Master Thesis

Capacity and Capacity-Achieving Input Distribution of the Energy Detector

Erik Leitinger

Signal Processing and Speech Communication Laboratory
Graz University of Technology
Head: O. Univ.-Prof. Dipl.-Ing. Dr. techn. Gernot Kubin

Assessor: Assoc. -Prof. Dipl.-Ing. Dr. Klaus Witrisal
Supervisor: Dipl.-Ing. Bernhard Geiger

Graz, December 2011

Kurzfassung

Nichtkohärente Empfängerstrukturen, zu denen der Energiedetektor zählt, haben aufgrund ihrer geringen Komplexität große Bedeutung für Ultra-Wideband Kommunikationssysteme. Ein wichtiges informationstheoretisches Bewertungskriterium stellt die Kanalkapazität, die maximal erreichbare Informationsrate eines Übertragungskanals bei einem bestimmten Signal-zu-Rausch Verhältnis, dar. Um diese zu erreichen, muss das Eingangsalphabet bestimmte statistische Eigenschaften erfüllen. Für einen nichtlinearen Kanal wie den Energiedetektor weist ein derartiges Alphabet eine diskrete Amplitudenverteilung auf und kann z.B. durch Amplitudenumtastung dargestellt werden.

Die Herausforderung besteht darin, solch eine nichtlineare Optimierungsaufgabe für einen kontinuierlichen Kanal (Energiedetektor) zu lösen. Die vorliegende Arbeit stellt ein numerisches Verfahren zur Lösung dieser Problemstellung vor, welches die Kanalkapazität und das dazugehörige optimale Eingangsalphabet bestimmt. Darüber hinaus soll eine ausführliche Literaturrecherche Ansätze liefern, um schließlich zu zeigen, dass die Wahrscheinlichkeitsverteilung des Alphabets diskret sein muss und eine endliche Anzahl von Massenpunkten besitzt.

Abstract

Non-coherent receiver structures, which include the energy detector, have great importance for ultra-wideband systems due to their low complexity. The channel capacity, i.e. the maximum amount of information that can be transmitted at a particular signal-to-noise ratio (SNR), represents a fundamental concept in information theory. In order to achieve the channel capacity, the input alphabet has to satisfy certain statistical properties. For a non-linear channel like the energy detector, such an alphabet shows a discrete amplitude distribution and can be represented by on-off keying for low SNR or amplitude-shift-keying for higher SNR.

The difficulty lies in providing a solution for the underlying non-linear optimization problem for the continuous channel (energy detector). This master thesis presents a numerical algorithm to compute the channel capacity and the corresponding capacity-achieving input distribution over a wide range of SNR. Furthermore, it will be demonstrated that the input distribution has to be discrete with a finite number of mass points. A detailed literature research supports these conclusions.

Deutsche Fassung:

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz,am ....
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

....
date (signature)

Acknowledgements

I would like to thank my supervisors Klaus Witrisal and Bernhard Geiger for supporting me and inspiring me with new ideas. Especially, I want to thank Bernhard Geiger for all the hours he spent discussing difficulties which had to be dealt within the course of this work.

Graz, December 2011 Erik Leitinger

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The increasing interest in ultra-wide bandwidth (UWB) systems in the recent past arises from the benefits of large bandwidth — for UWB systems larger than 500 MHz — which offer the possibility of extremely high data rate or high-accuracy ranging and positioning even in indoor environments. Additionally, such systems should show lower power consumption and low complexity. Especially high accuracy-positioning and tracking by using impulse radio (IR) UWB systems — where the large bandwidth is produced by a train of short pulses — are growing fields of interest. The fine time resolution, which leads to a large number of resolvable multipath components (MPC), allows sub-meter ranging accuracy. Another substantial benefit of UWB systems is the reduction of small-scale fading effects by using the rich multipath diversity.

The 802.15.4a standard provides signalling schemes for such applications and allows the usage of high-complexity coherent receivers and low-complexity energy detectors. The very high sampling rate and the elaborate processing tasks, being a consequence of the large signal bandwidth, render the implementation of coherent receivers extremely complex. Tasks like synchronization, accurate phase tracking, the estimation of the amplitude and the phase of each multipath component at very low SNR are important issues in such architectures. Contrary to coherent receivers, non-coherent receivers like the energy detector are low-complexity but sub-optimum alternatives with reasonably lower power consumption, so that they became an interesting topic of research in this domain. A reason for the reduction of complexity is that the carrier phase of the signal does not have to be recovered.

The energy detector consists of a squaring device and an integrate-and-dump (I&D) device ([WLJ+09], [Gei09]), that is, a moving average (MA) filter with integration time T_i followed by a sampling device. Such a receiver structure can only observe the envelope, i.e. the instantaneous power, of the signal, but not the signal phase that changes from MPC to MPC. The energy detector simply accumulates the energy of all MPCs over the integration time T_i.

Due to the fact that any phase information is lost, modulation schemes like phase-shift keying are useless. Instead, modulation schemes like on-off keying (OOK) and

pulse-position modulation (PPM) are used. [WLJ+09] gives a very good overview of non-coherent UWB systems and treats topics such as the performance analysis of non-coherent structures like the energy detector, using binary PPM and OOK. Complexity and implementation aspects of the hardware architecture, synchronization, and time-of-arrival estimation are discussed as well.

At this point it is an interesting question to ask to what extent the deployed modulation schemes are similar to the modulation alphabets, e.g. amplitude distributions, which achieve the channel capacity, i.e. maximum data rate, of the energy detector. Out of this, an interest grows to analyse the energy detector from an information-theoretical point-of-view in order to find the channel capacity and the optimal amplitude distribution of the input alphabet. Therefore, a statistical model of the energy detector has to be derived. Using the obtained statistic, the energy detector can be treated as a transmission channel with an input and output alphabet, whereas the multipath propagation of the input signal is not considered. The main task of this master thesis was to compute the channel capacity and the capacity-achieving input distribution, i.e. the optimal input amplitude distribution, over a wide range of SNR of this continuous transmission channel. To solve this non-linear optimization problem, the so-called Blahut-Arimoto algorithm has been extended to continuous channels ([Dau05] and [Dau06]).

Another aspect will be to deliver a mathematical method to show that the achieving input distribution of the energy detector, which is a non-linear channel, is discrete with a finite number of mass points. Such discrete amplitude distribution can be represented by e.g. on-off keying for low SNR or amplitude-shift-keying for higher SNR. Due to the fact that the optimization problem of finding the optimal input distribution is very complex, I also will present a simpler method to approximate the channel capacity over certain regions of SNR.

1.1 Outline

In chapter 2 the statistical model of the energy detector is discussed in detail. The first section is dedicated to the signal model of the energy detector and how it can be simplified so that a statistical model of the energy detector's output can be established. Afterwards, I will give an exact mathematical description of the statistics of the output signal, i.e. the energy detector as information-theoretical channel. Furthermore, I formulate an approximation of the channel model of the energy detector.

The first part of chapter 3 discusses the concept of the channel capacity and the Blahut-Arimoto algorithm will be introduced. The next part of this section forms the main part of this work, due to the fact that the energy detector is seen as a continuous channel with a continuous input distribution. I will present an improved algorithm for solving the channel capacity for such continuous channels

with constrained input alphabets. The basic ideas for this algorithm were inspired by the paper [Dau05].

Chapter 4 treats discrete capacity-achieving input distributions of linear and non-linear channels with constraints. On the basis of a detailed literature research, I will briefly explain why the achieving input-distribution of such channels is discrete and outline the approach for proving the discreteness for the input distribution of the energy detector.

In chapter 5, the simulation results for the capacity of the energy detector and the corresponding achieving input distribution are presented and analysed. A comparison between the exact and the approximated statistical models for different integration times will be made. In addition, I will analyse the behaviour of the achieving input distributions for different peak amplitude-to-average power ratios of the input alphabet. After that, the results of the simulation will be evaluated using simple models of the energy detector.

Finally, chapter 6 will introduce a method to obtain a capacity-approaching input distribution over a range of SNR, at which the corresponding mutual information has nearly the same value as the channel capacity. The impulse for this idea comes from [FH09], where the input distribution, whose mutual information closely approaches the channel capacity, can be evaluated based on source entropy maximization.

Chapter 2

Statistical Model of the Energy Detector

The energy detector is a well-known receiver structure in wireless communications, which simply captures the energy of the transmitted signal in a certain period of time. The energy detector consists of a filter, a squaring unit and an energy integrator, as shown in Fig. 2.1. This relatively simple structure implicates a large complexity advantage compared to coherent receiver structures. It is obvious that such a non-coherent receiver can only observe the envelope of the received signal and any information of the symbol phase is lost.

To find a statistical representation of the energy detector, we have to find a proper analytical signal model. This means that, if we want to get the probability density function (PDF) of the energy detector output, we have to take into account the noise-bandwidth reduction because of filtering, the non-linear operation of the squaring unit, and the integration over the time interval.

This is the purpose of the following sections. The first step is to find a signal representation for the output of the energy detector, for which we can obtain a closed-form expression for the PDF. In the following one can show that this PDF can be approximated by Gaussian distribution with signal-dependent noise variance for large integration times. In the final step I want to demonstrate with a set of simulations that the theoretically correct PDF and the approximation match very well for sufficiently large integration times T_i.

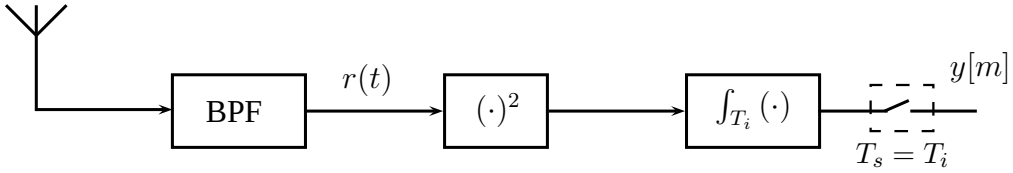


Figure 2.1: Structure of the energy detector, where T_i is the integration time and T_s is the symbol duration.

2.1 Energy Detector

For the following derivations it will be assumed that the input signal $r(t) = x(t)+\eta(t)$ of the energy detector does not depend on the transmission channel. Furthermore, it is assumed that the input signal is given in the baseband representation. The input signal $x(t)$, for example, can be represented as a pulse amplitude modulation (PAM) signal. For the noise signal $\eta(t)$ a Gaussian noise process with zero-mean and power spectral density (PSD) N_0 was considered.

The output of the integrator after the sampling device can be written as follows

$$y[m] = \int_{(m-1)T_i}^{mT_i} r(t)^2 \mathrm{d}t. \tag{2.1}$$

Out of this we get

$$y[m] = \int_{(m-1)T_i}^{mT_i} \Big(x(t) + \eta(t)\Big)^2 \mathrm{d}t = \int_{(m-1)T_i}^{mT_i} x(t)^2 + 2x(t)\eta(t) + \eta(t)^2 \mathrm{d}t \tag{2.2}$$

where T_i represents the integration interval and W_{rx} the bandwidth of the receiver bandpass-filter.

After the filtering, the input signal $x(t)$ and the noise $\eta(t)$ can be approximated by a finite sum of $T_i W_{rx}$ terms [Urk67]. For simplicity, a rectangular baseband-filter is considered with the bandwidth W_{rx}. So the filtered noise and signal are expressed by

$$\eta(t) = \sum_{n=-\infty}^{\infty} \eta[n]\mathrm{sinc}(tW_{rx} - n), \tag{2.3}$$

where

$$\eta[n] = \eta\left(\frac{n}{W_{rx}}\right). \tag{2.4}$$

Under these considerations, it is obvious that the noise samples are independent and identically distributed Gaussian random variables with zero-mean and variance

${\sigma_i}^2 = \sigma^2 = N_0 W_{rx}$. For the signal we can write

$$x(t) = \sum_{n=-\infty}^{\infty} x[n]\mathrm{sinc}(tW_{rx} - n) \tag{2.5}$$

where

$$x[n] = x\left(\frac{n}{W_{rx}}\right). \tag{2.6}$$

By plugging these results in Eq. 2.1 and using the fact that the sinc-functions at different time steps are orthogonal, the equation can be rewritten as

$$\begin{aligned} y[m] &= \int_{(m-1)T_i}^{mT_i} \left(\sum_{n=-\infty}^{\infty} (\eta[n]\mathrm{sinc}(tW_{rx} - n) + x[n]\mathrm{sinc}(tW_{rx} - n) \right)^2 \mathrm{d}t \\ &= \int_{(m-1)T_i}^{mT_i} \sum_{n=-\infty}^{\infty} \left(x[n]\mathrm{sinc}(tW_{rx} - n)dt + \eta[n]\mathrm{sinc}(tW_{rx} - n) \right)^2 \mathrm{d}t \\ y[m] &\approx \sum_{n=0}^{T_i W_{rx}-1} \left(x[mM - n] + \eta[mM - n] \right)^2 = \sum_{n=0}^{T_i W_{rx}-1} r[mM - n]^2 \end{aligned} \tag{2.7}$$

The product $T_i W_{rx}$ of the integration interval and the bandwidth of the baseband-filter, which represents the length of the summation operator is called noise dimensionality. After the summation and the sampling we get a discrete time description of the input signal $x[n]$ and the noise signal $\eta[n]$. Here we assume that noise samples $\eta[n]$ with a different time index do not depend on each other. This means that $\eta[n]$ can be described by an iid RV (independent identically distributed random variable). Further we assume that the input signal $x[n]$ is independent of the noise signal $\eta[n]$.

To get a closer look at the pass-band representation of the signals, and at the interpolation of the output signal for channel estimation I would like to refer to the diploma thesis [Gei09] by Bernhard Geiger. Here the baseband filtered signal was considered as complex and the noise-term before filtering as a complex Gaussian process with zero-mean and a two-sided PSD N_0.

As mentioned before, the noise variable $\eta[n]$ is an iid RV with zero-mean and variance σ^2. So the input-signal $r[n]$ which appears in Eq. 2.7 has the following statistical distribution

$$r[n] \sim \mathcal{N}(x[n], \sigma^2) \tag{2.8}$$

By the reason that $x[n]$ can be seen as the new input signal of the system, which represents the mean of the RV $r[n]$, and $\eta[n]$ as additive noise-term, the model for the energy detector for the following statistical analysis can be reconsidered as shown in Fig. 2.2.

We will now treat the system from $r(t)$ to $y[m]$ as a channel, for which we will

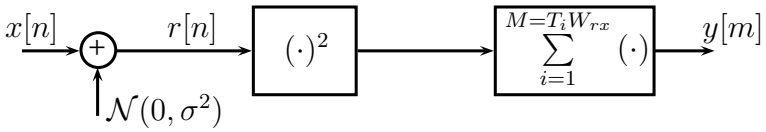


Figure 2.2: New structure of the energy detector for statistical considerations

determine the capacity and the capacity-achieving input distribution later. Since this channel is stationary and memoryless and the input-signal $x[n]$ can also be seen as memoryless, we simply can, with no loss of generality, suppress the time-index n if we speak about statistical representations of the input $x[n]$, the noise $\eta[n]$, the output $y[n]$, or the channel (transition PDF). For the following assumptions and calculations, the noise-dimensionality $T_i W_{rx}$, which represents the length of the sum, is replaced by the term M.

2.2 Non-Central Chi-Square Model

2.2.1 Derivation

The next natural task is to find a statistical representation for the derived model of the energy detector. As Fig. 2.2 shows, the input r of the squaring device has a Gaussian distribution $\mathcal{N}(x, \sigma^2)$, where x represents the mean value and σ^2 the variance. Such a non-linear operation, and the summation afterwards, transforms our RV with a Gaussian PDF into an RV with non-central χ^2 PDF.

It is important for this representation, that the variance of the transformed RV r is normalized to one, which is done by a division by the noise standard deviation σ. So one can write the output y of the energy detector as

$$y[m] = \sum_{n=0}^{M-1} \left(\frac{r[mM-n]}{\sigma} \right)^2 = \sum_{n=0}^{M-1} \tilde{r}[mM-n]^2 \tag{2.9}$$

where $\tilde{r}$ has the following distribution for $x[n] = 0$ or a constant value $x[n] = a$

$$\tilde{r}[n] \sim \begin{cases} \mathcal{N}(0,1), & x[n] = 0 \\ \mathcal{N}\left(\frac{a}{\sigma}, 1\right), & x[n] = a \end{cases}. \tag{2.10}$$

Out of this we can say that, if the samples $\tilde{r}[n]$ are an independent, normally distributed RVs like in Eq. 2.10, then the output signal $y[m]$ is distributed according to the non-central χ^2 distribution. Such a distribution is defined by two parameters,

the degree of freedom, which is specified by the noise dimensionality M and the so-called non-centrality parameter λ, which is calculated by the means μ_n of the RV $\tilde{r}[n]$, as the following shows

$$\lambda = \sum_{n=1}^{M} \mu_n^2 = \begin{cases} \sum_{n=1}^{M} \left(\frac{x[n]}{\sigma}\right)^2 = 0, & x[n] = 0 \\ \sum_{n=1}^{M} \left(\frac{x[n]}{\sigma}\right)^2 = M\frac{a^2}{\sigma^2}, & x[n] = a \end{cases}. \tag{2.11}$$

λ can be interpreted as SNR, where Ma^2 is the signal power and σ^2 is the noise power.

At this point I briefly want to talk about the energy detector input signal $x[n]$. I assume the input signal as constant over the entire integration time, so that we obtain the non-centrality parameter λ as shown in Eq. 2.11. But what happens to λ, if we send a single pulse with the same signal energy or if we assume a multipath propagation scenario for a UWB signal?
The energy of the signal over an integration period is given as

$$\sum_{i=1}^{M} x[n]^2 = P = Ma^2 \tag{2.12}$$

whereas the same signal energy can be obtained with the following input signals:

1. With the constant signal $x[n] = a$
2. With a single pulse $x[n] = \sqrt{P}\delta[n-m], \quad m \in 1, \ldots, M$
3. With series of pulses like a tapped-delay line model, as it is used for multipath propagation channels,

$$x[n] = \sqrt{P}\sum_{i=0}^{L-1} \alpha_i \delta[n-i], \tag{2.13}$$

where L can be seen as the number of MPC, α_i as the complex path gain and i as the discrete path delay of the lth component. (The symbol duration T_s has to be larger than the maximum excess delay τ_{max} of the channel, otherwise ISI (Inter-symbol Interference) occurs and we have to extend our channel model to a channel with memory.) A precondition for the channel to get the same signal power is that $\sum_{i=1}^{L} \alpha_i^2 = 1$. Such a multipath model can be assumed, if a signal with sufficiently large bandwidth, e.g. a UWB signal is assumed, at which every MPC can be considered as separable. (A closer look at the statistics of the energy detector with a multipath propagation channel is given in Appendix A.)

Due to the fact that all cases are equal, the choice of a constant input signal can be seen as correct and the $SNR = \frac{P}{\sigma^2}$.
With the knowledge of the non-centrality parameter, the PDF can be written as

$$p(y;\lambda,M) = \chi_M(\lambda) = \frac{1}{2}e^{-\frac{x+\lambda}{2}}\left(\frac{y}{\lambda}\right)^{\frac{M}{4}-\frac{1}{2}} I_{\frac{M}{2}}(\sqrt{\lambda y}) \tag{2.14}$$

where $I_k(z)$ is the modified Bessel function of the first kind. The mean and the variance of y are given by

$$E\{y\} = M + \lambda = \begin{cases} M, & x[n] = 0 \\ M + M\frac{a^2}{\sigma^2}, & x[n] = a \end{cases} \tag{2.15}$$

and

$$var\{y\} = 2(M + 2\lambda) = \begin{cases} 2M, & x[n] = 0 \\ 2M + 4M\frac{a^2}{\sigma^2}, & x[n] = a \end{cases}. \tag{2.16}$$

These results show that the mean and the variance of the output signal y of the energy detector increase both with the integration time, i.e. with the length of the sum M, and with the energy of the input signal x. This behaviour is shown in Fig. 2.3, where the distribution of the output signal is shown for some input values x.

The corresponding channel transition PDF is therefore the conditional non-central χ^2-distribution, conditioned on the input x and with the channel output y, so that we can write:

$$P_{Y|X}(y|x,M,\lambda) \sim \chi_M(\lambda) = \chi_M\left(\frac{Mx^2}{\sigma^2}\right) \tag{2.17}$$

2.2.2 Verification

In this section, the simulation results of the energy detector are shown to verify the statistical parameters of the analysed model. Fig. 2.3 also shows the experimental analysis of the output of the energy detector y for a specific input value x and the analytically calculated channel transition PDF for the same input value x. As we can see, the experimental results match with the analytical results.

The simulated mean and variance over a data-set shown as the dashed line in Fig. 2.4 follows the analytical calculation as in Eq. 2.15 and 2.16 of the mean and variance shown as the solid line. From this we can conclude that the derived model of the energy detector holds. Table 2.1 shows the numerical values of the mean and the variance for the input signal value $x = 1$.

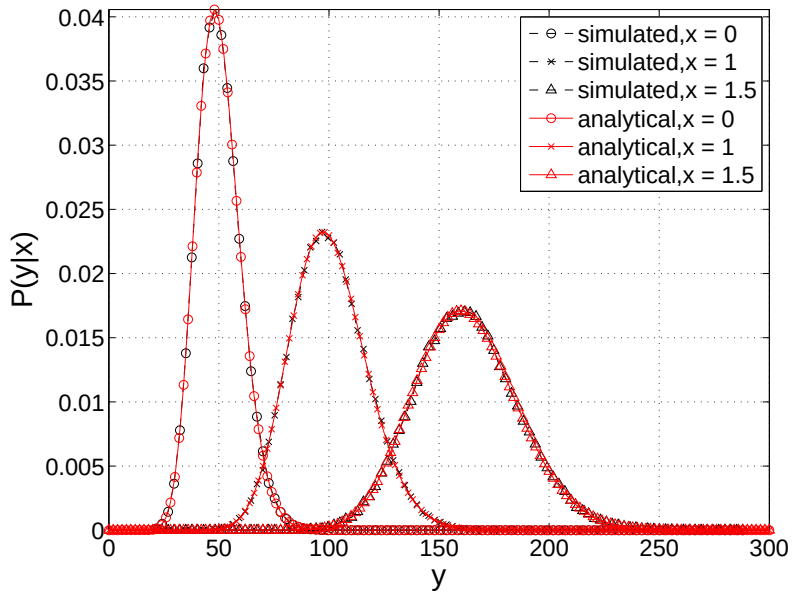
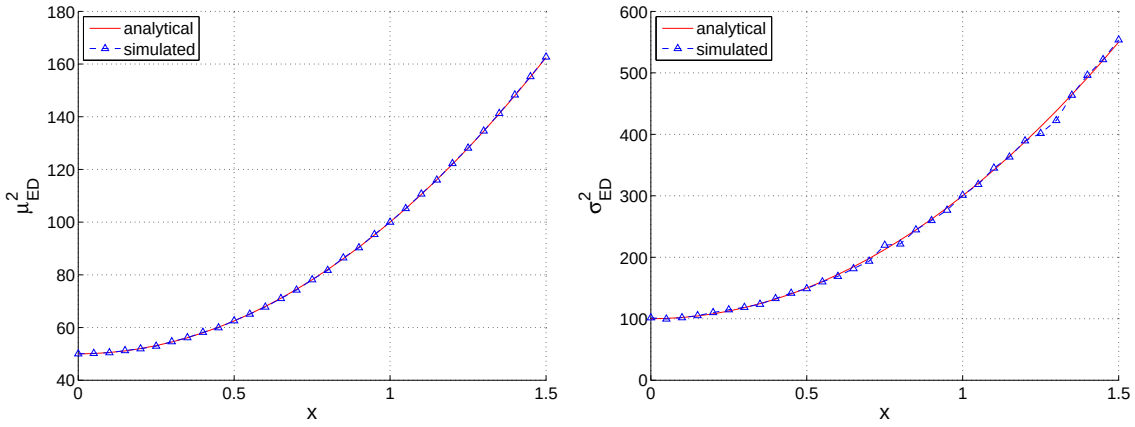


Figure 2.3: Non-central χ^2-channel transition PDF for $M = 50$. The solid red lines show the analytical PDF and the black dashed lines the simulated channel transition PDF.



(a) Mean μ_{ED} (b) Variance σ^2_{ED}

Figure 2.4: Analytical (red solid line) and simulated (blue dashed line) mean and variance of the energy detector output y over certain input values x for $M = 50$.

	M	σ^2	μ
analytical	50	300	100
simulated	50	301.4	99.6

Table 2.1: Mean and Variance for $x = 1$ (analytical and simulated)

2.3 Gaussian Approximation

2.3.1 Derivation

To simplify the problem for the further tasks, the exact statistical model of the energy detector can be modelled by another distribution. If the noise dimensionality $M = T_i W_{rx}$ is large enough, the non-central χ^2-distribution can be approximated by a Gaussian PDF with an input signal dependent variance $\sigma(x)$ and mean μ.

This fact holds because of the central limit theorem, which says that a sufficiently large number of independent identically distributed RV will be approximately normal distributed. For example a χ^2-distributed RV with a degree of freedom greater than 40 can be described by a Gaussian RV with a confidence better than 5% [SH06].

By taking over the variance and the mean derived for the non-central χ^2-distribution, the corresponding Gaussian approximation of the channel transition PDF can be written as

$$P_{Y|X}(y|x, M, \lambda) \sim \mathcal{N}\left(M + M\frac{x^2}{\sigma^2}, 2M + 4M\frac{x^2}{\sigma^2}\right). \tag{2.18}$$

Later in this section it will be shown with some qualitative simulations at which M the two models significantly differ from each other.

For a closer look at this approximation, I present the mathematical derivation for the mean and the variance. We take the derived signal model for the energy detector where $x[n]$ is normalized by σ:

$$\begin{aligned} y[m] &= \sum_{n=0}^{T_i W_{rx}-1} r[mM-n]^2 = \sum_{n=0}^{T_i W_{rx}-1} \Big(x[mM-n] + \eta[mM-n]\Big)^2 \\ &= \underbrace{\sum_{n=0}^{M-1} x[mM-n]^2}_{I} + \underbrace{2\sum_{n=0}^{M-1} x[mM-n]\eta[mM-n]}_{II} + \underbrace{\sum_{n=0}^{M-1} \eta[mM-n]^2}_{III} \end{aligned} \tag{2.19}$$

where

$$I = \begin{cases} 0, & x[n] = 0 \\ M\frac{a^2}{\sigma^2}, & x[n] = a \end{cases}$$

$$II \sim \begin{cases} 0, & x[n] = 0 \\ \mathcal{N}(0, 4M\frac{a^2}{\sigma^2}), & x[n] = a \end{cases} \tag{2.20}$$

$$III \sim \mathcal{N}(M, 2M)$$

and the distribution of the output signal y[n] under the Gaussian approximation is

$$y[n] \sim \begin{cases} \mathcal{N}(M, 2M), & x[n] = 0 \\ \mathcal{N}(M + M\frac{a^2}{\sigma^2}, 2M + 4M\frac{a^2}{\sigma^2}), & x[n] = a \end{cases}. \tag{2.21}$$

With these results the channel transition PDF for the Gaussian approximation can be written as

$$P_{Y|X}(y|x, M, \lambda) \sim \mathcal{N}\left(M + M\frac{x^2}{\sigma^2}, 2M + 4M\frac{x^2}{\sigma^2}\right). \tag{2.22}$$

2.3.2 Verification

At first I want to show that for $M = 50$ (same value as for the evaluation of the non-central χ^2-distribution model) the simulation of the energy detector output y and the Gaussian transition PDF for specific input values x are qualitatively the same, although some differences can be seen (see Fig. 2.5). However, we can say that for large M the Gaussian model with signal-dependent variance and mean models the behaviour of the energy detector very well.

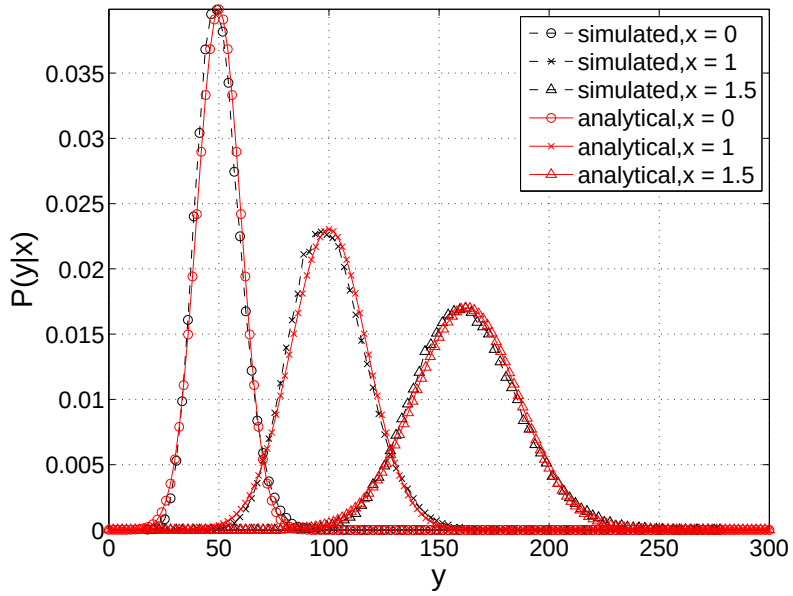
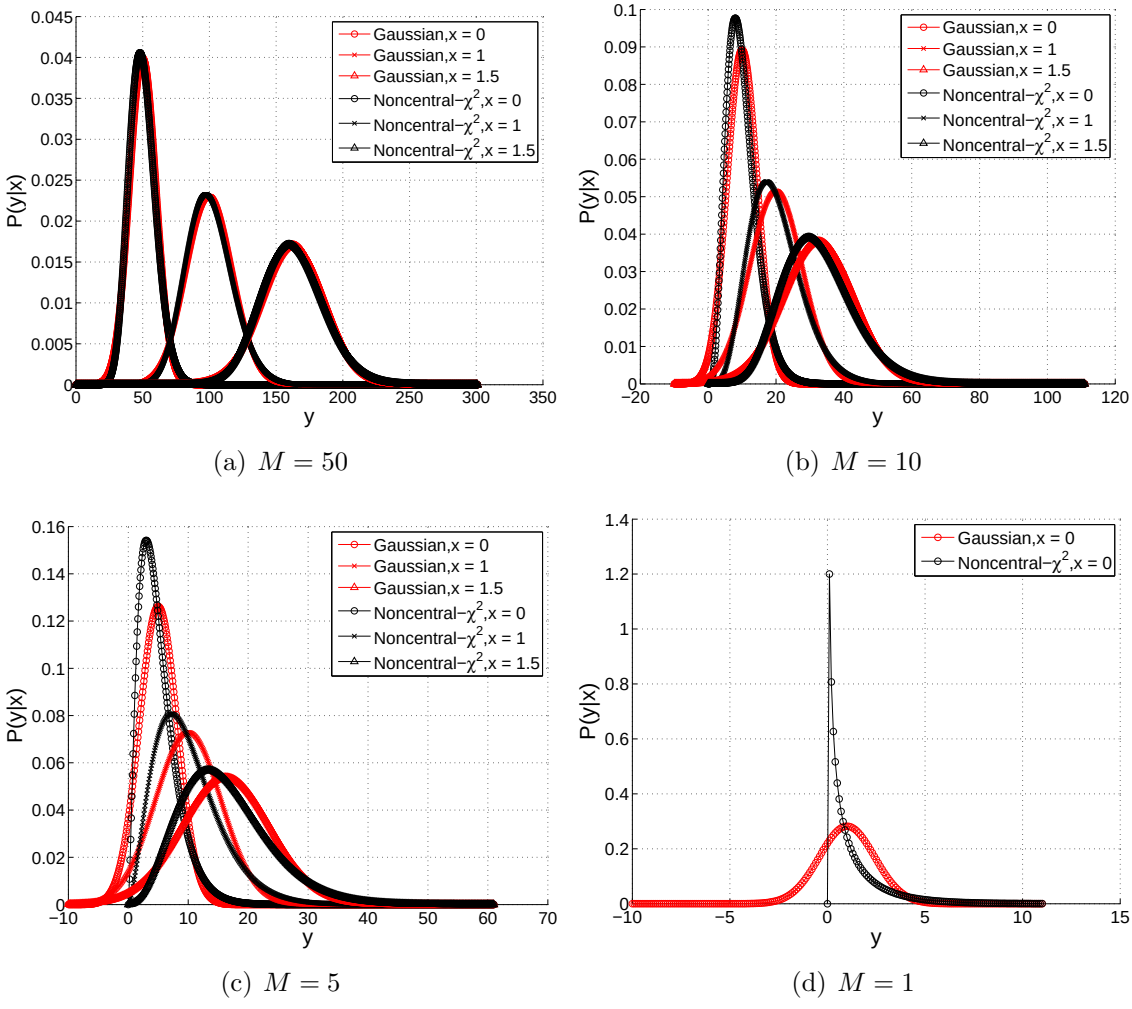


Figure 2.5: Gaussian approximation transition PDF for $M = 50$. The solid red lines show the analytical PDF and the black dashed lines the simulated transition PDF.

If M is getting smaller, the differences between the exact model and the Gaussian approximation are getting significant, as shown in Fig. 2.6. For $M = 1$ the energy detector works like a squaring device without summing up the values. In this case, we can see in Fig. 2.6(d) that the exact non-central χ^2 representation of the output signal still has zero probability for negative output values y, whereas the Gaussian approximation does not show this behaviour.



(a) $M = 50$ (b) $M = 10$

(c) $M = 5$ (d) $M = 1$

Figure 2.6: Non-central χ^2 transition PDF and Gaussian approximation transition PDF for $M = 1, 5, 10$, and 50.

Chapter 3

Channel Capacity of a Continuous Channel

The channel capacity is a fundamental concept in information theory, introduced by Claude Shannon [Sha48]. It gives an asymptotic bound on the maximum rate at which information can be reliably transmitted over a channel by using a specific input distribution, i.e. which information can be sent with arbitrarily low probability of error [CT91].

In [Bla72] and [Ari72] the so-called Blahut-Arimoto algorithm is introduced for computing the channel capacity of discrete channels and the probability mass function (PMF), which achieves this maximum mutual information. Blahut furthermore introduces the same algorithm for channels with certain constraints, e.g. an average power constraint, and also for continuous channels and sources, but gives no detailed solutions for such continuous channels.

Given that for continuous channels the Blahut-Arimoto algorithm is not directly applicable, one of the main tasks of this master thesis was to write an algorithm which calculates the channel capacity and the capacity-achieving input distribution for the energy detector, which can be represented as a continuous channel over a wide range of SNR. The conceptual idea of the algorithm is based on the paper [Dau05] and the PhD-thesis [Dau06] written by Justin Dauwels, where a numerical solution is given for continuous channels with constraints. Basically, this method works with two independent algorithms, which are iteratively executed. The first part calculates the masses of the input PMF, using the Blahut-Arimoto algorithm. The second part computes the positions of the mass points and is carried out by a gradient method, such as steepest decent.

In this section I will first give a short overview of the concept of channel capacity. Afterwards I introduce the extended version of the Blahut-Arimoto algorithm for continuous channels based on the ideas of Dauwels, which were used to calculate the channel capacity of the energy detector.

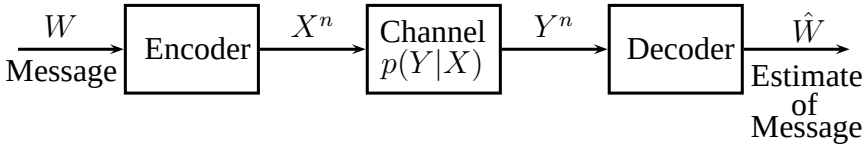


Figure 3.1: Communication system [CT91]

3.1 The Channel Capacity

3.1.1 Basics

For communication channels the channel capacity is considered as the limiting information rate that can be achieved with arbitrarily small error probability. This important concept of information theory states that the capacity is given by the maximum of the mutual information between the input and the output signal of the channel, where the maximization is done over the set of all input distributions. The input distribution, which achieves this maximum, is called the capacity-achieving input distribution. Any source that operates at a higher information rate than the capacity, will cause a loss of data because of uncorrectable errors, so that the real data rate can not be greater than the capacity.

Another way to describe this behaviour is to find the maximum number of distinguishable signals for a certain number n of uses of the communication channel. The channel capacity is defined as the exponent that leads to an exponential growth of the number of signals with n. So we can say that the channel capacity is the logarithm of the number of distinguishable signals, which in a first thought can be related to the entropy, i.e. the mutual information [CT91]. For the following theoretical derivations, the natural log is always chosen, but for practical results, the log is taken to the base 2, so that the channel capacity is expressed in terms of bits per channel use.

The mathematical block diagram of a communication system is shown in Fig. 3.1, where a message is mapped to the channel symbols, which are related to the output symbols after the channel. The output symbols occur randomly but have a distribution that depends on the distribution of the input symbols and the channel. We can choose the subset of input sequences with a specific distribution, so that every input sequence leads to a particular output sequence with negligible ambiguity. Then every input sequence can be reconstructed from the output sequences with an arbitrarily small error and the maximum information rate at which this can be done is called channel capacity.

The channel shown in Fig. 3.1 is defined as the conditional distribution function $P_{Y|X}(x, y) = p(y|x)$ which describes the probability of observing the output y given that the symbol x was sent. We implicitly assume that the channel is memoryless, i.e. the output symbols are independent of previous and future input and output

symbols. Let X and Y be discrete RVs with alphabets $\mathcal{X}$ and $\mathcal{Y}$ and probability mass functions $p(x) = Pr\{X = x\}$, $x \in \mathcal{X}$ and $p(y) = Pr\{Y = y\}$, $y \in \mathcal{Y}$. The channel capacity for discrete channels is defined as

$$\begin{aligned} C &= \max_{p(x)\in P_{\mathcal{X}}} I(X;Y) = \max_{p(x)\in P_{\chi}} (h(y) - h(y|x)) \\ &= \max_{p(x)\in P_{\mathcal{X}}} \sum_{x,y} p(x)p(y|x) \log \frac{p(y|x)}{p(y)} \end{aligned} \tag{3.1}$$

where $h(\cdot)$ is the differential entropy and $P_{\mathcal{X}}$ is the set of all probability mass functions on the countable set $\mathcal{X}$.

For continuous channels, i.e. continuous input and output alphabets, the RVs X and Y are said to be continuous and described by PDFs. So the summations are replaced by integrals and the channel capacity is defined as

$$C = \max_{p(x)\in P_{\mathbb{R}}} \iint_{x,y} p(x)p(y|x) \log \frac{p(y|x)}{p(y)} dxdy \tag{3.2}$$

where $P_{\mathbb{R}}$ is the set of all PDFs supported on the set $\mathbb{R}$ of real numbers.

If the input symbols are constrained with an averaging function $e(x)$, the maximization has to be done over the input distributions which satisfy the constraints. Thus, the new domain $P_{\mathbb{R}}(E)$ is the set of PDFs satisfying $E \geq \int_x e(x)p(x)dx$.

3.1.2 Example: Gaussian Channel

The Gaussian channel as shown in Fig. 3.2 is the most important continuous communication channel. The channel is assumed to be time-discrete and the output is the sum of the input X and the noise N. The noise N is an iid Gaussian continuous RV, which is independent of the input X. So that

$$Y = X + N, \qquad N \sim \mathcal{N}(0, \sigma^2) \tag{3.3}$$

Due to the fact that we assume an average power (AP) constraint, the signal X

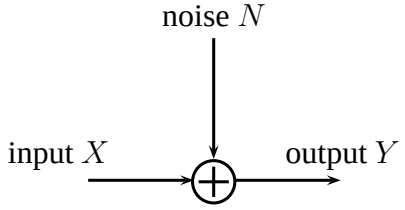


Figure 3.2: Additive Gaussian noise channel [CT91]

has to be restricted by

$$P \geq E\left\{X^2\right\} = \int_x x^2 p(x) dx. \tag{3.4}$$

The channel capacity is defined as the maximum of the mutual information between the input and output over all input distributions that satisfy the AP constraint. The capital letters X, Y and N represent continuous RVs. So the channel capacity can be calculated as follows

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+N|X) \\ &= h(Y) - h(N|X) \\ &= h(Y) - h(N) \end{aligned} \tag{3.5}$$

where $h(\cdot)$ is the differential entropy. As the noise N is Gaussian distributed, the entropy can be calculated very easily by $h(N) = \frac{1}{2}\log(2\pi e\sigma^2)$. The second moment of Y is

$$E\left\{Y^2\right\} = E\left\{X^2\right\} + 2\underbrace{E\left\{XN\right\}}_{=0} + 2\underbrace{E\left\{N^2\right\}}_{=\sigma^2} \leq P + \sigma^2. \tag{3.6}$$

$E\left\{Y^2\right\} = P + \sigma^2$ forms an upper bound for the entropy of Y by $\frac{1}{2}\log(2\pi e(P+\sigma^2))$. Applying this in Eq. 3.5 we obtain an upper bound for the mutual information, which represents the channel capacity

$$I(X;Y) = h(Y) - h(Y|X) \leq \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right). \tag{3.7}$$

This maximum is achieved if the input distribution X is also Gaussian distributed with $\mathcal{N}(0, P)$ [CT91].

$$C = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right) \tag{3.8}$$

3.2 Blahut-Arimoto Algorithm

Blahut [Bla72] and Arimoto [Ari72] independently developed algorithms to find the channel capacity and the corresponding capacity-achieving input distribution for discrete memoryless channels. Blahut also presented methods to calculate the rate distortion function and to calculate the capacity for channels with multiple constraints. Furthermore, he has shown that his newly developed mathematical tool can also be adapted to continuous channels. It is very interesting that Blahut has already mentioned that the achieving input distribution for continuous channels does not have to be continuous in general, but also can be discrete. Due to the fact that in most cases the capacity of continuous channels can not be solved analytically,

but with numerical methods, the mentioned discreteness of the input distribution can be a great benefit. The reason for that will be explained later. A fundamental discussion about the discreteness and finiteness of achieving input distributions for some channels will be given in Chapter 4.

To gain a better understanding of how the algorithm works, at first I will discuss the Blahut-Arimoto for discrete, constrained transmission channels. Afterwards, I will develop methods to solve these non-linear optimization problems for continuous channels based on [Dau05].

3.2.1 The BA-Algorithm for Unconstrained Discrete Channels

The evaluation[1] of the channel capacity C is carried out as the solution of a convex optimization problem. In most cases an analytical solution cannot be found. For this very reason an iterative algorithm was developed for discrete input- and output-alphabets, based on the fact that the mutual information can be formulated in two ways

$$\begin{aligned} I(X;Y) &= \min_{q(y)} \sum_x \sum_y p(x)p(y|x) \log \frac{p(y|x)}{q(y)} \\ I(X;Y) &= \max_{p(y|x)} \sum_x \sum_y p(x)p(y|x) \log \frac{\phi(x|y)}{p(x)} \end{aligned} \tag{3.9}$$

where X and Y are discrete RVs. $p(y|x)$ is the channel transition probability (matrix) from the channel input to the channel output, which represents the underlying channel, $p(x)$ is the input distribution (PMF), $q(y)$ is called the arbitrary output probability (PMF) and $\phi(x|y)$ is called the arbitrary channel transition probability (matrix) from the channel output to the channel input. The channel capacity is defined as

$$C = \max_{p(x)\in P_{\mathcal{X}}} \sum_x \sum_y p(x)p(y|x) \log \frac{p(y|x)}{\sum_x p(x)p(y|x)} \tag{3.10}$$

where

$$P_{\mathcal{X}} = \left\{ p(x) : \mathcal{X} \to [0,1] : \sum_x p(x) = 1 \right\} \tag{3.11}$$

is a set of all probability mass functions on the channel input alphabets. As Eqs. 3.9 and 3.10 show, $\phi(x|y)$ can be determined by the channel transition probability $p(y|x)$, the input distribution $p(x)$ and the arbitrary output distribution $q(y)$, which also can be expressed by the input probability and the channel transition probability. Due to the fact that the arbitrary transition probability depends only on the given

[1] The contents of sections 3.2.1 and 3.2.2 are based on [Bla72].

channel transition probability and the input probability, we can reformulate the optimization problem shown by Eq. 3.10 in an alternating double maximization problem given by

$$\begin{aligned} C &= \max_{p(x)} \max_{\phi(x|y)} J(p, P, \phi) \\ &= \max_{p(x)} \max_{\phi(x|y)} \sum_x \sum_y p(x)p(y|x) \log \frac{\phi(x|y)}{p(x)}. \end{aligned} \tag{3.12}$$

The alternating maximization works in the following way:

- For a fixed input distribution $p(x)$, $J(p, P, \phi)$ is maximized by

$$\phi(x|y) = \frac{p(y|x)}{\sum_x p(x)p(y|x)} \tag{3.13}$$

- For a fixed arbitrary channel transition probability $\phi(x|y)$, $J(p, P, \phi)$ is maximized over $p(x)$

The maximization of $J(p, P, \phi)$ over $p(x)$ can be obtained by using Lagrange multiplier to the constraint

$$\sum_x p(x) = 1. \tag{3.14}$$

We obtain

$$\frac{\partial}{\partial p(x)} \left\{ \sum_x \sum_y p(x)p(y|x) \log \frac{\phi(x|y)}{p(x)} + \lambda(\sum_x p(x) - 1) \right\} = 0 \tag{3.15}$$

which yields

$$-\log p(x) - 1 + \sum_x p(y|x) \log \phi(x|y) + \lambda = 0. \tag{3.16}$$

By combining the result 3.16 and the constraint 3.14 we can formulate the unknown parameter λ and can express the input distribution $p(x)$ by

$$p(x) = \frac{e^{\sum_y p(y|x) \log \phi(x|y)}}{\sum_{x'} e^{\sum_y p(y|x') \log \phi(x'|y)}}. \tag{3.17}$$

If the optimal arbitrary channel transition probability $\phi(x|y)$ is inserted into Eq. 3.17, the input distribution is capacity-achieving and can be written as

$$p(x) = \frac{p(x) e^{\sum_y p(y|x) \log \frac{p(y|x)}{\sum_x p(x)p(y|x)}}}{\sum_{x'} p(x') e^{\sum_y p(y|x') \log \frac{p(y|x')}{\sum_x p(x)p(y|x)}}} \tag{3.18}$$

where $\sum_x p(x)p(y|x)$ can be replaced by the output distribution $p(y)$. The sum in the exponent of Eq. 3.18 can be replaced by the Kullback-Leibler divergence or relative entropy $D(\cdot||\cdot)$ which is defined as

$$D(f(\cdot)||g(\cdot)) = \sum_x f(x) \log \frac{f(x)}{g(x)}. \tag{3.19}$$

This way we get the following update rules for the iterative alternating maximization algorithm:

1. Update the Kullback-Leibler divergence between the channel transition probability and the output distribution

$$\begin{aligned} D(p(y|x)||p^{k-1}(y)) &= \sum_y p(y|x) \log \frac{p(y|x)}{p^{k-1}(y)} \\ &= \sum_y p(y|x)) \log \frac{p(y|x)}{\sum_x p^{k-1}(x)p(y|x)} \end{aligned} \tag{3.20}$$

2. Update the input distribution

$$p^k(x) = \frac{p^{k-1}(x)e^{D(p(y|x)||p^{k-1}(y))}}{\sum_{x'} p^{k-1}(x')e^{D(p(y|x')||p^{k-1}(y))}} \tag{3.21}$$

These two optimization steps will be iterated until a stopping criterion is fulfilled.

3.2.2 Constrained Continuous Channels

If we turn attention to continuous channels and sources that are described by PDFs, this means that every sum in the algorithm turns into an integral and X and Y are continuous RVs. In this case, the Blahut-Arimoto algorithm is not directly applicable. A solution for such channels is given in the next section.

It is a fact that the input symbols of the channels get only a specific amount of power delivered by the source and can not exceed a certain amplitude. This means that the input distribution, which maximizes the mutual information, has to be constrained. Such a constrained channel is a channel which requires an average expense less than or equal to some specific number E, and the amplitude of an input symbol cannot exceed a specific number A. $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ are realisations of the RVs X and Y, which represent the values of the input and output symbols. Therefore, the domains[2] have to be restricted by the constraints given to the symbols.

[2]In the following equations I only indicate the domain of x and y if it is a final result of the equation

The channel capacity is then defined as

$$C = \sup_{p(x)\in P_{\mathbb{R}}(E,A)} I(X;Y) = \sup_{p(x)\in P_{\mathbb{R}}(E,A)} \int_{x\in\mathcal{X}} \int_{y\in\mathcal{Y}} p(x)p(y|x) \log \frac{p(y|x)}{\int_x p(x)p(y|x)} dxdy \tag{3.22}$$

where

$$P_{\mathbb{R}}(E,A) = \left\{ p : [0,A] \to [0,\infty] : \int_{x\in\mathcal{X}} p(x)dx = 1; 0 \leq x \leq A; \int_{x\in\mathcal{X}} e(x)p(x)dx \leq E \right\} \tag{3.23}$$

is a convex set of probability density functions on the channel input which satisfy the constraints. Due to the fact that the channel capacity of the energy detector will be solved for a set of input distributions which satisfy an AP and a peak amplitude (PA) constraint, the Blahut-Arimoto algorithm has to be formulated for these restrictions. The average power of the input alphabet can be seen as the second moment and is defined as

$$\int_x x^2 p(x)dx \leq P \tag{3.24}$$

where

$$\int_{x\in\mathcal{X}} p(x)dx = 1. \tag{3.25}$$

The PA constraint restricts the input symbols to $0 \leq x \leq A$. If in the initial input alphabet $x \in \mathcal{X}$ an amplitude appears which does not meet that constraint, the probabilities of such symbols have to be set to zero. A constraint-function which satisfies this condition and can be embedded in the algorithm, is

$$e(x) = \begin{cases} 0, & 0 \leq x \leq A \\ \infty, & x < 0 \wedge x > A \end{cases}. \tag{3.26}$$

This constraint-function is not directly embedded in the derivation of the algorithm, but is included in the final result of the update steps. Another option is a direct restriction of the initial input alphabet to values which satisfy the PA constraint $0 \leq x \leq A$.

The algorithm starts with an initial guess for the input distribution $p(x)$, so that the arbitrary channel transition probability $\phi(x|y)$ is given by

$$\phi(x|y) = \frac{p(x)p(y|x)}{\int_x p(x)p(y|x)dx} \tag{3.27}$$

Using the constraints 3.25 and 3.24, we can obtain the maximization 3.22 by using Lagrange multipliers s and λ (like in the previous section)

$$I(X;Y) = \int_x \int_y p(x)p(y|x) \log \frac{\phi(x|y)}{p(x)} dy + \lambda \left(\int_x p(x)dx - 1 \right) + s \left(\int_x x^2 p(x)dx - P \right) \tag{3.28}$$

The next step is to maximize $I(X;Y)$ with respect to $p(x)$, which is done by

$$\frac{\partial I(X;Y)}{\partial p(x)} = \int_y p(y|x) \log \frac{\phi(x|y)}{p(x)} dy - 1 + \lambda + sx^2 = 0 \tag{3.29}$$

Thus, we can obtain $p(x)$

$$\begin{aligned} e^{1-\lambda-sx^2} &= e^{\int_y p(y|x) \log \frac{\phi(x|y)}{p(x)} dy} \\ &= e^{-\log p(x) \int_y p(y|x) dy + \int_y p(y|x) \log \phi(x|y) dy} \\ &= p(x)^{-1} e^{\int_y p(y|x) \log \phi(x|y) dy} \\ \\ \rightarrow p(x) &= \frac{p(x) e^{\int_y p(y|x) \log \frac{p(y|x)}{\int_x p(x) p(y|x)} dy}}{e^{1-\lambda-sx^2}} \end{aligned} \tag{3.30}$$

By substituting $p(x)$ in constraint 3.25, we obtain

$$e^{1-\lambda} = \int_x \frac{p(x) \int_y p(y|x) \log \frac{p(y|x)}{\int_x p(x) p(y|x)} dy}{e^{-sx^2}} dx. \tag{3.31}$$

and by substituting $p(x)$ in constraint 3.24, we obtain

$$1 \geq \int_x \frac{x^2}{P} \frac{p(x) \int_y p(y|x) \log \frac{p(y|x)}{\int_x p(x) p(y|x)} dy}{e^{1-\lambda-sx^2}} dx. \tag{3.32}$$

If we combine 3.31 and 3.32 we get

$$\begin{aligned} \int_x \Big[1 - \frac{x^2}{P}\Big] p(x) e^{sx^2 + \int_y p(y|x) \log \frac{p(y|x)}{\int_x p(x) p(y|x)} dy} dx \geq 0 \\ \int_x \Big[1 - \frac{x^2}{P}\Big] p(x) e^{sx^2 + D(p(y|x)||p(y))} dx \geq 0 \end{aligned} \tag{3.33}$$

This non-linear equation can be solved by using the Newton-Raphson method as it is mentioned in [NO07b] and [NO07a]. Thus, we get an iterative solution for solving s as follows

$$s^{(n)} = s^{(n-1)} - \frac{\int_{x \in \mathcal{X}} e^{s^{(n-1)} x^2} \Big[1 - \frac{x^2}{P}\Big] p^{(k-1)}(x) e^{D(p(y|x)||p^{(k-1)}(y))} dx}{\int_{x \in \mathcal{X}} x^2 e^{s^{(n-1)} x^2} \Big[1 - \frac{x^2}{P}\Big] p(x)^{(k-1)} e^{D(p(y|x)||p^{(k-1)}(y))} dx} \tag{3.34}$$

where n defines the iteration-index to get the parameter s (iteration-index k is constant) and $D(p(y|x)||p^{(k-1)}(y)) = \int_{y \in \mathcal{Y}} p(y|x) \log \frac{p(y|x)}{\int_x p^{(k-1)}(x) p(y|x)} dy$ is the Kullback-Leibler divergence computed with the input distribution $p^{(k-1)}(x)$ from the previous iteration step k. The parameter s is always negative and defines the weight of the

AP constraint. The smaller the weight of the AP constraint, the smaller is s. If the iterative computation 3.34 produces positive values for s, this means that the AP constraint is inactive and only the PA constraint works. In this case the parameter s has to be set to zero. If we consider the solution for s in 3.30, the input distribution can be calculated with

$$p^{(k)}(x) = \frac{e^{(s^{(n)}x^2 - e(x))} p^{(k-1)}(x) e^{D(p(y|x)||p^{(k-1)}(y))}}{\int_{x\in\mathcal{X}} x^2 e^{(s^{(n)}x^2 - e(x))} p^{(k-1)}(x) e^{D(p(y|x)||p^{(k-1)}(y))} dx}. \tag{3.35}$$

The mutual information $I^{(k)}$ corresponding to the optimal input distribution $p^{(k)}(x)$ represents a lower bound for the channel capacity and is given by

$$I^{(k)} = \int_{x\in\mathcal{X}} p^{(k)}(x) D(p(y|x)||p^{(k)}(y)) dx. \tag{3.36}$$

The upper bound of the channel capacity is given by

$$U^{(k)} = \max_{x\in\mathcal{X}} \left\{ D(p(y|x)||p^{(k)}(y)) + s^{(n)}x^2 \right\} - s^{(n)}P^{(k)} \tag{3.37}$$

where

$$P^{(k)} = \int_{x\in\mathcal{X}} x^2 p^{(k)}(x) dx \tag{3.38}$$

which is the actual AP value. The function $e(x)$ for the PA constraint does not appear directly in the upper bound. It does not because the probability $p(x)$ for x not satisfying the PA constraint becomes zero. The probability $p(x)$ for x, which lies in the allowed region, does not get affected by this constraint. If after a certain number of iterations the gap between the lower and the upper bound is sufficiently small, one halts the algorithm and the mutual information $I^{(k)}$ can be seen as an estimate for the channel capacity. The iterative approximation of s seen in 3.34 yields an s so that the condition $P^{(k)} \leq P$ is satisfied and the gap between lower and upper bound is as small as possible. The Blahut-Arimoto algorithm for continuous channels with AP and PA constraint input can be seen as pseudo-code in code-box 1. In this form the algorithm can not be carried out directly, due to the fact that the continuous input PDF $p(x)$ leads to an infinite-dimensional optimization problem and to the fact that the integral has to be calculated. To solve this problem, we have to present a method to solve the integrals over the input and output distributions numerically. Furthermore the input distribution $p(x)$ has to be expressed in a numerical solvable form.

Algorithm 1: Constraint Blahut-Arimoto algorithm

Initialize:
(1) compute $p(y|x)$
(2) choose $p^{(0)}(x)$ so that $0 < p^{(0)}(x) < 1$ and $\int_x p^{(0)}(x)dx = 1$
(3) choose $I^{(0)}$ and $U^{(0)}$
(4) choose ϵ_c and ϵ_s
while $U^{(k)} - I^{(k)} \leq \epsilon_c$ **do**
  Step 1: compute:
  (1) $D(p(y|x)||p^{(k-1)}(y)) = \int_{y\in\mathcal{Y}} p(y|x) \log \frac{p(y|x)}{\int_x p^{(k-1)}(x)p(y|x)} dy$
  Step 2: Initialize $s^{(0,k)}$, $s^{(1,k)}$
  while $s^{(n,k)} - s^{(n-1,k)} \leq \epsilon_s$ **do**
    Step 3: compute:
    (1) $s^{(n-1,k)} = s^{(n,k)}$
    (2) $s^{(n,k)} = s^{(n-1,k)} - \frac{\int_{x\in\mathcal{X}} e^{s^{(n-1,k)}x^2}\left[1-\frac{x^2}{P}\right] p^{(k-1)}(x) e^{D(p(y|x)||p^{(k-1)}(y))}dx}{\int_{x\in\mathcal{X}} x^2 e^{s^{(n-1,k)}x^2}\left[1-\frac{x^2}{P}\right] p(x)^{(k-1)} e^{D(p(y|x)||p^{(k-1)}(y))}dx}$
    if $s^{(n)} > 0$ **then**
      $s^{(n)} = 0$
    end
  end
  Step 4: compute:
  if $x < 0$ *or* $x > A$ **then**
    $e(x) = \inf$
  else
    $e(x) = 0$
  end
  (1) $p^{(k)}(x) = \frac{e^{(s^{(n)}x^2 - e(x))} p^{(k-1)}(x) e^{D(p(y|x)||p^{(k-1)}(y))}}{\int_{x\in\mathcal{X}} x^2 e^{(s^{(n)}x^2 - e(x))} p^{(k-1)}(x) e^{D(p(y|x)||p^{(k-1)}(y))}dx}$
  (2) $I^{(k)} = \int_{x\in\mathcal{X}} p^{(k)}(x) D(p(y|x)||p^{(k)}(y))dx$
  (3) $P^{(k)} = \int_{x\in\mathcal{X}} x^2 p^{(k)}(x)dx$
  (4) $U^{(k)} = \max_{x\in\mathcal{X}} \left\{D(p(y|x)||p^{(k)}(y) + s^{(n)}x^2\right\} - s^{(n)}P^{(k)}$
end

3.2.3 The Particle Method

The presented method for solving such an optimization problem is developed by Dauwels in [Dau05] and [Dau06]. He gives a very simple and effective way to skirt the original infinite-dimensional problem and reduce it to a finite-dimensional optimization problem.
His suggestion was to present the input distribution $p(x)$ with a finite but long list of particles

$$\mathcal{P}_X = \{(\hat{x}_1, w_1), (\hat{x}_2, w_2), \ldots, (\hat{x}_N, w_N)\} \tag{3.39}$$

where $w = \{w_1, \ldots, w_N\}$ are the weights, i.e. the probabilities, of the particles and $\hat{x} = \{\hat{x}_1, \ldots, \hat{x}_N\}$ are the positions of the particles. This list can be seen as a PMF, which approximates the original input PDF $p(x)$. For some channels, as we will discuss in Chapter 4, the capacity achieving input distribution is known to be discrete and finite, so that a finite list of particles can represent this distribution exactly. For continuous achieving input distributions like the Gaussian distribution, this obviously does not work, but with a large list of particles the original input distribution is approximated well enough.

We get the following finite-dimensional optimization problem

$$C_{\mathcal{P}} = \max_{(\hat{x},w)} I(\hat{x}, w) \tag{3.40}$$

where

$$I(\hat{x}, w) = \sum_{i=1}^{N} \int_{y \in \mathcal{Y}} w_i p(y|\hat{x}_i) \log \frac{p(y|\hat{x}_i)}{p(y)} dy. \tag{3.41}$$

The corresponding output distribution $p(y)$ is given by the channel transition PDF $p(y|x)$, sampled on the particle positions $\hat{x}_i$ and the appearing probability w_i of the particles

$$p(y) = \sum_{i=1}^{N} w_i p(y|\hat{x}_i). \tag{3.42}$$

The remaining integral in the mutual information can not be solved analytically in most cases. Therefore, it may be evaluated by numerical or Monte-Carlo integration. So we can compute the mutual information in the following way

$$\begin{aligned} I(\hat{x}, w) &= h \sum_{i=1}^{N} \sum_{l=1}^{H} w_i p(\hat{y}_l|\hat{x}_i) \log \frac{p(\hat{y}_l|\hat{x}_i)}{\hat{p}(\hat{y}_l)} \\ &= h \sum_{i=1}^{N} w_i D(p(\hat{y}|\hat{x}_i)||p(\hat{y})) \end{aligned} \tag{3.43}$$

where $\hat{y}$ are the quantized levels and the output distribution $p(\hat{y})$ is then approximated by a histogram with bins of width h (see Fig. 3.3). Due to the fact that we have only a finite number of particles with a weight w_i and a position $\hat{x}_i$, we have to optimize the mutual information with respect to these weights and positions. The Blahut-Arimoto algorithm can be solved with this particle method and numerical integration, but it can only find the optimal probability of the particles which represent the input distribution. It will not find the optimal position.

The mutual information is concave w.r.t. the convex set of input distribution $p(x)$, whereas $I(\hat{x}, w)$ is non-convex w.r.t. the positions and weights. This can be explained by the fact that the mutual information is a concave function of $p(x)$ for fixed $p(y|x)$ and a convex function of $p(y|x)$ for fixed $p(x)$ [CT91]. If both

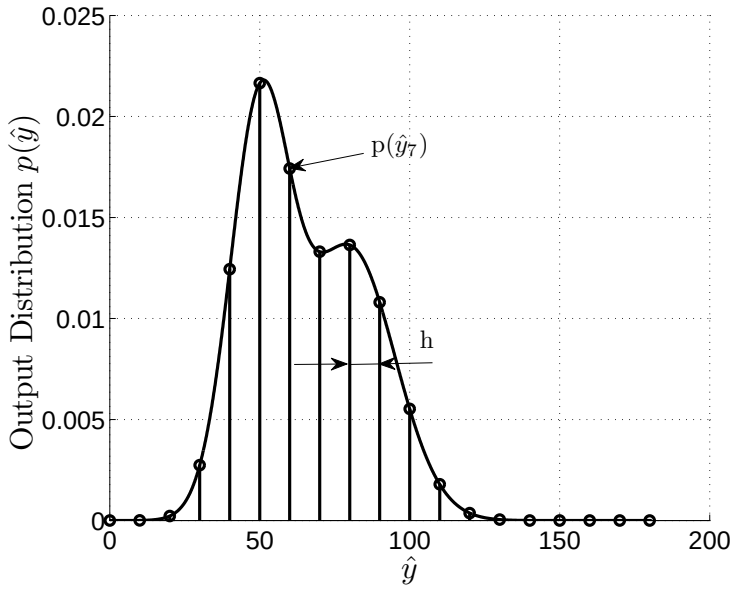


Figure 3.3: Quantized output distribution $p(\hat{y})$

distributions $p(x)$ and $p(y|x)$ are not fixed, then the mutual information is non-convex. The list of particles, which achieve the capacity, can be written as

$$\mathcal{P}_X = \{(\hat{x}_1^*, w_1^*), (\hat{x}_2^*, w_2^*), \ldots, (\hat{x}_N^*, w_N^*)\} = \arg\max_{\hat{x},w} I(\hat{x}, w) \tag{3.44}$$

As it is suggested in [Dau06], one can solve 3.44 by an alternating maximization:

$$\begin{aligned} W - \text{step}: \quad & w^{(k)} = \arg\max_{w} I(\hat{x}^{(k-1)}, w) \\ X - \text{step}: \quad & \hat{x}^{(k)} = \arg\max_{\hat{x}} I(\hat{x}, w^{(k)}) \end{aligned} \tag{3.45}$$

where k is the iteration-index for the main loop. This means that the weights $w^{(k)}$ in the k-th step are computed with the positions of the previous step $\hat{x}^{(k-1)}$ and the positions $\hat{x}^{(k)}$ are computed with the weight $w^{(k)}$ from the same step. In the W-step, the weights are optimized to maximize the mutual information while the positions are kept fixed. This is a concave optimization problem and can be carried out by the Blahut-Arimoto algorithm, due to the fact that the input alphabet $\hat{x}^{(k-1)} \in \mathcal{X}$ is discrete.

In the X-step, the weights are kept fixed while the positions are optimized to maximize the mutual information. Given that this is, as mentioned before, a non-convex optimization, the maximization in 3.45 is difficult to compute. Instead, one can select the positions of the particles such as

$$D(p(\hat{y}|\hat{x}_i^{(k)})||p^{(k)}(\hat{y})) + s^{(n)}(\hat{x}_i^{(k)})^2 \geq D(p(\hat{y}|\hat{x}_i^{(k-1)})||p^{(k-1)}(\hat{y})) + s^{(n)}(\hat{x}_i^{(k-1)})^2 \tag{3.46}$$

where

$$p(\hat{y})^{(k-1)} = \sum_{i=1}^{N} w_i^{(k)} p(\hat{y}|\hat{x}_i^{(k-1)}). \tag{3.47}$$

Eq. 3.46 shows the cost-function one has to maximize w.r.t. the positions. It is composed of the Kullback-Leibler divergence and the AP constraint function. The parameter s was adapted in the W-step for a fixed position. The selection of the positions can be obtained by applying some sort of gradient method, e.g. gradient ascent or steepest ascent.

3.2.4 Gradient Method

The general approach of a gradient method is to alternate between three steps [BV04]:

1. Determine an ascent direction $\Delta x^{(j)}$
2. Select the step size λ so that the Armijo rule is satisfied
3. Update the position values

Here j is the iteration-index for the gradient method. If we choose a gradient ascent, the alphabet $\hat{x}$ is updated according to the following rule

$$\begin{aligned}\hat{x}_i^{(j)} &= \hat{x}_i^{(j-1)} + \lambda \Delta x^{(j-1)} \\ &= \hat{x}_i^{(j-1)} + \lambda \frac{\partial}{\partial \hat{x}_i} \left(D(p(\hat{y}|\hat{x}_i)||\hat{p}^{(k)}(\hat{y})) + s^{(n)}(\hat{x}_i)^2 \right) \Big|_{w^{(k)}, \hat{x}_i^{(j-1)}}\end{aligned} \tag{3.48}$$

where the output PDF $p^{(k)}(\hat{y})$ depends on $p(\hat{y}|\hat{x})$ which is evaluated at $\hat{x}$. So it is important to keep $p(\hat{y}|\hat{x})$ constant for every separate gradient search iteration step. The adaptation of the step size λ can be done by the backtracking line search algorithm, whereas the step size is reduced until the Armijo rule

$$\begin{aligned}&D(p(\hat{y}|\hat{x}_i^{(j)})||p^{(j)}(\hat{y})) + s^{(n)}(\hat{x}_i^{(j)})^2 > \\ &D(p(\hat{y}|\hat{x}_i^{(j-1)})||p^{(j-1)}(\hat{y})) + s^{(n)}(\hat{x}_i^{(j-1)})^2 + \alpha\lambda \left(\Delta x^{(j-1)}\right)^2\end{aligned} \tag{3.49}$$

is satisfied for every step j . The partial derivation is given by:

$$\begin{aligned}&\frac{\partial}{\partial \hat{x}_i} \left(D(p(\hat{y}|\hat{x}_i)||p^{(j)}(\hat{y})) + s^{(n)}\hat{x}_i \right) \\ &= \sum_{l=1}^{H} \frac{\partial}{\partial \hat{x}_i} p(\hat{y}_l|\hat{x}_i) \Big(1 + \log \frac{p(\hat{y}_l|\hat{x}_i)}{p^{(j)}(\hat{y}_l)} - w_i \frac{p(\hat{y}_l|\hat{x}_i)}{p^{(j)}(\hat{y}_l)} \Big) + 2s^{(n)}\hat{x}_i^{(j)}\end{aligned} \tag{3.50}$$

As shown before, the AP constraint is directly integrated in the Blahut-Arimoto algorithm and in the cost-function used in the gradient method. The PA constraint

appears in the initial domain of the input alphabet $\hat{x}$, $0 < \hat{x}_i < A$. So before the update of position $\hat{x}$ in every gradient method step can be done, we have to verify that every new position $\hat{x}_i^{(j)}$ is in this domain.

Algorithm 2: Gradient search algorithm with backtracking line search

Initialize:
(1) take $\hat{x}^{(k-1)}$ from the previous X-step and $w^{(k)}$ from the W-step
(2) choose α and β for the backtracking line search algorithm
(3) $\lambda = 1$
(4) choose iteration number I
gradient iterations:
for $j = 1 : I$ **do**
  for every particle position:
  for $i = 1 : N$ **do**
    Step 1: compute: ascent direction
    $\Delta x_i{}^{(j-1)} = \frac{\partial}{\partial \hat{x}_i} f(w^{(k)}, \hat{x}_i^{(j-1)})$
    $= \frac{\partial}{\partial \hat{x}_i} \left(D(p(\hat{y}|\hat{x}_i)||\hat{p}^{(j-1)}(\hat{y})) + s^{(n)}(\hat{x}_i)^2 \right) \Big|_{w^{(k)}, \hat{x}_i^{(j-1)}}$
    Step 2: check if feasible point
    while $0 > (\hat{x}_i^{(j-1)} + \lambda \Delta x_i{}^{(j-1)})$ *and* $(\hat{x}_i^{(j-1)} + \lambda \Delta x_i{}^{(j-1)}) > A$ **do**
      $\lambda = \beta\lambda$
    end
    Step 3: backtracking line search
    while $f(w^{(k)}, \hat{x}_i^{(j-1)} + \lambda \Delta x_i{}^{(j-1)}) < f(w^{(k)}, \hat{x}_i^{(j-1)}) + \alpha\lambda \left(\Delta x^{(j-1)}\right)^2$ **do**
      $\lambda = \beta\lambda$
    end
    Step 4: update:
    (1) $\hat{x}_i^{(j)} = \hat{x}_i^{(j-1)} + \lambda \Delta x_i{}^{(j-1)}$
    (2) $\lambda = 1$
    (3) compute: $p(\hat{y}|\hat{x}_i^{(j)})$ and $\hat{p}^{(j)}(\hat{y})$
  end
end

The parameter α is typically chosen between 0.01 and 0.3. The parameter β is often chosen between 0.1 and 0.8, where a higher value means that the searching is more accurate [BV04].

3.2.5 The Entire Algorithm

Given that we know that the input distribution is discrete with a finite number of mass points, we can reduce the computation time by discarding particles, which have very low probability $w^{(k)}$, and fuse particles, if their positions $\hat{x}^{(k)}$ are close enough to each other. To define the thresholds for these methods is not a trivial task. One must have a prior knowledge of the capacity-achieving input distribution, which is defined by $\hat{x}$ and w after the upper bound $U^{(n)}$ and the lower bound $I^{(n)}$ are converged, which means

$$\begin{aligned} & I^{(n)} \leq C(P,A) \leq U^{(n)} \\ & I^{(n)} \leq C(P,A) \leq \max_{x \in \mathcal{X}} \Big[D(p(\hat{y}|\hat{x}^{(n)})||\hat{p}(\hat{y})^{(n)}) + s(\hat{x}^{(n)})^2 \Big] - s \sum_{i=1}^{N} w_i \hat{x}_i^{(n)} \\ & \rightarrow \qquad U^{(n)} - I^{(n)} < \epsilon. \end{aligned} \tag{3.51}$$

Dependent on the SNR, the channel parameters M and σ^2, and the input alphabet constraints, there are more or less mass points necessary to achieve the channel capacity. E.g., if the SNR is large, the capacity achieving input distribution consists of many mass points. Due to this, the space between the mass points is small and some of the mass points have a very low probability, so we have to choose the thresholds very carefully. If the threshold for the probability is chosen too large, we might discard particles which could be relevant. In almost the same manner we would adulterate the achieving input-distribution, if we chose the threshold for the fusing too large.

There is also a difference between the threshold for the probability and the one for the fusion: On the one hand, if the threshold for the probability is selected small enough, so that for the simulated SNR region no achieving distribution mass points get rejected, we have no loss of accuracy. But on the other hand, no matter how small we select the threshold for the fusion, for some special values of SNR we fuse two points, which have to be separate, because by reducing the SNR, particular pairs of mass points approach each other until one mass point appears instead. But if the threshold is chosen small enough, this inaccuracy is of no consequence compared to the accuracy of the upper and lower bound, if the algorithm has converged.

The two new methods to improve the convergence time of the algorithm are called, “kick particles” and “fuse particles” method.

Algorithm 3: Extended Blahut-Arimoto algorithm for continuous channels

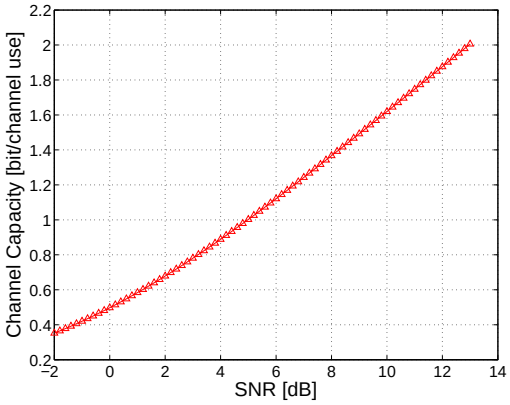
Initialize:
(1) define $\hat{x}^{(0)}$ and $w^{(0)}$ (uniformly distributed and spaced)
(2) define AP and PA constraint: P,A
(3) compute SNR and step h for numerical integration
(4) compute $p(\hat{y}|\hat{x})$
(5) choose Blahut-Arimoto algorithm and Gradient method iterations: BA-iter, GM-iter
(6) define w-thresh, x-thresh
while $U^{(k)} - I^{(k)} \leq \epsilon_c$ **do**
  Step 1: Blahut-Arimoto algorithm($\hat{x}^{(k-1)}$, $w^{(k-1)}$, $p(\hat{y}|\hat{x}^{(k-1)})$, BA-iter, P)
  Step 2: kick particles (only for discrete achieving input alphabets)
  for $i = 1 : N$ **do**
    if $w_i < w\text{-}thresh$ **then**
      $x_i = [empty]$
      $p(\hat{y}|\hat{x}_i^{(k-1)}) = [empty]$
      $w_i = [empty]$
    end
  end
  Step 3: fuse particles (only for discrete achieving input alphabets)
  for $i = 1 : N - 1$ **do**
    for $j = 1 + i : N$ **do**
      if $|x_i - x_j| < x\text{-}thresh$ **then**
        if $w_i > w_j$ **then**
          $w_i = w_i + w_j$
          $w_j = [empty]$
          $x_j = [empty]$
          $p(\hat{y}|\hat{x}_j^{(k-1)}) = [empty]$
        else
          $w_j = w_i + w_j$
          $w_i = [empty]$
          $x_i = [empty]$
          $p(\hat{y}|\hat{x}_i^{(k-1)}) = [empty]$
        end
      end
    end
  end
  Step 4: compute $p^{(k)}(\hat{y})$
  Step 5: gradient-search($\hat{x}^{(k-1)}$, $w^{(k)}$, $p(y|x)$, GM-iter, A)
  Step 6: update:
  (1) $p(\hat{y}|\hat{x}^{(k)})$
  (2) $I^{(k)} = \sum_{i=1}^{N} w_i D(p(\hat{y}|\hat{x}_i^{(k)})||\hat{p}^{(k)}(\hat{y}))$
  (3) $P^{(k)} = \sum_{i=1}^{N} w_i^{(k)} (\hat{x}_i^{(k)})^2$
  (4) $U^{(k)} = \max_{x \in \mathcal{X}} \left\{ D(p(\hat{y}|\hat{x}^{(k)})||\hat{p}^{(k)}(\hat{y})) + s^{(n)}(\hat{x}^{(k)})^2 \right\} - s^{(n)} P^{(k)}$
end

3.2.6 Example: Gaussian Channel with AP and PA Constraint

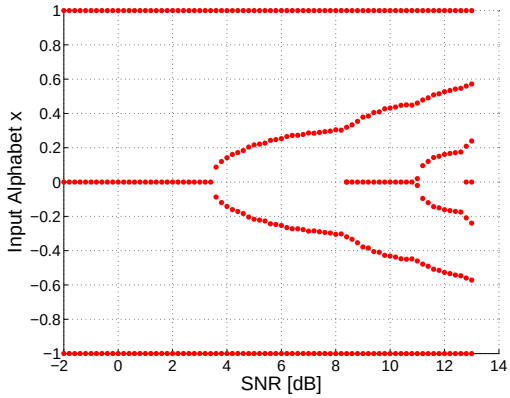
To verify the algorithm, the channel capacity of a Gaussian Channel was computed and compared to the result in [Dau05]. Fig. 3.4 shows the capacity and the achieving input distribution with a peak power to average power ratio (PAPR) $\rho = \frac{A^2}{P} = 2$ (amplitude A = 1 and average power P = 0.5). Fig. 3.4(a) shows the channel capacity over a wide range of SNR. As a matter of fact, the lower bound $I^{(n)}$ and upper bound $U^{(n)}$ are shown, but the difference between the bounds is negligibly small. The corresponding achieving input distribution is seen in Fig. 3.4(b). For every SNR value we get a PMF, where the dots represent the positions of the mass points. The probability of the mass points can be seen in Fig. 3.4(c).

This first example illustrate well that the algorithm computes the number and the positions of the mass points with their probability. On whatever channel the algorithm is applied, we always get the capacity and the input distribution with a certain accuracy, which depends on the step size of the numerical integration and, for continuous channels, on the initial number of mass points.

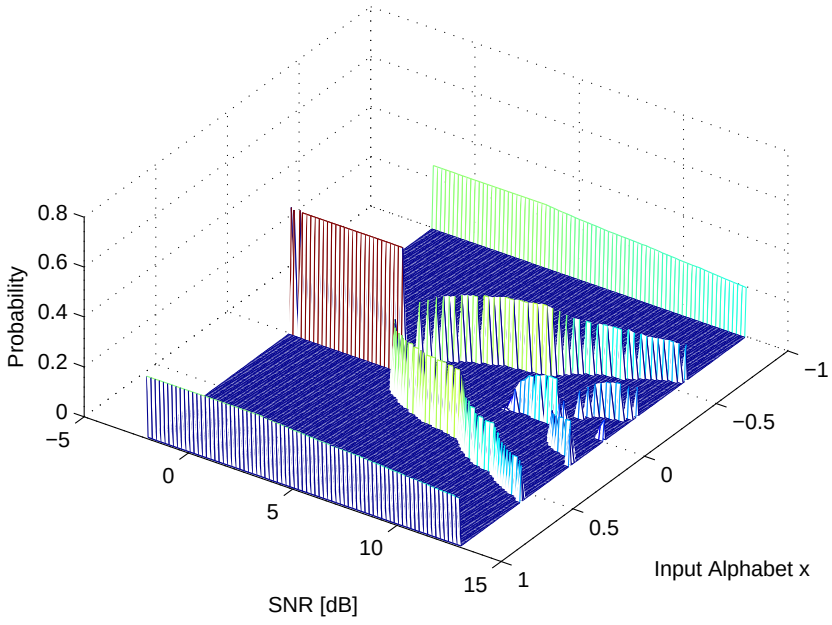
Another benchmark for the algorithm was a Gaussian channel with AP constraint, which is available in closed form. As we already know, the channel capacity of that channel is given by Eq. 3.7. It turns out that the simulated upper and lower bound and the analytical calculated channel capacity are practically the same.



(a) Channel capacity



(b) Positions of the mass points of the achieving input distribution



(c) 3-D Plot of achieving input distribution

Figure 3.4: Channel capacity and capacity-achieving input distribution of the Gaussian channel for $\rho = 2$

Chapter 4

Discrete Capacity-Achieving Input Distribution

Shannon [Sha48] already demonstrated in 1948 that, in case of an AP constraint, an additive white Gaussian noise channel has a continuous Gaussian RV as capacity-achieving input distribution. It is obvious that such distributions can not be realized in practice, since the input symbols have an unbounded amplitude. Therefore, a constraint on the PA has to be introduced, in addition to the constraint on the AP. Such considerations lead us to a class of capacity-achieving input distributions which is discrete and has in many cases a finite number of mass points. Especially the latter have a great importance for practical communication systems.

Already in 1971 Joel G. Smith (as quoted in [AFTS01], [SBD95], [CHK05]) delivered the proof that the capacity of a scalar additive white Gaussian noise channel with AP and PA constraints is achieved by a discrete input distribution with a finite number of mass points.

Honary and Darnell propose a numerical solution for computing the capacity-achieving input distribution in [HAD90] for such channels, if the output signal additionally is quantized and can therefore be represented by a finite number of mass points at the quantized output amplitude positions.

Shamai and Bar-David show in [SBD95] that a quadrature Gaussian channel with PA (norm), i.e. peak power and AP constraint, has a capacity-achieving input alphabet with a discrete amplitude (norm), a finite number of mass points, and an independent uniformly distributed phase.

In [CHK05] the capacity-achieving input probability under a PA and an AP constraint is studied for conditionally Gaussian channels, with input signal-dependent mean and, in general, input signal-dependent covariance. Examples for communication channels which fall into this category, are classical white Gaussian noise channels, multiple-input multiple-output (MIMO) Rayleigh-fading channels, some interference channels, and optical channels with signal-dependent noise (the latter has a similar model compared to the Gaussian approximation model of the energy detector, except that the noise depends linearly on the input x and not quadratic

x^2 like at the energy detector). This paper provides an extensive mathematical framework to show that the channel capacity is achievable and to derive necessary and sufficient conditions for a probability function to be capacity-achieving. Furthermore, under certain conditions, the achieving input distribution is proved to be discrete. Additionally, with this framework several practical channels are analysed, especially the optical channel with signal-dependent noise, for which an algorithm is proposed to find the discrete capacity-achieving input distribution. If one takes a look at the observations from the computed achieving input alphabet of the amplitude, one will discover many similarities to the achieving input alphabet of the energy detector (as we will see in Chapter 5), e.g.

1. there are always mass points at zero amplitude $x = 0$ and maximum amplitude $x = A$
2. the significant variations in the distance between two neighbour mass points by changing the SNR
3. the decreasing number of mass points by decreasing the SNR
4. the fact that with smaller SNR the probability of the mass point at $x = 0$, but also the probability of the mass point at $x = A$ increases, which can be explained by the increase of the "average distance" between the codewords.

Abou-Faycal and Fahs [AFF10] have studied the optimal input distribution for non-linear channels with additive white Gaussian noise, proved them to be discrete and two of them to be discrete with a finite number of mass points. To achieve this, they used mathematical tools based on standard decompositions in a Hilbert space and developed a method to use Hermite polynomials to study the capacity-achieving input distributions of such channels. Generally, it becomes evident that non-linear channels with at least an AP-constrained input signal, have capacity-achieving input distributions with discrete character. Under the additional PA constraint the discrete character is also given. Due to the non-linear character of the energy detector, one concludes that the achieving input distribution is also discrete.

The following provides further results of a literature research about capacity and (discrete) capacity-achieving input distribution of channels with bounded and average constraint input, in order to better understand the fact that the achieving input distribution of the energy detector is also discrete. I will rudimentarily present the mathematical approaches of two papers to prove that the capacity-achieving input distribution for some channels is discrete with a finite number of mass points.

4.1 Kuhn-Tucker Condition

The Kuhn-Tucker conditions are a generalization of the Lagrange method for finding the extrema of functions subject to a family of constraints. Where the Lagrange

method considers only equality constraints, the Kuhn-Tucker conditions allow also inequality constraints. The following section shows rudimentarily the derivation of the Kuhn-Tucker condition of a general channel with AP-constrained input distribution and is based on the appendix of [AFTS01]. Many other papers about this topic refer to the results of [AFTS01] and use the Kuhn-Tucker condition to find the properties of the input distribution.

To recall the maximization problem, the channel capacity is the supremum

$$C = \sup_{F \in \mathcal{F}} I(X;Y) = \sup_{F \in \mathcal{F}} \int \int p(y|x) \log \frac{p(y|x)}{p(y;F)} dy dF(x) \tag{4.1}$$

of the mutual information between input X and output Y over all input distributions $F \in \mathcal{F}$ that meet the constraint, where

$$p(y;F) = \int p(y|x) dF(x). \tag{4.2}$$

The existence and uniqueness of the input which achieves the maximum, result from the following theorem of optimization:

Theorem: If I is a real-valued, continuous function on a compact set $\Omega \subseteq X$, then I achieves its maximum on Ω. If furthermore Ω is convex, and I is strictly concave, then the maximum

$$C = \max_{F \in \Omega} I(F) \tag{4.3}$$

is achieved by a unique F_0 in Ω. The complete proof of this theorem is given in the appendix of [AFTS01] and leads

1. to the fact that the capacity is achieved by a unique input distribution and
2. to the Kuhn-Tucker condition, which applies to any average (or peak power) constrained channel for which the mutual information is continuous and differentiable.

The following listing shortly describes the steps of proof

- Show that the set of $\Omega \subset \mathcal{F}$ distribution function of non-negative RVs ($F(0^-) = 0$) with second moment constraint ($\int_0^\infty x^2 dF(x) \leq P$) is convex and compact.
- Show that the mutual information is a continuous and concave function.
- Use the Lagrangian theorem to integrate the AP constraint P into the optimization problem

$$C = \sup_{F \in \mathcal{F}} \{I(F) - \gamma g(F)\} \tag{4.4}$$

where γ is a Lagrange multiplier and

$$g(F) = \int_0^\infty x^2 dF(x) - P. \tag{4.5}$$

F_0 achieves the supremum 4.4, then $\gamma g(F_0) = 0$.

- Show the differentiability of the mutual information and the average power constraint in a convex space.
- Show that

$$\int p(y|x) \log \frac{p(y|x)}{p(y; F_0)} dy \leq C + \gamma(x^2 - P), \qquad \forall x \tag{4.6}$$

and

$$\int p(y|x) \log \frac{p(y|x)}{p(y; F_0)} dy = C + \gamma(x^2 - P), \qquad \forall x \in E_0 \tag{4.7}$$

where E_0 are mass points of F_0.

So we have shown that a random variable X^* with distribution F^* achieves the capacity of the channel with average power constraint if, and only if, there exists a $\gamma \geq 0$ so that

$$\gamma(x^2 - P) + C - \int p(y|x) \log \frac{p(y|x)}{p(y; F^*)} dy \geq 0 \tag{4.8}$$

for all x, with equality if x is in the support of X^*. Eq. 4.8 is called the Kuhn-Tucker condition.

4.2 Examples: Channel Capacity and Optimal Input Distributions

4.2.1 The Rayleigh-Fading Channel [AFTS01]

The memoryless discrete-time Rayleigh-fading channel model can be written as

$$V = AU + N \tag{4.9}$$

where U is the channel input, V is the channel output, and the fading parameter A and the noise N are a complex Gaussian RV with zero mean and variance σ_A^2 and σ_N^2 (phase is uniform distributed).

Subject to an AP constraint $E\{|U|^2\} \leq P_U$, here the capacity-achieving input distribution is studied and proved to be discrete with a finite number of mass points. Before analysing the mathematical methods to find the optimal input distribution, the first obvious step is to define the channel model, i.e. setting up the channel

transition PDF, which is a conditional output distribution given the channel input. The statistical model of the Rayleigh-Fading channel can be described by

$$p(v|u) = \frac{1}{\pi(\sigma_A^2|u|^2 + \sigma_N^2)} \exp\left\{\frac{-|v|^2}{\sigma_A^2|u|^2 + \sigma_N^2}\right\} \tag{4.10}$$

By a transformation of the RVs to $Y = |V^2|/\sigma_N^2$ and $X = |U|\sigma_A/\sigma_N$ and further to $S = 1/(1+X^2)$ we get

$$p(y|s) = se^{-ys}, \qquad s \in (0,1] \tag{4.11}$$

where $s = 1/(1+x^2)$ with the corresponding AP constraint $E\{1/s - 1\} \leq P$. If the optimal input RV X^* is discrete, also $S^* = 1/(1+X^*)$ is discrete.

The general idea is to find a sufficient and necessary condition for the input distribution to be optimal. With this knowledge one can verify whether the capacity of the channel can be achieved at all and what properties the optimal input alphabet must have. These properties can be shown by the Kuhn-Tucker condition 4.8, which can be written for the Rayleigh channel 4.11 as

$$\gamma\big(\frac{1}{s} - 1 - P\big) + C - \log s + 1 + \int_0^\infty se^{-sy} \log p(y; F^*)dy \geq 0 \tag{4.12}$$

where λ is the Lagrange multiplier and with equality, if s is in the support of S^*.

The discrete character of X^*

The optimal input RV X^* can have the following properties:

1. Have an infinite number of points of increase (mass points) on a bounded interval
2. Be discrete and have an infinite number of points of increase, but only finitely on a bounded interval
3. Be discrete and have a finite number of points of increase

The basic idea of proving the discreteness of the input distribution is to rule out cases 1. and 2., so that only the possibility of a discrete distribution with finite mass points remains.

Case 1: Infinite number of points of increase on a bounded interval

This assumption is satisfied by a continuous distribution. It is suggested to extend the Left-Hand-Side (LHS) term of 4.12 to the complex domain ($p(y; F^*) = p(y)$)

$$h(z) = \gamma\big(\frac{1}{z} - 1 - P\big) + C - \log z + 1 + \int_0^\infty ze^{-zy} \log p(y)dy \tag{4.13}$$

where $s = \Re\{z\}$. It is shown in [AFTS01] that $h(z)$ is equal to zero if z is in the support of S^*, not only on $z \in (0, 1]$ but also in general for all complex domain $s \in D$. For equality, 4.12 can be rewritten as

$$\int_0^\infty se^{-sy} \log p(y) dy = -\frac{1}{s}\big[\gamma\big(\frac{1}{s} - 1 - P\big) + C - \log s + 1\big] \tag{4.14}$$

where the LHS can be interpreted as the Laplace-transform of $s \log p(y)$. The inverse transform can be easily done by a table of Laplace transform pairs, so that

$$\begin{aligned} \log p(y) &= -\gamma y + \gamma(1 + P) - C - 1 - E_C - \log y \\ \rightarrow \qquad p(y) &= K\frac{e^{-\gamma y}}{y}, \qquad y \in (0, \infty). \end{aligned} \tag{4.15}$$

Due to the fact that the integral over $y \in (0, \infty)$ is infinite, 4.15 can not be a probability function. This rules out continuous input RV.

Case 2: Infinite number of points of increase but finitely many on a bounded interval

S^* can be seen as a RV with a countable but infinite number of mass points, so that $Pr\{S^* = s_i\} = p_i \neq 0$. The output density is given as

$$p(y) = \sum_{i=0}^{\infty} p_i s_i e^{-y s_i} \tag{4.16}$$

which has the lower bound $p(y) > p_i s_i e^{-y s_i}$ and therefore

$$\int_0^\infty se^{-ys} dy > \log(p_i s_i) - \frac{s_i}{s}. \tag{4.17}$$

With 4.17 the lower bound of the LHS of 4.12 can be written as

$$LHS > \gamma\big(\frac{1}{s} - 1 - P\big) + C - \log s + 1 + \log(p_i s_i) - \frac{s_i}{s}. \tag{4.18}$$

For $s \to 0$ the lower bound diverges to ∞, which means that the Lagrangian parameter γ has to be zero. This means that the constraint is inactive, which is impossible for the fading channel. Consequently, the 2. case is also ruled out. So the input distribution has to be discrete with a finite number of mass points.

4.2.2 The non-coherent Rician-Fading Channel [GPV05] and [GPV03]

Subject to an AP constraint ($E\{|x|^2\} \leq P$) and fourth moment constraint ($E\{|x|^4\} \leq \kappa P^2$) and, separately to a PA constraint[1] A, in [GPV05] and [GPV03] the capacity-achieving input distribution is proven to be discrete with a finite number of mass points. An interesting fact is that the channel PDF can be described by

$$p(y|x) = \frac{1}{\pi(\gamma^2|x|^2 + N_0)} \exp\left\{ \frac{|y - mx|^2}{\gamma^2|x|^2 + N_0} \right\} \tag{4.19}$$

i.e., a circular complex Gaussian PDF with a signal dependent variance. Here x is the complex channel input, y is the complex output, and a_i the fading parameter and the noise are independent complex circular Gaussian RV with zero mean and variances γ^2 and N_0. If the phase is not considered, this channel model can be considered as similar to the Gaussian approximation model of the energy detector. This leads us to the conclusion that the energy detector also has a capacity-achieving (amplitude) input distribution with finite mass points. If the phase of the input and output is considered as uniformly distributed and independent, $p(y|x)$ can be transformed into a Rician distribution

$$g_{R|r}(R|r) = \frac{1}{1 + r^2} \exp\left\{ -\frac{R + Kr^2}{1 + r^2} \right\} I_0\left(\frac{2\sqrt{K} r \sqrt{R}}{1 + r^2} \right) \tag{4.20}$$

where $R = (1/N_0)|y|^2$ and $r = (\gamma/\sqrt{N_0})|x|$ are the new RV and $K = |m|^2/\gamma^2$ is the Rician factor. The output distribution of the channel can be described by

$$f_R(R, F_r) = \int_0^\infty g_{R|r}(R|r) dF_r(r). \tag{4.21}$$

Here, the basic idea is to find, in a first step, a lower bound in the case of an AP and fourth moment constrained input alphabet and an upper bound in the case of a PA i.e. peak power constraint for the Kuhn-Tucker condition. In a second step it is shown that these bounds only are valid for a discrete input distribution with finite mass points. Since the approaches for the AP and fourth moment constraint and PA constraint are very similar, it suffices to describe one of the cases. The following should give a brief idea of the concepts presented in [GPV05] and [GPV03].

The Kuhn-Tucker condition for this channel and a PA-constrained input $x_i \leq A$ can be written as

$$\int_0^\infty g_{R|r}(R|r) \log f_R(R, F_0) dR + \log(1 + r^2) + C + 1 \geq 0, \qquad \forall r \in [0, A] \tag{4.22}$$

[1] As a third scenario the phase noise of the channel was also considered

with equality, if $r \in E_0$. F_0 is a capacity-achieving amplitude distribution and E_0 is the set of points of increase (mass points). To show the finite discrete character of F_0, one has to prove that

- the assumption of a distribution with an infinite number of points of increase on a bounded interval (continuous distribution)

contradict the Kuhn-Tucker condition, as it is done in section 4.2.1 for the Rayleigh-Fading channel. (The second case of section 4.12, infinite number of points of increase on an unbounded interval but finitely many on a bounded interval, is ruled out by the fact that the input signal amplitude is bounded.) To obtain a upper bound for the LHS of the Kuhn-Tucker condition, first a upper bound for $f_R(R, F_0)$ has to be found.

$$\begin{aligned} f_R(R, F_0) &= \int_0^A \frac{1}{1+r^2} \exp\left\{-\frac{R+Kr^2}{1+r^2}\right\} I_0\left(\frac{2\sqrt{K}r\sqrt{R}}{1+r^2}\right) dF_0(r) \\ &\leq \exp\left\{-\frac{R}{1+A^2} + \sqrt{KR}\right\} \int_0^A \frac{1}{1+r^2} \exp\left\{-\frac{Kr^2}{1+r^2}\right\} dF_0(r) \\ &= D_{F_0} \exp\left\{-\frac{R}{1+A^2} + \sqrt{KR}\right\} \end{aligned} \tag{4.23}$$

where $0 < D_{F_0} < 1$ is a constant depending on the optimal input distribution F_0. So one gets as upper bound for the LHS of the Kuhn-Tucker condition

$$LHS \leq \log D_{F_0} - \frac{1+(1+K)r^2}{1+A^2} + \sqrt{K}\sqrt{1+(1+K)r^2} + \log(1+r^2) + C + 1, \qquad \forall r \geq 0 \tag{4.24}$$

The LHS of the Kuhn-Tucker condition is extended in the complex domain (as in section 4.2.1)

$$h(z) = \int_0^\infty g_{R|z}(R|z) \log f_R(R, F_0) dR + \log(1+z^2) + \lambda(z^2 - P) + C + 1 \tag{4.25}$$

where $z \in \mathbb{C}$. If an optimal distribution is given, one has to assume that the Kuhn-Tucker condition should be satisfied with equality at the points of increase, and so $h(z) = 0$. Furthermore, it is shown (in [GPV03]) that $h(z) = 0$ is given in the whole region where it is analytic, and that $h(z)$ is analytic on the positive real line. Therefore, it is obvious that $h(r) = 0$, $\forall r \geq 0$. This is not possible, because if $r \to \infty$, the upper bound of the Kuhn-Tucker condition diverges. This rules out the case, infinite number of points of increase on a bounded interval. Consequently, the capacity-achieving amplitude distribution has to be discrete with finite mass points.

4.3 Application on The Energy Detector

With the mathematical approach of section 4.2.2, I want to give an idea of how a mathematical proof can be done for the energy detector. I want to emphasize that the following only reflects the ideas of [GPV03] and [GPV05] and should not be mistaken for a complete proof.

First of all, I assume the existence of a capacity-achieving input distribution for the energy detector. Furthermore, I take the Gaussian approximation as a channel model for the energy detector and consider the input alphabet only as PA constrained. With no loss of generality, the noise dimensionality M and variance of the noise σ can be considered as 1, so we get the following model

$$p(y|x) = \frac{1}{\sqrt{(4\pi(1+2x^2))}} \exp\left\{-\frac{1}{2}\frac{(y-(1+x^2))^2}{2+4x^2}\right\} \tag{4.26}$$

The Kuhn-Tucker condition for this channel and the PA constraint $0 \leq x \leq A$ can be written as

$$\int_{-\infty}^{\infty} p(y|x)\log p(y;F_0)dy + \frac{1}{2}\log[4\pi(1+2x^2)] + C + 1 \geq 0 \qquad \forall x \in [0,A] \tag{4.27}$$

where

$$p(y;F_0) = \int_0^A p(y|x)dF_0(x). \tag{4.28}$$

with equality, if $x \in E_0$. F_0 is a capacity-achieving amplitude distribution and E_0 is the set of points of increase (mass points). The next step is to find a upper bound for $p(y;F_0)$ written as

$$\begin{aligned} p(y;F_0) &= \int_0^A \frac{1}{\sqrt{(4\pi(1+2x^2))}} \exp\left\{-\frac{1}{2}\frac{(y-(1+x^2))^2}{2+4x^2}\right\} dF_0(x) \\ &\leq \int_0^A \frac{1}{\sqrt{(4\pi(1+2x^2))}} dF_0(x) \\ &= K_{F_0} \end{aligned} \tag{4.29}$$

where K_{F_0} is a constant which depends on the optimal input distribution F_0. The inequality can be obtained by

$$\exp\left\{-\frac{1}{2}\frac{(y-(1+x^2))^2}{2+4x^2}\right\} \leq 1 \qquad \forall x \geq 0 \tag{4.30}$$

Using the upper bound 4.29, the integral in the Kuhn-Tucker condition can be upper bounded as

$$\int_{-\infty}^{\infty} p(y|x) \log p(y; F_0) dy \leq \log K_{F_0} \tag{4.31}$$

where the second integral is the second moment and the third integral is the first moment of a Gaussian distributed RV. With 4.31 a upper bound on the LHS of the Kuhn-Tucker condition is derived by

$$LHS \leq \log K_{F_0} + \frac{1}{2} \log[4\pi(1 + 2x^2)] + C + 1 \qquad \forall x \geq 0. \tag{4.32}$$

If we assume that we can proceed like in subsection 4.2.2 for the Rician-fading channel, then we extend the Kuhn-Tucker condition in the complex domain and assume an input distribution with infinite mass points on a bounded interval. That way it can be shown that $h(z) = 0$ for all positive-real z, which means that $h(x) = 0$ $\forall x > 0$. But for $x \to \infty$ the upper bound of the LHS of the Kuhn-Tucker condition is diverging (the second term $\frac{1}{2} \log[4\pi(1+2x^2)] \to \infty$ diverge), which is a contradiction and so we can rule out alphabets with infinite mass points on a bounded interval. This leads to the conclusion that the capacity-achieving input distribution of the Gaussian approximation model (of the energy detector) with PA-constrained input distribution is discrete and has a finite number of mass points.

Chapter 5

Simulation Results and Evaluation

In this chapter it is time to pay attention to the simulation of the energy detector for both established mathematical models. For all simulations a constant input signal $x[n]$ is assumed. The SNR on the input of the energy detector is given by $SNR = \frac{P}{\sigma^2}$, where $P = E\{x[n]^2\}$ is the AP of the input signal and σ^2 is the variance of the additive Gaussian noise. The PA constraint is considered to be $0 \leq x \leq A$. An interesting fact for this input signal scenario is, that the performance of the energy detector, i.e. the channel capacity, increases with increasing integration time T_i, i.e. the noise dimensionality M. The simulations will show us how M and the PAPR ρ influence the channel capacity C and of course the capacity-achieving input distribution $p(x)$ over a wide range of SNR. Furthermore, a comparison between the exact statistic of the energy detector, the non-central χ^2 distribution and the approximated statistic, the Gaussian distribution with signal dependent variance, will show us the quality of the approximation model.

Another important fact that will be discovered, is that the capacity achieving input distribution is a PMF with finite mass points (discrete and finite) and that mass points at zero and maximum amplitude are always expected, as it was discussed in detail in the previous chapter.

Finally, the correctness of the two statistical models of the energy detector will be verified by some simple evaluation models, which are fed with capacity achieving input distributions. The performance parameter will be the mutual information between the achieving distribution and the output of the evaluation model, which has to be roughly the same value as the channel capacity.

The first step in this chapter will be to give a short analysis of the algorithm itself and of the convergence behaviour of the algorithm. Especially the impact of the "kick particles" and "fuse particles" method on the convergence time will be analysed.

5.1 The Algorithm

Fig. 5.1 shows the cost-function $D(p(\hat{y}|\hat{x})||p(\hat{y})) + s\hat{x}$ (blue curve) which is applied in both optimization steps, the gradient of the cost-function (magenta curve) and the positions and weights of the particles (red stem-plot), which represent the PMF of the input alphabet. Particularly in the Blahut-Arimoto step the cost-function is used to update the weights of the particles to reduce the gap between the maximum of the cost-function, which is the upper bound $U^{(n)}$, and the lower bound $I^{(n)}$. Furthermore, in the gradient step the cost-function is maximized by selecting the correct position of the particles, whereas the direction is given by the derivative of the cost-function.

Generally, we can say that the channel transition PDF with its parameters and the constraints of the input alphabet define the maxima of the cost-function. The Blahut-Arimoto algorithm shapes the weights of the particles in a way that the maxima coincide with the maxima of the cost-function. This means that the maxima of the cost-function are the positions of the mass points. The second step of the algorithm, the gradient method, moves the positions of the particles to the positions of the maxima of the cost-function. Since the algorithm uses a finite number of particles, i.e. mass points, the cost-function displayed has not the right shape, which means the maxima are not on the right positions. But with optimization of the weights and the positions, the maxima of the cost-function move to the right positions.

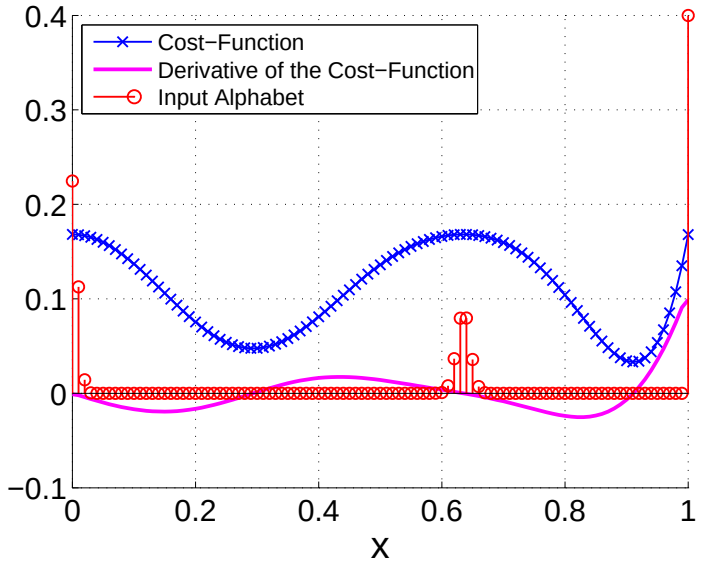
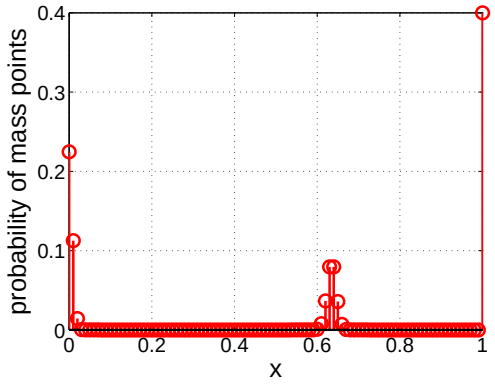


Figure 5.1: Cost-function for gradient method and the Blahut-Arimoto step, gradient of the cost-function, and positions and probabilities of the mass points after the first Blahut-Arimoto step for $M = 40$, $\rho = 2$, and $SNR = -0.2$ dB

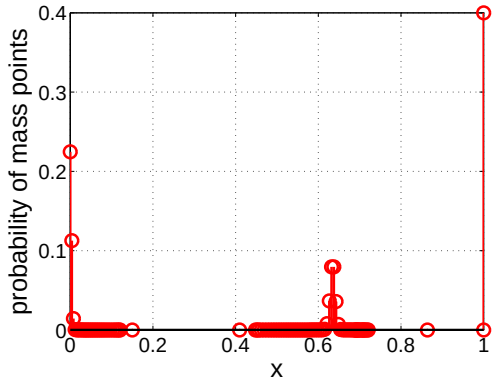
The algorithm starts with an initial list of mass points

$$\mathcal{P}^{(0)} = \left\{(\hat{x}_1^{(0)}, \hat{x}_1^{(0)}), (\hat{x}_2^{(0)}, \hat{x}_2^{(0)}), \ldots, (\hat{x}_N^{(0)}, \hat{x}_N^{(0)})\right\} \tag{5.1}$$

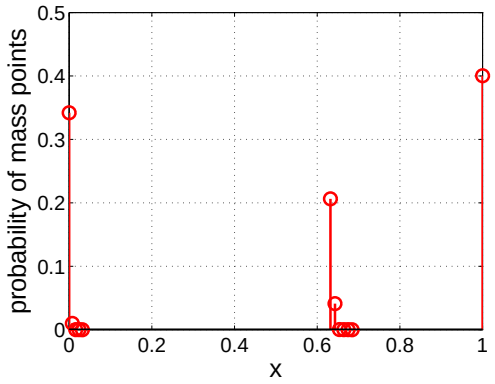
with uniformly distributed weights $w_i^{(0)} = \frac{1}{N}$ and uniform spacing between the mass points so that $\hat{x}_i^{(0)} = i\frac{A}{N}$, where N is sufficiently large. Fig. 5.2 shows the evolution of the weights and the positions of the particles over the steps of the algorithm. Fig. 5.2(a) shows the particles after the first Blahut-Arimoto step, Fig. 5.2(b) the particles after a gradient step, Fig. 5.2(c) after the "kick particles" and "fuse particles" methods and Fig. 5.2(d) the mass points after convergence of upper and lower bound, which represent the capacity-achieving input distribution. (Normally, the "kick particles" and "fuse particles" methods are executed before the gradient method, see algorithm 3.)



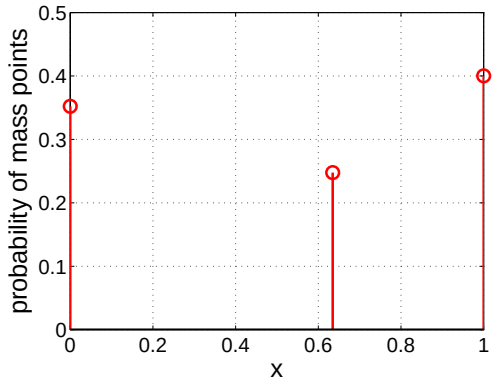
(a) Input alphabet after first Blahut-Arimoto step



(b) Input alphabet after first gradient method step

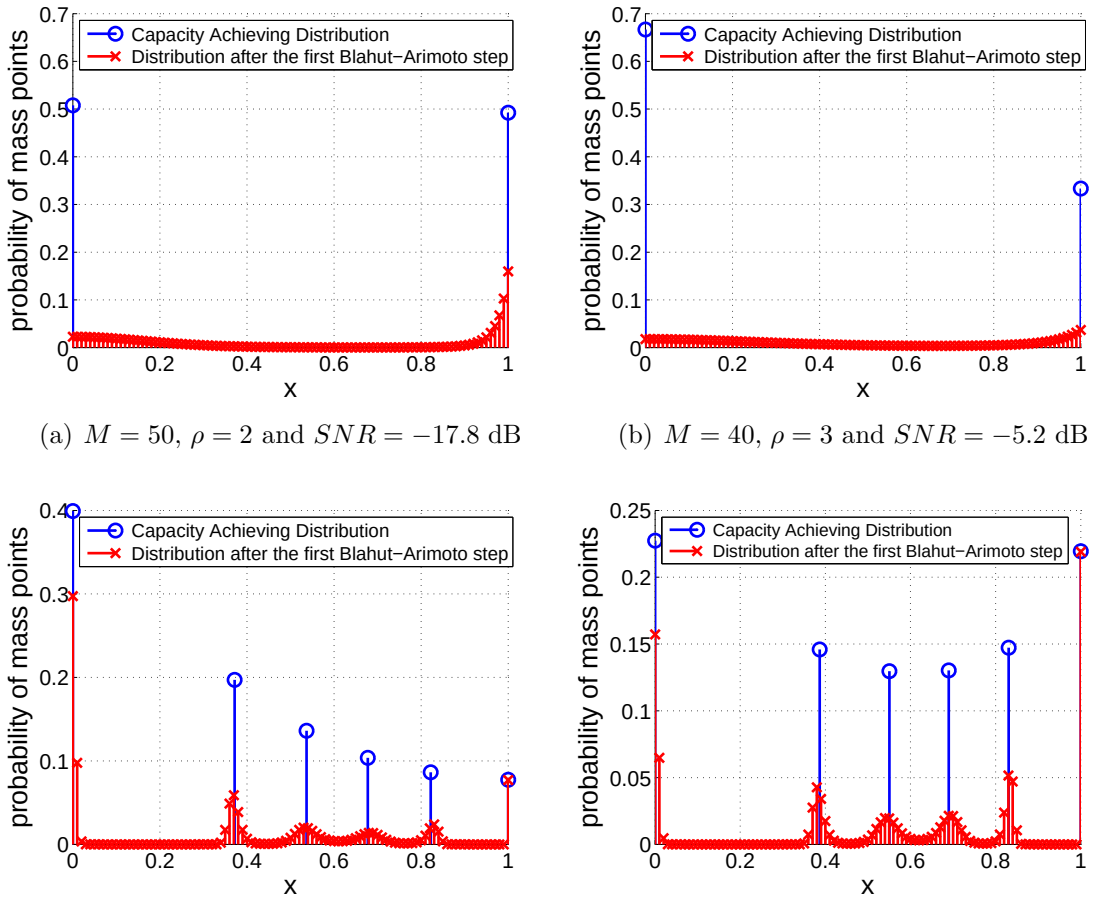


(c) Input alphabet after "kick particles" and "fuse particles" method



(d) Capacity achieving input alphabet after convergence in 6 iterations of main-loop

Figure 5.2: Evolution of the input alphabet over the algorithm steps for $M = 40$, $\rho = 2$, and $SNR = -0.2$ dB



(a) $M = 50$, $\rho = 2$ and $SNR = -17.8$ dB

(b) $M = 40$, $\rho = 3$ and $SNR = -5.2$ dB

(c) $M = 40$, $\rho = 4$ and $SNR = 1.4$ dB

(d) $M = 40$, $\rho = 2$ and $SNR = 0.2$ dB

Figure 5.3: Examples of capacity-achieving input distributions after the first Blahut-Arimoto step and after convergence of upper bound $U^{(n)}$ and lower bound $I^{(n)}$for several M values, PAPR ρ, and SNR values.

Fig. 5.3 shows some examples of capacity-achieving input distributions after the first Blahut-Arimoto step and after convergence of upper and lower bounds for different values of M, PAPR ρ, and SNR. The figure illustrates the discreteness of the input alphabet, and also that the probabilities and positions of the mass points depend on the noise dimensionality M, the PAPR ρ, and the SNR. Furthermore, it points out that there are always mass points at zero and maximum amplitude. Another interesting aspect is that, if ρ increases — and therefore the AP constraint gets stricter compared to the PA constraint — the mass points with bigger amplitude decrease in probability. If the SNR gets smaller, the probability of most separated mass points is increasing. This means (not really surprisingly), that the probability of the mass points at maximum amplitude also increases for a noisy environment, so that the symbols with the biggest distance are the most probable.

For $\rho > 2$ the AP is always active and the probabilities of the mass points with higher amplitudes show decreasing weights. This behaviour is not strict, because for SNR values for which new mass points appear, mass points with higher amplitude could have a higher probability than some with a lower amplitude. Another fact is, that the probability of the two remaining mass points does not change with changing SNR. E.g., the distribution $p(x) = 0.67\delta(x) + 0.33\delta(x - A)$ shown in Fig. 5.3(b) for $M = 40$ and $\rho = 3$ does not change with a further decrease of the SNR.
If $\rho \leq 2$, the AP constraint is not active, which means that

$$\sum_{i=1}^{N} \hat{x}_i^2 w_i < P. \tag{5.2}$$

In this case, the distribution of the mass points shows the tendency to be uniformly distributed. (seen in Fig. 5.3(a) and 5.3(d)). If the AP constraint is not active, the distribution for two mass points does change with changing SNR. As shown in Fig. 5.3(a) the input alphabet almost reached a uniform distribution $p(x) = 0.5\delta(x) + 0.5\delta(x - A)$ at $SNR = -17.8$ dB, but if one takes a closer look at the probabilities, one will see that $p(0) = 0.5077$ and $p(A) = 0.4923$. With decreasing SNR the probability of the two mass points approaches the value 0.5, until at a certain SNR a uniform distribution is reached, dependent on the PAPR ρ (≤ 2) and M (e.g. for $\rho = 2$ and $M = 40$ at $SNR = -49.999$ dB mass points are uniformly distributed).

Table 5.1 demonstrates, that for $\rho \leq 2$ the calculated average power over the computed input alphabet is smaller than the AP constraint. If the AP constraint is active ($\rho > 2$ and $s < 0$), the calculated average power shows the same value as the AP constraint.

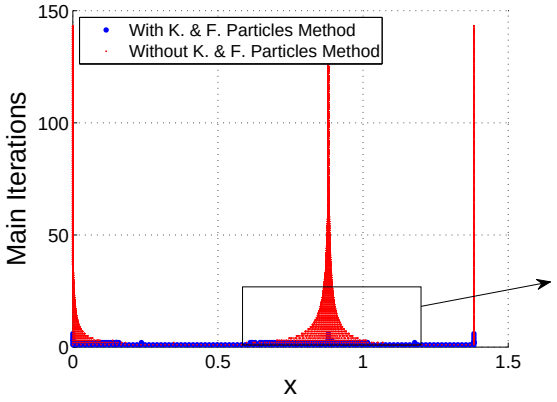
ρ	**parameter s**	**SNR [dB]**	**AP:** $P = 10^{\frac{SNR}{10}}$	**PA:** $A = \sqrt{\rho P}$	**calc. power**
1.5	0	-0.2	0.955	1.38	0.645
2	0	-0.2	0.955	1.38	0.889
3	-0.358	-0.2	0.955	1.38	0.955
4	-0.529	-0.2	0.955	1.38	0.955

Table 5.1: Comparison between the chosen AP constraint value and the calculated average power $\sum_{i=1}^{N} \hat{x}_i^2 w_i$ of the capacity-achieving input distribution for different values of ρ and $M = 40$

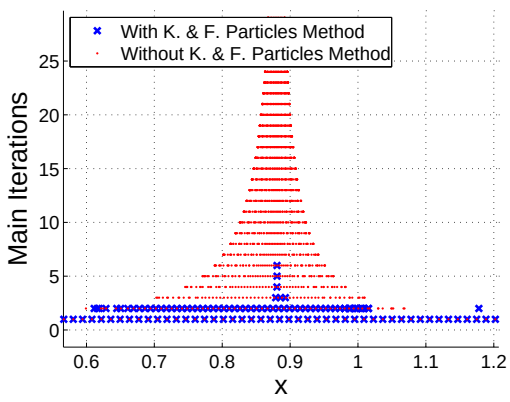
Convergence behaviour In order to make the algorithm work, the initial list of particles has to be sufficiently large. Even if the input alphabet is known to be discrete, the Blahut-Arimoto step needs a large enough number of particles to find all probability maxima. This chosen number of initial particles also affects the convergence time, as does the step size of the numerical integration, which defines the number of quantized output samples. The number of particles and of output samples defines in particular the size of the sampled channel transition PDF and of

course the length of the input and the output alphabet. This size significantly affects the convergence time, i.e. the number of iterations until upper and lower bounds converge. Not only the number of main-iterations increases, also the computation times of the W-step and X-step increases.

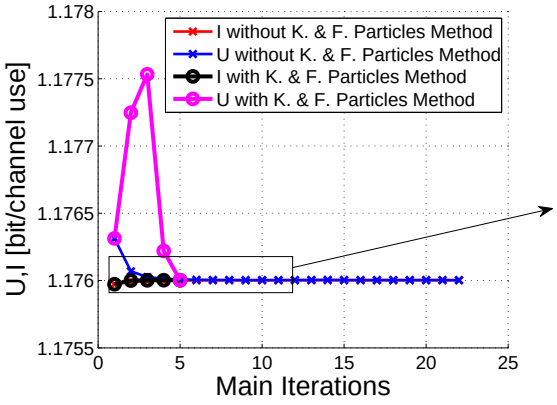
To reduce the convergence time, the number of particles can be decimated by the "kick particles" and "fuse particles" methods. Figs. 5.4(a) and 5.4(b) show the evolution of the positions of the particles $\hat{x}$ over the main-iterations to the positions f the mass points of the capacity-achieving input distribution. The figure illustrates that with the "kick particles" and "fuse particles" methods the gap between upper bound and lower bound is, after already 6 main-iterations, smaller than the one defined by the convergence bound $\epsilon = 10^{-6}$. In contrast, without these methods, the convergence bound is not reached until 143 main-iterations are executed. Figs. 5.4(c) and 5.4(d) show the convergence of the upper bound U and the lower bound I over the first 20 main-iterations. An interesting fact is that the bounds with the "kick particles" and "fuse particles" methods for the first two iteration steps diverge. This always appears if many particles are discarded. E.g., at the first iterations all particles which are not near an anticipated mass point, have a very low weight and get potentially discarded.



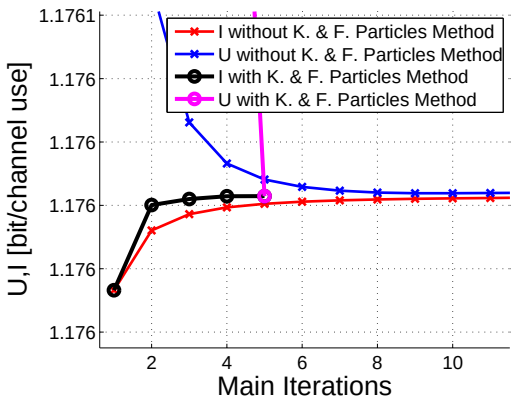
(a) Convergence of the particles



(b) Convergence of the particles (zoom)



(c) Convergence of the upper and lower bound



(d) Convergence of the upper and lower bound (zoom)

Figure 5.4: Convergence of the particle positions and of the upper bound $U^{(n)}$ and lower bound $I^{(n)}$ over the main-iterations of the algorithm

5.2 Gaussian Approximation

The following results for the Gaussian approximation and the non-central χ^2 model are simulated with a convergence parameter $\epsilon = 10^{-5}$, i.e. $U^{(n)} - I^{(n)} < \epsilon$, for higher SNR and with $\epsilon = 10^{-8}$ for lower SNR values, where the channel capacity is small. The reason for this is that the sensitivity of the mutual information for low SNR appears to be small compared to the exact number and location of the mass points as it also was mentioned in [AFTS01]. This generally means that neither a slightly wrong position and/or probability of single mass points, nor a representation of the mass points by more than one particle distributed around the real position of the achieving mass points, will affect the value of the channel capacity that strongly. Due to that fact, the convergence parameter $\epsilon = 10^{-5}$ does not lead to a good estimate of

the capacity-achieving input distribution for low SNR. Fig. 5.5 shows the channel

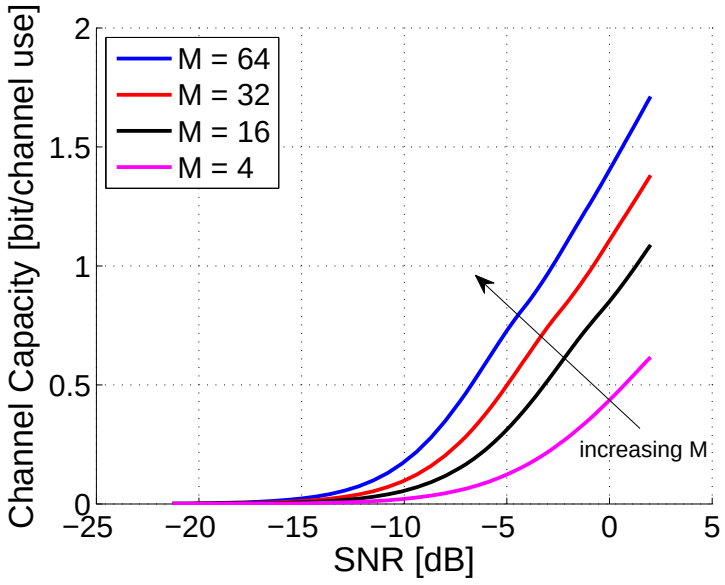


Figure 5.5: Capacity of the Gaussian approximation model for several values of M and $\rho = 2$

capacity of the energy detector represented by the Gaussian approximation model for $\rho = 2$ and several M values over a wide range of SNR. The input signal is assumed to be constant $x[n] = a$, so that the average power for the constraint is given by $P = Ma^2$. For this scenario the channel capacity is getting higher with increasing integration time T_i, i.e. noise dimensionality M, because more signal energy is captured. Fig. 5.6 shows the channel capacity for $M = 40$ and several ρ values over a wide range of SNR. The capacity is getting higher by increasing the PAPR ρ. This can be explained by the fact that for a constant $SNR = \frac{P}{\sigma^2} = \frac{A^2}{\rho\sigma^2}$, i.e. constant AP, an increasing PARP ρ leads to a higher maximum amplitude, so that the distance between the symbols is getting larger. The positions of the capacity-achieving input distribution mass points over the SNR are shown in Figs. 5.7(a)-(f), where the PAPR $\rho = 2$ and M is varied. Figs. 5.8 and 5.9 show the probabilities of the mass points for $M = 100$ and $\rho = 2$ and $M = 40$ and $\rho = 5$ respectively. A fact that we already has discussed in the previous section is that for a given PAPR ($\rho = 2$) the number of mass points of $p(x)$ is monotonically non-decreasing with the SNR and the noise dimensionality M, and there exists always a mass point at zero amplitude $\hat{x} = 0$ and a mass point at maximum amplitude $\hat{x} = A$. Another fact is that the distances between two neighbouring mass points vary significantly with changes of the SNR and that with smaller SNR the probability of the mass points at $\hat{x} = 0$ and $\hat{x} = A$ increases. This can see in Figs. 5.8 and 5.9, which can be explained by the increase of the "average distance" between the codewords (in a similar way as for the optical channel in Chapter 4). There are two ways new mass

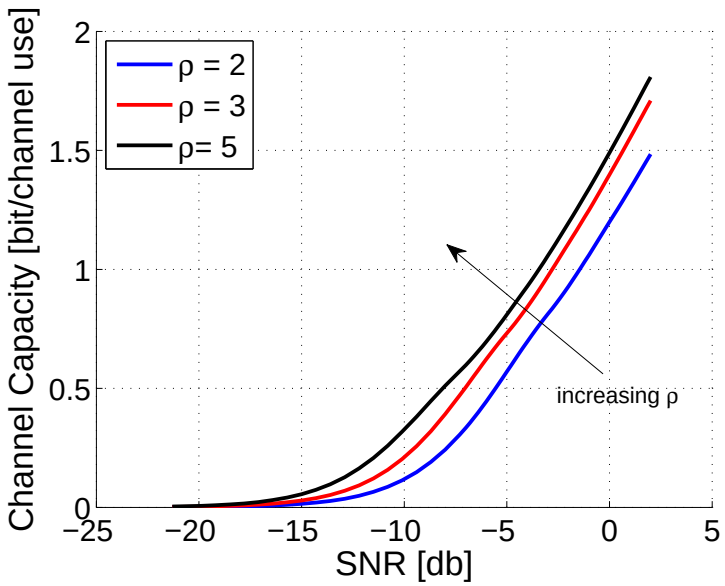


Figure 5.6: Capacity of the Gaussian approximation model for several PAPR ρ and $M = 40$

points can appear if the SNR increases:

1. a new mass point appears apart from the other mass points with a low probability,

2. or an existing mass point with relatively high probability splits up into two new mass points with low probabilities.

The relations between appearing new mass points and their probabilities are well illustrated in Fig. 5.8, which represents the 3-D plot of Fig. 5.7(f). For example, between $SNR = -5.4$ dB and $SNR = -5.2$ dB a new mass point appears and between $SNR = -2.6$ dB and $SNR = -2.4$ dB a mass point is split up into two new mass points. The exact values of the probabilities of the mass points in this SNR regions are listed in Table 5.1(a) and 5.1(b). As Figs. 5.7(a)-(f) show, the capacity-achieving input alphabets seem to have roughly the same shape for different M ($\rho = 2$) , but for increasing M alphabets with more mass points are shifted to lower SNR values. Table 5.2(a) gives the capacity-achieving input distribution for $M = 40$ and $M = 100$ for SNR values for which the first time three mass points appear. The comparison points out that the mass points for both alphabets have nearly the same probabilities, but the mass points (of course except $p(0)$) move slightly towards the mass point at maximum amplitude $x = A$. The PMF listed in Table 5.2(b) confirms the assumption by comparing two achieving PMF at another SNR value. For $\rho > 2$ the input alphabet changes with M in the same way as for $\rho \leq 2$, with two differences: With increasing PAPR ρ the input alphabet gets more and more distorted and the positions of the inner mass points (except $p(0)$

(a) Probabilities of mass points, if a new mass point appears (2 mass points → 3 mass points)

SNR	[dB]	-5.4	-5.2	-5	-4.8	-4.6
$p(x_0)$	[ ]	0.512	0.482	0.465	0.45	0.436
$p(x_1)$	[ ]	0.488	0.034	0.064	0.091	0.116
$p(x_2)$	[ ]	-	0.484	0.471	0.459	0.448

(b) Probabilities of mass points, if a mass point splits up into two new mass points (3 mass points → 4 mass points)

SNR	[dB]	-2.6	-2.4	-2.2	-2	-1.8
$p(x_0)$	[ ]	0.35	0.345	0.334	0.324	0.314
$p(x_1)$	[ ]	0.262	0.097	0.134	0.146	0.154
$p(x_2)$	[ ]	0.388	0.174	0.156	0.162	0.17
$p(x_2)$	[ ]	-	0.384	0.376	0.369	0.362

Table 5.2: Probabilities of new appearing mass points for $M = 100$ and $\rho = 2$

(a) 3 mass points

M	SNR	**PMF**
[ ]	[dB]	
40	-2.8	$p(x) = 0.482\delta(x) + 0.032\delta(x - 0.652) + 0.487\delta(x - 1)$
100	-5.2	$p(x) = 0.482\delta(x) + 0.034\delta(x - 0.671) + 0.485\delta(x - 1)$

(b) 4 mass points

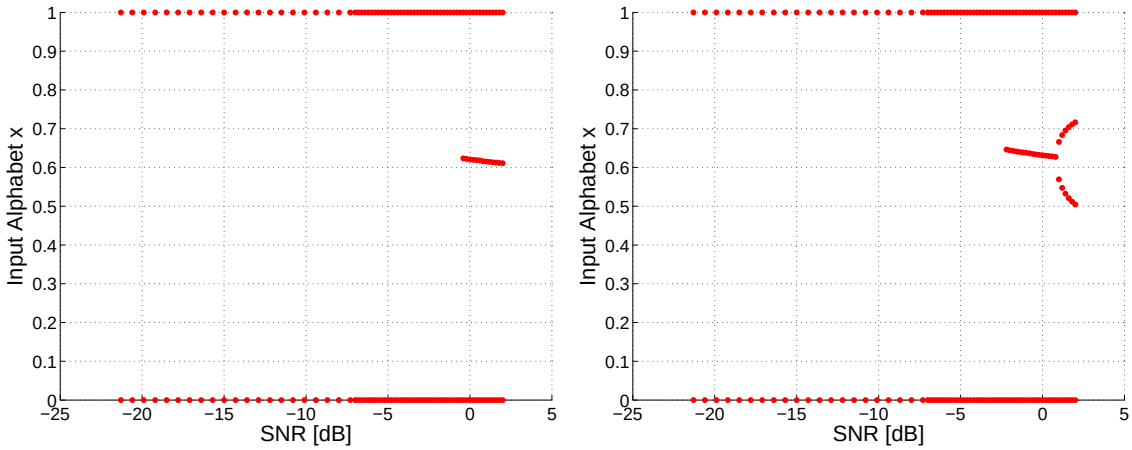
M	SNR	**PMF**
[ ]	[dB]	
40	2	$p(x) = 0.269\delta(x) + 0.182\delta(x - 0.493) + 0.203\delta(x - 0.733) + 0.347\delta(x - 1)$
100	-0.6	$p(x) = 0.269\delta(x) + 0.193\delta(x - 0.516) + 0.208\delta(x - 0.753) + 0.33\delta(x - 1)$

Table 5.3: Comparison between capacity-achieving input distributions for $M = 100$ and $M = 40$ for $\rho = 2$ and specific SNR values

and $p(A)$) advance stronger towards maximum amplitude (Figs. 5.10(a)-(d)). The second difference is that the larger ρ, the more mass points are at the same SNR to achieve the channel capacity. That a new mass points appears apart from the others is more likely than that an existing mass point splits up into two new ones, which is well illustrated in Fig. 5.9. Figs. 5.11(a)-(d) show the output distribution $p(y)$ (black curve) of the energy detector. To recall the calculation of the output distribution:

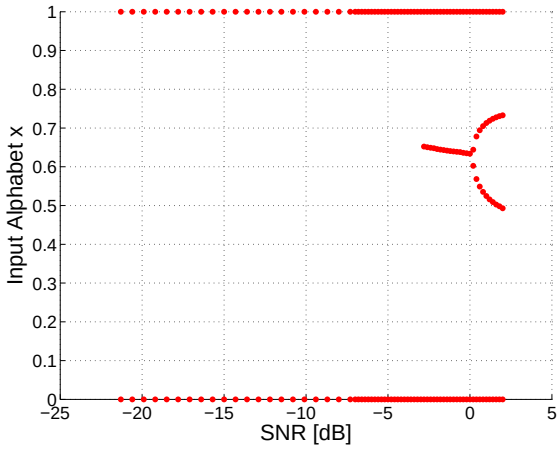
$$p(y) = \sum_{i=0}^{N} w_i p(y|x_i) \tag{5.3}$$

The coloured curves represent the channel transition PDFs $p(y|\hat{x}_i)$ evaluated at the positions of the capacity-achieving input mass points $\hat{x}_i$. These curves are weighted with the probability of the mass points w_i.

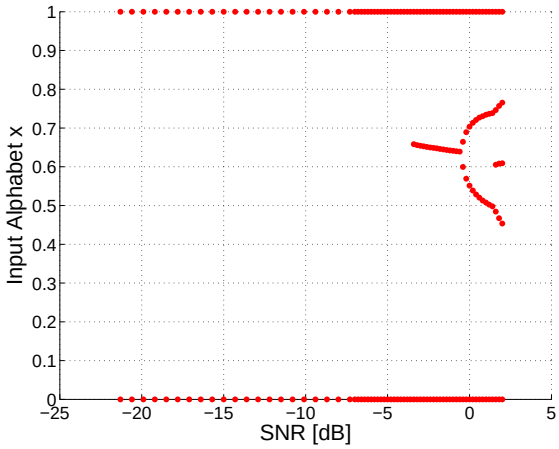


(a) M = 16 and $\rho = 2$

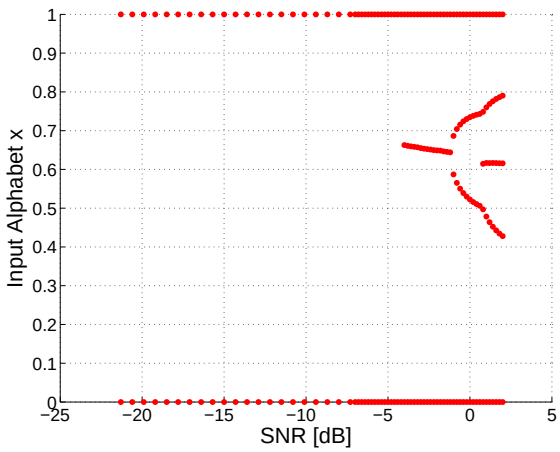
(b) M = 32 and $\rho = 2$



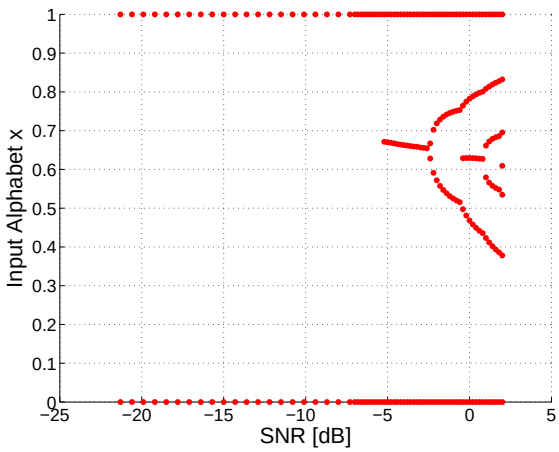
(c) M = 40 and $\rho = 2$



(d) M = 50 and $\rho = 2$



(e) M = 64 and $\rho = 2$



(f) M = 100 and $\rho = 2$

Figure 5.7: Positions of the mass points of the capacity-achieving input distribution of the Gaussian model for different M

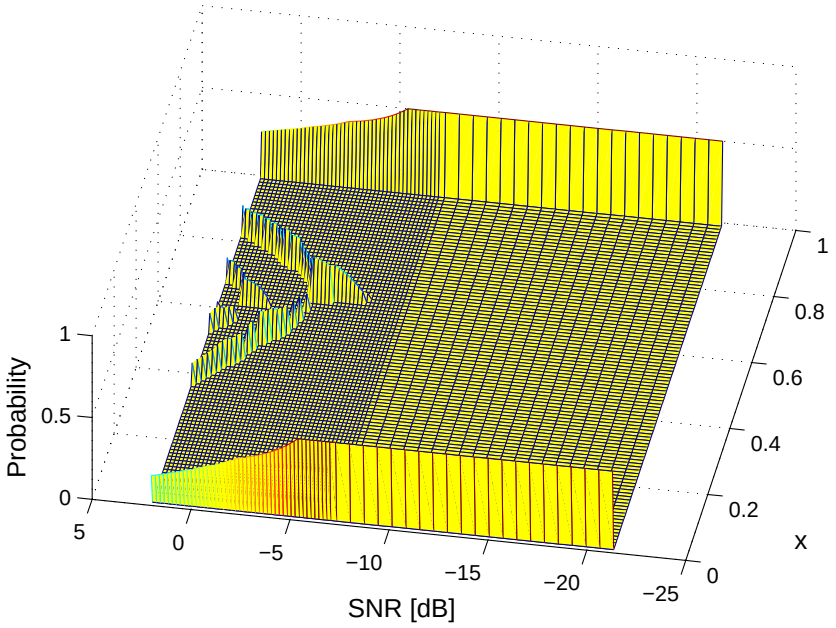


Figure 5.8: Capacity-achieving input distribution of the Gaussian approximation model for $M = 100$ and $\rho = 2$

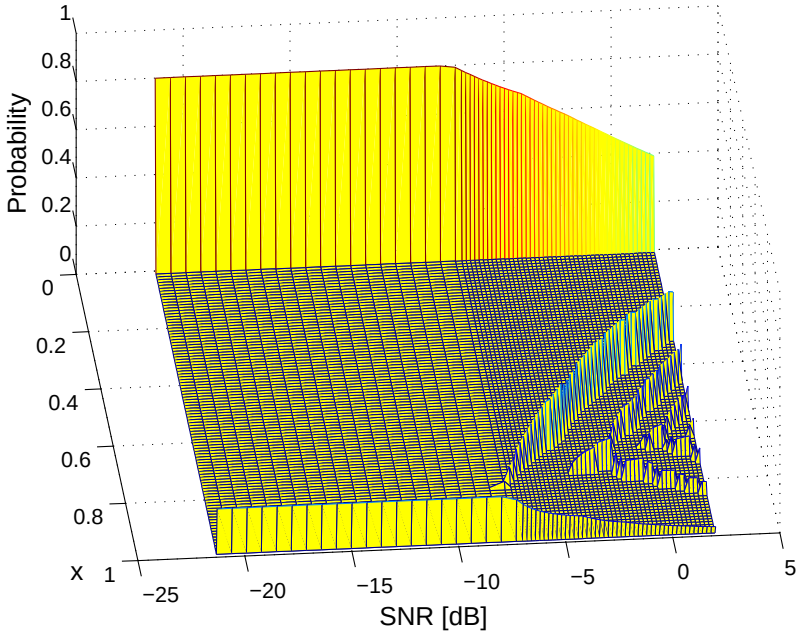
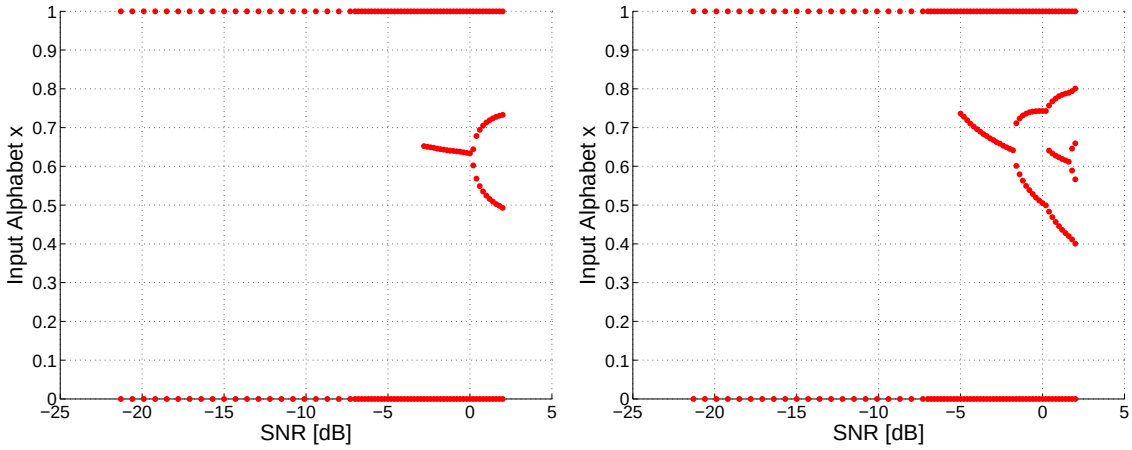
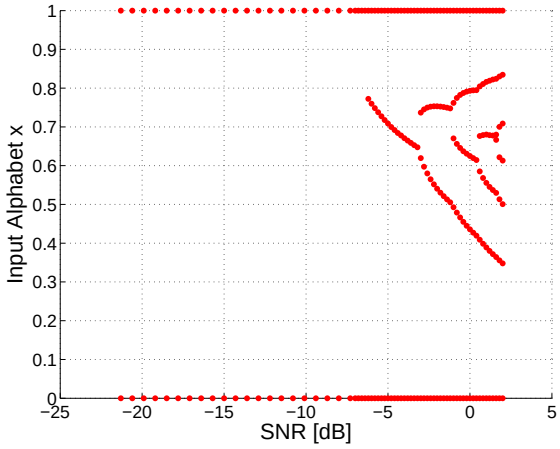


Figure 5.9: Capacity-achieving input distribution of the Gaussian approximation model for $M = 40$ and $\rho = 5$

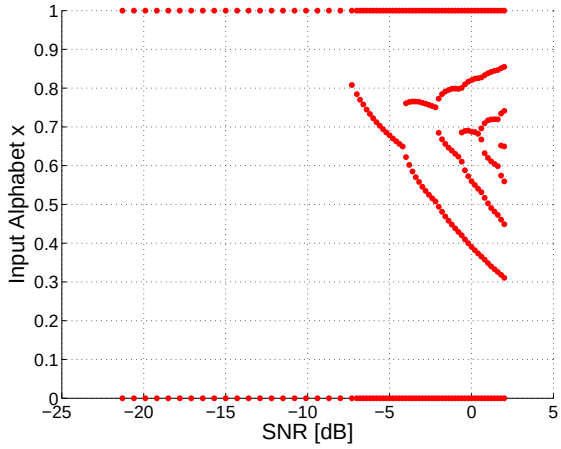


(a) $\rho = 2$ and $M = 40$

(b) $\rho = 3$ and $M = 40$

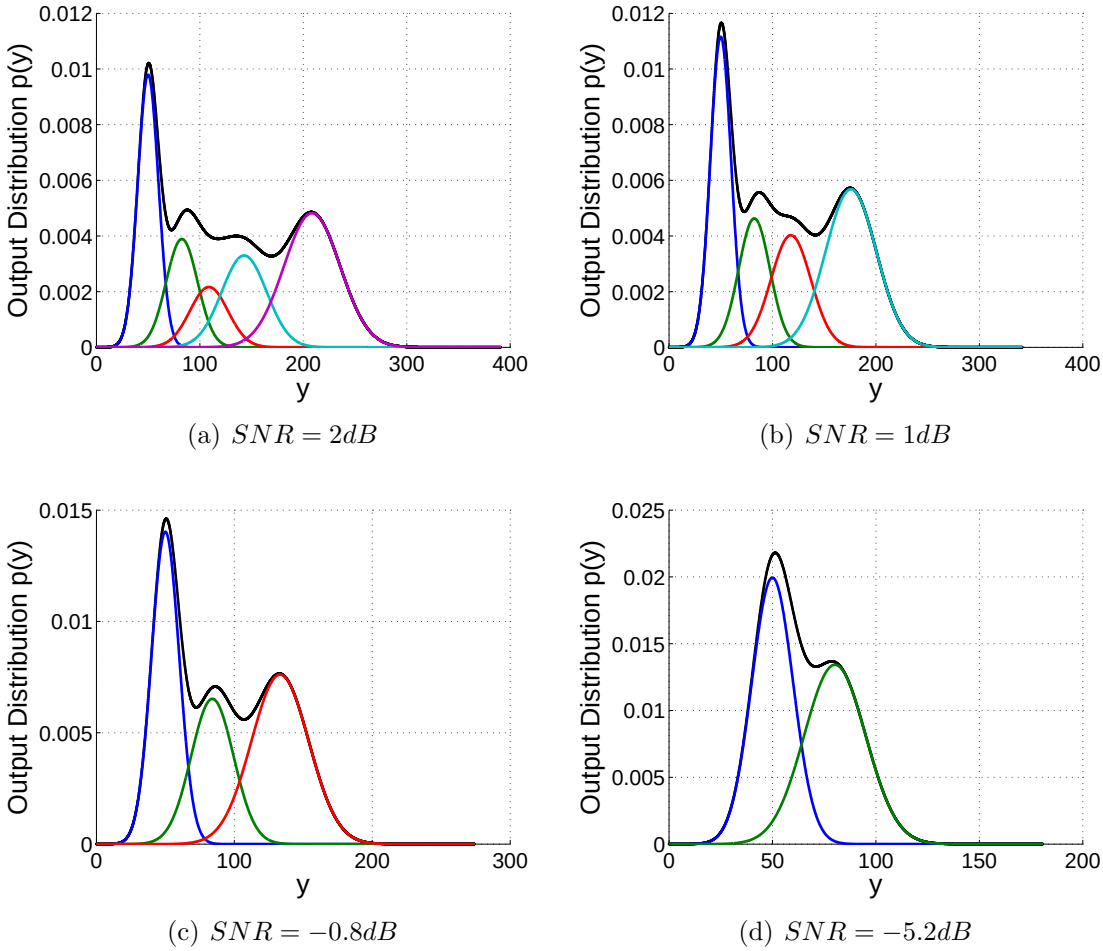


(c) $\rho = 4$ and $M = 40$



(d) $\rho = 5$ and $M = 40$

Figure 5.10: Positions of the mass points of the capacity-achieving input distribution of the Gaussian model for different PAPR ρ and $M = 40$



(a) $SNR = 2dB$

(b) $SNR = 1dB$

(c) $SNR = -0.8dB$

(d) $SNR = -5.2dB$

Figure 5.11: Output distribution $p(y)$ for $M = 50$ and $\rho = 2$

5.3 Non-Central Chi-Square Model

The simulation results for the non-central χ^2 model, as we can see in the following figures, are basically the same for a noise dimensionality $M \geq 16$. The following figures demonstrate this by showing the channel capacity (Fig. 5.12), the capacity-achieving input distribution (Fig. 5.13) for $M = 50$ and $\rho = 2$ and the mass points (Fig. 5.14) for several values of M and $\rho = 2$.

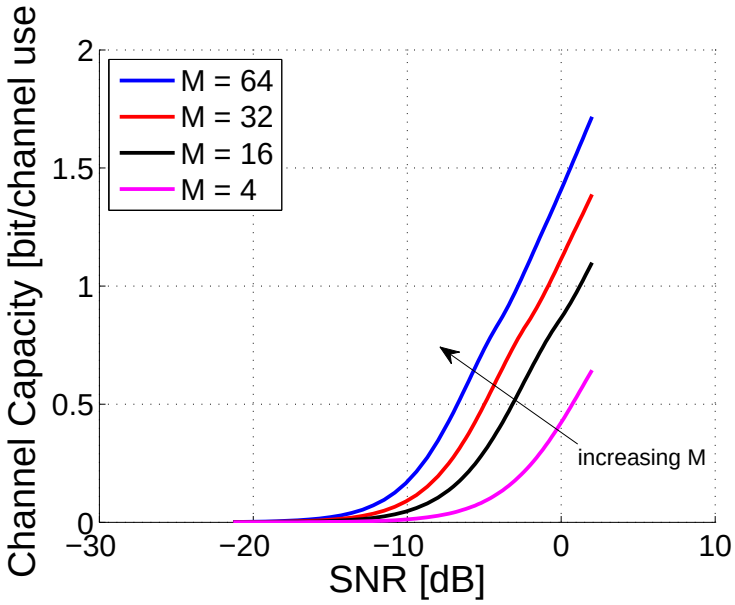


Figure 5.12: Capacity of the non-central χ^2 channel for different M and $\rho = 2$.

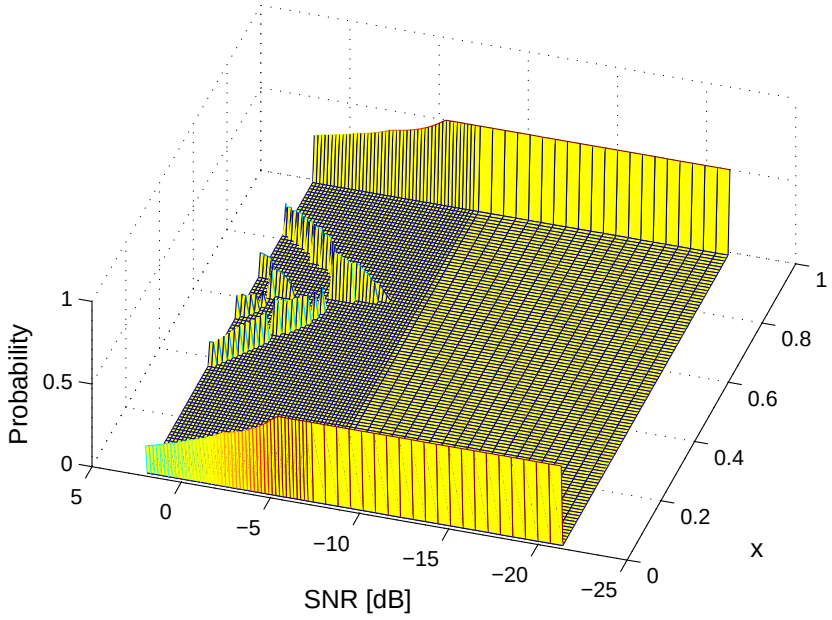
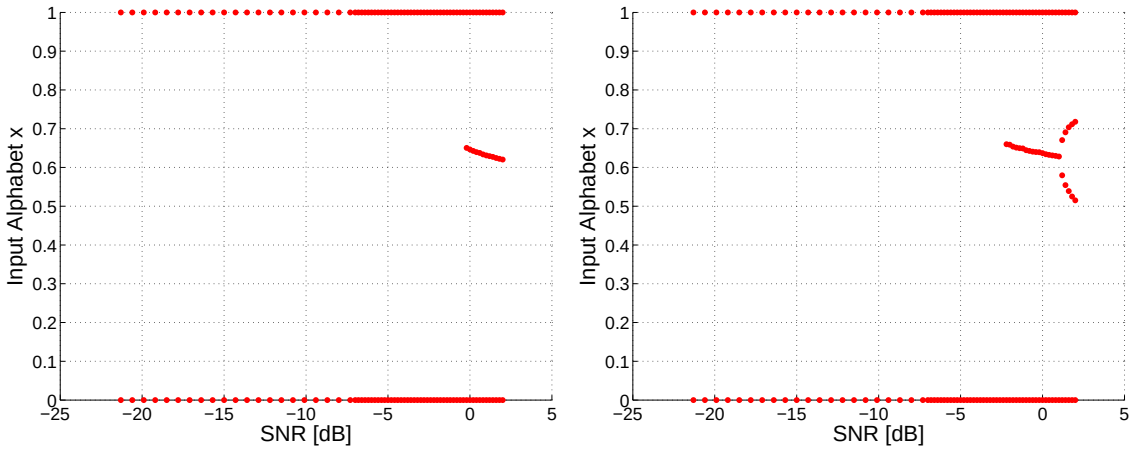
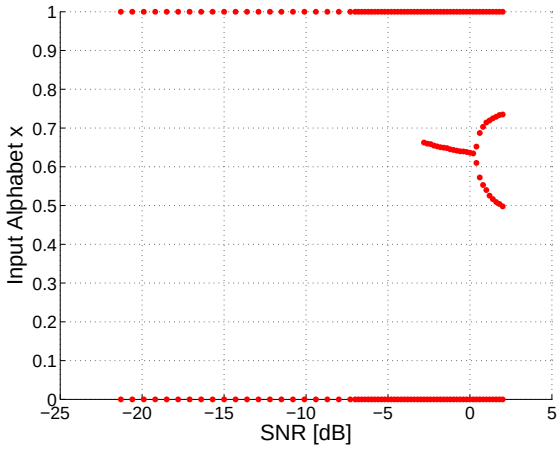


Figure 5.13: Capacity-achieving input distribution of the Non-central χ^2 model for $M = 100$ and $\rho = 2$

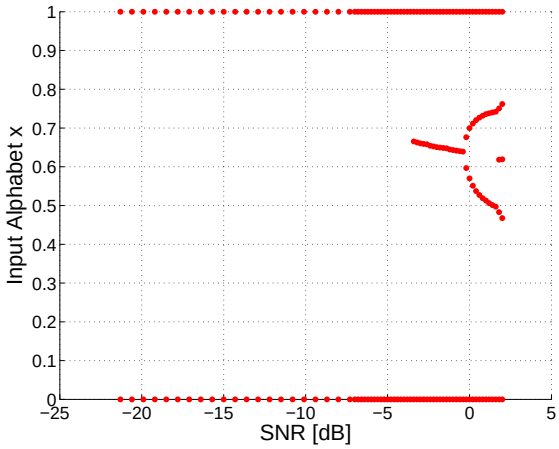


(a) $M = 16$ and $\rho = 2$

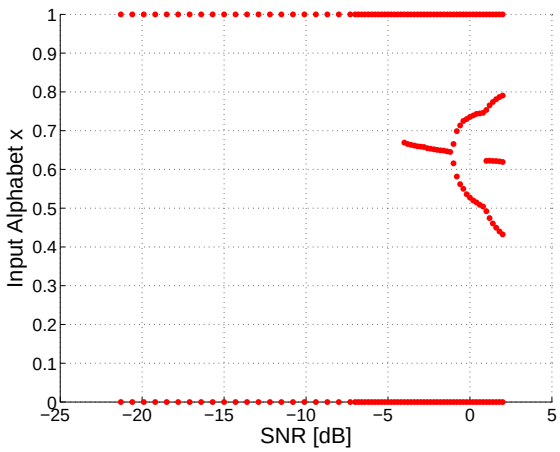
(b) $M = 32$ and $\rho = 2$



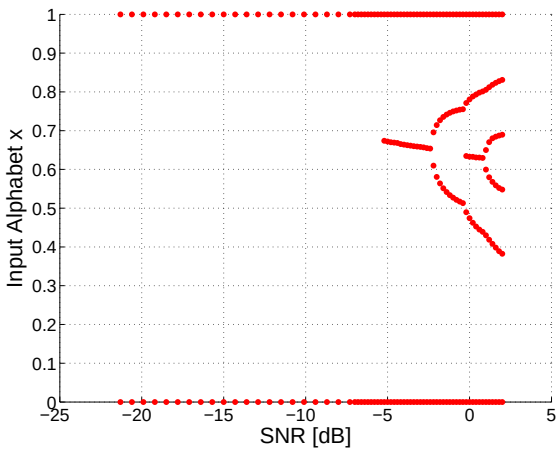
(c) $M = 40$ and $\rho = 2$



(d) $M = 50$ and $\rho = 2$



(e) $M = 64$ and $\rho = 2$



(f) $M = 100$ and $\rho = 2$

Figure 5.14: Positions of the mass points of the capacity-achieving input distribution of the non-central χ^2 model for different M

5.4 Comparison and Evaluation

Comparison

As Fig. 2.6 in Section 2.3 already has shown, the Gaussian approximation model and the non-central χ^2 model coincide relatively well for $M \geq 10$. This is the reason why the simulation results for these models are virtually the same for $M \geq 10$. Fig. 5.15 shows the channel capacity over the noise dimensionality M. It stands out that the channel capacity increases with M. But more importantly, the channel capacities of both models matches very well. It is not until $M < 5$, that a recognizable but nevertheless small difference between the channel capacities of both models is seen. This has its reason in the reduced sensitivity of the mutual information in the exact probabilities and positions of the mass points, especially for small SNR, if there are only two mass points.

Fig. 5.16 shows that the capacity achieving input distributions of the two models differ significantly for small values of M. For Fig. 5.16(a) the SNR region was chosen for higher SNR values to show the difference between the input alphabets for more than 2 mass points at zero and maximum amplitude, so that a distinction becomes visible. The figure shows also that the input alphabets not only shift with the SNR but also the mass points move slightly towards the mass point at $x = 0$.

Another recognizable difference between the models is the calculation time of the channel transition PDF in MATLAB. The calculation of the non-central χ^2 distribution takes significantly more time, especially for long input vectors $\hat{x}$ and output vectors $\hat{y}$ (quantized values for the quantized output PDF $\hat{p}(\hat{y})$). This leads to a considerably longer simulation time for the non-central χ^2 model.

Evaluation

To verify the simulation results, it was reasonable to design three different evaluation models. The first one is a model of the energy detector, i.e. it squares and sums up the input signal over the noise dimensionality M as

$$\boldsymbol{s_i} = \begin{pmatrix} \hat{x}i \\ \dots \\ \hat{x}i \end{pmatrix}_{1\times M} + \sigma \begin{pmatrix} k_1 \sim \mathcal{N}(0,1) \\ \dots \\ k_M \sim \mathcal{N}(0,1) \end{pmatrix}_{1\times M}$$

$$\boldsymbol{y_1} = \begin{pmatrix} \sum_{j=1}^{M} s_1(j)^2 \\ \dots \\ \sum_{j=1}^{M} s_N(j)^2 \end{pmatrix} \tag{5.4}$$

The second model calculates the distribution of the output y according to the Gaussian approximation model (MATLAB: *normpdf*) and the third one according to the

non-central χ^2 distribution (MATLAB: *ncx2rnd*) as

$$\begin{aligned} \boldsymbol{y_2} &= \mu(\boldsymbol{x}) + \sigma(\boldsymbol{x})(k \sim \mathcal{N}(0,1)) \\ &= \big(M + M\frac{\boldsymbol{x}^2}{\sigma^2}\big) + \big(2M + 4M\frac{\boldsymbol{x}^2}{\sigma^2}\big)(k \sim \mathcal{N}(0,1)) \end{aligned} \tag{5.5}$$

and

$$\boldsymbol{y_3} = \chi^2_L(\lambda) = \chi^2_M(\frac{M\boldsymbol{x}^2}{\sigma^2}). \tag{5.6}$$

The capacity-achieving input distributions from the simulations serve as input symbols for the models. So the mutual information between input symbols x and output symbols y of the models has to be approximately the channel capacity obtained from the simulation. Table 5.4 shows (for some chosen SNR values) the numerical results of the channel capacity and the mutual information calculated for the three models. In all three models the mutual information has roughly the same value as the channel capacity. On closer inspection, one notices that the mutual information for the modelled energy detector (first model) and the non-central χ^2 evaluation model (third model) represent the capacity more accurately than the Gaussian model. Given that the first model is the exact mathematical representation of the energy detector and the third one uses the exact statistics of the energy detector, it becomes obvious why the first and the third model coincide better.

The Figs. 5.17(a)-(d) illustrates that the output probability $p(y)$ of the models (coloured curves) coincides well with the exact output probability $p(y)$ (black curve).

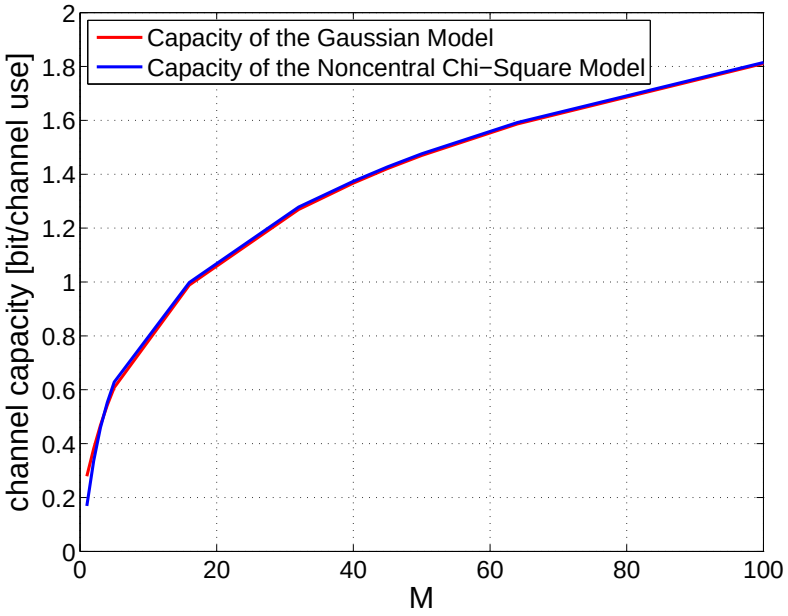
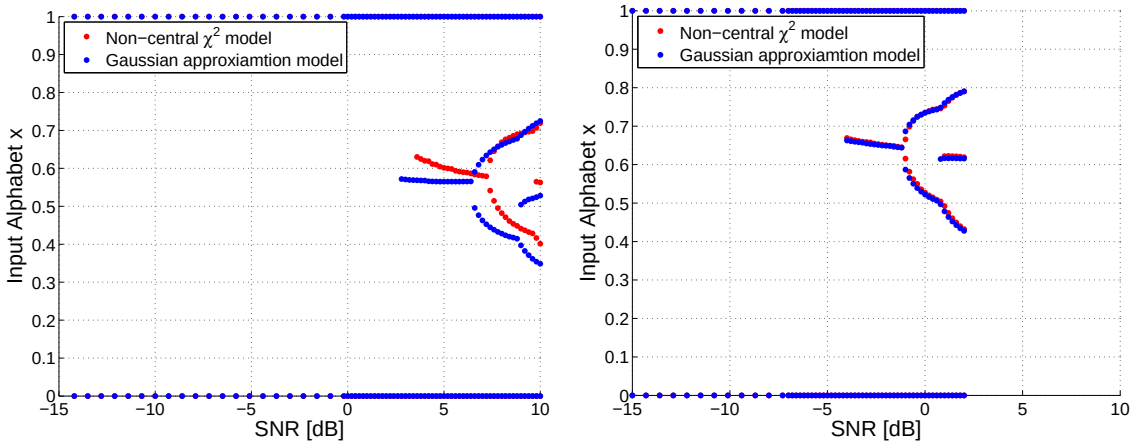


Figure 5.15: Channel Capacity of the Gaussian approximation model and the non-central χ^2 model over M for $SNR = 1.2dB$ and $\rho = 2$

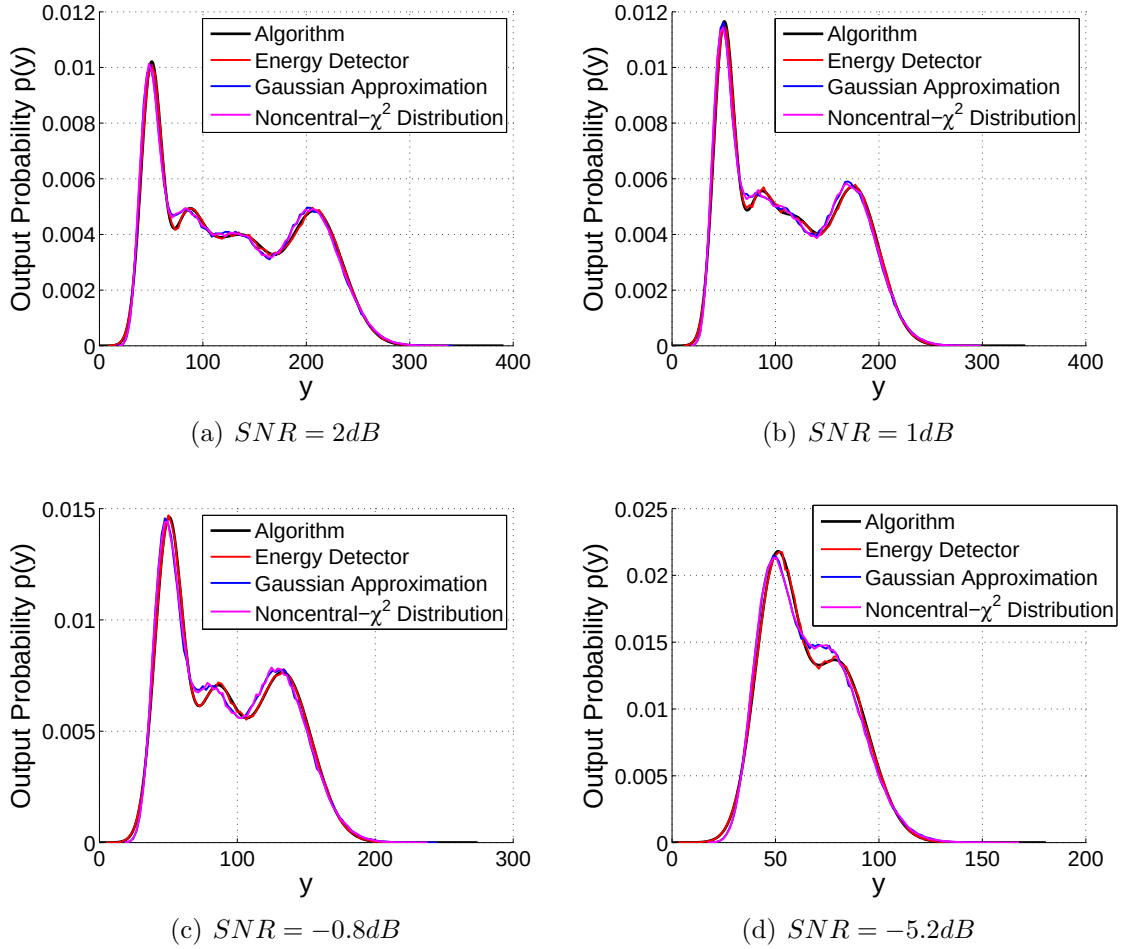


(a) $M = 5$ and $\rho = 2$ (b) $M = 64$ and $\rho = 2$

Figure 5.16: Comparison of Achieving input distributions between the non-central χ^2 model and the Gaussian approximation model

SNR	**C (algorithm)**	I_{ED}	I_{GM}	I_{NCX2M}
[dB]	[bit/channel]	[bit/channel]	[bit/channel]	[bit/channel]
2	1.5896	1.5844	1.582	1.586
1	1.44	1.4366	1.431	1.4367
-0.8	1.1807	1.1785	1.172	1.1781
-5.2	0.61882	0.61458	0.60951	0.61278

Table 5.4: Comparison of the channel capacity and the mutual information of the models: I_{ED} energy detector, I_{GM} Gaussian model, I_{NCX2M} non-central χ^2 model



(a) $SNR = 2dB$ (b) $SNR = 1dB$

(c) $SNR = -0.8dB$ (d) $SNR = -5.2dB$

Figure 5.17: Comparison between the output distributions $p(y)$ ot the different models for $M = 50$ and $\rho = 2$

Chapter 6

Introducing a Lower Bound

After we have analysed the non-linear optimization problem and seen the simulation results, it is obvious that for every parameter set — PAPR, SNR value and noise dimensionality M — the optimization problem must be solved to an exact input distribution $p(x)$. The complexity of finding this optimal input distribution grows with increasing SNR and number of appearing mass points.

For practical reasons it is often not necessary to solve this complex optimization problem. Instead, it is possible to find a good estimate for the capacity-achieving input distribution for a specific range of SNR. The basic concept was to find an input distribution, for which the mutual information approaches the channel capacity over a limited range of SNR. In [FH09] the authors proposed to find a simpler family of capacity-approaching distributions by solving the source entropy maximization problem. The mutual information calculated with this maxentropic input distribution can be seen as a lower bound of the channel capacity for defined intervals of SNR.

By keeping the number of mass points fixed over this range of SNR a realistic evaluation for practical coding schemes can be done. We will see that a simple OOK input alphabet almost achieves the capacity for small values of SNR.

My idea was to expand the concept introduced in the paper [FH09] of finding equally spaced mass points with a defined probability, to mass points which are not equally spaced. By analysing the capacity-achieving input distribution of the energy detector, it turned out that computing the square-root of the equally spaced mass point position leads to a very good approximation of the mass point spacing of the energy detector for certain SNR values.

6.1 Capacity-Approaching Input Distribution

The following derivations are based on [FH09]. At first we define a set of discrete input distributions $q(x)$ with $K+1$ mass points, which are not equally spaced, i.e.

$$P_{\mathcal{X}} = \Big\{ q(x) : [0, A] \to [0, 1] : \; b_k = \sqrt{k\frac{A^2}{K}}, \; q(x) = \sum_{k=0}^{K} a_k \delta(x - kb_k), \sum_{k=0}^{K} a_k = 1 \Big\} \tag{6.1}$$

where b_k are the positions, a_k the probabilities of the mass points and A is the maximum amplitude. The variable parameters of the distribution are the number of mass points and their probability, but the spacing between the points is fixed. For this class of distributions the maximum entropy and the corresponding input distribution with $K+1$ mass points can be defined as

$$\overline{q}(x; K) = \arg\max_{P_{\mathcal{X}}} \mathcal{H}(X) \tag{6.2}$$

where

$$\mathcal{H}(X) = \sum_{k=0}^{K} a_k \log \frac{1}{a_k}. \tag{6.3}$$

Out of this set of distributions, the capacity approaching input distribution, which maximizes mutual information, can be found in the same manner. The important difference is that the number of mass points $\overline{K}(\rho, SNR)+1$ is dependent on the channel parameters $\rho = \frac{A^2}{P}$ and $SNR = \frac{P}{\sigma^2}$. So we can define the capacity approaching input distribution as

$$\overline{q}^*(x; \overline{K}(SNR; \rho)) = \arg\max_{q(x) \in P} I(X; Y) \tag{6.4}$$

where

$$P = \Big\{ \overline{q}(x) : \forall K \in \mathbb{N}; \; \overline{q}(x) = \overline{q}(x; K) \Big\}. \tag{6.5}$$

To solve the optimization problem involving the channel parameter, we have to find the probabilities of the mass points and the total number of mass points $\overline{K}$. By the reason that $\overline{K}$ varies with the channel parameters SNR and ρ, it is difficult to define an analytical expression, but for a given number of mass points $K+1$, an entropy maximizing input $\overline{q}(x; K)$ can be found by solving

$$\max_{a_k} \mathcal{H}(X), \tag{6.6}$$

under the following constraints:

$$\sum_{k=0}^{K} a_k = 1, \; \sum_{k=0}^{K} b_k^2 a_k \leq P, \; A = b_K \tag{6.7}$$

By using Lagrange multipliers the optimization can be done by

$$\begin{aligned}\mathcal{J} &= \sum_{k=0}^{K} a_k \log\frac{1}{a_k} - \lambda_1\Big(\sum_{k=0}^{K} a_k - 1\Big) - \lambda_2\Big(\sum_{k=0}^{K}\Big(\sqrt{k\frac{A^2}{K}}\Big)^2 a_k - P\Big) \\ &= \sum_{k=0}^{K} a_k \log\frac{1}{a_k} - \lambda_1\Big(\sum_{k=0}^{K} a_k - 1\Big) - \lambda_2\Big(\sum_{k=0}^{K} k\frac{A^2}{K} a_k - P\Big)\end{aligned} . \tag{6.8}$$

To find the input distribution that the achieves maximum of $\mathcal{H}(X)$, Eq. 6.8 has to be differentiated w.r.t. a_k and the constraints have to be inserted. So we get

$$\sum_{k=0}^{K}\Big(1 - k\frac{A^2}{KP}\Big)e^{-k\lambda_2\frac{A^2}{K}} = \sum_{k=0}^{K}\Big(1 - k\frac{\rho}{K}\Big)e^{-k\lambda_2\frac{P\rho}{K}} \tag{6.9}$$

which can be rewritten as

$$S(t) = \sum_{k=0}^{K}\Big(1 - k\frac{\rho}{K}\Big)t^k \leq 0, \tag{6.10}$$

where

$$t = e^{-\lambda_2\frac{P\rho}{K}}. \tag{6.11}$$

For a given number of mass points K, Eq. 6.10 can be solved to obtain the real valued root $t_0 \in [0, 1]$. The discrete maxentropic input distribution with spacing b_k is then defined as

$$\overline{q}(x; K) = \sum_{k=0}^{K} \overline{a}_k \delta(x - kb_k), \qquad b_k = \sqrt{k\frac{A^2}{K}} \tag{6.12}$$

where

$$\overline{a}_k = \begin{cases} \frac{1}{K+1}, & A^2 \leq 2P \\ \frac{t_0^k}{1+t_0+t_0^2+\ldots+t_0^K}, & A^2 \geq 2P \end{cases} \tag{6.13}$$

In the case $A^2 < 2P$, the AP constraint is not active, which means that $\sum_{k=0}^{K} k\frac{A^2}{K} < P$ and hence $\lambda_2 = 0$. If the optimization problem is solved without the AP constraint, a uniform distribution is obtained. When $A \geq 2P$, both constraints are active and we have to compute t_0 as shown before. $A^2 = 2P$ is a limit case, whether the AP constraint is active or not, so that both equations in 6.13 lead to uniform distribution.

6.2 Evaluation

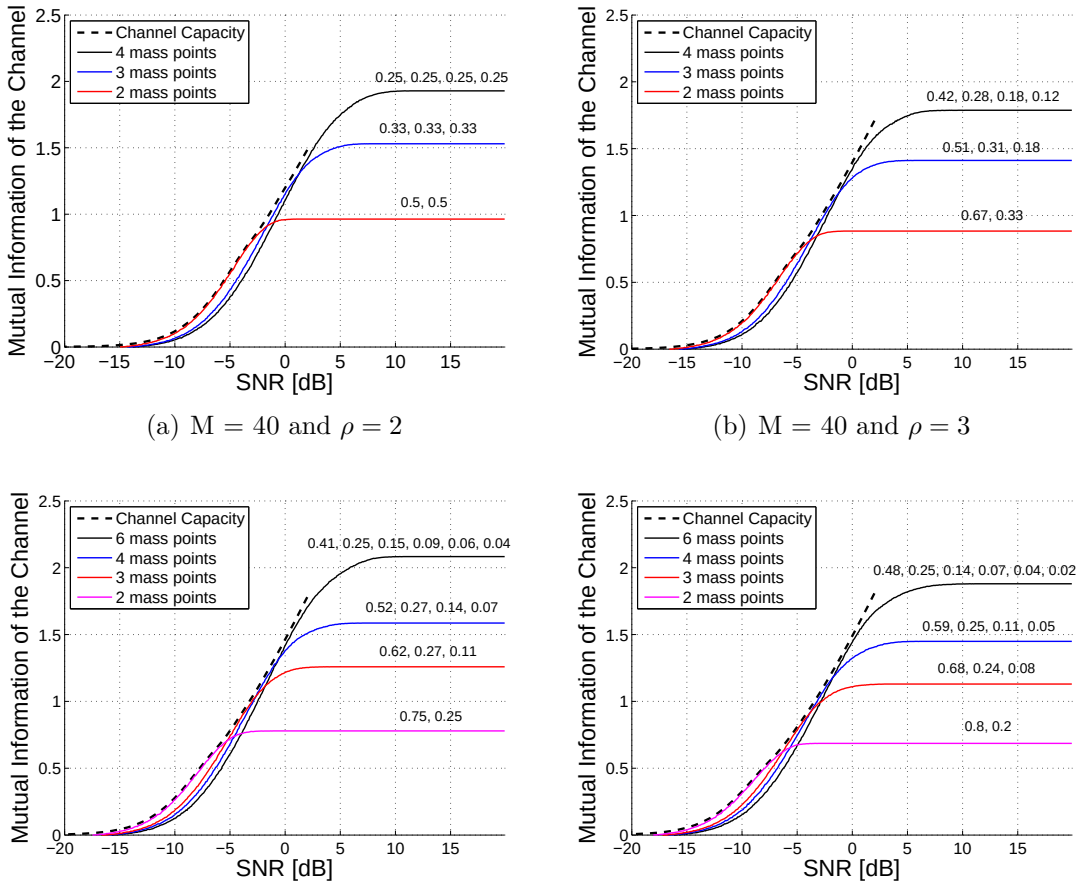
The Figs. 6.1(a)-(d) show the channel capacity and the mutual information of the introduced maxentropic input distribution versus SNR for different PAPR ρ and $M = 40$. Fig. 6.1(a) shows these curves for $\rho = 2$ and $K = 1$, 2 and 3. As we proposed before, the maxentropic input distribution for $\rho = 2$ is uniformly distributed with $K+1$ probability mass points. By comparing the channel capacity with the mutual information, a very small gap between them can be noticed for specific regions of SNR. This tells us, which coding scheme can be deduced from the input distribution in a certain range of SNR. It is especially interesting that for small SNR values the resulting capacity-achieving input distribution $p(x) = 0.5\delta(x) + 0.5\delta(x - A)$ computed by the original non-linear optimization problem coincides with the maxentropic input distribution obtained with Eq. 6.13. This distribution can be implemented directly using binary codes as it is suggested in [FH09].

In case that $\rho > 2$, the maxentropic input distribution has a non-uniform character, but at low SNR the distribution also coincides with the capacity-achieving input distribution. As we see in Fig. 6.1(b), 6.1(c) and 6.1(d) the bigger ρ is, the more weight is assigned to the mass point at zero, e.g. for $\rho = 4$, $p(x) = 0.75\delta(x) + 0.25\delta(x - A)$. Another fact is that with increasing ρ the input distribution gets more non-uniform. The probability of the mass points for this binary OOK alphabet can easily be found by the value set

$$[\overline{q}(0), \overline{q}(A)] = \left[\frac{\rho - 1}{\rho}, \frac{1}{\rho}\right]. \tag{6.14}$$

Figs. 6.2(a)-(d) show the capacity-achieving and the maxentropic distribution for several SNR values and $\rho = 4$. For $SNR = -6.4dB$ it can be seen in Fig. 6.2(a) that for two mass points the maxentropic distribution is also capacity-achieving. If the SNR increases the spacing between the mass points of the maxentropic distribution begins to differ from the optimal positions and the probabilities differ from the optimal probabilities, as shown in Fig. 6.2(b), 6.2(c) and 6.2(d).

In order to get the capacity-approaching input distribution, $\overline{K}(\rho, SNR)$ has to be obtained, which is analytically not a trivial task. So I provide a simpler method to compute the mutual information of the capacity-approaching input distribution over a wide range of SNR. Since I have computed the mutual information for different numbers of mass points, it was easy to find the $SNR((K+1)+1)$ values, where the mutual information for the maxentropy distribution with $K + 1$ mass points gets smaller than the mutual information of the distribution with $(K+1)+1$ mass points. Given that the number of mass points can be switched at the according $SNR((K + 1) + 1)$ value, the optimal maxentropic input distribution over a wide range of SNR and hence the capacity-approaching input distribution can be computed. The mutual information computed with the capacity-approaching input distribution is



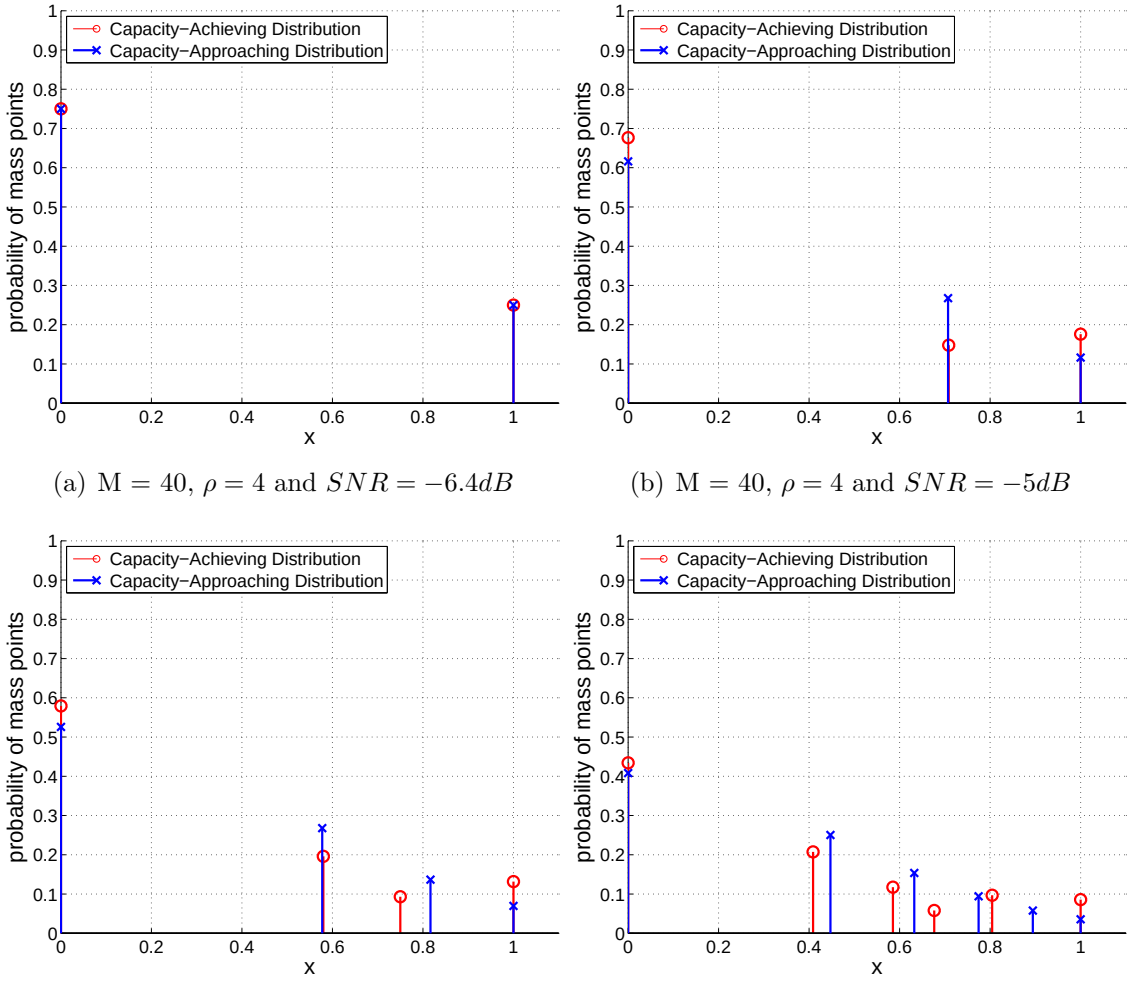
(a) M = 40 and $\rho = 2$

(b) M = 40 and $\rho = 3$

(c) M = 40 and $\rho = 4$

(d) M = 40 and $\rho = 5$

Figure 6.1: Channel capacity and mutual information of the energy detector calculated with maxentropic input distribution for different PAPR and $M = 40$



(a) M = 40, $\rho = 4$ and $SNR = -6.4dB$

(b) M = 40, $\rho = 4$ and $SNR = -5dB$

(c) M = 40, $\rho = 4$ and $SNR = -2.6dB$

(d) M = 40, $\rho = 4$ and $SNR = 0.6dB$

Figure 6.2: Maxentropic input distribution of the energy detector for different SNR values, $\rho = 4$ and $M = 40$

very close to the channel capacity as seen in Fig 6.3.

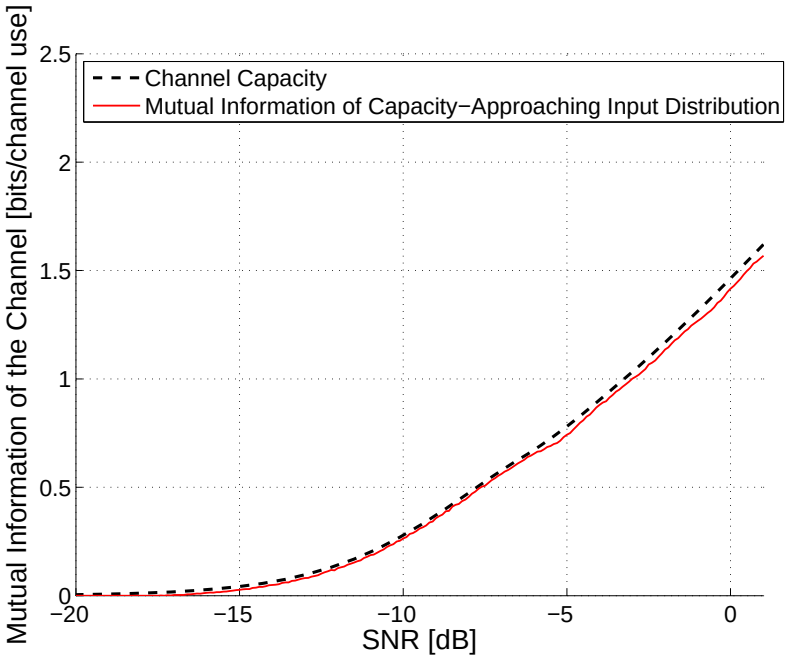


Figure 6.3: Mutual information of the capacity-approaching input distribution and channel capacity for $\rho = 4$ and $M = 40$

Chapter 7

Conclusion

In this thesis a detailed analysis was performed of the energy detector in an information-theoretical way. A numerical algorithm was provided to calculate the channel capacity and achieving input distribution of the energy detector, which can be seen as a continuous constrained channel. Therefore, the statistic of the output was analysed to provide a statistical model for the energy detector seen as transmission channel.

At first, I used the results from Urkowitz [Urk67] to approximate the output of the energy detector as the summation of the squared input signal, to obtain a simpler signal representation for finding the statistics of the output signal. The mathematical derivation has led to a non-central χ^2 RV with a non-centrality parameter dependent on the input signal x and the noise dimensionality M, i.e. length of the sum of the energy detector. Although the channel is non-linear, for a sufficiently large integration time T_i, i.e. sufficiently large noise dimensionality M, it can be approximated by an additive white Gaussian noise channel with input-signal-dependent noise variance. The evaluations have shown that the simulation results for both models of the energy detector coincide already very well at $M > 10$. Another fact was that the computation time of the algorithm shows significantly smaller values for the Gaussian approximation model than for the non-central χ^2 model.

In a second step, the numerical algorithm to solve such non-linear optimization problems was developed based on [Dau05] and [Dau06]. For an extended version of the Blahut-Arimoto algorithm, which is applicable to continuous channels, the key-idea is to represent the probability density (mass) function by a list of particles (mass points), where every particle has a weight and a position. To maximize the mutual information, the algorithm alternates two separate steps, where in the first step the weights (probabilities) are updated and in the second step the positions of the particles are updated. If the gap between the mutual information as lower bound of the channel capacity and the upper bound (maximum of the Kullback-Leibler divergence $D(p(y|x)||p(x))$) is small enough, the algorithm has converged and provides the particles as the mass points of the capacity-achieving input distribution and the maximized mutual information as channel capacity. The main problem of

the algorithm was long computation time until convergence. The reason for this was that every computation step of the algorithm was executed with the number of particles of the initial step, although the weights of many particles are insignificantly small. To take this fact into account and reduce the computation time, I further extended the algorithm by the "kick particles" and "fuse particles" methods.

The following step was the analysis of the resulting channel capacity and achieving input distribution of energy detector delivered by the algorithm. The simulations with both statistical models of the energy detector (non-central χ^2 and Gaussian approximation) have shown, amongst others, the following results

- The channel capacity increases with increasing PAPR ρ and noise dimensionality M, i.e. increasing integration time
- The alphabets are discrete with a finite number of mass points
- The number of mass points is monotonically non-decreasing with the SNR

The positions and probabilities of the mass points have to vary significantly with changing SNR to achieve the channel capacity. Therefore, a chosen alphabet (resulting from a modulation scheme, e.g. ASK, and a channel coding scheme) is only able to approach the optimal input distribution for a small region of SNR. Under a certain SNR value (depending on M and ρ), the achieving input distribution only consists of two mass points, where the probability of both is dependent of PAPR ρ. In this region of low SNR values, OOK is always an optimal modulation alphabet.

The comparison of the non-central χ^2 and the Gaussian model led to the conclusion that their simulation results begin to differ significantly from $M < 10$ downwards. To verify the numerical values of the channel capacity, I designed three evaluation models, which were fed with the simulated achieving input distribution. I then observed that the mutual information between the achieving input distribution and the output of the evaluation models practically coincided.

Finally, as such optimization problems are very complex, I developed a method to find a lower bound for the channel capacity for defined intervals of SNR by solving the source entropy maximization problem. I noticed that the capacity-achieving and capacity-approaching input distributions for low SNR match exactly, which led again to the conclusion that OOK is an optimal modulation alphabet for low SNR.

In summary, this master thesis was successful in finding the channel capacity and the corresponding achieving input distribution of the energy detector modelled with the exact non-central χ^2 model and the Gaussian approximation model. Additionally, it gave an idea of how to prove that the input distribution is discrete with a finite number of mass points. Finally, it delivered an approximation for the achieving input distribution.

There are still many questions that remain unanswered and would provide interesting areas of research. I briefly would like to propose two topics. Firstly, a detailed

comparing analysis between actual modulation and coding schemes used in UWB standards and the optimal input distribution alphabets calculated in this thesis, would contribute to finding optimal modulation and coding schemes for the energy detector used for UWB applications. Secondly, the algorithm could be extended to channels with memory so that we can take inter-symbol interference account.

Appendix A

Statistical Model of the Energy Detector with a Multipath Propagated Input Signal

The transmitted signal $s(t)$ is obtained by a train of pulses as

$$s(t) = \sqrt{P} \sum_{k=0}^{K-1} c_k p(t - kT_s) \tag{A.1}$$

where T_s is the symbol duration, K is the number of pulses and T_p is the length of a pulse. The pulses are orthogonal if $T_p \ll T_s$. A typical impulse response of a multipath fading channel can be written as a weighted sum of delta-pulses [ACB06]:

$$h(t) = \sum_{l=0}^{L-1} \alpha_l \delta(t - \tau_l) \tag{A.2}$$

where L is the number of multipath components, α_l the complex path gain and τ_l the path delay of the lth MPC. The received signal is the convolution of the transmitted signal with the channel impulse response and is written as

$$\begin{aligned} x(t) &= s(t) * h(t) + n(t) \\ &= \sqrt{P} \sum_{k=0}^{K-1} p(t - kT_s) * h(t) + n(t) \\ &= \sqrt{P} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \alpha_l p(t - kT_s - \tau_l) + n(t). \end{aligned} \tag{A.3}$$

The equation of the energy detector can be written as

$$y[m] = \int_{(m-1)T_i}^{mT_i} x(t)^2 dt = \int_{(m-1)T_i}^{mT_i} \left(\sqrt{P} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \alpha_l p(t - kT_s - \tau_l) + n(t) \right)^2 dt. \tag{A.4}$$

If no ISI occurs, this means that the symbol duration T_s is larger than the maximum excess delay $T_s \gg \tau_{max}$. This can be approximated by [Urk67]

$$y[m] = \sum_{n=0}^{M-1} \left(\sqrt{P} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \alpha_l p\left(\frac{mM-n}{W_{rx}} - kT_s - \tau_l \right) + n[mM-n] \right)^2 \tag{A.5}$$

where the noise dimensionality $M = T_i W_{rx} = T_s W_{rx}$. With $n[n] \sim \mathcal{N}(0, \sigma^2)$, $x[n]$ has the following statistic, if it is normalized with σ

$$x[n] \sim \mathcal{N}\left(\frac{\sqrt{P}}{\sigma} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \alpha_l p\left(\frac{n}{W_{rx}} - kT_s - \tau_l \right), 1 \right). \tag{A.6}$$

So the non-centrality parameter of the non-central χ^2 distribution can be written as

$$\lambda = \sum_{n=1}^{M} \mu_n^2 = M \frac{P}{\sigma^2} \left(\sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \alpha_l p\left(\frac{n}{W_{rx}} - kT_s - \tau_l \right) \right)^2. \tag{A.7}$$

This can be rewritten, if the pulse duration is less than $T_p < \tau_{l_{min}}$ ($\tau_{l_{min}} = \min_{l,j} |\tau_l - \tau_j|$), as

$$\lambda = \sum_{i=1}^{M} \mu_i^2 = M \frac{P}{\sigma^2} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \alpha_l^2 p^2\left(\frac{n}{W_{rx}} - kT_s - \tau_l \right). \tag{A.8}$$

A UWB system with transmitted-reference signalling in multipath propagation environment is analyzed in [QW05], which provides a more detailed statistical analysis of multipath propagated signals, received with a non-coherent receiver structure.

List of Abbreviations

AP	Average Power
ASK	Amplitude-Shift Keying
I&D	Integration and Dump
iid	independent identically distributed
ISI	Inter-symbol Interference
LHS	Left Hand Side
MA	Moving Average
MPC	Multipath Components
OOK	On-Off Keying
PA	Peak Amplitude
PAM	Pulse Amplitude Modulation
PAPR	Peak to Average Power Ratio
PDF	Probability Density Function
PMF	Probability Mass Function
PPM	Pulse Position Modulation
PSD	Power Spectral Density
RV	Random Variable
SNR	Signal to Noise Ratio
UWB	Ultra Wide Band

Bibliography

[ACB06] H. Arslan, Z.N. Chen, and M.G.D. Benedetto. *Ultra wideband wireless communication*. Wiley-Interscience, 2006.

[AFF10] I. Abou-Faycal and J. Fahs. On the capacity of some deterministic nonlinear channels subject to additive white gaussian noise. In *Telecommunications (ICT), 2010 IEEE 17th International Conference on*, pages 63 –70, april 2010.

[AFTS01] I.C. Abou-Faycal, M.D. Trott, and S. Shamai. The capacity of discrete-time memoryless rayleigh-fading channels. *Information Theory, IEEE Transactions on*, 47(4):1290 –1301, may 2001.

[Ari72] S. Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *Information Theory, IEEE Transactions on*, 18(1):14 – 20, jan 1972.

[Bla72] R. Blahut. Computation of channel capacity and rate-distortion functions. *Information Theory, IEEE Transactions on*, 18(4):460 – 473, jul 1972.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[CHK05] T.H. Chan, S. Hranilovic, and F.R. Kschischang. Capacity-achieving probability measure for conditionally gaussian channels with bounded inputs. *Information Theory, IEEE Transactions on*, 51(6):2073 – 2088, june 2005.

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 1991.

[Dau05] Justin Dauwels. Numerical computation of the capacity of continuous memoryless channels, 2005.

[Dau06] Justin H. G. Dauwels. *On Graphical Models for Communications and Machine Learning: Algorithms, Bounds, and Analog Implementation*. PhD thesis, 2006.

[FH09] A. Farid and S. Hranilovic. Channel capacity and non-uniform signalling for free-space optical intensity channels. *Selected Areas in Communications, IEEE Journal on*, 27(9):1553 –1563, december 2009.

[Gei09] Bernhard Geiger. Enhanced accuracy channel estimation and ranging for energy detectors. Master's thesis, 2009.

[GPV03] Mustafa C. Gursoy, H. Vincent Poor, and Sergio Verd. On the capacity-achieving distribution of the non-coherent rician fading channel,? canadian workshop on information theory. In *Canadian Workshop on Inform. Theory*, pages 18–21, 2003.

[GPV05] M.C. Gursoy, H.V. Poor, and S. Verdu. The noncoherent rician fading channel-part i: structure of the capacity-achieving input. *Wireless Communications, IEEE Transactions on*, 4(5):2193 – 2206, sept. 2005.

[HAD90] B. Honary, F. Ali, and M. Darnell. Information capacity of additive white gaussian noise channel with practical constraints. *Communications, Speech and Vision, IEE Proceedings I*, 137(5):295 –301, oct. 1990.

[NO07a] Obianuju Ndili and Tokunbo Ogunfunmi. Achieving maximum possible download speed on adsl systems. In *Signal Processing Systems, 2007 IEEE Workshop on*, pages 407 –411, oct. 2007.

[NO07b] Obianuju Ndili and Tokunbo Ogunfunmi. Achieving maximum possible speed on constrained block transmission systems. *EURASIP J. Adv. Sig. Proc.*, 2007, 2007.

[QW05] T.Q.S. Quek and M.Z. Win. Analysis of uwb transmitted-reference communication systems in dense multipath channels. *Selected Areas in Communications, IEEE Journal on*, 23(9):1863 – 1874, sept. 2005.

[SBD95] S. Shamai and I. Bar-David. The capacity of average and peak-power-limited quadrature gaussian channels. *Information Theory, IEEE Transactions on*, 41(4):1060 –1071, jul 1995.

[SH06] Lothar Sachs and Jürgen Hedderich. *Angewandte Statistik*. Springer, Berlin [u.a.], 12., vollst. neu überarb. aufl. edition, 2006.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.

[Urk67] H. Urkowitz. Energy detection of unknown deterministic signals. *Proceedings of the IEEE*, 55(4):523 – 531, april 1967.

[WLJ+09] Klaus Witrisal, Geert Leus, Gerald J. M. Janssen, Marco Pausini, Florian Trösch, Thomas Zasowski, and Jac Romme. Noncoherent ultra-wideband systems: An overview of recent research activities. *IEEE Signal Processing Magazine*, July 2009.