Gray Codes and the Thue - Morse - Hedlund Sequence

H. Fredricksen

Mathematics Department Code MA
Naval Postgraduate School
Monterey, CA 93943

1. Introduction

A gray code is a sequence of all 2^k binary k-tuples (code words) arranged in order so that successive code words differ only in a single coordinate. The name "gray codes" is assigned to this arrangement of the codewords [8] but the idea of these code words is very old and goes back at least to Cardano in 1550.

Gray codes are used to minimize the number of bit errors in transmitted signals [3]. The single change in consecutive code words helps detect some multiple errors when they occur. The code also provides a solution to a type of combination lock [14] as well as appearing in a number of ways to provide combinatorial objects [4], [6], [13]. Finally, gray codes provide the theory for solution of a variety of mathematical puzzles [7]. In sections 2 and 3 of this paper we further describe the gray code and methods of its construction as well as a sequence definable from the gray code called the change sequence.

In section 4 we define another sequence called the Thue - Morse - Hedlund sequence [9]. This sequence appears in the study of almost periodic sequences in symbolic dynamics [2]. Another sequence derived from the Thue - Morse - Hedlund sequence has interesting properties in its own right. In section 5 we describe a relationship which exists between this sequence and the change sequence of the gray code. This interesting relationship is the major contribution of the current paper.

2. The Reflected Gray Code

A gray code of order k is a permutation of the integers, $0, 1, \cdots, 2^k - 1$ each written in binary. The properties that apply on the permutation require that consecutive images, $\pi(i)$ and $\pi(i+1)$, differ in exactly one coordinate. The number of different gray codes which exist for a given order k is not known. Related questions also go by the name snake-in-the-box codes [10].

We consider a specific gray code called the reflected gray code. For the reflected gray code of order $k, \pi(0) = 0$ and the rest of the code can be determined recursively from the gray code of order $k-1$ as described by the following procedure:

Procedure Gray (k).

1. To the code, Gray ($k - 1$), append a column of 0's to the left of the $k - 1$ columns of the code.

JCMCC 11 (1992), pp. 3-11

2. Copy the rows of Gray $(k-1)$, in reverse order, starting with row $2^{k-1}-1$ and ending with row 0.
3. To the last 2^{k-1} rows append a column of 1's to the left of the $k-1$ columns.

If Gray (1) is defined as the 2 long column vector containing 0 and 1, then we illustrate the procedure to generate $G(2)$, $G(3)$ and $G(4)$ in figure 1, [4].

		000	0000	1100
		001	0001	1101
	00	011	0011	1111
0	01	010	0010	1110
1	11	110	0110	1010
	10	111	0111	1011
		101	0101	1001
		100	0100	1000
G(1)	G(2)	G(3)	G(4)	

Figure 1. Reflected Gray Codes, $G(k)$, $k = 1,2,3,4$.

As the last vector of $G(k)$ also differs from the first vector in exactly one coordinate, the gray code is circular. Traversing the vectors of the code in order forms a Hamiltonian circuit of the points of binary k-space, [8].

An alternate way of defining the reflected gray code is by explicitly showing the particular permutation associated with the reflected gray code. We represent the integer i as the binary vector $b_k b_{k-1} \cdots b_1$. Then, the permutation $\pi(i)$ of i that represents ith gray code vector is represented as

$$\pi(i) = g_k g_{k-1} \cdots g_1$$

where the g_j are defined as

$$g_k = b_k$$

and

$$g_l = g_{l+1} \oplus g_l \quad \text{for } l = k-1, k-2, \cdots, 1$$

and where $\oplus$ represents modulo 2 addition. Also, $\pi^{-1}(g)$ can be described as the binary vector $b_k b_{k-1} \cdots b_1$ where

$$b_k = g_k$$

and

$$b_l = \oplus \sum_{t=l}^{k} g_t \quad \text{for } l = k-1, k-2, \cdots, 1$$

and again the summation is modulo 2 on the bits of g.

3. The Change Sequence

The definition of the gray code states that only one bit coordinate of a codeword changes to produce the next word in the code. Fom the recursive definition of the code it is easy to see that the sequence of coordinates that change, as Gray (k) is developed, is a palindrome of length 2^{k-1} with the 2^{k-1} th position of the sequence equal to k . Thus, for small values of k, the change sequences are as given in table 1 for the respective reflected gray codes, [6].

$k = 1$	1
$k = 2$	121
$k = 3$	1213121
$k = 4$	121312141213121
$k = 5$	1213121412131215121312141213121

Table 1 The change sequences for $k = 1,2,3,4,5$

The change sequence for the reflected gray code of order k appears frequently in mathematical puzzles, such as the Towers of Hanoi and the Chinese Rings [7]. In the former case, the complete cycle is exhausted, while in the latter, a "solution" is found after $\lfloor 2^{k+1}/3 \rfloor$ moves [14].

It can be noted that the change sequence for the reflected gray code of order k can be generated by a set of k finite automata each of which is capable of counting to 2 and writing a digit. We imagine a blank tape along which the automata are to march. They read the contents of the cells they encounter and the automaton numbered j prints a j in the first blank space it encounters and again prints a j in each alternate blank space it encounters. The various automata print the digits on the blank tape as shown in figure 2.

Positions ⟶

										1										2										3	
	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
Automata 1	1		1		1		1		1		1		1		1		1		1		1		1		1		1		1		1
Automata 2		2				2				2				2				2				2				2				2	
Automata 3				3								3								3								3			
Automata 4								4																4							
Automata 5																5															
Final tape	1	2	1	3	1	2	1	4	1	2	1	3	1	2	1	5	1	2	1	3	1	2	1	4	1	2	1	3	1	2	1

Figure 2. The positions numbered by the automata.

Obviously, the same approach can number an infinite tape if an infinite number of automata are available. It is not hard to see how a single automata can number a tape of length 2^{k-1} if it is capable of counting as high as 2^{k-1} and can recognize integers as large as k.

In fact, the changed coordinate of the ith gray code vector can be computed as the number $j + 1$ where $2^j || i$. That is, i is of the form $i = 2^j \cdot l$ and l is odd, so that i is divisible exactly j times by 2. Thus all the odd vectors are divisible by 2^0, every 4th vector, starting with the second, is divisible by 2^1, etc. This is evidenced by the example of figure 2.

4. Thue - Morse - Hedlund Sequence

The Thue - Morse - Hedlund (TMH) sequence is an infinite length binary sequence which can be used to prove the existence of "recurrent" non-periodic geodesics [11] on certain surfaces of negative curvature. It also is a solution to a problem [12] concerning a drawing rule in a variant of chess. The sequence has interesting properties. For example, TMH has the BBb property, i.e. the subsequence $b_1 b_2 \cdots b_n b_1 b_2 \cdots b_n b_1$ does not appear in the sequence for any positive integer value of n [11]. Subsequent investigations [5] have resulted in classes of sequences which do not contain the blocks $B\pi(B)$, where B is any block of length n and $\pi(B)$ is any permutation of that block, whose elements come from the alphabet $A = \{1, 2, \cdots, k\}$.

The TMH sequence can be generated in a variety of ways. A recursive method [9] proceeds as follows:

$$(A) \qquad T_1 = t_1 = 0 \quad \text{and}$$
$$T_{2^n} = T_{2^{n-1}} \; \overline{T_{2^{n-1}}} \quad \text{for } n = 1, 2, \cdots$$

where $\overline{T}$ is the binary complement of T.

An equivalent method to generate the sequence is:

$$(B) \qquad T_1 = t_1 = 0 \quad \text{and}$$
$$T_{2^n} = T_{2^{n-1}} \, r(\overline{T_{2^{n-1}}}) \quad \text{for } n \text{ odd}$$
$$T_{2^n} = T_{2^{n-1}} \, r(T_{2^{n-1}}) \quad \text{for } n \text{ even}.$$

where $r(T)$ reverses the bits of vector T.

Finally, the TMH sequence can be generated by making a computation on the bits of an appropriate vector.

(C) If $i = \sum b_j 2^j$, then the bit t_i of the TMH sequence T can be expressed as the modulo 2 sum $t_i = \oplus \sum b_j$.

It is shown in [2] that TMH can be self-generated in the sense that bits of T determine subsequent bits. Thus when a bit 0 is read, the bits 01 are appended to the end of the sequence and when 1 is read, 10 is written. Getting started with $t_1 = 0$ is tricky as we read 0 and overwrite that 0 and the next 1. It is also shown [2], that if TMH is viewed as a fraction preceeded by a binary point, that the TMH sequence is a solution of a quadratic equation. Thus, according to the theory developed in [9], the TMH sequence is almost periodic.

5. The Delta Thue Sequence

We define the delta-j Thue sequence $\{\delta_i^{(j)}\}$ of the TMH sequence as

$$\delta_i^{(j)} = t_i \oplus t_{i+j} \quad i = 0, 1, 2, \cdots .$$

From the previous section, by following any of the generation rules. we produce and display in figure 3, the TMH sequence and the sequences $\delta^{(1)}$ and $\delta^{(2)}$ as defined above.

TMH	0 1 1 0 1 0 0 1 1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0 0 1 1 0 1 0 0 1
$\delta^{(1)}$	1 0 1 1 1 0 1 0 1 0 1 1 1 0 1 1 1 0 1 1 1 0 1 0 1 0 1 1 1 0 1
$\delta^{(2)}$	1 1 0 0 1 1 1 1 1 1 0 0 1 1 0 0 1 1 0 0 1 1 1 1 1 1 0 0 1 1

Figure 3. The TMH sequence and its first two delta sequences.

We establish the following property of the sequence $\delta^{(1)}$. For simplicity we refer to $\delta^{(1)}$ as δ.

Theorem 1. *δ_i is equal to 1 if and only if $i + 1 = (1 + 2j)2^{2k}$ for some non-negative integers j, k.*

Proof: By the description (C) of the TMH sequence, the bit t_i differs from the bit t_{i+1} when i is even, since then its binary representation terminates in a 0. In this case $k = 0$.

If i is odd, and the binary representation of i terminates with $2k$ ones, then $i + 1$ will have a different parity of ones in its binary representation. Thus $\delta_i = 1$ for such an i. If i terminates with $2k + 1$ ones, then the number of ones in $i + 1$ will have the same parity as does i and $\delta_i = 0$ in this case. ∎

Corollary. *Corresponding to the positions where i is even, $\delta_i = 1$. Where $i \equiv 2 \pmod 4$, $\delta_i = 0$. Where $i \equiv 4 \pmod 8$, $\delta_i = 1$, etc.*

Thus δ can be constructed by following the procedure below.

(D) Write a 1 in every other position of a blank tape. In every other of the blank spaces remaining, write a 0. In the remaining blank spaces write a 1 in every alternate space, etc.

The procedure (D) should sound familiar. The sequence displayed in figure 2, when written in modulo 2 notation, is the sequence of procedure (D) Thus we have established a relationship between the reflected gray code change sequence and the sequence of the TMH sequence.

There is also a recursive procedure to generate δ which proceeds as follows:

$$(E) \qquad D_0 = \delta_0 = 1 \qquad \text{and}$$

$$D_{2^n} = D_{2^{n-1}} (D_{2^{n-1}})^u \qquad \text{for } n = 0, 1 \cdots$$

where D_m^u is the same m-tuple as D_m except the *last bit* is complemented.

In figure 3, evidently $\delta^{(2)}$ is a dilated version of $\delta^{(1)}$ with every bit of $\delta^{(1)}$ appearing twice. We verify this in the following theorem.

Theorem 2. *The sequence $\delta^{(2)}$ is the dilated by two version of the sequence $\delta^{(1)}$.*

Proof: In theorem 1 we showed that $\delta_i^{(1)}$ was equal to 1 if and only if $(i+1) = (1+2j)2^{2k}$ for some integers k, j. We show that the bits of $\delta^{(1)}$ in position i are the same bits in $\delta^{(2)}$ for positions $2i$ and $2i+1$. We proceed by considering the separate cases.

From theorem 1, the bit $\delta_i^{(1)}$ is 1 if i is even or if i is of the form $x_k \cdots x_{k-j}$ $011 \cdots 1$ and the number of terminal 1's is even. If i is even then $2i$ ends in at least two zeros so $2i+2$ has one more 1 than $2i$ does in their respective binary representations. Similarly $2i+1$ and $2i+3$ have a different parity of ones in their respective binary representations. Thus $\delta_l^{(2)} = 1$ for $l = 2i$ and $2i+1$ where i is even.

If i is of the form $xxx \cdots x011 \cdots 1$ with an even, positive number of terminal 1's, then $2i$ and $2i+2$ are represented as $xx \cdots x011 \cdots 10$ and $xx \cdots x100 \cdots 0$ respectively and have a different parity of 1's in their binary representations. In a like manner, $2i+1$ and $2i+3$ are represented as $xx \cdots x011 \cdots 11$ and $xx \cdots x10 \cdots 01$, respectively, to satisfy the claim for the case $\delta_i^{(1)} = 1$.

The case for $\delta_i^{(1)} = 0$ is handled exactly similarly where now i is represented as $xx \cdots x011 \cdots 1$ ending in an odd, positive number of 1's. The rest of the argument follows precisely as before, mutatis mutandis. ∎

When one considers the sequence $\delta^{(4)}$ it is evident that each respective bit of $\delta^{(1)}$ is dilated 4 times in $\delta^{(4)}$. Examining the various cases for the bits $\delta_i^{(1)}$, the same arguments as above on $\delta_l^{(4)}$ for $l = 4i, 4i+1, 4i+2$ and $4i+3$ show that the observed behavior always occurs. We are led to make the following conjecture.

Conjecture. *The delta sequence $\delta^{(2j)}$ is the dilated by two sequence of the delta sequence $\delta^{(j)}$.*

In section 4 it was stated that TMH has the BBb property. Here we give an alternate proof of this fact by operating on the δ sequence.

Theorem 3. *The subsequence $e_1 e_2 \cdots e_n e_1 e_2 \cdots e_n$ does not appear as $2n$ consecutive bits in the sequence δ if $\oplus \sum_{i=1}^{n} e_i = 0$.*

Note. *The restriction on $\oplus \sum e_i = 0$ is necessary to ensure we do not have the case $B\overline{B}b$ in the TMH sequence.*

Proof: Suppose $d_0 d_1 d_2 \cdots$ is the delta sequence of the TMH sequence. Suppose further that $d_p d_{p+1} \cdots d_{p+n-1} = d_{p+n} d_{p+n+1} \cdots d_{p+2n-1}$ is an n-long repeat and

$$\oplus \sum_{i=p}^{p+n-1} d_i = 0.$$

Let d_{p+q} be the first 0 in $d_p d_{p+1} \cdots d_{p+n-1}$. Then $d_{p+q+n} = 0$ by assumption. So, by theorem 1 we must have

$$1 + p + q = (1 + 2j_1)2^{2k_1+1}$$

and

$$1 + p + q + n = (1 + 2j_2)2^{2k_2+1}$$

We proceed by assuming that there is an n-long repeat in the δ sequence of the TMH sequence and then find a contradiction. We actually exhibit a position that is 0 in one half of the $2n$-tuple and is 1 in the corresponding bit of the other half.

Case 1. $(k_1 = k_2)$.

We write $1+2j_1$ and $1+2j_2$ in binary as $a_s a_{s-1} \cdots a_0$ and $b_s b_{s-1} \cdots b_0$, respectively. We find the smallest t such that $a_t \neq b_t$ and define $m = \overline{a}_{t-1}\overline{a}_{t-2} \cdots \overline{a}_0 + 1$. Note, $m < n$. Then if we form $1 + 2j_1 + m$ and $1 + 2j_2 + m$ we see that exactly one of these is exactly divisible by an even power of 2 and the other is divisible by an odd power of 2. Thus, by theorem 1, the first of these is a 0 in the sequence and the second of these is a 1. This is a contradiction and case 1 is proved.

Case 2. $k_1 \neq k_2$.

Let $k_b = \max(k_1, k_2)$ and $k_s = \min(k_1, k_2)$ and associate j_b and j_s with the appropriate values of j_1 and j_2. Then the bit $d_z = 0$ where $z = 1 + p + q + \binom{0}{n} + 2^{2 \cdot k_s + 2}$ where by the notation $\binom{a}{b}$ we mean take the value a if $k_s = k_1$ and take the value b if $k_s = k_2$.

Case 2a. $n > 2^{2k_s+2}$.

By assumption, $d_y = 0$ where y is defined as $y = 1 + p + q + \binom{n}{0} + 2^{2k_s+2}$ as these positions, y and z, differ by n and the two n-tuples are supposed to agree. Then

$$y = (1 + 2j_b)2^{2k_b+1} + 2^{2k_s+2}$$

must be exactly divisible by an odd power of 2 since $d_y = 0$. But,

$$y = 2^{2k_s+1}((1 + 2j_b)2^{2(k_b-k_s)} + 2) = 2^{2k_s+1}(4t + 2)$$

which yields a contradiction and case $2a$ is established.

Case 2b. $n \leq 2^{2k_s+2}$.

Then, the number n can be expressed as the absolute difference

$$|(1 + 2j_1)2^{2k_1+1} - (1 + 2j_2)2^{2k_2+1}| \leq 2^{2k_s+2}.$$

Dividing through by 2^{2k_s+1} yields

$$|(1+2j_1)2^{2(k_1-k_s)} - (1+2j_2)2^{2(k_2-k_s)}| \leq 2.$$

But k_s has to equal one of k_1 and k_2. Thus one of these numbers is odd and the other is even. It follows that

$$|(1+2j_1)2^{2(k_1-k_s)} - (1+2j_2)2^{2(k_2-k_s)}| = 1$$

and therefore

$$|(1+2j_1)2^{2k_1+1} - (1+2j_2)2^{2k_2+1}| = 2^{2k_s+1}$$

so that if $n \leq 2^{2k_s+2}$, then $n = 2^{2k_s+1}$.

Now we follow the (E) construction of the delta sequence of the sequence TMH. It follows that every initial block D_{2^n} of length 2^n contains an odd number of 1's. Now, if $p \equiv 0$ (modulo $2^{2k_s+1} = n$), then the initial subsequence $D_n D_n^u D_n D_n$ contains the subsequences $d_p d_{p+1} \cdots d_{p+n-1}$ and $d_{p+n} \cdots d_{p+2n-1}$ and each of the two subsequences as one of either D_n or D_n^u. But at most one of these, D_n^u, contains an odd number of 1's to lead to a contradiction.

Finally, the last case to consider is when p is not a multiple of n. Then the subsequence $d_p d_{p+1} \cdots d_{p+2n-1}$ overlaps one of the divisions between the blocks D_{2n} and D_{2n}^u, or D_{2n}^u and D_{2n}, or D_{2n} and D_{2n}. We consider only the first of these possibilities and leave the other two possibilities to the reader.

	D_{2n}		D_{2n}^u	
	$(d_p \cdots$	$d_{p+n-1})$	$(d_{p+n} \cdots d_{p+2n-1})$	
1	2	3	4	5

In the diagram there are five regions to be considered, defined by the division between D_{2n} and D_{2n}^u and also by the location of d_p. If region 1 has an odd (even) number of ones, then region 2 must have an even (odd) number of ones since $||D_{2n}||$ is odd where $||x||$ denotes the number of ones in the vector x. Then also region 3 has an even (odd) number of ones, since $||d_p \cdots d_{p+n-1}||$ is supposed to be even. Also $||d_{p+n} \cdots d_{p+2n-1}||$ is supposed to be even so region 4 has an even number of ones. But since $||D_{2n}^u||$ is even, region 5 has an even (odd) number of ones. But D_{2n} and D_{2n}^u agree, except for the last bit. Now since $n = 1/2(2n)$, region 5 must have the opposite parity from region 2 so is therefore odd (even). The contradiction completes this subcase. The other two possibilities are handled the same, mutatis mutandis.

6. Conclusions

We have described the generation of gray codes, their change sequence, the Thue—Morse—Hedlund sequence and the delta sequence of this sequence. We show that the change sequence of the reflected gray code and the delta sequence of the TMH sequence are intimately connected. Perhaps other interesting applications can be found for these sequences when this relationship is completely exploited.

References

1. L.E. Baum and M.M. Sweet, *Badly approximable power series in characteristic 2*, Ann. of Math. (2) **105, #3** (1977), 573–580.
2. L.E. Baum, N. Herzberg, S. Lomanaco and M.M. Sweet, *Fields of almost periodic sequences*, Jour. of Combin. Th (A) **22** (1977), 176–180.
3. W.R. Bennett and J.R. Davey, *Data Transmission*, McGraw Hill, New York (1965).
4. J.R. Bitner, G. Ehrlich and E.M. Reingold, *Efficient generation of the binary reflected code and its applications*, CACM **19** (1976), 517–521.
5. L.J. Cummings, *Gray codes and strongly square-free strings*, Proceedings of Sequences '91, Positano (1991).
6. G. Ehrlich, *Loopless algorithms for generating permutations, combinations and other combinatorial configurations*, JACM 3 (1973), 500–513.
7. M. Gardner, *Mathematical Puzzles and Diversions*, Simon and Schuster, New York (1959).
8. E.N. Gilbert, *Gray codes and paths on the n-cube*, Bell Syst. Tech. Jour. **37, 9** (1958), 815–826.
9. W.H. Gottschalk and G.A. Hedlund, *Topological Dynamics*, Amer. Math. Soc. Colloquium Publ. **XXXVI** (1955).
10. V. Klee, *What is the maximum length of a d-dimensional snake*, Amer. Math. Monthly 77 (1970), 63–65.
11. M. Morse, *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. 22 (1921), 84–100.
12. M. Morse and G.A. Hedlund, *Unending chess, symbolic dynamics and a problem in semigroups*, Duke Math. Journal **11, #1** (1944).
13. E.M. Reingold, L. Nievergelt and N. Deo, *Combinational Algorithms: Theory and Practice*, Prentice Hall, Englewood Clis, N.J. (1977).
14. C.K. Yuen, *The separability of Gray codes*, IEEE Trans. Infor. Th. **IT-20** (1974), 668.