On the optimality of the binary reflected Gray code

Erik Agrell, Johan Lassing, Erik G. Ström, and Tony Ottosson

***Abstract*— This paper concerns the problem of selecting a binary labeling for the signal constellation in M-PSK, M-PAM, and M-QAM communication systems. Gray labelings are discussed and the original work by Frank Gray is analyzed. As is noted, the number of distinct Gray labelings that result in different bit-error probability grows rapidly with increasing constellation size. By introducing a recursive Gray labeling construction method called expansion, the paper answers the natural question of what labeling, among all possible constellation labelings, that will give the lowest possible average probability of bit errors for the considered constellations. Under certain assumptions on the channel, the answer is that the labeling proposed by Gray, the binary reflected Gray code, is the optimal labeling for all three constellations, which has, surprisingly, never been proved before.**

***Index Terms*— binary reflected Gray code, constellation labeling, phase shift keying, pulse amplitude modulation, quadrature amplitude modulation, average distance spectrum**

I. INTRODUCTION

THIS PAPER concerns the problem of selecting a binary labeling for the signal vectors that will minimize the probability of bit errors in communication systems employing PSK, PAM, or QAM signal constellations. The aim of the work is to find the best way of labeling these constellations if a low bit-error probability is desired. The discussion starts with the case of an M-PSK constellation and, although this constellation is known to be impractical for large M, it will be shown to provide a natural foundation for the PAM and QAM cases.

First, we introduce a system model and make a brief review of some useful expressions for the error probability of general signal constellations. A signal constellation $\mathcal{S}$ is a set of M points in n-dimensional space, $\mathcal{S} = \{\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{M-1}\}$. During each transmission interval, the transmitter selects a signal point $\mathbf{s} = \mathbf{s}_k$ for transmission over the communication channel. The transmitted signal point $\mathbf{s}$ is displaced from its original position due to the influence from the channel, and the decision device in the receiver observes the received signal point $\mathbf{r} \in \mathbb{R}^n$. The decision device operates based on a non-overlapping partitioning of the signal space into decision regions. The decision region for a signal point $\mathbf{s}_k$ is labelled Ω_k, and whenever the received signal point $\mathbf{r}$ falls in Ω_k, the receiver makes the decision that $\mathbf{s}_k$ was sent.

In this paper, we make the assumption that the distortion from the channel amounts to an additive (signal-independent) and symmetrical noise component and that the receiver operates based on a maximum-likelihood (ML) partitioning of the signal space [1]. We will regard the channel and the receiver as fixed and focus on the effect that the mapping of binary strings onto symbols has on the error performance.

The authors are with the Communication Systems group at the department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden (e-mail: johan.lassing@s2.chalmers.se).

A. Probability of Symbol Errors

If the received signal point $\mathbf{r}$ belongs to Ω_k, the decision device will decide that the transmitted signal was $\mathbf{s}_k$. An error will occur whenever $\mathbf{r}$ falls in another decision region than the decision region belonging to the actual transmitted signal point. When this occurs we experience a *symbol* error and the probability of this is

$$P_{\mathrm{s}} = \sum_{j=0}^{M-1} \Pr\left(\mathbf{s} = \mathbf{s}_j, \mathbf{r} \notin \Omega_j\right).$$

B. Probability of Bit Errors

In this paper, we pay particular attention to the problem where the signal constellation contains $M = 2^m$ signal points. This communication system transmits m bits with each transmission and we assume that each signal point $\mathbf{s}_k$ is assigned a label, $\mathbf{c}_k$, of m binary digits. For such a system, a *bit* error will occur whenever a bit differs between the label $\mathbf{c}_k$ associated with the transmitted point $\mathbf{s}_k$ and the label $\mathbf{c}_l$ of the decision $\mathbf{s}_l$. The bit error probability is

$$P_{\mathrm{b}} = \sum_{l=0}^{M-1} \sum_{k=0}^{M-1} \frac{d_H(\mathbf{c}_k, \mathbf{c}_l)}{m} \Pr\left(\mathbf{s} = \mathbf{s}_k, \mathbf{r} \in \Omega_l\right) \qquad (1)$$

where $d_H(\mathbf{c}_k, \mathbf{c}_l)$ is the Hamming distance (i.e., the number of differing bits) between labels $\mathbf{c}_k$ and $\mathbf{c}_l$. In (1) we have used the fact that by definition $d_H(\mathbf{c}_k, \mathbf{c}_k) = 0$. From this expression it is evident that there are two parts that affect the bit error probability; one part depends on the labeling of the signal points and the other depends on the channel, the transmitter, and the receiver.

C. Gray constellation labelings

The labeling of the signal constellation is at the hands of the system designers and their choices affect the performance of the communication system. The most commonly encountered binary labeling, both in theory and in practice, is the binary reflected Gray code suggested by Frank Gray in a patent from 1953 as a means of reducing the coding error in a pulse code communication system [2]. In Gray's patent the labels of the signal constellation are referred to as a *code*. A more appropriate name for this sequence of labels is a *labeling*, emphasizing the fact that the ordering of the labels is important for the labeling, whereas an unordered set of labels is usually called a code. In this paper, we keep the established term "Gray code" for historical reasons, although we consider it a labeling, not a code.

The system described by Gray in [2] can be viewed as an analog-to-digital converter, in which an analog signal controls the deflection of a sweeping electron beam. This electron beam sweeps during each sampling interval over a row of a coding mask, which allows the electrons to pass in certain slots while blocking them in others. Electrons that actually hit the collector anode give rise to an output current, while this current is essentially zero if the electrons are blocked. This system converts an analog signal into a signal representing a string of binary digits. The solution proposed by Gray addresses three main issues with this system. First, the problem of reducing the distortion of the decoded analog signal arising from a small error in the deflection of the electron beam. Second, simplifying the manufacturing of the coding mask by making the smallest apertures of the coding mask larger, and third, improving the timing properties of the recovery circuitry. The way that Gray solves this problem is by simply listing the binary numbers in a different order, so that adjacent numbers differ in a single bit position. This approach solves the first and most important issue, giving a small decoding error (a single bit) for small errors in the beam deflection. In addition, the particular mapping Gray proposes also doubles the size of the smallest apertures of the coding mask. Gray calls the proposed the reflected binary code, due to its recursive construction method (see Section IV-A). Gray identifies the trivial operations defined in Section III below, but his treatment only concerns Gray codes with the symmetric properties imposed by the recursive reflection construction method.

The outline of the paper is as follows. In Sections II to VI we address the problem of selecting an optimal labeling for M-PSK systems; Section II provides an introduction to the M-PSK specific aspects of the problem at hand, Section III gives the necessary nomenclature and definitions, Sections IV and V give the proofs, and Section VI introduces some interesting properties of the optimal labeling. The proofs for M-PAM constellations inherit most of their formulations from the M-PSK proofs, and only minor changes in these proofs are necessary. In Sections VII to IX the differences are elaborated on and the modified proofs are outlined. The proof for the M-QAM case follows almost directly from the M-PAM discussion, so Section X is kept short. Finally, Section XI provides a discussion and conclusion.

II. BINARY LABELINGS FOR M-PSK

The problem of evaluating the average BER of M-PSK modulation schemes has been studied extensively in the literature. In [3–6], approximate and exact values of the BER for certain values of M are given and in [7] the exact values are given for all M. All these references assume that the binary, reflected Gray code (BRGC) is used.

Assuming that all symbols are equally likely for transmission over the channel during each transmission interval, i.e., $\Pr\{\mathbf{s} = \mathbf{s}_k\} = 1/M$ for $k = 0, 1, \ldots, M-1$, the bit-error probability for M-PSK with any labeling of the $M = 2^m$ signal points is given by (1), which we can rewrite as

$$\begin{aligned} P_{\mathrm{b}} &= \frac{1}{mM} \sum_{l=0}^{M-1} \sum_{k=0}^{M-1} d_H\left(\mathbf{c}_l, \mathbf{c}_k\right) \Pr\left(\mathbf{r} \in \Omega_l \mid \mathbf{s} = \mathbf{s}_k\right) \\ &= \frac{1}{m} \sum_{k=0}^{M-1} \left(\frac{1}{M} \sum_{l=0}^{M-1} d_H\left(\mathbf{c}_l, \mathbf{c}_{(l+k) \bmod M}\right) \right) \\ &\qquad \cdot \Pr\left(\mathbf{r} \in \Omega_l \mid \mathbf{s} = \mathbf{s}_k\right) \\ &\triangleq \frac{1}{m} \sum_{k=0}^{M-1} \bar{d}\left(k\right) P(k), \end{aligned} \tag{2}$$

where $P(k)$ is the probability that the received vector is in the decision region of a symbol k steps counter-clockwise away from the transmitted signal. The main focus in this paper is on the function $\bar{d}(k)$, the average distance spectrum (ADS) of the binary labeling. The number $\bar{d}_(k)$ is the average number of bits that differ in symbols separated by k steps, averaged over all M symbols. The probabilities $P(k)$ are not functions of the constellation labeling, so the BER dependence on the labeling is captured entirely by $\bar{d}(k)$. For most channels of interest, the function $P(k)$ decreases rapidly with k (see e.g., [8, p. 201], for an expression for $P(k)$ for the case of an additive white Gaussian noise channel), making it reasonable to chose a labeling that assigns binary patterns to the constellation symbols in such a way that adjacent patterns differ in a single bit. These labelings are known as Gray codes.

In the literature, the binary reflected Gray code is usually referred to simply as "the Gray code", without further specification. However, for $m > 3$, there exist several Gray codes that have different ADS's and as m increases, the number of such labelings rapidly becomes very large [9–11]. To find the labeling that gives the lowest possible BER, it is therefore necessary to consider the entire class of binary labelings having the Gray property. For illustration, in Table I are given two binary labelings having the Gray property, along with their respective ADS. By comparing the ADS's of the two labelings it is seen from (2) that these labelings will indeed result in different BER.

The natural approach to finding the labeling that minimizes (2) is to make sure that the chosen labeling results in a $\bar{d}(k)$ that grows slowly with k. To be more precise, we will address the problem of finding the optimal labeling under the assumption that $P(k)$ decays sufficiently fast with k to make the minimization of the BER equivalent to sequential minimization of the components of the ADS. Under this assumption, considering two labelings a and b with ADS $\bar{a}(k)$ and $\bar{b}(k)$, respectively, the labeling a will result in a lower BER according to (2) if and only if

$$\begin{aligned} \bar{a}(i) &= \bar{b}(j), \qquad 0 \le i < j \\ \bar{a}(j) &< \bar{b}(j) \end{aligned}$$

for some integer $j > 0$. In this paper we will show that the binary reflected Gray code is the unique labeling that results in the slowest increasing ADS among all possible bit-to-symbol mappings (the precise meaning of uniqueness is defined in Section III). For the channels described above this mapping will be optimal in the sense of providing the lowest possible value of the BER.

TABLE I
TWO DIFFERENT BINARY MAPPINGS a AND b, BOTH HAVING THE GRAY PROPERTY AND THEIR RESPECTIVE AVERAGE DISTANCE SPECTRUM $\bar{a}(k)$ AND $\bar{b}(k)$.

a	b	k	$\bar{a}(k)$	$\bar{b}(k)$
0 0 0 0	0 0 0 0	0	0	0
0 0 0 1	0 0 0 1	1	1	1
0 1 0 1	0 0 1 1	2	2	2
0 1 0 0	0 0 1 0	3	2.375	2
0 1 1 0	0 1 1 0	4	2.5	2
0 0 1 0	0 1 1 1	5	2.5	2.5
1 0 1 0	0 1 0 1	6	2.5	3
1 1 1 0	0 1 0 0	7	2.125	2.5
1 1 0 0	1 1 0 0	8	2	2
1 1 0 1	1 1 0 1	9	2.125	2.5
1 1 1 1	1 1 1 1	10	2.5	3
0 1 1 1	1 1 1 0	11	2.5	2.5
0 0 1 1	1 0 1 0	12	2.5	2
1 0 1 1	1 0 1 1	13	2.375	2
1 0 0 1	1 0 0 1	14	2	2
1 0 0 0	1 0 0 0	15	1	1

In a related work [12], the effect of the constellation labeling on the constellation's *edge profile* is evaluated. The aim of the work in [12] is to provide a formal answer to what labelings are sensible for using in trellis-coded modulation systems. The edge profile is related to a *union bound* on the BER of the system using a particular constellation labeling, and since bit error probability is not in the scope of [12] it is consequently not mentioned. The edge profile cannot be used in the search for the optimal Gray code of an M-PSK system.

III. PRELIMINARIES

To simplify the discussion, we start by making some necessary definitions.

Definition 1—Binary Labeling: A binary labeling $\mathcal{C}$ of order $m \in \mathbb{Z}^+$ is a sequence of $M = 2^m$ distinct vectors (codewords or labels), $\mathcal{C} = (\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M-1})$, where each $\mathbf{c}_i \in \{0,1\}^m$.

Definition 2—Binary Cyclic Gray Code: A binary, cyclic Gray code of order m is a binary labeling with $M = 2^m$ codewords, where adjacent codewords, including the first and the last codeword, differ in only one of the m positions.

Throughout this paper, it will be implicit that all labelings mentioned are binary. Also, since it is assumed here that the Gray codes used for M-PSK constellation labeling are both cyclic and binary, we will use the term Gray codes to denote binary cyclic Gray codes.

Definition 3—Average Distance Spectrum: The average distance spectrum (ADS), $\bar{d}(k)$, of a binary labeling $\mathcal{C} = (\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M-1})$ is the average number of bit positions that differ in codewords separated by k steps, averaged cyclically over all the codewords, i.e.,

$$\bar{d}(k) \triangleq \frac{1}{M} \sum_{l=0}^{M-1} d_H\left(\mathbf{c}_l, \mathbf{c}_{(l+k) \bmod M}\right) \tag{3}$$

for all $k \in \mathbb{Z}$.

Remark: By definition, the ADS of a binary cyclic Gray code satisfies $\bar{d}(0) = 0$ and $\bar{d}(1) = 1$.

Remark: As a result of the modulo-operator and the absolute value function in (3), the ADS is an even function $\left(\bar{d}(k) = \bar{d}(-k)\right)$ and periodic with period M.

Definition 4—Superior and Equivalent ADS: The ADS $\bar{d}(k)$ of a binary labeling $\mathcal{C}_1$ is said to be superior to the ADS $\bar{h}(k)$ of a binary labeling $\mathcal{C}_2$ of the same order, if the following relations hold for some integer $j > 0$,

$$\begin{aligned} \bar{d}(i) &= \bar{h}(i), \qquad 0 \le i < j \\ \bar{d}(j) &< \bar{h}(j). \end{aligned}$$

If $\bar{d}(i) = \bar{h}(i)$ for all integers i, $\mathcal{C}_1$ and $\mathcal{C}_2$ are said to have equivalent ADS.

Definition 5—Optimality: The ADS of a binary labeling is said to be optimal if it is superior or equivalent to the ADS of any other binary labeling of the same order. An optimal labeling is a labeling with an optimal ADS.

In this paper the term optimal will always mean optimality in the sense of Definition 5.

Definition 6—Trivial Operations: Trivial operations on a binary labeling are

- cyclic shifts and reflection of the codeword sequence,
- permutation of the codeword coordinates,
- binary inversion of any coordinates.

Remark: Trivial operations on a labeling do not affect the ADS of the labeling.

To increase the readability of the text, we also define uniqueness of labelings in relation to trivial operations.

Definition 7—TO-uniqueness: A binary labeling $\mathcal{C}_m$ with ADS $\bar{d}_m(k)$ is said to be TO-unique if all labelings with the same ADS can be obtained from $\mathcal{C}_m$ by applying trivial operations.

In the the discussion to follow, it is sometimes convenient to relate a binary labeling to a path on a hypercube.

Definition 8—The Hypercube Q_m: The graph whose vertex-set is the set of all binary strings of length m, with an edge between two vertices if and only if they differ in exactly one position, is called the m-dimensional hypercube Q_m.

A binary cyclic Gray code of order m is formed by listing the binary strings corresponding to the vertices of a cycle in Q_m that contains all vertices. Such a path is known as a Hamiltonian cycle [13, pp. 226]. It is known that there exist Hamiltonian cycles of all orders $m \geq 1$, which is also evident from the constructions in Section IV.

IV. Recursive construction of binary labelings

In this section, we provide two different methods of how to recursively construct binary labelings of any order m from binary labelings of order $m-1$. Both these methods show that it is possible to construct binary cyclic Gray codes of any order $m \geq 1$.

As was noted in the introduction, for a given order m, the number of Gray codes that do not have equivalent ADS is usually very large. Only a fraction of these labelings can be generated from the recursive methods proposed below. However, we show in this paper that it is possible to generate an optimal labeling by these recursions.

A. Construction by labeling reflection

To generate a labeling of order m from a labeling of order $m-1$ by means of *reflection* we proceed as follows. To the labeling of order $m-1$, denoted by $\mathcal{C}_{m-1} = (\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M/2-1})$, we append a sequence of $M/2$ vectors formed by repeating the codewords of $\mathcal{C}_{m-1}$ in reverse order; $(\mathbf{c}_0, \ldots, \mathbf{c}_{M/2-1}, \mathbf{c}_{M/2-1}, \ldots, \mathbf{c}_0)$. To this new sequence of binary vectors, an extra coordinate is added to each vector from the left. This extra coordinate is 0 for the first half of the M vectors and 1 for the second half. The so obtained sequence $\mathcal{C}_m$ consists of *distinct* codewords, hence it is a labeling, and $\mathcal{C}_m$ is said to be obtained by reflection of $\mathcal{C}_{m-1}$. Labeling reflection is possible for $m \geq 2$ and is illustrated in Figure 1.

If $\mathcal{C}_{m-1}$ is a Gray code, then so is $\mathcal{C}_m$, which proves that Gray codes of any order exist. The originally proposed Gray code [2], which is still the most commonly encountered Gray code in communications, can be defined as follows.

Definition 9—Binary reflected Gray code: The labeling $\mathcal{G}_m$ obtained by $m-1$ recursive reflections of the trivial labeling $\mathcal{G}_1 = (0, 1)$ is the binary reflected Gray code of order m, for any $m \geq 1$.

B. Construction by labeling expansion

The second method of construction we will consider is termed labeling *expansion*. To generate a labeling $\mathcal{C}_m$ from a labeling $\mathcal{C}_{m-1}$ by expansion we do the following; from $\mathcal{C}_{m-1} = (\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M/2-1})$, repeat each codeword once to obtain a new sequence of M vectors $(\mathbf{c}_0, \mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_1 \ldots, \mathbf{c}_{M/2-1}, \mathbf{c}_{M/2-1})$. Now, an extra coordinate is added to each codeword from the right, taken in turn from the vector $(0, 1, 1, 0, 0, 1, 1, 0, \ldots, 0, 1, 1, 0)$ of length M. Labeling expansion is possible for $m \geq 2$, and the procedure is illustrated in Figure 2.

If $\mathcal{C}_{m-1}$ is a Gray code, then so is $\mathcal{C}_m$. By induction, it is possible to verify that $m-1$ recursive expansions of the trivial labeling $\mathcal{G}_1 = (0, 1)$ leads to a Gray code in which the codewords corresponds to the same path on Q_m as the BRGC.

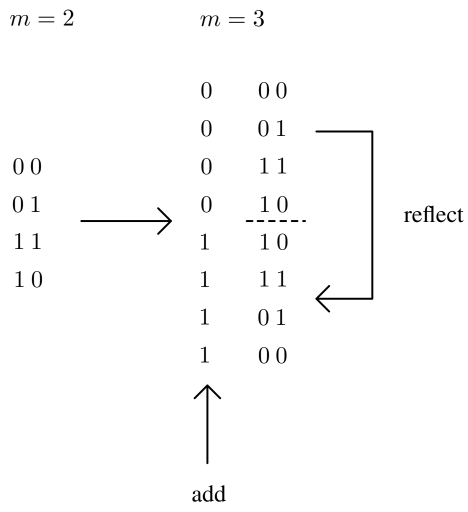

Fig. 1. Construction of a Gray code of order $m = 3$ from a Gray code of order $m = 2$ by means of reflection.

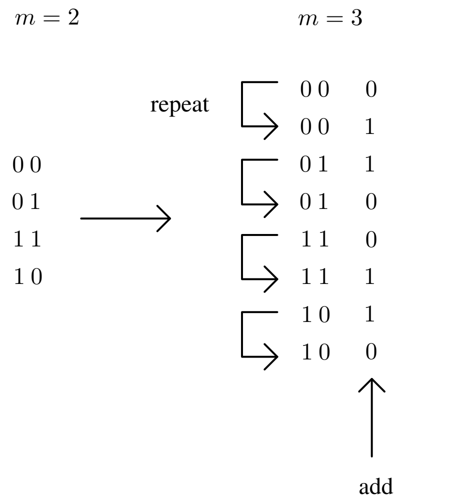

Fig. 2. Construction of a Gray code of order $m = 3$ from a Gray code of order $m = 2$ by means of expansion.

V. Optimality of the binary reflected Gray code

The main result of this paper is captured by the following theorem, which will be proved in Section V-B.

Theorem 1—Optimality of BRGC for M-PSK: The binary reflected Gray code of order m is the optimal, TO-unique, labeling for 2^m-PSK.

A. The first components of the ADS

In order to prove Theorem 1 we will rely on the following theorem, which relates the ADS of a labeling $\mathcal{C}_{m-1}$ of order $m-1$ to the ADS of its expanded labeling $\mathcal{C}_m$. The proof is given in Appendix A.

Theorem 2—Recursion for ADS of an Expanded Labeling: Let $\bar{d}_m(k)$ for $m \geq 2$ denote the ADS of the labeling $\mathcal{C}_m$ obtained from expansion of a labeling $\mathcal{C}_{m-1}$ having an ADS $\bar{d}_{m-1}(k)$. Then, for all integers k, the distance spectra of $\mathcal{C}_m$ and $\mathcal{C}_{m-1}$ satisfy

$$\bar{d}_m(4k) = \bar{d}_{m-1}(2k) \tag{4}$$

$$\bar{d}_m(4k+2) = \bar{d}_{m-1}(2k+1) + 1 \tag{5}$$

$$\bar{d}_m(2k+1) = \frac{1}{2}\bar{d}_{m-1}(k) + \frac{1}{2}\bar{d}_{m-1}(k+1) + \frac{1}{2} \tag{6}$$

with $\bar{d}_1(i) = 0$ for even i and $\bar{d}_1(i) = 1$ for odd i.

The following two lemmas give the first components of the ADS. They will be used in the proof of Lemma 5 and are proved in Appendices B and C.

Lemma 3: For any Gray code $\mathcal{C}_m$ with $m \geq 3$, the ADS satisfies $\bar{d}(1) = 1$, $\bar{d}(2) = 2$, and $\bar{d}(3) \geq 2$.

Lemma 4: Expanding a Gray code of order $m-1 \geq 2$ results in a Gray code of order m having $\bar{d}(3) = 2$. Conversely, all Gray codes $\mathcal{C}_m$ with $m \geq 3$ and $\bar{d}(3) = 2$ can be constructed by expanding a Gray code of order $m-1$, possibly followed by trivial operations.

B. Proof of the optimality of BRGC for M-PSK

We now address the problem of which particular labeling will give the slowest increasing ADS among all possible labelings, or more precisely, which labeling has the *optimal* ADS. According to the discussion in Section II, a mapping with optimal ADS will result in the lowest possible BER for the case when $P(k)$ decays sufficiently fast with k. We will show in the following that the binary reflected Gray code (BRGC) is the TO-unique labeling with optimal ADS.

Lemma 5: If $\mathcal{C}_{m-1}$ is an optimal labeling of order $m-1$, with $m \geq 2$, then an optimal, TO-unique, labeling $\mathcal{C}_m$ of order m is obtained by expanding $\mathcal{C}_{m-1}$.

Proof: The lemma is trivial for $m = 2$, since only one TO-unique Gray code of order 2 exists. From Lemmas 3 and 4, any optimal labeling $\mathcal{C}_m$ for $m \geq 3$ can be constructed by expanding a labeling $\mathcal{C}_{m-1}$ and applying trivial operations. Hence, the ADS of $\mathcal{C}_m$ satisfies (4)–(6). Since, for all integers i, $\bar{d}_m(2i-1)$ and $\bar{d}_m(2i)$ are increasing functions of $\bar{d}_{m-1}(i)$, and independent of $\bar{d}_{m-1}(j)$ for $j > i$, sequential minimization of $\bar{d}_m(1), \bar{d}_m(2), \ldots$ is equivalent to sequential minimization of $\bar{d}_{m-1}(1), \bar{d}_{m-1}(2), \ldots$ Since $\mathcal{C}_{m-1}$ is optimal by assumption, this proves that $\mathcal{C}_m$ is also an optimal labeling. □

The proof of the main theorem now follows straightforwardly from Lemma 5.

Proof of Theorem 1: The BRGC of order m can be obtained by $m-1$ recursive expansions of the trivial labeling $(0, 1)$. The proof of optimality for the BRGC is trivial for $m = 1$. By induction and Lemma 5, optimality of the BRGC is guaranteed for $m \geq 2$. □

VI. PROPERTIES OF THE OPTIMAL ADS FOR M-PSK

A closed-form expression for the ADS of the BRGC of order $m \geq 1$ is given in [7] as

$$\bar{d}_m(k) = 2\left|\frac{k}{M} - \left\lfloor\frac{k}{M}\right\rceil\right| + 2\sum_{i=2}^{m}\left|\frac{k}{2^i} - \left\lfloor\frac{k}{2^i}\right\rceil\right| \tag{7}$$

for all $k \in \mathbb{Z}$, where $\lfloor x \rceil$ denotes the closest integer to x (ties can be broken arbitrarily). In this section some properties of this optimal ADS is given that provide further insight into the labeling of signal constellations and how to obtain bounds on the average BER of systems using M-PSK modulation.

A. Constant sequences of optimal ADS

By considering a labeling having $M = 2^m$ signal points it is possible to obtain a recursive relation in terms of the optimal ADS of a labeling having 2^{m-1} points. For the optimal labeling of size 2^{m-1} we have from (7), for all $k \in \mathbb{Z}$ and $m \geq 2$,

$$\bar{d}_{m-1}(k) = 2\left|\frac{2k}{M} - \left\lfloor\frac{2k}{M}\right\rceil\right| + 2\sum_{i=2}^{m-1}\left|\frac{k}{2^i} - \left\lfloor\frac{k}{2^i}\right\rceil\right|. \tag{8}$$

By comparing (7) and (8), we see that

$$\bar{d}_m(k) = \bar{d}_{m-1}(k) + 4\left|\frac{k}{M} - \left\lfloor\frac{k}{M}\right\rceil\right| - 2\left|\frac{2k}{M} - \left\lfloor\frac{2k}{M}\right\rceil\right|.$$

Since $|x - \lfloor x \rceil|$ represents a triangular waveform of period 1 and amplitude $1/2$, the difference

$$4\left|\frac{k}{M} - \left\lfloor\frac{k}{M}\right\rceil\right| - 2\left|\frac{2k}{M} - \left\lfloor\frac{2k}{M}\right\rceil\right|$$

is a piecewise linear function. Using the even and periodic properties of the ADS and dividing the range $k = 0, \ldots, M$ uniformly into four segments yields for $m \geq 2$

$$\bar{d}_m(k) = \bar{d}_{m-1}(k), \qquad k \in \{0, \ldots, M/4\} \tag{9}$$

$$\bar{d}_m(k) = \bar{d}_{m-1}(k) + \frac{8k}{M} - 2, \qquad k \in \{M/4, \ldots, M/2\} \tag{10}$$

$$\bar{d}_m(k) = \bar{d}_{m-1}(k) + 6 - \frac{8k}{M}, \qquad k \in \{M/2, \ldots, 3M/4\} \tag{11}$$

$$\bar{d}_m(k) = \bar{d}_{m-1}(k), \qquad k \in \{3M/4, \ldots, M\}. \tag{12}$$

The recursion (9)–(12) should be compared with the recursion (4)–(6), as they both define $\bar{d}_m(k)$ in terms of $\bar{d}_{m-1}(k)$, but from different perspectives. In Figures 3 and 4, four ADS's are shown. Figure 3 is an illustration of the increasing resolution and jaggedness of the ADS as predicted by (4)–(6) and in Figure 4, the ADS of a BRGC of order $m = 12$ illustrates the self-repeating structure and the constant sequences of the ADS as predicted by (9)–(12).

B. Upper bound on optimal ADS

The average value of $\bar{d}_m(k)$, for any $m \geq 1$, taken over all $k = 0, \ldots, M-1$, is $\alpha = m/2$. To see this, return to (3) and average over k to obtain

$$\alpha \triangleq \frac{1}{M}\sum_{k=0}^{M-1}\bar{d}(k) = \frac{1}{M^2}\sum_{k=0}^{M-1}\sum_{l=0}^{M-1} d_H\left(\mathbf{c}_l, \mathbf{c}_{(l+k) \bmod M}\right).$$

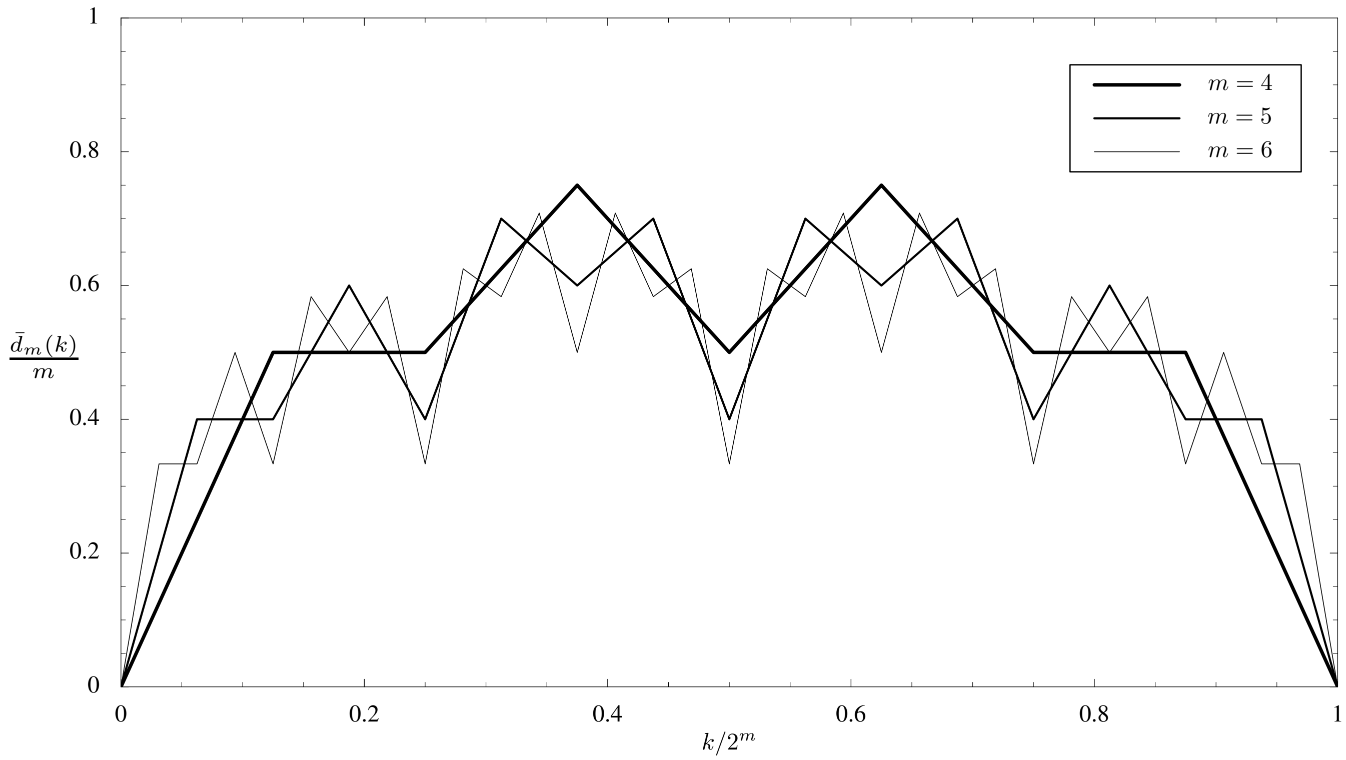

Fig. 3. The normalized ADS of the BRGC of orders 4, 5, and 6. The thickest curve, $m = 4$, represents $\bar{b}$ in Table I. The curves follow the same general pattern, with an increasing resolution and jaggedness, as anticipated by the recursion (4)–(6). The area under all the curves is the same, 1/2.

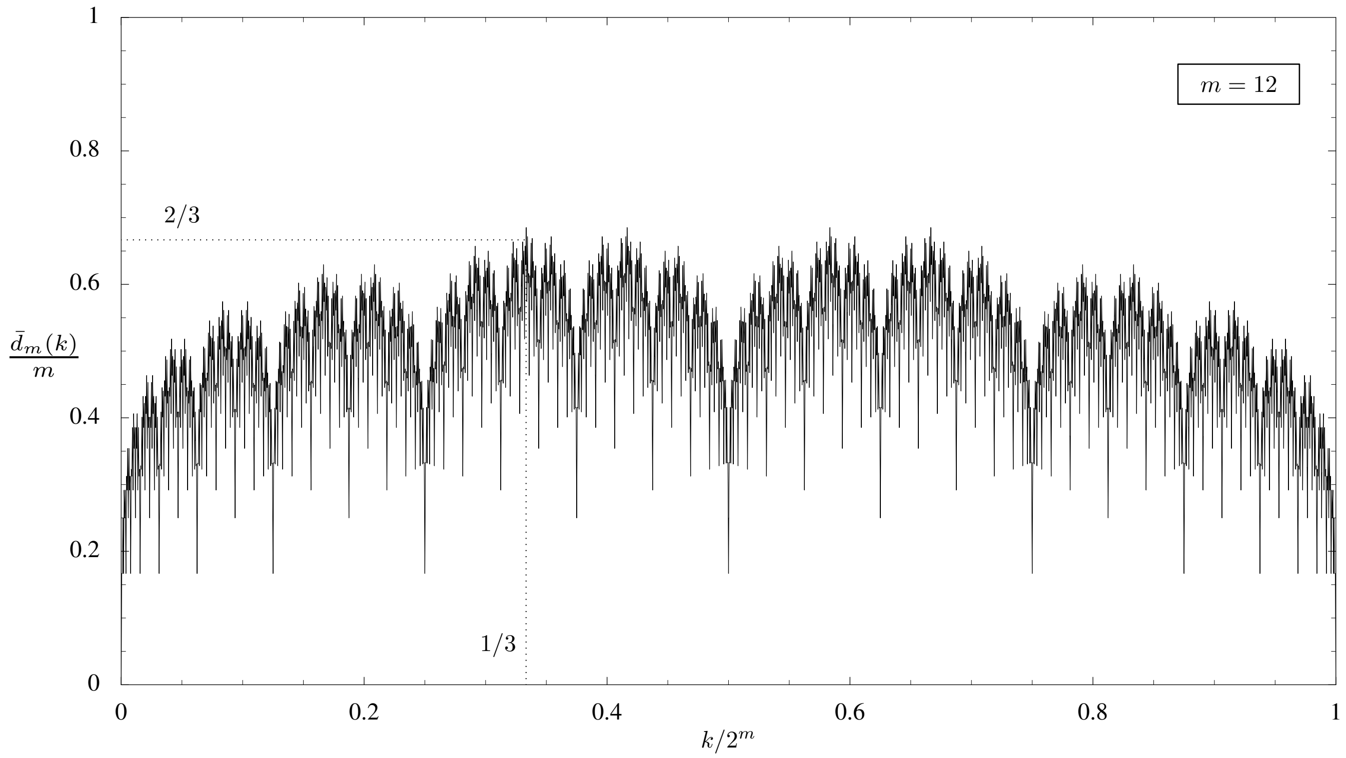

Fig. 4. This fractal curve represents the limit of the curve family of Figure 3. It was constructed as $\bar{d}_{12}(k)/12$, but the ADS's of higher orders look essentially the same to the eye. Observe the self-similarity: The first quarter of the curve is a rescaled copy of the first half, as anticipated by the recursion (9)–(12).

Interchange the order of summation and observe that the sum of all labels (and over all components) on the vertices of a hypercube Q_m is $Mm/2$ to obtain the result $\alpha = \frac{m}{2}$. This result indicates that the (normalized) area under all the curves in Figures 3 and 4 is $1/2$.

Since the average of $\bar{d}_m(k)$ over all k is $m/2$ and the optimization assigns small values to components with low k, one might ask if some components of the optimal $\bar{d}_m(k)$ with higher k have values close to m. The answer is no, as is shown in this subsection.

Theorem 6—Maximum value of optimal $\bar{d}_m(k)$: The maximum value of the optimal ADS

$$\hat{d}_m \triangleq \max_k \bar{d}_m(k)$$

is given recursively by $\hat{d}_2 = \hat{d}_3 = 2$ and

$$\hat{d}_m = \frac{1}{2}\hat{d}_{m-1} + \frac{1}{2}\hat{d}_{m-2} + 1, \qquad m \geq 4. \tag{13}$$

The proof is given in Appendix D. The following corollaries can be obtained from this theorem and its proof.

Corollary 7: The recursion (13) with initial conditions $\hat{d}_3 = \hat{d}_2 = 2$ is solved by

$$\hat{d}_m = \frac{1}{9}\left(6m + 2 + (-2)^{4-m}\right), \qquad m \geq 2 \tag{14}$$

and as $m \to \infty$, $\hat{d}_m/m \to 2/3$.

Proof: Direct substitution of (14) into (13) gives the result. □

Corollary 8: The maximum value of an optimal $\bar{d}_m(k)$ occurs for $k = j_m$, where

$$j_m = \frac{1}{3}\left(2^m + 2(-1)^m\right), \qquad m \geq 2 \tag{15}$$

and as $m \to \infty$, $j_m/2^m \to 1/3$.

Proof: Direct substitution of (15) into (34) of the proof of Theorem 6 yields the result. □

The asymptotic position ($M/3$) and magnitude ($2m/3$) of the maximum value of the ADS are indicated in Figure 4. The position of the maximum is, however, not unique (not even for $k \in \{0, \ldots, M-1\}$).

VII. BINARY LABELINGS FOR M-PAM

In this section we move on to find which labeling is optimal for an M-PAM system. This problem differs from the M-PSK case, since the M-PAM problem is not cyclic. This, in turn, means that the first and the last binary strings in the labeling not necessarily need to differ in a single bit. However, we show in the coming sections that this extra degree of freedom is not enough to allow for the construction of a labeling that is superior to the BRGC.

In an M-PAM communication system, the M signal points are distributed along a line, separated by a distance ρ. By defining

$$d^-(k) \triangleq \frac{1}{M}\sum_{l=k}^{M-1} d_H(\mathbf{c}_{l-k}, \mathbf{c}_l), \quad k = 1, \ldots, M-1 \tag{16}$$

$$d^-(k) \triangleq 0, \qquad \text{otherwise} \tag{17}$$

and

$$d^+(k) \triangleq \frac{1}{2M}\sum_{l=1}^{\min(k,M)-1} [d_H(\mathbf{c}_0, \mathbf{c}_l) + d_H(\mathbf{c}_{M-1}, \mathbf{c}_{M-1-l})], \qquad k \geq 2 \tag{18}$$

$$d^+(k) \triangleq 0, \qquad k \leq 1 \tag{19}$$

and assuming that the receiver is an ML receiver, it is shown in Appendix E that the exact average BER of M-PAM can be expressed as

$$P_b = \frac{2}{m}\sum_{k=1}^{M-1} P(k)d^-(k) + \frac{2}{m}\sum_{k=1}^{\infty} P(k)d^+(k). \tag{20}$$

Clearly, we can rewrite (20) as

$$P_b = \frac{2}{m}\sum_{k=1}^{\infty} P(k)\bar{d}(k)$$

where $\bar{d}(k) = d^-(k) + d^+(k)$. We have now obtained a formulation of the BER of a general M-PAM communication system which is similar to the formulation used in Section II for an M-PSK system. As we show in the next section, a similar argument as in the M-PSK case will prove that the BRGC is optimal for the M-PAM case as well.

VIII. MODIFIED DEFINITIONS FOR M-PAM

In this section, we repeat or modify slightly the definitions given in Section III for the M-PSK system to suit the M-PAM discussion. Definition 1 is used as it stands while Definition 2 is changed slightly into

Definition 2b—Binary Gray Code: A binary Gray code of order m is a binary labeling with $M = 2^m$ codewords, where adjacent codewords, not necessarily including the first and the last codeword, differ in only one of the m positions.

The definition for the average distance spectrum must change in several respects according to the discussion in Section VII.

Definition 3b—Average Distance Spectrum: The average distance spectrum (ADS), $\bar{d}(k)$, appropriate for M-PAM, is defined as

$$\bar{d}(k) \triangleq d^-(k) + d^+(k) \tag{21}$$

where $d^-(k)$ and $d^+(k)$ are defined in (16)–(17) and (18)–(19), respectively.

Note that the remarks made after Definition 3 do not hold true for the ADS defined in Definition 3b. Definitions 4 and 5 remain unchanged, but Definition 6 must change to remove the reference to a cyclic labeling.

Definition 6b—Trivial Operations: Trivial operations on a binary labeling used for M-PAM are

- reflection of the codeword sequence,
- permutation of the codeword coordinates,
- binary inversion of any coordinates.

Remark: Trivial operations on a labeling do not affect the ADS of the labeling. Since there are fewer types of trivial operations for M-PAM than for M-PSK, the number of nonequivalent labelings is larger. There are even two nonequivalent Gray codes for 8-PAM, which is not the case for 8-PSK.

Definition 7 and 8 stay unchanged for the discussion of M-PAM labelings.

Remark: Referring to Definitions 8 and 2b, a binary Gray code of order m is formed by any path, not necessarily a closed loop, that visits all vertices of the hypercube Q_m once.

We will now use these modified definitions to show that the TO-unique labeling that results in the optimal distance spectrum for M-PAM systems is the BRGC (which is still defined by Definition 9).

IX. OPTIMALITY OF THE BRGC FOR M-PAM

We now address the problem of finding the optimal labeling for M-PAM signal constellations. This subject will be treated more briefly than the analogous problem for M-PSK, since the general line of proof is the same as in Section V. We will focus on the details in which the M-PAM case differs from M-PSK. The main result of the M-PAM case is the following theorem.

Theorem 1b—Optimality of BRGC for M-PAM: The binary reflected Gray code of order m is the optimal, TO-unique, labeling for 2^m-PAM.

To prove Theorem 1b, we need as before a set of recursions for the ADS and tight lower bounds on the first components of the ADS. Then the result follows by induction. Not surprisingly, the recursions in the next theorem follow the same pattern as (4) – (6), with the addition of a correction term to account for the edge effects of M-PAM. The important point for the optimization is that these correction terms are independent on the labeling.

Theorem 2b—Recursion for ADS of an Expanded Labeling: Let $\bar{d}_m(k)$ for $m \geq 2$ denote the ADS of a labeling $\mathcal{C}_m$ obtained from expansion of a labeling $\mathcal{C}_{m-1}$ having an ADS $\bar{d}_{m-1}(k)$. Then, for all integers $0 \leq k \leq 2^{m-2}-1$, the ADS satisfies

$$\bar{d}_m(4k) = \bar{d}_{m-1}(2k) + \frac{k}{2^{m-1}} \tag{22}$$

$$\bar{d}_m(4k+1) = \frac{1}{2}\bar{d}_{m-1}(2k) + \frac{1}{2}\bar{d}_{m-1}(2k+1) + \frac{1}{2} \tag{23}$$

$$\bar{d}_m(4k+2) = \bar{d}_{m-1}(2k+1) + 1 - \frac{2k+1}{2^m} \tag{24}$$

$$\bar{d}_m(4k+3) = \frac{1}{2}\bar{d}_{m-1}(2k+1) + \frac{1}{2}\bar{d}_{m-1}(2k+2) + \frac{1}{2}. \tag{25}$$

Outline of proof: Recursions are derived for $d^-(k)$ and $d^+(k)$ separately. First, we proceed as in the proof of Theorem 2 in Appendix A, but consider $d_m^-(k)$ for $k = 0, \ldots, M-1$ instead of $\bar{d}_m(k)$ for all integers k. The analysis in the appendix holds in this case too, except that for the sequence $(0, 1, 1, 0, 0, 1, \ldots, 1, 0)$ in the last coordinate we do not obtain (30)–(32) but an expression with four cases and k-dependent correction terms subtracted from the right-hand sides. These correction terms, which follow straightforwardly from the definition (16), propagate directly to the recursions for $d_m^-(k)$, which are

$$\begin{aligned}
d_m^-(4k) &= d_{m-1}^-(2k)\\
d_m^-(4k+1) &= \frac{1}{2}d_{m-1}^-(2k) + \frac{1}{2}d_{m-1}^-(2k+1) + \frac{1}{2} - \frac{k}{2^{m-1}}\\
d_m^-(4k+2) &= d_{m-1}^-(2k+1) + 1 - \frac{2k+1}{2^{m-1}}\\
d_m^-(4k+3) &= \frac{1}{2}d_{m-1}^-(2k+1) + \frac{1}{2}d_{m-1}^-(2k+2)\\
&\qquad + \frac{1}{2} - \frac{k+1}{2^{m-1}}
\end{aligned}$$

for $m \geq 2$ and all integers $0 \leq k \leq 2^{m-2}-1$.

Second, recursions for $d^+(k)$, defined in (18)–(19), are derived in a similar manner. The result is, again for $m \geq 2$ and all integers $0 \leq k \leq 2^{m-2}-1$,

$$\begin{aligned}
d_m^+(4k) &= d_{m-1}^+(2k) + \frac{k}{2^{m-1}}\\
d_m^+(4k+1) &= \frac{1}{2}d_{m-1}^+(2k) + \frac{1}{2}d_{m-1}^+(2k+1) + \frac{k}{2^{m-1}}\\
d_m^+(4k+2) &= d_{m-1}^+(2k+1) + \frac{2k+1}{2^m}\\
d_m^+(4k+3) &= \frac{1}{2}d_{m-1}^+(2k+1) + \frac{1}{2}d_{m-1}^+(2k+2) + \frac{k+1}{2^{m-1}}
\end{aligned}$$

which completes the proof. □

From these recursions, we can derive the modified versions of Lemmas 3 and 4 given below.

Lemma 3b: For any Gray code $\mathcal{C}_m$ with $m \geq 3$, the ADS satisfies $\bar{d}(1) = 1 - 1/2^m$, $\bar{d}(2) = 2 - 3/2^m$, and $\bar{d}(3) \geq 2 - 1/2^{m-2}$.

Outline of proof: We define $d(l,k) \triangleq d_H(\mathbf{c}_l, \mathbf{c}_{l+k})$ for $l = 0, 1, \ldots, M-1-k$. Since $d(l,1) = 1$ and $d(l,2) = 2$ for all Gray codes and all l, the values of $d_m^-(1)$ and $d_m^-(2)$ are given directly by (16) and $d_m^+(k)$ is given by (18) for $k = 1, 2, 3$. We proceed to show that the sequence $d(l,3)$ for $l = 0, 1 \ldots, M-4$ contains at most $M/2-1$ ones, with the same method as in the proof of Lemma 3, which establishes $d_m^-(3) \geq 2 - 7/M$. The lower bound on $\bar{d}(3)$ then follows from (21). □

Lemma 4b: Expanding a Gray code of order $m-1 \geq 2$ results in a Gray code of order m having $\bar{d}(3) = 2 - 1/2^{m-2}$. Conversely, all Gray codes $\mathcal{C}_m$ with $m \geq 3$ and $\bar{d}(3) = 2 - 1/2^{m-2}$ can be constructed by expanding a Gray code of lower order $\mathcal{C}_{m-1}$, possibly followed by trivial operations.

Outline of proof: The first statement of the lemma follows immediately from (22)–(25). For the second statement, we study the sequence $d(l,3)$ for $l = 0, 1, \ldots, M-4$ and observe from the proof of Lemma 3b that the *only* sequence that meets the lower bound on $d_m^-(3)$ is $(1, 3, 1, 3, \ldots, 1)$ (the sequence $(3, 1, \ldots, 3)$ gives a higher value of $d_m^-(3)$ for PAM) and proceed exactly as in the proof of Lemma 4. □

Having established the necessary preliminaries, we are now ready to conclude the induction with our last lemma.

Lemma 5b: If $\mathcal{C}_{m-1}$ is an optimal labeling of order $m-1 \geq 1$, then an optimal, TO-unique, labeling $\mathcal{C}_m$ of order m is obtained by expanding $\mathcal{C}_{m-1}$.

The proof is identical to the proof of Lemma 5 in Section V-B.

Applying Lemma 5b $m-1$ times to the labeling $(0,1)$, which obviously is an optimal, TO-unique, labeling of order 1, proves that the BRGC is an optimal, TO-unique, labeling for 2^m-PAM of any order $m \geq 1$, which completes the proof of Theorem 1b.

X. Binary labelings for M-QAM

Following the discussion concerning the M-PAM constellations we now turn our attention to the closely related M-QAM constellations. We will restrict the discussion to rectangular $(M_1 \times M_2)$-QAM constellations, which are obtained from the direct product of two PAM constellations, one with $M_1 = 2^{m_1}$ points and one with $M_2 = 2^{m_2}$ points, with the same signal point separation ρ. Furthermore, we make the assumption that the components of the two-dimensional noise vector are statistically independent.

Let $P(k,l)$ denote the probability that the noise carries the transmitted signal point to a decision region that is k steps away along one axis and l steps away along the other. Again, for channels of interest, the most likely error events are associated with $P(\pm 1, 0) = P(0, \pm 1)$, and we want these error events to result in the smallest number of bit errors, i.e., a single bit error. Therefore, we want to use a two-dimensional Gray code, which is a labeling that has labels that differ in a single bit for any two signal points that are separated by a distance ρ.

The only way to assign labels to a rectangular QAM constellation that results in a Gray code is given by the following lemma, which is stated and proved in [12].

Lemma 9—Gray labeling of rectangular QAM constellations: The only ways to assign a labeling with the Gray property to a $(2^{m_1} \times 2^{m_2})$-point rectangular constellation is the direct product of a 2^{m_1}-point and a 2^{m_2}-point Gray code.

Since we made the assumption of statistically independent noise in each dimension, we may use a result from [14]: an M-QAM system obtained from the direct product of an M_1-PAM constellation and an M_2-PAM constellation has BER

$$P_{\mathrm{b}} = \frac{m_1}{m_1 + m_2} P_{\mathrm{b}}(M_1) + \frac{m_2}{m_1 + m_2} P_{\mathrm{b}}(M_2),$$

where $P_{\mathrm{b}}(M_i)$ is the BER of the M_i-PAM system. This shows that the optimal labeling for the M-QAM constellation is found by selecting the optimal labeling for each of the two PAM constellations independently. Therefore, from Theorem 1b, the optimal labeling for an $(M_1 \times M_2)$-QAM constellation is found from the labeling that is the direct product of the BRGC of order m_1 and the BRGC of order m_2.

In passing, we note that the so-called cross-QAM constellations cannot be labelled with a Gray code [12], therefore the selection of an optimal labeling for these constellations falls outside the scope of this paper.

XI. Discussion and conclusion

We have addressed the problem of finding which constellation labeling will produce the lowest possible BER among all possible labelings for M-PSK, M-PAM, and M-QAM. The search is done under the assumption that the communication takes place over channels for which $P(k)$ decays quickly enough to ensure that sequential minimization of the components of the ADS yields the minimum BER. We have shown that the best labeling under this assumption is the binary reflected Gray code.

The relevance of this discussion and the proof can be verified by consulting almost any widely spread textbook on communications in which the problem of calculating the average BER of systems using these constellation is treated. In most cases, the BRGC is used, but referred to simply as "the Gray code" and the fact that a wealth of different Gray codes exist and their impact on the BER is often neglected. The proofs in this paper validates the use of the BRGC for constellation labeling and allows for a clearer presentation of the topic of BER calculation for this type of communication system.

Appendix A
Proof of Theorem 2

The average distance spectrum (ADS) of any binary periodic sequence b_l with period P is defined, for all integers k, as

$$\bar{\delta}(b,k) \triangleq \frac{1}{P} \sum_{l=0}^{P-1} |b_l - b_{l+k}|.$$

Now, from b_l we form another sequence $u_l = (b_{-1}, b_{-1}, b_0, b_0, b_1, b_1, \ldots)$, u_l being simply an upsampled version of b_l, where each element of b_l is repeated once. The sequence u_l is a binary, periodic sequence with period $P' = 2P$, satisfying $u_{2l} = u_{2l+1} = b_l$, for all integers l. For this new sequence we have

$$\bar{\delta}(u,k) = \frac{1}{P'} \sum_{l=0}^{P'-1} |u_l - u_{l+k}| = \frac{1}{2P} \sum_{l=0}^{2P-1} |u_l - u_{l+k}|.$$

By rearranging terms in the second sum we obtain

$$\bar{\delta}(u,k) = \frac{1}{2P} \left(\sum_{l=0}^{P-1} |u_{2l} - u_{2l+k}| + \sum_{l=0}^{P-1} |u_{2l+1} - u_{2l+1+k}| \right).$$

For $k = 2i$, where i is an integer, we have

$$\begin{aligned}
\bar{\delta}(u,2i) &= \frac{1}{2P} \left(\sum_{l=0}^{P-1} |u_{2l} - u_{2l+2i}| + \sum_{l=0}^{P-1} |u_{2l+1} - u_{2l+1+2i}| \right) \\
&= \frac{1}{2P} \left(\sum_{l=0}^{P-1} |b_l - b_{l+i}| + \sum_{l=0}^{P-1} |b_l - b_{l+i}| \right) \\
&= \frac{1}{2P} \left(P\bar{\delta}(b,i) + P\bar{\delta}(b,i) \right) = \bar{\delta}(b,i) \qquad (26)
\end{aligned}$$

and, similarly, for $k = 2i+1$, we have

$$\begin{aligned}\bar{\delta}(u, 2i+1) &= \frac{1}{2P}\left(\sum_{l=0}^{P-1} |u_{2l} - u_{2l+2i+1}| \right.\\ &\qquad \left. + \sum_{l=0}^{P-1} |u_{2l+1} - u_{2l+2i+2}|\right)\\ &= \frac{1}{2P}\left(\sum_{l=0}^{P-1} |b_l - b_{l+i}| + \sum_{l=0}^{P-1} |b_l - b_{l+i+1}|\right)\\ &= \frac{1}{2P}\left(P\bar{\delta}(b,i) + P\bar{\delta}(b,i+1)\right)\\ &= \frac{1}{2}\bar{\delta}(b,i) + \frac{1}{2}\bar{\delta}(b,i+1). \end{aligned} \tag{27}$$

Consider now the ADS $\bar{d}_m(k)$ of a labeling $\mathcal{C}_m$ obtained by expanding a labeling $\mathcal{C}_{m-1}$ with ADS $\bar{d}_{m-1}(k)$. Let $[\mathbf{c}_l]_i$ be the ith bit of the codeword $\mathbf{c}_l$. By denoting the contribution to the ADS from coordinate i of all codewords with

$$\bar{d}_m^{(i)}(k) \triangleq \frac{1}{M}\sum_{l=0}^{M-1} \left|[\mathbf{c}_l]_i - [\mathbf{c}_{(l+k) \bmod M}]_i\right|, \qquad \forall k \in \mathbb{Z}$$

we have from (3) for the ADS of $\mathcal{C}_m$

$$\begin{aligned}\bar{d}_m(k) &= \sum_{i=0}^{m-1} \bar{d}_m^{(i)}(k)\\ &= \sum_{i=0}^{m-2} \bar{d}_m^{(i)}(k) + \bar{d}_m^{(m-1)}(k)\\ &\triangleq \bar{\nu}_m(k) + \bar{d}_m^{(m-1)}(k)\end{aligned}$$

where $i = m-1$ corresponds to the last coordinate in the codewords (cf. Table I). Now, we observe that the term $\bar{\nu}_m(k)$ is simply the ADS of the list of binary strings that results from simply repeating each codeword of $\mathcal{C}_{m-1}$ once. By noting that the modulo operator in (3) can be removed without affecting the result, by instead considering the periodic repetition of the codewords, we can use (26) and (27) above to obtain, for all integers k,

$$\bar{\nu}_m(2k) = \bar{d}_{m-1}(k) \tag{28}$$

$$\bar{\nu}_m(2k+1) = \frac{1}{2}\bar{d}_{m-1}(k) + \frac{1}{2}\bar{d}_{m-1}(k+1). \tag{29}$$

To obtain the desired result, we note that the term $\bar{d}_m^{(m-1)}(k)$ is the ADS of the (periodically repeated) sequence $(0, 1, 1, 0)$. For this sequence we have trivially, for all integers k,

$$\bar{d}_m^{(m-1)}(4k) = 0 \tag{30}$$

$$\bar{d}_m^{(m-1)}(4k+2) = 1 \tag{31}$$

$$\bar{d}_m^{(m-1)}(2k+1) = 1/2. \tag{32}$$

Combining (28)–(32) we have

$$\begin{aligned}\bar{d}_m(4k) &= \bar{d}_{m-1}(2k)\\ \bar{d}_m(4k+2) &= \bar{d}_{m-1}(2k+1) + 1\\ \bar{d}_m(2k+1) &= \frac{1}{2}\bar{d}_{m-1}(k) + \frac{1}{2}\bar{d}_{m-1}(k+1) + \frac{1}{2}.\end{aligned}$$

The optimal, TO-unique, labeling for $m = 1$ is $\mathcal{C}_1 = (0, 1)$. Clearly, $\bar{d}_1(0) = 0$ and $\bar{d}_1(1) = 1$, and since $\bar{d}_1(k)$ is periodic with period 2, $\bar{d}_1(2i) = \bar{d}_1(0) = 0$ and $\bar{d}_1(2i+1) = \bar{d}_1(1) = 1$, which completes the proof of Theorem 2.

APPENDIX B
PROOF OF LEMMA 3

Since, for a Gray code $\mathcal{C}_m$, all adjacent codewords differ in a single position, we have $\bar{d}_m(1) = 1$. Codewords separated by two steps can either differ in 0 or 2 positions, and since all codewords are distinct, $\bar{d}_m(2) = 2$ for $m \geq 2$. Let $d(l,k) \triangleq d_H(\mathbf{c}_l, \mathbf{c}_{(l+k) \bmod M})$. To show $\bar{d}_m(3) \geq 2$ for $m \geq 3$, we first show that no two consecutive terms $d(l,3)$ and $d(l+1,3)$ in (3) can both be 1 for $m \geq 3$. To see this, consider any sequence of five consecutive codewords, which, without loss of generality, can be taken to be $(\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4)$. Since $\mathbf{c}_1$ and $\mathbf{c}_3$ differ in two positions, there are exactly two points in Q_m that are adjacent to both $\mathbf{c}_1$ and $\mathbf{c}_3$. One of these points is $\mathbf{c}_2$; the other may be $\mathbf{c}_0$ or $\mathbf{c}_4$, but not both. Assume that $d(0,3) = 1$. Since $\mathbf{c}_1$ and $\mathbf{c}_4$ are separated by an odd number of steps this implies that $d(1,4) \geq 3$. Generalizing this argument shows that the sequence $d(l,3)$ for $l = 0, 1, \ldots, M-1$ contains at most $M/2$ ones and at least $M/2$ values that are 3 or higher. Hence their average $\bar{d}(3) \geq 2$.

APPENDIX C
PROOF OF LEMMA 4

The first statement of the lemma follows immediately from (6).

For the second statement of the lemma, we know from the proof of Lemma 3 that, for any Gray code of order $m \geq 3$, the sequence $d_H(\mathbf{c}_l, \mathbf{c}_{l+3})$ for $l = 0, 1, \ldots$, consists of odd positive integers such that no two consecutive values are both 1. Hence, the only sequence that results in $\bar{d}(3) = 2$ is $(1, 3, 1, 3, \ldots)$ (or $(3, 1, 3, 1, \ldots)$, which will not be further considered, since it simply corresponds to a cyclic shift of the codewords). This means that the codeword pairs $\{\mathbf{c}_0, \mathbf{c}_3\}, \{\mathbf{c}_2, \mathbf{c}_5\}, \{\mathbf{c}_4, \mathbf{c}_7\} \ldots$ are adjacent vertices of Q_m, while the codeword pairs $\{\mathbf{c}_1, \mathbf{c}_4\}, \{\mathbf{c}_3, \mathbf{c}_6\}, \{\mathbf{c}_5, \mathbf{c}_8\}, \ldots$, differ in three coordinates. Since $d_H(\mathbf{c}_l, \mathbf{c}_{l+3}) = 1$ for any even $0 \leq l \leq M-4$, $(\mathbf{c}_l, \mathbf{c}_{l+1}, \mathbf{c}_{l+2}, \mathbf{c}_{l+3})$ forms a square in Q_m. Hence, $\mathbf{c}_{l+1} - \mathbf{c}_l$ and $\mathbf{c}_{l+2} - \mathbf{c}_{l+3}$ are equal, or

$$\boldsymbol{\Delta} \triangleq \mathbf{c}_1 - \mathbf{c}_0 = \mathbf{c}_2 - \mathbf{c}_3 \ldots = \mathbf{c}_{M-2} - \mathbf{c}_{M-1}.$$

Refer to Figure 5 for an example of the relation between the codewords. The difference vector $\boldsymbol{\Delta}$ has only one nonzero position, say, position j. By performing trivial operations on $\mathcal{C}_m$ we can obtain a code $\mathcal{C}'_m$ for which

$$\begin{aligned}\boldsymbol{\Delta}' \triangleq \mathbf{c}'_1 - \mathbf{c}'_0 = \mathbf{c}'_2 - \mathbf{c}'_3 = \ldots = \mathbf{c}'_{M-2} - \mathbf{c}'_{M-1}\\ = (0, \ldots, 0, 1)\end{aligned}$$

without affecting its ADS. We now partition the codewords of $\mathcal{C}'_m$ according to the value of rightmost bit. This creates two subsets

$$(\mathbf{c}'_0, \mathbf{c}'_3, \mathbf{c}'_4, \mathbf{c}'_7, \mathbf{c}'_8, \ldots, \mathbf{c}'_{M-1})$$

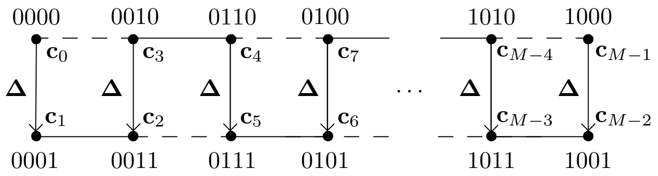

Fig. 5. Example of a path on Q_m that forms a cyclic Gray code. The labels correspond to the case $M = 16$. The solid lines form a path on Q_m and represent an expanded labeling, and the dashed lines identify some other pairs of vertices on Q_m that are neighbors.

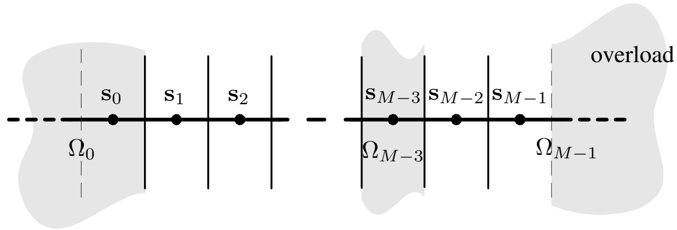

Fig. 6. Decision regions and notation used for an M-PAM signal constellation. Note how the overload region is a subset of the decision region for either of the end signal points $\mathbf{s}_0$ or $\mathbf{s}_{M-1}$ after removing a symmetrical strip surrounding the signal point.

and

$$(\mathbf{c}'_1, \mathbf{c}'_2, \mathbf{c}'_5, \mathbf{c}'_6, \mathbf{c}'_9, \ldots, \mathbf{c}'_{M-2})$$

which represent two cycles on Q_m. By picking any of the two subsets and puncturing the rightmost bit in this subset (which geometrically corresponds to a projection orthogonal to $\boldsymbol{\Delta}'$, so that the two subsets become identical after puncturing) we generate a cyclic Gray code of order $m-1$. It is easily verified that expanding this labeling using the procedure given in Section IV-B yields $\mathcal{C}'_m$ again.

APPENDIX D
PROOF OF THEOREM 6

From Theorem 2 we have (take the average of (4) and (5) to see that it equals (6))

$$\bar{d}_m(2k+1) = \frac{1}{2}\bar{d}_m(2k) + \frac{1}{2}\bar{d}_m(2k+2). \tag{33}$$

Define $j_2 \triangleq 2$ and

$$j_m \triangleq 2\left(j_{m-1} + (-1)^m\right), \qquad m \geq 3. \tag{34}$$

Then j_m is of the form $4k+2$ and from (5) of Theorem 2, for $m \geq 4$,

$$\begin{aligned}
\bar{d}_m(j_m) &= \bar{d}_{m-1}\underbrace{\left(j_{m-1} + (-1)^m\right)}_{\text{odd}} + 1 \\
&\overset{(33)}{=} \frac{1}{2}\bar{d}_{m-1}\left(j_{m-1} + (-1)^m - 1\right) \\
&\quad + \frac{1}{2}\bar{d}_{m-1}\left(j_{m-1} + (-1)^m + 1\right) + 1 \\
&= \frac{1}{2}\bar{d}_{m-1}\left(j_{m-1}\right) + \frac{1}{2}\bar{d}_{m-1}\underbrace{\left(j_{m-1} + 2(-1)^m\right)}_{\text{divisible by 4}} + 1 \\
&\overset{(4)}{=} \frac{1}{2}\bar{d}_{m-1}\left(j_{m-1}\right) + \frac{1}{2}\bar{d}_{m-1}\left(\frac{j_{m-1}}{2} + (-1)^m\right) + 1 \\
&\overset{(34)}{=} \frac{1}{2}\bar{d}_{m-1}\left(j_{m-1}\right) + \frac{1}{2}\bar{d}_{m-1}\left(j_{m-2}\right) + 1.
\end{aligned}$$

To show that $\hat{d}_2 = \hat{d}_3 = 2$ is straightforward from (7). By induction we will prove that if $\hat{d}_{m-2} = \bar{d}_{m-2}(j_{m-2})$ and $\hat{d}_{m-1} = \bar{d}_{m-1}(j_{m-1})$, then $\hat{d}_m = \bar{d}_m(j_m)$. We have from (5) and (33) that

$$\begin{aligned}
\bar{d}_m(4k+2) &= \frac{1}{2}\bar{d}_{m-1}(2k) + \frac{1}{2}\bar{d}_{m-1}(2k+2) + 1 \\
&\leq \frac{1}{2}\hat{d}_{m-1} + \frac{1}{2}\hat{d}_{m-2} + 1
\end{aligned} \tag{35}$$

where the inequality follows from (4) since either $2k$ or $2k+2$ is divisible by 4. Furthermore,

$$\begin{aligned}
\bar{d}_m(4k) &= \frac{1}{2}\bar{d}_{m-1}(2k) + \frac{1}{2}\bar{d}_{m-1}(2k) \\
&\leq \frac{1}{2}\hat{d}_{m-1} + \frac{1}{2}\hat{d}_{m-2} + 1
\end{aligned} \tag{36}$$

and using the interpolation formula (33) combined with (35) and (36) yields

$$\bar{d}_m(2k+1) \leq \frac{1}{2}\hat{d}_{m-1} + \frac{1}{2}\hat{d}_{m-2} + 1. \tag{37}$$

Combining (35) – (37) and using the definition of $\hat{d}_m$ gives

$$\hat{d}_m \leq \frac{1}{2}\hat{d}_{m-1} + \frac{1}{2}\hat{d}_{m-2} + 1 = \bar{d}_m(j_m).$$

But $\hat{d}_m \geq \bar{d}_m(j_m)$ by definition which implies that $\hat{d}_m = \bar{d}_m(j_m)$.

APPENDIX E
DERIVATION OF THE BER FOR M-PAM

In Figure 6, a general M-PAM signal constellation is shown and the notation that will be used to calculate the exact bit error rate of communication systems using this constellation is introduced. The signal points are separated by a distance ρ and each signal point $\mathbf{s}_i$ is associated with a decision region Ω_i. Assuming equal a priori probability for the transmitted symbols and an ML receiver, the decision regions will be as in Figure 6. For internal points (i.e., $\mathbf{s}_1, \ldots, \mathbf{s}_{M-2}$) these regions are strips in the real plane and the decision regions for the two edge points ($\mathbf{s}_0$ and $\mathbf{s}_{M-1}$) are half-planes (this is for ease of illustration, the ML receiver can just as well operate on a scalar, in which case the decision regions are intervals on the real line). In Figure 6, Ω_0 is the half-plane associated with $\mathbf{s}_0$ and Ω_{M-3} is the strip associated with the internal signal point $\mathbf{s}_{M-3}$. If all signal points are assigned strips congruent to Ω_{M-3}, the *overload* regions are the two regions that are outside the strips surrounding the outer points, $\mathbf{s}_0$ and $\mathbf{s}_{M-1}$.

To calculate the exact bit-error rate of this signal constellation, we again start from (1). From Figure 6 we see that, assuming a symmetrical channel, the M-PAM problem has two types of decision regions, so we define for all $k \in \mathbb{Z}$

$$P(k) \triangleq \Pr\left(\mathbf{s} - \left(k + \frac{1}{2}\right)\rho \leq \mathbf{r} \leq \mathbf{s} - \left(k - \frac{1}{2}\right)\rho\right)$$

to be the probability that the received vector falls within a strip k steps away from the transmitted signal point s and

$$P'(k) \triangleq \Pr\left(\mathbf{r} \leq \mathbf{s} - \left(k + \frac{1}{2}\right)\rho\right) = \sum_{i=k+1}^{\infty} P(i) \qquad (38)$$

to be the probability that the received vector ends up in the half-plane with a boundary at a distance $(k + 1/2)\,\rho$ from the transmitted signal point s. Since the noise is assumed to be symmetrical, $P(k) = P(-k)$ and $P'(k) = P'(-k)$, so that $P(k)$ and $P'(k)$, $k = 1, 2, \ldots$, are the only probabilities needed to complete the BER expression. If we briefly use the simplifying notation $d(i, j)$ to denote $d_H(\mathbf{c}_i, \mathbf{c}_j)$, we obtain for the average BER of the M-PAM constellation

$$\begin{aligned} P_{\text{b}} &= \frac{1}{mM}\sum_{k=1}^{M-1} d(k,0)\left(P(k) + P'(k)\right) \\ &+ \frac{1}{mM}\sum_{l=1}^{M-2}\sum_{k=0}^{M-1} d(k,l)P(k-l) \\ &+ \frac{1}{mM}\sum_{k=0}^{M-2} d(k, M-1)\left(P(M-1-k) + P'(M-1-k)\right). \end{aligned}$$

Now, by collecting terms relating to the overload regions (i.e., all terms involving $P'(k)$) and terms relating to internal points (all terms involving $P(k)$) we can, by making use of (38), express the above equality on the form

$$P_{\text{b}} = \frac{2}{m}\sum_{k=1}^{M-1} P(k)d^-(k) + \frac{2}{m}\sum_{k=2}^{\infty} P(k)d^+(k)$$

where we have defined for $k = 1, \ldots, M-1$

$$d^-(k) \triangleq \frac{1}{M}\sum_{l=k}^{M-1} d_H(\mathbf{c}_{l-k}, \mathbf{c}_l)$$

and also for $k = 2, 3, \ldots$

$$d^+(k) \triangleq \frac{1}{2M}\sum_{l=1}^{\min(k,M)-1}\left(d_H(\mathbf{c}_0, \mathbf{c}_l) + d_H(\mathbf{c}_{M-1}, \mathbf{c}_{M-1-l})\right).$$

Note that in the sum for $d^-(k)$, the number of terms are decreasing as k is increasing and also that the sum for $d^+(k)$ is not depending on k.

REFERENCES

[1] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, 3rd edition, 1995.

[2] F. Gray, "Pulse code communications," U. S. Patent No. 2632058, Mar. 1953.

[3] P. J. Lee, "Computation of the bit error rate of coherent M-ary PSK with Gray code bit mapping," *IEEE Transactions on Communications*, vol. COM-34, no. 5, pp. 488–491, May 1986.

[4] M. I. Irshid and I. S. Salous, "Bit error probability for coherent M-ary PSK systems," *IEEE Transactions on Communications*, vol. COM-39, no. 3, pp. 349–352, Mar. 1991.

[5] M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques*, Prentice Hall, Englewood Cliffs, New Jersey, 1994.

[6] J. Lu, K. B. Letaief, J. C.-I. Chuang, and M. L. Liou, "M-PSK and M-QAM BER computation using signal-space concepts," *IEEE Transactions on Communications*, vol. COM-47, no. 2, pp. 181–184, Feb. 1999.

[7] J. Lassing, E. G. Ström, T. Ottosson, and E. Agrell, "Computation of the exact bit error rate of coherent M-ary PSK with Gray code bit mapping," *IEEE Transactions on Communications*, vol. COM-51, no. 11, pp. 1758–1760, Nov. 2003.

[8] M. K. Simon and M. S. Alouini, *Digital Communications over Fading Channels—A Unified Approach to Performance Analysis*, John Wiley & Sons, 2000.

[9] E. N. Gilbert, "Gray codes and paths on the n-cube," *Bell System Technical Journal*, vol. 37, pp. 815–826, May 1958.

[10] J. E. Ludman, "Gray code generation for MPSK signals," *IEEE Transactions on Communications*, vol. COM-29, no. 10, pp. 1519–1522, Oct. 1981.

[11] C. E. Sundberg, "Bit error probability properties of Gray-coded M.P.S.K signals," *IEE Electronics Letters*, vol. 11, no. 22, pp. 542–544, Oct. 1975.

[12] R. D. Wesel, X. Liu, J. M. Cioffi, and C. Komninakis, "Constellation labeling for linear encoders," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2417–2431, Sept. 2001.

[13] J. Gross and J. Yellen, *Graph Teory and its applications*, CRC Press, Boca Raton, 1999.

[14] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Transactions on Communications*, vol. COM-50, no. 7, pp. 1074–1080, July 2002.