Construction of Efficient Amplitude Phase Shift Keying Constellations

Christoph Schmitz
Institute for Theoretical Information Technology
RWTH Aachen University
52056 Aachen, Germany
schmitz@umic.rwth-aachen.de

Anke Schmeink
Institute for Theoretical Information Technology
RWTH Aachen University
52056 Aachen, Germany
schmeink@umic.rwth-aachen.de

***Abstract*—Well-known modulation schemes for additive white Gaussian noise (AWGN) channels are based on uniform input distributions. Thus, their number of constellation points normally is a power of two. In this work, a method for the construction of efficient amplitude phase shift keying (APSK) modulation schemes is presented. The number of constellation points are not restricted to powers of two, and the referring input distributions are not uniform, although dyadic. Compared to the well-known quadrature amplitude modulation (QAM) schemes, they provide shaping gain both by arranging the constellation points in a circle, and due to the usage of non-uniform input distributions. Additionally, they allow a finer adaptation to the channel state. Performance evaluations show that these schemes are often better than the QAM schemes.**

I. Introduction

The transmission of discrete data via a complex AWGN channel requires the use of modulation schemes. The data stream is mapped to a stream of complex input symbols which are transmitted via the channel, and the channel output is decoded to achieve the original input data. To use the channel capacity efficiently, the modulation scheme has to be adapted to the channel state. If the signal-to-noise ratio (SNR) is low, schemes with a small number of constellation points keep the complexity as well as the symbol error probability low, while schemes with a higher number of constellation points are able to transmit more information per symbol when the SNR is higher, see [1].

Conventional approaches such as the well-known quadrature amplitude modulation schemes (QAM) often use a power of two as number of constellation points, and they map a fixed number of bits to a symbol, leading to a uniform channel input distribution in case of a memoryless source producing equiprobable bits. This works quite well and allows for a simple implementation, but is not optimal. To approximate the normal distribution as good as possible, the constellation points should rather be arranged in a circle than in a square or a cross, and the constellation points farther from the origin should be used less frequently than those close to the origin. The basic ideas of signal shaping are well-known, for an overview see for example [2] and [3]. Apart from the signal shaping aspect, restricting the number of constellation points to powers of two (or even to powers of four for square QAM schemes) leads to a somehow coarse adaptation to the channel state, which impedes the optimal usage of a given channel.

In this paper, we present a construction method for APSK modulation schemes which can be employed to transmit discrete data via a complex AWGN channel. The constellation points are placed on equally spaced rings around the origin. The optimal input distribution (which is a normal distribution) is then approximated by a dyadic distribution on these points, using the geometric Huffman coding algorithm proposed in [4]. To evaluate the performance of the different modulation schemes depending on the SNR, we compute the mutual information between the input and the decoded output, and the maximal symbol error probability. In doing so we show that our new approaches are often better than the QAM schemes and some other schemes proposed in the literature, and allow a finer adaptation of the modulation to the channel state.

II. System Model

In this work, our model is a complex AWGN channel with input X, noise term W and output $Y = X + W$, whereupon X and W are stochastically independent, and the latter follows a zero-mean circular symmetric complex normal distribution with variance $E(WW^*) = \sigma_W^2$. Thus, the real and the imaginary part of W are stochastically independent, and both are $\mathrm{N}(0, \sigma_W^2/2)$-distributed. The probability density function of W is

$$f_W(w) = \frac{1}{\pi\sigma_W^2} \exp\left(-\frac{\|w\|^2}{\sigma_W^2}\right). \tag{1}$$

For the discrete input distribution with the finite support set $\mathcal{X} = \{x_1, \ldots, x_M\}$, a power constraint

$$E(XX^*) = \sum_{i=1}^{M} P(X = x_i) \cdot x_i x_i^* = \sigma_X^2 \tag{2}$$

is given.

A coding function $g : \{0,1\}^* \to \mathcal{X}^*$ is used to map the bit stream from the source to a stream of input symbols. The bit stream from the source itself is assumed to be memoryless with a uniform distribution. Although some source encoders like plain Huffman coding often produce distributions that are not exactly uniform, this assumption is reasonable. It should be noted that there even exists a modification of the

ISBN 978-3-8007-3659-1

Huffman coding that enhances the uniformness of the resulting distribution, see [5]. Also, it should be mentioned that there exists a completely different idea to make use of a non-uniform distribution from the source encoder for shaping, see [6].

In the context of the definition of the coding function, $\{0,1\}^*$ denotes a sequence of bits with arbitrary length, and $\mathcal{X}^*$ denotes a sequence of input symbols from $\mathcal{X}$, also with arbitrary length. As one symbol can represent a variable number of bits in our approach, it is not possible to give a simpler definition of the coding function. It is, however, invertible, the sequence of input bits can be reconstructed from the sequence of symbols. This reconstruction is not in the scope of this work, the evaluation of the channel performance is done on the symbol level.

The decoding function $d : \mathbb{C} \to \{1,\ldots,M\}$ is defined by sets $D_1,\ldots,D_M \subseteq \mathbb{C}$ forming a partition of $\mathbb{C}$, such that $d(y) = i$ if and only if $y \in D_i$. In this work, we predominantly use the maximum likelihood (ML) decoding approach, the decoding regions are chosen such that

$$f_{Y|X}(y|x_i) \geq f_{Y|X}(y|x_k) \tag{3}$$

for all $k = 1,\ldots,M$ if $d(y) = i$. This means the decoder chooses the input symbol x_i that maximises the (infinitesimal) probability of the observed channel output y. Due to the zero-mean circular symmetric distribution of the noise term W, the conditional density

$$f_{Y|X}(y|x_i) = f_W(y - x_i) \tag{4}$$

is strictly monotonically decreasing in the Euclidean distance between y and x_i. Thus, the decoding regions correspond to the Voronoi regions around the constellation points. For a given constellation point, the associated Voronoi region is defined as the set of all elements of the complex plane that are closer to this point than to any other constellation point.

Another decoding approach is the maximum a posteriori (MAP) decoding rule, it chooses the decoding regions such that

$$P(X = x_i) \cdot f_{Y|X}(y|x_i) \geq P(X = x_k) \cdot f_{Y|X}(y|x_k) \tag{5}$$

for all $k = 1,\ldots,M$ if $d(y) = i$. That is, the decoder maximises the probability of the input symbol x_i under the observation of the output symbol y with its choice. Using the density of the noise term given in (1), it can be shown that this is equivalent to

$$\|y - x_i\|^2 \leq \|y - x_k\|^2 + \sigma_W^2 \ln\left(\frac{P(X = x_i)}{P(X = x_k)}\right). \tag{6}$$

For uniform input distributions both decoding rules are identical, as

$$\ln\left(\frac{P(X = x_i)}{P(X = x_k)}\right) = \ln(1) = 0 \tag{7}$$

holds for all i and k in that case. When a non-uniform input distribution is used, the decoding regions for constellation points with a higher probability are larger. The border between the decoding regions of two constellation points is still a straight line perpendicular to the connecting line of the points, but it is shifted towards the less probable point (and might even be located *beyond* the less probable point in extreme cases). Note that the amount of this shift does not only depend on the ratio between the input probabilities, but also on the variance of the noise. In practice this means that the receiver must know this value, or at least have an appropriate estimation for it, to apply this decoding rule.

Applying the decoding function to the complex channel output $Y = X + W$ yields the decoded output $\widetilde{Y} = d(Y)$. The conditioned distribution of $\widetilde{Y}$, given the input X, can be computed as

$$P(\widetilde{Y} = k|X = x_i) = \int_{D_k} f_W(y - x_i)dy. \tag{8}$$

A closed-form solution for this integral does not exist. In some rare cases, when the decoding region D_k is the Cartesian product of intervals, existing numerical approximations for the cumulative density function of the one-dimensional normal distribution could be used. In most cases, however, the decoding regions are arbitrarily shaped polygons. So, generally these values have to be calculated by numerical integration or Monte Carlo methods.

The mutual information between the input X and the decoded output $\widetilde{Y}$,

$$I(X,\widetilde{Y}) = H(\widetilde{Y}) - H(\widetilde{Y}|X), \tag{9}$$

is used as the primary performance measure, while the maximal symbol error probability

$$\varepsilon = \max_{1\leq i\leq M} P(\widetilde{Y} \neq i|X = x_i) \tag{10}$$

is also considered.

III. Modulation Schemes

A. Conventional Modulation Schemes

Well-tried modulation schemes for the transmission of discrete input via an AWGN channel are for example the quadrature amplitude modulation (QAM) schemes, with constellation points arranged in a square tiling, see [7]. The whole constellation has either the form of a square, if the number of points is a power of 4 (e.g., 16-QAM or 64-QAM), or the form of a cross (e.g., 32-QAM or 128-QAM). As the number of constellation points always is a power of 2, these schemes are normally used to encode a fixed number of bits per symbol. Assuming a memoryless source producing equiprobable bits, this leads to a uniform distribution of the channel input X.

B. Novel Modulation Schemes

Our process of constructing new modulation schemes consists of two steps. In the first step, we choose an arrangement of constellation points. The points are placed on K rings around the origin, see for example Fig. 2 which will be explained in detail later. The distance between adjacent rings is d_0, this value is also used as an approximate lower bound for the distance between the constellation points. Keeping the spacing of the constellation points more or less uniform ensures that the symbol error probability is also roughly uniform among the different points. This is especially important when using its maximum as a performance measure. To allow the first ring (counting from the origin) to carry four points with this spacing, its radius is chosen as $r_1 = 0.7 \cdot d_0$. Thus, the radius of the i-th ring is $r_i = (i - 0.3) \cdot d_0$. The number of constellation points on the i-th ring, n_i, is chosen such that $n_i \leq 2\pi \cdot r_i$, to ensure the desired spacing. For reasons which will be explained later, n_i is always a power of two. Thus, the total number of constellation points is already determined by K, the number of rings, although there exists no closed formula for this relationship.

The second step consists of constructing the input distribution for this constellation. The idea is to approximate a normal distribution. Due to the ring structure of the APSK schemes, the construction of the distribution does not have to consider the individual constellation points. Instead, only the distribution for the rings has to be determined. Within one ring, the points have a uniform distribution. This approach also ensures that constellation points with the same distance from the origin have the same probability, too. This would not be ensured if we would calculate the probabilities for the individual points, as the approximation procedure sometimes maps equal probabilities to different approximated values when generating a dyadic distribution.

To approximate a normal distribution in the complex plane, the distribution for the rings has to approximate a Rayleigh distribution. This can be seen as follows: Let X be circular symmetric complex normal distributed with $E(X) = 0$ and $E(XX^*) = \sigma_X^2$. Then the radius $R = \sqrt{\|X\|^2}$ is Ray$(\sigma_X^2/2)$-distributed with the probability density function

$$f_R(r) = \frac{2r}{\sigma_X^2} \exp\left(-\frac{r^2}{\sigma_X^2}\right) \quad \text{for } r \geq 0. \tag{11}$$

As about 98% of the mass of this distribution is concentrated between 0 and $2\sqrt{\sigma_X^2}$, the algorithm starts by placing the K rings equally spaced in this interval, setting

$$d_0 = \frac{2\sqrt{\sigma_X^2}}{K + 0.2} \tag{12}$$

initially. Then, the optimal distribution, which is given by

$$p_i' = \begin{cases} \frac{1}{q}\int_0^{r_1+0.5\cdot d_0} f_R(r)dr & i = 1 \\ \frac{1}{q}\int_{r_i-0.5\cdot d_0}^{r_i+0.5\cdot d_0} f_R(r)dr & i = 2, \dots, K \end{cases} \tag{13}$$

with

$$q = \int_0^{r_K+0.5\cdot d_0} f_R(r)dr, \tag{14}$$

is approximated by a dyadic distribution $(p_1, \dots, p_K)$, using the geometric Huffman coding (GHC) algorithm given in [4]. With knowledge of this distribution, the parameter d_0 is finally adjusted such that the power constraint $E(XX^*) = \sigma_X^2$ is fulfilled. As the distribution is chosen as an approximation to a continuous distribution which already fulfils the power

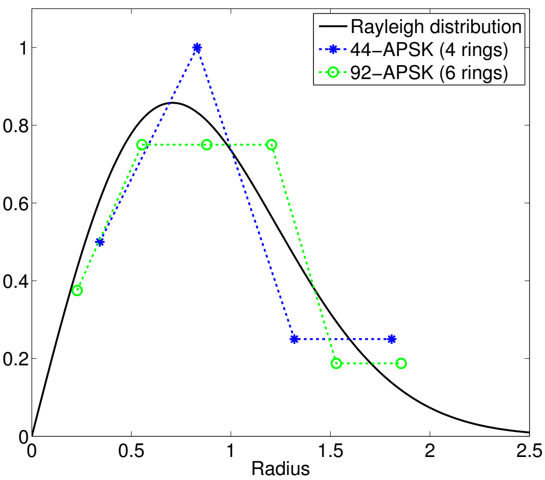

Fig. 1. Rayleigh distribution and approximations

constraint, this adjustment is typically small. Nevertheless, the small shifting of the rings induced by this might lead to the case that the chosen distribution is not the optimal approximation anymore, so the algorithm is repeated if neccessary.

A different choice of the interval for the initial placing does not have a great impact on the resulting distribution. This is due to the fact that the power constraint has to be fulfilled. When starting with a smaller interval, the adjustment to the power constraint expands the distribution by increasing the parameter d_0. When the initial distribution is generated using a much larger interval, the very small probabilities of the outmost rings are often set to zero by the GHC algorithm, leading to a distribution with fewer rings in a smaller interval. While it is possible to produce two or even three different distributions for a given number of rings in some cases, some quick examinations show that their performance does not differ significantly.

Two examples of the resulting distributions are shown in Fig. 1, together with the continuous Rayleigh distribution that is to be approximated. The first example contains four rings, resulting in 44 constellation points, the second example has six rings with 92 constellation points altogether. Note that the probabilities of the rings are scaled proportional to the number of rings to make them comparable to the continuous Rayleigh distribution. The power constraint is $\sigma_X^2 = 1$ in this case. Although the probability of the first ring p_1 is smaller than those of the second ring p_2 in both examples, this does not hold for the individual constellation points, as the first ring carries four points and the second one eight. In general, the probability of an individual constellation point located on the i-th ring is p_i/n_i, the probability of the whole ring divided by the number of points on that ring.

The binary codes assigned to the constellation points consist of two parts, a prefix that identifies the ring, and a second part that identifies the individual point within that ring. The prefixes for the rings are determined by the GHC algorithm that creates the dyadic distribution of the rings. Due to the

ISBN 978-3-8007-3659-1

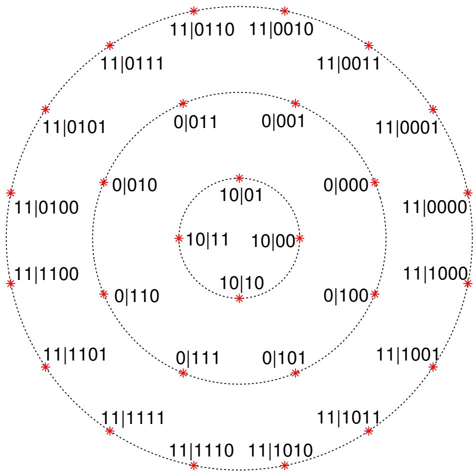

Fig. 2. APSK constellation with three rings

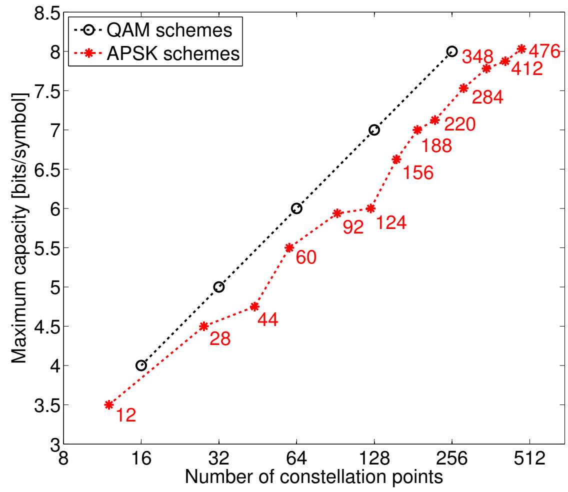

Fig. 3. Maximum capacity of the different modulation schemes

nature of the desired distribution they commonly have different lengths. The length of the second part depends on the number of constellation points in the ring. Thus, it is constant for all points within one ring, but usually differs between the rings. As the number of points in each ring is a power of 2, a binary code is able to induce a uniform distribution.

An example is shown in Fig. 2, in this case the constellation has 28 points placed on three rings. The vertical lines in the code words, which are only shown for the sake of clarity, separate the ring prefix and the second part of the code. The second ring has a probability of 2^{-1}, so it is addressed by a one-digit prefix, in this example '0'. The first and the third ring both have a probability of 2^{-2}, and thus, they have two-digit prefixes, in this case '10' for the first ring and '11' for the third. The length of the second part is determined by the number of constellation points within the ring. Thus, it has two digits for the first ring containing four points, three digits for the second ring containing eight points, and four digits for the third ring containing 16 points. The total length of the code is the same for the constellation points on the first and the second ring in this example, so the probability of the individual points is also the same, namely 2^{-4}. In contrast, the constellation points on the third ring have a probability of 2^{-6} each. Within each ring it is possible to use a Gray mapping, that is, the codes of neighbouring constellation points differ in exactly one digit, like shown in the example.

The maximum capacity of a modulation scheme is the number of bits that can be transmitted by one symbol over a perfect channel with no noise. This value is identical to the entropy $H(X)$ of the input if the latter is computed using the binary logarithm. Although the perfect complex channel itself has an unlimited capacity, the capacities of discrete modulation schemes (with a finite number of constellation points) are always finite. Fig. 3 shows the maximum capacity for 13 different schemes constructed by our proposed method, ranging from two rings (twelve constellation points) to 14 rings (476 constellation points). Of course the QAM schemes (or any other modulation schemes with uniform input distribution) can transmit slightly more bits for a given number of constellation points, as they achieve the theoretical limit given by the binary logarithm of the number of points, but this small advantage is not overly relevant. Although the number of constellation points affects the complexity of a transmission system, its performance on a given channel with a certain signal-to-noise ratio (SNR) is far more important. The results in Section IV show that our new schemes often outperform the QAM schemes in this regard.

All of the novel modulation schemes are proper, that is, they fulfil the property $E(XX) = 0$. The same holds for the QAM schemes mentioned before. This is not relevant for AWGN channels where the noise is assumed to be circular symmetric, but would be important for other channels where the noise does not have this property.

IV. Results and Discussion

To compare the performance of the constructed modulation schemes, the mutual information $I(X;\widetilde{Y})$ between the channel input X and the decoded channel output $\widetilde{Y}$ is computed for different values of the signal-to-noise ratio (SNR) σ_X^2/σ_W^2. This mutual information is upper bounded by two different values. On the one hand, we have

$$I(X;\widetilde{Y}) \le I(X;Y) \le C_{\text{cont.}} = \log\left(1+\frac{\sigma_X^2}{\sigma_W^2}\right), \qquad (15)$$

the mutual information is bounded by the capacity of the channel with continuous input and output. Note that the latter is reached if X follows a zero-mean circular symmetric normal distribution with variance $E(XX^*) = \sigma_X^2$. On the other hand, we know that

$$I(X;\widetilde{Y}) \le H(X), \qquad (16)$$

the mutual information is bounded by the entropy of the input. When computed using the binary logarithm, the latter

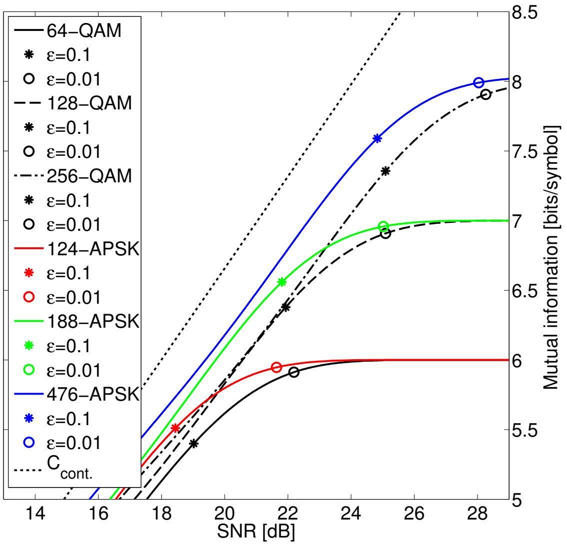

Fig. 4. Mutual information of different modulation schemes

is exactly the average number of transmitted bits per symbol, due to the fact that a symbol coding a bit sequence of length n is chosen with probability 2^{-n}.

Another important performance measure is the maximal symbol error probability ε, see (10). It is especially relevant in practical applications, where it is often desired to keep the error probability under a certain threshold. To reflect this, the points where ε reaches the values 0.1 and 0.01, respectively, are marked in the results.

For most of the 13 schemes considered here, the maximum capacity is not an integer, so it is difficult to compare their performance to the QAM schemes. But then, there are three schemes whose maximum capacity is an integer or at least close to an integer. These are 124-APSK with a maximum capacity of six bits per symbol, 188-APSK with seven bits per symbol and 476-APSK with 8.031 bits per symbol. Fig. 4 compares these three schemes to 64-QAM, 128-QAM and 256-QAM, respectively, showing the mutual information as a function of the SNR. The theoretical Shannon bound $C_{\text{cont.}}$ is also shown in the figure.

In general, the new schemes have a higher capacity than the corresponding QAM schemes when the SNR is low, and about the same when the SNR rises and the maximum capacity is approached. For 124-APSK, the SNR required for an error level of $\varepsilon = 0.1$ is clearly lower than for 64-QAM, the same holds for $\varepsilon = 0.01$. The capacity achieved by the new scheme at the corresponding point is also higher, especially in the former case. Regarding 188-APSK and 128-QAM, the required SNR values for the two relevant error levels are nearly the same, although 188-APSK has a slightly higher capacity at those points, especially in the case $\varepsilon = 0.1$. The comparison between 476-APSK and 256-QAM shows that the new scheme requires a lower SNR value for a given error level of $\varepsilon = 0.1$ and $\varepsilon = 0.01$, respectively. Additionally, the capacity achieved by the new scheme at this point is clearly higher, especially in the first case. Altogether, the new schemes are able to reduce the capacity gap to the Shannon bound substantially when used in the proper SNR range.

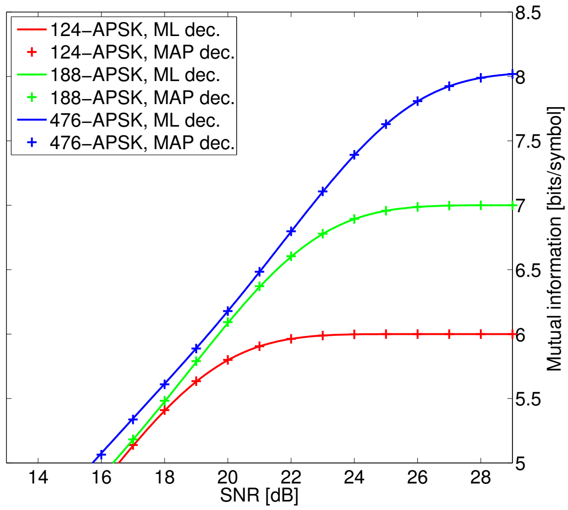

Fig. 5. Mutual information of different decoding rules

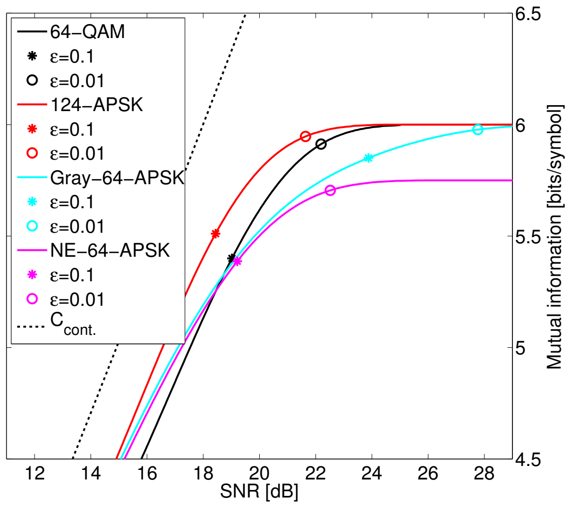

Fig. 6. Mutual information of some modulation schemes

The above results were achieved using the maximum likelihood (ML) decoding rule. While the maximum a posteriori (MAP) decoding rule might be better in theory, comparisons of the practical results show that the difference is negligible, and there are even cases when the MAP decoding is slightly worse. A comparison of both decoding rules is shown in Fig. 5, considering the modulation schemes 124-APSK, 188-APSK and 476-APSK. Due to the high computational complexity and the resulting runtime of the simulation, results for the MAP decoding were only calculated for the integer values of the SNR. Although a closer examination of the results shows small differences, the crosses of the MAP results appear to be located exactly on the lines of the ML results within the accuracy of the plot, so the difference is indeed negligible.

Some different new approaches to construct APSK mod-

ISBN 978-3-8007-3659-1

ulation schemes are given in [8] and [9]. In contrast to our approach, in [8] the authors propose the usage of a uniform input distribution, together with a number of constellation points that is a power of two. Also, the number of rings and the number of points in each ring (which is the same for all rings) are powers of two. Therefore it is possible to use Gray mapping not only inside each ring, but on the whole modulation scheme, like it can be done for square QAM schemes. The downside of this approach is the tight spacing of the constellation points on the inner rings. To overcome this, the authors in [9] propose to reduce the number of constellation points on the inner rings by mapping more than one input bit vector to one point. Although this reduces the maximum capacity, it leads to a more uniform spacing of the constellation points. The resulting input distribution on the constellation points is not uniform anymore in this case, although the distribution on the rings still is.

Fig. 6 compares two of these schemes to 64-QAM and our proposal 124-APSK, and to the theoretical Shannon bound $C_{\text{cont.}}$. The scheme designated as Gray-64-APSK here can be found in [8], it has four rings, each with 16 constellation points. This results in 64 constellation points with a uniform distribution, thus the maximum capacity is six bits per symbol. NE-64-APSK is described in [9]. It is based on Gray-64-APSK, but the number of constellation points on the innermost ring is reduced to eight, so in fact is has just 56 constellation points. While the points on the other rings still have a probability of 2^{-6}, it is 2^{-5} for the points on the innermost ring. Thus, the maximum capacity is reduced to 5.75. Although these schemes are a bit better than 64-QAM for low SNR values (with high symbol error probability), they behave worse when the SNR is higher. If an error level of $\varepsilon = 0.1$ is required, NE-64-APSK needs about the same SNR level as 64-QAM to fulfil this, and the resulting capacity is also roughly comparable. For $\varepsilon = 0.01$ the SNR level is still about the same among those two schemes, but NE-64-APSK has a much lower capacity in this case. Due to the tight spacing of some of the constellation points, Gray-64-APSK requires a much higher SNR level to achieve comparable error levels, and the capacity is lower than those of 64-QAM in the relevant SNR range. Compared to 124-APSK, these two schemes perform even slightly worse for very low SNR values. Please note that the authors in [8] and [9] use the average symbol error probability instead of the maximum as a performance measure, thus schemes with some tightly spaced constellation points seem to perform better than in our analysis.

V. Conclusion and Outlook

We have proposed a general method for the construction of APSK modulation schemes with dyadic input distributions. As such distributions can be generated by a prefix code, these schemes can be used efficiently for the transmission of discrete data via a complex channel. Compared to established modulation schemes like QAM, which place equiprobable constellation points in a square or a cross pattern, the schemes generated by our method place the constellation points on rings around the origin, which leads to a circular pattern. Additionally, the probability of the constellation points is variable, this allows for a better approximation of the normal distribution which would be the optimal input distribution. Nonetheless, our method ensures that the spacing of the constellation points is roughly uniform. Thus, there are no tightly spaced constellation points that lead to an unneccessary high symbol error probability.

Some of the new schemes have a maximum capacity comparable to existing QAM schemes or other recently proposed APSK modulation schemes. In these cases a comparison is possible, and it reveals that our new schemes perform better than the QAM schemes, and also better than the recent proposals for APSK schemes. The SNR value required for a given threshold for the symbol error probability is always a bit lower, and the capacity for the SNR value is higher, too. Those schemes that are not comparable to existing schemes can be seen as complements. As such, they allow a finer adaptation to the channel state, so for a given SNR value a higher capacity can often be reached.

Maximum a posteriori (MAP) decoding theoretically seems to make sense for modulation schemes with non-uniform input distributions, but our results show that the advantages over maximum likelihood (ML) decoding are marginal and do not justify the increased computational complexity that is required.

Acknowledgment

This work was supported by the UMIC Research Centre at RWTH Aachen University and by DFG grant SCHM 2643/4-1.

References

[1] A. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *Communications, IEEE Transactions on*, vol. 45, no. 10, pp. 1218 –1230, Oct 1997.

[2] G. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *Selected Areas in Communications, IEEE Journal on*, vol. 2, no. 5, pp. 632–647, Sep 1984.

[3] A. Calderbank and L. Ozarow, "Nonequiprobable signaling on the gaussian channel," *Information Theory, IEEE Transactions on*, vol. 36, no. 4, pp. 726–740, Jul 1990.

[4] G. Böcherer and R. Mathar, "Matching dyadic distributions to channels," in *Data Compression Conference (DCC)*, Snowbird, USA, Mar. 2011, pp. 23–32.

[5] F. Altenbach, G. Böcherer, and R. Mathar, "Short huffman codes producing 1s half of the time," in *International Conference on Signal Processing and Communication Systems (ICSPCS'11)*, Honululu, Hawaii, Dec. 2011, pp. 1–5.

[6] M. Valenti and X. Xiang, "Constellation shaping for bit-interleaved LDPC coded APSK," *Communications, IEEE Transactions on*, vol. 60, no. 10, pp. 2960–2970, October 2012.

[7] J. R. Davey, "Modems," *Proceedings of the IEEE*, vol. 60, no. 11, pp. 1284–1292, 1972.

[8] Z. Liu, Q. Xie, K. Peng, and Z. Yang, "APSK constellation with gray mapping," *Communications Letters, IEEE*, vol. 15, no. 12, pp. 1271–1273, 2011.

[9] F. Yang, K. Yan, Q. Xie, and J. Song, "Non-equiprobable APSK constellation labeling design for BICM systems," *Communications Letters, IEEE*, vol. 17, no. 6, pp. 1276–1279, June 2013.

[10] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley and Sons, Inc., 1991.

