CONNECTIVITY OF SYNCHRONIZABLE CODES IN THE N-CUBE

L. J. Cummings

University of Waterloo

Abstract. A binary code has bounded synchronization delay if there exists an integer s such that at most s consecutive bits are required to establish word synchronization in any message. The code whose words are lexicographically least in the non-periodic orbits determined by cyclic permutation of all words of length n is called the canonical bounded synchronization delay code. It has the maximal number of words possible in a synchronizable code of fixed word length. Any code of fixed word length n can be represented as a set of vertices in the n-cube. We prove that the canonical bounded synchronization delay code is a connected subset of the n-cube.

Introduction.

Let Σ_n denote the set of 2^n binary strings over $\{0,1\}$ of length n. Consider the action of the full cycle permutation $\pi = (1\ 2\ \ldots\ n)$ on Σ_n given by

$$w^{\pi} = w_{\pi(1)} w_{\pi(2)} \ldots w_{\pi(n)}$$

for $w = w_1 w_2 \ldots w_n \in \Sigma_n$. The relation $v \sim w$ if $v = w^{\pi^i}$ for some $i = 1, 2, \ldots, n$ is an equivalence relation on Σ_n. The resulting equivalence classes are sometimes referred to as "circular strings". In the sequel we are concerned with just those circular strings that are primitive. A string w is *primitive* if $w \neq u^k$, for any substring u and positive integer k. Here the exponential notation is used to indicate the concatenation of k copies of the substring u. For example, 01000101 is primitive but $(0100)^2 = 01000100$ is not. Note that if w is primitive and $v \sim w$ then v is primitive. An easy counting argument using elementary Moebius inversion [1] shows that the number of primitive binary strings with fixed length n is

$$S(n,2) = \sum \mu(n/d) 2^d \tag{1}$$

where the summation is over all positive divisors d of n and μ is the Moebius function of elementary number theory.

It will be convenient to consider the binary strings of length n as ordered by the usual lexicographical ordering.

Definition 1: Λ_n is the set of binary strings which are lexicographically least in the primitive equivalence classes determined by $\sim$. From (1) the cardinality of Λ_n is $1/n\, S(n,2)$.

Definition 2: A substring v of a string w is a *right factor* of w if $w = uv$ for some substring u of w. The substring v is a *proper right factor* of w if $w = uv$ and u is not the empty string. One may similarly define *left factors* and *proper left factors*.

This paper was submitted for inclusion in JCMCC II and should have appeared there.

Lemma 1 (R. Lyndon). *A binary string of length n is in Λ_n if and only if it is strictly less in lexicographical ordering than each of its proper right factors.*

Proof: A proof of this fact for strings over arbitrary alphabets appears in [3; p. 65].

Lemma 2 (Golomb and Gordon). *Λ_n is a code with bounded synchronization delay.*

Proof: See [2] for a proof over arbitrary alphabets.

Proposition 1. *Every word of Λ_n begins with 0 and ends with 1.*

Proof: Suppose $w = 1w_1 \in \Lambda_n$. Since $w \in \Lambda_n$ it contains at least one 0 because 1^n cannot be in Λ_n. Let $w_1 = u0v$. Then $0v1u$ is a conjugate of w less in the lexicographical ordering than w. The argument is similar if w terminates in 0.

Notice that Proposition 1 cannot be extended to alphabets other than the binary alphabet.

Proposition 2. *If $w \in \Lambda_n$ then $0w$ and $w1 \in \Lambda_{n+1}$.*

Proof: It follows easily from the definition of lexicographical ordering for all binary strings that $0w < w < w1$. Let v be a proper right factor of $0w$. If $v = w$ then $0w < v$ as required. Otherwise, v must be a proper right factor of w. By Lemma 1, $w < v$ and we can conclude $0w < v$. A similar proof shows that $w1 \in \Lambda_n$.

Definition 3: The *n-cube* is the graph whose vertices are the strings of Σ_n with an edge between distinct vertices α and β if $d(\alpha,\beta) = 1$, where $d(\alpha,\beta)$ denotes the Hamming distance between α and β; i.e., the number of bits in which α and β differ as binary strings. The *weight* of a single string is the number of ones it contains.

Properties of Λ_n.

While there are several characterizations of Λ_n available, an efficient algorithm for the construction of Λ_n is not yet known. A recursive algorithm suggested in [3], for example, produces duplicates and so requires frequent "lookups".

Proposition 3. *The words in Λ_n of weight 2 are precisely the $\lceil \frac{n-2}{2} \rceil$ words*

$$0^a 10^b 1 \qquad 0 \le b < a. \tag{2}$$

Proof: The proof depends on Lemma 1.

Case 1. Let $w^* = 0^i 10^b 1$ be a proper right factor of a word w of the form (2). Since $i < a$ we have $w < w^*$ in the lexicographical ordering because the first 1 in w^* precedes the first 1 in w.

Case 2. By Proposition 1, any $w \in \Lambda_n$ begins with 0. Hence, any w of the form (2) satisfies $w < 10^b1$.

Case 3. Consider a proper right factor for $w \in \Lambda_n$ of the form 0^j1 with $j \leq b$. Since $a > b$ and $j < a$ we have

$$w = 0^a10^b1 < 0^j1$$

in the lexicographical ordering.

Theorem 1. *Λ_n is a connected subgraph of the n-cube.*

Proof: If $n = 2$ then Λ_2 is the single vertex 01 and hence connected. Assume $w = w_1 \ldots w_n$ $(n \geq 3)$ is an arbitrary word of Λ_n.

Let i be the largest integer such that $w_i = 0$. Every word of Λ_n has at least one 0 since $1^n = 1 \ldots 1$ is not in Λ_n. Define

$$w' = w_1 \ldots w_{i-1} 1 w_{i+1} \ldots w_n.$$

Clearly, $d(w, w') = 1$. If $w = 01^{n-1}$ then $w' = 1^n \notin \Lambda_n$. Otherwise, $w' \in \Lambda_n$, since w' is strictly less in the lexicographical order than each of its proper right factors whenever w is except for $w = 01^{n-1}$. Now define $w'' = (w')'$. Clearly, $d(w', w'') = 1$ and $w'' \in \Lambda_n$ unless $w = 01^{n-1}$. Continuing in this way, we obtain a path containing not more than $n - 2$ vertices between w and 01^{n-1}, each vertex of which is in Λ_n. Now, if v is any other word of Λ_n distinct from w we find, in the same way, a path from v to 01^{n-1} of length at most $n - 2$. Therefore, there is a path all of whose vertices are in Λ_n of length at most $2n-4$ from v to w.

One reason for interest in Theorem 1 is that it opens the possibility of a listing for Λ_n in which there is only one bit changed between successive words.

Definition 4: A set of vertices $v_1, \ldots, v_k$ in the n-cube form a *Gray path* if there exists an ordering $v_{\sigma(1)}, \ldots, v_{\sigma(k)}$ of the vertices such that $d(v_{\sigma(i)}, v_{\sigma(i+1)}) = 1$, for $i = 1, \ldots, k-1$. A gray path is a *Gray cycle* if there is an ordering such that $d(v_{\sigma(k)}, v_{\sigma(1)}) = 1$. We observe that

$$\Lambda_3 : \begin{array}{c} 001 \\ 011 \end{array}$$

is a Gray cycle, while

$$\Lambda_4 : \begin{array}{c} 0001 \\ 0011 \\ 0111 \end{array}$$

is a Gray path but not a Gray cycle. However, Λ_5 is:

00001
00101
00111
01111
01011
00011.

Theorem 2. *Λ_6 can be written as a Gray path but not as a Gray cycle.*

Proof: In Λ_6 only $0^5 1$ has weight 1 and only 01^5 has weight 5. There are two words, $0^4 11$ and $0^3 101$, of weight 2 and two words $0^2 1^4, 0101^3$ of weight 4. Finally, $0^3 1^3, 0^2 10^2 1$, and $0^2 1^2 01$ have weight 3. Hence, if Λ_6 is a Gray path we can order the multiset $\{1,2,2,3,3,3,4,4,5\}$ in such a way that adjacent elements differ by 1. Indeed, the essentially unique ordering is 123454323. However, a straightforward argument shows that a circular ordering of this multiset is impossible, so Λ_6 is not a Gray cycle. The following is a listing of Λ_6 as a Gray path:

000001
000011
001011
001111
011111
010111
000111
000101
001101.

We remark in conclusion that our programs have verified that Λ_n is a Gray path for $n = 7, 8$ as well. We suspect that Λ_3 and Λ_5 are the only Gray cycles and for $n \geq 5$, Λ_n is a Gray path but not a Gray cycle.

References

1. L.J. Cummings, *Aspects of synchronizable coding*, The Journal of Combinatorial Mathematics and Combinatorial Computing 1 (1987), 67-84.
2. S.W. Golomb, *Codes with bounded synchronization delay*, Information and Control 8 (1965), 355-372.
3. M. Lothaire, "Combinatorics on Words," Addison-Wesley, Reading, Massachusetts, 1983.