Pairwise Optimization of Modulation Constellations*

Brendan Moore, Glen Takahara and Fady Alajaji
Dept. of Mathematics and Statistics
Queen's University
Kingston, Ontario, Canada
{brendan,takahara,fady}@mast.queensu.ca

Abstract

The design of two-dimensional signal constellations for the transmission of binary non-uniform memoryless sources over additive white Gaussian noise channels is investigated. The main application of this problem is the implementation of improved constellations where transmitted data is highly non-uniform. A simple algorithm, which optimizes a constellation by re-arranging its points in a pairwise fashion (i.e., two points are modified a time with the other points remaining fixed), is presented. In general, the optimized constellations depend on both the source statistics and the signal-to-noise ratio (SNR) in the channel. We show that constellations designed with source statistics considered can yield symbol error rate performance that is substantially better than rectangular quadrature amplitude modulation signal sets used with either Gray mapping or more recently developed maps. Gains as high as 5 dB in E_b/N_0 SNR are obtained for highly non-uniform sources.

1. Introduction

For uniformly distributed sources, rectangular quadrature amplitude modulation (QAM) using Gray mapping is known to perform well, and is shown as optimal in terms of bit error rate (BER) for high enough signal-to-noise ratios (SNR) [1]. As noted in [11], however, there are many real-world examples of data sources which are highly non-uniform, such as text (email and instant/short messages), medical images and encoded voice data [2]. Compression will often have residual redundancy in the output due to non-ideal coding methods [3]. Rather than using traditional source and channel coding (which can be sensitive to noise-related errors in decoding if optimal variable-length source coding is used), we can choose instead to directly exploit the non-uniformity of the source via the modulation scheme, while gaining noise-resiliency in many cases and significantly reducing system complexity and delay [3]. Such an approach can be quite attractive for complexity-constrained and delay-sensitive applications such as wireless sensor networks. In these non-uniform situations, the performance of Gray mapped M-ary rectangular QAM is sub-optimal. One simple improvement is to exploit the knowledge of symbol probability by implementing (optimal) maximum *a posteriori* (MAP) decoding (instead of maximum-likelihood decoding) at the receiver. In [11], new $M1$-mappings were developed to improve performance of M-ary rectangular QAM and phase-shift keying constellations. It is also noted in [11] that performance can be improved by translating each mapped constellation so that it has zero mean. Here we consider making further changes to the constellations in order to achieve lower symbol error rate (SER). In [5], such a constellation design problem was considered for uniform sources under additive white Gaussian noise (AWGN).

In this paper, we propose a method for redesigning M-ary constellations to better exploit the non-uniformity in the data, for large values of M, under AWGN and MAP decoding. The method, which is simple to implement, consists of iteratively improving the performance of a constellation by re-arranging its points two at a time, while keeping the other points fixed. We verify our work by comparing it to the known optimal constellations in [7] for $M = 2$, and in [9] for $M = 4$, before considering larger constellations. Other related works on constellation design include [4, 6, 10, 12].

The remainder of this paper is organized as follows.

*This work was supported in part by funding provided by the R. Samuel McLaughlin Fellowship and Herman K. Walter Award from Queen's University and NSERC of Canada.

In Section 2, we describe the problem. We explain our method for optimizing these constellations in Section 3. In Section 4, we present the results of our work, and compare performance to existing constellations and mappings. In Section 5, we draw our conclusions and present directions for future work.

2. Problem Statement

We consider a memoryless source $\{X_n\}$ which generates independent binary symbols $\{0, 1\}$ non-uniformly with $p = Pr\{X_n = 0\} > \frac{1}{2}$. We wish to transmit this data over an AWGN channel with noise variance of $\frac{N_0}{2}$ per dimension. We assume that an M-ary two-dimensional (2-D) modulation scheme is to be used, and that it is desirable to maximize data throughput per transmission while achieving the lowest possible SER. For convenience, we assume M to be a power of two. Binary symbols are grouped into sequences of $\log_2 M$ bits, forming a new symbol sequence $\{Y_n\}$ having M distinct values $\{s_1, s_2, ..., s_M\}$ with probabilities $\{p_1, p_2, ..., p_M\}$. The probabilities are defined by the number of zeros in the bit sequence. If sequence s_i has n_i zeros, then $p_i = p^{n_i}(1-p)^{\log_2 M - n_i}$. (In the constellation diagrams that come later, we refer to equiprobable symbols by the number of zeros, n, they have in their corresponding binary sequence.) Each channel symbol is then mapped to a signal point, $\vec{s_i}$, in some initial M-ary constellation, where $\vec{s_i} = (s_{i,x}, s_{i,y})$. Our objective is then to change the arrangement of the points in that constellation to achieve the lowest SER possible at a given SNR E_b/N_0, where E_b is the average energy per bit.

The search space to be considered is continuous and consists of all collections of points $\{\vec{s_1}, \vec{s_2}, ..., \vec{s_M}\}$ satisfying

(i) a zero mean constraint: $\sum_{i=1}^{M} p_i \vec{s_i} = 0$; and

(ii) an average power constraint: $\sum_{i=1}^{M} p_i \|\vec{s_i}\|^2 = E$,

where the average energy per symbol, E, is given. Note that E and E_b are related by $E_b = \frac{E}{\log_2 M}$. Our objective function is the SER. For $M = 2$, the optimal constellation was found analytically in [7], but as the constellation size grows, this quickly becomes difficult. In [9], the authors design optimal constellations for $M = 4$ by numerically evaluating tight error bounds developed in [8]. Our goal is to design signal point arrangements that are near-optimal for larger constellation sizes, such as $M = 16, 64, 256$, under MAP decoding.

3. Pairwise Optimization

In this section, we consider a new method for developing improved signal constellations for 2-D transmission. While the search space is continuous, the zero mean and average power constraints may be used to reduce the search complexity. The zero mean constraint is a necessary property of any optimal (in terms of minimal SER) constellation with constrained average energy, since SER performance under MAP decoding is not affected by translation or rotation of the constellation; it is only affected by changing the relative distances between points. It is of note that for non-uniform sources, rectangular (symmetric) constellations such as 16-, 64- and 256-QAM are not zero mean. It is trivial to improve such constellations slightly by translating them to be zero mean, and scaling them up to their original average energy (which will increase the separation between all points).

For a given initial constellation, it is not possible to adjust the position of a single point while adhering to the above two constraints. Taking any pair of points, however, allows us to move those points around while still adhering to the constraints. If $\vec{s_1}$ and $\vec{s_2}$ are the selected points, then the zero mean constraint implies that

$$p_1\vec{s_1} + p_2\vec{s_2} = -\sum_{i=3}^{M} p_i\vec{s_i}$$

so, if we let $\vec{b} = \sum_{i=3}^{M} p_i\vec{s_i}$, then

$$\vec{s_1} = \frac{1}{p_1}(-\vec{b} - p_2\vec{s_2})$$

or

$$\vec{s_1} = \vec{a} - c\vec{s_2}$$

where $\vec{a} = -\frac{\vec{b}}{p_1}$ and $c = \frac{p_2}{p_1}$. Thus

$$s_{1,x} = a_x - c \cdot s_{2,x} \text{ and } s_{1,y} = a_y - c \cdot s_{2,y}. \tag{1}$$

The average energy constraint implies the following:

$$p_1\|\vec{s_1}\|^2 + p_2\|\vec{s_2}\|^2 = E - \sum_{i=3}^{M} p_i\|\vec{s_i}\|^2. \tag{2}$$

Letting the constant $d = \sum_{i=3}^{M} p_i\|\vec{s_i}\|^2$ and substituting (1) in (2) yields

$$\begin{aligned} p_1\left((a_x - c \cdot s_{2,x})^2 + (a_y - c \cdot s_{2,y})^2\right) \\ +p_2(s_{2,x}^2 + s_{2,y}^2) = E - d. \end{aligned} \tag{3}$$

Expanding and completing the square gives us

$$\left(s_{2,x} - \frac{p_1 a_x}{p_1 + p_2}\right)^2 + \left(s_{2,y} - \frac{p_1 a_y}{p_1 + p_2}\right)^2 = r^2 \tag{4}$$

where $r^2 = \frac{p_1(E-d)}{p_2(p_1+p_2)} - \frac{p_1^3}{p_2(p_1+p_2)^2}\left(a_x^2 + a_y^2\right)$. Under the constraints, Eqn. (4) gives us a circle, centered at $\left(\frac{p_1 a_x}{p_1+p_2}, \frac{p_1 a_y}{p_1+p_2}\right)$ with radius r, on which $\vec{s_2}$ may travel, and the relationship given by Eqn. (1) defines a corresponding circle for $\vec{s_1}$. With (4), for each pair of signals $(\vec{s_1}, \vec{s_2})$, the problem of searching over four variables $(s_{1,x}, s_{1,y}, s_{2,x}, s_{2,y})$ is effectively reduced to searching over a single variable, θ, which is the angle parametrizing this circle for $\vec{s_2}$, measured counterclockwise relative to the positive x-axis for the center of the circle. For a given value of θ, $\vec{s_2}$ is defined, and $\vec{s_1}$ has a corresponding position. It is over this parameter θ that each pair of points can be optimized for performance.

With regards to the performance for a potential constellation, we consider the union upper bound[1] on the SER P_s, which is fairly tight for medium to high SNRs:

$$
\begin{aligned}
P_s &= \sum_{u=1}^{M} P(\epsilon|\vec{s_u})P(\vec{s_u}) \\
&= \sum_{u=1}^{M} P\left(\bigcup_{i\neq u}\epsilon_{iu}\right)P(\vec{s_u}) \\
&\leq \sum_{u=1}^{M}\sum_{i\neq u} P(\epsilon_{iu})P(\vec{s_u}) \qquad (5)
\end{aligned}
$$

where

$$P(\epsilon_{iu}) = Q\left(\frac{\|\vec{s_i}-\vec{s_u}\|}{\sqrt{2N_0}} + \frac{\sqrt{2N_0}\ln\frac{P(\vec{s_u})}{P(\vec{s_i})}}{2\|\vec{s_i}-\vec{s_u}\|}\right)$$

and $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy$ is the Gaussian Q-function. Note that $P(\epsilon_{iu})$ is the probability that $\vec{s_i}$ has a larger MAP decoding metric than $\vec{s_u}$ given that $\vec{s_u}$ was sent [8].

When considering only the pair of points $\vec{s_1}$ and $\vec{s_2}$, we can ignore the terms in Eqn. (5) for $u \neq 1, 2$ and $i \neq 1, 2$ as they will remain constant. The remaining terms we use as an upper bound are

$$
\begin{aligned}
F_{12} &= \textstyle\sum_{i\neq 1} P(\epsilon_{i1})P(\vec{s_1}) + \sum_{i\neq 2} P(\epsilon_{i2})P(\vec{s_2}) \\
&\quad + \textstyle\sum_{u=3}^{M} P(\vec{s_u})\left(P(\epsilon_{1u}) + P(\epsilon_{2u})\right) \qquad (6)
\end{aligned}
$$

which is the objective function to be minimized for each pair.

3.1. Algorithm

The implemented algorithm is as follows:

[1]To keep things simple, we herein employ the union bound which can be inaccurate for low SNRs. However, the tight upper and lower bounds of [8] can also be used to further improve system performance.

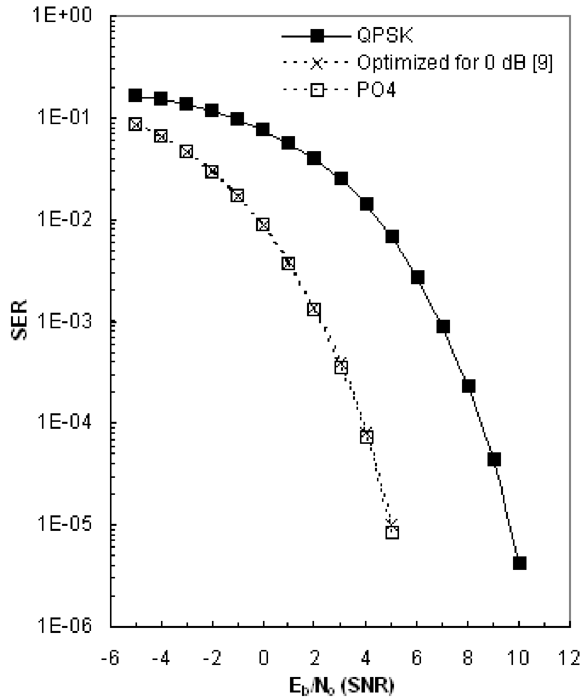

Figure 1. Performance of $\mathrm{M} = 4$ constellations for $\mathrm{p} = 0.9$. Optimized from [9] and PO4 are both designed for $\mathrm{SNR} = 0\,\mathrm{dB}$.

1. Configure some initial constellation, ensuring it adheres to the zero mean and average energy constraints.

2. Randomly (uniformly) select a pair of points $(\vec{s_1}, \vec{s_2})$.

3. Calculate the constrained circles from (4) and (1).

4. Find the positions of $(\vec{s_1}, \vec{s_2})$ by minimizing (6).

5. Go back to Step 2 and repeat until the constellation stabilizes.

The initial constellation used in Step 1 contains the source information implicitly through the symbol probabilities. Tests using different initial constellations (rectangular, circular, asymmetric) all yielded similar results. In Step 4, we calculate the circle noted in Eqn. (4) and set angle θ to be 0 relative to the x-axis, and take discrete steps counterclockwise. At each step of θ, F_{12} is calculated using the corresponding $\vec{s_1}$ and $\vec{s_2}$ on their respective circles, and the design SNR (E_b/N_0), which is set as a constant. This is a simple and brute-force approach, but it works well enough for our needs. The complexity of the algorithm can be approximated by the number of times we calculate the Gaussian Q-function. For each pair of points being optimized, we

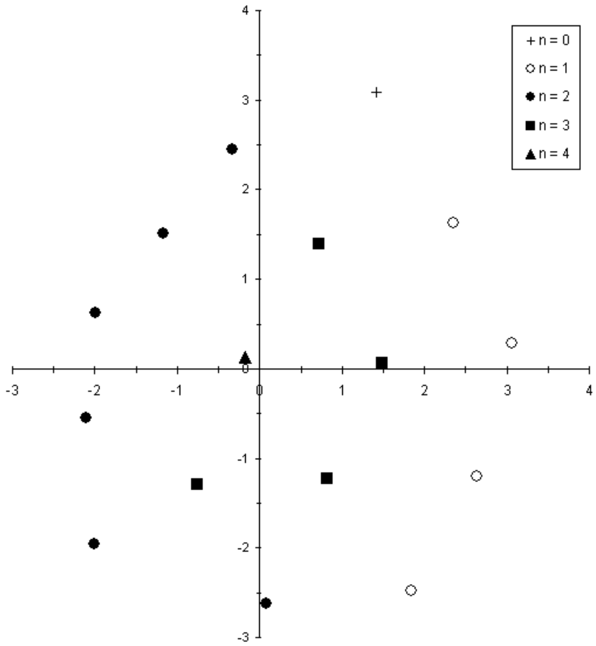

Figure 2. Pairwise optimized constellation for $\mathbf{M = 16}$, $\mathbf{p = 0.9}$ and design $\mathbf{SNR = 1\,dB}$.

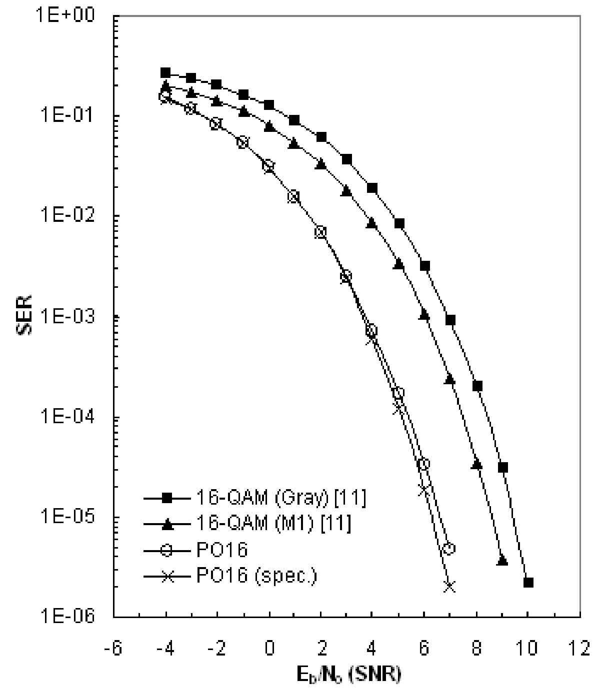

Figure 3. Performance of $\mathbf{M = 16}$ constellations for $\mathbf{p = 0.9}$ and design $\mathbf{SNR = 1\,dB}$. Performance of a specialized constellation (i.e., with design SNR identical to true SNR) also shown.

calculate F_{12} for 50 steps of θ, each of which requires $4M$ calls to $Q(\cdot)$ as in (6), or $200M$ calls per pair. We need roughly M^2 pairs before good constellations are achieved, for a total of $200M^3$ calls (each call takes approx. 3 μs on our 3.0 GHz AMD hardware). When executed, our algorithm stabilizes in a matter of seconds for sizes up to $M = 16$, and scales up to three or four hours for $M = 256$. Stabilization, as used in Step 5, means visual inspection of the constellation at this point. When considering the speed of convergence, it is difficult to be precise, since we do not know what the optimal constellation looks like, or even the final PO constellation for larger sizes. In general, the more likely symbols settle quickly, but the large number of unlikely symbols in large constellations tend to continue rearranging (with better performance at each step) for much longer.

4. Numerical Results and Discussion

We consider the memoryless non-uniform binary source with distribution p for transmission over an AWGN channel for $M = 2, 4, 16, 64, 256$ and compare the performance (in terms of SER simulations) under symbol-by-symbol MAP decoding of our pairwise optimized constellations (which are denoted by PO2, PO4,$\cdots$, PO256) to existing constellations. We use $p = 0.9$ in the simulations, except for the discussion at the end of the Section 4.2.

4.1. Binary and Quaternary Constellations

We begin by comparing to the known optimal constellation presented in [7] for $M = 2$. Our algorithm directly arrives at the same final constellation as the work in [7], as shown in Eqn. (3) with both $\vec{a} = \vec{0}$ and $d = 0$ (since we have no symbols beyond $\vec{s_1}$ and $\vec{s_2}$):

$$p\left((-c \cdot s_{2,x})^2 + (-c \cdot s_{2,y})^2\right) + (1-p)(s_{2,x}^2 + s_{2,y}^2) = E$$

and we choose the point with $s_{2,y} = 0$, so

$$p(-c \cdot s_{2,x})^2 + (1-p)s_{2,x}^2 = E$$

$$s_{2,x} = \sqrt{\frac{E}{pc^2 + (1-p)}} = \sqrt{\frac{E \cdot p}{(1-p)}}$$

$$s_{1,x} = -c \cdot s_{2,x} = -\sqrt{\frac{E \cdot (1-p)}{p}}$$

which is the result obtained in [7]. Note that for $M = 2$, the union bound in (5) yields the exact SER. While the pairwise algorithm is not limited to one dimension, the results are equivalent after rotation. Simulation confirms an exact SER performance match, as expected. There is no consideration of design SNR for $M = 2$,

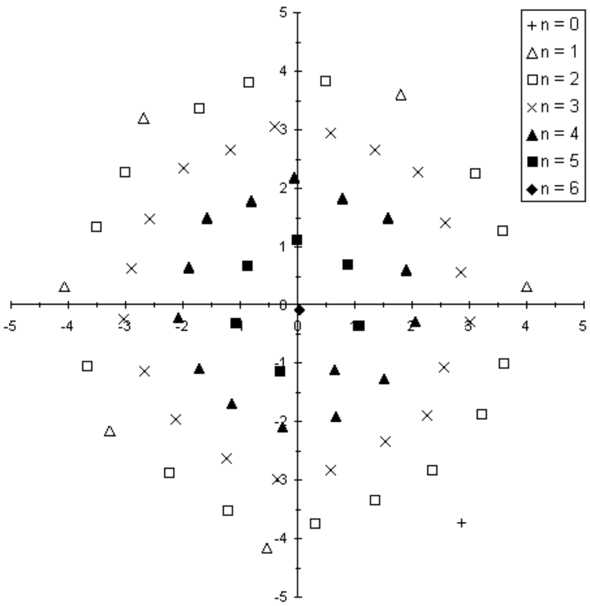

Figure 4. Pairwise optimized constellation for $\mathbf{M = 64}$, $\mathbf{p = 0.9}$ and design $\mathbf{SNR = 2\,dB}$.

because the constraints alone fix the relative positions of $\vec{s_1}$ and $\vec{s_2}$, and we have no other points with respect to which we may optimize.

We next consider the constellations found in [9] for $M = 4$. When the pairwise optimization stabilizes, the resulting constellation is very similar to those arrived at in [9] for the given design SNR (in this case $SNR = 0\,dB$), up to a rotation and/or reflection. In Fig. 1, it is clear that the pairwise optimized constellation PO4 performs identically to the optimized $M = 4$ constellation of [9]. Both constellations perform considerably better than quaternary phase shift keying (QPSK) for highly non-uniform sources, with nearly 5 dB gain at any SNR. The above results indicate that the algorithm does in fact tend towards optimal constellations, and we may proceed to apply it to larger modulation constellations, where optimal constellations are not known.

4.2. 16-ary Constellations and Robustness

Before investigating large constellations, we will examine the performance of the 16 point constellation. The pairwise optimized constellation is shown in Fig. 2. In Fig. 3, the M1 mapping of [11] already improves the performance of (rectangular) 16-QAM by approximately 1 dB. We can also see that the pairwise optimized constellation PO16 achieves a further improvement of 2 dB over the M1 mapping, for a total gain of 3 dB over Gray mapped 16-QAM. Also included in Fig. 3 is the performance at each true SNR step of a specialized constellation designed specifically for that noise level. This specialized configuration does not provide considerable gains over a constellation designed at a single low-mid SNR (in this case, $1\,dB$) and used across all noise levels. This shows that a constellation designed using a single appropriately chosen design SNR provides robust performance *vis-a-vis* changes in the true SNR.

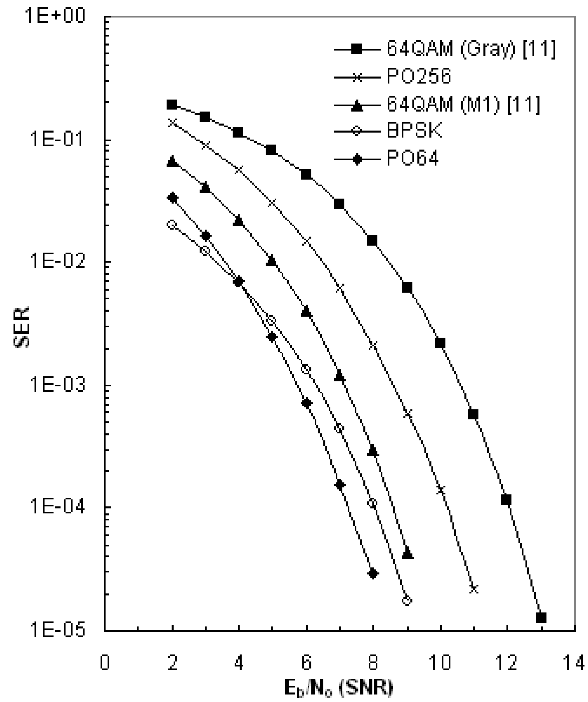

Figure 5. Performance of constellations for $\mathbf{M = 64}$, $\mathbf{p = 0.9}$ and design $\mathbf{SNR = 2\,dB}$ and the pairwise optimized constellation for $\mathbf{M = 256}$ with design $\mathbf{SNR = 4\,dB}$. BPSK also shown as reference point.

While we have not included the plots here, gains over Gray-mapped 16-QAM are also achieved by PO16 for smaller values of p. For $p = 0.5$, the gain achieved was negligible, as expected. With $p = 0.6$, the gain is about $0.25dB$. For $p = 0.7$ and $p = 0.8$, more significant gains of $0.5\,dB$ and $1.5\,dB$ are achieved, respectively.

4.3. 64-ary and 256-ary Constellations

The result of pairwise optimization of a 64 point constellation using design $SNR = 2\,dB$ is shown in Fig. 4. Again we see the tendency of more likely points to lay closer to the origin. This keeps the average energy low, allowing less likely points to sit farther away, thus creating more distance between points overall. In Fig. 5, we compare the performance of this constellation to 64-QAM with the M1 and Gray mappings. The $M1$ 64-

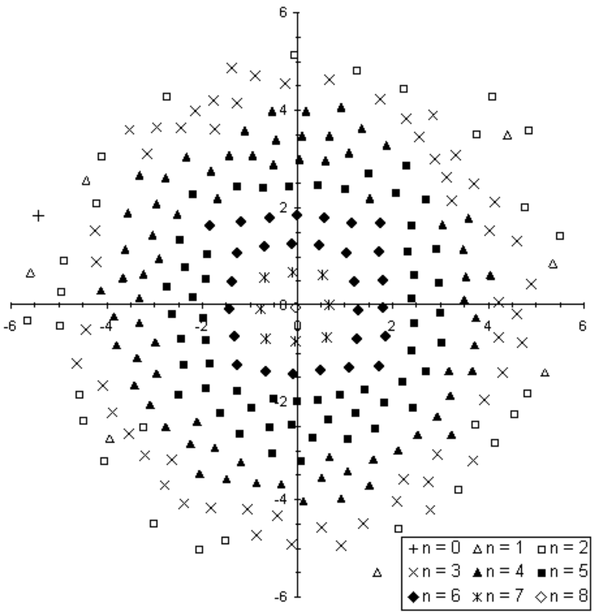

Figure 6. Pairwise optimized constellation for $\mathbf{M = 256}$, $\mathbf{p = 0.9}$ and design $\mathbf{SNR = 4\,dB}$.

QAM mapping developed in [11] already outperforms Gray mapped 64-QAM by approximately 3.5 dB for any given SER. The pairwise optimized constellation we develop here, PO64, outperforms 64-QAM with $M1$ mapping by another $1.5\,dB$ at a given SER, for a total improvement of about $5\,dB$ over 64-QAM with Gray mapping. It is interesting to note that for medium and high SNRs (above $4\,dB$), the PO64 constellation achieves better SER than the BER of binary phase shift keying (BPSK). It is likely that the BER of PO16 will be lower than that of BPSK for sufficiently high values of p. The performance of the pairwise optimized constellation for $M = 256$ (PO256 in Fig. 6) is better than 64-QAM with Gray map by approximately $2dB$ for any SER. Note that PO256 has both a higher data rate and a lower SER than Gray mapped rectangular 64-QAM at all SNRs, thus improving both system performance and throughput.

5. Conclusions

It is clear that the pairwise optimized constellations offer significant gains over traditional (rectangular QAM) modulation constellations for highly non-uniform sources. This is especially true for high rate constellations, where much energy is spent needlessly on likely symbols. We recognize that asymmetric non-rectangular constellations introduce additional complexity in the (de)modulation process. Smaller improvements can be easily obtained by re-centering the traditional rectangular constellations to be zero mean. Here we have only considered SER performance of these constellations.

Further work will involve designing good mappings for these improved constellations to achieve even better results in terms of BER.

References

[1] E. Agrell, E. Strom, and T. Ottossom. Gray Coding for Multilevel Constellations in Gaussian Noise. *IEEE Transactions on Information Theory*, IT-53(1):224–235, January 2007.

[2] F. Alajaji, N. Phamdo, and T. Fuja. Channel Codes that Exploit the Residual Redundancy in CELP-Encoded Speech. *IEEE Transactions on Speech Audio Processing*, 4:325–336, Sept. 1996.

[3] F. Behnamfar, F. Alajaji, and T. Linder. MAP Decoding for Multi-Antenna Systems with Non-Uniform Sources: Exact Pairwise Error Probability and Applications. *IEEE Transactions on Communications*, 57(1):242–254, January 2009.

[4] S. Emami and S. L. Miller. Nonsymmetric Sources and Optimum Signal Selection. *IEEE Transactions on Communications*, 44(4):440–447, April 1996.

[5] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein. Optimization of Two-Dimensional Signal Constellations in the Presence of Gaussian Noise. *IEEE Transactions on Communications*, COM-22(1):28–38, January 1974.

[6] J. Huang, S. Meyn, and M. Médard. Error Exponents for Channel Coding With Applications to Signal Constellation Design. *IEEE Journal on Selected Areas of COmmunications*, 24(8):1647–1661, August 2006.

[7] I. Korn, J. P. Fonseka, and S. Xing. Optimal Binary Communication With Nonequal Probabilities. *IEEE Transactions on Communications*, 51(9):1435–1438, September 2003.

[8] H. Kuai, F. Alajaji, and G. Takahara. Tight Error Bounds for Nonuniform Signaling over AWGN Channels. *IEEE Transactions on Information Theory*, 46(7):2712–2718, November 2000.

[9] H. Nguyen and T. Nechiporenko. Quarternary Signal Sets for Digital Communications with Nonuniform Sources. *IEEE CCECE/CCGEI*, pages 2085–2088, May 2005.

[10] Y. Sun. Stochastic Iterative Algorithms for Signal Set Design for Gaussian Channels and Optimality of the L2 Signal Set. *IEEE Transactions on Information Theory*, 43(5):1574–1587, September 1997.

[11] G. Takahara, F. Alajaji, N. C. Beaulieu, and H. Kuai. Tight Error Bounds for Nonuniform Signaling over AWGN Channels. *IEEE Transactions on Information Theory*, 51(3):400–408, March 2003.

[12] N. Wei and Y. Wan. Optimal Constellation for General Rectangular PAM/QAM with Arbitrary Code Mapping. In *Proceedings of IEEE International Conference on Communications*, pages 2749–2754, Glasgow, Scotland, June 24-28 2007.