Paper

A construction of all CR sequences in the de Bruijn sequences of length 2^{2p+1} where p is a prime number

Hiroshi Fujisaki^{1 a)}

¹ Graduate School of Natural Science and Technology, Kanazawa University
Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan

^{a)} fujisaki@t.kanazawa-u.ac.jp

Received October 1, 2013; Revised December 2, 2013; Published April 1, 2014

Abstract: We partially solve a fundamental problem posed by Fredricksen on existence of CR (complement reverse) sequences in the de Bruijn sequences of length 2^{2p+1} ($p \geq 1$). For the case that p is a prime number, we construct the set of CR graphs, which yields all CR sequences. As an application of this result, we discuss enumeration of the total number of distinct auto-correlation functions for the set of de Bruijn sequences of length 2^{2p+1} . Since the worst cases of the normalized cross-correlation functions for pairs of de Bruijn sequences are characterized by the normalized auto-correlation functions for sequences of the worst pairs, we obtain upper bounds of the total number of distinct auto-correlation functions for the set of de Bruijn sequences of length 2^n ($n \geq 4$).

Key Words: de Bruijn sequences, CR (complement reverse) sequences, correlation functions, discretized Markov transformations

1. Introduction

Pseudo-random sequences have found extensive applications in cryptography and digital communication systems. Recently a novel application of pseudo-random sequences in LDPC (low-density parity-check) coding system has been proposed in [1]. In these systems, the performance depends upon correlational properties of the sequences. Moreover, from the view point of complexity, pseudo-random sequences appropriate for application in such systems are supposed to fulfill the property of full-length or maximal-period.

A LFSR (linear feedback shift register) has been widely used to generate full-length sequences. On the other hand, all the following full-length sequences can be generated by NFSR (nonlinear feedback shift register) but not by LFSR. In view of randomness in chaotic dynamics of one-dimensional ergodic transformations, sequences based on discretized Bernoulli transformations were proposed in [2] and [3]. The latter sequences have a great advantage in terms of their family size. For instance, for binary sequences of length 2^n ($n \geq 1$), while the total number $\varphi(2^n - 1)/n$ of the former sequences is less than $2^n/n$, where φ is called the Euler function, the total number of the celebrated de Bruijn sequences is known to be $2^{2^{n-1}-n}$.

Motivated by the sequences proposed in [2] and [3], we generally defined discretized Markov transformations and found an algorithm to give the total number of full-length sequences based on discretized Markov transformations in [4]. The discretized Markov transformations, which can be regarded as examples of *ultradiscrete* dynamical systems [5], are permutations of subintervals in Markov partitions determined from the underlying transformations. From this viewpoint, de Bruijn sequences are merely special examples of full-length sequences in the discretized Markov transformations.

In [6], we defined the piecewise-monotone-increasing Markov transformations and gave the bounded monotone truth-table algorithm for generating *all* full-length sequences which are based on the discretized piecewise-monotone-increasing Markov transformations.

In light of the results [6], we can freely construct all full-length sequences, including all de Bruijn sequences, which are based on the discretized piecewise-monotone-increasing Markov transformations. Unfortunately, however, we know little of the statistical properties of full-length sequences which are based on the discretized transformations.

With the help of linearity, the algebraic structure of LFSR enables us to evaluate the correlational properties of full-length sequences based on the LFSR. On the other hand, because of the nature of nonlinearity, it is intractable to characterize the correlational properties of full-length sequences based on the discretized piecewise-monotone-increasing Markov transformations. Even for de Bruijn sequences, only bounds of the maximum values of the normalized auto-correlation functions have been known [7].

The problem of finding a family of good full-length sequences in terms of the correlational properties is not only mathematically challenging but also practically important as pointed out in the beginning of the Introduction. As the first step, we focus on a fundamental example of such full-length sequences, namely the de Bruijn sequences.

In the previous research [8], we have provided a novel lower bound of the minimum values of the normalized *auto*-correlation functions for de Bruijn sequences of length 2^n ($n \geq 3$). The lower bound is tight in the sense that the equality holds for $n = 3$ and $n = 4$. In this research, we consider the worst cases of the normalized *cross*-correlation values of the de Bruijn sequences. The cross-correlation properties of full-length sequences are at least as important as the auto-correlation properties in the systems where the sequences are used.

This report is composed of five sections. Considering the bounds of the normalized cross-correlation values of the de Bruijn sequences, we immediately observe that the worst cases of the normalized cross-correlation functions for pairs of de Bruijn sequences are characterized by the normalized auto-correlation functions for sequences of the worst pairs. In order to designate such worst cases in advance, we discuss enumeration of the total number of distinct auto-correlation functions for the set of de Bruijn sequences of length 2^n ($n \geq 4$). In Sect. 2, we introduce the reverse of sequence to study this enumeration problem. Then we give a general formula for the normalized auto-correlation function for a sequence to be invariant under permutations of the sequence. In the previous research closely related to the reverse of sequences, it was pointed out in [9] that for $n = 5$, there exist CR (complement reverse) sequences in the de Bruijn sequences of length 2^n . Naturally the following problem was posed by Fredricksen in [9]: Show that there exists a CR sequence whenever $n (\geq 3)$ is odd. This problem has often been discussed. For instance, in [10], a characterization of CR sequences were presented. In Sect. 3, we partially solve the Fredricksen's problem. For the case that p is a prime number, we construct the set of CR graphs, which yields all CR sequences. In Sect. 4, we obtain upper bounds of the total number of distinct auto-correlation functions for the set of de Bruijn sequences of length 2^n in the case that $n (\geq 4)$ is even; and the case that $n = 2p + 1$, where p is any prime number. The report concludes with the summary in Sect. 5.

2. Preliminaries

The correlation functions for sequences are measures of the similarity, or relatedness, between two sequences. Mathematically they are defined as follows.

Definition 1 The *cross-correlation function* of time delay ℓ for the sequences $\mathbf{X} = (X_i)_{i=0}^{N-1}$ and

$\mathbf{Y} = (Y_i)_{i=0}^{N-1}$ over $\{-1, 1\}$ is defined by

$$R_N(\ell; \mathbf{X}, \mathbf{Y}) = \sum_{i=0}^{N-1} X_i Y_{i+\ell \pmod{N}},$$

where $\ell = 0, 1, \dots, N-1$ and, for integers a and $b (\geq 1)$, $a \pmod{b}$ denotes the least residue of a to modulus b . The *normalized cross-correlation function* of time delay ℓ for the sequences $\mathbf{X}$ and $\mathbf{Y}$ is defined by

$$r_N(\ell; \mathbf{X}, \mathbf{Y}) = \frac{1}{N} \sum_{i=0}^{N-1} X_i Y_{i+\ell \pmod{N}}.$$

If $\mathbf{X} = \mathbf{Y}$, we call $R_N(\ell; \mathbf{X}, \mathbf{X})$ and $r_N(\ell; \mathbf{X}, \mathbf{X})$ the *auto-correlation function* and the *normalized auto-correlation function*, and simply denote them by $R_N(\ell; \mathbf{X})$ and $r_N(\ell; \mathbf{X})$, respectively.

In this study, we are concerned with correlational properties of the de Bruijn sequences. As we see below, a de Bruijn sequence is usually defined as a sequence over $\{0, 1\}$ while the correlation functions are defined for a sequence over $\{-1, 1\}$. Throughout this report, when we compute the values of the normalized cross-correlation functions $r_N(\ell; \mathbf{X}, \mathbf{Y})$ for de Bruijn sequences $\mathbf{X}$ and $\mathbf{Y}$, we regard 0 in the de Bruijn sequences as -1 . In other words, we transform de Bruijn sequence $\mathbf{X}$ and $\mathbf{Y}$ of length N over $\{0, 1\}$ to sequences of length N over $\{-1, 1\}$ by one-to-one correspondence between 0 and -1 , respectively.

To treat a *time-reversal* of sequences, we introduce

Definition 2 For a sequence $\mathbf{X} = (X_i)_{i=0}^{N-1}$ over $\{-1, 1\}$, the *reverse* ${}^r\mathbf{X}$ of $\mathbf{X}$ is defined by ${}^r\mathbf{X} = (X_i)_{i=N-1}^0$.

In Sect. 4, we discuss enumeration of the total number of distinct auto-correlation functions for the set of de Bruijn sequences of length $N = 2^n$ ($n \geq 4$). To study this enumeration problem, we need to know which pair of two sequences $\mathbf{X}$ and $\mathbf{Y}$ satisfy the condition $R_N(\ell; \mathbf{X}) = R_N(\ell; \mathbf{Y})$ for every $\ell = 0, 1, \dots, N-1$. Recalling de Bruijn sequences are examples of full-length sequences in the discretized Markov transformations, let $\mathcal{T}$ be the set of all full-length sequences of length N over $\{0, 1\}$ based on a discretized Markov transformation. For each pair $\mathbf{X}$ and $\mathbf{Y}$ of full-length sequences in $\mathcal{T}$, there exists a permutation transformation $P : \{0, 1\}^N \rightarrow \{0, 1\}^N$ such that $\mathbf{Y} = P(\mathbf{X})$. Thus, given $\mathbf{X} \in \mathcal{T}$, what we have to do is to specify permutation transformations P which satisfy the condition $R_N(\ell; \mathbf{X}) = R_N(\ell; P(\mathbf{X}))$ for all $\ell = 0, 1, \dots, N-1$.

Since $R_N(0; \mathbf{X}, \mathbf{Y})$ is nothing but an inner product of $\mathbf{X}$ and $\mathbf{Y}$ in a vector space $\mathcal{V} = \mathbb{R}^N$, denoting $R_N(0; \mathbf{X}, \mathbf{Y})$ simply by $(\mathbf{X}, \mathbf{Y})$, we consider an inner product space $(\mathcal{V}, (\cdot, \cdot))$. The transformation $S : \mathcal{V} \rightarrow \mathcal{V}$ is called the *shift transformation* on $\mathcal{V}$ if $(\mathbf{X}, S(\mathbf{Y})) = R_N(1; \mathbf{X}, \mathbf{Y})$. Defining $S^0(\mathbf{Y}) = \mathbf{Y}$ and $S^\ell(\mathbf{Y}) = S(S^{\ell-1}(\mathbf{Y}))$ for $\ell = 1, \dots, N-1$, we have $(\mathbf{X}, S^\ell(\mathbf{Y})) = R_N(\ell; \mathbf{X}, \mathbf{Y})$ for $\ell = 0, 1, \dots, N-1$. In this general setting, we obtain

Lemma 1 For any $\mathbf{X} \in \mathcal{V}$, $(P(\mathbf{X}), S(P(\mathbf{X}))) = (\mathbf{X}, S(\mathbf{X}))$ if and only if there exists k ($0 \leq k \leq N-1$) such that $P = S^k$ or $P = S^k \circ P_r = P_r \circ S^k$, where P_r is the permutation transformation which converts $\mathbf{X}$ into the reverse ${}^r\mathbf{X}$, namely $P_r(\mathbf{X}) = {}^r\mathbf{X}$.

Proof: The sufficiency is obvious. We shall prove the necessity. Suppose $(P(\mathbf{X}), S(P(\mathbf{X}))) = (\mathbf{X}, S(\mathbf{X}))$ for $\mathbf{X} \in \mathcal{V}$. We use σ to denote the permutation of the index set $\{0, 1, \dots, N-1\}$ associated with P , which is defined by $P(\mathbf{X}) = (X_{\sigma(0)}, X_{\sigma(1)}, \dots, X_{\sigma(N-1)})$.

Since $(P(\mathbf{X}), S(P(\mathbf{X}))) = (\mathbf{X}, P^{-1}(S(P(\mathbf{X}))))$, we obtain $(\mathbf{X}, (P^{-1} \circ S \circ P - S)(\mathbf{X})) = 0$, where $(P^{-1} \circ S \circ P - S)(\mathbf{X}) = P^{-1}(S(P(\mathbf{X}))) - S(\mathbf{X})$. Noting $\delta_{j, i+1 \pmod{N}} = \delta_{j, \sigma^{-1}(\sigma(i)+1 \pmod{N})} = 0$ for $j = i$, we have

$$\sum_{\substack{0 \leq i, j \leq N-1 \\ i \neq j}} \{(\delta_{j, \sigma^{-1}(\sigma(i)+1 \pmod{N})} - \delta_{j, i+1 \pmod{N}}) + (\delta_{i, \sigma^{-1}(\sigma(j)+1 \pmod{N})} - \delta_{i, j+1 \pmod{N}})\} X_i X_j = 0,$$

where $\delta_{j,i}$ is the Kronecker δ ; $\delta_{j,i} = 1$ if $i = j$ and $\delta_{j,i} = 0$ if $i \neq j$. Thus we obtain

$$\sum_{i=0}^{N-1} (\delta_{i+1 \pmod N, \sigma^{-1}(\sigma(i)+1 \pmod N)} + \delta_{i, \sigma^{-1}(\sigma(i+1 \pmod N)+1 \pmod N)} - 1) X_i X_{i+1 \pmod N} = 0. \quad (1)$$

Since $\mathbf{X} \in \mathcal{V}$ is arbitrarily chosen, (1) holds if and only if, for each $i = 0, 1, \dots, N-1$, either $\delta_{i+1 \pmod N, \sigma^{-1}(\sigma(i)+1 \pmod N)} = 1$ or $\delta_{i, \sigma^{-1}(\sigma(i+1 \pmod N)+1 \pmod N)} = 1$ holds but not the both.

If $\delta_{i+1 \pmod N, \sigma^{-1}(\sigma(i)+1 \pmod N)} = 1$ holds for some i ($0 \leq i \leq N-1$), then (1) implies that $\delta_{i+1 \pmod N, \sigma^{-1}(\sigma(i)+1 \pmod N)} = 1$ holds for all $i = 0, 1, \dots, N-1$, equivalently there exists k ($0 \leq k \leq N-1$) such that $P = S^k$. On the other hand, if $\delta_{i, \sigma^{-1}(\sigma(i+1 \pmod N)+1 \pmod N)} = 1$ holds for some i , then (1) implies that there exists k ($0 \leq k \leq N-1$) such that $P = S^k \circ P_r = P_r \circ S^k$, which completes the proof. $\square$

We shall use the following direct consequence of Lemma 1.

Corollary 1 For any $\mathbf{X} = (X_i)_{i=0}^{N-1}$ over $\mathbb{R}$, the normalized auto-correlation function $r_N(\ell; \mathbf{X})$ satisfies

$$r_N(\ell; \mathbf{X}) = r_N(\ell; {}^r\mathbf{X}), \quad 0 \leq \ell \leq N-1.$$

In order to study the above-mentioned enumeration problem, we also need to consider a multiplication of $\mathbf{X} \in \mathcal{V}$ by scalars $c \in \mathbb{R}$. Immediately we see

Observation 1 For $\mathbf{X} \in \mathcal{V}$ and $c \in \mathbb{R}$, $(c\mathbf{X}, S(c\mathbf{X})) = (\mathbf{X}, S(\mathbf{X}))$ if and only if $c = \pm 1$.

In this research, the focus is on the de Bruijn sequences, which are typical examples of full-length sequences in the discretized Markov transformations as stated in the Introduction. The de Bruijn sequences can be defined in terms of the discretized Markov transformations [4]. However, we here simply define them irrespective of the discretized Markov transformations as follows.

A *binary word* is a finite binary sequence. A word of length k is called a *k-word*.

A (*binary*) *cycle* of length k is a sequence of binary k -word $a_1 a_2 \dots a_k$ taken in a circular order. In the cycle $a_1 a_2 \dots a_k$, a_1 follows a_k , and $a_2 \dots a_k a_1, \dots, a_k a_1 \dots a_{k-1}$ are all the same cycle as $a_1 a_2 \dots a_k$. Two sequences $\mathbf{X} = (X_i)_{i=0}^{N-1}$ and $\mathbf{Y} = (Y_i)_{i=0}^{N-1}$ are said to be *equivalent*, in symbols $\mathbf{X} \simeq \mathbf{Y}$, if $\mathbf{X}$ and $\mathbf{Y}$ are the same cycle.

A (*binary*) *complete cycle* of length 2^n is a cycle of binary 2^n -words, such that the 2^n possible ordered sequences of binary n -words of that cycle are all different. Any binary n -word occurs exactly once in the complete cycle.

Example 1 We give examples of complete cycles of length 2^n :

$$\begin{aligned} n = 1, & \quad 01, \\ n = 2, & \quad 0011, \\ n = 3, & \quad 00010111, \quad 00011101. \end{aligned}$$

Because of the following theorem, the complete cycles are sometimes called de Bruijn sequences.

Theorem 1 (de Bruijn [11], Flye Sainte-Marie [12]) For each positive integer n , there are exactly $2^{2^{n-1}-n}$ complete cycles of length 2^n .

3. A construction of all CR sequences in the de Bruijn sequences

For $a \in \{0, 1\}$, we use $\bar{a}$ to denote the binary complement of a , i.e. $\bar{0} = 1$ and $\bar{1} = 0$. To simplify notations, we write $N = 2^n$ ($n \geq 1$).

Noting the complement $\bar{\mathbf{X}}$ of $\mathbf{X}$ over $\{0, 1\}$ corresponds to the scalar multiplication $-\mathbf{Y}$ of $\mathbf{Y}$ over $\{-1, 1\}$, Observation 1 readily leads to

Remark 1 Let $\mathbf{X}$ be a de Bruijn sequence of length 2^n , then

$$r_N(\ell; \mathbf{X}) = r_N(\ell; \overline{\mathbf{X}}), \quad 0 \leq \ell \leq N - 1.$$

Together with a general formula for $r_N(\ell; \mathbf{X})$ in Lemma 1, taking account of the equivalence relation which arises from the cycle, for a de Bruijn sequence $\mathbf{X}$ of length 2^n , we obtain

$$r_N(\ell; \mathbf{X}) = r_N(\ell; {}^r\mathbf{X}) = r_N(\ell; \overline{\mathbf{X}}) = r_N(\ell; {}^r\overline{\mathbf{X}}), \quad 0 \leq \ell \leq N - 1. \quad (2)$$

Now we turn to consider cross-correlation functions of de Bruijn sequences. By Theorem 1, the cross-correlation functions of de Bruijn sequences of length 2^n can be defined for $n \geq 3$. As mentioned in the Introduction, for de Bruijn sequences, the bounds of auto-correlation functions were already clarified in [8]. The following general bounds of cross-correlation functions of de Bruijn sequences were obtained in [7]. We start with this result.

Theorem 2 (Zhang and Chen [7]) If $\mathbf{X}$ and $\mathbf{Y}$ are distinct de Bruijn sequences of length 2^n ($n \geq 3$), then

$$-1 \leq r_N(\ell; \mathbf{X}, \mathbf{Y}) \leq 1 - \frac{4}{2^n}, \quad 0 \leq \ell \leq N - 1.$$

The equality on the left-hand side holds if and only if $\mathbf{Y} = \overline{\mathbf{X}}$ and $\ell = 0$.

The conditions that the equality holds for the lower bound in Theorem 2 characterize the worst pair of de Bruijn sequences in terms of the cross-correlation functions and the worst case of the cross-correlation functions for the pairs. Based on the conditions, we refer to the pairs $(\mathbf{X}, \overline{\mathbf{X}})$ of de Bruijn sequences as the worst pairs. Thus, from (2), we can generally obtain

Observation 2 Let $\mathbf{X}$ be a de Bruijn sequence of length 2^n ($n \geq 3$), then

$$\begin{aligned} r_N(\ell; \mathbf{X}, \overline{\mathbf{X}}) &= r_N(\ell; \overline{\mathbf{X}}, \mathbf{X}) = r_N(\ell; {}^r\mathbf{X}, {}^r\overline{\mathbf{X}}) = r_N(\ell; {}^r\overline{\mathbf{X}}, {}^r\mathbf{X}) \\ &= -r_N(\ell; \mathbf{X}) = -r_N(\ell; {}^r\mathbf{X}) = -r_N(\ell; \overline{\mathbf{X}}) = -r_N(\ell; {}^r\overline{\mathbf{X}}), \quad 0 \leq \ell \leq N - 1. \end{aligned}$$

Taking this observation into account, Theorem 2 implies that the worst cases of the normalized cross-correlation functions for pairs of de Bruijn sequences are characterized by the normalized auto-correlation functions for sequences of the worst pairs.

Our goal is to construct a good family of de Bruijn sequences in terms of the correlational properties. To this end, it is preferable to know in advance the total number of distinct auto-correlation functions for the set of de Bruijn sequences of length 2^n ($n \geq 1$). By virtue of Theorem 1, we have $2^{2^{n-1}-n}$ de Bruijn sequences of length 2^n . Thus, for each $n \geq 1$, we have $2^{2^{n-1}-n}$ auto-correlation functions by definition. In view of (2), we study the relations among $\mathbf{X}$, ${}^r\mathbf{X}$, $\overline{\mathbf{X}}$, and ${}^r\overline{\mathbf{X}}$.

3.1 Previous results

We again set $N = 2^n$ ($n \geq 1$). Let $\mathbf{X} = (X_i)_{i=0}^{N-1}$ be a de Bruijn sequence of length 2^n . We summarize the known results which are closely related to this research as follows.

Observation 3 (Fredricksen [9]) Since $\mathbf{X}$ is a complete cycle, $\overline{\mathbf{X}} = (\overline{X_i})_{i=0}^{N-1}$ is also a complete cycle, i.e. a de Bruijn sequence. Similarly, ${}^r\mathbf{X}$ is a de Bruijn sequence as well.

Lemma 2 (Chan, Games, and Key [13]) For $n \geq 3$, $\mathbf{X} \neq \overline{\mathbf{X}}$.

Lemma 3 (Etzion and Lempel [10]) For $n \geq 3$, $\mathbf{X} \neq {}^r\mathbf{X}$.

By the definitions, ${}^r({}^r\mathbf{X}) = \mathbf{X}$, $\overline{(\overline{\mathbf{X}})} = \mathbf{X}$, and ${}^r\overline{\mathbf{X}} = \overline{{}^r\mathbf{X}}$. If $\mathbf{X} \simeq {}^r\overline{\mathbf{X}}$ or equivalently $\overline{\mathbf{X}} \simeq {}^r\mathbf{X}$, then $\mathbf{X}$ is called *CR (complement reverse) sequence*. By the definition, if $\mathbf{X}$ is a CR sequence, so are $\overline{\mathbf{X}}$ and ${}^r\mathbf{X}$.

Lemma 4 (Fredricksen [9]) For even $n \geq 4$, $\mathbf{X} \not\simeq {}^r\overline{\mathbf{X}}$.

On the other hand, it was pointed out in [9] that for $n = 5$, $\mathbf{X} \simeq {}^r\overline{\mathbf{X}}$ occurs. In fact, 32 pairs of CR sequences exit for $n = 5$. An example of such CR sequences is given:

Example 2 (Fredricksen [9]) $\mathbf{X} = 11111001000101011101100000110100$ has $\overline{\mathbf{X}} \simeq {}^r\mathbf{X}$.

Naturally the following problem was posed by Fredricksen in [9]: Show that there exists a CR sequence whenever $n (\geq 3)$ is odd. This problem has often been discussed. Particularly, in [10], a characterization of CR sequences were presented.

Lemma 5 (Etzion and Lempel [10]) Let $\mathbf{Y} = (Y_i)_{i=0}^{N-1}$ be a sequence over $\{0, 1\}$, which is not necessarily a de Bruijn sequence. The sequence $\mathbf{Y}$ is a CR sequence if and only if N is even and $\mathbf{Y} \simeq {}^r\overline{w}w$ for an $N/2$ -word w .

For words u and v , we use uv to denote a concatenation of u and v .

Unfortunately, however, to the best of the author's knowledge, the above Fredricksen's problem is still open. In this research, we solve this problem.

3.2 Construction of CR graphs

Let $G_n = (\mathcal{V}_n, \mathcal{A}_n)$ be the de Bruijn graph with the set $\mathcal{V}_n = \{0, 1\}^{n-1}$ of vertexes and the set $\mathcal{A}_n = \{0, 1\}^n$ of arcs. In order to solve the Fredricksen's question, we construct Eulerian graphs associated with the de Bruijn graph for odd $n (\geq 3)$. The resulting Eulerian graphs are called CR graphs since Eulerian circuits in the obtained graphs yield CR sequences in the de Bruijn sequences of length 2^n . Besides, the set of CR graphs supplies all CR sequences in the de Bruijn sequence.

We set $n = 2m + 1$ ($m \geq 1$). Since $n - 1 = 2m$ is even, in view of Lemma 5, the set $\mathcal{V}_{2m+1} = \{0, 1\}^{2m}$ of vertices includes all 2^m CR sequences of length $2m$. To distinguish such CR sequences of length $2m$ from CR sequences in question of length 2^n , we refer to such CR sequences as CR vertices or CR $2m$ -words. We use $\mathcal{V}_n^{CR} (\subset \mathcal{V}_n)$ to denote the set of CR vertices. Since CR $2m$ -words are in the form of ${}^r\overline{w}w$ where $w \in \{0, 1\}^m$, a total order relation $\leq$ on $\mathcal{V}_n^{CR}$ is defined by the following: for any ${}^r\overline{u}u$ and ${}^r\overline{v}v$ in $\mathcal{V}_n^{CR}$, ${}^r\overline{u}u \leq {}^r\overline{v}v$ if and only if

$$u_1 2^{m-1} + u_2 2^{m-2} + \cdots + u_m \leq v_1 2^{m-1} + v_2 2^{m-2} + \cdots + v_m,$$

where $u = u_1 u_2 \cdots u_m$ and $v = v_1 v_2 \cdots v_m$ are in $\{0, 1\}^m$. Thus we number all the elements in $\mathcal{V}_n^{CR}$: $v^{(0)} < v^{(1)} < \cdots < v^{(2^m-1)}$.

Using the notion of CR vertex, we obtain a refinement of Lemma 5 as follows, which plays crucially important roles in constructions of CR sequences.

Lemma 6 Let $\mathbf{X} \simeq {}^r\overline{w}w$ be a CR sequence in the de Bruijn sequence of length 2^{2m+1} , where $w = w_1 w_2 \cdots w_{2^m} \in \{0, 1\}^{2^m}$. Then there exists a unique CR vertex $v \in \mathcal{V}_{2m+1}^{CR}$ such that

$$v = {}^r\overline{w_1 w_2 \cdots w_m} w_1 w_2 \cdots w_m = w_{2^m-m+1} \cdots w_{2^m-1} w_{2^m} {}^r\overline{w_{2^m-m+1} \cdots w_{2^m-1} w_{2^m}}. \quad (3)$$

Moreover, the unique v occurs in $\mathbf{X}$ twice in the form of $0v1$ and $1v0$ while the other CR vertices $u \in \mathcal{V}_{2m+1}^{CR}$ occurs only once in w in the form of $1u1$ or $0u0$.

Proof: The existence of such a vertex v is clear by the definitions of CR sequences and CR vertices for the de Bruijn sequences and their graphs. Its uniqueness comes from the second assertion. Since $\mathbf{X}$ is a complete cycle, every $u \in \mathcal{V}_{2m+1}^{CR}$ with $u \neq v$ occurs only once in w . If $u (\neq v)$ occurs in w in the form of $0u1$ or $1u0$, then ${}^r\overline{w}w$ cannot be a CR sequence. Similarly, if v occurs in $\mathbf{X}$ in the form of $0v0$ or $1v1$, then, under the above condition (3), ${}^r\overline{w}w$ cannot be a CR sequence. This completes the proof. $\square$

To proceed constructions of CR graphs, we introduce

Definition 3 The weight $W(\mathbf{Y})$ of a sequence $\mathbf{Y} = (Y_i)_{i=0}^{N-1}$ over $\{0, 1\}$ is defined to be the number of nonzero digits among the N Y_i 's, i.e.,

$$W(\mathbf{Y}) = \sum_{i=0}^{N-1} Y_i.$$

Using this, we divide $\mathcal{V}_n$ into three disjoint subsets $\mathcal{V}_n^- = \{v \in \mathcal{V}_n : W(v) < m\}$, $\mathcal{V}_n^0 = \{v \in \mathcal{V}_n : W(v) = m\}$, and $\mathcal{V}_n^+ = \{v \in \mathcal{V}_n : W(v) > m\}$. Note that $\mathcal{V}_n^{CR} \subset \mathcal{V}_n^0$ since $W(v) = m$ for $v \in \mathcal{V}_n^{CR}$.

For each $v^{(i)} \in \mathcal{V}_n^{CR}$ ($i = 0, 1, \dots, 2^m - 1$), we construct the CR graph $H_{v^{(i)}}$ associated with the de Bruijn graph G_n as follows. Since, for each pair $v^{(i)}$ and $v^{(j)}$ in $\mathcal{V}_n^{CR}$ with $i + j = 2^m$, $\overline{v^{(i)}} = v^{(j)}$ holds, we will see $H_{v^{(i)}}$ and $H_{v^{(j)}}$ with $i + j = 2^m$ are graph isomorphic.

First we construct a directed graph G_n^0 associated with the de Bruijn graph G_n as follows. We set $\mathcal{W}_n = \{\lambda\} \cup \mathcal{V}_n \setminus \mathcal{V}_n^+$. For two vertices of the forms $u = a_1 a_2 \dots a_{n-1}$ and $v = a_2 a_3 \dots a_n$ in $\mathcal{W}_n$, the binary n -word $a_1 a_2 \dots a_n$ is defined as an arc from u to v . The obtained subgraph of G_n is not Eulerian since two types of arcs in G_n are not presented: $u1$ where $u \in \mathcal{V}_n^0$ is in the form of $u = 0v$; and $1u$ where $u \in \mathcal{V}_n^0$ is in the form of $u = v0$. Except these two types, for $w \in \mathcal{W}_n \setminus \{\lambda\}$, all arcs $0w$, $1w$, $w0$, and $w1$ in G_n are furnished in the resulting subgraph. In order to make the subgraph Eulerian, corresponding these two types of absent arcs, we add arcs from or to λ as follows. Corresponding all such arcs each in G_n : $u1$ where $u = 0v \in \mathcal{V}_n^0$; and $1u$ where $u = v0 \in \mathcal{V}_n^0$, we add an arc $u\lambda$ from u to λ for every $u = 0v \in \mathcal{V}_n^0$; and an arc λu from λ to u for every $u = v0 \in \mathcal{V}_n^0$.

Thus we obtain a connected graph, which we denote it by G_n^0 . By the above construction, λ has equal in-degree and out-degree, which is $\binom{2m-1}{m-1}$, and every vertex in $\mathcal{V}_n \setminus \mathcal{V}_n^+$ has equal in-degree and out-degree, which is two. Here we recall that the *in-degree* of a vertex is the number of vertices adjacent to it and that the *out-degree* of a vertex is the number of vertices adjacent from it. Hence the resulting directed graph G_n^0 is Eulerian.

Second we modify the directed graph G_n^0 to obtain a prototype of CR graphs. To this end, we divide $\mathcal{V}_n^0$ into four disjoint subsets $\mathcal{V}_n^{00} = \{v \in \mathcal{V}_n^0 : v = 0w0, w \in \{0, 1\}^{2(m-1)}\}$, $\mathcal{V}_n^{01} = \{v \in \mathcal{V}_n^0 : v = 0w1, w \in \{0, 1\}^{2(m-1)}\}$, $\mathcal{V}_n^{10} = \{v \in \mathcal{V}_n^0 : v = 1w0, w \in \{0, 1\}^{2(m-1)}\}$, and $\mathcal{V}_n^{11} = \{v \in \mathcal{V}_n^0 : v = 1w1, w \in \{0, 1\}^{2(m-1)}\}$. In the case that $m = 1$, we promise that $w \in \{0, 1\}^0$ means $w = \epsilon$ and that $u\epsilon v = uv$ for words u and v , where we use ϵ to denote the empty word. We note again that $\mathcal{V}_n^{CR} \subset \mathcal{V}_n^{01} \cup \mathcal{V}_n^{10}$. In particular, for $m = 1$, we have $\mathcal{V}_n^{CR} = \mathcal{V}_n^{01} \cup \mathcal{V}_n^{10}$ and $\mathcal{V}_n^{00} = \mathcal{V}_n^{11} = \emptyset$.

To proceed our constructions, we recall some notions from elementary theory of numbers. For $m \geq 2$, let $e_p(m)$ be defined as the power with which the prime p appears in the (unique) representation of m as a product of powers of its prime divisors, i.e.,

$$m = \prod_p p^{e_p(m)}.$$

Using $d(m)$ to denote the number of the divisors of m , we obtain

$$d(m) = \prod_p (e_p(m) + 1).$$

For each $m \geq 2$, the following vertices in $\mathcal{V}_n^{CR}$ particularly play essential roles in constructions of CR sequences. For a word w , we use w^k to denote the concatenation of k copies of w , i.e., $\underbrace{w \dots w}_k$.

For integers a and b , if a is a divisor of b , we write $a|b$. We use $[x]$ to denote the greatest integer not exceeding x . We use again S to denote the shift transformation on $\{0, 1\}^{2m}$ as above, i.e., $S(v_1, v_2, \dots, v_{2m-1}, v_{2m}) = (v_2, v_3, \dots, v_{2m}, v_1)$ for $v = v_1 v_2 \dots v_{2m} \in \{0, 1\}^{2m}$.

Definition 4 For $m (\geq 2)$, $2(d(m) - 1)$ vertices in $\mathcal{V}_n^{CR}$ in the form of $v^{(i(k))} = (1^k 0^k)^{\frac{m}{k}}$ and $\overline{v^{(i(k))}}$ with $k \geq 2$ are called the neutral vertices, where $k|m$ and $i(k) = \frac{2^{2k[\frac{m}{k}] + k} - 2^k}{2^k + 1}$. We use $\mathcal{V}_n^{CR, \nu}$ to denote the set of the neutral vertices in $\mathcal{V}_n^{CR}$. For each $j = 1, 2, \dots, k - 1$, $S^j(v^{(i(k))})$ is in $\mathcal{V}_n^{11}$. Such vertices in $\mathcal{V}_n^{11}$ are also called neutral. We use $\mathcal{V}_n^{11, \nu}$ to denote the set of the neutral vertices in $\mathcal{V}_n^{11}$. The set $\mathcal{V}_n^{00, \nu}$ of the neutral vertices in $\mathcal{V}_n^{00}$ is complementarily defined.

We turn to our constructions. Except the neutral vertices in $\mathcal{V}_n^{CR} \cup \mathcal{V}_n^{11}$, we split every vertex $v \in \mathcal{V}_n^0$ into two vertices: v with arcs $0v$ and $v0$; and v^+ with arcs $1v^+$ and v^+0 , as in the diagram:

$$\begin{array}{ccc} 0v \searrow & \nearrow & v0 \\ & \circ v & \\ 1v \nearrow & \searrow & v1 \end{array} \quad \text{into} \quad \begin{array}{ccc} \xrightarrow{0v} & \overset{v}{\circ} & \xrightarrow{v0} \\ \xrightarrow{1v^+} & \overset{v^+}{\circ} & \xrightarrow{v^+0} \end{array}$$

. Then, other than the neutral vertices, for every $v \in \mathcal{V}_n^0$, the copied vertex v^+ occurs in a single loop $\lambda 1^i v^+ 1^j \lambda$ where $0 \leq i + j \leq m$. We delete all such single loops. On the other hand, for each pair of neutral vertices $v^{(i(k))}$ and $\overline{v^{(i(k))}}$ in $\mathcal{V}_n^{CR}$, where $k|m$ and $i(k)$ is as in Definition 4, we have an arc $\overline{v^{(i(k))}} 0^k v^{(i(k))}$ from $\overline{v^{(i(k))}}$ to $v^{(i(k))}$. For each k , where $k|m$, we delete such an arc from $\overline{v^{(i(k))}}$ to $v^{(i(k))}$, where $i(k)$ is as in Definition 4. Then we add an arc from λ to $v^{(i(k))}$ and label it as $\lambda v^{(i(k))}$ while we add an arc from $\overline{v^{(i(k))}}$ to λ labeled as $\overline{v^{(i(k))}} \lambda$.

Thus we obtain a connected graph with the vertex set $\{\lambda\} \cup (\mathcal{V}_n^0 \setminus \mathcal{V}_n^{00,\nu}) \cup \mathcal{V}_n^-$, which we denote it by G_n^- . By the above construction, λ has equal in-degree and out-degree, which is $d(m) - 1$; every vertex in $\mathcal{V}_n^{CR,\nu} \cup \mathcal{V}_n^{11,\nu} \cup \mathcal{V}_n^-$ has equal in-degree and out-degree, which is two; and every vertex in $\mathcal{V}_n^0 \setminus (\mathcal{V}_n^{CR,\nu} \cup \mathcal{V}_n^{11,\nu} \cup \mathcal{V}_n^{00,\nu})$ has equal in-degree and out-degree, which is one. Hence the resulting directed graph G_n^- is Eulerian. We call it the prototype of CR graphs.

Now we are in a position to construct CR graphs by modifying the directed graph G_n^- . In what follows, we suppose $m (\geq 2)$ is a prime number. In order to indicate this assumption, we replace m by p in the rest of this research. We note here that, under this assumption, we obtain

$$\mathcal{V}_n^{CR,\nu} = \{v^{(i(p))} = 1^p 0^p, \overline{v^{(i(p))}} = 0^p 1^p\}.$$

First, we replace the vertex λ and its all four arcs labeled $\lambda v^{(i(p))}$ or $\overline{v^{(i(p))}} \lambda$, where $v^{(i(p))} \in \mathcal{V}_n^{CR,\nu}$, by two arcs from $\overline{v^{(i(p))}}$ to $v^{(i(p))}$ each labeled the same as $\overline{v^{(i(p))}} \lambda v^{(i(p))}$.

Choose $v^{(i)} \in \mathcal{V}_n^{CR}$ in G_n^- and fix it. If $v^{(i)}$ is not the neutral vertex, i.e., $v^{(i)} \in \mathcal{V}_n^{CR} \setminus \mathcal{V}_n^{CR,\nu}$, then we add a loop, an arc from $v^{(i)}$ to $v^{(i)}$, labeled $\overline{v^{(i)}} \lambda v^{(i)}$. If $v^{(i)}$ is the neutral vertex, i.e., $v^{(i)} = v^{(i(p))}$ or $v^{(i)} = \overline{v^{(i(p))}}$, then do nothing since two arcs labeled $\overline{v^{(i(k))}} \lambda v^{(i(k))}$: from $\overline{v^{(i)}}$ to $v^{(i)}$ when $v^{(i)} = v^{(i(p))}$; or from $v^{(i)}$ to $\overline{v^{(i)}}$ when $v^{(i)} = \overline{v^{(i(p))}}$, are already provided in the graph in construction.

Next, if $v^{(i)} \in \mathcal{V}_n^{CR} \setminus \mathcal{V}_n^{CR,\nu}$, then we split both neutral vertices $v^{(i(p))}$ and $\overline{v^{(i(p))}}$ in $\mathcal{V}_n^{CR,\nu}$ each into two vertices similarly as in the above diagram, which leads to

$$\begin{array}{ccccccc} \xrightarrow{0\overline{v^{(i(p))}}} & \overline{v^{(i(p))}} & \xrightarrow{\overline{v^{(i(p))}} \lambda v^{(i(p))}} & v^{(i(p))} & \xrightarrow{v^{(i(p))} 0} & & \\ & \circ & & \circ & & & \\ \xrightarrow{1\overline{v^{(i(p))}+}} & \overline{v^{(i(p))}+} & \xrightarrow{\overline{v^{(i(p))}+} \lambda v^{(i(p))}+} & v^{(i(p))}+ & \xrightarrow{v^{(i(p))}+1} & & \\ & \circ & & \circ & & & \end{array}$$

On the other hand, if $v^{(i)}$ is the neutral vertex, i.e., $v^{(i)} = v^{(i(p))}$ or $v^{(i)} = \overline{v^{(i(p))}}$, then we split both neutral vertices $v^{(i(p))}$ and $\overline{v^{(i(p))}}$ in $\mathcal{V}_n^{CR,\nu}$ each into two vertices as in the following diagram:

$$\begin{array}{ccccccc} \xrightarrow{1\overline{v^{(i(p))}+}} & \overline{v^{(i(p))}+} & \xrightarrow{\overline{v^{(i(p))}+} \lambda v^{(i(p))}} & v^{(i(p))} & \xrightarrow{v^{(i(p))} 0} & & \\ & \circ & & \circ & & & \\ \xrightarrow{0\overline{v^{(i(p))}}} & \overline{v^{(i(p))}} & \xrightarrow{\overline{v^{(i(p))}} \lambda v^{(i(p))}+} & v^{(i(p))}+ & \xrightarrow{v^{(i(p))}+1} & & \\ & \circ & & \circ & & & \end{array}$$

Eventually, for each $v^{(i)} \in \mathcal{V}_n^{CR}$, we obtain a connected graph with the vertex set $\mathcal{V}_n^0 \cup \mathcal{V}_n^- \cup \mathcal{V}_n^{CR,\nu+}$, where $\mathcal{V}_n^{CR,\nu+} = \{v^{(i(p))}+, \overline{v^{(i(p))}+}\}$, which we denote it by $H_{v^{(i)}}$. Noting that the vertex sets are the same for all $v^{(i)} \in \mathcal{V}_n^{CR}$, we write $\mathcal{W}_n^{CR} = \mathcal{V}_n^0 \cup \mathcal{V}_n^- \cup \mathcal{V}_n^{CR,\nu+}$. For each $v^{(i)} \in \mathcal{V}_n^{CR}$, every vertex in $H_{v^{(i)}}$ has equal in-degree and out-degree. Thus the resulting directed graph $H_{v^{(i)}}$ is Eulerian. We call it the CR graph associated with $v^{(i)}$ since Eulerian circuits in $H_{v^{(i)}}$ yield CR sequences. Using $\mathcal{B}_{v^{(i)}}$ to denote the set of arcs in $H_{v^{(i)}}$, we write $H_{v^{(i)}} = (\mathcal{W}_n^{CR}, \mathcal{B}_{v^{(i)}})$.

At this stage we have 2^p CR graphs. It is worth noting that $H_{v^{(i)}}$ and $H_{\overline{v^{(i)}}}$ are graph isomorphic. In symbols, we write $H_{v^{(i)}} \simeq H_{\overline{v^{(i)}}}$. In fact, we obtain a graph isomorphism $(\partial\Phi, \Phi) : H_{v^{(i)}} \rightarrow H_{\overline{v^{(i)}}}$ which consists of a pair of maps $\partial\Phi : \mathcal{W}_n^{CR} \rightarrow \mathcal{W}_n^{CR}$ and $\Phi : \mathcal{B}_{v^{(i)}} \rightarrow \mathcal{B}_{\overline{v^{(i)}}}$ as follows. We define the map $\partial\Phi$ by $\partial\Phi(v) = {}^r v$ for $v \in \mathcal{W}_n^{CR}$, as well as the map Φ by, for an arc va from a vertex $v \in \mathcal{W}_n^{CR}$, $\Phi(va) = {}^r(va) = a {}^r v$ which is an arc to the vertex ${}^r v$, and for an arc av to a vertex $v \in \mathcal{W}_n^{CR}$, $\Phi(av) = {}^r(av) = {}^r va$ which is an arc from the vertex ${}^r v$, where $a \in \{0, 1\}$. In particular,

recalling that $v^{(i)}$ is a CR word, we have $\partial\Phi(v^{(i)}) = \overline{v^{(i)}}$. Since both $\partial\Phi$ and Φ are one-to-one and onto, $(\partial\Phi, \Phi) : H_{v^{(i)}} \rightarrow H_{\overline{v^{(i)}}}$ is a graph isomorphism.

3.3 Construction of all CR sequences

In [6], we gave the bounded monotone truth-table algorithm for generating all full-length sequences which are based on the discretized piecewise-monotone-increasing Markov transformations. The algorithm is efficient in the sense that it guarantees to generate all full-length sequences without computing their total number. The algorithm proposed in [6] is applicable to generation of all de Bruijn sequences. Unfortunately, however, we might not apply this algorithm to generation of all CR sequences based on the CR graphs since we could not find in the CR graphs the expanded cycles defined in [6], which plays a vital role in the algorithm. Currently we need to compute the total number of the CR sequences.

In order to compute the total number of CR sequences in the de Bruijn sequences of length 2^{2p+1} , we first review the matrix tree theorem.

Let G be a directed graph with vertices $v_1, v_2, \dots, v_M$, and with t_{ij} arcs leading from v_i to v_j ($i, j = 1, 2, \dots, M$). The matrix $A = (t_{ij})_{i,j=1}^M$ is called the *adjacency matrix*. Writing $\text{diag}(d_1, d_2, \dots, d_M)$ for a diagonal matrix whose diagonal entries are $d_1, d_2, \dots, d_M$, let $D = \text{diag}(\text{odeg}(v_1), \text{odeg}(v_2), \dots, \text{odeg}(v_M))$, where $\text{odeg}(v)$ is the out-degree of a vertex v . For later use, we denote the in-degree of a vertex v by $\text{ideg}(v)$. The matrix $C = D - A$ is called the *matrix of admittance*. An *oriented spanning tree* of G with root v_i is a set of $M - 1$ arcs $a_1, a_2, \dots, a_{M-1}$ such that for $j = 1, 2, \dots, M$, there is a directed path along these arcs from v_j to v_i . The following theorem is well-known as the matrix tree theorem.

Theorem 3 (Tutte [14]) The number of oriented spanning trees of G with root v_i is the cofactor of c_{ii} in the matrix of admittance $C = (c_{ij})_{i,j=1}^M$.

If the directed graph G given above is Eulerian, then the following theorem shows that the number of Eulerian circuit in such G is given by the cofactor of c_{ii} for any vertex v_i .

Theorem 4 (van Aardenne-Ehrenfest and de Bruijn [15]) Let $G = (\mathcal{V}, \mathcal{A})$ be a directed graph with $\text{odeg}(v) = \text{ideg}(v)$ for every vertex $v \in \mathcal{V}$, and let G' be an oriented spanning tree of G . Let r be the root of G' and let $a(v)$ be the arc of G' with initial vertex v . Let a_1 be any arc with initial vertex r . Then $v_0 a_1 v_1 \dots a_N v_N$, where $v_0 = r, v_i \in \mathcal{V}$ and $a_i = v_{i-1} v_i \in \mathcal{A}$ ($i = 1, 2, \dots, N$) is an Eulerian circuit if it is an oriented path for which

- i) no arc is used more than once.
- ii) $a(v)$ is not used in $a_1, a_2, \dots, a_N$ unless it is the only choice consistent with rule (i).
- iii) $ra_1 v_1 \dots a_N v_N$ terminates only when it cannot be continued by rule (i).

For each CR graph $H_{v^{(i)}} (v^{(i)} \in \mathcal{V}_n^{CR}, i = 0, 1, \dots, 2^p - 1)$, we use $\Delta^{(v^{(i)})}$ to denote the cofactor of the $(1, 1)$ -th entry in the matrix of admittance of $H_{v^{(i)}}$. Since the CR graphs are Eulerian, by virtue of Theorem 4 together with the matrix tree theorem, the number of Eulerian circuit in $H_{v^{(i)}}$ is given by $\Delta^{(v^{(i)})}$.

In view of Theorem 4 in conjunction with the matrix tree theorem, in order to obtain all the Eulerian circuit in $H_{v^{(i)}}$, what we have to do is to find all minimum spanning tree for $H_{v^{(i)}}$. For such purposes, a wide variety of algorithms has been proposed. For instance, Kruskal's algorithm is a well-known algorithm, which is a greedy algorithm for finding a minimum spanning tree for a connected weighted graph [16].

Henceforth we may suppose that, once given a CR graph $H_{v^{(i)}}$, we obtain all $\Delta^{(v^{(i)})}$ Eulerian circuit in $H_{v^{(i)}}$. Let $\mathbf{Y}$ be an Eulerian circuit in $H_{v^{(i)}}$. We identify the circuit $\mathbf{Y}$ as a sequence over $\{\lambda, 0, 1\}$.

Let us consider a periodic sequence generated by the sequence $\mathbf{Y}$, which we use $\mathbf{Y}^\infty$ to denote.

- i) If $v^{(i)}$ is not the neutral vertex in $\mathcal{V}_n^{CR}$, then $\mathbf{Y}^\infty$ may be written in the form of

$$v^{(i)} 0 \alpha 0 \overline{v^{(i(p))}} \lambda v^{(i(p))} 0 \beta 1 \overline{v^{(i(p))}} \lambda v^{(i(p))} 1 \gamma 0 \overline{v^{(i)}} \lambda v^{(i)} 0 \alpha \dots \quad (4)$$

or

$$v^{(i)}0\alpha1\overline{v^{(i(p))}}\lambda v^{(i(p))}1\beta0\overline{v^{(i(p))}}\lambda v^{(i(p))}0\gamma0\overline{v^{(i)}}\lambda v^{(i)}0\alpha\cdots,$$

where α , β , and γ are words over $\{0, 1\}$ which occur in the first $(2^{2p} + 4p + 3)$ -word in $\mathbf{Y}^\infty$. We have to consider both cases. However, we consider only the former case that $\mathbf{Y}^\infty$ is in the form of (4) since the processes of constructing a CR sequence from $\mathbf{Y}$ are exactly the same in both cases. Hereafter, we assume that $\mathbf{Y}^\infty$ is in the form of (4). We transform the first $(2^{2p} + 4p + 3)$ -word in $\mathbf{Y}^\infty$ into

$$v^{(i)}0\alpha0\overline{v^{(i(p))}}\lambda \overline{r v^{(i(p))}0\beta1\overline{v^{(i(p))}}} \lambda v^{(i(p))}1\gamma0\overline{v^{(i)}}\lambda.$$

Noting that $v^{(i(p))}$ and $\overline{v^{(i(p))}}$ are CR words, we obtain

$$v^{(i)}0\alpha0\overline{v^{(i(p))}}\lambda\overline{v^{(i(p))}}0 \overline{r\beta1}v^{(i(p))}\lambda v^{(i(p))}1\gamma0\overline{v^{(i)}}\lambda.$$

After deleting λ , replace repetitions $\overline{v^{(i(p))}}\overline{v^{(i(p))}}$ and $v^{(i(p))}v^{(i(p))}$ each by single words $\overline{v^{(i(p))}}$ and $v^{(i(p))}$ respectively, then we obtain

$$v^{(i)}0\alpha0\overline{v^{(i(p))}}0 \overline{r\beta1}v^{(i(p))}1\gamma0\overline{v^{(i)}},$$

which we use $\mathbf{Z}$ to denote. Noting again that $v^{(i(p))}$ and $\overline{v^{(i(p))}}$ are CR words, we obtain

$$\mathbf{Z} \overline{r\mathbf{Z}} = v^{(i)}0\alpha0\overline{v^{(i(p))}}0 \overline{r\beta1}v^{(i(p))}1\gamma0\overline{v^{(i)}}\overline{v^{(i)}}1 \overline{r\gamma0}v^{(i(p))}0\beta1\overline{v^{(i(p))}}1 \overline{r\alpha}1v^{(i)}.$$

Replacing repetitions $\overline{v^{(i)}}\overline{v^{(i)}}$ and $v^{(i)}v^{(i)}$ in a circular order each by single words $\overline{v^{(i)}}$ and $v^{(i)}$ respectively, we obtain a CR sequence

$$\mathbf{X} = v^{(i)}0\alpha0\overline{v^{(i(p))}}0 \overline{r\beta1}v^{(i(p))}1\gamma0\overline{v^{(i)}}1 \overline{r\gamma0}v^{(i(p))}0\beta1\overline{v^{(i(p))}}1 \overline{r\alpha}1$$

of length 2^{2p+1} .

By the construction, in the word $v^{(i)}0\alpha0\overline{v^{(i(p))}}0\beta1v^{(i(p))}1\gamma0\overline{v^{(i)}}$, every vertex in $\mathcal{V}_n^{11,\nu} \cup \mathcal{V}_n^-$ occurs exactly twice while every vertex in $\mathcal{V}_n^0 \setminus \mathcal{V}_n^{11,\nu}$ occurs just once. Hence the obtained CR sequence $\mathbf{X}$ is in the de Bruijn sequences of length 2^{2p+1} .

ii) We consider the case that $v^{(i)}$ is the neutral vertex, i.e., $v^{(i)} = v^{(i(p))}$ or $v^{(i)} = \overline{v^{(i(p))}}$. We have to consider both cases. However, we consider only the case that $v^{(i)} = v^{(i(p))}$ since the processes of constructing a CR sequence from $\mathbf{Y}$ are exactly the same in both cases. Then, taking account of the equivalence relation which arises from the cycle, $\mathbf{Y}^\infty$ may be written uniquely in the form of

$$\overline{v^{(i(p))}}\lambda v^{(i(p))}0\alpha0\overline{v^{(i(p))}}\lambda v^{(i(p))}1\beta1\overline{v^{(i(p))}}\lambda v^{(i(p))}0\alpha\cdots,$$

where α and β are words over $\{0, 1\}$ which occur in the first $(2^{2p} + 4p + 3)$ -word in $\mathbf{Y}^\infty$.

Once we transform the first $(2^{2p} + 4p + 3)$ -word in $\mathbf{Y}^\infty$ into

$$\overline{v^{(i(p))}}\lambda v^{(i(p))}0\alpha0\overline{v^{(i(p))}}\lambda \overline{r v^{(i(p))}1\beta1\overline{v^{(i(p))}}} \lambda,$$

by using exactly the same procedure as in the case i) above, we obtain a CR sequence $\mathbf{X}$ in the de Bruijn sequences of length 2^{2p+1} .

Conversely, when we are given a CR sequence $\mathbf{X}$ in the de Bruijn sequences of length 2^{2p+1} , in view of Lemma 3, we find in $\mathbf{X}$ or $\overline{\mathbf{X}}$ a unique $v^{(i)} \in \mathcal{V}_{2m+1}^{CR}$ that satisfies the condition (3). Depending on whether $v^{(i)}$ is neutral or not, if we reverse the above procedure for the case i) or ii), we obtain an Eulerian circuit in $H_{v^{(i)}}$ from $\mathbf{X}$ or $\overline{\mathbf{X}}$. This correspondence is two-to-one and onto. In view of Lemma 2, for a CR sequence $\mathbf{X}$, $\overline{\mathbf{X}}$ gives a distinct CR sequence. Hence the above procedures for $v^{(i)} \in \mathcal{V}_{2m+1}^{CR}$ as a whole exhaust all pairs $(\mathbf{X}, \overline{\mathbf{X}})$ of CR sequences in the de Bruijn sequences of length 2^{2p+1} .

By using this correspondence between the Eulerian circuits and the CR sequences, together with Lemma 2, the total number of CR sequences in the de Bruijn sequences of length 2^{2p+1} is given by $2 \sum_{i=0}^{2^p-1} \Delta^{(v^{(i)})}$ which is equal to $4 \sum_{i=0}^{2^{p-1}-1} \Delta^{(v^{(i)})}$ in view of the graph isomorphism between $H_{v^{(i)}}$ and $H_{\overline{v^{(i)}}}$ described in the end of the previous subsection. Eventually we have partially solved the fundamental problem posed by Fredricksen in [9] on existence of CR sequences in the de Bruijn sequences of length 2^{2p+1} ($p \geq 1$):

Fig. 1. The prototype of CR graphs for $p = 2$.

Fig. 2. The CR graph associated with 1100.

Theorem 5 There exist $4 \sum_{i=0}^{2^{p-1}-1} \Delta(v^{(i)})$ CR sequences in the de Bruijn sequences of length 2^{2p+1} whenever p is prime, where $v^{(i)} \in \mathcal{V}_{2p+1}^{CR}$.

We immediately obtain a simple lower bound:

Remark 2

$$4 \sum_{i=0}^{2^{p-1}-1} \Delta(v^{(i)}) \geq 2^{p+1}$$

In the following example, to simplify notations, we omit the superscript $+$ from $v^{(i(p))}$.

Example 3 Figure 1 shows the prototype of CR graphs for $p = 2$. Figures 2 and 3 show the CR graphs associated with 1100 and with 1010 respectively. By these two CR graphs, simple computations lead to $\Delta^{(1100)} = 12$ and $\Delta^{(1010)} = 4$. Thus the number of CR sequences in the de Bruijn sequences of length 2^5 is 64 which confirm the experimental observation in [9].

4. On enumeration of auto-correlation functions of de Bruijn sequences

Based on the results obtained above together with the previous results, we discuss enumeration of the total number of distinct auto-correlation functions for the set of de Bruijn sequences of length 2^n ($n \geq 4$) in this section.

We have already introduced the equivalence relation $\simeq$ which arises from the cycle in the set of de Bruijn sequences at the end of Sect. 2. Let $\mathcal{T}_n$ be the quotient set by $\simeq$ of the set of de Bruijn sequences of length 2^n ($n \geq 4$). We introduce other equivalence relation $\sim$ in $\mathcal{T}_n$ as follows. For each

$$r_N(\ell; \mathbf{X}) \geq -1 + \frac{4}{2^n}.$$

This is equivalent to $d(\mathbf{X}, \overline{\mathbf{X}}) \geq 2$ for $n \geq 3$. From Theorem 3, we obtain $d(\mathbf{X}, {}^r\mathbf{X}) \geq 2$ for $n \geq 3$, and $d(\mathbf{X}, {}^r\overline{\mathbf{X}}) \geq 2$ for even $n \geq 4$. Thus, the common minimum values is 2 among these three distances, which suggests that in order to verify the existence of $\mathbf{Y}$ stated in the above observation, as the first step, we work on the question whether the set $\{\mathbf{Y} \in \mathcal{T}_n : \mathbf{Y} \sim \mathbf{X}, d(\mathbf{X}, \mathbf{Y}) = 2\} \setminus \{{}^r\mathbf{X}, \overline{\mathbf{X}}, {}^r\overline{\mathbf{X}}\}$ is empty for any $\mathbf{X} \in \mathcal{T}_n$ or not. To this end, let us characterize the elements in $\{\mathbf{Y} \in \mathcal{T}_n : \mathbf{Y} \sim \mathbf{X}, d(\mathbf{X}, \mathbf{Y}) = 2\}$ for $\mathbf{X} \in \mathcal{T}_n$.

Lemma 7 Let $\mathbf{X} = (X_i)_{i=0}^{2^n-1}$ and $\mathbf{Y} = (Y_i)_{i=0}^{2^n-1}$ be in $\mathcal{T}_n$ ($n \geq 5$). Then $d(\mathbf{X}, \mathbf{Y}) = 2$ and $\mathbf{X} \sim \mathbf{Y}$ if and only if there exist i ($0 \leq i \leq 2^n - 1$) and ℓ ($0 \leq \ell \leq 2^n - 1$) such that $X_i \neq Y_{i+\ell \pmod{2^n}}$, $X_{i+2^{n-1} \pmod{2^n}} \neq Y_{i+\ell+2^{n-1} \pmod{2^n}}$, and $X_j = Y_{j+\ell \pmod{2^n}}$ for $0 \leq j \leq 2^n - 1$ where $j \neq i$ and $j \neq i + 2^{n-1} \pmod{2^n}$; and for $1 \leq k \leq 2^{n-1} - 1$,

$$X_{i-k+2^n \pmod{2^n}} + X_{i+k \pmod{2^n}} = X_{i-k+2^{n-1} \pmod{2^n}} + X_{i+k+2^{n-1} \pmod{2^n}}. \quad (7)$$

Proof: The sufficiency is obvious. We shall prove the necessity. Suppose $d(\mathbf{X}, \mathbf{Y}) = 2$ and $\mathbf{X} \sim \mathbf{Y}$. Taking account of the equivalence relation in the cycle, we may choose $i = 0$ and $\ell = 0$. Namely $X_0 \neq Y_0$. This choice simplifies notations, but it does not affect the proof.

Thus we set $X_0 \neq Y_0$. Then $d(\mathbf{X}, \mathbf{Y}) = 2$ implies that there exists m ($1 \leq m \leq 2^n - 1$) such that $X_m \neq Y_m$ and $X_j = Y_j$ for $1 \leq j \leq 2^n - 1$ where $j \neq m$. Since $\mathbf{X}$ is a complete cycle, we have $X_m \neq X_0$.

We shall prove $m = 2^{n-1}$. Suppose $m \neq 2^{n-1}$. In conjunction with $\mathbf{X} \sim \mathbf{Y}$, we obtain

$$X_{2^n-k} + X_k = X_{m-k} + X_{k+m \pmod{2^n}} \text{ for } 1 \leq k \leq 2^n - 1 \text{ where } k \neq m; \quad (8)$$

and

$$X_{2^n-m} = X_{2m} \text{ for } k = m. \quad (9)$$

Then, from (8) and (9), we inductively obtain a sequence of equations: $X_{2^n-jm \pmod{2^n}} = X_{(j+1)m \pmod{2^n}}$ for $j = 1, 2, \dots$. In particular, if $j = 2^n - 1$, then we obtain $X_m = X_0$, which contradicts $X_m \neq X_0$. This completes the proof. $\square$

By virtue of Lemma 7, since $\mathbf{X} \sim {}^r\mathbf{X}$, we can specify the sequence $\mathbf{X}$ that satisfies $d(\mathbf{X}, {}^r\mathbf{X}) = 2$ as follows.

Corollary 3 If $d(\mathbf{X}, {}^r\mathbf{X}) = 2$, then $\mathbf{X}$ satisfies $X_0 \neq X_{2^{n-1}}$, and for $1 \leq k \leq 2^{n-1} - 1$, (7) as well as

$$X_{i+k \pmod{2^n}} = X_{i-k+2^{n-1} \pmod{2^n}} \quad \text{and} \quad X_{i-k+2^n \pmod{2^n}} = X_{i+k+2^{n-1} \pmod{2^n}}.$$

As we see in this corollary, the necessary and sufficient condition in Lemma 7 explicitly characterizes $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{T}_n$ ($n \geq 5$) that satisfy $d(\mathbf{X}, \mathbf{Y}) = 2$ and $\mathbf{X} \sim \mathbf{Y}$. However, it does not guarantee the existence of $\mathbf{Y}$ stated in the above observation. The set $\{\mathbf{Y} \in \mathcal{T}_n : \mathbf{Y} \sim \mathbf{X}, d(\mathbf{X}, \mathbf{Y}) = 2\} \setminus \{{}^r\mathbf{X}, \overline{\mathbf{X}}, {}^r\overline{\mathbf{X}}\}$ could be an empty set for any $\mathbf{X} \in \mathcal{T}_n$. We still have to verify the existence of such $\mathbf{Y}$. To this end, we need further studies on the characterization of the elements in $\{\mathbf{Y} \in \mathcal{T}_n : \mathbf{Y} \sim \mathbf{X}, d(\mathbf{X}, \mathbf{Y}) \neq 0\}$. Fortunately, we have

Example 4 For $n = 5$,

$$\mathbf{X} = 00000100101100110101000111110111 \text{ and } \mathbf{Y} = 00000100101101110101000111110011$$

satisfy $d(\mathbf{X}, \mathbf{Y}) = 2$ and $\mathbf{X} \sim \mathbf{Y}$ but $\mathbf{Y} \neq \mathbf{X}$, $\mathbf{Y} \neq {}^r\mathbf{X}$, and $\mathbf{Y} \neq \overline{\mathbf{X}}$.

This example shows that the set $\{\mathbf{Y} \in \mathcal{T}_n : \mathbf{Y} \sim \mathbf{X}, d(\mathbf{X}, \mathbf{Y}) = 2\} \setminus \{{}^r\mathbf{X}, \overline{\mathbf{X}}, {}^r\overline{\mathbf{X}}\}$ is not empty for the given $\mathbf{X} \in \mathcal{T}_n$ when $n = 5$, which implies that the equality does not hold in (6) for $p = 2$.

Accordingly, it also suggests the following problem: Show that whenever $n \geq 5$, there exist $\mathbf{X} \in \mathcal{T}_n$ such that $\{\mathbf{Y} \in \mathcal{T}_n : \mathbf{Y} \sim \mathbf{X}, d(\mathbf{X}, \mathbf{Y}) \neq 0\} \setminus \{^r\mathbf{X}, \overline{\mathbf{X}}, ^r\overline{\mathbf{X}}\}$ is not empty. It is noteworthy that, if we verify such an existence for some fixed $d(\mathbf{X}, \mathbf{Y})$, say $d(\mathbf{X}, \mathbf{Y}) = 2$, then the existence for other values of $d(\mathbf{X}, \mathbf{Y})$ is automatically implied as follows.

Remark 4 Noting that $\mathbf{Y} \neq ^r\mathbf{X}$ and $\mathbf{Y} \neq \overline{\mathbf{X}}$ are equivalent to $\mathbf{X} \neq ^r\mathbf{Y}$ and $\mathbf{X} \neq \overline{\mathbf{Y}}$, respectively, and that $\mathbf{Y} \sim ^r\mathbf{Y} \sim \overline{\mathbf{Y}} \sim ^r\overline{\mathbf{Y}}$, for $\mathbf{X}$ and $\mathbf{Y}$ in Example 4, we obtain $^r\mathbf{Y}$ and $\overline{\mathbf{Y}}$ which satisfy the conditions stated in the above observation with $d(\mathbf{X}, ^r\mathbf{Y}) = d(\mathbf{X}, \overline{\mathbf{Y}}) = 10$. Moreover, since $\mathbf{Y} \neq ^r\overline{\mathbf{Y}}$, we also obtain $^r\overline{\mathbf{Y}}$ which satisfies the conditions stated in the above observation with $d(\mathbf{X}, ^r\overline{\mathbf{Y}}) = 8$.

5. Summary

We have partially solved a fundamental problem posed by Fredricksen on existence of CR (complement reverse) sequences in the de Bruijn sequences of length 2^{2p+1} ($p \geq 1$). For the case that p is a prime number, we have constructed the set of CR graphs, which yields all CR sequences. As an application of this result, we discussed enumeration of the total number of distinct auto-correlation functions for the set of de Bruijn sequences of length 2^{2p+1} . We immediately observed that the worst cases of the normalized cross-correlation functions for pairs of de Bruijn sequences are characterized by the normalized auto-correlation functions for sequences of the worst pairs. In order to estimate such worst cases in advance, we obtained upper bounds of the total number of distinct auto-correlation functions for the set of de Bruijn sequences of length 2^n ($n \geq 4$).

Acknowledgments

This study was supported in part by the Telecommunications Advancement Foundation and in part by the Grant-in-Aid for Scientific Research (C) under Grant No. 24560445 from the Japan Society for the Promotion of Science. This paper is the resubmitted version of the paper with ID 2013NLP0003 received January 10, 2013; revised May 13, 2013; rejected June 11, 2013, and the paper with ID 2013ENP0005 received June 12, 2013; revised August 21, 2013; rejected September 22, 2013. Since May 13, 2013, the main results in Sect. 3 have not been changed. The author is grateful to all associate editors for their careful examinations and their making fair judges. The author thanks all reviewers of all the papers for their constructive suggestions and remarks. In particular, the reference to [10] was suggested by the reviewer B of the paper with ID 2013NLP0003. Without his suggestion, the Fredricksen's problem would not been challenged by the author. The reviewer A of the paper with ID 2013ENP0005 found Example 4, which posed the open problem described at the end of Sect. 4.

References

- [1] S. Kozic and M. Hasler, "Low-density codes based on chaotic systems for simple encoding," *IEEE Trans. Circuits and Systems I*, vol. 56, pp. 405–415, 2009.
- [2] N. Masuda and K. Aihara, "Chaotic cipher by finite-state baker's map," *Trans. of IEICE*, vol. 82-A, pp. 1038–1046, 1999 (in Japanese).
- [3] A. Tsuneda, Y. Kuga, and T. Inoue, "New maximal-period sequences using extended nonlinear feedback shift registers based on chaotic maps," *IEICE Trans. on Fundamentals*, vol. E85-A, pp. 1327–1332, 2002.
- [4] H. Fujisaki, "Discretized Markov transformations – an example of ultradiscrete dynamical systems –, " *IEICE Trans. Fundamentals*, vol. E88-A, pp. 2684–2691, 2005.
- [5] R. Hirota and D. Takahashi, *Discrete and Ultradiscrete Systems*, Kyoritsu Shuppan, 2003 (in Japanese).
- [6] H. Fujisaki, "An algorithm for generating all full-length sequences which are based on discretized Markov transformations," *NOLTA*, vol. 1, pp. 166–175, 2010.
- [7] Z. Zhang and W. Chen, "Correlation properties of de Bruijn sequences," *Systems Science and Mathematical Sciences*, vol. 2, pp. 170–183, 1989.
- [8] H. Fujisaki and Y. Nabeshima, "On auto-correlation values of de Bruijn sequences," *NOLTA*, vol. 3, pp. 400–408, 2011.

- [9] H. Fredricksen, "A survey of full length nonlinear shift register cycle algorithm," *SIAM Review*, vol. 24, pp. 195–221, 1982.
- [10] T. Etzion and A. Lempel, "On the distribution of de Bruijn sequences of given complexity," *IEEE Trans. on Information Theory*, vol. 30, pp. 611–614, 1984.
- [11] N.G. de Bruijn, "A combinatorial problem," *Nederl. Akad. Wetensch. Proc.*, vol. 49, pp. 758–764, 1946.
- [12] C. Flye Sainte-Marie, "Solution to problem number 48," *L'Intermédiaire des Mathématiciens*, vol. 1, pp. 107–110, 1894.
- [13] A.H. Chan, R.A. Games, and E L. Key, "On the complexities of de Bruijn sequences," *J. Comb. Theory, Ser. A*, vol. 33, pp. 233–246, 1982.
- [14] W.T. Tutte, "The dissection of equilateral triangles into equilateral triangles," *Proc. Cambridge Phil. Soc.*, vol. 44, pp. 463–482, 1948.
- [15] T. van Aardenne-Ehrenfest and N.G. de Bruijn, "Circuits and trees in oriented linear graphs," *Simon Stevin*, vol. 28, pp. 203–217, 1951.
- [16] J.B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proceedings of the American Mathematical Society*, vol. 7, pp. 48–50, 1956.