Constellation Optimization for Coherent Optical Channels Distorted by Nonlinear Phase Noise

Christian Häger†, Alexandre Graell i Amat†, Alex Alvarado‡, and Erik Agrell†
†Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden
‡Department of Engineering, University of Cambridge, UK
{christian.haeger, alexandre.graell, agrell}@chalmers.se, alex.alvarado@ieee.org

***Abstract*—We consider the design of amplitude phase-shift keying (APSK) constellations, targeting their application to coherent fiber-optical communications. Phase compensation is used at the receiver to combat nonlinear phase noise caused by the Kerr-effect. We derive the probability density function of the post-compensated observation for multilevel constellations. Optimal APSK constellations in terms of symbol error probability (SEP) are found assuming a two-stage detector. Performance gains of $3.2\ \mathrm{dB}$ can be achieved compared to 16-QAM at a SEP of 10^{-2}. We optimize the number of rings, the number of points per ring, as well as the radius distribution of the constellation. For low to moderate nonlinearities, radius optimization only yields minor improvements over an equidistant spacing of rings. In the highly nonlinear regime, however, a smaller SEP can be achieved by "sacrificing" the outer ring of the constellation, in favor of achieving good SEP in the remaining rings.**

I. Introduction

Fiber nonlinearities are considered to be one of the limiting factors for achieving high data rates in coherent long-haul optical transmission systems [1]–[3]. Therefore, a good understanding of the influence of nonlinearities on the system behavior is necessary in order to increase data rates of future optical transmission systems. In this paper, we consider dispersionless transmission and ideal distributed Raman amplification. The probability density function (PDF) of the corresponding discrete memoryless channel has been derived in [3]–[6]. This model captures the interaction of Kerr-nonlinearities with the signal itself and the inline amplified spontaneous emission noise, giving rise to nonlinear phase noise (NLPN) [7].

The design of signal constellations in the presence of NLPN has been studied before. In [8], the authors applied several predistortion as well as postcompensation techniques in combination with minimum-distance detection for quadrature amplitude modulation (QAM) to mitigate NLPN. They also proposed a two-stage (TS) detector consisting of a radius detector, an amplitude-dependent phase compensation, and a phase detector. Moreover, parameter optimization was performed with respect to four 4-point, custom constellations under maximum likelihood (ML) detection. In [9], the TS detector was used to optimize the radius distribution of four 16-point constellations for two power regimes. It was shown in [9] that the optimal distribution highly depends on the transmit power.

In the first part of this paper, we analyze in detail the (suboptimal) TS detector in [8]. We regard radius detection and phase compensation as a separate processing block that yields a postcompensated observation and we derive the corresponding PDF for multilevel constellations. To the best of our knowledge, this PDF has not been previously derived, possibly due to the fact that the symbol error probability (SEP) under TS detection can be calculated with a simplified PDF [10]. We use the new PDF to gain insights into the performance behavior of the TS detector with respect to optimal detection. We also explain why this PDF is necessary to calculate the average bit error probability for systems employing this type of compensation at the receiver.

In the second part of this paper, we find optimal amplitude phase-shift keying (APSK) constellations in terms of SEP under TS detection for a given input power and fiber length. In contrast to [9], we optimize the number of rings, the number of points per ring, as well as the radius distribution. Our results suggest that the widely employed 16-QAM constellation has poor performance compared to the best found constellations over a wide range of input powers for this channel model. We also provide numerical support for an optimal signal shaping strategy in the presence of strong nonlinearities, which is analytically described in [11].

II. System Model

We consider the discrete memoryless channel model

$$Y = (X + Z)e^{-\jmath\Phi_{\mathrm{NL}}}, \tag{1}$$

where $X \in \mathcal{X}$ denotes the complex channel input, $\mathcal{X}$ the signal set or constellation, Z the total additive noise, Y the observation, and Φ_{NL} the NLPN, which is given by [3, Ch. 5]

$$\Phi_{\mathrm{NL}} = \frac{\gamma L}{K} \sum_{i=1}^{K} |X + Z_i|^2. \tag{2}$$

Here, γ is the nonlinear Kerr-parameter, L is the total length of the fiber, K denotes the number of fiber segments, and Z_i is the noise contribution of all fiber segments up to segment i. More precisely, $Z_i \triangleq N_1 + \ldots + N_i$ is defined as the sum of i independent and identically distributed complex Gaussian random variables with zero mean and variance σ_0^2

This work was partially funded by the Swedish Agency for Innovation Systems (VINNOVA) under the P36604-1 MAGIC project, Sweden, and by The British Academy and The Royal Society (via the Newton International scheme), UK. The calculations were performed in part on resources provided by the Swedish National Infrastructure for Computing (SNIC) at C3SE.

TABLE I
CONSTANTS AND PARAMETERS

symbol	value	meaning
n_{sp}	1.41	spontaneous emission factor
ν	$1.936 \cdot 10^{14}$ Hz	optical carrier frequency
h	$6.626 \cdot 10^{-34}$ J s	Planck's constant
α	0.0578 km^{-1}	fiber loss (0.25 dB/km)
γ	1.2 W^{-1}km^{-1}	nonlinearity parameter
$\Delta\nu$	42.7 GHz	optical bandwidth
L	5500 km	fiber length

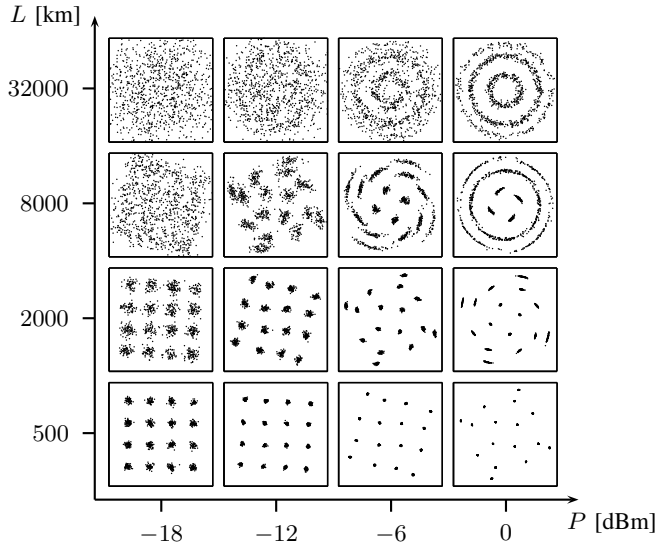


Fig. 1. Scatter plots of Y assuming $X \in \mathcal{X}_{\text{16-QAM}}$ for several combinations of transmit power and fiber length.

per dimension (real and imaginary parts). We denote the total additive noise of all K fiber segments by $Z \triangleq Z_K$, which has variance $\sigma^2 \triangleq \mathbb{E}\left[|Z|^2\right] = 2K\sigma_0^2$, where $\mathbb{E}[\cdot]$ is the expected value. For ideal distributed amplification, we consider the case $K \to \infty$. The total noise variance can be calculated as $\sigma^2 = 2n_{sp}h\nu\alpha\Delta\nu L$ [12, Sec. IX-B], where all parameters are taken from [8] and summarized in Table I.

An important aspect of this channel model is the fact that the variance of the phase noise is dependent on the channel input (cf. (2)), or equivalently on the average transmit power P, defined as $P \triangleq \mathbb{E}\left[|X|^2\right]$. In order to gain some insight into the qualitative behavior of the channel, in Fig. 1 we show the received scatter plots assuming that $X \in \mathcal{X}_{\text{16-QAM}}$, where $\mathcal{X}_{\text{16-QAM}} \triangleq \{\sqrt{P/10}(a+\jmath b) \,:\, a,b \in \{\pm 1, \pm 3\}\}$ is the 16-QAM constellation, and $K = 100$ for several combinations of P and L. It can be observed that for very low input power and fiber length, nonlinearities are negligible and the channel behaves as a regular additive white Gaussian noise (AWGN) channel. The scatter plots along a diagonal in Fig. 1 correspond to a constant signal-to-(additive-)noise ratio (SNR), defined as SNR $\triangleq P/\sigma^2$. In contrast to an AWGN channel for which the scatter plots along any diagonal would look similar, the received constellation points in Fig. 1 start to rotate in a deterministic fashion and the effect of the NLPN becomes pronounced for large L and P. Therefore, in order to specify the operating point of the channel, the SNR alone is not sufficient, because the parameter space of the channel is two-dimensional, cf. [6, Sec. VII]. In this paper, we assume a fixed fiber length and variable transmit power, i.e., we move along a horizontal line in Fig. 1.

We adopt the term *amplitude phase-shift keying (APSK)* for the discrete-input constellations considered in this paper[1]. APSK constellations can be described as a union of phase-shift keying (PSK) signal sets with different amplitude levels. We consider APSK constellations with $M = 16$ points and define the APSK signal set as

$$\mathcal{X} \triangleq \left\{ r_k e^{\jmath \frac{2\pi i}{l_k}} \,:\, 0 \le i \le l_k - 1, \quad 1 \le k \le N \right\}, \quad (3)$$

where N denotes the number of amplitude levels or rings, r_k is the radius of the kth ring, $l_k \ge 1$ denotes the number of points in the kth ring, and $\sum_{k=1}^{N} l_k = M$. This is a particular case of the general APSK signal set (cf. [13, Sec. III]) where each ring has zero phase offset. This can be done without loss of generality, because, as we explain in Sec. III-D, the SEP for the assumed TS detector is not affected by a phase offset in any of the rings. We assume a uniform distribution on the channel input X over all symbols, and thus $P = (1/M)\sum_{k=1}^{N} l_k r_k^2$. The radii are assumed to be ordered such that $r_1 < \ldots < r_N$ and we use the vector $\boldsymbol{r} \triangleq (r_1, \ldots, r_N)$ to denote the *radius distribution*. The radius distribution is uniform if $r_{k+1} - r_k = \Delta$ for $1 \le k \le N-1$, where $\Delta = r_2$ if $l_1 = 1$ and $\Delta = r_1$ if $l_1 \ge 2$.

We also define the vector $\boldsymbol{l} \triangleq (l_1, \cdots, l_N)$ and use the notation $\boldsymbol{l}$-APSK for an APSK constellation with N rings and l_k points in the kth ring, e.g., (4,4,4,4)-APSK. Note that this notation does not specify a particular constellation without ambiguity, due to the missing information about the radii.

III. DETECTION METHODS

A. Maximum Likelihood Detection

Let the polar representation of the channel input and the observation be given by $X = R_0 e^{\jmath\Theta_0}$ and $Y = Re^{\jmath\Theta}$, respectively. The PDF of Y can be written in the form [6, Sec. III], [3, p. 225], [8]

$$f_{Y|X=x}(y) = \frac{f_{R|R_0=r_0}(r)}{2\pi r} + \frac{1}{\pi r}\sum_{m=1}^{\infty} \Re\left\{ C_m(r, r_0) e^{\jmath m(\theta - \theta_0)} \right\}, \quad (4)$$

where $x = r_0 e^{\jmath\theta_0}$, $y = re^{\jmath\theta}$, $\Re\{z\}$ is the real part of $z \in \mathbb{C}$, the coefficients $C_m(r, r_0)$ can be found in [6, Sec. III], and the PDF of the received amplitude R given the transmitted amplitude $R_0 = r_0$ is

$$f_{R|R_0=r_0}(r) = \frac{2r}{\sigma^2} \exp\left(-\frac{r^2 + r_0^2}{\sigma^2} \right) \mathrm{I}_0\left(\frac{2rr_0}{\sigma^2} \right), \quad (5)$$

[1]APSK constellations are used for example in recent satellite communication standards due to their flexibility and robustness against nonlinear amplifier distortions, see [13] and references therein.

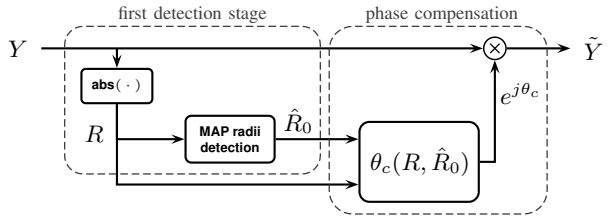


Fig. 2. Block diagram of the postcompensation.

where $\mathrm{I}_0(\cdot)$ is the modified Bessel function of the first kind. The ML detector can be described in the form of decision regions $\mathcal{R}_i^{\mathrm{ML}} \subset \mathbb{C}$ for each symbol $x_i \in \mathcal{X}$, $1 \le i \le M$:

$$\mathcal{R}_i^{\mathrm{ML}} \triangleq \bigcup_{\substack{j=1 \\ j \neq i}}^{M} \{y \in \mathbb{C} \,:\, f_{Y|X=x_i}(y) \ge f_{Y|X=x_j}(y)\}. \tag{6}$$

Note that for any detector, the average SEP is defined as

$$\mathrm{SEP} \triangleq 1 - \frac{1}{M} \sum_{i=1}^{M} \int_{\mathcal{R}_i} f_{Y|X=x_i}(y)\, \mathrm{d}y, \tag{7}$$

where $\mathcal{R}_i$ is the decision region for the symbol x_i. If we take the ML decision regions defined in (6), (7) becomes a lower bound on the achievable SEP with suboptimal detectors. However, evaluating the SEP by numerically integrating over the PDF is computationally expensive. Moreover, unlike for an AWGN channel, the ML decision regions (6) depend on the transmit power [8].

B. Two-Stage Detection

The detector we study in this paper is a slightly modified version of the suboptimal TS detector proposed in [8]. This detector has much lower complexity compared to ML detection. In particular, the TS detector employs one-dimensional decisions: first in the amplitude direction (first detection stage), followed by a phase compensation, and then in the phase direction (second detection stage).

In Fig. 2, we show a block diagram of the first detection stage together with the phase compensator. We refer to the processing depicted in Fig. 2 as *postcompensation*. Based on the absolute value of the observation R, maximum a posteriori (MAP) radius detection is performed. The radius detector implements the rule: choose $\hat{R}_0 = r_k$, when $\mu_{k-1} \le R < \mu_k$; where μ_k, $0 \le k \le N$, denote the decision radii or thresholds with $\mu_0 \triangleq 0$ and $\mu_N \triangleq \infty$. In contrast to [8], we use MAP instead of ML radius detection and make use of the a priori probability for selecting a certain ring at the transmitter, thereby achieving a small performance advantage. After that, based on the radius $\hat{R}_0$ of the detected ring and the received amplitude R, a correction angle θ_c is calculated, by which the observation Y is rotated to obtain the postcompensated observation $\tilde{Y}$ as shown in Fig. 2. The correction angle is given by

$$\theta_c(R, \hat{R}_0) = -\angle C_1(R, \hat{R}_0), \tag{8}$$

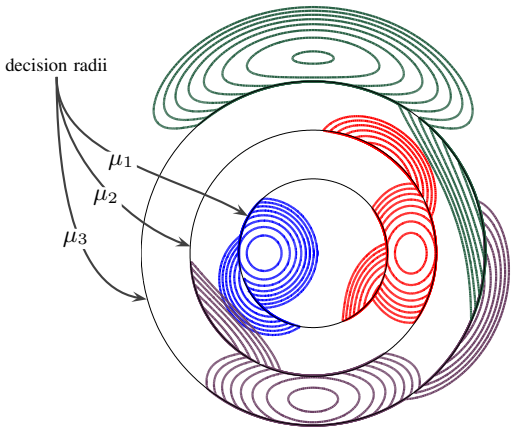


Fig. 3. PDF of $\tilde{Y}$ for $P = -5$ dBm conditioned on one particular point in each ring of the uniform (4,4,4,4)-APSK constellation. Color is helpful.

which is approximately a quadratic function in R [8]. The second detection stage (not shown in Fig. 2) is then performed with respect to $\tilde{Y}$: a phase detector chooses the constellation point with radius $\hat{R}_0$ that is closest to $\tilde{Y}$.

It is shown in [8] that for PSK signal sets (where $\hat{R}_0 = \sqrt{P} = \text{const.}$), a minimum-distance detector for $\tilde{Y}$ is equivalent to ML detection. However, this is not the case for multilevel constellations such as APSK, as we discuss in Sec. III-D.

C. PDF of the Postcompensated Observation

An important observation is that the postcompensation in Fig. 2 is invertible, and thus it is a lossless operation. This implies that, in principle, *optimal* detection of X is still possible based on $\tilde{Y}$ due to the fact that every invertible function forms a sufficient statistic for detecting X based on Y [14, p. 443]. In the following, we derive the PDF of the postcompensated observation $\tilde{Y}$. It is clear from the block diagram of Fig. 2 that the PDF can be written as

$$f_{\tilde{Y}|X=x}(\tilde{y}) = f_{Y|X=x}\left(\tilde{y} \cdot e^{-\jmath\theta_c(R,\hat{R}_0)}\right). \tag{9}$$

The correction angle θ_c is a discontinuous function of the amplitude R because it depends on the detected ring $\hat{R}_0$. Let us assume that a constellation point from the kth ring is transmitted. For simplicity, we only consider the possibility that the radius detector decides either for the transmitted ring or the adjacent rings. At most, a ring has two adjacent ring neighbors, whose radii are given by r_{k-1} and r_{k+1}. The decision thresholds between r_k and these rings are denoted by μ_{k-1} and μ_k. The correction angle can then be written as

$$\theta_c(R, \hat{R}_0) = \begin{cases} \theta_c(R, r_{k+1}) & \text{for } R \ge \mu_k \\ \theta_c(R, r_k) & \text{for } \mu_{k-1} \le R < \mu_k \\ \theta_c(R, r_{k-1}) & \text{for } R < \mu_{k-1} \end{cases}. \tag{10}$$

When the inner ($k = 1$) and outer ring ($k = N$) of the constellation are considered, only one adjacent ring is present and (10) has to be adjusted accordingly.

For illustration purposes, we plot in Fig. 3 the PDF resulting from (9) and (10) conditioned on one particular point in each ring of the uniform $(4, 4, 4, 4)$-APSK constellation[2]. If we consider the PDF conditioned on $X = r_2$, i.e., $R_0 = r_2$ and $\Theta_0 = 0$ (shown in red), it can be observed that the contour lines look as though they have been sliced up along the decision radii of the radius detector, because for $R < \mu_1$ and $R \geq \mu_2$, the correction angle is calculated with respect to r_1 and r_3, respectively. We will refer to the PDF of the postcompensated observation $\tilde{Y}$ in (9) as a *sliced* PDF.

D. Decision Regions

In Fig. 4(a), the PDF in (4) is plotted for the $(4, 4, 4, 4)$-APSK constellation together with the ML decision regions. In Fig. 4(b), the sliced PDF in (9) is used instead. Finally, Fig. 4(c) shows the same PDF as Fig. 4(b) together with the suboptimal decision regions implemented by the TS detector. For clarity reasons, the decision region corresponding to the point $X = r_3 e^{\jmath\pi/2}$ is shaded in Fig. 4. Comparing the optimal decision regions in Fig. 4(b) with the decision regions in Fig. 4(c), it can be seen that TS detection is clearly suboptimal for this constellation and input power. However, one would expect the two small shaded regions in Fig. 4(b) to become smaller for higher power. Intuitively, this is explained by the increasing accuracy of the radius detector for increasing transmit power, due to the Rice distribution (5) of the amplitude. In fact, it can be shown that the SEP under TS detection converges to the SEP under ML detection as $P \to \infty$, assuming an APSK constellation with uniform radius distribution.

Note that the SEP under TS detection can be computed analytically using (7) with the TS decision regions and the PDF (9). To this end, it is sufficient to assume a simplified PDF as was done in [10], where the slicing effect is not taken into account because it only affects the PDF *outside* the TS decision regions. Also, the SEP under TS detection does not change assuming a phase offset in any of the rings: the PDF of $\tilde{Y}$ is simply rotated by the phase offset, but so is the detector region itself, and hence the integrals in (7) are not affected.

We remark that the sliced PDF is, however, crucial to accurately calculate the average bit error probability, because the evaluation of symbol *transition* probabilities, i.e., $\Pr[\hat{X} = x_j | X = x_i]$, $1 \leq i, j \leq M$, $i \neq j$, where $\hat{X}$ denotes the detected symbol, requires knowledge of the correct PDF conditioned on $X = x_i$ *outside* the decision region for x_i.

IV. Constellation Optimization

In this section, we optimize the parameters of APSK constellations by minimizing the SEP under TS detection. The optimization problem can be stated as: Given M,

$$\begin{aligned} \underset{N,\boldsymbol{l},\boldsymbol{r}}{\text{minimize}} \quad & \text{SEP} && (11)\\ \text{subject to} \quad & 1 \leq N \leq M && (12)\\ & l_1 + \ldots + l_N = M && (13)\\ & l_1 r_1^2 + \ldots + l_N r_N^2 = PM && (14)\\ & l_k \leq 1, 1 \leq k \leq N. && (15) \end{aligned}$$

The number of ways to distribute M points to N rings is given by $\binom{M-1}{N-1}$. At most there are $M = 16$ rings, which gives $2^{M-1} = 2^{15} = 32768$ possibilities to choose N and $\boldsymbol{l}$.

A. Uniform Radius Distribution

In a first study, we constrain $\boldsymbol{r}$ to be uniform and optimize only over N and $\boldsymbol{l}$, cf. (11)−(15). The optimization problem then becomes an integer program which we solve in an exhaustive fashion. The input power P is varied from -14 dBm to 10 dBm in steps of 2 dBm and the results are shown in Fig. 5, where we indicate the optimal constellation next to the corresponding marker of the curve. To avoid cumbersome notation, we define $\mathbf{1}_N \triangleq 1, \ldots, 1$ (N times). Note that for the optimized constellation $r_1 = 0$ whenever $l_1 = 1$, i.e., for $l_1 = 1$ it is always better to place the constellation point at the origin than in the first ring with nonzero radius. The SEP under TS detection for 16-QAM, $(4, 4, 4, 4)$-APSK, and $(\mathbf{1}_{16})$-APSK are shown for comparison, while the SEP under ML detection for 16-QAM in an AWGN channel is shown as a reference. $(\mathbf{1}_{16})$-APSK can be regarded as a type of pulse amplitude modulation (PAM), for which the SEP can be calculated as

$$\text{SEP} = 1 - \frac{1}{M}\left(1 - \sum_{i=1}^{15} Q_1\left(\tilde{r}_i, \tilde{\mu}_i\right) + \sum_{i=2}^{16} Q_1\left(\tilde{r}_i, \tilde{\mu}_{i-1}\right)\right), \quad (16)$$

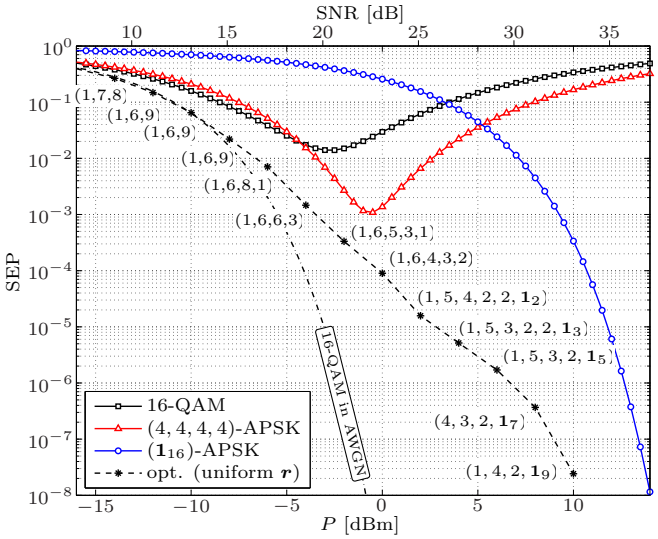
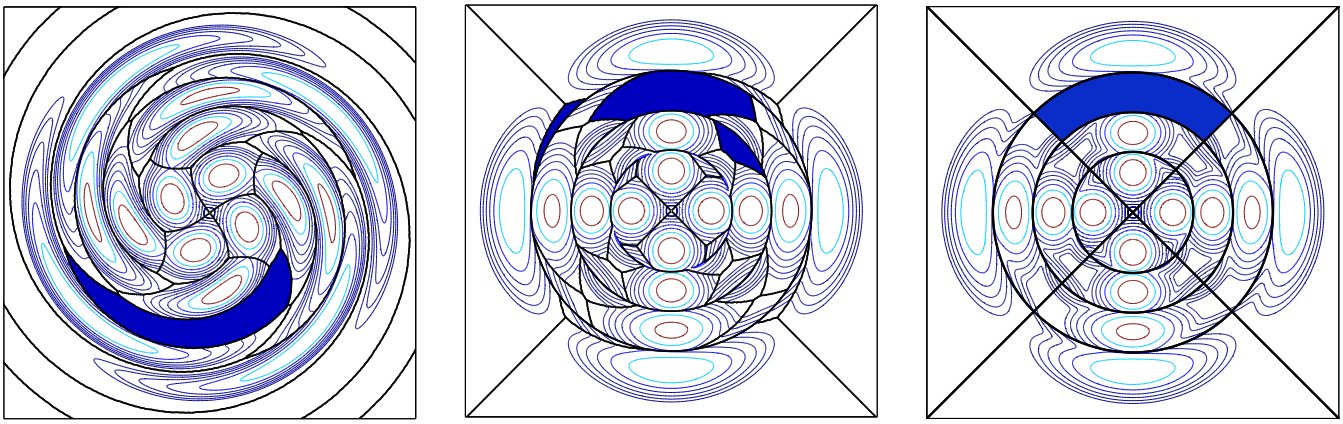


Fig. 5. Results for the APSK constellation optimization. The SEP under ML detection for 16-QAM in an AWGN channel (dash-dotted line) is shown as a reference.

[2]The PDFs of the points which are not shown look identical to the PDF of the corresponding point in the same ring up to a phase rotation.



(a) ML decision regions w.r.t. Y (b) ML decision regions w.r.t. $\tilde{Y}$ (c) TS decision regions

Fig. 4. Decision regions for the uniform $(4, 4, 4, 4)$-APSK constellation at $P = -4$ dBm. Shaded regions correspond to the point $X = r_3 e^{\jmath\pi/2}$.

where $Q_1(\cdot)$ is the Marcum Q-function and $\tilde{r}_i \triangleq \sqrt{2} r_i/\sigma$ and $\tilde{\mu}_i \triangleq \sqrt{2}\mu_i/\sigma$.[3] In [8], it was argued that this constellation is optimal for very high input power. Moreover, we remark that this constellation is the *only* APSK constellation with a uniform radius distribution for which the SEP is decreasing with input power, because the error probability of the phase detector will eventually increase for more than one point per ring. The results in Fig. 5 show that up to an input power of -8 dBm, the performance of the optimized constellations follows closely the performance of 16-QAM in AWGN. In other words, for this power regime it is possible to find APSK constellations with TS detection that perform as well as 16-QAM for a channel without nonlinear impairments under optimal detection. For higher input power, the optimized constellations gradually utilize more amplitude levels, due to the increase in NLPN. If we take as a baseline the minimum SEP achieved by the 16-QAM constellation ($P \approx -2.8$ dBm and SEP $\approx 10^{-2}$), and interpolate the optimal APSK performance for the same SEP, we observe that a performance gain of 3.2 dBm is achieved. Moreover, while the performance of 16-QAM and $(4, 4, 4, 4)$-APSK exhibit a minimum after which the performance degrades with increasing input power, the optimized constellations have a SEP that always decreases.

B. Radius Optimization

In this section, we consider the optimization of the radius distribution $\boldsymbol{r}$ for a constellation with fixed N and $\boldsymbol{l}$. The optimization problem then becomes a nonlinear program where the dimensionality of the search space is $N - 1$, due to the power constraint. We conjecture that this problem is nonconvex, i.e., a local optimum does not necessarily imply a global solution. We use sequential quadratic programming (SQP) [15] to solve the optimization, although the solution is not guaranteed to be the globally optimal. However, we verified the global optimality for some constellations with a brute-force grid search.

[3]Note that for $(\mathbf{1}_{16})$-APSK, the TS detector can be replaced by a one-stage detector without phase compensation and phase detection.

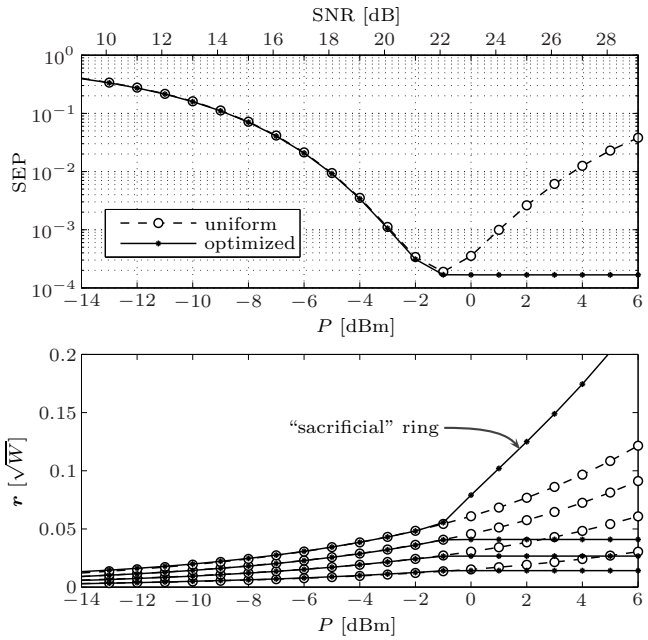


Fig. 6. Performance of the $(1, 6, 5, 4, 1)$-APSK constellation with a uniform and optimized radius distribution (top) and the corresponding radius distribution (bottom).

As an example, in Fig. 6 we show the results for $(1, 6, 5, 4, 1)$-APSK, which is the optimal constellation with a uniform radius distribution for $P = -2$ dBm (cf. Fig. 5). For $P < -1$ dBm, the optimal radius distribution closely follows a uniform distribution and the SEP curves consequently overlap. For $P > -1$ dBm, however, the optimal ring spacing changes significantly. For this power regime, all rings except the outer one remain at the same amplitude level and any increase in average power is completely absorbed by putting the outermost ring further away from the other rings. This has an interesting effect on the SEP, which then flattens out, instead of increasing. We performed an optimization of the radius distribution for several APSK constellations with similar results for the

highly nonlinear regime as shown in Fig. 6, i.e., for high enough input power, it becomes optimal to "sacrifice" the outer ring of the constellation. The SEP flattening effect is, however, limited to constellations with $l_N = 1$.

Although the findings above may have little practical utility (optimally wasting power is still a waste), they do have theoretical implications for the problem of constellation shaping in the presence of severe nonlinear distortion [11]. Our results show that radius optimization may lead to rather unexpected results for high nonlinearities. Therefore, a priori assumptions on the radius distribution as in [16], where the authors impose a constraint on the radius distribution of so-called ring constellations in order to reduce the complexity of the optimization, might be too restrictive to adequately reflect the true performance that is possible to achieve for very high input power.

V. Conclusion

In this paper, we optimized APSK constellations for a simplified fiber-optical channel model without dispersion. We derived the PDF of the postcompensated observation which can for example be used to correctly calculate bit error probabilities for arbitrary constellations and subsequently find good coding schemes. Compared to a baseline 16-QAM constellation, significant performance improvements in terms of uncoded SEP can be achieved by choosing an optimized APSK constellation. Radius optimization yields only negligible improvements for low and moderate input powers in terms of SEP. For very high input power, it may become beneficial to sacrifice a constellation ring in order to reduce the average SEP of the constellation.

Acknowledgment

The authors would like to thank L. Beygi and F. Brännström for many helpful discussions.

References

[1] I. Djordjevic, W. Ryan, and B. Vasic, *Coding for Optical Channels*. Springer, 2010.

[2] G. Agrawal, *Nonlinear Fiber Optics, Fourth Edition*, 4th ed. Academic Press, 2006.

[3] K. Ho, *Phase-modulated Optical Communication Systems*. Springer, 2005.

[4] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Physical Review Letters*, vol. 91, no. 20, p. 203901, Nov. 2003.

[5] A. Mecozzi, "Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers," *J. Lightw. Technol.*, vol. 12, no. 11, pp. 1993–2000, Nov. 1994.

[6] M. I. Yousefi and F. R. Kschischang, "On the per-sample capacity of nondispersive optical fibers," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7522–7541, Nov. 2011.

[7] J. P. Gordon and L. F. Mollenauer, "Phase noise in photonic communications systems using linear amplifiers," *Optics Letters*, vol. 15, no. 23, pp. 1351–1353, Dec. 1990.

[8] A. P. T. Lau and J. M. Kahn, "Signal design and detection in presence of nonlinear phase noise," *J. Lightw. Technol.*, vol. 25, no. 10, pp. 3008–3016, Oct. 2007.

[9] L. Beygi, E. Agrell, and M. Karlsson, "Optimization of 16-point ring constellations in the presence of nonlinear phase noise," in *Optical Fiber Communication Conference (OFC)*, Los Angeles, CA, Mar. 2011.

[10] L. Beygi, E. Agrell, P. Johannisson, and M. Karlsson, "A novel multilevel coded modulation scheme for fiber optical channel with nonlinear phase noise," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Miami, FL, Dec. 2010.

[11] E. Agrell, "The channel capacity increases with power," *arXiv:1108.0391v2*, Aug. 2011. [Online]. Available: http://arxiv.org/abs/1108.0391

[12] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[13] R. De Gaudenzi, A. Guillén i Fàbregas, and A. Martinez, "Performance analysis of turbo-coded APSK modulations over nonlinear satellite channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2396–2407, Sep. 2006.

[14] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge University Press, 2009.

[15] P. Boggs and J. Tolle, "Sequential quadratic programming," *Acta numerica*, vol. 4, no. 1, pp. 1–51, 1995.

[16] T. Freckmann, R.-J. Essiambre, P. J. Winzer, G. J. Foschini, and G. Kramer, "Fiber capacity limits with optimized ring constellations," *IEEE Photon. Technol. Lett.*, vol. 21, no. 20, pp. 1496–1498, Oct. 2009.