Gray Codes for A-Free Strings

Matthew B. Squire*
IBM Corporation
Dept NFVA/Bldg 664
P.O. Box 12195
Research Triangle Park, NC 27709
Email: msquire@vnet.ibm.com
Submitted: July 10, 1995; Accepted February 14, 1996.

Abstract

For any $q \geq 2$, let $\Sigma_q = \{0, \ldots, q-1\}$, and fix a string A over Σ_q. The *A-free strings of length n* are the strings in Σ_q^n which do not contain A as a contiguous substring. In this paper, we investigate the possibility of listing the A-free strings of length n so that successive strings differ in only one position, and by ± 1 in that position. Such a listing is a Gray code for the A-free strings of length n.

We identify those q and A such that, for infinitely many $n \geq 0$, a Gray code for the A-free strings of length n is prohibited by a parity problem. Our parity argument uses techniques similar to those of Guibas and Odlyzko (*Journal of Combinatorial Theory A* 30 (1981) pp. 183–208) who enumerated the A-free strings of length n. When q is even, we also give the complementary positive result: for those A for which an infinite number of parity problems do not exist, we construct a Gray code for the A-free strings of length n for all $n \geq 0$.
Mathematical Review Subject Numbers: 68R15, 05A15.

1 Introduction

Fix $q \geq 2$ and let $\Sigma_q = \{0, \ldots, q-1\}$. When we refer to strings, we mean strings over Σ_q. For any strings A and B, an *A-factor* of B is a contiguous substring of B equal to A. Fix a string A, and let $\mathcal{F}(n) = \mathcal{F}(n; A)$ be the strings of length n which do not have an A-factor. The set $\mathcal{F}(n)$ consists of the A-free strings of length n.

We are interested in *Gray codes* for $\mathcal{F}(n)$. A Gray code for a set is a listing of its elements so that successive elements are "similar." Define two strings in Σ_q^n to be

*Research supported by National Science Foundation Grant No. DMS9302505.

similar if they differ in exactly one position, and by ± 1 in that position. When q is even, the addition is done modulo q, but the addition is *not* modular for odd q.

Given $S \subseteq \Sigma_q^n$, we can form a graph whose vertices are the strings in S, and which has edges between all pairs of similar strings. Let $G_q(S)$ denote this graph. Then $G_q(S)$ is a bipartite subgraph of the q-ary n-cube. A Gray code for S is equivalent to a Hamilton path of $G_q(S)$. For a bipartite graph with bipartition (X, Y) to have a Hamilton path, X and Y must have cardinalities which differ by at most one. Call a string $B = b_0 \dots b_{n-1}$ *even* if $\sum b_i$ is even, and *odd* if $\sum b_i$ is odd. If S^+ and S^- denote the even and odd strings in S, respectively, then (S^+, S^-) is a bipartition of $G_q(S)$. Thus, a Gray code for $\mathcal{F}(n)$ is only possible if the numbers of even and odd A-free strings differ by at most one.

In this paper, we investigate the existence of Gray codes for A-free strings. More specifically, we characterize the q and A for which a Gray code for $\mathcal{F}(n)$ *cannot* exist for *infinitely many* n due to problems with the cardinalities of the even and odd strings. When the number of even and odd strings in $\mathcal{F}(n)$ differ by more than one, we say that *$\mathcal{F}(n)$ has a parity problem.* We say that *$\mathcal{F}(n)$ has infinite parity problems* to indicate that $\mathcal{F}(n)$ has a parity problem for infinitely many n.

The set $\mathcal{F}(n)$ can be defined more generally. Let $\mathcal{A}$ be any finite set of strings. The $\mathcal{A}$-free strings of length n, $\mathcal{F}(n; \mathcal{A})$, are the strings of length n which do not contain any $A \in \mathcal{A}$ as a substring. The previously mentioned definition of $\mathcal{F}(n)$ restricted $\mathcal{A}$ to be a singleton set. Let $f(n) = |\mathcal{F}(n; \mathcal{A})|$, and let

$$F(z) = \sum_{n \geq 0} f(n) z^{-n}$$

be its generating function. Many authors [7, 6, 12, 3, 8, 10] have shown $F(z)$ to be a rational generating function of the form

$$F(z) = \frac{P(z)}{Q(z)},$$

where $P(z)$ and $Q(z)$ are polynomials in z. Guibas and Odlyzko [6] provide efficient means of determining the polynomials $P(z)$ and $Q(z)$ for arbitrary $\mathcal{A}$.

When $\mathcal{A}$ is a singleton set, Guibas and Odlyzko [6] present a beautiful and concise formula for $F(z)$ which depends only upon how the string A "overlaps" itself. The authors use the *correlation* of a string to measure self-overlaps. Gardner [1] attributes the notion of correlation to Conway. Let $A = a_0 \dots a_{m-1}$ and $B = b_0 \dots b_{n-1}$. The *correlation of A over B*, denoted AB, is the binary string of length m where bit i, $0 \leq i < m$, is defined as follows. Place B under A with b_0 under a_i. If the overlapping parts of the string match, then bit i of AB is one, otherwise it is zero. Refer to Figure 1 for an example. The *autocorrelation of A* is AA.

It is often helpful to view AB as the coefficient sequence of a polynomial. If $AB = c_0 \dots c_{m-1}$, let $AB_z = \sum_{i=0}^{m-1} c_i z^{m-1-i}$. See Figure 1 for an example. Guibas and Odlyzko [6] show that when $\mathcal{A} = \{A\}$,

$$F(z) = \frac{zAA_z}{(z-q)AA_z + 1}. \tag{1}$$

$A = 1001010 \qquad B = 101011$

AB	=	0	0	0	1	0	1	0	$AB^+ =$	0	0	0	1	0	0	0
AB_z	=				z^3		$+z$		$AB^- =$	0	0	0	0	0	1	0
									$AB^* =$	0	0	0	1	0	-1	0
									$AB^*_z =$				z^3		-z	

Figure 1: Correlations and correlation polynomials

We are more concerned with determining the parity difference for $\mathcal{F}(n)$ rather than the cardinality of $\mathcal{F}(n)$. The *parity difference* of $\mathcal{F}(n)$ is the difference between the number of even and odd strings in $\mathcal{F}(n)$. When the parity difference is greater than one, no Gray code for $\mathcal{F}(n)$ can exist. To this end, we define the *positive* correlation of A over B, denoted AB^+, in an analogous manner to the correlation of A over B. For bit i of AB^+ to be one, bit i of AB must be one *and* $a_i \ldots a_{m-1}$ must be even. The *negative* correlation of A over B, AB^-, is defined similarly except that $a_i \ldots a_{m-1}$ must be odd. Define the *parity correlation of A over B*, AB^*, to be the string in $\{0, \pm 1\}^m$ equal to $AB^+ - AB^-$, where the subtraction is performed bitwise. Refer to Figure 1 for an example. The parity correlation determines the parity difference for $\mathcal{F}(n)$ in much the same way as the correlation determines the cardinality of $\mathcal{F}(n)$. Let AB^*_z (AB^+_z, AB^-_z) denote the polynomial in z whose coefficient sequence is given by AB^* (AB^+, AB^-).

In Section 2, we develop recursive definitions and generating functions for the parity difference of $\mathcal{F}(n)$. In Section 3, the generating functions are used to characterize those q and A for which $\mathcal{F}(n)$ has infinite parity problems. For these q and A, we know that Gray codes for $\mathcal{F}(n)$ often cannot exist. For even q and for those A which do not have infinite parity problems, Gray codes for $\mathcal{F}(n)$ are constructed for all n. This is done in Section 4. When q is odd and $\mathcal{F}(n)$ has no parity problems, we leave it open to discover Gray codes for $\mathcal{F}(n)$. In Section 5, we summarize the results of the paper and indicate several related questions. Throughout this paper, $A = a_0 \ldots a_{m-1}$ represents the chosen string for $\mathcal{F}(n) = \mathcal{F}(n; A)$, and m denotes the length of A. We assume $m > 0$ because every string contains an ϵ-factor, where ϵ denotes the null string.

2 Generating Functions and Recurrences

As mentioned in Section 1, the generating function $F(z)$ has been well studied. In this section, a similar argument is used to obtain the generating function for the parity difference of $\mathcal{F}(n)$.

Let $\mathcal{F}^+(n)$ and $\mathcal{F}^-(n)$ denote the even and odd strings of $\mathcal{F}(n)$, and let $f^+(n)$ and $f^-(n)$ denote the cardinalities of these sets. Denote the parity difference of $\mathcal{F}(n)$ by

$f^*(n) = f^+(n) - f^-(n)$. We wish to determine the generating function

$$F^*(z) = \sum_{n\geq 0} f^*(n) z^{-n}.$$

To do so, we develop two identities for $F^*(z)$. The two identities are then combined to solve for $F^*(z)$. Our approach follows that of Guibas and Odlyzko [6].

Before developing the first identity, we introduce several more concepts. For $i \in \Sigma_q$, let $\mathcal{F}_i(n)$ denote the strings in $\mathcal{F}(n)$ which end with i. Also, let $\mathcal{G}_A(n)$ be the strings of length n which contain *exactly one* A-factor, and the A-factor is a suffix of the string. Our notation for cardinalities and even and odd strings is extended to the sets $\mathcal{F}_i(n)$ and $\mathcal{G}_A(n)$. So, for example, $\mathcal{F}_i^+(n)$ are the even strings in $\mathcal{F}_i(n)$, and $g_A^*(n)$ is the parity difference of $\mathcal{G}_A(n)$. Let $G_A^*(z) = \sum_{n\geq 0} g_A^*(n) z^{-n}$ be the generating function for $g_A^*(n)$. For any strings X and Y, $X * Y$ denotes the concatenation of X and Y.

We now develop the first identity for $F^*(z)$. Assume that a_{m-1} is even, and suppose we are given $n \geq 0$ and $B \in \mathcal{F}^+(n)$. Then $C = B * a_{m-1}$ is also an even string, and C either ends with A or it does not. Since this is true for all $B \in \mathcal{F}^+(n)$, $f^+(n) = g_A^+(n+1) + f^+_{a_{m-1}}(n+1)$. Every string in $\mathcal{F}(n+1)$ must terminate with some character, so $\mathcal{F}^+(n+1) = \cup_{i=0}^{q-1} \mathcal{F}_i^+(n+1)$. Let $S = \Sigma_q \setminus \{a_{m-1}\}$, and let S^+ and S^- denote the subsets of S consisting of the even and odd characters, respectively. For all $i \in S$, $B \in \mathcal{F}(n)$ if and only if $B * i \in \mathcal{F}_i(n+1)$. Therefore,

$$\begin{aligned}
f^+(n) &= g_A^+(n+1) + f^+_{a_{m-1}}(n+1) \\
&= g_A^+(n+1) + f^+(n+1) - \sum_{i\in S} f_i^+(n+1) \\
&= g_A^+(n+1) + f^+(n+1) - \sum_{i\in S^+} f^+(n) - \sum_{i\in S^-} f^-(n) \\
&= g_A^+(n+1) + f^+(n+1) - |S^+| f^+(n) - |S^-| f^-(n). \qquad (2)
\end{aligned}$$

We similarly arrive at

$$f^-(n) = g_A^-(n+1) + f^-(n+1) - |S^+| f^-(n) - |S^-| f^+(n). \qquad (3)$$

Subtracting eq. (3) from eq. (2) yields

$$(|S^+| - |S^-| + 1) f^*(n) = g_A^*(n+1) + f^*(n+1). \qquad (4)$$

If A terminates with an odd character, the corresponding equation is

$$(|S^+| - |S^-| - 1) f^*(n) = g_A^*(n+1) + f^*(n+1). \qquad (5)$$

We now show that eqs. (4) and (5) depend only on the parity of q, not on the actual value of q or on a_{m-1}. Suppose $q = 2k$ and a_{m-1} is even. Then $|S^+| = k-1$, $|S^-| = k$, and the coefficient of $f^*(n)$ in eq. (4) collapses to zero. If $q = 2k$ and a_{m-1} is odd, then the coefficient of $f^*(n)$ in eq. (5) is again zero. When q is even, $f^*(n) = -g_A^*(n)$ for $n > 0$. By multiplying both sides by z^{-n}, summing over $n \geq 0$, and using $f^*(0) = 1$ and $g_A^*(0) = 0$, we arrive at

$$F^*(z) = -G_A^*(z) + 1 \qquad \text{when } q \text{ is even.} \qquad (6)$$

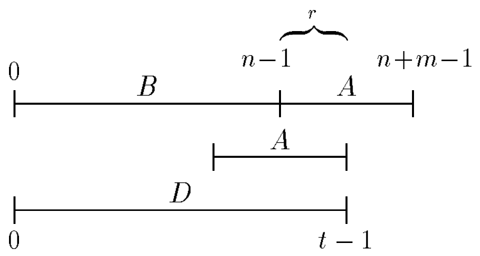


Figure 2: Overlaps for the second identity

When q is odd, the coefficient of $f^*(n)$ in both eqs. (4) and (5) collapses to one. For odd q, $f^*(n) = g_A^*(n+1) + f^*(n+1)$ for all $n \geq 0$. Following the same procedure as before, we get

$$(z-1)F^*(z) = -zG_A^*(z) + z \qquad \text{when } q \text{ is odd.} \tag{7}$$

Equations (6) and (7) constitute the first identity for $F^*(z)$.

We now develop the second relationship between $F^*(z)$ and $G_A^*(z)$. Fix $n \geq 0$ and let $B \in \mathcal{F}(n)$. Then $B * A = C = c_0 \dots c_{n+m-1}$ certainly contains an A-factor. Let t be such that $c_{t-m} \dots c_{t-1}$ is the leftmost A-factor in C. Refer to Figure 2 for an illustration. Then $t > n$ because B does not have an A-factor, and $D = c_0 \dots c_{t-1} \in \mathcal{G}_A(t)$. Let $r = t - n \geq 1$. Then $c_{t-r} \dots c_{t-1} = a_0 \dots a_{r-1}$ because these characters begin the A in $B * A$. However, $c_{t-r} \dots c_{t-1} = a_{m-r} \dots a_{m-1}$ because D ends in A. Therefore, $a_{m-r} \dots a_{m-1} = a_0 \dots a_{r-1}$, and bit $(m-r)$ of AA is one. For $1 \leq i \leq m$, let $i \in AA$ denote that bit $(m-i)$ of AA is 1 (similarly for AA^+ and AA^-). If $a_{m-r} \dots a_{m-1}$ has even parity ($r \in AA^+$), then the parity of D is equal to the parity of B. If $a_{m-r} \dots a_{m-1}$ has odd parity ($r \in AA^-$), then the parities of B and D are different. In this way, each $B \in \mathcal{F}^+(n)$ maps to exactly one D in $(\cup_{r\in AA^+}\mathcal{G}_A^+(n+r)) \cup (\cup_{r\in AA^-}\mathcal{G}_A^-(n+r))$. Conversely, given a D in $(\cup_{r\in AA^+}\mathcal{G}_A^+(n+r)) \cup (\cup_{r\in AA^-}\mathcal{G}_A^-(n+r))$, removing the last r characters gives a unique B in $\mathcal{F}^+(n)$. Therefore,

$$\begin{aligned} f^+(n) &= \sum_{r\in AA^+} g_A^+(n+r) + \sum_{r\in AA^-} g_A^-(n+r) \\ f^-(n) &= \sum_{r\in AA^+} g_A^-(n+r) + \sum_{r\in AA^-} g_A^+(n+r) \end{aligned}$$

Subtracting the latter equation from the former yields

$$f^*(n) = \sum_{r\in AA^+} g_A^*(n+r) - \sum_{r\in AA^-} g_A^*(n+r).$$

Observe that $g_A(n) = 0$ for $n < m$ because m characters are required for an A-factor. Isolating on the first summation above, we see that

$$\sum_{n\geq 0}\sum_{r\in AA^+} g_A^*(n+r)z^{-n} = \sum_{r\in AA^+} z^r \sum_{n\geq 0} g_A^*(n+r) z^{-(n+r)} = \sum_{r\in AA^+} z^r G_A^*(z).$$

The last term is just $zAA_z^+G_A^*(z)$, so

$$F^*(z) = zAA_z^+G_A^*(z) - zAA_z^-G_A^*(z) = zAA_z^*G_A^*(z). \tag{8}$$

Eq. (8) can be combined with eqs. (6) and (7) to obtain

$$F^*(z) = \frac{zAA_z^*}{zAA_z^*+1} \quad \text{when } q \text{ is even}, \tag{9}$$

$$F^*(z) = \frac{zAA_z}{(z-1)AA_z^*+1} \quad \text{when } q \text{ is odd}. \tag{10}$$

Since $F^*(z)$ is rational with a degree m polynomial in the denominator, $f^*(n)$ satisfies a degree m recurrence [8]. This recurrence can be used to efficiently calculate $f^*(n)$. In the following section, $F^*(z)$ is used to identify those q and A for which the set $\mathcal{F}(n)$ has infinite parity problems. Constructive results are presented in Section 4.

3 Parity problems

In this section, we examine the generating function $F^*(z)$ and determine for which q and which A a Gray code for $\mathcal{F}(n)$ is not always possible. For each q, we characterize all A for which $\mathcal{F}(n)$ has infinite parity problems.

Let $H(z) = \sum a_n z^{-n}$. It is well known that the rate of growth of the a_n is highly dependent upon the modulus of the greatest singularity of $H(z)$ (see, for example, Wilf [11]).

Theorem 3.1 *Let $H(z) = \sum_{n\geq 0} a_n z^{-n}$ be analytic on the disk $|z| > R$, let a singularity of $H(z)$ of largest modulus be at a point α_0, and let $\epsilon > 0$ be given. For infinitely many n we have*

$$|a_n| > (|\alpha_0| - \epsilon)^n .$$

In particular, if $H(z)$ has a singularity of modulus strictly greater than one, then the coefficients of $H(z)$ are unbounded in absolute value.

The singularities of $F^*(z)$ are the roots of the polynomial in the denominator of $F^*(z)$. This polynomial is either zAA_z^*+1 (when q is even) or $(z-1)AA_z^*+1$ (when q is odd). If the polynomial in question has a root of modulus greater than one, then Theorem 3.1 implies $\mathcal{F}(n)$ has infinite parity problems.

Let $p(z)$ be a polynomial with integer coefficients and a unit leading coefficient. Suppose that all of the roots of $p(z)$ are contained in the closed unit disk of the complex plane. This implies (see [9, Problem 200, page 145]) that the non-zero roots of $p(z)$ are roots of unity. The n^{th} *cyclotomic polynomial*, $\phi_n(z)$, can be defined as $\phi_n(z) = \prod(z-\theta)$, where the product is taken over all primitive n^{th} roots of unity. Therefore, $p(z)$ can be expressed as a monomial z^k times a product of cyclotomic polynomials. A polynomial $q(z)$ is *self-reciprocal* if α being a non-zero root of $q(z)$ implies $1/\alpha$ is also a root. Cyclotomic polynomials are self-reciprocal, as are products of self-reciprocal polynomials. Thus, if $p(z)$ has all roots in the closed unit disk, then $p(z)$ is self-reciprocal. It can be easily checked that if $q(z) = \sum_{i=0}^{m} q_i z^{m-i}$ is self-reciprocal then $(q_0, \ldots, q_m) = \pm(q_m, \ldots, q_0)$.

Let $(p_0, \ldots, p_m)$ be the coeffients of the polynomial in the denominator of $F^*(z)$. To prove that $\mathcal{F}(n)$ has infinite parity problems, we need only show that $(p_0, \ldots, p_m) \neq \pm(p_m, \ldots, p_0)$. This is the approach taken in Sections 3.1 and 3.2.

3.1 Parity problems when q is even

Throughout Section 3.1, q is assumed even. The polynomial in the denominator of $F^*(z)$ is then $p(z) = p(z; A) = zAA^*_z + 1$. This polynomial either has a root outside the closed unit disk, or it is self-reciprocal. Given any string A, one can determine whether $p(z)$ is self-reciprocal by simply examining its coefficients.

We wish to characterize those A for which the roots of $p(z)$ all lie in the closed unit disk, with the characterization based directly on A and not $p(z)$. A *period* of A is an integer p such that $a_i = a_{i+p}$ for all $0 \leq i < m - p$. The periods of A are the positions of the ones in AA. With each non-zero period p, there is a corresponding prefix $a_0 \ldots a_{p-1}$ of A. Call B a *base of* A if $B = a_0 \ldots a_{p-1}$ for some period $p > 0$ of A. For any base B of A, $A = B * \cdots * B * C$, where C is a (possibly empty) prefix of B. The *minimal period* of A is the smallest non-zero period of A. If A has no non-zero period, then its minimal period is defined to be m. The *minimal base* of A is the base corresponding to the minimal period.

Let p be the minimal period of A and let B be its minimal base. The string A is *periodic* if $p < m$ and p divides m. In this case, $A = B * \cdots * B$. Note that the minimal base of any string cannot itself be periodic. If A is not periodic, then it is *aperiodic*.

Not every binary string is an autocorrelation. For example, 000 is not the autocorrelation of any string. Binary strings which are autocorrelations are known to satisfy many properties, including the *Forward Propagation Rule* [5]:

Definition 3.1 *A length m bit vector satisfies the* Forward Propagation Rule *if bits i and $i + k$ being one ($k \geq 0$) implies bit $i + jk$ is one for all $i \leq i + jk < m$.*

A corollary of the Forward Propagation Rule is that if $A = B * \cdots * B * C$, where B is the minimal base of A and C is a non-empty (and possibly non-proper) prefix of B, then $AA = 10^{p-1} * \cdots * 10^{p-1} * CC$. The string $10 \cdots 0$ is the *trivial autocorrelation*, and if $AA = 10 \cdots 0$, then A is *trivially autocorrelated*, or t.a.

Suppose that $p(z) = zAA^*_z + 1$ has all of its roots in the closed unit disk, so $p(z)$ is self-reciprocal. If $X = AA * 1$, then it must be that $X = X^R$, where X^R denotes the reverse of X. We claim that this can only occur if A is t.a., or if A is periodic with a t.a. minimal base. It is simple to verify that in either of these cases, $X = X^R$ and the roots of $p(z)$ are roots of unity.

Now suppose that A is not t.a., and that A is not periodic with a t.a. minimal base. For instance, suppose that A is periodic, but that its minimal base B is not t.a. Then bit $p - j$ of BB is one for some $0 < j < p$. In this case, bit $m - j$ of AA is also one, but bits 1 through $p - 1$ of AA are zero. This implies $X \neq X^R$, where $X = AA * 1$. Thus, $p(z)$ cannot be self-reciprocal, so it must have a root outside the closed unit disk. The case when A is aperiodic but not t.a. can be handled similarly. Therefore, $p(z)$ has all roots in the closed unit disk iff A is t.a., or if A is periodic with a t.a. minimal base. For all other A, $\mathcal{F}(n)$ has infinite parity problems.

We could prove directly that if A is t.a. or A is periodic with a t.a. minimal base, then $|f^*(n)| \leq 1$ for all n. In this case, the roots of $p(z)$ are all roots of unity, so $f^*(n)$ is a periodic function, ie $f^*(n) = f^*(n+r)$ for some $r > 0$. By showing that $|f^*(n)| \leq 1$ for $0 \leq n < r$, we could conclude that $|f^*(n)| \leq 1$ for all n. In Section 4, we actually construct Gray codes for $\mathcal{F}(n)$ for all n and any such A. This implies $\mathcal{F}(n)$ never has a parity problem. We therefore omit this latter part of the proof and simply state:

Theorem 3.2 *Fix an even $q \geq 2$, $m > 0$, and $A \in \Sigma_q^m$. If A is t.a. or A is periodic with a t.a. minimal base, then $\mathcal{F}(n)$ has no parity problems. Otherwise, $\mathcal{F}(n)$ has infinite parity problems.*

It may seem to the reader that t.a. strings are rare. In fact, they are not. Guibas and Odlyzko [5, 4] have studied the number of strings with a given autocorrelation. When $q = 2$, about 27% of all strings [5] are t.a. For $q \geq 3$, the majority of strings are t.a., and as q gets larger, the *vast* majority of strings are t.a. For large q, the fraction of t.a. strings is about $(q-2)/(q-1)$.

In Theorem 3.2, we have characterized those A for which $\mathcal{F}(n)$ has infinite parity problems when q is even. The odd q case is examined in the next section.

3.2 Parity problems when q is odd

When q is odd, the polynomial in the denominator of $F^*(z)$ is $p(z) = p(z; A) = (z-1)AA_z^* + 1$. If $p(z)$ is not self-reciprocal, then $\mathcal{F}(n)$ has infinite parity problems.

We first show that, for certain polynomials $c(z)$, if $(z-1)c(z)+1$ is self-reciprocal, then $c(z)$ has two unit coefficients followed by a zero coefficient. If we interpret $c(z)$ as AA_z^*, then for certain strings A, if $p(z)$ is self-reciprocal, then AA contains the substring 110.

Lemma 3.1 *Let $c(z) = \sum_{i=0}^{m-1} c_i z^{m-1-i}$ satisfy $c_0 \neq 0$ and $c_1 = c_2 = c_{m-1} = 0$. If $(z-1)c(z)+1$ is self-reciprocal, then $c_{m-2} = c_{m-3} = 1$.*

Proof. Let $p(z) = \sum_{i=0}^{m} p_i z^{m-i} = (z-1)c(z) + 1$ be self-reciprocal. Then $p_0 = c_0$, $p_m = 1 - c_{m-1} = 1$, and for i, $0 < i < m$, $p_i = c_i - c_{i-1}$. If $p_i = p_{m-i}$ for $0 \leq i \leq m$, then

$$\begin{aligned} c_0 = p_0 &= p_m = 1 \\ -c_0 = c_1 - c_0 = p_1 &= p_{m-1} = c_{m-1} - c_{m-2} = -c_{m-2} \\ 0 = c_2 - c_1 = p_2 &= p_{m-2} = c_{m-2} - c_{m-3} \end{aligned}$$

Therefore, $c_{m-2} = c_{m-3} = 1$. The same result follows when $p_i = -p_{m-i}$ for all i. $\square$

Suppose $c(z) = AA_z^*$ in Lemma 3.1. If the conditions are satisfied and $p(z)$ is self-reciprocal, then last three positions of AA^* are 110. Applying the Forward Propagation Rule, it is easy to see that no autocorrelation can contain the substring 110. Thus, if AA is of the form $100X0$ for some binary string X, then $p(z)$ cannot be self-reciprocal.

We now examine other possible forms of AA and determine when $p(z)$ is self-reciprocal. If $AA = X1$ and a_{m-1} is odd, then $p(z)$ has constant term two, and is thus not self-reciprocal. Let p be the minimal period of A and let B be its minimal base. The rest of the analysis is divided into three cases.

Case 1: When $p > 2$. In this case, $AA = 100X$. We have already discussed the possibility that $AA = 100Y0$, and the possibility that $AA = 100Y1$ with a_{m-1} odd. We are left to consider when $AA = 100Y1$ and a_{m-1} is even. Suppose a_{m-1} is even and AA ends in 01^k, $k > 0$. Since $a_0 = a_{m-k} = \cdots = a_{m-1}$ is even, $AA^*_z = z^k c(z) + z^{k-1} + \ldots + 1$, where $c(z)$ is the degree $(m-k-1)$ polynomial whose coefficients are given by the first $m-k$ positions of AA^*. The constant term of $c(z)$ is zero. In this case,

$$p(z) = (z-1)AA^*_z + 1 = z^k((z-1)c(z) + 1).$$

If $p'(z) = (z-1)c(z) + 1$ is self-reciprocal, then the lower three coefficients of $c(z)$ are $1, 1, 0$ by Lemma 3.1. This implies bits $(m-k-3)$ through $(m-k-1)$ of AA are 110, again leading to a contradiction. Thus $p'(z)$, and $p(z)$, are not self-reciprocal.

Case 2: When $p = 2$. We separately examine when m is even and m is odd. First suppose m is even. Then $AA = (10)^{m/2}$ and $A = B * \cdots * B$ for an aperiodic string B of length two. Say B is an even string, so $p(z) = z^m - z^{m-1} + z^{m-2} \cdots - z + 1$. In this particular subcase, the roots of $p(z)$ are roots of unity ($p(z)$ divides $z^{m+1} + 1$). If B is an odd string and $m = 2$, then $p(z) = -z^2 + z + 1$. Here, the parity differences are the Fibonacci numbers (doesn't the generating function look familiar?). Finally, consider when B is odd and $m \geq 4$. The smallest term of AA^*_z is $-z$, so $p(z) = \pm z^m \mp z^{m-1} \cdots + z + 1$, where the signs of z^m and z^{m-1} are different, but the signs of z and 1 are both positive. This $p(z)$ is not self-reciprocal. Thus, when $p = 2$ and m is even, $\mathcal{F}(n)$ has infinite parity problems if the minimal base is an odd string. Otherwise, the roots of $p(z)$ are roots of unity.

If m is odd, a similar analysis yields that $\mathcal{F}(n)$ has infinite parity problems if the minimal base does not consist of two even characters. If B does consist of two even characters, then the roots of $p(z)$ all lie in the closed unit disk.

Case 3: When $p = 1$. We have already considered the possibility that $AA = X1$ and a_{m-1} is odd. We therefore need only consider when a_{m-1} is even. In this case, $p(z) = z^m$, and $F^*(z)$ is a simple polynomial with coefficients in $\{0, \pm 1\}$. When $p = 1$ and $a_0 = a_{m-1}$ is odd, $\mathcal{F}(n)$ has infinite parity problems. When a_0 is even, $\mathcal{F}(n)$ does not have a parity problem for any n.

As mentioned in Section 3.2, if $p(z)$ has all roots in the closed unit disk, then $f^*(n)$ is a periodic function. By showing that $|f^*(n)| \leq 1$ for $0 \leq n < r$, where r is the period of the function, we could prove that when the roots of $p(z)$ are all in the closed unit disk, $\mathcal{F}(n)$ has no parity problems. As in Section 3.1, we omit this part of the proof, and state the following theorem.

Theorem 3.3 *Fix an odd $q \geq 3$, $m > 0$, and $A \in \Sigma_q^m$ with minimal period p. If*

- *$p > 3$, or*
- *$p = 2$, m is even, and $a_0 a_1$ is odd, or*
- *$p = 2$, m is odd, and either a_0 or a_1 is odd, or*
- *$p = 1$ and a_0 is odd,*

then $\mathcal{F}(n)$ has infinite parity problems. Otherwise, $\mathcal{F}(n)$ has no parity problems.

In Section 3, we have examined the generating function $F^*(z)$ and identified the q and A for which the coefficients of $F^*(z)$ are unbounded. In Section 4, Gray codes are constructed for $\mathcal{F}(n)$ when q is even and A is t.a. or A is periodic with a t.a. minimal base. These are the only A for which $\mathcal{F}(n)$ has no parity problems when q is even. For odd q, it remains open to find Gray codes for $\mathcal{F}(n)$ when $\mathcal{F}(n)$ has no parity problems.

4 Gray Codes for A-Free Strings when q is even

Throughout Section 4, q is assumed even. The goal of this section is to construct Gray codes for $\mathcal{F}(n)$ for all $n \geq 0$ when A is t.a. or A is periodic with a t.a. minimal base. For all other A, Theorem 3.2 showed that $\mathcal{F}(n)$ has infinite parity problems. We begin by presenting the recursive identity on which our constructions are based.

Let C be a string over Σ_q. Define $\mathcal{H}(n; C)$ to be the $D \in \Sigma_q^n$ such that $C * D$ has no A-factor. It is clear that for any $k \leq n$,

$$\mathcal{H}(n; C) \quad = \bigcup_{D \in \mathcal{H}(n-k; C)} D * \mathcal{H}(k; C * D). \tag{11}$$

For certain A, we construct Gray codes for $\mathcal{H}(n; C)$ for all n and C. Letting ϵ denote the empty string, $\mathcal{F}(n) = \mathcal{H}(n; \epsilon)$. Thus, for certain A, we get Gray codes for $\mathcal{F}(n)$.

Our method of constructing Gray codes is recursive. In Section 4.1, we describe the construction methods and present the main recursive step. Sections 4.2 through 4.5 supply the construction of the basis. The basis construction is also recursive.

4.1 Construction Overview

In this section, we overview our construction methodology. The main inductive step is given in the following lemma. Let $h^*(n; C)$ be the parity difference of $\mathcal{H}(n; C)$. Bold letters ($\mathbf{L}$) represent sequences of strings. If $\mathbf{L}$ is a list of strings, then $\mathbf{L}^R$ denotes the reverse of $\mathbf{L}$.

Lemma 4.1 *Fix an even $q \geq 2$, $m > 0$, and $A \in \Sigma_q^m$. Suppose there exists $k > 0$ such that both of the following conditions are true.*

1. *For any n, $1 \leq n \leq k$, and for any C with no A-factor, there is a Gray code for $\mathcal{H}(n; C)$. When $h^*(n; C) = 0$, there is a cyclic Gray code.*
2. *There exist strings $S, T \in \Sigma_q^k$, of equal parity, such that for any D with no A-factor, there is a Gray code for $\mathcal{H}(k; D)$ from S to T.*

Then, for all $n \geq 1$ and for all C with no A-factor, there is a Gray code for $\mathcal{H}(n; C)$. When $h^*(n; C) = 0$, there is a cyclic Gray code.

Proof. We argue by induction on n. Let C be any string with no A-factor. When $1 \leq n \leq k$, condition (1) of the lemma provides Gray codes for $\mathcal{H}(n; C)$. Assume $n > k$ and recall the recursive definition of eq. (11).

By induction, there exists a Gray code $\mathbf{L}' = D_0, \ldots, D_{N-1}$ for $\mathcal{H}(n-k;C)$. For $0 \le i < N$, let $\mathbf{L}_i$ be the Gray code for $\mathcal{H}(k; C * D_i)$ from S to T which exists by condition (2). A Gray code for $\mathcal{H}(n;C)$ is given by

$$\mathbf{L} = \quad D_0 * \mathbf{L}_0, \quad D_1 * \mathbf{L}_1^R, \quad D_2 * \mathbf{L}_2, \quad \ldots$$

Since S and T have equal parity, $|h^*(k;D)| = 1$ for all D. Therefore, $h^*(n;C) = 0$ iff $h^*(n-k;C) = 0$, and $\mathbf{L}$ is cyclic iff $\mathbf{L}'$ is cyclic. $\Box$

We are left with the difficult task of proving that the conditions of the lemma can be satisfied when A is t.a. and when A is periodic with a t.a. minimal base. Sections 4.2 through 4.5 show this to be true. If A is t.a., then the conditions are satisfied when $k = m$. If A is periodic with a t.a. minimal base, then the conditions are satisfied when $k = m + p$, where p is the minimal period of A.

We construct cyclic Gray codes for $\mathcal{H}(n;C)$ whenever $h^*(n;C) = 0$. It is therefore important to know when $h^*(n;C) = 0$ and when $|h^*(n;C)| = 1$. For $n \le m$, Lemma 4.2 characterizes these two cases based upon whether $\mathcal{H}(n;C)$ contains certain strings. For $i < j$, $a_j \ldots a_i$ is defined to be the empty string. Every $C \in \Sigma_q^*$ has an ϵ-suffix.

Lemma 4.2 *Fix an even $q \ge 2$, $m > 0$ and $A \in \Sigma_q^m$. For any n, $1 \le n \le m$, and for any $C \in \Sigma_q^*$, $|h^*(n;C)| = 1$ if and only if $a_{m-n} \ldots a_{m-2}x \in \mathcal{H}(n;C)$ for all $x \ne a_{m-1}$ and C ends with $a_0 \ldots a_{m-n-1}$. Otherwise, $h^*(n;C) = 0$.*

Proof. We attempt to pair the even and odd strings in $\mathcal{H}(n;C)$. If every string in $\mathcal{H}(n;C)$ *except one* can be paired with a string of opposite parity, then $|h^*(n;C)| = 1$. Arbitrarily pair the even and odd characters in Σ_q. For $x \in \Sigma_q$, the *mate* of x is the element with which x is paired. For $D = d_0 \ldots d_{n-1} \in \mathcal{H}(n;C)$, let $D' = d_0 \ldots d_{n-2}y$, where y is the mate of d_{n-1}. Then (D, D') is a perfect pairing of $\mathcal{H}(n;C)$ unless there exists a $D \in \mathcal{H}(n;C)$ with $C * D'$ having an A-suffix. In this case, C ends with $a_0 \ldots a_{m-n-1}$ and $D = a_{m-n} \ldots a_{m-2}z$, where z is the mate of a_{m-1}. Since $C * D$ does not have an A-factor, $C * a_{m-n} \ldots a_{m-2}x$ does not have an A-factor for any $x \ne a_{m-1}$, and D is the only unpaired string in $\mathcal{H}(n;C)$. $\Box$

Let q, m, and A be as in the lemma. Suppose that n, $1 < n \le m$, and $C \in \Sigma_q^*$ are such that $|h^*(n;C)| = 1$. By the preceding lemma, $C * a_{m-n} \ldots a_{m-1}$ has an A-suffix, but $C * a_{m-n} \ldots a_{m-2}x$ does not have an A-factor for $x \ne a_{m-1}$. Then $(C * a_{m-n}) * a_{m-n+1} \ldots a_{m-1}$ also has an A-suffix, and $(C * a_{m-n}) * a_{m-n+1} \ldots a_{m-2}x$ does not have an A-factor for $x \ne a_{m-1}$. Thus $|h^*(n-1; C * a_{m-n})| = 1$. For $x \ne a_{m-n}$, since $C * x$ does not end with a_{m-n}, $h^*(n-1; C * x) = 0$.

Now suppose n and C are such that $h^*(n;C) = 0$. Then either $a_{m-n} \ldots a_{m-2}x \in \mathcal{H}(n;C)$ for all $x \in \Sigma_q$, or $a_{m-n} \ldots a_{m-2}x \notin \mathcal{H}(n;C)$ for all $x \in \Sigma_q$. So $h^*(n-1; C * a_{m-n}) = 0$. For $x \ne a_{m-n}$, $C * x$ does not end with a_{m-n}, so $h^*(n-1; C * x) = 0$.

We thus get the following corollary of Lemma 4.2.

Corollary 4.1 *Fix an even $q \ge 2$, $m > 0$, and $A \in \Sigma_q^m$. For any n, $1 < n \le m$, and any $C \in \Sigma_q^*$, $h^*(n;C) = 0$ if and only if $h^*(n-1; C * x) = 0$ for all $x \in \Sigma_q$. Also, $|h^*(n;C)| = 1$ if and only if $|h^*(n-1; C * a_{m-n})| = 1$ and $h^*(n-1; C * x) = 0$ for all $x \in \Sigma_q \setminus \{a_{m-n}\}$.*

When $q = 2$ and $n \leq m$, Gray codes for $\mathcal{H}(n; C)$ are studied in Section 4.2. When $q \geq 4$ is even and $n \leq m$, Gray codes are examined in Section 4.3. Gray codes when $n \leq m$ do not require A to be t.a. or to be periodic with a t.a. minimal base. By Lemma 4.2, $\mathcal{F}(n)$ cannot have a parity problem when $n \leq m$. Only as n gets larger can parity problems arise. Gray codes for arbitrary n are discussed in Sections 4.4 and 4.5. The former section considers when A is t.a., while the latter section considers when A is periodic with a t.a. minimal base.

4.2 Gray codes when $q = 2$ and $n \leq m$

In this section we construct Gray codes for $\mathcal{H}(n; C)$ when $q = 2$ and $n \leq m$. Throughout Section 4.2, it is assumed that $q = 2$. *The constructions of this section also assume that A ends with two identical characters.* In Sections 4.4 and 4.5, we show how to remove this restriction. For $x \in \Sigma_q$, let $\overline{x} = x + 1 \bmod q$.

The Gray code constructions of this section are based on eq. (11) when $k = n - 1$. For any C, define $r(C) = r(C; A)$ to be the minimum length of any *non-empty* string D such that $C * D$ ends in A-factor. Such a D must be of the form $a_{m-r} \dots a_{m-1}$ for some $0 < r \leq m$. When $k = n - 1$ and C does not have an A-factor, eq. (11) becomes

$$\begin{aligned} \mathcal{H}(n; C) &= \overline{a}_{m-1} * \mathcal{H}(n-1; C * \overline{a}_{m-1}) && \text{when } r(C) = 1, && (12) \\ &= 0 * \mathcal{H}(n-1; C * 0) \ \cup \ 1 * \mathcal{H}(n-1; C * 1) && \text{otherwise.} && (13) \end{aligned}$$

We build cyclic Gray codes for $\mathcal{H}(n; C)$ when $h^*(n; C) = 0$, and Gray codes between fixed endpoints when $|h^*(n; C)| = 1$. Given $h^*(n; C)$, Corollary 4.1 yields $h^*(n-1; C*0)$ and $h^*(n-1; C*1)$. If C does not have an A-factor, then $\mathcal{H}(n; C) = \Sigma_q^n$ when $r(C) > n$ because $r(C)$ characters are needed to create an A-factor. We say that $\mathbf{L}$ *contains the path* $[X, \dots, Z]$ if $X, \dots, Z$ appears as a contiguous subsequence of $\mathbf{L}$ or $\mathbf{L}^R$.

Theorem 4.1 *Fix $q = 2$, $m > 1$ and $A \in \Sigma_2^m$ such that A ends with two identical characters. Let n be in the range $2 \leq n \leq m$. Then there exist strings S_n, T_n, X_n, Y_n, Z_n, and Z'_n such that for any $C \in \Sigma_2^*$ with no A-factor,*

1. *If $r(C) = 1$ and $n \geq 3$, then there exists a Gray code for $\mathcal{H}(n; C)$ containing the path $[X_n, Y_n, Z_n]$. When $h^*(n; C) = 0$, the Gray code is cyclic. Otherwise, $|h^*(n; C)| = 1$, and the Gray code starts at S_n and ends at T_n.*
2. *If $r(C) > 1$ and $n \geq 2$, then there exists a Gray code for $\mathcal{H}(n; C)$ containing the path $[X_n, Y_n, Z'_n]$. When $h^*(n; C) = 0$, the Gray code is cyclic. Otherwise, $|h^*(n; C)| = 1$, and the Gray code starts at S_n and ends at T_n.*

For $2 \leq n \leq m$, S_n and T_n have parity opposite that of $A_n = a_{m-n} \dots a_{m-1}$, and $A_n \in \mathcal{H}(n; C)$ implies $S_n, T_n \in \mathcal{H}(n; C)$.

Proof. Assume that A ends in 00; similar techniques apply if A ends in 11. We prove Parts 1 and 2 by induction on n using the recursive decomposition of eqs. (12) and (13), and using the strings

$$\begin{aligned} S_n &= a_{m-n} \dots \overline{a}_{m-1} & \qquad X_n &= 101^{n-2} \\ T_n &= a_{m-n} \dots \overline{a}_{m-2} a_{m-1} & Y_n &= 1^n \\ Z'_n &= 01^{n-1} & Z_n &= 1101^{n-3} \end{aligned}$$

Let $C_0 = C * 0$ and $C_1 = C * 1$.

Basis. First consider when $n = 2$ and $r(C) > 1$. If $r(C) > 2$, then $\mathcal{H}(2; C) = \Sigma_2^2$. If $r(C) = 2$ then $\mathcal{H}(2; C) = \{01, 11, 10\}$. The Gray codes in these cases are $01, 11, 10, 00$ and $01, 11, 10$, respectively. These Gray codes satisfy Part 2.

Next consider when $n = 3$ and $r(C) = 2$, so C ends with $a_0 \ldots a_{m-3}$. By eq. (13)

$$\mathcal{H}(3; C) = 0 * \mathcal{H}(2; C_0) \quad \cup \quad 1 * \mathcal{H}(2; C_1).$$

Note that C_0 ends with $a_0 \ldots a_{m-2}$, so $r(C_0) = 1$ and $\mathcal{H}(2; C_0) = \{11, 10\}$. Since C_1 ends with 1, $r(C_1) \geq 2$. If $r(C_1) = 2$, then $\mathcal{H}(2; C_1) = \{01, 11, 10\}$, and if $r(C_1) > 2$, then $\mathcal{H}(2; C_1) = \Sigma_2^2$. In the former case, let $\mathbf{L} = 101, 111, 011, 010, 110$. Then $\mathbf{L}$ contains the path $[X_3, Y_3, Z_3']$. The fact that $r(C_1) = 2$ implies that C_1 ends in $a_0 \ldots a_{m-3}$, so $a_{m-3} = 1$, and $\mathbf{L}$ runs between S_3 and T_3. In the latter case, $\mathbf{L} = 101, 111, 011, 010, 110, 100$ is a cyclic Gray code containing $[X_3, Y_3, Z_3']$.

Inductive step for Part 1. Suppose that n is in the range $3 \leq n \leq m$, $C \in \Sigma_2^*$ has no A-factor, and $r(C) = 1$. Then $\mathcal{H}(n; C) = 1 * \mathcal{H}(n-1; C_1)$ by eq. (12). Since C_1 ends with 1, $r(C_1) \geq 2$ and $\mathcal{H}(n-1; C_1)$ falls into Part 2 of the theorem. Inductively assume that the theorem holds for $\mathcal{H}(n-1; C_1)$.

Suppose $h^*(n; C) = 0$, so $h^*(n-1; C_1) = 0$ also. By induction, there exists a cyclic Gray code $\mathbf{L}_1$ for $\mathcal{H}(n-1; C_1)$ containing the path $[X_{n-1}, Y_{n-1}, Z_{n-1}'] = [101^{n-3}, 1^{n-1}, 01^{n-2}]$. Then $1 * \mathbf{L}_1$ is a Gray code for $\mathcal{H}(n; C)$ satisfying Part 1.

Next suppose $|h^*(n; C)| = 1$, so $|h^*(n-1; C_1)| = 1$ also. We can inductively find a Gray code $\mathbf{L}_1$ for $\mathcal{H}(n-1; C_1)$ such that $\mathbf{L}_1$ starts at S_{n-1} and ends at T_{n-1}. Further, $\mathbf{L}_1$ contains the path $[X_{n-1}, Y_{n-1}, Z_{n-1}']$. If $a_{m-n} = 1$, then $1 * \mathbf{L}_1$ satisfies Part 1. By Lemma 4.2, C_1 must end with $a_0 \ldots a_{m-n}$, so $a_{m-n} = 1$. This concludes the inductive step for Part 1 of the theorem.

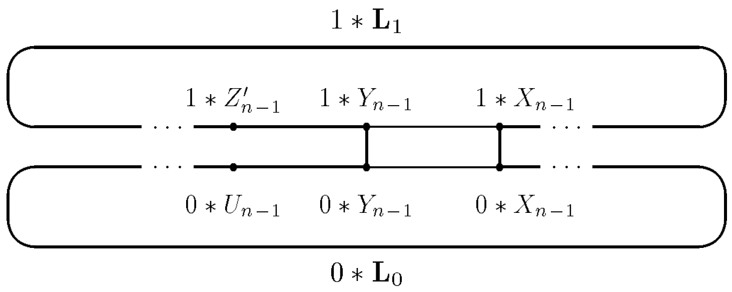
Inductive step for Part 2. Suppose that n is in the range $3 \leq n \leq m$, $C \in \Sigma_2^*$ has no A-factor, and $r(C) \geq 2$. Inductively assume that the theorem holds for $\mathcal{H}(n-1; C_0)$ and $\mathcal{H}(n-1; C_1)$. We separately examine when $h^*(n; C) = 0$ (Case 2a), and when $|h^*(n; C)| = 1$ (Case 2b).

Case 2a. Suppose $h^*(n; C) = 0$. By Corollary 4.1, both $h^*(n-1; C_0)$ and $h^*(n-1; C_1)$ are also zero. Note that $r(C_1) \geq 2$ because C_1 ends with 1. There is a cyclic Gray code $\mathbf{L}_1$ for $\mathcal{H}(n-1; C_1)$ containing the path $[X_{n-1}, Y_{n-1}, Z_{n-1}']$. Also, there is a cyclic Gray code $\mathbf{L}_0$ for $\mathcal{H}(n-1; C_0)$ containing the path $[X_{n-1}, Y_{n-1}, U_{n-1}]$, where $U_{n-1} = Z_{n-1}$ if $r(C_0) = 1$, and $U_{n-1} = Z_{n-1}'$ otherwise. Since $\mathbf{L}_0$ and $\mathbf{L}_1$ both contain the path $[X_{n-1}, Y_{n-1}]$, a Gray code $\mathbf{L}$ for $\mathcal{H}(n; C)$ can be formed by joining $0 * \mathbf{L}_0$ and $1 * \mathbf{L}_1$ by these strings as in Figure 3. Then $\mathbf{L}$ contains the path $[1 * Z_{n-1}', 1 * Y_{n-1}, 0 * Y_{n-1}] = [X_n, Y_n, Z_n']$, satisfying Part 2.

Case 2b. Suppose $|h^*(n; C)| = 1$. Let $D = C * \overline{a}_{m-n}$ and $E = C * a_{m-n}$, so

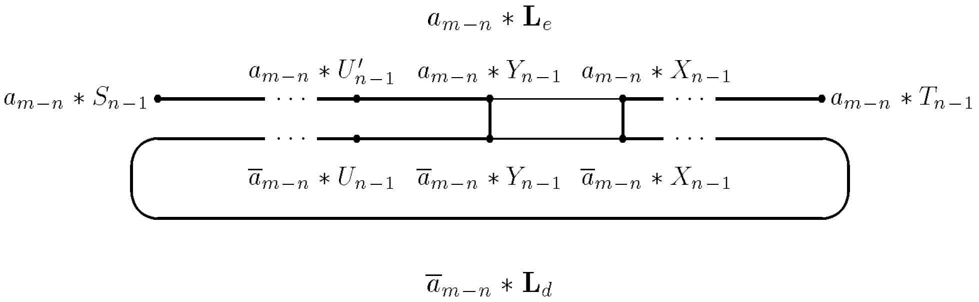
$$\mathcal{H}(n; C) \quad = \quad \overline{a}_{m-n} * \mathcal{H}(n-1; D) \quad \bigcup \quad a_{m-n} * \mathcal{H}(n-1; E).$$

Using Corollary 4.1, $h^*(n-1; D) = 0$ and $|h^*(n-1; E)| = 1$. The possibility that $n = 3$ and $r(C) = 2$ was considered in the basis. Inductively, there exists a cyclic Gray code $\mathbf{L}_d$ for $\mathcal{H}(n-1; D)$ containing the path $[X_{n-1}, Y_{n-1}, U_{n-1}]$, where $U_{n-1} = Z_{n-1}$ if $r(D) = 1$



If $r(C_0) = 1$, then $U_{n-1} = 1101^{n-4}$.
If $r(C_0) > 1$, then $U_{n-1} = 01^{n-2}$.

Figure 3: Gray code construction for Case (2a)



If $r(D) = 1$, then $U_{n-1} = Z_{n-1}$. Else $U_{n-1} = Z'_{n-1}$.
If $r(E) = 1$, then $U'_{n-1} = Z_{n-1}$. Else $U'_{n-1} = Z'_{n-1}$.

Figure 4: Gray code construction for Case (2b)

and $U_{n-1} = Z'_{n-1}$ otherwise. There also exists a Gray code $\mathbf{L}_e$ for $\mathcal{H}(n-1; E)$ starting at S_{n-1} and ending at T_{n-1}. The Gray code $\mathbf{L}_e$ contains the path $[X_{n-1}, Y_{n-1}, U'_{n-1}]$, where $U'_{n-1} = Z_{n-1}$ if $r(E) = 1$ and $U'_{n-1} = Z'_{n-1}$ otherwise.

Let $\mathbf{L}$ be the Gray code obtained from $\overline{a}_{m-n} * \mathbf{L}_d$ and $a_{m-n} * \mathbf{L}_e$ as shown in Figure 4. Then $\mathbf{L}$ starts and ends on the correct vertices, and it contains the path

$$[a_{m-n} * U'_{n-1}, \quad a_{m-n} * Y_{n-1}, \quad \overline{a}_{m-n} * Y_{n-1}, \quad \overline{a}_{m-n} * U_{n-1}].$$

If $a_{m-n} = 0$, then D ends with 1, so $r(D) > 1$ and $U_{n-1} = Z'_{n-1}$. If $a_{m-n} = 1$, then E ends with 1, so $r(E) > 1$ and $U'_{n-1} = Z'_{n-1}$. In either case, $\mathbf{L}$ contains the path $[X_n, Y_n, Z'_n]$. This completes the inductive argument.

Note that S_n and T_n have parity opposite that of A_n. If $C * a_{m-n} \ldots a_{m-1}$ does not have an A-factor, then neither does $C * a_{m-n} \ldots \overline{a}_{m-1}$. Likewise for $C * T_n$. □

By considering the small n cases not covered by Theorem 4.1, we get the following.

Corollary 4.2 *Fix $q = 2$, $m > 1$ and $A \in \Sigma_2^m$ such that A ends with two identical characters. For any n, $1 \leq n \leq m$, and for any $C \in \Sigma_2^*$ with no A-factor, there is a*

Gray code for $\mathcal{H}(n;C)$. When $h^*(n;C)=0$, the Gray code is cyclic. Otherwise, its endpoints are independent of C and of equal parity.

Proof. We need only consider when $n \leq 2$ and Theorem 4.1 does not apply. □

In the next section, we present results analogous to Theorem 4.1 and Corollary 4.2 for even $q \geq 4$.

4.3 Gray codes when $q \geq 4$ is even and $n \leq m$

In this section, we construct Gray codes for $\mathcal{H}(n;C)$ when $1 \leq n \leq m$ and $q \geq 4$ is even. For $x \in \Sigma_q$, we defined $\overline{x} \equiv x+1 \bmod q$. We similarly define $\underline{x} \equiv x-1 \bmod q$. Recall from Section 1 that when q is even, two strings are "similar" if they differ at one position, and by $\pm 1 \bmod q$ in that position.

Theorem 4.2 *Fix an even $q \geq 4$, $m > 0$ and $A \in \Sigma_q^m$. For any n, $1 \leq n \leq m$, there exist strings S_n, T_n, X_n, Y_n, and Z_n such that for any $C \in \Sigma_q^*$ with no A-factor, there exists a Gray code for $\mathcal{H}(n;C)$ containing the path $[X_n, Y_n, Z_n]$. When $h^*(n;C)=0$, the Gray code is cyclic, and when $|h^*(n;C)|=1$, the Gray code is from S_n to T_n. Further, S_n and T_n have parity opposite that of $A_n = a_{m-n}\ldots a_{m-1}$, and $A_n \in \mathcal{H}(n;C)$ implies $S_n, T_n \in \mathcal{H}(n;C)$.*

Proof. Let x satisfy $|x - a_{m-1}| \geq 2$. We argue inductively on n using the strings

$$\begin{array}{rclcrcl} S_n &=& a_{m-n}\ldots\overline{a}_{m-1} & \qquad & X_n &=& \underline{x}x^{n-1} \\ T_n &=& a_{m-n}\ldots\underline{a}_{m-1} & & Y_n &=& x^n \\ & & & & Z_n &=& \overline{x}x^{n-1} \end{array}$$

Let C be a string over Σ_q with no A-factor. When $r(C) > 1$, $\mathcal{H}(1;C) = \Sigma_q$. When $r(C)=1$, $\mathcal{H}(1;C) = \Sigma_q \setminus \{a_{m-1}\}$. Note that $\{\underline{x}, x, \overline{x}\} \subseteq \mathcal{H}(1;C)$. The Gray codes in these cases are, respectively, $a_{m-1}, \overline{a}_{m-1}, \ldots, \underline{a}_{m-1}$ and $\overline{a}_{m-1}, \ldots, \underline{a}_{m-1}$. These Gray codes satisfy the theorem.

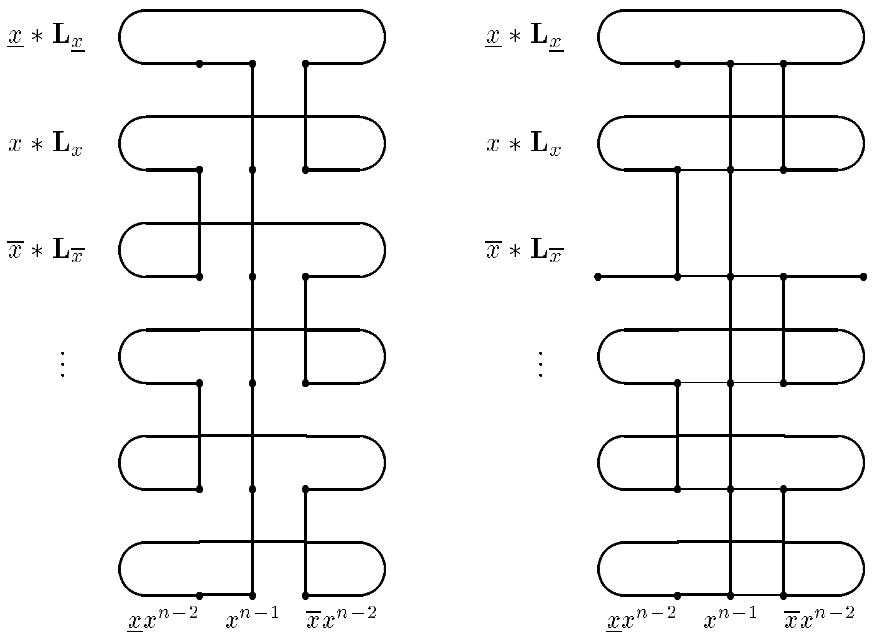
Suppose that $1 < n \leq m$ and $C \in \Sigma_q^*$ has no A-factor. From (11) we obtain

$$\mathcal{H}(n;C) = \bigcup_{y \in \mathcal{H}(1;C)} y * \mathcal{H}(n-1; C*y).$$

For each $y \in \mathcal{H}(1;C)$, assume the theorem holds for $\mathcal{H}(n-1;C*y)$. Let $\mathbf{L}_y$ be a Gray code for $\mathcal{H}(n-1;C*y)$ containing the path $[X_{n-1}, Y_{n-1}, Z_{n-1}]$. These Gray codes exist inductively, and are either cyclic, or between S_n and T_n.

Suppose $h^*(n;C)=0$. By Corollary 4.1, $h^*(n-1;C*y)=0$ for all $y \in \mathcal{H}(1;C)$, so the Gray codes $\mathbf{L}_y$ are cyclic. The $y * \mathbf{L}_y$ can be connected as in Figure 5 (a) to form a Gray code for $\mathcal{H}(n;C)$. First, join all $y*\mathbf{L}_y$ using the string $y*Y_{n-1}$, ensuring that $[\underline{x}*Y_{n-1}, x*Y_{n-1}, \overline{x}*Y_{n-1}]$ is a subpath. This can be done because $\{\underline{x}, x, \overline{x}\} \subseteq \mathcal{H}(1;C)$. Next, use the outer two strings on the fixed path to finish the joining the $y*\mathbf{L}_y$. This strategy can be employed independent of $r(C)$.

A similar strategy works when $|h^*(n;C)|=1$. In this case, by Corollary 4.1, $|h^*(n-1;C*a_{m-n})|=1$ and $h^*(n-1;C*y)=0$ for $y \in \mathcal{H}(1;C)\setminus\{a_{m-n}\}$. For



(a) when $h^*(n;C) = 0$ (a) when $|h^*(n;C)| = 1$

Figure 5: Gray code construction when $q \geq 4$ is even.

$y \in \mathcal{H}(1;C)$, $y \neq a_{m-n}$, $\mathbf{L}_y$ is cyclic, while $\mathbf{L}_{a_{m-n}}$ goes from S_{n-1} to T_{n-1}. Connect all of the $y * \mathbf{L}_y$ using $y * Y_{n-1}$. Then, as before, use the outer two strings to finish joining the $y * \mathbf{L}_y$. Refer to Figure 5 (b) for an illustration.

Observe that S_n and T_n have parity opposite that of A_n, and if $C * A_n = C * a_{m-n} \ldots a_{m-1}$ has no A-factor, then neither do $C * S_n$ and $C * T_n$. $\square$

As with Theorem 4.1, we get a corollary to Theorem 4.2.

Corollary 4.3 *Fix an even $q \geq 4$, $m > 0$ and $A \in \Sigma_q^m$. For any n, $1 \leq n \leq m$, and any $C \in \Sigma_q^*$ with no A-factor, there is a Gray code for $\mathcal{H}(n;C)$. When $h^*(n;C) = 0$, the Gray code is cyclic. Otherwise, its endpoints are independent of C and of equal parity.*

The techniques used to construct Gray codes in Theorems 4.1 and 4.2 do not apply as n is increased. In particular, Lemma 4.2 and Corollary 4.1 are not true for all $n > m$. For larger n, a new construction technique must be employed.

In Section 3.1 we showed that, when q is even, unless A is either t.a. or periodic with a t.a. minimal base, then $\mathcal{F}(n)$ has infinite parity problems. In Section 4.4, Gray codes for $\mathcal{F}(n)$ are constructed for all n when A is t.a and q is even. When A is periodic with a t.a. minimal base and q is even, Gray codes for $\mathcal{F}(n)$ also exist for all $n \geq 0$. The proof of this result is in Section 4.5.

4.4 A trivially autocorrelated

Recall that A is trivially autocorrelated, or t.a., if $AA = 10\cdots 0$, or, equivalently, if $r(A) = m$. *A trivially autocorrelated string can overlap itself in only the trivial ways.* In this section, we construct Gray codes for $\mathcal{F}(n)$ for all $n \geq 0$ when A is t.a. and q is even. The main result of this section is Theorem 4.3.

Theorem 4.3 *Fix an even $q \geq 2$, $m > 0$, and a t.a $A \in \Sigma_q^m$. For all $n \geq 1$ there is a Gray code for $\mathcal{F}(n)$. If $f^*(n) = 0$, and either $A \notin \{01, 10\}$ or $q \geq 4$, then there is a cyclic Gray code.*

We prove Theorem 4.3 a little later. First, we examine the parity difference of $\mathcal{H}(m; C)$.

Lemma 4.3 *Fix an even $q \geq 2$, $m > 0$ and a t.a $A \in \Sigma_q^m$. For any $C \in \Sigma_q^*$ with no A-factor, $|h^*(m; C)| = 1$.*

Proof. Let $C \in \Sigma_q^*$ be a string with no A-factor. It is clear that $A \notin \mathcal{H}(m; C)$. By Lemma 4.2, if, for all $x \neq a_{m-1}$, $A' = a_0 \ldots a_{m-2}x \in \mathcal{H}(m; C)$, then $|h^*(m; C)| = 1$. Suppose $D = C * A'$ has an A-factor. Any A factor in D must overlap A' in a non-trivial way, which contradicts the hypothesis that A is t.a. Therefore, $C * A'$ does not have an A-factor and the lemma is proved. □

We now prove Theorem 4.3. Recall Lemma 4.1 of Section 4.1. First suppose that A is any non-empty t.a. string and $q \geq 4$. Then Corollary 4.3 provides Gray codes for $\mathcal{H}(n; C)$ for all n, $1 \leq n \leq m$, and all C with no A-factor. It provides cyclic Gray codes when $h^*(n; C) = 0$. This same corollary, in conjuction with Lemma 4.3 above, also proves the existence of strings S and T, of equal parity, such that for any D with no A-factor, there is a Gray code for $\mathcal{H}(m; D)$ from S to T. Hence, conditions (1) and (2) of Lemma 4.1 are satisfied when $k = m$. We can therefore conclude that for all $n \geq 1$ and for all C with no A-factor, there is a Gray code for $\mathcal{H}(n; C)$. Further, there is a cyclic Gray code when $h^*(n; C) = 0$. Since $\mathcal{F}(n) = \mathcal{H}(n; \epsilon)$, Theorem 4.3 is proved when $q \geq 4$ is even.

Next suppose that $q = 2$ and A is a t.a. string ending in two identical characters. In this case, Corollary 4.2 and Lemma 4.3 prove that, when $k = m$, conditions (1) and (2) of Lemma 4.1 are again satisfied. Thus, for all $n \geq 1$ and all C with no A-factor, there is a Gray code for $\mathcal{H}(n; C)$. There is a cyclic Gray code when $h^*(n; C) = 0$. Again, Theorem 4.3 follows by $\mathcal{F}(n) = \mathcal{H}(n; C)$.

We are left to consider when $q = 2$ and A is t.a. but does not end in two identical characters. We claim that any t.a. binary string of length $m \geq 3$ has either $a_{m-2} = a_{m-1}$ or $a_0 = a_1$. To see this, suppose A is a t.a. binary string of length $m \geq 3$ with $a_0 \neq a_1$ and $a_{m-2} \neq a_{m-1}$. WLOG, say $a_0 = 0$. If $a_{m-1} = 0$, then bit $(m-1)$ of AA is one. If $a_{m-1} = 1$, then A begins and ends with 01, so bit $(m-2)$ of AA is one. In either case, A is not t.a.

So every t.a. binary string of length $m \geq 3$ must begin or end with two identical characters. We have already considered when A ends with two identical characters, so suppose that A begins with two identical characters, and let Z be the reverse of A. Then Z is also t.a., and Z ends with two identical characters. For $n \geq 1$, we can find a Gray code $\mathbf{L}$ for the strings with no Z-factor. Reversing every string on $\mathbf{L}$ gives a Gray code for $\mathcal{F}(n)$, the strings with no A-factor.

We have now proved Theorem 4.3 in all cases except when $q = 2$ and $m < 3$. The only t.a. binary strings of length less than or equal to two are $\{0, 1, 01, 10\}$. When $A = 0$, $\mathcal{F}(n) = \{1^n\}$. When $A = 01$, $\mathcal{F}(n) = \{1^k 0^{n-k} | 0 \le k \le n\}$. When $A = 1$ or $A = 10$, the set $\mathcal{F}(n)$ is similarly defined. Gray codes in these cases are clear and unique. If $A \in \{01, 10\}$, then no cyclic Gray code exists even when the parity difference is zero. This completes the proof of Theorem 4.3.

In Section 4.5, we prove an analogous result when A is periodic with a t.a. minimal base.

4.5 A periodic with a t.a. minimal base

In this section, we examine $\mathcal{H}(n; C)$ when $A = B * \cdots * B$ and B is t.a. Throughout Section 4.5, B refers to the minimal base of A and p to the minimal period. Again, $q \ge 2$ is assumed even. The overall strategy employed in this section is identical to that of the previous section. We identify k such that $|h^*(k; C)| = 1$ independent of C. We then apply Lemma 4.1 to obtain Gray codes for all n.

As seen in Section 4.4, when A is t.a. and $k = m$, $|h^*(k; C)| = 1$ independent of C. When A is periodic with a t.a. minimal base, this is no longer true, as we see below.

Lemma 4.4 *Fix an even $q \ge 2$, $m > 0$ and a periodic $A \in \Sigma_q^m$ with a t.a minimal base B. For any $C \in \Sigma_q^*$ with no A-factor, $h^*(m; C) = 0$ if and only if C ends with B. Otherwise, $|h^*(m; C)| = 1$.*

Proof. Let $C \in \Sigma_q^*$ be any string with no A-factor. By Lemma 4.2 with $n = m$, $h^*(m; C) = 0$ if and only if $A' = a_0 \ldots a_{m-2} x \notin \mathcal{H}(m; C)$ for some $x \ne a_{m-1}$. Suppose $C * A'$ has an A-factor. Some B-term of this A-factor must overlap the first B-term of A'. Since B is t.a., the B-terms must overlap completely. This implies that C ends with B. Otherwise, $A' \in \mathcal{H}(m; C)$ for all $x \ne a_{m-1}$, and $|h^*(m; C)| = 1$. □

Suppose that $A \in \Sigma_q^m$ is periodic with a t.a. minimal base and with minimal period p. By the previous lemma, there cannot exist strings S and T such that for any C with no A-factor, there is a Gray code for $\mathcal{H}(m; C)$ from S to T. However, we will show that such strings S and T do exist for $\mathcal{H}(m + p; C)$. As an intermediate step, we extend Theorems 4.1 and 4.2 to provide Gray codes for $\mathcal{H}(n; C)$ for n in the range $m < n \le m + p$. This is done below.

Lemma 4.5 *Fix an even $q \ge 2$, $m > 0$, and a periodic $A \in \Sigma_q^m$ with a t.a. minimal base B and with minimal period p. When $q = 2$, further assume that $m \ge 2$ and $a_{m-2} = a_{m-1}$. For any $C \in \Sigma_q^*$ with no A-factor and any n, $1 \le n \le p$, there is a Gray code for $\mathcal{H}(m+n; C)$. When $h^*(m+n; C) = 0$, there is a cyclic Gray code. Otherwise, $|h^*(m+n; C)| = 1$, and there is a Gray code between two strings independent of C and of equal parity.*

Proof. Let $C \in \Sigma_q^*$ be a string with no A-factor, and let n be in the range $1 \le n \le p$. By Corollaries 4.2 and 4.3, for $1 \le l \le m$, there exist strings S_l and T_l, of equal parity, such that $\mathcal{H}(l;C)$ has a Gray code between S_l and T_l whenever $|h^*(l;C)| = 1$. Define S and T by $S = S_n * S_m$ and $T = T_n * T_m$. We show that when $|h^*(n+m;C)| = 1$, there is a Gray code for $\mathcal{H}(n+m;C)$ between S and T. Observe that S and T have equal parity.

For a recursive decomposition of $\mathcal{H}(m+n;C)$, we use eq. (11) with $k = m$. Since $n \leq p$, there exists *at most one* $D \in \mathcal{H}(n;C)$ such that $C * D$ ends with B, namely $A_n = b_{p-n} \dots b_{p-1}$. By Lemma 4.4, for at most one $D \in \mathcal{H}(n;C)$ does $h^*(m;C*D) = 0$. When $C * D$ does not end with B, there is a Gray code $\mathbf{L}_D$ for $\mathcal{H}(m;C*D)$ between S_m and T_m. When $C*D$ ends with B, there is a cyclic Gray code $\mathbf{L}_D$ for $\mathcal{H}(m;C*D)$. Consider: when $|h^*(n;C)| = 1$ (Case 1), and when $h^*(n;C) = 0$ (Case 2).

Case 1. Assume $|h^*(n;C)| = 1$. By Corollaries 4.2 and 4.3, there exists a Gray code $\mathbf{L}' = D_0, \dots, D_{N-1}$ for $\mathcal{H}(n;C)$ between S_n and T_n. Note that N is odd. By Lemma 4.2, C ends with $a_0 \dots a_{m-n-1}$, so $A_n \notin \mathcal{H}(n;C)$. Therefore, for no D in $\mathcal{H}(n;C)$ does $C * D$ end with B. By Lemma 4.4, $|h^*(m;C*D_i)| = 1$ for all D_i. For each i, we can find a Gray code $\mathbf{L}_i$ for $\mathcal{H}(m;C*D_i)$ from S_m to T_m. Let

$$\mathbf{L} = D_0 * \mathbf{L}_0, \quad D_1 * \mathbf{L}_1^R, \quad \cdots, \quad D_{N-1} * \mathbf{L}_{N-1}.$$

Then $\mathbf{L}$ is a Gray code for $\mathcal{H}(m+n;C)$ from S to T.

Case 2. Assume $h^*(n;C) = 0$. By Corollaries 4.2 and 4.3, there is a cyclic Gray code $\mathbf{L}' = D_0, \dots, D_{N-1}$ for $\mathcal{H}(n;C)$. We have two subcases to consider: when there exists a D_i such that $C * D_i$ ends with B, and when no such D_i exists. In the latter case, by Lemma 4.4, $|h^*(m;C*D_i)| = 1$ for all D_i. There exists a Gray code $\mathbf{L}_i$ for each $\mathcal{H}(m;C*D_i)$ from S_m to T_m by Corollaries 4.2 and 4.3. Then

$$\mathbf{L} = D_0 * \mathbf{L}_0, \quad D_1 * \mathbf{L}_1^R, \quad \dots, D_{N-1} * \mathbf{L}_{N-1}^R$$

is a cyclic Gray code for $\mathcal{H}(m+n;C)$.

Next, consider when there exists an i such that $C * D_i$ ends with B. WLOG we can assume that $i = 0$, so $D_0 = A_n$. First, assume that either $n \geq 2$ or $q \geq 4$. By Theorems 4.1 and 4.2, there exist strings $X_m, Y_m \in \Sigma_q^m$ such that for all i, there is a Gray code for $\mathcal{H}(m;C*D_i)$ containing the path $[X_m, Y_m]$. When $i = 0$, there is a cyclic Gray code, and for all other i, there is a Gray code from S_m to T_m. By Theorems 4.1 and 4.2, S_n and T_n are in $\mathcal{H}(n;C)$ because $A_n = D_0 \in \mathcal{H}(n;C)$. Further, S_n and T_n have parity opposite that of A_n.

Let l and r be such that $D_l = S_n$ and $D_r = T_n$. WLOG assume $l < r$. By the parity condition, l and r must be odd. The construction of the Gray code for $\mathcal{H}(m+n;C)$ has three phases. Refer to Figure 6 for a graphical illustration. We refer to X_m and Y_m as *special* strings.

In phase one, we construct a Gray code for $\cup_{i=0}^{l} D_i * \mathcal{H}(m;C*D_i)$. Starting with the S_m end of $\mathbf{L}_l$, follow $\mathbf{L}_l$ until reaching the first special string. Then jump to $\mathbf{L}_{l-1}$ and move to S_m or T_m, whichever end is reachable without passing through the other special string. At this endpoint, jump to $\mathbf{L}_{l-2}$, and continue this process of using only the part of $\mathbf{L}_i$ between an endpoint its closest special string. At D_0, follow the cyclic Gray code $\mathbf{L}_0$ between the special strings, and then work back up the $\mathbf{L}_i$ using the part of $\mathbf{L}_i$ not used on the way down. Continue this process until the T_m end of $\mathbf{L}_l$

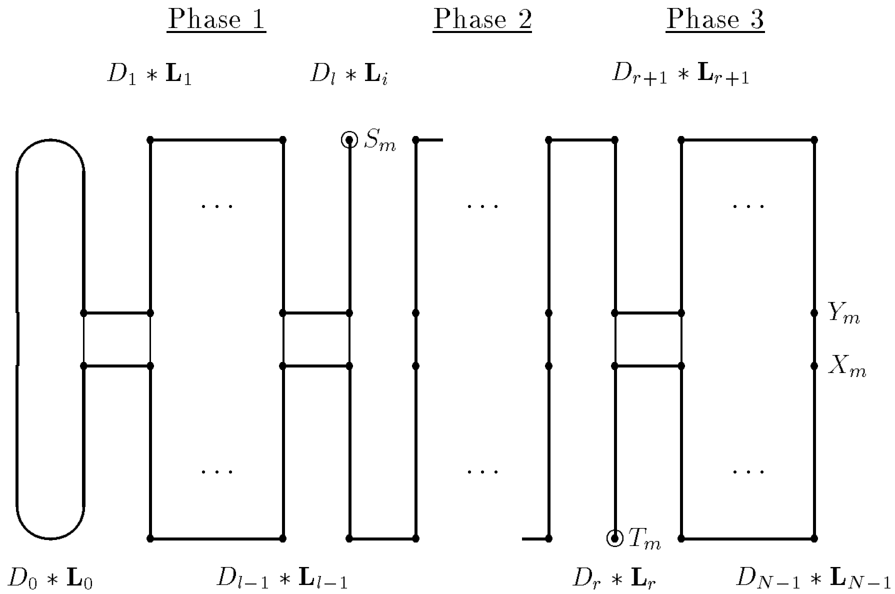


Figure 6: Constructing a Gray code for $\mathcal{H}(m+n;C)$ when $1 \le n \le p$

is reached. This constructs a Gray code for $\cup_{i=0}^{l} D_i * \mathcal{H}(m; C * D_i)$ from $S_n * S_m$ to $S_n * T_m$.

In phase two, a Gray code for $\cup_{i=l+1}^{r-1} D_i * \mathcal{H}(m; C * D_i)$ is constructed. The Gray code begins with $D_{l+1} * T_m$ and ends with $D_{r-1} * S_m$. The Gray code is defined by

$$D_{l+1} * \mathbf{L}_{l+1}^R, \quad D_{l+2} * \mathbf{L}_{l+2}, \quad \ldots, \quad D_{r-1} * \mathbf{L}_{r-1}^R.$$

Phase three constructs a Gray code for $\cup_{i=r}^{N-1} D_i * \mathcal{H}(m; C * D_i)$ from $T_n * S_m$ to $T_n * T_m$. This Gray code is constructed in a similar manner to the phase one Gray code. Starting with S_m in $\mathbf{L}_r$, follow $\mathbf{L}_r$ from S_m to the nearest special string, then jump to $\mathbf{L}_{r+1}$. In $\mathbf{L}_{r+1}$, move to the endpoint reachable without passing through the other special string, and jump to $\mathbf{L}_{r+2}$. When $\mathbf{L}_{N-1}$ is reached, use it to turn around by going between S_m and T_m. Recall that r is odd and N is even, so this is possible. Then travel back down the unused portions of the $\mathbf{L}_i$ until reaching $D_r * T_m$.

Concatenating together the Gray codes from each phase yields a Gray code for $\mathcal{H}(m+n;C)$ from S to T.

We have not yet considered when $q = 2$ and $n = 1$. It is a simple matter to get a Gray code for $\mathcal{H}(m+1;C)$ from Gray codes for $\mathcal{H}(m;C_0)$ and $\mathcal{H}(m;C_1)$. We omit the details. In all cases, when $h^*(m+n;C) = 0$, the Gray codes constructed are cyclic. Otherwise, the Gray codes are between S and T. □

Suppose $A \in \Sigma_q^m$ is periodic with minimal period p and with a t.a. minimal base. Let $C \in \Sigma_q^*$ be a string with no A-factor. The preceding lemma supplies a Gray code for $\mathcal{H}(n;C)$ for n in the range $m < n \le m+p$. This result can be combined with Corollaries 4.2 and 4.3 to get Gray codes for $\mathcal{H}(n;C)$ whenever $1 \le n \le m+p$. We next show that at the upper bound of this range, there is always a Gray code between two endpoints independent of C.

Lemma 4.6 *Fix an even $q \geq 2$, $m > 0$, and a periodic $A \in \Sigma_q^m$ with a t.a. minimal base and with minimal period p. If $C \in \Sigma_q^*$ has no A-factor, then $|h^*(m+p;C)| = 1$.*

Proof. As in the proof of Lemma 4.2, we pair even and odd strings. Say $C \in \Sigma_q^*$ has no A-factor, and let $D = d_0 \dots d_{m+p-1}$ be a string in $\mathcal{H}(m+p;C)$. If D does not end with $A' = a_0 \dots a_{m-2}x$, where x is the mate of a_{m-1}, then pair D with the string obtained by replacing the last character of D with its mate.

Let $\mathcal{S}$ denote the set of unpaired strings in $\mathcal{H}(m+p;C)$. We claim each $D \in \mathcal{S}$ can be paired with the string obtained by replacing the character in position $(p-1)$ of D with its mate *unless* D also begins with $B' = b_0 \dots b_{p-2}x$, where x is the mate of $b_{p-1} = a_{m-1}$. Suppose $D = d_0 \dots d_{p-1} * A' \in \mathcal{S}$ but $E = D' * A' = d_0 \dots d_{p-2}y * A' \notin \mathcal{S}$, where y is the mate of d_{p-1}. Every A-factor in $C * E$ must use last character of D'. If the A-factor of $C * E$ ends on this character, then $D' = B$. If a B-term of the A-factor overlaps the first B-term of A', then the B-terms must overlap completely. Again, $D' = B$. Thus, the only possible unmatched string is $B' * A'$.

We claim $B' * A' \in \mathcal{H}(m+p;C)$ for all $C \in \Sigma_q^*$ with no A-factor. If $D = C * B' * A'$ contained an A-factor, then either the A-factor is in $C * B'$, or the A-factor overlaps the A'-suffix of D. In either case, B overlaps itself in a non-trivial way. Thus $B' * A' \in \mathcal{H}(m+p;C)$, and $h^*(m+p;C)$ is ± 1 depending on the parity of $B' * A'$. □

We now prove the main result of this section.

Theorem 4.4 *Fix an even $q \geq 2$, $m > 0$, and a periodic $A \in \Sigma_q^m$ with a t.a. minimal base. For $n \geq 1$, there exists a Gray code for $\mathcal{F}(n)$. When $f^*(n) = 0$, there is a cyclic Gray code.*

Proof. This proof is similar to the proof of Theorem 4.3 in Section 4.4. Suppose that $A \in \Sigma_q^m$ is periodic with a t.a. minimal base B and with minimal period p, and that $q \geq 4$ is even. Say $C \in \Sigma_q^*$ has no A-factor. Refer to Lemma 4.1 of Section 4.1. By Corollary 4.3 and Lemma 4.5, there exist a Gray code for $\mathcal{H}(n;C)$ for all $1 \leq n \leq m+p$. There is a cyclic Gray code whenever $h^*(n;C) = 0$. This satisfies condition (1) of Lemma 4.1. By Lemmas 4.5 and 4.6, for any D with no A-factor, there exist a Gray code for $\mathcal{H}(m+p;D)$ between two strings independent of D and of equal parity. This satisfies condition (2) of Lemma 4.1. Therefore, for all $n \geq 1$ and for all C with no A-factor, there exists a Gray code for $\mathcal{H}(n;C)$. This Gray code is cyclic when $h^*(n;C) = 0$. This, of course, implies the Gray code result for $\mathcal{F}(n) = \mathcal{H}(n;\epsilon)$.

Now suppose $q = 2$. If A ends in two identical characters, then the arguments of the preceding paragraph can be repeated by replacing Corollary 4.3 with Corollary 4.2. We have not yet considered when A is periodic with a t.a. minimal base and A does not end in two identical characters. As in Section 4.4, if $p \geq 3$, then A either begins or ends in two identical characters. If A begins with two identical characters, then Gray codes for $\mathcal{F}(n)$ can be found by considering the reverse of A.

We are left to consider when $p \leq 2$. If $p = 2$, then $B \in \{01, 10\}$. In this case, neither A nor its reverse satisfies the conditions of Corollary 4.2. If $p = 1$ and $m \geq 2$, then A satisfies the conditions of Corollary 4.2 and the conclusions follow as before.

The remaining possibility is that $B \in \{01, 10\}$. To construct Gray codes for $\mathcal{H}(n; C)$ in this case requires complicating the constructions in Section 4. In particular, there are more individual cases in the basis. The details can be found in Appendix A, and the conclusions of the theorem still follow. $\square$

When q is even, we have constructed Gray codes for $\mathcal{F}(n)$ for all $n \geq 1$ when A is t.a. or A is periodic with a t.a. minimal base. The results of Section 3.1 prove that these are the only A for which Gray codes are always possible.

5 Conclusions

In this paper, we have investigated the existence of Gray codes for A-free strings. We have determined for which q and A Gray codes for $\mathcal{F}(n)$ are always possible, and in the even q case, we have constructed the Gray codes. When q is odd, it remains open to construct Gray codes for $\mathcal{F}(n)$ when no parity problems exist. Recall that when q is odd, two strings are similar if they differ in one position, and by ± 1 in that position, where the addition is *not* modulo q. If modular arithmetic were used, the resulting auxiliary graph $G_q(\mathcal{F}(n))$ would not be bipartite, thus invalidating the parity arguments of Section 3. Note that when q is odd, it is possible for the graph $G_q(\mathcal{F}(n))$ to be disconnected. For example, consider when $q = 5$, $A = 2$, and $n \geq 1$. However, we conjecture that, with few exceptions, there do exist a Gray codes for $\mathcal{F}(n)$ when $\mathcal{F}(n)$ does not have infinite parity problems.

Similarly, we can consider removing the modular arithmetic in the even q case. Again, we conjecture that, for the most part, there exist Gray codes for $\mathcal{F}(n)$ when $\mathcal{F}(n)$ does not have infinite parity problems. We can also consider using modular arithmetic in the odd q case. The parity arguments of Section 3 would no longer apply, so Gray codes may exist in situations not covered in Section 3

As mentioned in Section 1, the A-free string concept has been defined more generally. It would be interesting to develop conditions on a set $\mathcal{A}$ of strings such that $\mathcal{F}(n; \mathcal{A})$ has a Gray code for all $n \geq 0$. In this general case, even if q is even and we use modular addition, the graph $G_q(\mathcal{F}(n))$ need not be connected. For example, consider when $q = 2$, $\mathcal{A} = \{01, 10\}$, and $n \geq 2$.

There have also been many variations on the A-free string concept. Given $\mathcal{A} = \{A_1, \ldots, A_p\}$ and a vector $\mathbf{m} = (m_1, \ldots, m_p)$, let $\mathcal{F}(n; \mathcal{A}; \mathbf{m})$ denote the set of $X \in \Sigma_q^n$ such that, for $1 \leq i \leq p$, X has m_i occurrences of A_i. Goulden and Jackson [3] prove that $\mathcal{F}(n; \mathcal{A}; \mathbf{m})$ has a rational generating function. Zeilberger [12] and Goulden and Jackson [2] have studied similar sets when each $i \in \Sigma_q$ is restricted to appear a certain number of times. Zeilberger [13] has also studied the concept of A-free strings when A cannot appear as a sub*sequence* (not necessarily contiguous). Gray codes for these variations deserve future attention.

We briefly mention the topic of generating A-free strings. Given $\mathcal{A}$, one can define a deterministic finite automaton M so that M accepts $\cup_n \mathcal{F}(n; \mathcal{A})$ [10]. Once we have a DFA for a language, generating its elements is a simple procedure.

The enumeration and generation of objects with restricted subobjects is an area rich for study. If strings are genetic sequences and certain subsequences are "diseased," what is the probability of being disease-free? How does one efficiently generate all graphs that do not contain certain subconfigurations? How many ways can G be colored so that there does not exist a unicolored edge? These and many other problems can be characterized as enumerating and generating sets of objects with forbidden subobjects. As seen in this paper, introducing a Gray code requirement to the generation problem can complicate and enrich even the simple examples.

References

[1] Martin Gardner. On paradoxical situations that arise from nontransitive relations. *Scientific American*, 231(4):120–125, 1974.

[2] I. P. Goulden and D. M. Jackson. An inversion theorem for cluster decompositions of sequences with distinguished subsequences. *Journal of the London Mathematical Society*, 20:567–576, 1979.

[3] I. P. Goulden and D. M. Jackson. *Combinatorial Enumeration*. John Wiley & Sons, 1983.

[4] Leo J. Guibas. Periodicities in strings. In Alberto Apolstolico and Zvi Galil, editors, *Combinatorial Algorithms on Words*, pages 257–269. Springer-Verlag, 1984.

[5] Leo J. Guibas and Andrew M. Odlyzko. Periods in strings. *Journal of Combinatorial Theory A*, 30:19–42, 1981.

[6] Leo J. Guibas and Andrew M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory A*, 30:183–208, 1981.

[7] Ki Hang Kim, Mohan S. Putcha, and Fred W. Rousch. Some combinatorial properties of free semigroups. *Journal of the London Mathematical Society*, (2) 16:397–402, 1977.

[8] Andrew M. Odlyzko. Enumeration of strings. In Alberto Apostolico and Zvi Galil, editors, *Combinatorial Algorithms on Words*, pages 205–228. Springer-Verlag, 1984.

[9] George Pólya and Gabor Szegö. *Problems and Theorems in Analysis*, volume II. Springer-Verlag, 1976.

[10] Howard Straubing. Applications of the theory of automata in enumeration. *Discrete Mathematics*, 64:269–279, 1987.

[11] Herbert S. Wilf. *Generatingfunctionology*. Academic Press, Inc., 1990.

[12] Doron Zeilberger. Enumerating words by their number of mistakes. *Discrete Mathematics*, 34:89–91, 1981.

[13] Doron Zeilberger. Enumerating totally clean words. *Discrete Mathematics*, 64:313–315, 1987.

A Appendix: A Special Case Construction

In the main body of this paper, we classified $A \in \Sigma_2^*$ such that, when $q = 2$, $\mathcal{F}(n; A)$ either had infinite parity problems, or $\mathcal{F}(n; A)$ had no parity problems. Further, for most of the A for which $\mathcal{F}(n; A)$ had no parity problems, Gray codes for $\mathcal{F}(n; A)$ were constructed for all $n \geq 0$. When A is periodic with minimal base 01 or 10, we proved that $\mathcal{F}(n; A)$ had no parity problems, but we did not provide the Gray code construction. In this appendix, we remedy the situation by supplying the construction.

We continue to use $A = a_0 \ldots a_{m-1}$ to denote the excluded substring, where m denotes the length of A. For this appendix, we assume that A is periodic with minimal base 10. Hence, $m \geq 4$. If the minimal base of A were 01, then a Gray code for $\mathcal{F}(n; A)$ is found by reversing all of the strings on a Gray code for $\mathcal{F}(n; A^R)$.

We actually construct Gray codes for the set $\mathcal{H}(n; C)$, thus implying Gray codes for $\mathcal{F}(n) = \mathcal{H}(n; \epsilon)$. As in Section 4, the construction of Gray codes for $\mathcal{H}(n; C)$ has three stages: when $n \leq m$, when $m < n \leq m + 2$, and when $n > m + 2$. Gray codes for $\mathcal{H}(n; C)$ when $n \leq m$ are constructed in Section A.1. Gray codes for the latter two stages are described in Section A.2.

A.1 Gray codes when $n \leq m$

In this section, we construct Gray codes for $\mathcal{H}(n; C)$ for all C with no A-factor and all $n \leq m$. Lemma A.1, below, provides most of the Gray codes. This lemma is analogous to Theorem 4.1 of Section 4.

Lemma A.1 *Fix $q = 2$, $m \geq 4$ even, and let A be the periodic string of length m with minimal base* 10. *Let n be in the range $4 \leq n \leq m$. There exist strings S_n, T_n, V_n, W_n, X_n, Y_n, and Z_n such that for any $C \in \Sigma_2^*$ with no A-factor,*

1. *If $r(C) = 1$ and $n \geq 5$, then there exists a Gray code for $\mathcal{H}(n; C)$ containing the path $[V_n, W_n, X_n]$. When $h^*(n; C) = 0$, the Gray code is cyclic. Otherwise, $|h^*(n; C)| = 1$, and the Gray code starts at S_n and ends at T_n.*
2. *If $r(C) = 2$ and $n \geq 4$, then there exists a Gray code for $\mathcal{H}(n; C)$ containing the path $[Y_n, W_n, Z_n]$. When $h^*(n; C) = 0$, the Gray code is cyclic. Otherwise, $|h^*(n; C)| = 1$, and the Gray code starts at S_n and ends at T_n.*
3. *If $r(C) > 2$ and $n \geq 4$, then there exists a Gray code for $\mathcal{H}(n; C)$ containing the path $[V_n, W_n, Y_n]$. When $h^*(n; C) = 0$, the Gray code is cyclic. Otherwise, $|h^*(n; C)| = 1$, and the Gray code starts at S_n and ends at T_n.*

Proof. Let $C_0 = C * 0$ and $C_1 = C * 1$. We argue by induction on n using the strings

$$
\begin{array}{lcl@{\qquad\qquad}lcl}
S_n & = & a_{m-n} \ldots \overline{a}_{m-1} & X_n & = & 1101^{n-3} \\
T_n & = & a_{m-n} \ldots \overline{a}_{m-2} a_{m-1} & Y_n & = & 01^{n-1} \\
V_n & = & 101^{n-2} & Z_n & = & 11101^{n-4} \\
W_n & = & 1^n & & &
\end{array}
$$

Basis. The basis includes when $n = 4$ and $r(C) > 1$, when $n = 5$ and $2 \leq r(C) \leq 3$, and when $n = 6$ and $r(C) = 3$. First consider when $n = 4$. If $r(C) > 4$, then

$r(C) > 4$	$r(C) = 4$	$r(C) = 3$ $A \neq 1010$	$r(C) = 3$ $A = 1010$	$r(C) = 2$
0000	$\underline{1011}$	0000	$\underline{1011}$	0000
0010	$\underline{1111}$	0010	$\underline{1111}$	0100
0110	$\underline{0111}$	0110	$\underline{0111}$	1100
1110	0011	1110	0110	1101
1010	0001	1010	1110	0101
1000	1001	1000	1100	0001
1100	1101	1100	1101	0011
0100	0101	1101	1001	$\underline{0111}$
0101	0100	1001	0001	$\underline{1111}$
1101	0000	$\underline{1011}$	0011	$\underline{1110}$
1001	0010	$\underline{1111}$	0010	0110
$\underline{1011}$	0110	$\underline{0111}$	0000	0010
$\underline{1111}$	1110	0011	1000	
$\underline{0111}$	1100	0001		
0011	1000			
0001				

Table 1: Gray codes for $\mathcal{H}(4;C)$ in basis

$\mathcal{H}(4;C) = \Sigma_2^4$. If $r(C) = 4$, then $\mathcal{H}(4;C) = \Sigma_2^4 \backslash \{1010\}$. Suppose that $r(C) = 3$. If $A \neq 1010$, then $\mathcal{H}(4;C) = \Sigma_2^4 \backslash \{010b\}$, where $010b$ represents the strings in Σ_2^4 with prefix 010. If $A = 1010$, then $\mathcal{H}(4;C) = \Sigma_2^4 \backslash \{010b, 1010\}$. Finally, when $r(C) = 2$, $\mathcal{H}(4;C) = \Sigma_2^4 \backslash \{10bb\}$. Whenever $|h^*(4;C)| = 1$, the starting and ending points of the Gray code must be 1011 and 1000. Gray codes for these cases are given in Table 1. The strings on the specified subpath are underlined.

Next consider when $n = 5$. Note that $m \geq 6$ because $m \geq n = 5$ and m is even. If $r(C) = 2$, then $\mathcal{H}(5;C) = \Sigma_2^5 \backslash \{10bbb\}$. If $r(C) = 3$, then $\mathcal{H}(5;C) = \Sigma_2^5 \backslash \{010bb\}$. Gray codes for $\mathcal{H}(5;C)$ when $r(C) = 2, 3$ are given in Table 2.

We leave the case when $n = 6$ and $r(C) = 3$ as an exercise for the reader. Again, $m \geq n = 6$. One must separately consider when $A = 101010$ and when $A \neq 101010$.

For any $k \leq n$,

$$\mathcal{H}(n;C) = \bigcup_{D \in \mathcal{H}(n-k;C)} D * \mathcal{H}(k; C * D). \tag{14}$$

The inductive steps use equation (14) when $k = n - 1$ and when $k = n - 2$.

Inductive step for Part 1. Let $C \in \Sigma_2^*$ be a string with no A-factor and with $r(C) = 1$. Let n satisfy $5 \leq n \leq m$. Then $\mathcal{H}(n;C) = 1 * \mathcal{H}(n-1;C_1)$, and $r(C_1) = m - 1 > 2$. By induction, there is a Gray code $\mathbf{L}_1$ for $\mathcal{H}(n-1;C_1)$ containing the path $[V_{n-1}, W_{n-1}, Y_{n-1}] = [101^{n-3}, 1^{n-1}, 01^{n-2}]$. Note that $\mathbf{L} = 1 * \mathbf{L}_1$ contains the path $[1101^{n-3}, 1^n, 101^{n-2}] = [X_n, W_n, V_n]$. If $\mathbf{L}_1$ is cyclic, then $\mathbf{L}$ is cyclic and we are done. Otherwise, $h^*(n;C) = 1$. In this case, $|h^*(n-1;C_1)| = 1$, and $\mathbf{L}_1$ is between S_{n-1} and T_{n-1} by induction. By Lemma 4.2, C_1 ends with $a_0 \ldots a_{m-n}$, so $a_{m-n} = 1$. Thus, $\mathbf{L}$

$r(C)=2$			$r(C)=3$		
00000	$\underline{01111}$	01010	00000	11101	10010
00001	$\underline{11111}$	01011	00001	11100	10011
00011	$\underline{11101}$	01001	00011	10100	10001
00010	11100	01000	00010	10101	11001
00110	11000		00110	$\underline{10111}$	11011
00111	11001		00111	$\underline{11111}$	11010
00101	11011		00110	$\underline{01111}$	11000
00100	11010		00100	01110	10000
01100	11110		01100	11110	
01101	01110		01101	10110	

Table 2: Gray codes for $\mathcal{H}(5;C)$ in basis

starts and ends on the correct strings.

Inductive step for Part 2. Let $C \in \Sigma_2^*$ be a string with no A-factor and with $r(C)=2$. Let n satisfy $6 \leq n \leq m$. Then

$$\mathcal{H}(n;C) = 0 * \mathcal{H}(n-1;C_0) \quad \bigcup \quad 1 * \mathcal{H}(n-1;C_1),$$

$r(C_0) = m > 2$, and $r(C_1) = 1$. Inductively, there is a Gray code $\mathbf{L}_0$ for $\mathcal{H}(n-1;C_0)$ containing the path $[V_{n-1}, W_{n-1}, Y_{n-1}] = [101^{n-3}, 1^{n-1}, 01^{n-2}]$, and a Gray code $\mathbf{L}_1$ for $\mathcal{H}(n-1;C_1)$ containing the path $[V_{n-1}, W_{n-1}, X_{n-1}] = [101^{n-3}, 1^{n-1}, 110^{n-4}]$. A Gray code $\mathbf{L}$ for $\mathcal{H}(n;C)$ can be formed by joining $0 * \mathbf{L}_0$ and $1 * \mathbf{L}_1$ using the path $[V_{n-1}, W_{n-1}]$. The resulting Gray code $\mathbf{L}$ contains the path $[01^{n-1}, 1^n, 11101^{n-4}] = [Y_n, W_n, Z_n]$. If both $\mathbf{L}_0$ and $\mathbf{L}_1$ are cyclic, then $\mathbf{L}$ is cyclic. Otherwise, $|h^*(n-1; C * a_{m-n})| = 1$ and $h^*(n-1; C * \overline{a}_{m-n}) = 0$ by Corollary 4.1. Thus, $\mathbf{L}_{\overline{a}_{m-n}}$ is cyclic, and $\mathbf{L}_{a_{m-n}}$ is between S_{n-1} and T_{n-1}. The resulting $\mathbf{L}$ is between S_n and T_n.

Inductive step for Part 3. Let $C \in \Sigma_2^*$ be a string with no A-factor and with $r(C) > 2$. Let n satisfy $5 \leq n \leq m$. First consider when $r(C) = 3$ and $n > 6$ ($n = 5, 6$ were examined in the basis). Let $C_{00} = C * 00$, and similarly define C_{01}, C_{10}, and C_{11}. Since $m \geq n > 6$, we know $r(C_{01}) = 1$, and $r(C_{ij}) > 2$ for $ij \neq 01$. There exist Gray codes $\mathbf{L}_{ij}$ for $\mathcal{H}(n-2; C_{ij})$, where $\mathbf{L}_{01}$ contains the path $[V_{n-2}, W_{n-2}, X_{n-2}] = [101^{n-4}, 1^{n-2}, 1101^{n-5}]$, and, for $ij \neq 01$, $\mathbf{L}_{ij}$ contains the path $[V_{n-2}, W_{n-2}, Y_{n-2}] = [101^{n-4}, 1^{n-2}, 01^{n-3}]$.

The individual Gray codes can be joined as in Figure A.1. We first form the subpath $[01^{n-1}, 1^n, 101^{n-2}, 001^{n-2}]$. We then use the outer two strings on the specified subpaths of the $\mathbf{L}_{ij}$ to finish connecting the $\mathbf{L}_{ij}$. During this process, we must use the paths $[011^{n-2}, 011101^{n-5}]$ and $[01101^{n-4}, 11101^{n-4}]$. This ensures that the rest of the connections can be made. If $h^*(n;C) = 0$, then $h^*(n-2;C_{ij}) = 0$ for all ij by repeated applications of Corollary 4.1. Thus, all of the $\mathbf{L}_{ij}$ are cyclic, and the resulting $\mathbf{L}$ is cyclic. By the same corollary, if $|h^*(n;C)| = 1$, then $|h^*(n-2;C_{xy})| = 1$, where $xy = a_{m-n}a_{m-n+1}$, and $h^*(n-2;C_{ij}) = 0$ for $ij \neq xy$. Then $\mathbf{L}_{xy}$ is between S_{n-2} and T_{n-2}, and $\mathbf{L}_{ij}$ is cyclic for $ij \neq xy$. In this case, $\mathbf{L}$ starts and ends on the correct

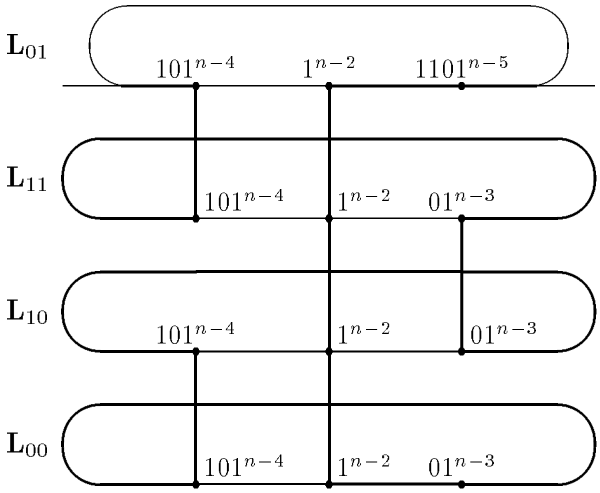


Figure 7: Constructing Gray code when $r(C) = 3$

strings. In Figure A.1, we illustrate the construction when $xy = 01$. We include paths for both cyclic and non-cyclic $\mathbf{L}_{xy}$.

Finally, consider when $r(C) \geq 4$ and $5 \leq n \leq m$. Then $r(C_0) \geq 3$ and $r(C_1) \geq 3$. There exist Gray codes $\mathbf{L}_0$ and $\mathbf{L}_1$ for $\mathcal{H}(n-1; C_0)$ and $\mathcal{H}(n-1; C_1)$, respectively. Each $\mathbf{L}_i$ contains the path $[V_{n-1}, W_{n-1}, Y_{n-1}] = [101^{n-3}, 1^{n-1}, 01^{n-2}]$. Join the $\mathbf{L}_i$ using the path $[V_{n-1}, W_{n-1}]$. This forms a Gray code $\mathbf{L}$ for $\mathcal{H}(n; C)$ containing the path $[V_n, W_n, Y_n]$. If both $\mathbf{L}_i$ are cyclic, then $\mathbf{L}$ is cyclic. Otherwise, as in the other cases, $\mathbf{L}$ is between the specified endpoints.

This completes the proof of the lemma. □

We can then obtain the following corollary of Lemma A.1.

Corollary A.1 *Fix $q = 2$, $m \geq 4$ even, and let A be the periodic string of length m with minimal base 10. For any n, $1 \leq n \leq m$, and for any $C \in \Sigma_2^*$ with no A-factor, there is a Gray code for $\mathcal{H}(n; C)$. When $h^*(n; C) = 0$, the Gray code is cyclic. Otherwise, its endpoints are independent of C and of equal parity.*

Proof. Most of the Gray codes are supplied by Lemma A.1. The possibilities not covered by Lemma A.1 are when $r(C) = 1$ and $n \leq 4$, or when $r(C) > 1$ and $n \leq 3$. In the former case, $\mathcal{H}(n; C) = 1 * \mathcal{H}(n-1; C_1)$, and $r(C_1) = m - 1 \geq 3$. So $\mathcal{H}(n; C) = 1 * \Sigma_2^{n-1}$ if either $m > 4$ or $n < 4$, and $\mathcal{H}(n; C) = 1 * (\Sigma_2^3 \setminus \{010\})$ if $n = m = 4$. There certainly exist Gray codes in these cases. In the latter case, if $r(C) > n$, then $\mathcal{H}(n; C) = \Sigma_2^n$. If $r(C) = k \leq n$, then

$$\mathcal{H}(n; C) = \{D \in \Sigma_2^n | D \text{ does not begin with } a_{m-k} \ldots a_{m-1}\}.$$

Again, Gray codes are easy to find. □

A.2 Gray codes when $n > m$

In this section, we prove the following results. These results are analogs to Lemma 4.5 and Theorem 4.4 of Section 4.

Lemma A.2 *Fix $q = 2$, $m \geq 4$ even, and let A be the periodic string of length m with minimal base 10. For all C with no A-factor and all n, $m + 1 \leq n \leq m + 2$, there exist a Gray code for $\mathcal{H}(n; C)$. When $h^*(n; C) = 0$, the Gray code is cyclic, and when $|h^*(n; C)| = 1$, the Gray code is between two strings independent of C and of equal parity.*

Theorem A.1 *Fix $q = 2$, $m \geq 4$ even, and let A be the periodic string of length m with minimal base 10. For $n \geq 1$, there exists a Gray code for $\mathcal{F}(n)$. When $f^*(n) = 0$, there exists a cyclic Gray code.*

We first prove Lemma A.2. We prove only the $n = m+1$ case, the $n = m+2$ proof being similar. Let $C \in \Sigma_2^*$ be a string with no A-factor. By Lemma 4.2, $h^*(m; C) = 0$ if and only if C ends in 10. Otherwise, $|h^*(m; C)| = 1$. Since C_1 cannot end in 10, $|h^*(m; C_1)| = 1$. Likewise, $h^*(m; C_0) = 0$ iff C ends in 1.

By Corollary A.1, there exist strings $S', T' \in \Sigma_2^m$, of equal parity, such that for any D with no A-factor and not ending in 10, there is a Gray code for $\mathcal{H}(m; D)$ between S' and T'. When $|h^*(m + 1; C)| = 1$, we prove there is a Gray code for $\mathcal{H}(m + 1; C)$ between $S = 1 * S'$ and $T = 1 * T'$.

First suppose $r(C) = 1$. By Corollary A.1, there is a Gray code $\mathbf{L}_1$ for $\mathcal{H}(m; C_1)$ between S' and T'. Then $\mathbf{L} = 1 * \mathbf{L}_1$ is a Gray code for $\mathcal{H}(m + 1; C)$ satisfying the conditions of the lemma.

Next suppose $r(C) = 2$, so $r(C_0) > 2$ and $r(C_1) = 1$. In this case, C ends in 10, so $|h^*(m; C_0)| = 1$ and $|h^*(m; C_1)| = 1$. By Corollary A.1, for $i = 0, 1$, there exists a Gray code $\mathbf{L}_i$ for $\mathcal{H}(n; C_i)$ between S' and T'. Then $0 * \mathbf{L}_0, 1 * \mathbf{L}_1$ is a cyclic Gray code for $\mathcal{H}(m + 1; C)$.

Finally, suppose $r(C) > 2$ so $r(C_i) > 1$ for $i = 0, 1$. By Lemma A.1, there exist strings W_m and Y_m such that, for $i = 0, 1$, there exists a Gray code $\mathbf{L}_i$ for $\mathcal{H}(m; C_i)$ containing the path $[W_m, Y_m]$. Since C_1 does not end in 10, $|h^*(m; C_1)| = 1$ and $\mathbf{L}_1$ runs between S' and T'. The Gray code $\mathbf{L}_0$ can either be cyclic (when C ends in 1) or between S' and T' (when C ends in 0 or $C = \epsilon$). If $\mathbf{L}_0$ is between S' and T', then $0 * \mathbf{L}_0, 1 * \mathbf{L}_1$ is a cyclic Gray code for $\mathcal{H}(m + 1; C)$. If $\mathbf{L}_0$ is cyclic, then $0 * \mathbf{L}_0$ can be spliced into $1 * \mathbf{L}_1$ using the path $[W_m, Y_m]$. In either case, the resulting Gray code satisfies the restrictions of the lemma.

This completes the construction of the Gray code for $\mathcal{H}(m + 1; C)$. A similar case by case analysis can be performed when $n = m + 2$. This finishes the proof of Lemma A.2.

We now prove Theorem A.1. As in Section 4.5, we use Lemma 4.1 to prove this theorem. Let $C \in \Sigma_2^*$ be any string with no A-factor. Corollary A.1 and Lemma A.2 provide Gray codes for $\mathcal{H}(n; C)$ when $1 \leq n \leq m + 2$. They provide cyclic Gray codes whenever $h^*(n; C) = 0$. Lemma 4.6 proves that $|h^*(m + 2; C)| = 1$ for any C with no A-factor, thus yielding Gray codes for $\mathcal{H}(m + 2; C)$ between two strings independent of C and of equal parity. We can then apply Lemma 4.1 to conclude that, for all $n \geq 1$ and for all C with no A-factor, there exists a Gray code for $\mathcal{H}(n;C)$. When $h^*(n;C) = 0$, there is a cyclic Gray code.

In this appendix, we have constructed Gray codes for $\mathcal{H}(n;C)$ when A is periodic with minimal base 10. This, in essence, completes the proof of Theorem 4.4 in Section 4 which stated that if A is periodic with a t.a. minimal base, then there exists a Gray code for $\mathcal{F}(n)$ for all $n \geq 1$.