{6.14}$$

6.3.1 Error Rates in the Z-Channel

In this section we apply the tools presented previously to analyze the error rates in the Z-channel, one of the simplest channel models in information theory. In addition to its intrinsic interest, the Z-channel will prove of importance in the analysis of the additive energy channels.

A Z-channel has two possible inputs, say 0 and 1, two possible outputs, say α and β, and the following transition matrix

$$Q(\alpha|0) = 1, \qquad Q(\alpha|1) = \epsilon \tag{6.15}$$

$$Q(\beta|0) = 0, \qquad Q(\beta|1) = 1 - \epsilon, \tag{6.16}$$

where $0 \leq \epsilon \leq 1$. A diagram is depicted in Fig. 6.1. The name of the channel becomes apparent from inspecting the diagram.

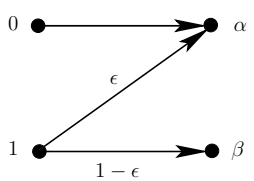


Figure 6.1: Transitions in the Z-channel.

We estimate the error rates of the Z-channel with randomized inputs: the input is mapped with probability 1/2 to either 0 or 1; this choice is known by the receiver. In general, this randomization may sacrifice performance with respect to an optimum code design. If the input is 0, the output is α, and the bit score is $\log \frac{Q(\alpha|1)}{Q(\alpha|0)} = \log \epsilon$. When the input is 1, the output is either α, in which case Ξ_b is $\log \frac{Q(\alpha|0)}{Q(\alpha|1)} = -\log \epsilon$, an event with probability ϵ, or the output is β, and the bit score is $\log \frac{Q(\beta|0)}{Q(\beta|1)} = -\infty$. Hence, the bit score takes the values

$$\xi = \begin{cases} -\infty, & \text{with probability } \frac{1}{2}(1-\epsilon), \\ \log \epsilon, & \text{with probability } \frac{1}{2}, \\ -\log \epsilon, & \text{with probability } \frac{1}{2}\epsilon. \end{cases} \tag{6.17}$$

The cumulant transform of Ξ_b is $\kappa_1(r) = \log \mathrm{E}[e^{r\Xi}] = \log\big(\frac{1}{2}\epsilon^r + \frac{1}{2}\epsilon^{1-r}\big)$. Note that $\kappa_1(0) = \log\big(\frac{1}{2}(1+\epsilon)\big) < 1$, since the random variable Ξ_b is defective, in the sense that a possible value it may take $(-\infty)$ is not a real number.

Using Proposition 6.3, the saddlepoint is located at $\hat{r} = 1/2$, which also can be verified by direct computation. The second derivative at the saddlepoint is readily found to be $\kappa_1''(\hat{r}) = \log^2 \epsilon$. We have the following

Theorem 6.7. *In the Z-channel, the bit score takes only two real values, namely* $\pm \log \epsilon$. *Its cumulant transform is*

$$\kappa_1(r) = \log\left(\frac{1}{2}\epsilon^r + \frac{1}{2}\epsilon^{1-r}\right). \tag{6.18}$$

At the saddlepoint, located at $\hat{r} = 1/2$, $\kappa_1(\hat{r}) = \frac{1}{2}\log\epsilon$ *and* $\kappa_1''(\hat{r}) = \log^2 \epsilon$. *For two codewords at Hamming distance* d, $\kappa_{pw}(r) = d\kappa_1(r)$.

The pairwise error probability in the Z-channel is given by

$$\mathrm{pep}(d) = \begin{cases} \dfrac{2}{\sqrt{2\pi d}}\left(\sum_{j=0}^{\frac{d-1}{2}} \epsilon^j\right)\epsilon^{(d+1)/2}, & d \text{ odd}, \\ \dfrac{2}{\sqrt{2\pi d}}\left(\frac{1}{2} + \sum_{j=1}^{\frac{d}{2}} \epsilon^j\right)\epsilon^{d/2}, & d \text{ even}. \end{cases} \tag{6.19}$$

Proof. The first part of the theorem, on the cumulant transform, is a restatement of the previous analysis. The approximation to the pairwise error probability is derived in Appendix 6.C. □

6.4 Error Probability in Binary-Input Gaussian Channels

6.4.1 Channel Model

We now consider BPSK modulation over the Gaussian channel. The set-up is as follows. A codeword $\mathbf{b}$, with n bits, $\mathbf{b} = (b_1, \ldots, b_n)$, is mapped onto a sequence of binary symbols x_k, with $x_k = \{-1, +1\}$; the mapping rule is

$$x_k = -1, \text{ if } \quad b_k = 0; \quad x_k = +1, \text{ if } \quad b_k = 1. \tag{6.20}$$

At the channel output, a complex-valued vector $\mathbf{y} = (y_1, \ldots, y_n)$ is detected. For each index k, $k = 1, \ldots, n$, the output y_k is given by the sum

$$y_k = \sqrt{\mathrm{SNR}}\, h_k\, x_k + z_k, \tag{6.21}$$

where z_k is Gaussian noise, $Z_k \sim \mathcal{N}_{\mathbf{C}}(0, 1)$, SNR is the average signal-to-noise ratio at the receiver, and h_k is a fading coefficient. We assume the fully-interleaved fading model of Section 4.2.5 on page 83. The squared fading

coefficient $\chi_k = |h_k|^2$ follows a gamma distribution

$$p_{\chi_k}(\chi) = \frac{m_f^{m_f} \chi^{m_f - 1}}{\Gamma(m_f)} e^{-m_f \chi}, \tag{6.22}$$

where m_f is a real non-negative number. The phase of h_k is assumed known at the receiver. Recall that unfaded AWGN is recovered by letting $m_f \to +\infty$ and Rayleigh fading by setting $m_f = 1$.

This is the standard model for BPSK modulation. In the absence of fading, it also models the transmission of QPSK signals; whenever there is fading and QPSK is used there is some residual correlation among the channel outputs through the fading coefficient h_k. This effect will be studied in Section 6.5.3.

6.4.2 Exact Pairwise Error Probability in the absence of Fading

In this section, we present some standard results for BPSK modulation. The material is presented in terms of the bit score in order to ease the generalization to other channels. With no fading, $h_k = 1$, and we have

Theorem 6.8. *In the binary-input AWGN channel the bit score has a normal distribution, $\Xi_b \sim \mathcal{N}(-4\,\mathrm{SNR}, 8\,\mathrm{SNR})$, independent of the value of the transmitted bit. The bit score is drawn randomly and independently at each channel realization and the pairwise score for two codewords at Hamming distance d also has a normal distribution, $\Xi_{pw} \sim \mathcal{N}(-4d\,\mathrm{SNR}, 8d\,\mathrm{SNR})$.*

The pairwise error probability is given by

$$\mathrm{pep}(d) = \Pr(\Xi_{pw} > 0) = Q(\sqrt{2d\,\mathrm{SNR}}), \tag{6.23}$$

where $Q(x)$ is the Gaussian tail function $Q(x) = (2\pi)^{-1/2} \int_x^{+\infty} e^{-t^2/2}\, dt$.

This formula is the well-known formula for the pairwise error probability for BPSK [32, 69]. It is exact, a trait not shared by the equivalent results we shall derive for more general channels. We derive it by using the bit score.

Proof. Assume symbol x is sent. The value of Ξ_b depends on the noise z as

$$\xi_b = \log \frac{\exp\bigl(-|\sqrt{\mathrm{SNR}}(x - x') + z|^2\bigr)}{\exp(-|z|^2)} = -4\,\mathrm{SNR} - 4\,\mathrm{Re}\bigl(\sqrt{\mathrm{SNR}} x z^*\bigr), \tag{6.24}$$

where we used the mapping rule in Eq. (6.20). Since the noise z is distributed as $\mathcal{N}_{\mathbf{C}}(0, 1)$, the bit score has a normal distribution $\mathcal{N}(-4\,\mathrm{SNR}, 8\,\mathrm{SNR})$. Further, the bit score does not depend on the value of b.

For the d values where b and b' differ the bit scores Ξ_b are drawn randomly and independently. Since the sum of d Gaussian variables is a Gaussian with mean and variance the sum of means and variances, so does the distribution of the pairwise score. Using Proposition 6.2, we have that

$$\mathrm{pep}(d) = \Pr\left(\sum_{j=1}^{d} \Xi_j > 0\right) = \frac{1}{\sqrt{16\pi d\,\mathrm{SNR}}} \int_0^{+\infty} e^{-\frac{(t+4d\,\mathrm{SNR})^2}{16d\,\mathrm{SNR}}}\, dt. \tag{6.25}$$

and Eq. (6.23) follows from the definition $Q(x) = (2\pi)^{-1/2} \int_x^{+\infty} e^{-t^2/2}\, dt$.

□

A well-known approximation to the $Q(\cdot)$ function [32, 70] is

$$Q(\sqrt{2d\,\mathrm{SNR}}) = \frac{e^{-d\,\mathrm{SNR}}}{2\sqrt{\pi d\,\mathrm{SNR}}}\left(1 - \frac{1}{2d\,\mathrm{SNR}} + \frac{1\cdot 3}{(2d\,\mathrm{SNR})^2} + \dots\right), \tag{6.26}$$

an equation used sometimes to extract information on the asymptotic behaviour of the error probability with increasing d or SNR. In the following section, we show that such an expression naturally appears when the exact error probability is computed by means of the saddlepoint approximation.

6.4.3 Pairwise Error Probability in Nakagami Fading

In general, the bit score Ξ_b depends on all the random elements in the channel. For fading channels we use that the random elements in the channel output y are the noise and the fading realizations, respectively z and h.

Theorem 6.9. *In the binary Nakagami-m_f fading channel, $\kappa_1(r)$ is given by*

$$\kappa_1(r) = -m_f \log\left(1 + \frac{4r\,\mathrm{SNR}}{m_f} - \frac{4r^2\,\mathrm{SNR}}{m_f}\right). \tag{6.27}$$

In first-order approximation, the pairwise error probability is

$$\mathrm{pep}(d, \mathrm{SNR}) \simeq \frac{1}{2\sqrt{\pi d\ \mathrm{SNR}}}\left(1 + \frac{\mathrm{SNR}}{m_f}\right)^{-m_f d + \frac{1}{2}}. \tag{6.28}$$

The second-order approximation is

$$\mathrm{pep}(d, \mathrm{SNR}) \simeq \frac{1}{2\sqrt{\pi d\ \mathrm{SNR}}}\left(1 + \frac{\mathrm{SNR}}{m_f}\right)^{-m_f d + \frac{1}{2}}\left(1 - \frac{1}{2d\,\mathrm{SNR}} - \frac{1}{8dm_f}\right). \tag{6.29}$$

The pairwise error probability is upper bounded by

$$\mathrm{pep}(d) \leq \left(1 + \frac{\mathrm{SNR}}{m_f}\right)^{-m_f d}. \tag{6.30}$$

In particular, for AWGN, when $m_f \to \infty$,

$$\mathrm{pep}(d) \leq e^{-d\,\mathrm{SNR}}. \tag{6.31}$$

Even though an exact form of the density of Ξ_b may be relatively difficult to obtain, the cumulant transform admits a rather simple derivation. As a complement, in Appendix 6.D we determine the exact bit error probability for uncoded transmission in Rayleigh channels by using the bit score.

Remark 6.10. For non-stationary channels, e. g. in the presence of memory across the fading coefficients h, the bit scores $\Xi_{b,j}$ are not independent and the appropriate joint density should be used to compute the total cumulant transform $\kappa_{\mathrm{pw}}(r)$, which is no longer $d\kappa_1(r)$. Furthermore, the code spectrum A_d in the union bound, Eq. (6.3), should be expanded to properly take into account the distribution of the Hamming weight d within the codeword.

Proof. Using the definition of the bit score, we write $\kappa_1(r)$ as

$$\kappa_1(r) = \log \mathrm{E}\,\mathrm{E}\big[e^{r\Xi_b} | H = h\big], \tag{6.32}$$

where the inner expectation is for a fixed fading realization. Conditioned on a realization of the fading h, Eq. (6.24) shows that Ξ_b is normally distributed, $\Xi_b \sim \mathcal{N}(-4\chi\,\mathrm{SNR}, 8\chi\,\mathrm{SNR})$, where $\chi = |h|^2$. As for the unfaded case, Ξ_b does not depend on the transmitted bit b. The cumulant transform of a Gaussian random variable of mean μ and variance σ^2 is $\mu r + \frac{1}{2}\sigma^2 r^2$ [32], which gives

$$\mathrm{E}\big[e^{r\Xi_b} | H = h\big] = \mathrm{E}\big[e^{r\Xi_b} | \chi\big] = e^{-4r\chi\,\mathrm{SNR} + 4r^2\chi\,\mathrm{SNR}}, \tag{6.33}$$

from which we deduce in turn that

$$\kappa_1(r) = \log \int_0^{+\infty} \frac{m_f^{m_f} \chi^{m_f - 1}}{\Gamma(m_f)} e^{-m_f \chi}\, e^{-4r\chi\,\mathrm{SNR} + 4r^2\chi\,\mathrm{SNR}}\, d\chi \tag{6.34}$$

$$= -m_f \log\left(1 + \frac{4r\,\mathrm{SNR}}{m_f} - \frac{4r^2\,\mathrm{SNR}}{m_f}\right). \tag{6.35}$$

The first derivatives at the saddlepoint, located at $\hat{r} = 1/2$, are given by

$$\kappa_1''(\hat{r}) = \frac{8\,\mathrm{SNR}}{1 + \frac{\mathrm{SNR}}{m_f}}, \qquad \kappa_1^{IV}(\hat{r}) = \frac{192\,\mathrm{SNR}^2}{m_f\left(1 + \frac{\mathrm{SNR}}{m_f}\right)^2}. \tag{6.36}$$

Theorem 6.5 gives the error probability estimates. For the second-order approximation, some straightforward computations give

$$\alpha = 1 - \frac{1}{2d\,\mathrm{SNR}} - \frac{1}{8dm_f}. \tag{6.37}$$

The Chernoff bound is a restatement of Proposition 6.4. □

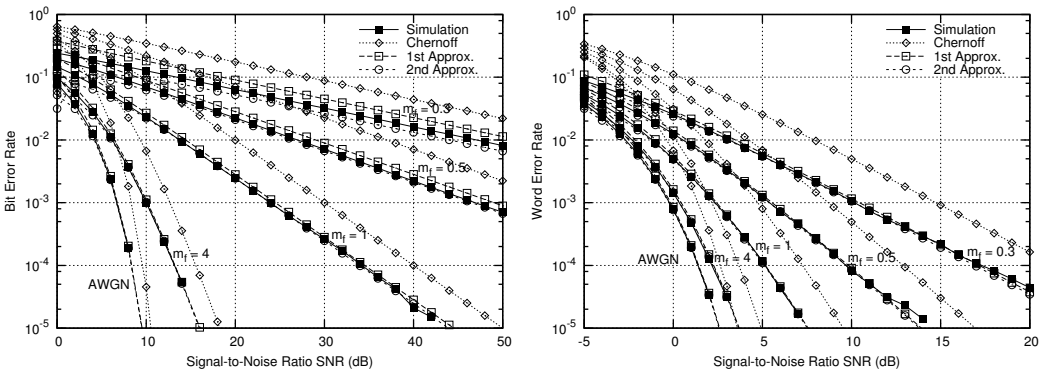
For Rayleigh fading ($m_f = 1$), Eqs. (6.28) and (6.29) constitute a generalization of the well-known Chernoff bound in Eq. (6.30), namely

$$\mathrm{pep}(d) \leq (1 + \mathrm{SNR})^{-d}. \tag{6.38}$$

Further, for $m_f \to \infty$, i. e. for the AWGN channel, the approximation gives the first two terms in the classical expansion of the $Q(\cdot)$ function, see Eq. (6.26). This suggests that the saddlepoint approximation generalizes the classical expansion of the $Q(\cdot)$ function from AWGN to fading channels.

We next show some numerical results that illustrate the accuracy of the proposed method as well as its asymptotic behavior. Figure 6.2 compares the error probability simulation with the saddlepoint approximations for BPSK in Nakagami fading with parameter $m_f = 0.3, 0.5, 1$ and 4. In particular, we show the following: the Chernoff bound Eq. (6.30), the saddlepoint approximation, both in first-order Eq. (6.28) and second-order, Eq. (6.29), and the simulation of the bit-error rate. The curves for simulation and saddlepoint approximation are very close, especially when the second order term is included. The second-order effect is noticeable only for low values of m_f. Figure 6.2a shows the results of uncoded transmission and Fig. 6.2b the word error rate for Hamming distance $d = 5$. Results for convolutional and turbo-like codes may be found in [71].

The saddlepoint approximation gives an accurate result at a fraction of the complexity required by alternative computation methods [72], such as the (exact) formula for the uncoded case (a Gauss hypergeometric function), or numerical integration of Craig's form of the $Q(\cdot)$ function. Furthermore, as opposed to the numerical integration, the saddlepoint approximation is useful in an engineering sense, as it highlights the individual role of all the relevant



(a) Uncoded Bit Error Rate ($d = 1$). (b) Word Error Rate ($d = 5$).

Figure 6.2: Comparison of simulation, Chernoff bound, and saddlepoint approximation to pep(d) for BPSK in Nakagami fading of parameter $m_f = 0.3, 0.5, 1, 4, \infty$.

variables (SNR, m_f, d) in the error probability. Finally, we note that the saddlepoint method is an approximation to direct integration in the complex plane, used in [73] to exactly compute the error probability.

6.5 Pairwise Error Probability for Bit-Interleaved Coded Modulation in Gaussian Noise

6.5.1 Channel Model

The concept of bit-interleaved coded modulation (BICM), a pragmatic coding scheme for non-binary modulations, was discussed in Section 4.3 in the context of Gaussian channels. Very briefly, binary codewords $\mathbf{b} = (b_1, \ldots, b_l)$ are mapped onto an array of channel symbols $\mathbf{x} = (x_1, \ldots, x_n)$ by bit-interleaving the binary codeword and mapping it on the signal constellation $\mathcal{X}$ with a binary labelling rule μ, as in Eq. (6.1). With no loss of generality, the constellation set is assumed to have 2^m elements, so that m bits are necessary to index one symbol. The binary decoder uses the metric function in Eqs. (6.4) and (6.5). We assume that interleaving is infinite except in Section 6.5.3, where we exploit the results in [68] to analyze QPSK modulation in fading channels.

The error probabilities of BICM in Gaussian channels were analyzed in [26].

However, the results reported there were either not tight or exceedingly complex to compute. In this section, we provide a simple, yet accurate, estimate of the pairwise error probability by making use of a saddlepoint approximation to the tail probability of the pairwise score. This work was partly published in [66].

6.5.2 An Estimate of the Error Probability

As in the binary case, the received signal y_k at time instant k, for $k = 1, \ldots, n$, is expressed as $y_k = \sqrt{\mathrm{SNR}}\, h_k\, x_k + z_k$, where h_k is a fading attenuation, z_k a sample of circularly-symmetric Gaussian noise of unit variance, and x_k the transmitted signal, chosen from a set with unit energy; the average received signal-to-noise ratio is SNR. All the variables are complex-valued. We consider Nakagami-m_f fading channels, as described in Section 6.4.1.

Each input bit b_j is matched with a bit score $\xi_{b,j}$, whose sample value depends on all the random elements in the channel: 1) the bit value b_j; 2) the bit position j, randomly mapped to any set of the possible label indices i, $i = 1, \ldots, m$; 3) the transmitted symbol x_k, drawn from the set $\mathcal{X}_i^b$, which contains all symbols with bit b at position i; and 4) the noise and fading realizations z_k and h_k. For convenience, we group all these random elements in a 5-tuple (B, I, X, H, Z). Using Definition 6.1 and Theorem 6.5 we have

Theorem 6.11. *For BICM with infinite interleaving in the AWGN fading channel, the pairwise error probability* $\mathrm{pep}(d)$ *admits the approximation*

$$\mathrm{pep}(d) \simeq \frac{2}{\sqrt{2\pi d}\,\mathrm{E}^{\frac{1}{2}}[\sqrt{\gamma}\log^2\gamma]}\ \mathrm{E}^{d+\frac{1}{2}}[\sqrt{\gamma}], \tag{6.39}$$

where γ is the following function of the 5-tuple (B, I, X, H, Z),

$$\gamma(b, i, x, h, z) = \frac{q_i(\bar{b}, hx + z)}{q_i(b, hx + z)} = \frac{\sum_{x' \in \mathcal{X}_i^{\bar{b}}} e^{-|h(x-x')\sqrt{\mathrm{SNR}}+z|^2}}{\sum_{x' \in \mathcal{X}_i^{b}} e^{-|h(x-x')\sqrt{\mathrm{SNR}}+z|^2}}. \tag{6.40}$$

The expectation is carried out over all possible values of (B, I, X, H, Z).

The explicit form of the expectations is

$$\mathrm{E}[\gamma^r] = \frac{1}{m2^m} \sum_{b\in\{0,1\}} \sum_{i=1}^{m} \sum_{x\in\mathcal{X}_i^b} \iint p_H(h) p_Z(z)\, \gamma^r\, dz\, dh. \tag{6.41}$$

This expectation can be easily evaluated by numerical integration using the appropriate quadrature rules. Unfortunately, there seems to be no tractable, simple expression for the final result.

Proof. From Definition 6.1 we extract the function γ in Eq. (6.40). Proposition 6.3 shows that the saddlepoint is located at $\hat{r} = 1/2$. As computed in Proposition 6.3, $\kappa_1''(\hat{r})$ is given by

$$\kappa_1''(\hat{r}) = \frac{\mathrm{E}[\Xi_b^2 e^{\hat{r}\Xi_b}]}{\mathrm{E}[e^{\hat{r}\Xi_b}]} = \frac{\mathrm{E}[\sqrt{\gamma}\log^2\gamma]}{\mathrm{E}\big[\sqrt{\gamma}\big]}. \tag{6.42}$$

□

We next compare the saddlepoint approximation with the alternative methods given in [26] to compute the pairwise error probability $\mathrm{pep}(d)$ for BICM: the Bhattacharyya union bound and the expurgated BICM union bound. Clearly, the Bhattacharyya union bound is simply the Chernoff bound in Proposition 6.4 using the definitions in Theorem 6.11, namely $\mathrm{pep}(d) \leq e^{d\kappa_1(\hat{r})} = \mathrm{E}^d[\sqrt{\gamma}]$.

Let $\mathrm{pep}_{\mathrm{ex}}(d)$ denote the pairwise error probability in the expurgated bound. We start the analysis with Eqs. (48-49) in [26] – p. 938 –, which read

$$\mathrm{pep}_{\mathrm{ex}}(d) = \frac{1}{2\pi j}\int_{\hat{r}-j\infty}^{\hat{r}+j\infty} \psi_{\mathrm{ex}}(r)^d \,\frac{dr}{r}, \tag{6.43}$$

where

$$\psi_{\mathrm{ex}}(r) = \frac{1}{m2^m}\sum_{i=1}^{m}\sum_{b=0}^{1}\sum_{x\in\mathcal{X}_i^b} \mathrm{E}\big[e^{r\Delta(x,\hat{x})}\big], \tag{6.44}$$

and the function $\Delta(x, x')$ had been defined in (their) Eq. (33) – p. 936 – as

$$\Delta(x,\hat{x}) = \log\frac{Q(y|\hat{x})}{Q(y|x)}. \tag{6.45}$$

Here x is the transmitted symbol, and $\hat{x}$ its nearest neighbour in $\mathcal{X}_i^{\bar{b}}$, i. e., with complementary bit $\bar{b}$ in label index i, and $Q(y|x)$ the channel transition matrix.

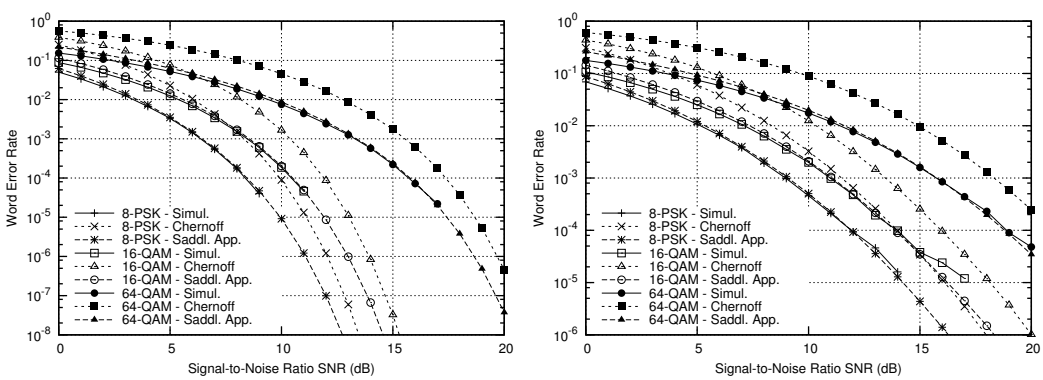
These equations are very close to our analysis. First, the saddlepoint approximation is an efficient method to carry out the integration in the complex plane in Eq. (6.43), by choosing $\hat{r}$ to be the saddlepoint. Then, the expectation in Eq. (6.44) coincides with the expectation used in Theorem 6.11 and the

cumulant transform of the bit score is (almost) $\kappa_1(r) = \log \psi_{\text{ex}}(r)$. It is not completely so because the role of $\Delta(x, \hat{x})$ is played in our analysis by

$$\log \frac{\sum_{x' \in \mathcal{X}_i^{\bar{b}}} Q(y|x')}{\sum_{x' \in \mathcal{X}_i^{b}} Q(y|x')}. \tag{6.46}$$

Caire's equation (6.45) can be obtained by taking only one term in each summation, respectively x and $\hat{x}$. From Definition 6.1 we see that this is equivalent to changing the decoder metric $q_i(b, y)$. Depending the specific metric and mapping rule Caire's approximation may be accurate or not. For example, in the simulation results for the set-partitioning mapping reported in [26], the union bound was not close to the simulation. This inaccuracy was solved in the simulations in [66] by using Eq. (6.46). To any extent, the added complexity from taking all terms in Eq. (6.46) is negligible.

Figure 6.3 depicts the pairwise error probability between two codewords at Hamming distance $d = 5$ for several modulations and channel models. The modulations are 8-PSK, 16-QAM and 64-QAM, all with Gray mapping, and the channel models AWGN and Rayleigh fading. In all cases the saddlepoint approximation shows an excellent match with the simulation results. Results for convolutional and turbo-like code ensembles may be found in [66].



(a) AWGN channel. (b) Rayleigh fading channel.

Figure 6.3: Comparison of simulation, Chernoff bound, and saddlepoint approximation to pep(d) for $d = 5$ over 8-PSK, 16-QAM, and 64-QAM.

6.5.3 On the Memory among the Channel Outputs

We have assumed in our analysis, as it was done in [26], that bit interleaving is infinite and that the l BICM sub-channels from bit b_j to log-likelihood ratio λ_j are independent, and so are the corresponding bit scores. For finite interleaving, some of the d bits in which the two codewords differ may belong to the same symbol and therefore suffer from the same noise, and some residual statistical dependence among the bit scores $\Xi_{b,j}$ appears, as we mentioned in Section 6.3. In this section we study what arguably constitutes the simplest case, namely QPSK with Gray mapping under Nakagami fading.

In general, the d bits fall into a number of symbols, each of them having between 1 and m bits. Since it appears almost impossible to know how these bits are distributed onto the n symbols, it seems more appropriate to compute an average pairwise error probability, by averaging over all possible distributions of d bits onto n symbols. Following [68], let us define $w = \min(m, d)$ and denote a possible pattern by $\pi_\ell = (\ell_0, \ldots, \ell_w)$, where ℓ_v is the number of symbols with v bits. Clearly, $d = \sum_{v=1}^{w} v\ell_v$ and $\ell_0 = n - \sum_{v=1}^{w} \ell_v$. A counting argument gives the probability of the pattern π_ℓ as

$$P\big(\pi_\ell = (\ell_0, \ldots, \ell_w)\big) = \frac{\binom{m}{1}^{\ell_1}\binom{m}{2}^{\ell_2}\cdots\binom{m}{w}^{\ell_w}}{\binom{mn}{d}} \frac{n!}{\ell_0!\ell_1!\ell_2!\cdots\ell_w!} \tag{6.47}$$

At this point, it would convenient to define the cumulant transforms of generic symbol scores Ξ_s for a given Hamming weight in the symbol. Then, $\kappa_v(r)$ is such cumulant transform for a Hamming weight v. Definition 6.1 corresponds to $v = 1$ and is thus consistent with this idea. Since we limit ourselves to QPSK, we postpone the definition of $\kappa_2(r)$. In Theorem 6.5 we assumed that the cumulant transform of the pairwise score is $\kappa_{\mathrm{pw}}(r)$ is given by $d\kappa_1(r)$. In general, and for a given pattern, one would have that

$$\kappa_{\mathrm{pw}}(r, \pi_\ell) = \sum_{v=1}^{w} \ell_v \kappa_v(r), \tag{6.48}$$

since the symbol scores are assumed to be independent. A theorem analogous to Theorem 6.5 can then be formulated. Moreover, the average over all possible patterns π_ℓ of the conditional pairwise error probability gives the average pairwise error probability.

When $l \to \infty$ and $m > 1$ the probability that the bit scores are dependent tends to zero, but does so relatively slowly, as proved in the following

Proposition 6.12. *The probability that all d bits are independent is* $\Pr_{ind}$

$$\Pr_{ind} = \frac{m^d \binom{l/m}{d}}{\binom{l}{d}}. \tag{6.49}$$

For large l, the probability of $\pi_{all\ one} = (1, \dots, 1)$ is

$$\Pr_{ind} \simeq e^{-\frac{d(d-1)(m-1)}{2l}} \simeq 1 - \frac{d(d-1)(m-1)}{2l}. \tag{6.50}$$

Proof. We use that $\ell_1 = d$, $\ell_v = 0$ for $v > 1$, and $\ell_0 = n - d$. The derivation of the approximation can be found in Appendix 6.E. □

For BPSK, or $m = 1$, there is no dependence, as it should. We next have

Proposition 6.13. *In the Nakagami fading channel with QPSK modulation and Gray mapping, the cumulant transform of the symbol score when two bits belong to the same symbol, denoted by $\kappa_2(r)$, is given by*

$$\kappa_2(r) = -m_f \log\left(1 + \frac{4r\,\mathrm{SNR}}{m_f} - \frac{4r^2\,\mathrm{SNR}}{m_f}\right). \tag{6.51}$$

Proof. Separating the real and imaginary parts, and respectively denoting them by 1 and 2, the channel has an equivalent output

$$y_1 = \sqrt{\mathrm{SNR}}\,\mathrm{Re}(x)h + \mathrm{Re}(z), \quad y_2 = \sqrt{\mathrm{SNR}}\,\mathrm{Im}(x)h + \mathrm{Im}(z). \tag{6.52}$$

Conditioned on h, Eq. (6.24) shows that, say, $\Xi_{b,1}$ takes the value

$$\xi_{b,1} = -\,\mathrm{SNR}\,\chi\big(\mathrm{Re}(x) - \mathrm{Re}(x')\big)^2 - 2\sqrt{\mathrm{SNR}}\sqrt{\chi}\,\mathrm{Re}(x - x')\,\mathrm{Re}(z), \tag{6.53}$$

that is $\Xi_{b,1} \sim \mathcal{N}(-2\chi\,\mathrm{SNR}, 4\chi\,\mathrm{SNR})$. The same analysis gives identical distribution for $\Xi_{b,2}$. Eq. (6.55) then follows from the binary case in Theorem 6.9. When $\chi = |h|^2$ is the same value for both 1,2, then

$$\kappa_2(r) = \log \mathrm{E}\,\mathrm{E}\big[e^{r\Xi_{b,1} + r\Xi_{b,2}}|\chi\big] = \log \mathrm{E}[e^{-2(2r\chi\,\mathrm{SNR} + 2r^2\chi\,\mathrm{SNR})}], \tag{6.54}$$

from which we proceed as in the binary case, in Theorem 6.9. □

When the bits belong to different symbols, the same proof shows that an individual bit behaves as BPSK with signal-to-noise ratio $\frac{1}{2}\,\mathrm{SNR}$. Therefore, the cumulant transform of one such bit score is given by

$$\kappa_1(r) = -m_f \log\left(1 + \frac{2r\,\mathrm{SNR}}{m_f} - \frac{2r^2\,\mathrm{SNR}}{m_f}\right). \tag{6.55}$$

For this case, we can index all possible partitions by considering ℓ_2, the number of symbols with 2 bits. Clearly, this number can take the values $\ell_2 = \max(0, d - \lceil l/2 \rceil), \ldots, \lfloor d/2 \rfloor$. Applying Eq. (6.47), we obtain

$$P(\ell_2) = \frac{2^{\ell_1}}{\binom{l}{d}} \frac{n!}{(n-\ell_1-\ell_2)!\ell_1!\ell_2!}. \tag{6.56}$$

Finally, combining this formula with Proposition 6.13, we average over all possible mappings of the d bits onto the $l/2$ QPSK symbols and derive

Theorem 6.14. *The average pairwise error probability* $\mathrm{pep}(d)$ *of QPSK with codeword length* l *in Nakagami fading is*

$$\mathrm{pep}(d) \simeq \sum_{\ell_2} \frac{2^{\ell_1}}{\binom{l}{d}} \frac{n!}{(n-\ell_1-\ell_2)!\ell_1!\ell_2!} \frac{\left(1+\frac{\mathrm{SNR}}{2m_f}\right)^{-m_f\ell_1}\left(1+\frac{\mathrm{SNR}}{m_f}\right)^{-m_f\ell_2}}{\sqrt{2\pi\left(\ell_1\frac{\mathrm{SNR}}{1+\frac{\mathrm{SNR}}{2m_f}}+\ell_2\frac{2\,\mathrm{SNR}}{1+\frac{\mathrm{SNR}}{m_f}}\right)}}, \tag{6.57}$$

where ℓ_2 *is limited to* $\max(0, d - \lceil l/2 \rceil) \leq \ell_2 \leq \lfloor d/2 \rfloor$, $\ell_1 + 2\ell_2 = d$.

The (averaged) Chernoff bound is

$$\mathrm{pep}(d) \leq \sum_{\ell_2} \frac{2^{\ell_1}}{\binom{l}{d}} \frac{n!}{(n-\ell_1-\ell_2)!\ell_1!\ell_2!} \left(1+\frac{\mathrm{SNR}}{2m_f}\right)^{-m_f\ell_1}\left(1+\frac{\mathrm{SNR}}{m_f}\right)^{-m_f\ell_2}. \tag{6.58}$$

Proof. By construction, the pairwise score is the sum of ℓ_1 symbols with Hamming weight 1 and ℓ_2 symbols with Hamming weight 2, with respective cumulant transforms given by Eqs. (6.55) and (6.51). The conditional transform at the saddlepoint is therefore given by

$$\kappa_{\mathrm{pw}}(\hat{r}) = -m_f\ell_1 \log\left(1+\frac{\mathrm{SNR}}{2m_f}\right) - m_f\ell_2 \log\left(1+\frac{\mathrm{SNR}}{m_f}\right). \tag{6.59}$$

The second derivative can be recovered from Theorem 6.9,

$$\kappa''_{\mathrm{pw}}(\hat{r}) = \ell_1 \frac{4\,\mathrm{SNR}}{1+\frac{\mathrm{SNR}}{2m_f}} + \ell_2 \frac{8\,\mathrm{SNR}}{1+\frac{\mathrm{SNR}}{m_f}}. \tag{6.60}$$

□

Figure 6.4 depicts the pairwise error probabilities of Eq. (6.57) for $d = 3$ and $d = 6$ in Nakagami fading with $m_f = 0.5$ and $m_f = 1$. Two short interleavers

are considered, $l = 10$ and $l = 200$. In all cases the effect from the correlation, appearing as a knee in the curves takes place at very low values of pairwise error probability, easily below 10^{-6}. Further, the transition takes place at ever lower values by increasing the length l or the distance d. In a practical scenario, the effect due to the correlation among the log-likelihood ratios is unlikely to determine the error probabilities, being other variables, for instance the perfect channel estimation, which possibly determine the error rates.

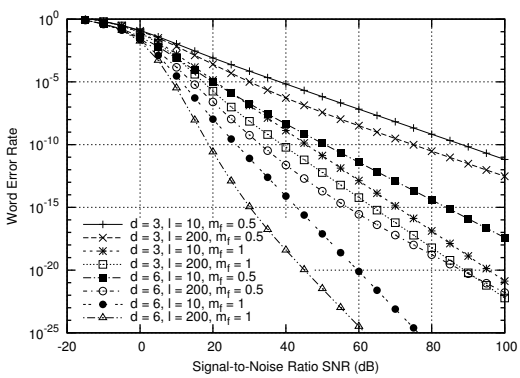


Figure 6.4: Saddlepoint approximation pep(d) for QPSK in Nakagami fading $m_f = 0.5, 1$; Hamming distance $d = 3, 6$, and codeword length $l = 10, 200$.

Going back to the general BICM case, the apparent closeness of the BICM probability analysis with the simulation results is likely due to the limited effect of the dependence across the log-likelihood ratios, compared to a situation with no statistical independence (infinite interleaving). It would be interesting to characterize this effect quantitatively and find the threshold signal-to-noise ratio at which the dependence starts being noticeable.

A second conclusion one can draw is that for finite interleaving and high-enough signal-to-noise ratio, the error probability will be determined by the worst possible π_ℓ, that having the largest tail for the log-likelihood ratio. This implies that any analysis based on infinite interleaving fails for large enough SNR. A caveat is in order, since this large enough may be extremely large.

6.5.4 Cumulant transform asymptotic analysis

An interesting result of [66], proved by Albert Guillén i Fàbregas, concerns the limit for large SNR of BICM. For the AWGN channel with no fading BICM behaves as a binary modulation, in the sense that

$$\lim_{\mathrm{SNR}\to\infty} \frac{\kappa_1(\hat{r})}{\mathrm{SNR}} = -\frac{d_{\min}^2}{4}, \tag{6.61}$$

where $d_{\min}^2 = \min_{x,x'\in\mathcal{X}} |x - x'|^2$ is the minimum squared Euclidean distance of the constellation. For some standard modulations, the minimum distance is given by (see for instance Table I of [74])

$$d_{\min}^2 = 4\sin^2\frac{\pi}{2^m} \qquad \text{for } 2^m\text{-PSK}, \tag{6.62}$$

$$d_{\min}^2 = \frac{12}{2^{2m}-1} \qquad \text{for } 2^m\text{-PAM}, \tag{6.63}$$

$$d_{\min}^2 = \frac{6}{(2^m-1)} \qquad \text{for } 2^m\text{-QAM}. \tag{6.64}$$

Note that the mapping does not affect the result.

The loss with respect to binary modulation ($d_{\min}^2 = 4$) for 8-PSK amounts to 8.34 dB, for 16-QAM is 10 dB, and for 64-QAM becomes 16.23 dB. At low SNR, these values match well with the simulations reported in Fig. 6.3, when compared with the binary case included in Fig. 6.2b.

We will later see that a similar result holds for the additive energy channels.

6.6 Error Probability in the Exponential Noise Channel

In this section, we study the pairwise error probability in the additive exponential noise (AEN) channel. As for the Gaussian channel, we build our analysis around the pairwise and bit scores. The results presented here seem to be new and are somewhat surprising, in the sense that the pairwise error probability is very close to that of an AWGN channel with identical signal-to-noise ratio:

- The Chernoff bound to the pairwise error probability for BPSK and binary modulation in the AEN channel coincide.
- For BICM at large signal-to-noise ratio SNR, and comparing the Chernoff bounds to the pairwise error probability, 2^{2m}-PEM in AEN incurs in a power loss of $\frac{3}{2}$ (about 1.76 dB) with respect to 2^{2m}-QAM in AWGN.

This similarity strongly suggests that binary codes will achieve essentially the same performance in these two channels.

6.6.1 Channel Model for Binary Modulation

We start with the study of binary modulation (2-PEM). At the transmitter, each bit b_k is mapped onto a binary symbol x_k, with $x_k = \{0, +2\}$; the mapping rule μ is

$$x_k = 0, \text{ if } \quad b_k = 0; \quad x_k = +2, \text{ if } \quad b_k = 1, \tag{6.65}$$

used with probability 1/2, and its complement $\bar{\mu}$, used with probability 1/2. The choice between μ and $\bar{\mu}$ is made known to the receiver.

At the channel output, and for each index k, the output y_k is given by

$$y_k = \mathrm{SNR}\, x_k \chi_k + z_k, \tag{6.66}$$

where z_k is a real-valued, exponentially distributed, noise $Z_k \sim \mathcal{E}(1)$, SNR is the average signal-to-noise ratio at the receiver, and χ_k is a fading coefficient having a gamma distribution, as in Eq. (6.22). For a fixed fading realization, assumed known at the receiver, the channel transition probability $Q(y|x)$ is

$$Q(y|x) = e^{-(y - \mathrm{SNR}\, x\chi)} u(y - \mathrm{SNR}\, x\chi). \tag{6.67}$$

The presence of the fading is a natural generalization of the exponential channel we have considered so far and fits smoothly with its Gaussian counterpart. In the additive energy channels, we use the name "gamma fading" rather than Nakagami fading, since the latter is usually expressed in terms of the complex amplitude, rather than the energy.

6.6.2 Error Probability in the absence of Fading

We first consider the case without fading, obtained by setting $\chi_k = 1$ in the model, Eq. (6.66). The natural counterpart of the Gaussian result is

Theorem 6.15. *The binary AEN channel is a Z-channel with transition probability $\epsilon = e^{-2\,\mathrm{SNR}}$. The cumulant transform of the bit score is*

$$\kappa_1(r) = \log\left(\frac{1}{2} e^{-2r\,\mathrm{SNR}} + \frac{1}{2} e^{-2(1-r)\,\mathrm{SNR}}\right) \tag{6.68}$$

$$= -\,\mathrm{SNR} + \log\cosh\big(\mathrm{SNR}(2r-1)\big), \tag{6.69}$$

and the pairwise error probability for Hamming distance d is

$$\mathrm{pep}(d) \simeq \begin{cases} \dfrac{2}{\sqrt{2\pi d}}\left(\sum_{j=0}^{\frac{d-1}{2}} e^{-2j\,\mathrm{SNR}}\right)e^{-(d+1)\,\mathrm{SNR}}, & d \text{ odd}, \\ \dfrac{2}{\sqrt{2\pi d}}\left(\frac{1}{2}+\sum_{j=1}^{\frac{d}{2}} e^{-2j\,\mathrm{SNR}}\right)\epsilon^{-d\,\mathrm{SNR}}, & d \text{ even} \end{cases}. \tag{6.70}$$

The Chernoff bound to the pairwise error probability is

$$\mathrm{pep}(d) \le e^{-d\,\mathrm{SNR}}. \tag{6.71}$$

Proof. Let symbol x be transmitted and $\bar{x}$ denote its complement. The bit score ξ_b depends on the noise realization z as

$$\xi_b = \log \frac{e^{-(\mathrm{SNR}\,x+z-\mathrm{SNR}\,\bar{x})}u(\mathrm{SNR}\,x+z-\mathrm{SNR}\,\bar{x})}{e^{-(\mathrm{SNR}\,x+z-\mathrm{SNR}\,x)}u(\mathrm{SNR}\,x+z-\mathrm{SNR}\,x)} \tag{6.72}$$

$$= \log\big(e^{-\,\mathrm{SNR}(x-\bar{x})}u(\mathrm{SNR}(x-\bar{x})+z)\big). \tag{6.73}$$

There are two possibilities, $u(\mathrm{SNR}(x-\bar{x})+z)=0$ or $u(\mathrm{SNR}(x-\bar{x})+z)=1$. The first takes place when $\mathrm{SNR}(x-\bar{x})+z<0$ and occurs with a probability

$$\int_0^{\mathrm{SNR}(\bar{x}-x)} e^{-t}\,dt = 1-e^{-\,\mathrm{SNR}(\bar{x}-x)}. \tag{6.74}$$

In this case $\xi_b=-\infty$. In the second case, $\mathrm{SNR}(x-\bar{x})+z\ge 0$, an event with probability $e^{-\,\mathrm{SNR}(\bar{x}-x)}$, and inducing a bit score $\xi_b=-\,\mathrm{SNR}(x-\bar{x})$.

Using the randomized mapping, there are three possible values of ξ_b,

$$\xi_b = \begin{cases} -\infty, & \text{with probability } \frac{1}{2}(1-e^{-2\,\mathrm{SNR}}), \\ -2\,\mathrm{SNR}, & \text{with probability } \frac{1}{2}, \\ 2\,\mathrm{SNR}, & \text{with probability } \frac{1}{2}e^{-2\,\mathrm{SNR}}. \end{cases} \tag{6.75}$$

This corresponds to a Z-channel with transition probability $\epsilon = e^{-2\,\mathrm{SNR}}$. The rest of the theorem is an application of Theorem 6.7. □

Remark that the use of the continuous saddlepoint approximation, given in Theorem 6.5 would lead to inexact results.

Since the Chernoff bound coincides with the value of BPSK in the AWGN channel, given in Eq. (6.31), the error performance of both modulation formats would be similar for identical levels of signal-to-noise ratio.

Also, the error probability decays as $e^{-(d+1)\,\mathrm{SNR}}$ when d is odd, slightly faster than in the Chernoff bound. This effect may create some room to design efficient codes for this channel, since the error probability for d odd is similar to the error probability with $d+1$, buying, so to speak, some Hamming distance from the channel itself. A similar effect was noticed by van de Meeberg in the analysis of binary symmetric channels [75]. A special case of d odd is uncoded transmission, for which $d = 1$, and the exact bit error rate is

$$P_b = \Pr(\Xi > 0) = \frac{1}{2} e^{-2\,\mathrm{SNR}}. \tag{6.76}$$

This value should be compared with the saddlepoint approximation, $P_b \simeq \frac{2}{\sqrt{2\pi}} e^{-2\,\mathrm{SNR}} \simeq 0.798 e^{-2\,\mathrm{SNR}}$, a reasonably good approximation.

Figure 6.5 depicts the word error rate with 2-PEM for several values of d. Simulations match well with the approximation in Theorem 6.15. The Chernoff bound does not give the correct dependence with d for odd d, especially for $d = 1$; this problem does not arise for the saddlepoint approximation.

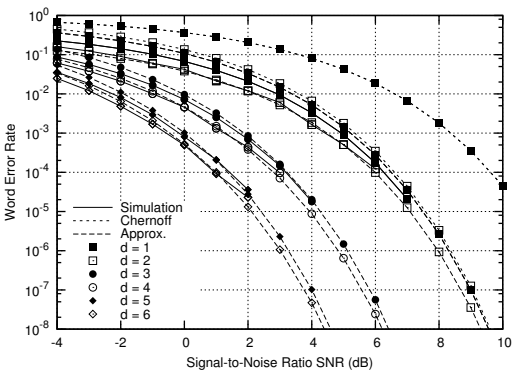


Figure 6.5: Comparison of simulation and saddlepoint approximation to pep(d) of 2-PEM in the AEN channel, $d = 1, \ldots, 6$.

6.6.3 Error Probability with Gamma Fading

We complete our analysis of the AEN channel by considering the effect of gamma fading. In this case, it is possible to compute an exact expression for the uncoded error probability. Conditioned to a fading realization, the error

probability is given by

$$\frac{1}{2}e^{-2\chi\,\mathrm{SNR}}, \tag{6.77}$$

and after averaging over the fading realizations,

$$P_b = \int_0^{+\infty} \frac{m_f^{m_f}\chi^{m_f-1}}{\Gamma(m_f)} e^{-m_f\chi}\frac{1}{2}e^{-2\chi\,\mathrm{SNR}}\,d\chi = \frac{1}{2}\left(1+\frac{2\,\mathrm{SNR}}{m_f}\right)^{-m_f}, \tag{6.78}$$

obtained from the definition of the gamma function, see Eq. (3.34). This nice formula does not seem to extend easily to the pairwise error probability. However, we have

Theorem 6.16. *In the binary AEN channel with gamma fading, the cumulant transform of the bit score Ξ_b is given by*

$$\kappa_1(r) = \log\left(\frac{1}{2}\left(1+\frac{2r\,\mathrm{SNR}}{m_f}\right)^{-m_f} + \frac{1}{2}\left(1+\frac{2(1-r)\,\mathrm{SNR}}{m_f}\right)^{-m_f}\right). \tag{6.79}$$

The Chernoff bound to the pairwise error probability is

$$\mathrm{pep}(d) \le \left(1+\frac{\mathrm{SNR}}{m_f}\right)^{-m_f d}. \tag{6.80}$$

The saddlepoint approximation to the pairwise error probability is

$$\mathrm{pep}(d) \simeq \begin{cases} \frac{2}{\sqrt{2\pi d}}\left(\sum_{j=0}^{\frac{d-1}{2}}\left(1+\frac{(d+1+2j)\,\mathrm{SNR}}{m_f}\right)^{-m_f}\right), & d \text{ odd}, \\ \frac{2}{\sqrt{2\pi d}}\left(\frac{1}{2}\left(1+\frac{d\,\mathrm{SNR}}{m_f}\right)^{-m_f} + \sum_{j=1}^{\frac{d}{2}}\left(1+\frac{(d+2j)\,\mathrm{SNR}}{m_f}\right)^{-m_f}\right), & d \text{ even} \end{cases}. \tag{6.81}$$

Proof. Conditioned on a fading realization χ, it is clear from the proof of Theorem 6.15 that Ξ_b has the discrete distribution of a Z-channel with transition probability $\epsilon = e^{-2\chi\,\mathrm{SNR}}$. Using the definition of Ξ_b, we rewrite $\kappa_1(r)$ as

$$\kappa_1(r) = \log \mathrm{E}\,\mathrm{E}\big[e^{r\Xi}|\chi\big], \tag{6.82}$$

where the inner expectation is for a fixed fading realization. The moment generating function, $e^{\kappa_1(r)}$, is derived from Theorem 6.15,

$$\mathrm{E}\big[e^{s\Xi_b}|\chi\big] = \frac{1}{2}e^{-2s\chi\,\mathrm{SNR}} + \frac{1}{2}e^{-2(1-s)\chi\,\mathrm{SNR}}, \tag{6.83}$$

from which we deduce that

$$\kappa_1(r) = \log \frac{1}{2} \int_0^{+\infty} \frac{m_f^{m_f} \chi^{m_f - 1}}{\Gamma(m_f)} e^{-m_f \chi} \left(e^{-2r\chi \,\mathrm{SNR}} + e^{-2(1-r)\chi \,\mathrm{SNR}} \right) d\chi \quad (6.84)$$

$$= \log \left(\frac{1}{2} \left(1 + \frac{2r \,\mathrm{SNR}}{m_f} \right)^{-m_f} + \frac{1}{2} \left(1 + \frac{2(1-r) \,\mathrm{SNR}}{m_f} \right)^{-m_f} \right). \quad (6.85)$$

From Proposition 6.3, the saddlepoint is located at $\hat{r} = 1/2$, and

$$\kappa_1(\hat{r}) = -m_f \log \left(1 + \frac{\mathrm{SNR}}{m_f} \right). \quad (6.86)$$

The Chernoff bound is then obtain by using Proposition 6.4.

The saddlepoint approximation follows from integrating the approximation in the absence of fading, given in Theorem 6.15, and using the definition of the Gamma function. □

In practice, the sums over j in Eq. (6.81) are dominated by the first term, and good approximations would be obtained by using only them.

In this result, we have not used a direct saddlepoint approximation to the tail probability of Ξ_{pw}, such as that of Theorem 6.5. The reason is that the latter approximation is not tight for large values of m_f, that is small fading, when the score becomes ever more "discrete". To any extent, the equation in the absence of fading can be integrated term by term and the final formula gives a good approximation, as can be seen from Fig. 6.6, where simulation and the approximation for uncoded 2-PEM in gamma fading of parameter $m_f = 0.3, 0.5, 1, 4, \infty$ are compared. Similarly good match is observed for the pairwise error probability $\mathrm{pep}(d)$.

6.6.4 Bit-Interleaved Coded Modulation

In Section 6.5, we discussed bit-interleaved modulation (BICM) for Gaussian noise and gave a saddlepoint approximation to its pairwise error probability. For BICM, the channel model in Eq. (6.66) is still valid, with the addition that the symbols x_k are not directly indexed by the mapping rule in Eq. (6.65), but rather the binary codeword $\mathbf{b}$ is mapped onto an array of channel symbols $\mathbf{x}$ by bit-interleaving the binary codeword and mapping it on the signal constellation $\mathcal{X}$ with a binary labelling rule μ. We have then

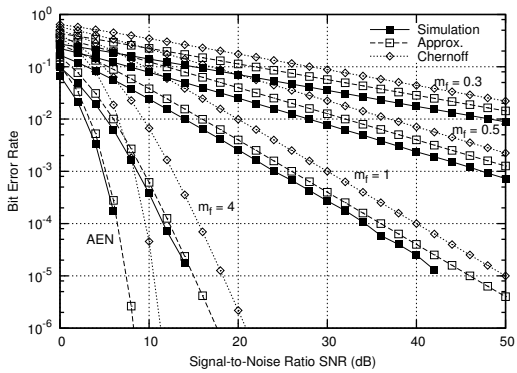


Figure 6.6: Comparison of simulation and saddlepoint approximation to $\mathrm{pep}(d=1)$ for 2-PEM in gamma fading of parameter $m_f = 0.3, 0.5, 1, 4$.

Theorem 6.17. *For BICM with infinite interleaving in the AEN channel, the pairwise error probability* $\mathrm{pep}(d)$ *admits the approximation*

$$\mathrm{pep}(d) \simeq \frac{2}{\sqrt{2\pi d}\,\mathrm{E}^{\frac{1}{2}}[\sqrt{\gamma}\log^2\gamma]}\ \mathrm{E}^{d+\frac{1}{2}}[\sqrt{\gamma}], \tag{6.87}$$

where γ *is the following function of the 5-tuple* (B, I, X, χ, Z)*,*

$$\gamma(b,i,x,\chi,z) = \frac{\sum_{x'\in\mathcal{X}_i^{\bar{b}},x'\leq x+\frac{z}{\mathrm{SNR}\,\chi}} e^{-\mathrm{SNR}\,\chi(x-x')-z}}{\sum_{x'\in\mathcal{X}_i^{b},x'\leq x+\frac{z}{\mathrm{SNR}\,\chi}} e^{-\mathrm{SNR}\,\chi(x-x')-z}}. \tag{6.88}$$

The expectation is carried out over all possible values of (B, I, X, χ, Z)*.*

The expectation over z can be carried out in closed form, as we did in Chapter 5. As for the fading, it seems to be less tractable, but numerical integration is straightforward.

We conclude our analysis by studying the asymptotic behaviour of the cumulant transform of the bit score for large SNR.

Theorem 6.18. *For large* SNR*, BICM in the AEN channel behaves as a binary modulation with distance* $d_{\min} = \min_{x,x'\in\mathcal{X}} |x - x'|$*, in the sense that*

$$\lim_{\mathrm{SNR}\to\infty} \frac{\kappa_1(\hat{r})}{\mathrm{SNR}} = -\frac{d_{min}}{2}. \tag{6.89}$$

Proof. The proof can be found in Appendix 6.F, bar for the trivial result on the minimum distance. □

For uniform 2^m-PEM, the cumulant transform is then approximately $\kappa_1(\hat{r}) \simeq -\frac{\mathrm{SNR}}{2^m-1}$. The optimum value of λ for the PEM constellations considered in Chapter 5 is $\lambda = 1$. Using the Chernoff bound, the pairwise error probability approximately decays as $\mathrm{pep}(d) \simeq e^{-d\frac{\mathrm{SNR}}{2^m-1}}$.

The modulations depicted in Fig. 6.7 asymptotically behave at large SNR as 2-PEM with respective losses of 4.77 dB for 4-PEM, 11.76 dB for 16-PEM, and 17.99 dB for 64-PEM. A comparison with the results for 2-PEM from Fig. 6.5 shows there is good agreement between the simulations and the asymptotic approximation suggested by Theorem 6.18.

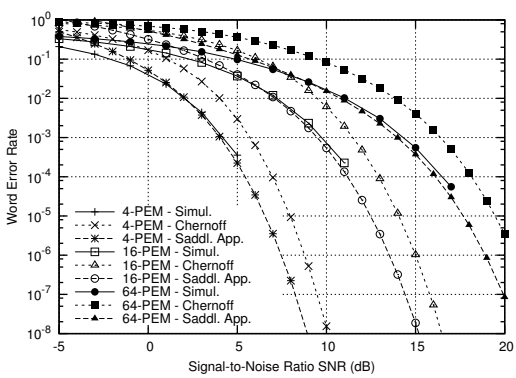


Figure 6.7: Comparison of simulation, Chernoff bound, and saddlepoint approximation to $\mathrm{pep}(d = 5)$ for 4-, 16-, and 64-PEM in the AEN channel.

Finally, the asymptotic loss with respect to 2^m-QAM modulation in the AWGN channel is $\frac{3}{2}$, or 1.76 dB, since we saw in Section 6.5.4 that $\kappa_1(\hat{r})$ grows asymptotically as $-\frac{3\,\mathrm{SNR}}{2(2^m-1)}$ for 2^m-QAM modulation in the AWGN channel.

6.7 Error Probability in the Binary Discrete-Time Poisson Channel

We now shift our attention to the discrete-time Poisson (DTP) channel. In this section we consider the error rates under binary modulation, 2-PEM. As in the AEN channel, each bit b_k is mapped onto a binary symbol x_k, with

$x_k = \{0, +2\}$; the mapping rule μ is

$$x_k = 0, \text{ if } \quad b_k = 0; \quad x_k = +2, \text{ if } \quad b_k = 1, \tag{6.90}$$

used with probability 1/2, and its complement $\bar{\mu}$, used with probability 1/2. The choice between μ and $\bar{\mu}$ is known at the receiver.

At time k, the channel output y_k follows a Poisson distribution with parameter $\varepsilon_s x_k \chi_k$, where ε_s is the average signal count, and χ_k is a fading coefficient having a gamma distribution, as in Eq. (6.22). For a fixed fading realization (known at the receiver) the channel transition probability $Q(y|x)$ is

$$Q(y|x) = e^{-\varepsilon_s x \chi} \frac{(\varepsilon_s x \chi)^y}{y!}. \tag{6.91}$$

We directly estimate the performance for the binary DTP channel in

Theorem 6.19. *In the absence of fading, the binary DTP channel is a Z-channel with transition probability $\epsilon = e^{-2\varepsilon_s}$. The pairwise error probability admits the saddlepoint approximation*

$$\text{pep}(d) = \begin{cases} \dfrac{2}{\sqrt{2\pi d}} \left(\sum_{j=0}^{\frac{d-1}{2}} e^{-2j\varepsilon_s} \right) e^{-(d+1)\varepsilon_s}, & d \text{ odd}, \\ \dfrac{2}{\sqrt{2\pi d}} \left(\frac{1}{2} + \sum_{j=1}^{\frac{d}{2}} e^{-2j\varepsilon_s} \right) \epsilon^{-d\varepsilon_s}, & d \text{ even} \end{cases}. \tag{6.92}$$

The Chernoff bound to the pairwise error probability is

$$\text{pep}(d) \leq e^{-d\varepsilon_s}. \tag{6.93}$$

In the presence of gamma fading, the Chernoff bound is

$$\text{pep}(d) \leq \left(1 + \frac{\varepsilon_s}{m_f}\right)^{-m_f d}, \tag{6.94}$$

and the saddlepoint approximation to the pairwise error probability is

$$\text{pep}(d) \simeq \begin{cases} \dfrac{2}{\sqrt{2\pi d}} \left(\sum_{j=0}^{\frac{d-1}{2}} \left(1 + \frac{(d+1+2j)\varepsilon_s}{m_f}\right)^{-m_f} \right), & d \text{ odd}, \\ \dfrac{2}{\sqrt{2\pi d}} \left(\frac{1}{2}\left(1 + \frac{d\varepsilon_s}{m_f}\right)^{-m_f} + \sum_{j=1}^{\frac{d}{2}} \left(1 + \frac{(d+2j)\varepsilon_s}{m_f}\right)^{-m_f} \right), & d \text{ even} \end{cases}. \tag{6.95}$$

The theorem coincides with the result for the AEN channel, replacing SNR by ε_s. Therefore, the comments and figures in Sections 6.6.2 and 6.6.3 apply unchanged to the DTP channel.

Proof. We prove the equivalence to a Z-channel. Under input randomization, a symbol $x_1 = 0$ is chosen with probability 1/2, in which case the output is $y = 0$, and the bit score is $\xi_b = \log e^{-2\varepsilon_s} = -2\varepsilon_s$. The symbol $x_2 = 2\varepsilon_s$ is chosen with probability 1/2. We distinguish two cases, $y = 0$, where $\xi_b = \log e^{2\varepsilon_s} = 2\varepsilon_s$, and $y > 0$, for which $\xi_b = -\infty$. This coincides with the binary AEN channel. Therefore the proofs of Theorems 6.15 and 6.16 apply. □

As for bit-interleaved coded modulation, we have the following

Theorem 6.20. *For BICM with infinite interleaving in the DTP channel, the pairwise error probability* $\mathrm{pep}(d)$ *admits the approximation*

$$\mathrm{pep}(d) \simeq \frac{2}{\sqrt{2\pi d}\,\mathrm{E}^{\frac{1}{2}}[\sqrt{\gamma}\log^2\gamma]}\,\mathrm{E}^{d+\frac{1}{2}}[\sqrt{\gamma}], \tag{6.96}$$

where γ *is the following function of the 5-tuple* (B, I, X, χ, Y),

$$\gamma(b, i, x, \chi, y) = \frac{\sum_{x'\in\mathcal{X}_i^{\bar{b}}} x'^y\, e^{-\mathrm{SNR}\,\chi x'}}{\sum_{x'\in\mathcal{X}_i^{b}} x'^y\, e^{-\mathrm{SNR}\,\chi x'}}. \tag{6.97}$$

The expectation is carried out over all possible values of (B, I, X, χ, Y).

We conclude our analysis by studying the asymptotic behaviour of the cumulant transform of the bit score for large ε_s.

Theorem 6.21. *As the average number of quanta* ε_s *goes to infinity, the cumulant transform of the bit score evaluated at the saddlepoint behaves as*

$$\lim_{\varepsilon_s\to\infty} \frac{\kappa_1(\hat{r})}{\varepsilon_s} = -\min_{x,x'\in\mathcal{X}} \frac{1}{2}\big(\sqrt{x} - \sqrt{x'}\big)^2. \tag{6.98}$$

The limit for uniform 2^m*-PEM modulation is given by*

$$\lim_{\varepsilon_s\to\infty} \frac{\kappa_1(\hat{r})}{\varepsilon_s} = -\left(1 - \sqrt{\frac{2^m-2}{2^m-1}}\right)^2. \tag{6.99}$$

For a given constellation set, the largest limit (in absolute value) is achieved when the modulation points x_j, $j = 1, \ldots, 2^m$, are placed at

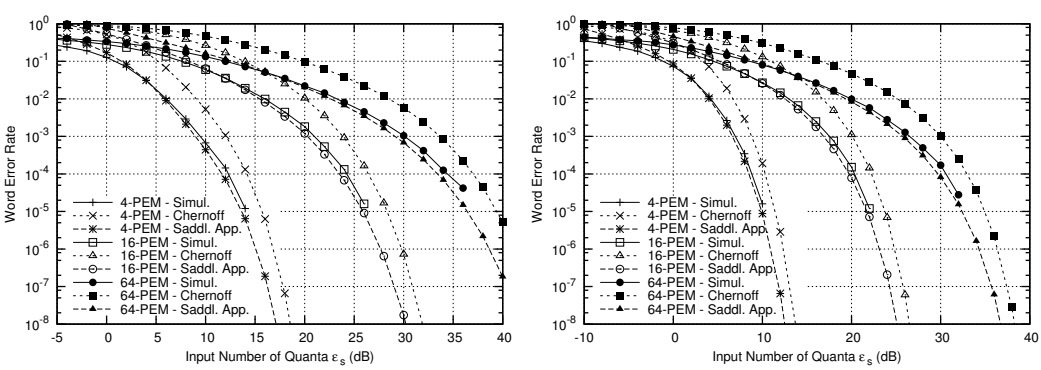
$$x_j = (j-1)^2 \frac{6}{(2^{m+1}-1)(2^m-1)}. \tag{6.100}$$

For this modulation set, the limit as $\varepsilon_s \to \infty$ is

$$\lim_{\varepsilon_s \to \infty} \frac{\kappa_1(\hat{r})}{\varepsilon_s} = \frac{3}{(2^{m+1}-1)(2^m-1)}. \tag{6.101}$$

The optimum constellation was also given in [64]. Note that the optimum value of λ for the PEM constellations considered in Chapter 5 is $\lambda = 2$.

Using the Chernoff bound, the pairwise error probability approximately decays as $\mathrm{pep}(d) \simeq e^{-d\frac{3\varepsilon_s}{(2^{m+1}-1)(2^m-1)}}$ for the optimum constellation and as $\mathrm{pep}(d) \simeq e^{-d\varepsilon_s\left(1-\sqrt{\frac{2^m-2}{2^m-1}}\right)^2}$ for the uniform constellation. For the cases depicted in Fig. 6.8a, the asymptotic energy loss of uniform 2^m-PEM with respect to binary modulation are 14.73 dB for 4-PEM, 29.39 dB for 16-PEM, and 41.97 dB for 64-PEM, in good agreement with Fig. 6.5.



(a) Uniformly spaced constellation.

(b) Optimized constellation.

Figure 6.8: Comparison of simulation, Chernoff bound, and saddlepoint approximation to $\mathrm{pep}(d = 5)$ for 4-, 16-, and 64-PEM in the DTP channel.

As for the optimum constellation, depicted in Fig. 6.8b, the energy loss with respect to binary modulation is reduced to 8.45 dB for 4-PEM, 21.90 dB

for 16-PEM, and 34.26 dB for 64-PEM. As m becomes large, the exponent is approximately $\frac{3}{2}2^{-2m}$, similar to the value $3\,2^{-2m}$ achieved by 2^m-PAM in Gaussian channels, as a function of SNR. In general, simulation results match well with the limit given by the theorem and with the saddlepoint approximation. On the other hand, the Chernoff bound is somewhat loose, as expected.

6.8 Conclusions

In this chapter, we have studied the computation of the pairwise error probability, an essential tool to estimate the error rates of practical channel codes. Unlike other common approaches to the subject, we have not derived true bounds to the error probability, such as the Chernoff bound, but have rather computed a saddlepoint approximation to the probability. In general, the approximation has been found to be very close to the exact simulation values. The saddlepoint approximation for continuous random variables has been derived anew, and its expansion up to second order determined. This approximation is valid for all values of the saddlepoint, even as it approaches zero.

We have found it useful to define a new quantity, the pairwise score, a random variable which depends on all the random elements in the channel and whose positive tail probability gives the pairwise error probability. For two codewords at Hamming distance d, the pairwise score is in turn the sum of d bit scores. For memoryless and stationary channels with binary input the bit scores are independent, but in general are statistically dependent. We have carried out our analysis of bit-interleaved coded modulation by assuming that they are independent, as proposed by Caire *et al.* [26].

Throughout the chapter we have considered a specific decoding metric, defined in Section 6.3, similar to the one used in BICM decoding. The methods presented in this chapter can however be easily extended to other metrics.

When applied to the AWGN channel, the formulas we have found seem to rank among the simplest yet closest approximations to the error probability. Two special cases are of particular interest, the fully-interleaved Nakagami fading channel, considered in Section 6.4, and the performance under BICM, discussed in Section 6.5. In the analysis of BICM, have solved a small inconsistency in the bounds presented in the paper [26] by Caire *et al.*. In Section 6.5.3 we have briefly considered the subject of finite interleaving and determined the performance of QPSK in fully-interleaved Nakagami fading channels. The effect of the finite interleaving has been found to be negligible.

We have seen how the AEN and the DTP channels with binary input admit a natural modelling as a Z-channel. We have determined the saddlepoint approximation to the error probability of the Z-channel, and therefore also to the other channels. We have considered gamma fading channels, a natural counterpart to Nakagami fading in the AWGN channel. Intriguingly, the performance of binary modulation in the exponential channel is almost identical to that of BPSK in Gaussian noise, for all the values of the fading parameter.

As for BICM, we have seen that the performance of 2^m-PEM at large signal-to-noise ratio in the AEN channel remains close to that of an AWGN channel with identical signal-to-noise ratio. More specifically, the pairwise error probability asymptotically decays in both channels as $e^{-a\,\mathrm{SNR}\,2^{-m}}$ for large SNR and large m; a is a constant factor of value $a = 3/2$ for 2^m-QAM modulation in the AWGN channel and $a = 1$ for 2^m-PEM modulation in the AEN channel.

Regarding non-binary transmission in the DTP channel, general formulas for the pairwise error probability have been given. These formulas show that equispaced 2^m-PEM modulation is non-optimal, as already discovered by Einarsson [64]. For the optimum constellation, a PEM modulation with parameter $\lambda = 2$, the asymptotic behaviour of the error probability as a function of the quanta count coincides with that of a Gaussian channel using 2^m-PAM modulation, i. e. it asymptotically decays as $e^{-a\varepsilon_s 2^{-2m}}$ for large ε_s and large m; here $a = 3$ for the AWGN channel and $a = 3/2$ for the DTP channel.

An interesting possible extension of the present work would include the analysis of the general quantized additive energy channel, in order to determine the way the functional form of the pairwise error probability changes between the additive exponential and discrete-time Poisson limits.

6.A Saddlepoint Location

For each i, and conditioned on b, the output density $Q_i(y|b)$ is given by Eq. (6.6). Since the metric $q_i(b, y)$ is proportional to $Q_i(y|b)$, we have that

$$\mathrm{E}[e^{r\Xi}] = \int \left(\frac{1}{2} Q_i(y|0) e^{r \log \frac{Q_i(y|1)}{Q_i(y|0)}} + \frac{1}{2} Q_i(y|1) e^{r \log \frac{Q_i(y|0)}{Q_i(y|1)}} \right) dy \tag{6.102}$$

$$= \int \left(\frac{1}{2} Q_i^{1-r}(y|0) Q_i^r(y|1) + \frac{1}{2} Q_i^{1-r}(y|1) Q_i^r(y|0) \right) dy. \tag{6.103}$$

This quantity is symmetric around $r = \frac{1}{2}$, since it remains unchanged if we replace r by $1 - r$.

The first derivative of $\kappa_1(r)$ is given by $\kappa_1'(r) = \frac{\mathrm{E}\left[\Xi_b e^{r\Xi_b}\right]}{\mathrm{E}\left[e^{r\Xi_b}\right]}$, and is thus zero when $\mathrm{E}\left[\Xi_b e^{r\Xi_b}\right] = 0$. We concentrate on $\mathrm{E}\left[\Xi_b e^{r\Xi_b}\right]$, readily computed as

$$\int \frac{1}{2}\left(Q_i(y|0)\log\frac{Q_i(y|1)}{Q_i(y|0)}\left(\frac{Q_i(y|1)}{Q_i(y|0)}\right)^r + Q_i(y|1)\log\frac{Q_i(y|0)}{Q_i(y|1)}\left(\frac{Q_i(y|0)}{Q_i(y|1)}\right)^r\right)dy \tag{6.104}$$

$$= \int \frac{1}{2}\left(Q_i(y|0)\left(\frac{Q_i(y|1)}{Q_i(y|0)}\right)^r - Q_i(y|1)\left(\frac{Q_i(y|1)}{Q_i(y|0)}\right)^r\right)\log\frac{Q_i(y|1)}{Q_i(y|0)}\,dy, \tag{6.105}$$

which is zero at $\hat{r} = \frac{1}{2}$ thanks to the symmetry.

As for the other derivatives, we have that

$$\kappa_1''(r) = \frac{\mathrm{E}\left[\Xi_b^2 e^{r\Xi_b}\right]}{\mathrm{E}\left[e^{r\Xi_b}\right]} - \frac{\mathrm{E}^2\left[\Xi_b e^{r\Xi_b}\right]}{\mathrm{E}^2\left[e^{r\Xi_b}\right]} \tag{6.106}$$

$$\kappa_1'''(r) = \frac{\mathrm{E}\left[\Xi_b^3 e^{r\Xi_b}\right]\mathrm{E}\left[e^{r\Xi_b}\right]}{\mathrm{E}^2\left[e^{r\Xi_b}\right]} - \frac{3\,\mathrm{E}\left[\Xi_b^2 e^{r\Xi_b}\right]\mathrm{E}\left[\Xi_b e^{r\Xi_b}\right]}{\mathrm{E}^2\left[e^{r\Xi_b}\right]} + \frac{2\,\mathrm{E}^3\left[\Xi_b e^{r\Xi_b}\right]}{\mathrm{E}^3\left[e^{r\Xi_b}\right]}, \tag{6.107}$$

which, evaluated at the saddlepoint, respectively become

$$\kappa_1''(\hat{r}) = \frac{\mathrm{E}\left[\Xi_b^2 e^{r\Xi_b}\right]}{\mathrm{E}\left[e^{r\Xi_b}\right]}, \quad \kappa_1'''(\hat{r}) = \frac{\mathrm{E}\left[\Xi_b^3 e^{r\Xi_b}\right]}{\mathrm{E}\left[e^{r\Xi_b}\right]}. \tag{6.108}$$

We thus find that the second derivative is proportional to

$$\frac{1}{2}\int\left(Q_i(y|0)\left(\frac{Q_i(y|1)}{Q_i(y|0)}\right)^r + Q_i(y|1)\left(\frac{Q_i(y|1)}{Q_i(y|0)}\right)^r\right)\log^2\frac{Q_i(y|1)}{Q_i(y|0)}\,dy, \tag{6.109}$$

which is positive at $r = \frac{1}{2}$. As for the third derivative, it is proportional to

$$\frac{1}{2}\int\left(Q_i(y|0)\left(\frac{Q_i(y|1)}{Q_i(y|0)}\right)^r - Q_i(y|1)\left(\frac{Q_i(y|1)}{Q_i(y|0)}\right)^r\right)\log^3\frac{Q_i(y|1)}{Q_i(y|0)}\,dy, \tag{6.110}$$

which is zero at $r = \frac{1}{2}$, thanks to the symmetry.

6.B A Derivation of the Saddlepoint Approximation

We wish to estimate the tail probability $\Pr(Z > z_0)$ for Z, a continuous random variable with density $p_Z(z)$. Instead of the density, we represent Z by its cumulant transform $\kappa(r)$, defined as $\kappa(r) = \log \mathrm{E}[e^{rZ}]$, r a complex number.

Often, the saddlepoint approximation is derived by assuming that Z is the sum of d random variables X_i, $Z = \sum_{j=1}^{d} X_j$, and the asymptotic behaviour as $d \to \infty$ is studied. Since in our problem there need be no asymptotics, we prefer to work with Z directly. To any extent, when the variables X_j are independent, the cumulant transform of Z is the sum of the transforms for each component, and the analysis applies unchanged. Bar for this change of emphasis, the presentation follows closely, and slightly generalizes, Olver's book [70], and Jensen's [67].

The tail probability can be recovered from $\kappa(r)$ by Fourier inversion:

$$\Pr(Z > z_0) = \frac{1}{2\pi j} \int_{r=-j\infty}^{j\infty} e^{\kappa(r) - r z_0} \frac{dr}{r}. \tag{6.111}$$

An application of Cauchy's integral theorem allows us to shift the integration path to the right, from the imaginary axis to a line $\mathcal{L} = (\hat{r} - j\infty, \hat{r} + j\infty)$ that crosses the real axis at another point $\hat{r}$.

It is most convenient to choose $\hat{r}$ the real number which verifies $\kappa'(\hat{r}) = z_0$. Since complex-variable analytic functions do not reach extreme points, this point is a saddlepoint. It exists and is unique due to the convexity of $\kappa(r)$ for r real. The shifted integration path can be parameterized by $r = \hat{r} + jr_i$, $-\infty < r_i < \infty$, or $(r - \hat{r}) = jr_i$. Using this new variable of integration we now expand the argument of the exponential term in a Taylor series around $\hat{r}$,

$$\kappa(r) - r z_0 = \kappa(\hat{r}) - \hat{r} z_0 + \frac{\kappa''(\hat{r})}{2!} (jr_i)^2 + R_2(r_i), \tag{6.112}$$

where we have used that the first derivative is zero and $R_2(r_i)$ includes the remaining terms in the expansion around $\hat{r}$,

$$R_2(r_i) = \sum_{\ell=3}^{\infty} \frac{\kappa^{(\ell)}(\hat{r})}{\ell!} (jr_i)^{\ell}. \tag{6.113}$$

Eq. (6.111) can then be rewritten as

$$\Pr(Z > z_0) = \frac{1}{2\pi} e^{\kappa(\hat{r}) - \hat{r} z_0} \int_{-\infty}^{+\infty} e^{-\frac{\kappa''(\hat{r})}{2} r_i^2} e^{R_2(r_i)} \frac{dr_i}{\hat{r} + j r_i} \tag{6.114}$$

$$= \frac{1}{2\pi} e^{\kappa(\hat{r}) - \hat{r} z_0} \int_{-\infty}^{+\infty} e^{-\frac{\kappa''(\hat{r})}{2} r_i^2} e^{R_2(r_i)} \frac{\hat{r} - j r_i}{\hat{r}^2 + r_i^2} \, dr_i, \tag{6.115}$$

where we have multiplied numerator and denominator times a factor $\hat{r} - j r_i$.

We next use the Taylor series of the exponential to expand $e^{R_2(r_i)}$, and express it in powers of r_i. Including the factor $\hat{r} - j r_i$ in the expansion,

$$(\hat{r} - j r_i) e^{R_2(r_i)} = (\hat{r} - j r_i) \sum_{m=0}^{\infty} \frac{1}{m!} \left(\sum_{\ell=3}^{\infty} \frac{\kappa^{(\ell)}(\hat{r})}{\ell!} (j r_i)^{\ell} \right)^m \tag{6.116}$$

$$= \sum_{m=0}^{\infty} \eta_m \, r_i^m, \tag{6.117}$$

where the terms with common factor r_i^m are grouped, and their corresponding coefficient denoted by η_m. The symmetry of the integrand in Eq. (6.115) implies that the integral of the terms with odd m is zero; we need thus consider only even values of m, which we parameterize by $m = 2n$. The first few terms are

$$\eta_0 = \hat{r}, \quad \eta_2 = 0, \tag{6.118}$$

$$\eta_4 = -\frac{\kappa^{(3)}(\hat{r})}{3!} + \hat{r} \frac{\kappa^{IV}(\hat{r})}{4!}, \tag{6.119}$$

$$\eta_6 = \frac{\kappa^{(5)}(\hat{r})}{5!} - \hat{r} \frac{\kappa^{(6)}(\hat{r})}{6!} - \hat{r} \frac{1}{2!} \left(\frac{\kappa^{(3)}(\hat{r})}{3!} \right)^2. \tag{6.120}$$

The next step is the normalization of the ℓ-th order derivatives, or ℓ-th cumulant, by $\kappa''(\hat{r})$. The normalized ℓ-th derivative, denoted by $\tilde{\kappa}^{(\ell)}(\hat{r})$, is

$$\kappa^{(\ell)}(\hat{r}) = \tilde{\kappa}^{(\ell)}(\hat{r}) \kappa''(\hat{r}), \tag{6.121}$$

and the coefficients η_{2n} become polynomials in $\kappa''(\hat{r})$. The degree of the polynomials will prove useful when tracking the various terms in the final expansion.

The tail probability in Eq. (6.115) becomes

$$\Pr(Z > z_0) = e^{\kappa(\hat{r}) - \hat{r} z_0} \sum_{n=0}^{\infty} \eta_{2n} \tau(2n), \tag{6.122}$$

a weighted sum of integrals $\tau(2n)$, for non-negative n, of the form

$$\tau(2n) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{\kappa''(\hat{r})}{2}x^2}\frac{x^{2n}}{\hat{r}^2+x^2}\,dx. \tag{6.123}$$

For $n = 0$,

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{\kappa''(\hat{r})}{2}x^2}\frac{1}{\hat{r}^2+x^2}\,dx = \frac{1}{2\hat{r}}\operatorname{erfc}\left(\hat{r}\sqrt{\frac{\kappa''(\hat{r})}{2}}\right)\exp\left(\hat{r}^2\frac{\kappa''(\hat{r})}{2}\right), \tag{6.124}$$

where $\operatorname{erfc}(x)$ is the error complementary function $\operatorname{erfc}(x) = \frac{2}{\pi}\int_x^\infty e^{-t^2}\,dt$. It has an asymptotic series

$$\operatorname{erfc}(x) = \frac{e^{-x^2}}{x\sqrt{\pi}}\sum_{m=0}^{\infty}(-1)^m\frac{1\cdot 3\cdot\cdots\cdot(2m-1)}{2^m x^{2m}}, \tag{6.125}$$

and therefore

$$\tau(0) = \frac{1}{\hat{r}^2\sqrt{2\pi\kappa''(\hat{r})}}\sum_{m=0}^{\infty}(-1)^m\frac{1\cdot 3\cdot\cdots\cdot(2m-1)}{\hat{r}^{2m}\big(\kappa''(\hat{r})\big)^m}. \tag{6.126}$$

In general, when $n \neq 0$, we can expand the fraction in the integrand by explicitly carrying out the division. We obtain then

$$\frac{x^{2n}}{\hat{r}^2+x^2} = \sum_{i=0}^{n-1}(-1)^{n-i-1}\hat{r}^{2(n-i-1)}x^{2i} + (-1)^n\frac{\hat{r}^{2n}}{\hat{r}^2+x^2}, \tag{6.127}$$

a sum which we next integrate term by term. Each summand,

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{\kappa''(\hat{r})}{2}x^2}x^{2i}\,dx, \tag{6.128}$$

is seen to be proportional to the $2i$-th moment of a normal random variable with zero mean and variance $\big(\kappa''(\hat{r})\big)^{-1}$, and therefore

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{\kappa''(\hat{r})}{2}x^2}x^{2i}\,dx = \frac{1\cdot 3\cdots(2i-1)}{\sqrt{2\pi}\big(\kappa''(\hat{r})\big)^{i+\frac{1}{2}}}. \tag{6.129}$$

When $i = 0$, we agree that the product $1\cdots(2i-1)$ is 1.

Combining the summands in Eq. (6.127) back into Eq. (6.123), and using the expansion (6.126) for the last summand in Eq. (6.127), we get

$$\begin{aligned}\tau(2n) &= \sum_{i=0}^{n-1}(-1)^{n-i-1}\hat{r}^{2(n-i-1)}\frac{1\cdot 3\cdots(2i-1)}{\sqrt{2\pi}\big(\kappa''(\hat{r})\big)^{i+\frac{1}{2}}} \\ &\qquad + (-1)^n\hat{r}^{2n}\frac{1}{\hat{r}^2\sqrt{2\pi\kappa''(\hat{r})}}\sum_{m=0}^{\infty}(-1)^m\frac{1\cdot 3\cdots\cdot(2m-1)}{\hat{r}^{2m}\big(\kappa''(\hat{r})\big)^m} \qquad (6.130)\\ &= \frac{1}{\hat{r}^2\sqrt{2\pi\kappa''(\hat{r})}}\left(\sum_{i=0}^{n-1}(-1)^{n-i-1}\hat{r}^{2(n-i)}\frac{1\cdot 3\cdots\cdot\big(2i-1\big)}{\big(\kappa''(\hat{r})\big)^i}\right. \\ &\qquad \left. + \sum_{m=0}^{\infty}(-1)^{n+m}\hat{r}^{2(n-m)}\frac{1\cdot 3\cdots\cdot(2m-1)}{\big(\kappa''(\hat{r})\big)^m}\right) \qquad (6.131)\\ &= \frac{1}{\hat{r}^2\sqrt{2\pi\kappa''(\hat{r})}}\sum_{m=n}^{\infty}(-1)^{n+m}\hat{r}^{2(n-m)}\frac{1\cdot 3\cdots\cdot(2m-1)}{\big(\kappa''(\hat{r})\big)^m}. \qquad (6.132)\end{aligned}$$

To derive Eq. (6.131) we used some obvious algebraic combinations, and the last equation follows by noting that the first $n-1$ terms in both summations exactly cancel each other. Factoring out the first term in the series, we obtain

$$\tau(2n) = \frac{1\cdot 3\cdots\cdot(2n-1)}{\hat{r}^2\sqrt{2\pi\kappa''(s)}\big(\kappa''(\hat{r})\big)^n}\left(1-\frac{2n+1}{\hat{r}^2\kappa''(\hat{r})}+\frac{(2n+1)(2n+3)}{\hat{r}^4\big(\kappa''(\hat{r})\big)^2}+\dots\right). \qquad (6.133)$$

The classical saddlepoint approximation is obtained by taking only $\tau(0)$, and then the leading term in Eq. (6.133),

$$\Pr(Z>z_0) \simeq \eta_0\tau(0)e^{\kappa(\hat{r})-\hat{r}z_0} = \frac{1}{\sqrt{2\pi\kappa''(\hat{r})}\hat{r}}e^{\kappa(\hat{r})-\hat{r}z_0}. \qquad (6.134)$$

Note that it loses its validity for small $\hat{r}$, in which case we may use Eq. (6.124),

$$\Pr(Z>z_0) = \frac{1}{2}\operatorname{erfc}\left(\hat{r}\sqrt{\frac{\kappa''(\hat{r})}{2}}\right)\exp\left(\frac{1}{2}\hat{r}^2\kappa''(\hat{r})^2\right)e^{\kappa(\hat{r})-\hat{r}z_0}. \qquad (6.135)$$

This approximation remains valid for small values of the saddlepoint, since the probability tends to $1/2$ for $\hat{r}\to 0$.

Higher-order approximations are obtained by extending the procedure. For instance, the following non-zero contribution comes from η_4, or $n=2$. The

leading coefficient of $\tau(4)$ in Eq. (6.133) is $\big(\kappa''(\hat{r})\big)^{-\frac{5}{2}}$, a term multiplied in by η_4, linear in $\kappa''(\hat{r})$. The effective leading coefficient is thus $\big(\kappa''(\hat{r})\big)^{-\frac{3}{2}}$, and coefficients up to the same order must be considered in $\tau(0)$,

$$\tau(0) \simeq \frac{1}{\hat{r}^2\sqrt{2\pi\kappa''(\hat{r})}}\left(1 - \frac{1}{\hat{r}^2\kappa''(\hat{r})}\right). \tag{6.136}$$

Similarly for $\tau(4)$,

$$\tau(4) \simeq \frac{1\cdot 3}{\hat{r}^2\sqrt{2\pi\kappa''(\hat{r})}\big(\kappa''(\hat{r})\big)^2}. \tag{6.137}$$

Since η_6 (Eq. (6.120)) has a term proportional to $\big(\kappa''(\hat{r})\big)^2$, the leading term of $\tau(6)$ must be included as well,

$$\tau(6) \simeq \frac{1\cdot 3\cdot 5}{\hat{r}^2\sqrt{2\pi\kappa''(\hat{r})}\big(\kappa''(\hat{r})\big)^3}. \tag{6.138}$$

The second-order saddlepoint approximation is

$$\Pr(Z > z_0) \simeq (\eta_0\tau_0 + \eta_4\tau_4 + \eta_6\tau_6)e^{\kappa(\hat{r})-\hat{r}z_0}, \tag{6.139}$$

where only one the last term in Eq. (6.120) for η_6 is to be included. Explicitly,

$$\Pr(Z > z_0) \simeq \alpha\frac{e^{\kappa(\hat{r})-\hat{r}z_0}}{\hat{r}\sqrt{2\pi\kappa''(\hat{r})}}, \tag{6.140}$$

where the factor α is given by

$$\begin{aligned}\alpha = \frac{\hat{r}}{\hat{r}}\left(1 - \frac{1}{\hat{r}^2\kappa''(\hat{r})}\right) + \left(-\frac{\kappa^{(3)}(\hat{r})}{3!} + \hat{r}\frac{\kappa^{IV}(\hat{r})}{4!}\right)\frac{1\cdot 3}{\hat{r}\big(\kappa''(\hat{r})\big)^2} \\ - \hat{r}\frac{1}{2!}\left(\frac{\kappa^{(3)}(\hat{r})}{3!}\right)^2\frac{1\cdot 3\cdot 5}{\hat{r}\big(\kappa''(\hat{r})\big)^3}\end{aligned} \tag{6.141}$$

$$= 1 - \frac{1}{\hat{r}^2\kappa''(\hat{r})} - \frac{\kappa^{(3)}(\hat{r})}{2\hat{r}\big(\kappa''(\hat{r})\big)^2} + \frac{\kappa^{IV}(\hat{r})}{8\big(\kappa''(\hat{r})\big)^2} - \frac{15\big(\kappa^{(3)}(\hat{r})\big)^2}{72\big(\kappa''(\hat{r})\big)^3}. \tag{6.142}$$

This additional term in the expansion also serves as an estimate of the error made by the approximation.

In general, the first term of the expansion gives a very good approximation to the real tail probability, with no need of considering extra terms.

6.C Pairwise Error Probability in the Z-Channel

The distribution of the pairwise score Ξ_{pw} is discrete and can be inscribed in a lattice with span $2|\log\epsilon| = -2\log\epsilon$, since $\epsilon < 1$. Also, the set of possible values of the pairwise score is $\big\{-d, -(d-2), \ldots, d-2, d\big\} \times |\log\epsilon|$.

The approximations differ depending on whether d is odd or even. When d is odd, the score is positive for $\xi_{\text{pw}} \in \{1, 3, \ldots, d\} \times |\log\epsilon|$, points of the form $(2j+1) \times |\log\epsilon|$, where j runs from 0 up to $\frac{d-1}{2}$. Using Theorem 6.6, we obtain

$$\text{pep}(d) = \Pr(\Xi_{\text{pw}} > 0) \simeq \sum_{j=0}^{\frac{d-1}{2}} \frac{\beta e^{-\hat{r}(2j+1)|\log\epsilon|}}{\sqrt{2\pi d \kappa_1''(\hat{r})}} e^{d\kappa_1(\hat{r})}. \tag{6.143}$$

And replacing the values of the lattice span β, $\kappa_1(\hat{r})$, and $\kappa_1''(\hat{r})$, we have that

$$\Pr(\Xi_{\text{pw}} > 0) \simeq \sum_{j=0}^{\frac{d-1}{2}} \frac{2|\log\epsilon| e^{\frac{1}{2}(2j+1)\log\epsilon}}{\sqrt{2\pi d \log^2\epsilon}} e^{d\frac{1}{2}\log\epsilon} \tag{6.144}$$

$$= \frac{2}{\sqrt{2\pi d}} \epsilon^{(d+1)/2} \sum_{j=0}^{\frac{d-1}{2}} \epsilon^j. \tag{6.145}$$

Similarly, if d is even, the values of the score ξ_{pw} leading to an erroneous decision are $\xi_{\text{pw}} = 0$, with an error probability of $1/2$, and $\xi_{\text{pw}} = 2, 4, \ldots, d$, when an error is surely made. This latter set has the form $2j$, where j runs from 1 to $d/2$. Then,

$$\text{pep}(d) = \tfrac{1}{2}\Pr(\Xi_{\text{pw}} = 0) + \Pr(\Xi_{\text{pw}} > 0) \tag{6.146}$$

$$\simeq \frac{1}{2} \frac{\beta}{\sqrt{2\pi d \kappa_1''(\hat{r})}} e^{d\kappa_1(\hat{r})} + \sum_{j=1}^{\frac{d}{2}} \frac{\beta e^{-\hat{r}2j|\log\epsilon|}}{\sqrt{2\pi d \kappa_1''(\hat{r})}} e^{d\kappa_1(\hat{r})} \tag{6.147}$$

$$= \frac{2}{\sqrt{2\pi d}} \epsilon^{d/2} \left(\frac{1}{2} + \sum_{j=1}^{\frac{d}{2}} \epsilon^j \right). \tag{6.148}$$

6.D Error Probability of Uncoded BPSK in Rayleigh Fading

The uncoded error probability in binary-input (BPSK), Rayleigh fading, Gaussian noise channels admits a closed-form expression [32]. We derive it in this

Appendix by using the bit score. Conditioned on the fading realization h, or $\chi = |h|^2$, we saw in the proof of Theorem 6.9 that $\Xi_b \sim \mathcal{N}(-4\chi\,\mathrm{SNR}, 8\chi\,\mathrm{SNR})$. Averaging over all possible values of χ, and using Mathematica, we obtain

$$p_{\Xi_b}(\xi_b) = \int_0^{+\infty} e^{-\chi} \frac{1}{\sqrt{16\chi\pi\,\mathrm{SNR}}} e^{-\frac{(\xi_b+4\chi\,\mathrm{SNR})^2}{16\chi\,\mathrm{SNR}}}\, d\chi \tag{6.149}$$

$$= \begin{cases} \frac{1}{4\sqrt{\mathrm{SNR}(1+\mathrm{SNR})}} \exp\left(-\frac{\xi_b}{2}\left(1+\sqrt{\frac{1+\mathrm{SNR}}{\mathrm{SNR}}}\right)\right), & \xi_b \geq 0 \\ \frac{1}{4\sqrt{\mathrm{SNR}(1+\mathrm{SNR})}} \exp\left(-\frac{\xi_b}{2}\left(1-\sqrt{\frac{1+\mathrm{SNR}}{\mathrm{SNR}}}\right)\right) & \xi_b < 0. \end{cases} \tag{6.150}$$

The distribution is a two-sided exponential; its decay is slower than that of a normal random variable.

The error probability P_w is the tail $\Xi_b > 0$ and takes the value

$$P_w = \int_0^{+\infty} \frac{1}{4\sqrt{\mathrm{SNR}(1+\mathrm{SNR})}} \exp\left(-\frac{\xi_b}{2}\left(1+\sqrt{\frac{1+\mathrm{SNR}}{\mathrm{SNR}}}\right)\right) d\xi_b \tag{6.151}$$

$$= \frac{1}{2}\left(1-\sqrt{\frac{\mathrm{SNR}}{1+\mathrm{SNR}}}\right), \tag{6.152}$$

where in the last equation we have multiplied times $\sqrt{1+\mathrm{SNR}} - \sqrt{\mathrm{SNR}}$. This equation coincides with the error probability of binary transmission in fading channels [32], as it should.

6.E Probability of All-One Sequence

We use Stirling's approximation to the factorial, $n! \simeq n^n e^{-n}\sqrt{2\pi n}$, to Eq. (6.49) in order to obtain

$$\mathrm{Pr}_{\mathrm{ind}} \simeq \frac{m^d \left(\frac{l}{m}\right)^{\frac{l}{m}} e^{-\frac{l}{m}} \sqrt{\frac{l}{m}} (l-d)^{l-d} e^{-(l-d)} \sqrt{l-d}}{\left(\frac{l}{m}-d\right)^{\frac{l}{m}-d} e^{-\left(\frac{l}{m}-d\right)} \sqrt{\frac{l}{m}-d}\, l^l e^{-l} \sqrt{l}} \tag{6.153}$$

$$= \frac{l^{\frac{l}{m}}}{(l-md)^{\frac{l}{m}-d+\frac{1}{2}}} \frac{(l-d)^{l-d+\frac{1}{2}}}{l^l}, \tag{6.154}$$

with the obvious simplifications and groupings of common terms. Extracting a factor l in $(l-d)$ and $(l-md)$, we get

$$\Pr_{\text{ind}} \simeq \left(1-\frac{md}{l}\right)^{-\frac{l}{m}+d-\frac{1}{2}}\left(1-\frac{d}{l}\right)^{l-d+\frac{1}{2}}. \tag{6.155}$$

Since the powers of l in numerator and denominator cancel.

Taking logarithms, the right-hand side of Eq. (6.155) becomes

$$\left(-\frac{l}{m}+d-\frac{1}{2}\right)\log\left(1-\frac{md}{l}\right)+\left(l-d+\frac{1}{2}\right)\log\left(1-\frac{d}{l}\right). \tag{6.156}$$

We now use Taylor's expansion of the logarithm $\log(1+t) \simeq t-\frac{1}{2}t^2+\mathrm{o}(t^3)$, and discard all powers of l higher than l^{-2},

$$\begin{aligned}\log \Pr_{\text{ind}} &\simeq \left(-\frac{l}{m}+d-\frac{1}{2}\right)\left(-\frac{md}{l}-\frac{m^2d^2}{2l^2}\right)\\ &\qquad+\left(l-d+\frac{1}{2}\right)\left(-\frac{d}{l}-\frac{d^2}{2l^2}\right) &(6.157)\\ &\simeq d+\frac{md^2}{2l}-\frac{md^2}{l}+\frac{md}{2l}-d-\frac{d^2}{2l}+\frac{d^2}{l}-\frac{d}{2l} &(6.158)\\ &= -\frac{md^2}{2l}+\frac{md}{2l}+\frac{d^2}{2l}-\frac{d}{2l}=-\frac{d(d-1)}{2l}(m-1). &(6.159)\end{aligned}$$

Finally, recovering the exponential,

$$\Pr_{\text{ind}} \simeq e^{-\frac{d(d-1)}{2l}(m-1)}. \tag{6.160}$$

6.F Cumulant Transform Asymptotic Analysis - AEN

The cumulant transform $\kappa_1(\hat{r})$ is

$$\kappa_1(\hat{r}) = \log \mathrm{E}\left[\left(\frac{\sum_{x'\in\mathcal{X}_i^1} e^{-\mathrm{SNR}(x-x')-z}u(\mathrm{SNR}(x-x')+z)}{\sum_{x'\in\mathcal{X}_i^0} e^{-\mathrm{SNR}(x-x')-z}u(\mathrm{SNR}(x-x')+z)}\right)^{\hat{r}}\right]. \tag{6.161}$$

In the limit $\mathrm{SNR}\to\infty$, we can take the dominant terms in the sums and

$$\lim_{\mathrm{SNR}\to\infty}\frac{\kappa_1(\hat{r})}{\mathrm{SNR}} = \lim_{\mathrm{SNR}\to\infty}\frac{1}{\mathrm{SNR}}\log \mathrm{E}\left[\left(\frac{e^{-\mathrm{SNR}(x-\tilde{x})-z}}{e^{-z}}\right)^{\hat{r}}\right] \tag{6.162}$$

where $\tilde{x}$ denotes the signal constellation symbol closest to and below x, in the complementary set $\mathcal{X}_i^1$. Cancelling the common term e^{-z}, we have that

$$\lim_{\mathrm{SNR}\to\infty} \frac{\kappa_1(\hat{r})}{\mathrm{SNR}} = \lim_{\mathrm{SNR}\to\infty} \frac{1}{\mathrm{SNR}} \log \mathrm{E}\big[e^{\hat{r}\,\mathrm{SNR}(\tilde{x}-x)}\big]. \tag{6.163}$$

The expectation now has the form of a sum over input bits, label positions, and symbols. The dominant summand is that with smallest distance $\tilde{x} - x$, which is also the (negative) minimum distance in the constellation, $-d_{\min}$, and

$$\lim_{\mathrm{SNR}\to\infty} \frac{\kappa_1(\hat{r})}{\mathrm{SNR}} = \lim_{\mathrm{SNR}\to\infty} \frac{1}{\mathrm{SNR}} \log \mathrm{E}\big[e^{-\hat{r}\,\mathrm{SNR}\,d_{\min}}\big] \tag{6.164}$$

$$= -\hat{r} d_{\min}. \tag{6.165}$$

6.G Cumulant Transform Asymptotic Analysis - DTP

The cumulant transform $\kappa_1(\hat{r})$ is

$$\kappa_1(\hat{r}) = \log \mathrm{E}\left[\left(\frac{\sum_{x'\in\mathcal{X}_i^1} e^{-\varepsilon_s x'}(\varepsilon_s x')^y/y!}{\sum_{x'\in\mathcal{X}_i^0} e^{-\varepsilon_s x'}(\varepsilon_s x')^y/y!}\right)^{\hat{r}}\right]. \tag{6.166}$$

In the limit $\varepsilon_s \to \infty$, we can take the dominant terms in the sums and

$$\lim_{\varepsilon_s\to\infty} \frac{\kappa_1(\hat{r})}{\varepsilon_s} = \lim_{\varepsilon_s\to\infty} \frac{1}{\varepsilon_s} \log \mathrm{E}\left[\left(\frac{Q(y|\tilde{x})}{Q(y|x)}\right)^{\hat{r}}\right] \tag{6.167}$$

where x denotes the transmitted symbol closest and $\tilde{x}$ denotes the closest symbol in the complementary set $\mathcal{X}_i^1$.

Carrying out the expectation over y, we get

$$\sum_{y=0}^{\infty}\big(Q(y|\tilde{x})Q(y|x)\big)^{1/2} = \sum_{y=0}^{\infty} \frac{1}{y!} e^{-\frac{1}{2}\varepsilon_s\tilde{x}}(\varepsilon_s\tilde{x})^{y/2} e^{-\frac{1}{2}\varepsilon_s x}(\varepsilon_s x)^{y/2} \tag{6.168}$$

$$= e^{-\frac{1}{2}\varepsilon_s\tilde{x}} e^{-\frac{1}{2}\varepsilon_s x} e^{\varepsilon_s\sqrt{\tilde{x}x}} \tag{6.169}$$

$$= e^{-\frac{1}{2}\varepsilon_s(\sqrt{\tilde{x}}-\sqrt{x})^2}. \tag{6.170}$$

As next step, we carry out the expectation over x, which has the form of a sum over input bits, label positions, and symbols. The dominant summand is

that with smallest distance $(\sqrt{\tilde{x}} - \sqrt{x})$, and we have that

$$\lim_{\varepsilon_s \to \infty} \frac{\kappa_1(\hat{r})}{\varepsilon_s} = -\min_{x,\tilde{x}} \frac{1}{2}\big(\sqrt{\tilde{x}} - \sqrt{x}\big)^2. \tag{6.171}$$

A natural problem is the determination of the constellation $\mathcal{X}$ which minimizes this exponent. One such constellation has points x_j, $j = 1, \ldots, 2^m$ at

$$x_j = (j-1)^2 \alpha, \tag{6.172}$$

where α is a factor to have normalized energy

$$1 = \alpha \frac{1}{2^m} \sum_{j=1}^{2^m} \frac{1}{2}(j-1)^2 = \alpha \frac{1}{2^m} \frac{1}{6} 2^m (2^m - 1)(2^{m+1} - 1) \tag{6.173}$$

and therefore $\alpha = \frac{6}{(2^{m+1}-1)(2^m-1)}$. The difference between adjacent symbols (and exponent of the pairwise error probability) is indeed independent of j,

$$\frac{1}{2}\big(\sqrt{x} - \sqrt{\tilde{x}}\big)^2 = \frac{1}{2}\alpha = \frac{3}{(2^{m+1}-1)(2^m-1)}. \tag{6.174}$$

7

Discussion and Recommendations

A detailed list of the main contributions of our study can be found in the Summary, as well as in the outline to the dissertation, included in the introductory chapter, starting at page 8. In this chapter, we critically discuss the results presented in previous chapters, their possible relevance for practical communication systems, and some links with other bodies of work. As a complement, we propose several possible extensions of the research into new directions.

As explained in some detail in Chapters 1 and 2, the additive energy channels incorporate traits of two standard channel models for the transmission of information via electromagnetic radiation, namely the discrete-time additive white Gaussian noise (AWGN) channel and the discrete-time Poisson (DTP) channel. Whereas in the AWGN channel the channel output is a complex-valued quadrature amplitude, given by the sum of a useful signal and an additive noise component, in the additive energy channels the output is, as for the DTP channel, a non-negative number, an energy. The non-negativity of the output is shared with the non-coherent discrete-time AWGN (NC-GN) channel, whose channel output is the squared modulus of the output of a coherent AWGN channel. Unlike the additive energy channels, for which the output is given by the sum of the signal and noise energies, the channel output in the NC-GN channel is affected by an additional beat term (cross-product) between signal and additive noise components.

We have considered two distinct cases, depending on whether the channel output is continuous (additive exponential noise channel —AEN—) or discrete (quantized additive energy channel —AE-Q—), in which case the energy comes as an integer number of quanta, each of them with energy ε_0. When the quantum of energy ε_0 vanishes, keeping the total energy constant, the AE-Q channel model degenerates into the AEN model. Finally, one can establish a link between the AE-Q channel and the DTP channel by noting that the

latter is an AE-Q channel whose additive noise is zero; equivalently the AE-Q channel is a generalization of the DTP channel whose output is affected by an additive noise component with geometric distribution. Figure 7.1 shows how the various channel models relate to each other.

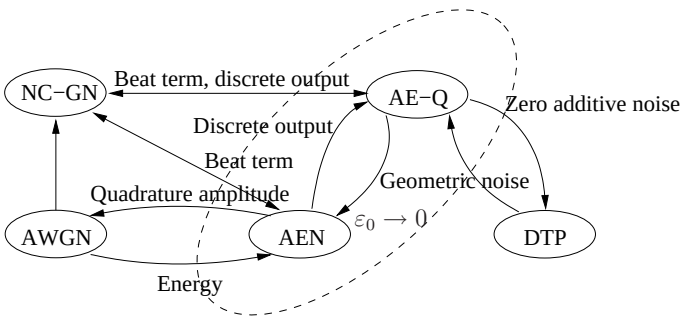


Figure 7.1: Additive energy channels (in dashed ellipse), their Gaussian relatives, and the discrete-time Poisson channel.

7.1 Elaboration of the Link with Practical Systems

Perhaps the most natural candidate for a real channel where the additive energy channel models are applicable is a radiation channel, be it at radio or optical frequencies. In Chapter 3, we considered an AE-Q channel where the quantum of energy is given by the energy of a photon, i. e. $\varepsilon_0 = h\nu$, where h is Planck's constant and ν the frequency, with the assumption that the additive noise component is distributed as thermal radiation at temperature T_0. We saw in that chapter the appearance of a natural threshold signal-to-noise ratio of the equivalent AWGN channel, denoted by SNR^*, and approximately given by

$$\mathrm{SNR}^* \simeq \frac{6 \cdot 10^{12}}{\nu}, \tag{7.1}$$

for a temperature $T_0 = 290\,\mathrm{K}$. For example, at frequencies 60 GHz and 600 MHz the respective thresholds are 20 dB and 40 dB. Below the threshold the capacity of the AE-Q channel is close to that of an equivalent AWGN channel, namely $\log\left(1 + \frac{E_s}{E_w}\right) = \log\left(1 + \frac{\varepsilon_s}{\varepsilon_n}\right)$, where $E_s = \varepsilon_s \varepsilon_0$ and $E_w = \varepsilon_n \varepsilon_0$ are the average signal and noise energy and ε_s and ε_n are the average number of photons in

signal and thermal noise. Above threshold, the capacity asymptotically approaches the value $\frac{1}{2}\log\varepsilon_s$. A similar change in the slope of the capacity from 1 to $\frac{1}{2}$ takes place in the NC-GN channel, as determined by Katz [23].

Another possible link could be made with the self-noise which appears in some channel measurements of wireless systems [76], or with phase noise, which is known to progressively becomes the limiting factor in performance for high enough frequencies. The general problem is to elaborate the possible relationship between phase noise, self-noise, and Poisson noise.

Since most radio receivers are based on some form of coherent detection, further research should be carried out on the possibility of designing energy-modulation transmitters and direct-detection receivers for waveform channels at either radio or optical frequencies.

Even though it is not reported in the dissertation, it is worthwhile mentioning that a linear amplifier model can be naturally defined for the AE-Q channel. In this model, each quantum at the input generates, independently of the remaining quanta, a random number of quanta at the output. It can be shown [77] that this model leads to a natural definition of noise figure and of Friis's formula for the noise figure of a chain of amplifiers.

It is intriguing that the capacity of coherent detection may be achieved even though no explicit use is made of the quadrature components of the signal. An explanation for this effect is likely to need some use of quantum information theory, possibly with an identification of the concepts "quantum" and "classical" capacity with some of the channels we have mentioned here. Some steps along this line were carried out by Helstrom [78], who however did not consider the additive energy channel models we have studied, but rather models based on a non-coherent detection, which leads to a chi-square or a Laguerre distribution, as we previously mentioned.

In a recent paper [79], Ben Schumacher, a pioneer in the field of quantum information theory, is quoted as saying that "interesting restrictions on experimental operations yield interesting information theories". The additive energy channels might thus be seen as "an interesting restriction on experimental operations" which naturally leads to an "information theory" deserving some study. In this dissertation, we have presented some elements of such an analysis.

7.2 Extensions of the Channel Model

The family of additive energy channels naturally admits more models than the ones we have considered. For example, in the AE-Q channel model the channel

output at time k y_k is given by the sum $y_k = s_k + z_k$, where s_k is a useful signal with Poisson distribution, and z_k an additive noise with geometric distribution.

A first natural extension of the model is the addition of a second Poisson noise source, say v_k, with Poisson distribution of a given mean, so that the channel output becomes $y_k = s_k + v_k + z_k$. In fact, in the analysis of optical channels, a DTP channel model of the form $y_k = s_k + v_k$ is often used; in this case the noise source v_k receives the name dark current and its origin is linked to an undesired source of ambient light or to the working of the photodiode.

Another possible extension relates to the effect of signal fading. Inclusion of fully-interleaved gamma-m_f fading is straightforward; as we saw in Chapter 3 in the context of DTP channels, the signal s_k has in this case a negative binomial distribution with coefficient m. Along this line, a physical reasoning for the appearance of a gamma distribution in the fading would also be of interest.

A natural question is the determination of the capacity in presence of a dark current or of fading. Concerning the dark current, its effect on the capacity of the DTP channel was studied by Lapidoth and Moser [42] and Brady and Verdú [35]. As for the fading, it is straightforward that the capacity of the AEN channel with gamma fading also coincides with that of an AWGN with gamma fading; in both cases the fading amplitude is assumed known at the receiver. This follows from the fact that the capacities of the AEN and AWGN channels with identical signal-to-noise ratio coincide.

Moreover, and as we mentioned previously, we saw in Chapter 3 that the capacity of the AE-Q channel has two distinct forms as a function of the signal-to-noise ratio. Below a threshold, additive geometric noise prevails over the signal-dependent Poisson noise; above the threshold the situation is reversed. In a similar vein, it would be interesting to determine the effect of gamma fading on the position of the threshold signal-to-noise ratio.

Other extensions refer to studying the capacity of the non-coherent Gaussian channel and its discrete counterpart, where the output has a Laguerre distribution [37]. Recently, the NC-GC channel was extensively studied by Katz and Shamai [23]; the capacity of the discrete non-coherent model seems to have received little attention in the literature.

Another research topic relates to finding the exact capacity or the capacity-achieving distributions of the DTP and AE-Q channels. Taking into account the experience from the DTP channel [35,42], this problem is likely to be hard. A possibly simpler problem is finding an alternative, simpler derivation of the bounds to the capacity of the DTP channel.

As we discussed in Chapter 1, multiple-antenna systems have received widespread attention in the past few years since their use increases the number

of degrees of freedom available for communication. In order to account for the presence of multiple antennas (or multiple-input multiple-output in general, MIMO), the scalar model $y_k = x_k + z_k$ should replaced by an equation relating signal and noise vectors through a mixing matrix. In contrast with the Gaussian channels, this mixing matrix has non-negative coefficients, and the optimality of standard matrix decomposition techniques to split the channel into a set of parallel sub-channels [4, 5] should be established. Otherwise, an extension of the matrix decomposition techniques to non-negative matrices could be necessary.

The matrix decomposition techniques would allow one to extend the MIMO results from microwave systems to similar optical designs, such as mode-group diversity multiplexing [20], a concept for the design of transparent parallel links proposed by the optical communications group at TU/e. Direct extensions of the Gaussian-channel results to optical frequencies have been reported by Stuart [19] and by Shah *et al.* [80], the latter using coherent detection.

Other model extensions include the consideration of multi-user scenarios. For the multiple-access AEN channel, Verdú found that its capacity region coincides with that of the AWGN channel [21]. Other multiple-user channels, such as the broadcast channel, still need to be studied.

We have recently obtained some results on the multiple-access additive energy channels [48]. Of special interest is the analysis of the effect of feedback on the multiple-access channel. Recall that, in an information-theoretic context, feedback means that the value of the channel outputs $y_1, y_2, \ldots, y_k$ are made known at the transmitter before sending the signal at time $k+1$, x_{k+1}. The signal x_{k+1} may therefore be a function of the channel outputs $y_1, y_2, \ldots, y_k$. For the single-user case, feedback does not increase the capacity [22], but it can do so for the multiple-user AWGN channel [22]. Despite the apparent similarity of the capacity regions of the AWGN and AEN channels, feedback does not enlarge the capacity region of the latter. For details, refer to [48]. Moreover, the threshold signal-to-noise ratio in the AE-Q channel depends on the number of active users. Essentially, the more users there are, the larger their aggregate Poisson noise is, and the lower the threshold.

7.3 Refinement of the Analysis of Coding and Modulation

Not only are the channel capacities of the AWGN and the AEN/AE-Q channels close under a broad set of conditions, we have also seen in Chapters 5 and 6 that there exist simple digital modulation formats whose constrained capacity

is similar to the channel capacity, and that the error rates of binary codes over these channels are close to those in the AWGN channel.

We have determined the first two derivatives of the constrained capacity at zero signal energy by direct computation. It would be worthwhile to explore further the link with estimation theory recently found by Guo, Shamai, and Verdú [60, 61] to provide an alternative computation for our results, possibly for arbitrary values of the signal energy. Along this line, we have found that the first-order coefficient of the constrained capacity over the AE-Q channel is zero, which implies that the minimum energy per bit is attained at a finite signal energy, whose value we have been unable to determine.

At the other range of signal energy, an exact form of the shaping gain of the DTP channel for our family of pulse energy modulations (PEM), with no use of a Gaussian approximation, is open. Since the approximation is quite good, the practical improvement is however likely to be small. Knowing the shaping gain for the DTP channel would allow for an extension of the quantitative analysis of PEM modulation to the AE-Q channel, and possibly for a determination of the threshold signal-to-noise ratio for the constrained capacity, rather than the channel capacity considered in the dissertation.

A side-line of our analysis of the wideband regime in the Gaussian channel was the extension of the trade-off between power and bandwidth to account for the possibility of variations in the power, next to the bandwidth change studied by Verdú [25]. It seems to be simple to extend our analysis to multiple-user Gaussian channels, thereby extending the results by Caire *et al.* [81].

As final element, we have estimated the pairwise error probability in some additive energy channels by using the union bound and a saddlepoint approximation to the pairwise error probability. A possible complement would be to use EXIT chart techniques [69] to optimize codes and modulation mappings.

From a more mathematical point of view, it would be useful to improve the accuracy of the saddlepoint approximation, possibly by estimating the error incurred by them. An extension of the saddlepoint approximation to general random variables, thereby removing the distinction between lattice and continuous random variables when using the approximation would have interest in itself, and would moreover lead to a method to consider the error probability in the AE-Q channel, which we have not considered in its full generality.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 623–656, Jul., Oct. 1948.

[2] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.

[3] A. S. Y. Poon, R. W. Brodersen, and D. N. C. Tse, "Degrees of freedom in multiple-antenna channels: A signal space approach," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 523–536, Feb. 2005.

[4] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–355, 1998.

[5] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[6] I. Bar-David, "Communication under the Poisson regime," *IEEE Trans. Inf. Theory*, vol. 15, no. 1, pp. 31–37, Jan. 1969.

[7] S. Verdú, "Poisson communication theory," in *Proc. Intl. Technion Comm. Day in honor of Israel Bar-David, March 25*, Haifa, Israel, 1999.

[8] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel - part I," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.

[9] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proc. - Part I*, vol. 137, no. 6, pp. 424–430, Dec. 1990.

[10] C. M. Caves and P. D. Drummond, "Quantum limits on bosonic communication rates," *Rev. Mod. Phys.*, vol. 66, no. 2, pp. 481–537, Apr. 1994.

[11] B. M. Oliver, "Thermal and quantum noise," *Proc. IEEE*, vol. 53, pp. 436–454, May 1965.

[12] J. P. Gordon, "Quantum effects in communication systems," *Proc. IRE*, vol. 50, no. 9, pp. 1898–1908, Sep. 1962.

[13] ——, "Noise at optical frequencies; information theory," in *Proc. Intl School of Physics "Enrico Fermi", Course XXXI.* London: Academic Press, 1964, pp. 156–181.

[14] Y. Yamamoto and H. A. Haus, "Preparation, measurement and information capacity of optical quantum states," *Rev. Mod. Phys.*, vol. 58, no. 4, pp. 1001–1020, Oct. 1986.

[15] J. M. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation/detection techniques for DWDM systems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 10, no. 2, pp. 259–272, Mar./Apr. 2004.

[16] V. W. S. Chan, "Coding and error correction in optical fiber communications systems," in *Optical Fiber Communications III*, I. P. Kaminow and T. L. Koch, Eds. San Diego: Academic Press, 1997, vol. A, pp. 42–62.

[17] A. H. Gnauck and P. J. Winzer, "Optical phase-shift-keyed transmission," *J. Lightw. Technol.*, vol. 23, no. 1, pp. 115–130, Jan. 2005.

[18] J. Conradi, "Bandwidth-efficient modulation formats for digital fiber transmission systems," in *Optical Fiber Telecommunications IVB: Systems and Impairments*, I. P. Kaminow and T. Li, Eds. San Diego: Academic Press, 2002, pp. 862–901.

[19] H. R. Stuart, "Dispersive multiplexing in multimode optical fiber," *Science*, vol. 289, pp. 281–283, Jul. 2000.

[20] T. Koonen, H. van den Boom, I. Tafur Monroy, and G.-D. Khoe, "High capacity multi-service in-house networks using mode group diversity multiplexing," in *Proc. of Opt. Fibre Comm. Conf., 2004, OFC 2004*, vol. 2, 2004, p. FG4.

[21] S. Verdú, "The exponential distribution in information theory," *Prob. Per. Inf.*, vol. 32, no. 1, pp. 86–95, Jan-Mar 1996.

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley and Sons, 1991.

[23] M. Katz and S. Shamai (Shitz), "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2257–2270, Oct. 2004.

[24] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[25] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[26] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[27] V. V. Prelov and E. C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals," *IEEE Trans. Inf. Theory*, vol. 39, no. 6, pp. 1728 – 1735, Sep. 1993.

[28] ——, "Higher order asymptotics of mutual information for nonlinear channels with nongaussian noise," in *Proc. 2003 Int. Symp. Inf. Th. ISIT 2003*, 2003, p. 83.

[29] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[30] J. R. Pierce, E. C. Posner, and E. R. Rodemich, "The capacity of the photon counting channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 61–77, Jan. 1981.

[31] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley and Sons, 1968.

[32] J. G. Proakis, *Digital Communications.* McGraw-Hill, 1995.

[33] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge: Cambridge Univ. Press, 2005.

[34] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–19, Jan. 1996.

[35] D. Brady and S. Verdú, "The asymptotic capacity of the direct detection photon channel with a bandwidth constraint," in *Proc. 28th Allerton Conf.*, Allerton House, Monticello, Sep. 1990, pp. 691–700.

[36] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. John Wiley and Sons, 1971, vol. 2.

[37] R. M. Gagliardi and S. Karp, *Optical Communications*. New York: John Wiley and Sons, 1976.

[38] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge University Press, 1995.

[39] R. E. Blahut, *Principles and Practice of Information Theory*. Addison-Wesley, 1987.

[40] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 1st ed. Academic Press, Inc. 1981.

[41] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[42] A. Lapidoth and S. M. Moser, "Bounds on the capacity of the discrete-time Poisson channel," in *Proc. 41st Allerton Conf.*, Oct. 2003.

[43] R. J. Evans, J. Boersma, N. M. Blachman, and A. A. Jagers, "The entropy of a Poisson distribution: Problem 87-6," *SIAM Rev.*, vol. 30, no. 2, pp. 314–317, Jun. 1988.

[44] G. E. Andrews, R. Askey, and R. Roy, *Special Functions*. Cambridge: Cambridge University Press, 1999.

[45] A. Lapidoth, "On phase noise channels at high SNR," in *Proc. IEEE Inf. Theory Workshop*, Bangalore (India), 2002, pp. 1–4.

[46] A. Martinez, "Spectral efficiency of optical direct detection," *J. Opt. Soc. Am. B*, vol. 24, no. 4, pp. 739–749, Apr. 2007.

[47] ——, "Capacity bounds for the Einstein radiation channel," in *Proc. of Int. Symp. Inf. Theory ISIT 2006, Seattle, USA*, pp. 366–370.

[48] ——, "Variations on the Gaussian multiple-access channel," in *Proc. of Int. Symp. Inf. Theory ISIT 2007, Nice, France*, pp. 2386–2390.

[49] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, A. Jeffrey, Ed. Academic Press, 1994.

[50] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.

[51] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[52] V. Prelov and S. Verdú, "Second order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[53] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Bit-interleaved coded modulation in the wideband regime," in *Proc. of Int. Symp. Inf. Theory ISIT 2007, Nice, France*, pp. 2131–2135.

[54] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation in the wideband regime," submitted to IEEE Trans. Inf. Th.

[55] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[56] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, Sep. 1984.

[57] G. Ungerboeck, "Channel coding with multi-level/phase signals," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[58] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[59] A. Martinez and F. Willems, "A coding theorem for bit-interleaved coded modulation," in *Proc. of the* 27th *WIC Benelux Symposium*, 2006.

[60] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1261–1282, Apr. 2005.

[61] D. P. Palomar and S. Verdú, "Representation of mutual information via input estimates," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 453–470, Feb. 2007.

[62] S. Golomb, "The limiting behavior of the Z-channel," *IEEE Trans. Inf. Theory*, vol. 26, no. 3, p. 372, May 1980.

[63] C. W. Helstrom, "Computing the performance of optical receivers with avalanche diode detectors," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 61–66, Jan. 1988.

[64] G. Einarsson, *Principles of Lightwave Communications*. John Wiley and Sons, 1996.

[65] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. McGraw-Hill, 1979.

[66] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Error probability of bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 262–271, Jan. 2006.

[67] J. L. Jensen, *Saddlepoint Approximations*. Oxford: Clarendon Press, 1995.

[68] P.-C. Yeh, S. Zummo, and W. Stark, "Error probability of bit-interleaved coded modulation in wireless environments," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 722–728, Mar. 2006.

[69] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Comm.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[70] F. W. J. Olver, *Asymptotics and Special Functions*. New York: Academic Press, 1974.

[71] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "A closed-form approximation for the error probability of coded BPSK fading channels," *IEEE Trans. Wir. Comm.*, vol. 6, no. 6, pp. 2051 – 2054, Jun. 2007.

[72] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*. Wiley-Interscience, 2000.

[73] E. Biglieri, G. Caire, G. Taricco, and J. Ventura-Traveset, "Simple method for evaluating error probabilities," *Electron. Lett.*, vol. 32, no. 3, pp. 191–192, Feb. 1996.

[74] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[75] L. van de Meeberg, "A tightened upper bound on the error probability of binary convolutional codes with Viterbi decoding," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 389 – 391, May 1974.

[76] R. Laroia, "Lessons unlearnt in wireless," in *Proc. of the 2006 Intl. Zurich Sem. Comm.*, 2006.

[77] A. Martinez, "Quantum noise in linear amplifiers revisited," in *Proc. Conf. on Information Systems and Sciences, CISS 2006*, 2006.

[78] C. W. Helstrom, *Quantum Detection and Estimation Theory.* Academic Press, 1976.

[79] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, "Reference frames, superselection rules, and quantum information," *Rev. Mod. Phys.*, vol. 79, no. 2, pp. 555–609, Apr.-Jun. 2007.

[80] A. R. Shah, R. C. J. Hsu, A. Tarighat, A. H. Sayed, and B. Jalali, "Coherent optical MIMO (COMIMO)," *J. Lightw. Technol.*, vol. 23, no. 8, pp. 2410 – 2419, Aug. 2005.

[81] G. Caire, D. Tuninetti, and S. Verdú, "Suboptimality of TDMA in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 608–620, Apr. 2004.

Index

Curriculum Vitae

A. Martinez was born in October 1973 in Zaragoza, Spain. In 1997 he received the M. Sc. in Telecommunications Engineering from the University of Zaragoza, in Spain. In the period 1998-2003 he was employed at the research centre of the European Space Agency in Noordwijk, the Netherlands, working on the design of digital communication systems for satellite communications. The main focus of his work was the analysis and design of coded modulation systems, ranging from modulation and channel coding to some aspects related to the implementation of channel decoders and demodulators.

In April 2003 he joined the Signal Processing group at the Faculty of Electrical Engineering in the Technische Universiteit Eindhoven, in the Netherlands to work on optical communication theory under the project mode-group diversity multiplexing. In this period, he analyzed different channel models for non-coherent communications. He also contributed to the design and implementation of a demonstrator of the mode-group diversity multiplexing technique, a method to transparently multiplex several simultaneous transmission channels over an optical fibre, using a single frequency band.

In 2007 he spent four months at University of Cambridge, in the United Kingdom, supported by the British Royal Society, working on the performance analysis and design of digital modulation systems.

His research interest is communication theory in general, with special attention to the interplay between communication theory and physics.