More restrictive Gray codes for necklaces and Lyndon words

Vincent VAJNOVSZKI
LE2I - UMR CNRS, Université de Bourgogne
B.P. 47 870, 21078 DIJON-Cedex France

October 4, 2007

Abstract

In the last years, the order induced by the Binary Reflected Gray Code or its generalizations shown an increasing interest. In this note we show that the BRGC order induces a cyclic 2-Gray code on the set of binary necklaces and Lyndon words and a cyclic 3-Gray code on the unordered counterparts. This is an improvement and a generalization to unlabeled words of the result in [10, 13]; however an algorithmic implementation of our Gray codes remains an open problem.
Keywords: necklaces, Lyndon words, Gray codes.

1 Introduction

For a word $x = uv$, the word vu is called a rotation of x. A *necklace* is a word which is lexicographically minimal under the rotation; and a *Lyndon word* is an aperiodic necklace. An *unlabeled necklace* is a word which is lexicographically minimal under both rotation and permutation of alphabet symbols; and an *unlabeled Lyndon word* is an aperiodic unlabeled necklace. Generation of necklaces and Lyndon words lexicographically have been widely studied in the literature; see for instance [6, 7, 8, 10, 13] and the references therein. We restrict our attention to binary words and we denote by N_n, L_n, UN_n and UL_n the set of *binary* length n necklaces, Lyndon words, unlabeled necklaces and unlabeled Lyndon words, respectively. These four sets are related by the inclusions:

$$\begin{array}{ccc} UL_n & \subset & UN_n \\ \cap & & \cap \\ L_n & \subset & N_n \end{array}$$

A *k-Gray code* for a set of words S is an *ordered* list for S such that the Hamming distance between any two consecutive words in the list is at most k. If the distance between the last and the first words in the list is also bounded by k, then the Gray code is called cyclic; in addition if k is minimal, then we say that the Gray code is minimal.

In [9, 12] 2-Gray codes for binary necklaces with fixed density (given number of ones) are given and in [10, 13] are presented 3-Gray codes and generating algorithms for binary and k-ary necklaces and Lyndon words with no density restriction. In [12] it is shown that, in general, there is no 1-Gray code for binary necklaces or Lyndon words. Here we give 2-Gray codes for length n binary necklaces and Lyndon words and 3-Gray codes for their unlabeled counterparts.

In the last years, the order induced by the Binary Reflected Gray Code (BRGC) [5] or its generalizations shown an increasing interest [1, 2, 4, 10, 13]. For example, in [4] the authors use the BRGC order *twice*: firstly they define Gray-necklaces as binary strings minimal (under rotation) in BRGC order instead of lexicographic order, then they list them in BRGC order; the obtained list is almost a Gray code for necklaces in a non-standard representation. As in [10, 13], our approach here is based on the BRGC order but the obtained Gray codes for binary necklaces and Lyndon words are 2-Gray codes (and so minimal) and can be extended to unlabeled necklaces and Lyndon words; however they seem less appropriate for an algorithmic implementation.

2 The main result

The main result of this note is Theorem 1. We begin by recalling the definition of the BRGC order [5] and giving three technical 'facts' whose proof can be easily recovered by the reader.

Definition 1. *Let $a = a_1a_2\cdots a_n$ and $b = b_1b_2\cdots b_n$ be words in $\{0,1\}^n$ and i the rightmost position in which a and b differ. We say that a is less than b in* BRGC *(denoted by $a \prec b$) if $\sum_{j=i}^{n} a_j$ is even.*

Any set $X \subset \{0,1\}^n$ of length n binary words listed in $\prec$ order gives a *suffix partitioned list*, that is, all the words in X with a common suffix are contiguous in the list. An order relation with this property is called *genlex* (as generalized lexicographical) order [11]. See Table 1 for the sets N_6, L_6, UN_6 and UL_6 listed in $\prec$ order.

N_6	L_6	UN_6	UL_6
0 0 0 0 0 0		✓	
0 0 0 0 1 1	✓	✓	✓
0 1 1 0 1 1			
0 0 1 0 1 1	✓	✓	✓
0 0 1 1 1 1	✓		
1 1 1 1 1 1			
0 1 1 1 1 1	✓		
0 1 0 1 1 1	✓		
0 0 0 1 1 1	✓	✓	✓
0 0 0 1 0 1	✓	✓	✓
0 1 0 1 0 1		✓	
0 0 1 1 0 1	✓		
0 0 1 0 0 1		✓	
0 0 0 0 0 1	✓	✓	✓

Table 1: The set of length 6 binary necklaces, Lyndon words, unlabeled necklaces and unlabeled Lyndon words in BRGC order.

Fact 1. *Let $\alpha \in \{0,1\}^n$ and j, $1 \le j < n$. If α satisfies the following three conditions: (i) 0^j is a prefix of α (ii) α has no other 0^j factor, then α is a Lyndon word. If, in addition, α has no 1^j factors, then α is an unlabeled Lyndon word.*

For a binary word $\alpha \in \{0,1\}^n$ let $|\alpha|_1$ denote the number of 1's in α.

Fact 2. *For $\lambda \in \{0,1\}^n$ with $|\lambda|_1 \geq 1$ let λ' be the binary word obtained from λ by changing its leftmost 1 bit to 0.*

- *If $X \in \{N_n, UN_n\}$ and $\lambda \in X$ with $|\lambda|_1 \geq 1$, then $\lambda' \in X$.*
- *If $X \in \{L_n, UL_n\}$ and $\lambda \in X$ with $|\lambda|_1 \geq 2$, then $\lambda' \in X$.*

Fact 3. *In $\prec$ order*

- *for $n \geq 1$ the first word in N_n and in UN_n is 0^n;*
- *for $n \geq 3$ (resp. for $n \geq 4$) the first word in L_n (resp. in UL_n) is $0^{n-2}1^2$;*
- *for $n \geq 2$ the last word in N_n, UN_n, L_n and UL_n is $0^{n-1}1$.*

Lemma 1. *Let $X \in \{L_n, N_n, UL_n, UN_n\}$ and $\lambda \in X$. Suppose that r is the rightmost position where λ differs from its successor (*resp. *its predecessor) in $\prec$ order, assuming that λ is not the last (*resp. *first) word in X in this order. Then $|\lambda_1\lambda_2\ldots\lambda_{r-1}|_1 \leq 1$.*

Proof. By contradiction. Let r be the rightmost position where λ differs from μ, its successor. If $|\lambda_1\lambda_2\ldots\lambda_{r-1}|_1 \geq 2$, then by the definition of BRGC order and by Fact 2 one of the words λ' or $(\lambda')'$ belongs to X and is larger than λ and smaller than μ in $\prec$ order. The proof is similar when we consider the predecessor of λ. □

The next corollary is a 'weak' version of Theorem 1; its proof is very similar with the one of Theorem 1 in [10] and we omit it.

Corollary 1. *$\prec$ order induces a 3-Gray code on the sets L_n, N_n, UL_n and UN_n.*

Now we will prove that $\prec$ induces a more restrictive Gray code. Let α be a binary word with $|\alpha|_1 \geq 1$ and let j be the leftmost position where $\alpha_j = 1$. If $\alpha_{j+1} = 0$, then let define $\tilde{\alpha}$ as the binary word $\tilde{\alpha}_i = \alpha_i$, except $\tilde{\alpha}_j = 0$ and $\tilde{\alpha}_{j+1} = 1$. α and $\tilde{\alpha}$ have the shape given by:

$\alpha = 0 \ldots 0\ 1\ 0\ \alpha_{j+2} \ldots \alpha_n$

$\tilde{\alpha} = 0 \ldots 0\ 0\ 1\ \alpha_{j+2} \ldots \alpha_n$.

If $\alpha \in N_n$, then $\tilde{\alpha} \in L_n$. This statement is formalized in the first point of the next lemma. This result is not true for unlabeled words. Indeed, $010101 \in UN_6$ but $001101 \notin UN_6$ since the representative of 001101 is $001011 \in UN_6$. Similarly, $00101101 \in UL_8$ but $00011101 \notin UL_8$ since the representative of 00011101 is $00010111 \in UL_8$. The second point of the next lemma gives a more restrictive unlabeled counterpart of its first point.

Lemma 2. *Let $X \subset \{0,1\}^n$, $\alpha \in X$ with $|\alpha|_1 \geq 1$, and let j be the leftmost position where $\alpha_j = 1$, and suppose that $\alpha_{j+1} = 0$.*

1. *If $X \in \{L_n, N_n\}$, then $\tilde{\alpha} \in L_n$.*
2. *If $X \in \{UL_n, UN_n\}$ and $\alpha_{j+2} = 0$, then $\tilde{\alpha} \in UL_n$.*

Proof. The length $j-1$ prefix of α is 0^{j-1} and α has no 0^j factor. $\tilde{\alpha}$ has a unique 0^j factor which is its length j prefix, and so, by Fact 1, it is a Lyndon word and the first point holds.

If $X \in \{UL_n, UN_n\}$, then α has no 1^j factor. In addition, if $\alpha_{j+2} = 0$, then $\tilde{\alpha}$ has no 1^j factor too. Again, by Fact 1, $\tilde{\alpha}$ is an unlabeled Lyndon word, and the second point holds. □

Let $X \in \{L_n, N_n, UL_n, UN_n\}$, $\lambda \in X$. If λ differs from its successor in $\prec$ order in at least two positions, then the leftmost and the rightmost of these positions can not be arbitrarily far. This situation is formally stated in the next lemma and summarized in the table at the end of its proof.

Lemma 3. *Let $X \subset \{0,1\}^n$ and $\lambda \in X$. Suppose that λ differs from its successor in $\prec$ order in at least two positions (assuming that λ is not the last word in X in this order). Let ℓ and r be, respectively, the leftmost and the rightmost of these positions.*

1. *If $\lambda_\ell \neq \lambda_r$ and $X \in \{L_n, N_n, UL_n, UN_n\}$, then $r = \ell + 1$.*
2. *If $\lambda_\ell = \lambda_r$ and $X \in \{L_n, N_n\}$, then $r = \ell + 1$.*
3. *If $\lambda_\ell = \lambda_r$ and $X \in \{UL_n, UN_n\}$, then $r = \ell + 1$ or $r = \ell + 2$. In addition, if $r = \ell + 2$, then $\lambda_{\ell+1} \neq \lambda_\ell$.*

Proof. The proof is by contradiction supposing that $r - \ell > 1$ or $r - \ell > 2$, respectively. Let μ be the successor of λ in $\prec$ order.

Proof of 1. Let suppose that $r - \ell > 1$.
When $\lambda_\ell = 1$ and $\lambda_r = 0$, then $\mu_\ell = 0$ and $\mu_r = 1$, and by Lemma 1, $\lambda_i = 0$ for all i, $1 \leq i < \ell$ and $\ell < i < r$. Since $\lambda_\ell = 1$ and $\lambda_{\ell+1} = \lambda_{\ell+2} = 0$, applying Lemma 2 we have that the word $\tilde{\lambda}$ is in X. Moreover, it is easy to check that $\lambda \prec \tilde{\lambda} \prec \mu$, which is in contradiction with μ being the successor of λ in $\prec$ order.
Similarly, if $\lambda_\ell = 0$ and $\lambda_r = 1$, then $\mu_\ell = 1$ and $\mu_r = 0$. In this case, again by Lemma 1, $\mu_i = 0$ for all i, $1 \leq i < \ell$ and $\ell < i < r$. Since $\mu_\ell = 1$ and $\mu_{\ell+1} = \mu_{\ell+2} = 0$, applying Lemma 2 it results that $\tilde{\mu}$ is in X and $\lambda \prec \tilde{\mu} \prec \mu$.

Proof of 2. Let suppose that $r - \ell > 1$.
When $\lambda_\ell = \lambda_r = 1$, then $\mu_\ell = \mu_r = 0$, and $\lambda_i = 0$ for all i, $1 \leq i < \ell$ and $\ell < i < r$. Since $\lambda_\ell = 1$ and $\lambda_{\ell+1} = 0$, by the first point of Lemma 2, $\tilde{\lambda}$ is in X and $\lambda \prec \tilde{\lambda} \prec \mu$.
When $\lambda_\ell = \lambda_r = 0$, then $\mu_\ell = \mu_r = 1$, $\mu_{\ell+1} = 0$ and similarly $\tilde{\mu}$ is in X with $\lambda \prec \tilde{\mu} \prec \mu$.

The proof of 3 is analogous to the one of point 2 by considering $r - \ell > 2$. In addition, if $r - \ell = 2$, then $\lambda_{\ell+1} \neq \lambda_\ell$ and $\mu_{\ell+1} \neq \mu_\ell$.

L_n, N_n, UL_n, UN_n	L_n, N_n	UL_n, UN_n
$\lambda_\ell \neq \lambda_r$	$\lambda_\ell = \lambda_r$	$\lambda_\ell = \lambda_r$
$\Downarrow$	$\Downarrow$	$\Downarrow$
$r = \ell + 1$	$r = \ell + 1$	$r \leq \ell + 2$

□

Recall that a Gray code is a k-Gray code if successive words differ in at most k positions; and cyclic if the last and the first words differ in the same way. As a consequence of the previous lemma, together with Fact 3, we can state the main result of this note.

Theorem 1.

1. *BRGC order induces a cyclic 2-Gray code on N_n and L_n.*
2. *BRGC order induces a cyclic 3-Gray code on UN_n and UL_n.*

Moreover, if two successive words, in BRGC order, in N_n, L_n, UN_n or UL_n differ in more than one position, then these positions are consecutive.

3 Final remarks

The BRGC order $\prec$ induces a cyclic 2-Gray code on N_n and L_n, and, as mentioned in Introduction, it is minimal. For example, in the Gray code list for N_6 in Table 1 there are four 2-changes between successive words. We do not know if our Gray codes for N_n and L_n minimize the number of 2-changes between successive words.

Also, our Gray code for UN_n and UL_n are 3-Gray codes, and in the list for UN_6 in Table 1 there is one 3-changes, namely between the *6th* and the *7th* word. A more restrictive (with no 3-changes) list for UN_6 is $(000000, 000001, 010101, 000111, 000101, 001001, 001011, 000011)$; this shows that our Gray codes for UN_n and UL_n might be not minimal.

The main difference between the Gray codes presented here and the ones in [10] is that the latter ones are prefix partitioned lists. The algorithmic implementation of the suffix partitioned Gray codes presented in this paper seems more difficult. Indeed, the recursive definitions for N_n and L_n [3, Theorem 2] (on which is based the algorithm in [10]), and for UN_n and UL_n [3, Theorem 5] work only for prefix partitioned lists. This last remark suggests

Question 1 Can the Gray codes presented here be efficiently implemented? That is, can N_n, L_n, UN_n and UL_n be efficiently listed in BRGC order (defined in Definition 1)?

A natural generalization of the BRGC order to a k-ary alphabet is given by (see [13])

Definition Let $k, n > 0$ and $a = a_1 a_2 \cdots a_n$ and $b = b_1 b_2 \cdots b_n$ be words in $\{0, 1, \ldots, k-1\}^n$ and i the rightmost position in which a and b differ. We say that a is less than b in reflected Gray code order (denoted by $a \prec b$) if either

- $a_i < b_i$ and $\sum_{j=i+1}^{n} a_j = \sum_{j=i+1}^{n} b_j$ is even, or
- $a_i > b_i$ and $\sum_{j=i+1}^{n} a_j = \sum_{j=i+1}^{n} b_j$ is odd.

Question 2 Can the previous results be extended to k-ary alphabet? Experimental results show that the successor, in $\prec$ order, of the 3-ary length 6 Lyndon word 001002 is 0**121**02, and the successor of 012202 is 0**2**22**1**2; and those are the only consecutive (in $\prec$ order) 3-ary length 6 Lyndon words which do not differ in at most two consecutive positions.

Acknowledgment

The author would like to thank one of the anonymous referees for helpful suggestions which have improved the accuracy of this paper.

References

[1] J.L. Baril, V. Vajnovszki. Minimal change list for Lucas strings and some graph theoretic consequences, *Theoretical Computer Science*, **346** (2-3), 189-199, 2005.

[2] M.W. Bunder, K.P. Tognetti, G.E. Wheeler. On binary reflected Gray codes and functions, to appear in *Discrete Mathematics*.

[3] K. Cattell, F. Ruskey, J. Sawada, M. Serra, C.R. Miers. Fast algorithms to generate necklaces, unlabeled necklaces, and irreducible polynomials over GF(2), *Journal of Algorithms* **37**(2) (2000), 267–282.

[4] C. Degni, A.A. Drisko. Gray-ordered binary necklaces, *The Electronic Journal of Combinatorics*, 14 (2007), ♯ R7.

[5] F. Gray. Pulse code communication. U. S. Patent 2632058 (1953).

[6] F. Ruskey, C. Savage, T. Wang. Generating necklaces, *J. Algorithms*, **13** (1992) 414–430.

[7] F. Ruskey. *Combinatorial generation.* Book in preparation.

[8] J. Sawada, F. Ruskey. An efficient algorithm for generating necklaces with fixed density, *SIAM J. Computing*, **29** (1999), 671–684.

[9] T. Ueda. Gray codes for necklaces, *Discrete Math.* **219**(1-3) (2000), 235–248.

[10] V. Vajnovszki. Gray code order for Lyndon words, *Discrete Mathematics and Theoretical Computer Science*, **9**(2) (2007), 145–152.

[11] T. Walsh. Generating Gray codes in $O(1)$ worst-case time per word, *Proceeding of 4th Conference on Discrete Mathematics and Theoretical Computer Science*, Dijon, France, 7–12 July 2003, pp. 73–78. Springer-Verlag, Berlin.

[12] T.M.Y. Wang, C.D. Savage. A Gray code for necklaces of fixed density, *SIAM J. Discrete Math.* **9**(4) (1996), 654–673.

[13] M. Weston, V. Vajnovszki. Gray codes for necklaces and Lyndon words of arbitrary base, to appear in *Pure Mathematics and Applications/Algebra and Theoretical Computer Science.*