The binary reflected Gray code is optimal for M-PSK

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 an M-PSK communication system. 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. Under certain assumptions on the channel values, the answer is that the labeling proposed by Gray, the binary reflected Gray code, is the optimal labeling for M-PSK systems, 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 in an M-ary phase shift keyed (M-PSK) communication system that will minimize the probability of bit errors. As was shown in [1], the average bit error probability (BER) of coherent M-PSK (M is assumed to be a power of two) is given by,

$$P_b = \frac{1}{m}\sum_{k=1}^{M-1} \bar{d}(k)\, P(k), \qquad (1)$$

where $m = \log_2 M$ and $P(k)$ is the probability that the received signal is displaced into 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 a binary labeling of the symbols in a signal constellation. This function is the average number of bits that differ in symbols separated by k steps, averaged over all M symbols. The probabilities $P(k)$ given, for example, by the expressions in [2, p. 201], 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 making it is 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. Such labelings are known as Gray labelings.

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 [1] the exact values are given for all M. Common for all these references is the use of the binary reflected Gray labeling of the constellation symbols. The binary reflected labeling was suggested by Frank Gray in a patent from 1953 [7] as a means of reducing the coding error in a pulse code communication system. This system referred to by Gray can be viewed as an analog-to-digital converter, in which an analog signal controls the deflection of a sweeping electron beam. The 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 that is essentially 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 mapping the reflected binary code, due to its recursive construction method (see Section III-A). Gray identifies the trivial operations defined in Section II below, but his treatment only concerns Gray labelings with the symmetric properties imposed by the recursive reflection construction method.

In most references in the literature the binary reflected labeling proposed by Gray is referred to simply as ‘the Gray’ labeling, without further specification. However, for $m > 3$ there exist several Gray labelings that have different ADS and as m increases the number of such labelings rapidly becomes very large [8–10]. To find the labeling that gives the lowest possible BER, it is necessary to consider the entire class of binary labelings having the Gray property. Only a fraction of the labelings in this class can be generated from the recursive methods defined below, but we will show that it is possible to generate the optimum labeling by these recursions. For illustration, in Table I are given two binary labelings having the Gray property along with their respective ADS. By comparing the ADS of the two labelings it is seen that, from (1), these labelings will indeed result in different BER.

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).

The natural approach to finding the labeling that minimizes (1) 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 the channel values $P(k)$ decay sufficiently quick with k to make the minimization of the BER equivalent to sequential minimization of the components of the ADS. Such channels will be called "well-behaved" in the rest of this text. 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 (1) 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 (BRGC) is the unique labeling that results in the slowest increasing ADS among all possible bit-to-symbol mappings. For well-behaved channels this mapping will be optimal in the sense of providing the lowest possible value of the BER.

In a related work [11], the effect of the constellation labeling on the constellation's *edge profile* is evaluated. The aim of the work in [11] is to provide a formal answer to what labelings are sensible for using in trellis-coded modulation systems. However, the edge profile is related to the *union bound* on the BER of the system using a particular constellation labeling, although bit error probability is not the scope of [11] and consequently not mentioned. The edge profile cannot be used to determine the effect on the exact BER of an M-PSK system.

The outline of the paper is as follows. The necessary nomenclature and definitions are given in section II along with some remarks that will prove useful. In section III, two recursive construction methods for binary Gray labelings, reflection and expansion, are introduced. The proof of the optimality of the BRGC is given in section IV and section V concludes the discussion.

II. PRELIMINARIES

To simplify the discussion, we start by making some necessary definitions. We will focus our discussion on binary, cyclic labelings.

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), $\mathcal{C} = (\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M-1})$, 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.

Note that we call the labeling with adjacent vectors differing in a single position a *Gray code* since this is the ubiquitous designation in the literature, although the term *Gray labeling* would be more appropriate.

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.,

$$\bar{d}(k) \triangleq \frac{1}{M} \sum_{i=0}^{m-1} \sum_{l=0}^{M-1} \left| [\mathbf{c}_l]_i - [\mathbf{c}_{(l+k) \bmod M}]_i \right| \tag{2}$$

for all $k \in \mathbb{Z}$, where the notation $[\mathbf{c}_l]_i$ denotes the bit in position i of $\mathbf{c}_l$.

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 (2), 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.

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 does not affect the ADS of the labeling.

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

Definition 7—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 [12, pp. 226]. It is known that there exist Hamiltonian cycles of all orders $m \ge 1$, which is also evident from the constructions in Section III.

III. 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 $m \ge 1$.

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	$\bar{a}(k)$	$b(k)$
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

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 the optimum 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, so 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 illustrated in Figure 1.

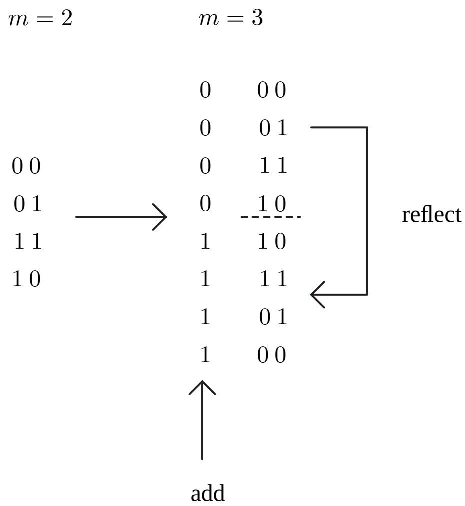

Fig. 1. Construction of Gray code of order $m = 3$ from a Gray code of order $m = 2$ by means of reflection.

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 [7], which is still the most commonly encountered Gray code in communications, can be defined as follows.

Definition 8—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.

IV. 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 IV-B.

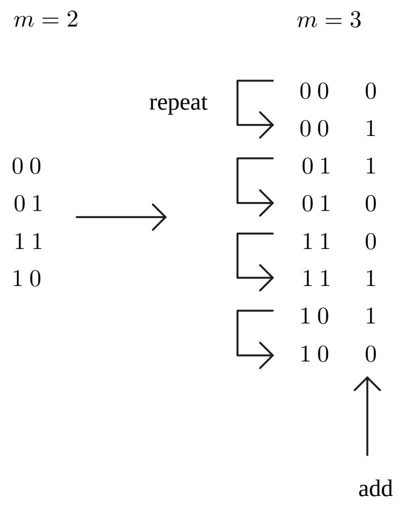

Fig. 2. Construction of Gray code of order $m = 3$ from a Gray code of order $m = 2$ by means of expansion.

Theorem 1—Optimality of BRGC for M-PSK: The binary reflected Gray code of order m is the optimal labeling for 2^m-PSK, in the sense of Definition 5. The labeling is unique up to trivial operations.

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 the Appendix.

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

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

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

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

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 the Appendix.

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 *optimum* ADS in the sense of Definition 5. According to the discussion in Section I, a mapping with optimum ADS will asymptotically result in the lowest possible BER for the M-PSK system. We will show in the following that the binary reflected Gray code (BRGC) is the unique labeling, up to trivial operations, with optimum ADS.

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

Proof: The lemma is trivial for $m = 2$, since, up to trivial operations, only one 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 (3)–(5). 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 $m \geq 2$. □

V. DISCUSSION AND CONCLUSION

We have addressed the problem of finding what constellation labeling that will produce the lowest possible BER among all possible labelings. The search is done under the assumption of a well-behaved channel, for which the channel values $P(k)$ decay quickly enough to ensure that sequential minimization of the components of the ADS yields the minimum BER. We have shown that the best way of labeling an M-PSK constellation under this assumption is by using 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 M-PSK is treated. In most cases, the BRGC is used, but referred to simply as 'the Gray' labeling and the fact that a wealth of different Gray labelings exist and their impact on the BER is often neglected. The proofs in this paper validates the assumption of the BRGC for M-PSK constellation labeling and allows for a more clear presentation of the topic of BER calculation for this type of communication system.

APPENDIX
PROOFS OF THEOREMS AND LEMMAS

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

$$\begin{aligned}
\bar{\delta}(u,2i) &= \frac{1}{2P} \left(\sum_{l=0}^{P-1} |u_{2l} - u_{2l+2i}| + \sum_{l=0}^{P-1} |u_{2l+1} - u_{2l+1+2i}| \right) \\
&= \frac{1}{2P} \left(\sum_{l=0}^{P-1} |b_l - b_{l+i}| + \sum_{l=0}^{P-1} |b_l - b_{l+i}| \right) \\
&= \frac{1}{2P} \left(P\bar{\delta}(b,i) + P\bar{\delta}(b,i) \right) = \bar{\delta}(b,i) \qquad (6)
\end{aligned}$$

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

$$\begin{aligned}
\bar{\delta}(u,2i+1) &= \frac{1}{2P} \left(\sum_{l=0}^{P-1} |u_{2l} - u_{2l+2i+1}| \right. \\
&\qquad \left. + \sum_{l=0}^{P-1} |u_{2l+1} - u_{2l+2i+2}| \right) \\
&= \frac{1}{2P} \left(\sum_{l=0}^{P-1} |b_l - b_{l+i}| + \sum_{l=0}^{P-1} |b_l - b_{l+i+1}| \right) \\
&= \frac{1}{2P} \left(P\bar{\delta}(b,i) + P\bar{\delta}(b,i+1) \right) \\
&= \frac{1}{2}\bar{\delta}(b,i) + \frac{1}{2}\bar{\delta}(b,i+1) \qquad (7)
\end{aligned}$$

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 $\bar{d}_{m-1}(k)$. 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 (2) for the ADS of $\mathcal{C}_m$

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

where $i = 0$ corresponds to the last coordinate in the codewords. 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 (2) can be removed without affecting the result, by instead considering the periodic repetition of the codewords, we can use (6) and (7) above to obtain, for all integers k,

$$\begin{aligned}
\bar{\nu}_m(2k) &= \bar{d}_{m-1}(k) \\
\bar{\nu}_m(2k+1) &= \frac{1}{2}\bar{d}_{m-1}(k) + \frac{1}{2}\bar{d}_{m-1}(k+1) \qquad (8)
\end{aligned}$$

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

$$\begin{aligned}
\bar{d}_m^{(0)}(4k) &= 0 \\
\bar{d}_m^{(0)}(4k+2) &= 1 \\
\bar{d}_m^{(0)}(2k+1) &= 1/2 \qquad (9)
\end{aligned}$$

Combining (8) and (9) we have

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

which completes the proof of Theorem 2. □

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 $m \geq 2$. To show $\bar{d}_m(3) \geq 2$ for $m \geq 3$, we start by rewriting the equation for the ADS given by (2) as

$$\begin{aligned}
\bar{d}(k) &= \frac{1}{M} \sum_{l=0}^{M-1} \sum_{i=0}^{m-1} \left| [\mathbf{c}_l]_i - [\mathbf{c}_{(l+k) \bmod M}]_i \right| \\
&= \frac{1}{M} \sum_{l=0}^{M-1} d(l,k), \qquad \forall k \in \mathbb{Z} \qquad (10)
\end{aligned}$$

where $d(l,k)$ is the number of bits that differ between $\mathbf{c}_l$ and $\mathbf{c}_{(l+k) \bmod M}$. From this we show that no two consecutive terms $d(l,3)$ and $d(l+1,3)$ in (10) 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 $|\mathbf{c}_0 - \mathbf{c}_3| = 1$, so that $d(0,3) = 1$. Since $\mathbf{c}_1$ and $\mathbf{c}_4$ are separated by an odd number of steps this implies that $|\mathbf{c}_1 - \mathbf{c}_4| \geq 3$, so that necessarily $\bar{d}(3) \geq 2$. □

Proof of Lemma 4: The first statement of the lemma follows immediately from (5).

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 $d(l,3)$ 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 corresponds simply 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 $|\mathbf{c}_i - \mathbf{c}_{i+3}| = 1$ for any even $0 \leq i \leq M-4$, $(\mathbf{c}_i, \mathbf{c}_{i+1}, \mathbf{c}_{i+2}, \mathbf{c}_{i+3})$ forms a square in Q_m. Hence, $\mathbf{c}_{i+1} - \mathbf{c}_i$ and $\mathbf{c}_{i+2} - \mathbf{c}_{i+3}$ are equal, or

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

The difference vector $\mathbf{\Delta}$ 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

$$\mathbf{\Delta}' \triangleq \mathbf{c}'_1 - \mathbf{c}'_0 = \mathbf{c}'_2 - \mathbf{c}'_3 = \ldots = \mathbf{c}'_{M-2} - \mathbf{c}'_{M-1} = 0\ldots01,$$

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

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

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 $\mathbf{\Delta}'$, 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 III-B yields $\mathcal{C}'_m$. □

REFERENCES

[1] 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. 51, no. 11, pp. 1758–1760, Nov. 2003.

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

[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. 47, no. 2, pp. 181–184, Feb. 1999.

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

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

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

[10] 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.

[11] 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.

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