Capacity and Capacity-Achieving Input Distribution of the Energy Detector

Erik Leitinger*, Bernhard C. Geiger*, and Klaus Witrisal*

*Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria

{erik.leitinger,geiger,witrisal}@tugraz.at

***Abstract*—This paper presents the capacity-achieving input distribution of an energy detection receiver structure. A proper statistical model is introduced which makes it possible to treat the energy detector as a constrained continuous communication channel. To solve this non-linear optimization we used the Blahut-Arimoto algorithm extended with a particle method, so that also continuous channels can be handled. To get a better convergence behavior of the algorithm, we also implement two new methods, which are called "fuse particles" and "kick particles" [1]. The results we present show that the capacity of the energy detector decreases with increasing integration time and decreasing peak-to-average power ratio. It is shown that the capacity-achieving input distribution is discrete with a finite number of mass points.**

I. Introduction

The increasing interest in UWB systems in the recent past arises from the benefits of large bandwidth which offer the possibility of extremely high data rate, high-accuracy ranging and positioning in indoor environments, and robustness against multipath fading. Especially high accuracy-positioning and tracking are growing fields of interest. 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. 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 [2].

In this paper, the energy detector is analyzed from an information-theoretical point-of-view in order to find the channel capacity and the capacity-achieving input distribution. For this purpose, a statistical model of the energy detector is needed [3], [4], which describes the whole system from the input to output as a memoryless channel. In order to compute the capacity and the corresponding optimal input distribution, the so-called Blahut-Arimoto algorithm [5] was combined with a particle method [6], [7], so the algorithm becomes applicable to continuous memoryless channel with input constraints.

If there exist constraints on the average power (AP) and/or the peak amplitude (PA) of the input distribution, the capacity-achieving input distributions can be discrete, with an infinite or a finite number of mass points, depending on the communication channel. Already in 1971 Joel G. Smith (as quoted in [8]) delivered the proof that the capacity of a scalar additive white Gaussian noise channel with AP and PA constraint is achieved by a discrete input distribution with a finite number of mass points. Furthermore, according to [9] many non-linear channels with at least an AP constraint have a discrete optimal input alphabet with an infinite or a finite number of mass points, depending on the non-linearity. In [10] and [8] a statistical model for an optical intensity channel is proposed, which is very similar to our model of the energy detector. Further it was shown that such Gaussian channels with signal-dependent noise variance, also have also a discrete and finite optimal input alphabet, if an AP and PA constraint is considered.

The remainder of this paper is organized as follows: Section II presents the signal model of the energy detector, its statistical model and an approximation model are found in Section III. Section IV shortly introduces the used algorithm to calculate the channel capacity and the corresponding achieving input distribution of memoryless constrained continuous channels. Finally, in Section V the simulation results are presented and discussed.

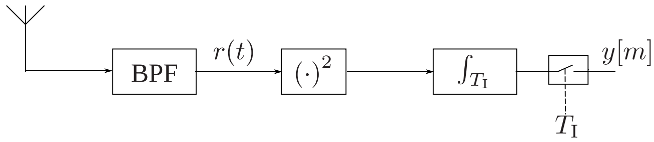


Fig. 1. Energy Detection Architecture

II. Signal Model

Fig. 1 shows the energy detector, where the received signal is first filtered, squared, integrated over the integration time T_I, and afterwards sampled at symbol period T_I. Let $r(t) = x(t) + \eta(t)$, where $x(t)$ is the transmitted signal convolved with the transmission channel impulse response and $\eta(t)$ is zero-mean Gaussian noise with power spectral density (PSD) $N_0/2$. So we get the output signal of the energy detector as

$$y[m] \quad = \quad \int_{mT_\mathrm{I}}^{(m+1)T_\mathrm{I}} \Big(x(t)+\eta(t)\Big)^2 \mathrm{d}t. \qquad (1)$$

According to Urkowitz [3] a sampled function with duration T_I and filtered with bandwidth W_rx can be approximated by a finite sum of $M = 2T_\mathrm{I}W_\mathrm{rx}$ sampled terms, where M is called noise dimensionality. In the case of a low-pass process the signal is sampled at times $1/(2W_\mathrm{rx})$ apart. For simplicity, a rectangular baseband-filter is considered with bandwidth W_rx.

The filtered noise signal $\eta(t)$ can be expressed as

$$\eta(t) = \sum_{n=-\infty}^{\infty} \eta[n]\mathrm{sinc}(2W_{\mathrm{rx}}t - n), \quad (2)$$

where $\eta[n] = \eta\left(n/\left(2W_{rx}\right)\right)$. Under these considerations, the noise samples $\eta[n]$ are independent and identically distributed (iid) Gaussian random variables (RV) with zero-mean and variance $\sigma^2 = N_0 W_{\mathrm{rx}}$. The signal $x(t)$ can be expressed in the same manner. By the fact that the sinc-functions at different time-steps are orthogonal, we can write for the output signal $y[m]$ of the energy detector [3]

$$\begin{aligned} y[m] &= \int\limits_{mT_{\mathrm{I}}}^{(m+1)T_{\mathrm{I}}} \left(\sum_{n=-\infty}^{\infty} (x[n] + \eta[n])\,\mathrm{sinc}(2W_{\mathrm{rx}}t - n) \right)^2 \mathrm{d}t \\ &\approx \frac{1}{2W_{\mathrm{rx}}} \sum_{n=1}^{M} \big(x[(m+1)M - n] + \eta[(m+1)M - n] \big)^2 \\ &= \frac{1}{2W_{\mathrm{rx}}} \sum_{n=1}^{M} r^2[(m+1)M - n]. \end{aligned} \quad (3)$$

The signal $x[n]$ for pulse amplitude modulation (PAM) is expressed as

$$x[n] = \sum_k z_k \tilde{\mathrm{p}} \left(\frac{n}{2W_{\mathrm{rx}}} - kT_{\mathrm{I}} \right) = \sum_k z_k \tilde{\mathrm{p}}\,[n - kM] \quad (4)$$

where z_k are the transmitted symbols with the time-index k and $\tilde{\mathrm{p}}(t) = \mathrm{p}(t) \star h(t)$ is the unit energy signal pulse convolved with the channel impulse response $h(t)$. It is assumed that $1/(2W_{\mathrm{rx}}) \sum_{n=0}^{M-1} \tilde{\mathrm{p}}^2[n] \approx \int_0^{T_{\mathrm{I}}} \tilde{\mathrm{p}}^2(t)\mathrm{d}t = 1$ and $\tilde{\mathrm{p}}(t) = 0$ for $t \notin [0, T_{\mathrm{I}}]$.

III. Statistical Model

A. Non-central $\mathcal{X}^2$ model

The filtered received signal $r[n]$ is modeled as independent RVs with distribution $\mathcal{N}\left(x[n], \sigma^2\right)$. It is well known that the sum of equal variance, squared Gaussian RVs can be represented by a non-central $\mathcal{X}_M^2\left(\lambda[m]\right)$-distribution that is defined by two parameters. Its degrees of freedom is specified by the noise dimensionality M and the non-centrality parameter $\lambda[m]$ is calculated from the squared means $x[n]$ of the signal $r[n]$, as

$$\begin{aligned} \lambda[m] &= \sum_{n=1}^{M} \frac{x^2[(m+1)M - n]}{\sigma^2} \\ &= \sum_{n=1}^{M} \frac{\sum_k z_k^2 \tilde{\mathrm{p}}^2\left[(m+1)M - kM - n\right]}{W_{\mathrm{rx}} N_0} = \frac{2z_m^2}{N_0} \end{aligned} \quad (5)$$

where z_m^2 is the energy of the mth transmitted symbol.

We will now treat the system from z_m to $y[m]$ as a stationary and memoryless channel and z_m as an instance of the iid RV $Z : \Omega \to \left\{\sqrt{E_{s_1}}, \sqrt{E_{s_1}} \cdots, \sqrt{E_{s_K}}\right\}$, where K is the number of symbols in the alphabet, i.e. number of mass points and $P[Z = \sqrt{E_{s_i}}] = p_i$ is the probability that the ith symbol from the modulation alphabet occurs, so that we simply can, with no loss of generality, suppress the time-index m. The average energy[1] of the modulation alphabet can be calculated by $E_s = \mathrm{E}\left\{Z^2\right\} = \sum_{i=1}^{K} E_{s_i} p_i$ and the PA, i.e. the maximum amplitude at the transmitter side is given by $A = \max_i \sqrt{E_{s_i}}$. The corresponding channel transition probability density function (PDF) is therefore the non-central $\mathcal{X}^2$-distribution conditioned on the input modulation symbols z and with the channel output y, so that we can write

$$\mathrm{P}_{Y|Z}^{(\mathcal{X}^2)}(y|z, M) \sim \mathcal{X}_M^2(\lambda) = \mathcal{X}_M^2\left(\frac{2z^2}{N_0}\right). \quad (6)$$

The mean and the variance of the conditioned output y are given by

$$\begin{aligned} \mathrm{E}_Y\left\{y|z\right\} &= M + \lambda = M + \frac{2z^2}{N_0} & (7) \\ \mathrm{var}_Y\left\{y|z\right\} &= 2\left(M + 2\lambda\right) = 2M + 4\frac{2z^2}{N_0}. & (8) \end{aligned}$$

These results show that both the mean and the variance of the output signal y of the energy detector increase with the noise dimensionality M and with the energy of the modulation symbols z.

B. Gaussian approximation model

To simplify the channel model for further tasks, the exact statistical model of the energy detector can be approximated by a Gaussian distribution. If the noise dimensionality M is large enough, the non-central $\mathcal{X}^2$-distribution can be approximated by a Gaussian PDF with a signal-dependent variance and mean (similar model as in [8], [10]). This fact holds because of the central limit theorem, which says that the sum of a sufficiently large number of independent identically distributed RVs will be approximately normal distributed.

By taking over the variance and the mean for the non-central $\mathcal{X}^2$-distribution, the corresponding Gaussian approximation of the channel transition PDF can be written as

$$\mathrm{P}_{Y|Z}^{(\mathrm{Gauss})}(y|z, M) \sim \mathcal{N}\left(M + \frac{2z^2}{N_0}, 2M + 4\frac{2z^2}{N_0}\right). \quad (9)$$

Figure 2 shows that for increasing M the Gaussian model coincides better with the non-central $\mathcal{X}^2$ model.

IV. Channel Capacity of constrained continuous channels

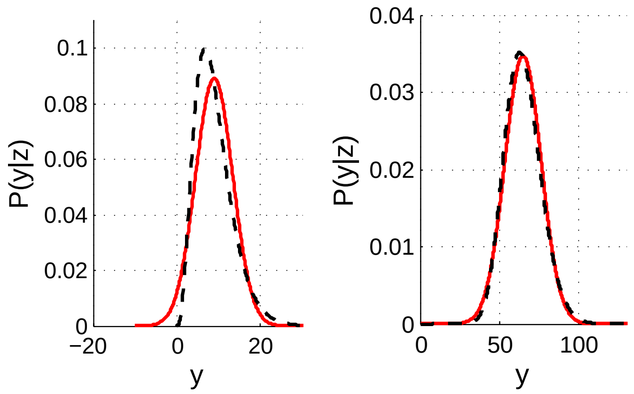
A. Channel Capacity

The mutual information, in bits/channels use, between the input $Z \in \mathcal{Z}$ and output $Y \in \mathcal{Y}$ alphabet, for a given channel transition PDF is given as [11]

$$\mathrm{I}(Z;Y) = \int_{z \in \mathcal{Z}} \int_{y \in \mathcal{Y}} \mathrm{p}(z)\mathrm{p}(y|z) \log \frac{\mathrm{p}(y|z)}{\mathrm{p}(y)} \mathrm{d}z\mathrm{d}y \quad (10)$$

where $\mathrm{p}(z)$ is the PDF of the input alphabet, $\mathrm{p}(y)$ the PDF of the output alphabet, and $\mathrm{p}(y|z)$ is the channel transition PDF.

[1] Due to the fact that the average power and average energy are simply related by the symbol rate, we use the term average power that is common in related literature.



(a) M = 8 (b) M = 64

Fig. 2. Non-central $\mathcal{X}^2$ (black dashed line) channel transition PDF and Gaussian approximation transition PDF (red solid line) for $M = 8$ and $M = 64$ and $\lambda = 1$.

The maximum of the mutual information, i.e. the maximum data rate that can be conveyed with an arbitrarily small error probability over the channel, is the channel capacity, which can be written as [11]

$$\mathrm{C} = \sup_{\mathrm{p}(z)\in \mathrm{P}_{\mathbb{R}}} \mathrm{I}(Z;Y) \tag{11}$$

where $\mathrm{P}_{\mathbb{R}}$ is the set of all input PDFs supported on the set $\mathbb{R}$ of real numbers. If the input alphabet is constrained in the AP and the PA, then the set of input distributions $\mathrm{p}(z)$ over which the maximization is done, has to be restricted. The new set which includes the constraints has the following structure [5] , [7]:

$$\mathrm{P}_{\mathbb{R}}(E_s, A) = \Big\{ \mathrm{p} : [0, A] \rightarrow [0, \infty] : \\ \int_{z\in\mathcal{Z}} \mathrm{p}(z)\mathrm{d}z = 1; \int_{z\in\mathcal{Z}} \mathrm{p}(z)z^2\mathrm{d}z \leq E_s \Big\} \tag{12}$$

Where E_s and A are certain numbers, which constrain the AP and the PA of the input alphabet. For given constraints E_s, A, a given noise variance σ^2, and a given channel PDF $\mathrm{p}(y|z)$ this non-linear optimization problem can be solved by numerical optimization methods.

B. The Algorithm

To solve such non-linear optimization problem, the Blahut-Arimoto algorithm [5] was extended with a particle method based on [6], [7]. To skirt the original infinite-dimensional optimization problem and reduce it to a finite-dimensional one, the continuous input distribution $\mathrm{p}(z)$ is reduced to a finite list of N particles

$$\mathcal{P}_{\mathrm{Z}} = \{(\hat{z}_1, w_1), (\hat{z}_2, w_2), \ldots , (\hat{z}_N, w_N)\} \tag{13}$$

where $w = \{w_1, \ldots , w_N\}$ are the weights, i.e. the probabilities of the particles, and $\hat{z} = \{\hat{z}_1, \ldots , \hat{z}_N\}$ are the positions of the particles. So we get the following finite-dimensional equation to optimize

$$\mathrm{C}(A, E_s) = \max_{\hat{z},w} \mathrm{I}(\hat{z}, w) = \max_{\hat{z},w} \sum_{i=1}^{N} w_i \mathrm{D}(\mathrm{p}(y|\hat{z}_i)||\mathrm{p}(y)) \tag{14}$$

where $\mathrm{D}(\mathrm{p}(y|\hat{z}_i)||\mathrm{p}(y)) = \int_{y\in\mathcal{Y}} \mathrm{p}(y|\hat{z}_i) \log \frac{\mathrm{p}(y|\hat{z}_i)}{\mathrm{p}(y)} \mathrm{d}y$ is the Kullback-Leibler divergence of the channel transition PDF $\mathrm{p}(y|\hat{z}_i)$ and the output distribution $\mathrm{p}(y)$. The corresponding output distribution $\mathrm{p}(y)$ is given by the channel transition PDF $\mathrm{p}(y|z)$, sampled on the particle positions $\hat{z}_i$ and the appearing weights w_i of the particles, $\mathrm{p}(y) = \sum_{i=1}^{N} w_i \mathrm{p}(y|\hat{z}_i)$. The remaining integral in the mutual information over the output distribution $\mathrm{p}(y)$ may be evaluated numerically or by Monte-Carlo integration. A fact is, that the mutual information is concave with respect to the set of input distributions $\mathrm{p}(z)$ for a fixed channel transition PDF $\mathrm{p}(y|z)$, whereas $\mathrm{I}(\hat{z}, w)$ is non-convex with respect to the postions $\hat{z}_i$ and weights w_i of the particles [11]. Due to that, the algorithm is separated in two alternating maximization steps [6], [7]:

1) W-step: The weights w_i are optimized to maximize the mutual information while the positions $\hat{z}_i$ are kept fixed. This is carried out by the Blahut-Arimoto algorithm.
2) X-step: The positions $\hat{z}_i$ are selected so that the Kullback-Leibler divergence $\mathrm{D}(\mathrm{p}(y|\hat{z}_i)||\mathrm{p}(y))$ between the channel transition PDF and the output distribution is maximized. This selection of the positions can be done by applying a gradient method.

Important is that the AP constraint and the PA constraint imposed to the input distribution of the modulation alphabet $\mathrm{p}(z)$ are considered in both steps. This means that the cost-function, i.e. the mutual information which has to be maximized, includes also the constraints, weighted with Lagrangian multipliers.

The W-step and the X-step are iterated until after q iterations the gap between the upper bound $U^{(q)} = \max_{\hat{z}} \mathrm{D}(\mathrm{p}\left(y|\hat{z}^{(q)}||\mathrm{p}(y)^{(q)}\right)$ and the lower bound $\mathrm{I}^{(q)} = \sum_{i=1}^{N} w_i^{(q)} \mathrm{D}\left(\mathrm{p}(y|\hat{z}_i^{(q)})||\mathrm{p}(y)^{(q)}\right)$ of the channel capacity $\mathrm{C}(A, E_s)$ is sufficiently small [1], [7], i.e. if

$$\mathrm{U}^{(q)} - \mathrm{I}^{(q)} \quad \leq \quad \epsilon \tag{15}$$

where ϵ is a “small” positive real number, e.g. $\epsilon = 10^{-6}$.

In [8] the capacity-achieving input distribution under an AP and a PA constraint is studied for conditionally Gaussian channels, with input-dependent mean and, in general, input signal-dependent covariance. The paper shows that optical Gaussian channels with signal-dependent noise variance and AP and PA constraints have a discrete input distribution with a finite number of mass points. Due to the fact that the Gaussian approximation model of the energy detector is very similar, we also expect a discrete input distribution with a finite number of mass points, so that the optimization problem can be exactly solved by using a finite list of particle.

Because of this, we can reduce the computation time of the algorithm by discarding particles which have very low weights w, and fusing particles if their positions $\hat{z}$ are close enough to one another. This methods to improve the convergence of the algorithm are called “kick particles” and “fuse particles” [1]. Important is that the number of initial particles is large enough so that every mass point of the optimal alphabet can be found. Furthermore, the threshold for discarding particles should not be too large, otherwise we might discard particles which could be relevant for the capacity-achieving input distribution. In almost the same manner we would adulterate the achieving input distribution, if we chose the threshold for fusing too large.

If the upper and the lower bounds are sufficiently close after q iterations, the mutual information $\mathrm{I}^{(q)}$ represents a good estimate of the channel capacity and the list of particles $\mathcal{P}_{\mathrm{Z}}^{(q)} = \left\{(\hat{z}_1^{(q)}, w_1^{(q)}), \ldots, (\hat{z}_K^{(q)}, w_K^{(q)})\right\}$ represents the capacity-achieving probability mass function (PMF) of the modulation alphabet, where K is the number of modulation symbols, i.e. mass points.

V. Simulations and Results

Due to the fact that already for $M = 8$ both models match well (Fig. 2) and our simulations are done for $M \geq 16$, it was sufficient to do all simulations only for the Gaussian model[2].

For the simulations the ratio of the signal energy E_s to the noise spectral density $N_0/2$ was chosen as variable parameter, which is a convenient definition of the SNR. The AP constraint is defined as $\mathrm{E}\left\{\hat{Z}^2\right\} \leq E_s$. The relation between the squared PA and the AP is given by the peak-to-average-power ratio (PAPR) $\rho = A^2/E_s$. The PA constraint is given by $\Pr\left[0 \leq \hat{z}_i \leq A\right] = 1$. The algorithm starts with an initial list of mass points $\mathcal{P}_{\mathrm{Z}}^{(0)} = \left\{(\hat{z}_1^{(0)}, w_1^{(0)}), (\hat{z}_2^{(0)}, w_2^{(0)}), \ldots, (\hat{z}_N^{(0)}, w_N^{(0)})\right\}$ with uniformly distributed weights $w_i^{(0)} = \frac{1}{N}$ and uniform spacing between the mass points so that $\hat{z}_i^{(0)} = iA/N$, where N, the number of initial particles, has to be sufficiently large. For the simulations, we use 100 particles and every main iteration consists of 2000 Blahut-Arimoto iterations and 30 gradient ascent update steps. The convergence parameter is chosen to be $\epsilon = 10^{-6}$.

Figure 3 shows the capacity for different M values over a wide range of E_s/N_0. The capacity is getting smaller with increasing noise dimensionality M, because more noise energy is accumulated. Figure 4 shows the capacity for $M = 64$ and for different values of PAPR ρ over a wide range of E_s/N_0. The capacity is getting higher by increasing ρ. This can be explained by the fact that for constant E_s/N_0, i.e. constant AP constraint, an increasing ρ leads to a higher allowed maximum amplitude, so that the distance between the symbols is getting larger and so that those are better separable.

Figure 5 shows the positions of the mass points of the capacity-achieving input distribution over E_s/N_0. First, the figure suggests that the achieving input distribution is discrete and has always a mass points at zero amplitude and maximum amplitude. The simulation results for the optical intensity channel in [8] confirm these properties of the optimal input distribution. Another fact is that the number of appearing mass points is monotonically non-decreasing with increasing E_s/N_0 to achieve the according capacity.

There are two ways new mass points can appear if the E_s/N_0 increases: First, a new mass point can appear apart from the other mass points with a low probability or, second, an existing mass point with relatively high probability can be split up into two new mass points with low probabilities. This characteristic of the probability of the mass points over the E_s/N_0 is shown Fig. 6.

Examples of capacity-achieving input distribution after the first Blahut-Arimoto cycle and after the upper and lower bound are sufficiently close are shown in Fig. 7. For $\rho > 2$ the AP is always active and the probabilities of the mass points with higher amplitudes show decreasing weights. This behavior is not strict, because when new mass points appear, mass points with higher amplitude could have a higher probability than some with a lower amplitude. Another interesting fact is that the probability of the two remaining mass points at low E_s/N_0 does not change. (For $M = 64$ and $\rho = 5$ the optimal distribution for small E_s/N_0 values is $\mathrm{p}(z) = 0.8\delta(z) + 0.2\delta(z - A)$.) For $\rho \leq 2$ the distribution of the mass points shows the tendency to be “tub-shaped”. In this case, if the AP constraint is not active, the probabilities of the two remaining mass points at low E_s/N_0 is changing with the E_s/N_0. The optimal distribution is approximately $\mathrm{p}(z) \approx 0.5\delta(z) + 0.5\delta(z - A)$, but the probabilities show small changes with E_s/N_0. The

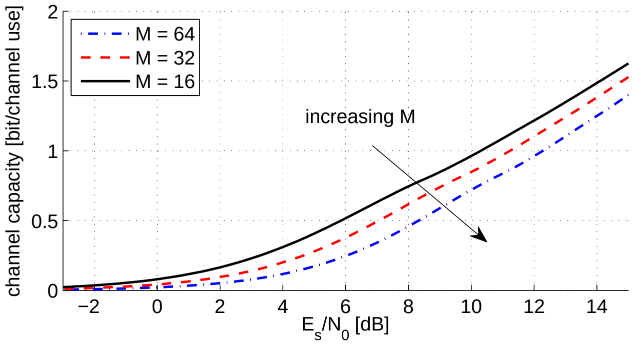


Fig. 3. Channel Capacity for different M and $\rho = 2$

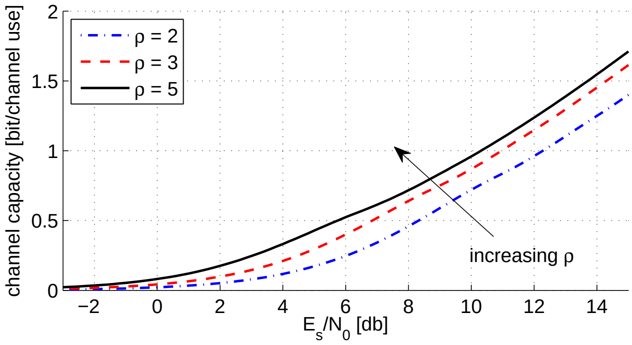


Fig. 4. Channel Capacity for different ρ and $M = 64$

[2]It is noteworthy that the algorithm also works for the non-central $\mathcal{X}^2$-distribution, but the simulation times are much longer.

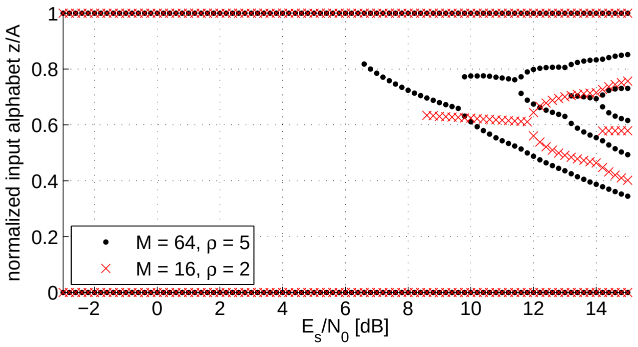


Fig. 5. Positions of mass point of the capacity-achieving input distribution

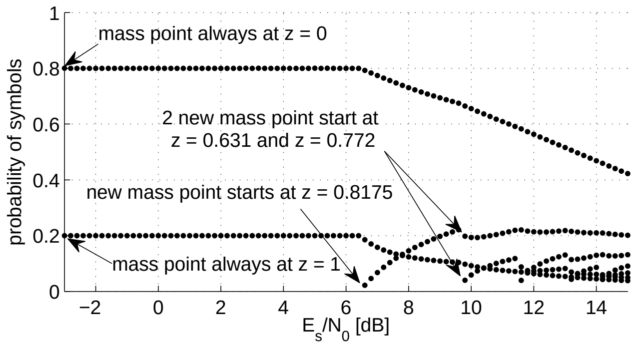
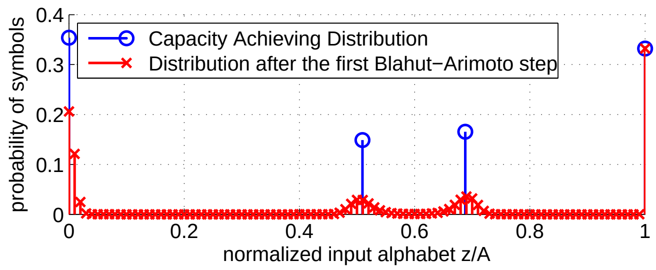
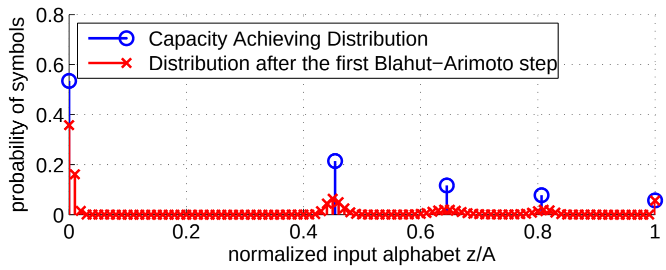


Fig. 6. Probabilities of mass point of the capacity-achieving input distribution for $M = 64$ and $\rho = 5$



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



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

Fig. 7. Capacity-achieving input distribution for $E_s/N_0 = 12.6$dB

simulation results show that the positions, probabilities, and numbers of the mass points vary significantly with changing E_s/N_0 to achieve the channel capacity. Therefore, a chosen alphabet is only able to approach the optimal input distribution for a small region of E_s/N_0.

We already mentioned that the capacity increases with the PAPR ρ. This fact is mirrored by the increasing asymmetry of the optimal input distribution. Under a certain minimum E_s/N_0 value (depending on M and ρ), the achieving input distribution only consists of two mass points, whose probabilities can be expressed by $[\mathrm{p}(0), \mathrm{p}(A)] = [(\rho - 1)/\rho, 1/\rho]$, which shows that the asymmetry of the probability increases with the PAPR ρ. In this region of low E_s/N_0 values, on-off keying (OOK) is always an optimal modulation alphabet. For higher E_s/N_0 values higher-order PAM is optimal, where the optimality is only given over a small region of E_s/N_0.

VI. Conclusion

We have presented the capacity and capacity-achieving input distribution of the energy detector. Using the Blahut-Arimoto algorithm combined with a particle method, the positions and probabilities of the optimal mass points were found. The simulation results have shown that the optimal input distribution, if an AP and PA constraint are applied, is always discrete with a finite number of mass points and that always mass points at zero and maximum amplitude appear. Also it was shown that the capacity increases with decreasing noise dimensionality M and increasing PAPR ρ. For alphabets with only two mass point the latter behavior can be seen through the increasing asymmetry of the optimal input alphabet. Furthermore, the results point out that for low E_s/N_0 OOK is always an optimal modulation scheme and for higher E_s/N_0 values higher-order PAM are optimal.

References

[1] E. Leitinger, "Capacity and capactiy-achieving input distribution of the energy detector," Master's thesis, Signal Processing and Speech Communication Laboratory, Graz University of Technology, 2012.

[2] K. Witrisal, G. Leus, G. Janssen, M. Pausini, F. Troesch, T. Zasowski, and J. Romme, "Noncoherent ultra-wideband systems," *Signal Processing Magazine, IEEE*, vol. 26, no. 4, pp. 48 – 66, July 2009.

[3] H. Urkowitz, "Energy detecction of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, April 1967.

[4] T. Quek and M. Win, "Analysis of UWB transmitted-reference communication systems in dense multipath channels," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 9, pp. 1863 – 1874, Septemper 2005.

[5] R. Blahut, "Computation of channel capacity and rate-distortion functions," *Information Theory, IEEE Transactions on*, vol. 18, no. 4, pp. 460 – 473, July 1972.

[6] J. H. G. Dauwels, "Numerical computation of the capacity of continuous memoryless channels," in *Proceedings of the 26th Symposium on Information Theory in the BENELUX*, May 2005, pp. 221 – 228.

[7] ——, "On graphical models for communications and machine learning: Algorithms, bounds, and analog implementation," Ph.D. dissertation, ETH Zürich, 2006.

[8] T. Chan, S. Hranilovic, and F. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *Information Theory, IEEE Transactions on*, vol. 51, no. 6, pp. 2073 – 2088, June 2005.

[9] 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*, April 2010, pp. 63 – 70.

[10] S. Moser, "Capacity results of an optical intensity channel with input-dependent Gaussian noise," *Information Theory, IEEE Transactions on*, vol. 58, no. 1, pp. 207 – 223, January 2012.

[11] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 1991.