Further Investigations on Nonlinear Complexity of Periodic Binary Sequences

Qin Yuan [*], Chunlei Li [†], Xiangyong Zeng, Tor Helleseth, and Debiao He [‡]

Abstract

Nonlinear complexity is an important measure for assessing the randomness of sequences. In this paper we investigate how circular shifts affect the nonlinear complexities of finite-length binary sequences and then reveal a more explicit relation between nonlinear complexities of finite-length binary sequences and their corresponding periodic sequences. Based on the relation, we propose two algorithms that can generate all periodic binary sequences with any prescribed nonlinear complexity.

Index Terms: Periodic sequence, nonlinear complexity, randomness

1 Introduction

Pseudorandom sequences have applications in various digital systems and communication technologies, such as radio communications, distance ranging, simulation, game theory, and cryptography [1]. The quality of randomness is a crucial factor for pseudorandom sequences in many of these applications, particularly for cryptographic applications. For assessing the randomness of pseudorandom sequences, different complexity measures

[*]Q. Yuan and X. Zeng are with Faculty of Mathematics and Statistics, Hubei Key Laboratory of Applied Mathematics, Hubei University, Wuhan 430062, Hubei, China. Email: yuanqin2020@aliyun.com, xzeng@hubu.edu.cn

[†]C. Li and T. Helleseth are with the Department of Informatics, University of Bergen, Bergen, N-5020, Norway. Email: chunlei.li@uib.no, tor.helleseth@uib.no

[‡]D. He is with Key Laboratory of Aerospace Information Security and Trusted Computing Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, Hubei, China. Email: hedebiao@whu.edu.cn

were proposed in the literature [2–5] and the most well-understood one is probably the linear complexity. The linear complexity of a sequence is defined as the length of the shortest linear feedback shift registers (LFSRs) that can generate the sequence [6, 7]. Suppose a sequence $\mathbf{s}$ of length n has linear complexity $l \leq n/2$. Given any of its $2l$-length subsequences, namely, $(s_i, s_{i+1}, \ldots, s_{i+2l-1})$ for any $i \geq 0$, the Berlekamp-Massey algorithm [8] can efficiently produce the linear recurrence of length l and thereby the whole sequence $\mathbf{s}$. Hence pseudorandom sequences for cryptographic applications must not have low linear complexity. Rueppel [9] conjectured that n-periodic binary sequences have expected linear complexity close to n. Meidl and Niederreiter [10] confirmed this conjecture for arbitrary finite fields; in particular, they showed that n-periodic binary sequences have expected linear complexity at least $\frac{3n-1}{4}$. Research has been done on the linear complexity of special sequences, for instance, Lempel-Cohn-Eastman sequences [11], Legendre sequences [12]. Meanwhile, variants and extensions of linear complexity, e.g., linear complexity profile [7], k-error linear complexity [7,10], quadratic span [13] and nonlinear complexity [14] have been studied. Interested readers may refer to [15,16] for more discussions on these complexity measures.

As an additional figure of merit to judge the randomness of a sequence, nonlinear complexity (also referred to as maximum-order complexity) was introduced by Jansen and Boekee in 1989 [14,17], where they defined it as the length of the shortest FSRs that generate a given sequence. Their work showed that the expected nonlinear complexity of random n-length q-ary sequences is approximately $2\log_q(n)$. Significant progress has been made on the derivation of the shortest feedback functions for a given sequence and the construction of sequences with high nonlinear complexity. Jansen and Boekee [14] initially related nonlinear complexity to the maximum depth of a directed acyclic word graph, which can be employed to determine the nonlinear complexity (profile) of a given sequence. Rizomiliotis and Kalouptsidis [18] in 2005 proposed an efficient algorithm, which exploited the special structure of associated linear equations, for finding the shortest feedback functions for a given sequence. Later Limniotis et al. [19] studied the relation between nonlinear complexity and Lempel-Ziv complexity, thereby presenting a recursive algorithm which has a similar procedure to the Berlekamp-Massey algorithm. As for the construction of sequences with high nonlinear complexity, researchers mainly exploited some algebraic tools and explored the structure of those constructed sequences [20–28]. Niederreiter et al. [20], Luo et al. [21] and Castellanos et al. [22] constructed sequences with high nonlinear complexity from function fields with many rational places. Based on

detailed investigations of internal structures of sequences, all q-ary sequences of length n and high nonlinear complexities $n-i$, $i=1,2,3,4$, were completely characterized in [23,24]. Very recently Liang et al. [25] completely determined n-length binary sequences with nonlinear complexity $\geq \frac{n}{2}$.

Let $\mathbf{s}_n^\infty$ be an n-periodic sequence over certain alphabet $\mathcal{A}$. Its nonlinear complexity is closely related to nonlinear complexities of its n-length subsequences, namely, $nlc(\mathbf{s}_n^\infty) = nlc(\mathbf{a}_n^2) \geq nlc(\mathbf{a}_n)$, where nlc denotes the nonlinear complexity of a given sequence, $\mathbf{a}_n$ is any n-length subsequence of $\mathbf{s}_n^\infty$ and $\mathbf{a}_n^2 = \mathbf{a}_n\mathbf{a}_n$ denotes the concatenation of two identical $\mathbf{a}_n$. As $\mathbf{a}_n$ can be chosen as any n-length subsequence of $\mathbf{s}_n^\infty$, this inequality is relatively vague and does not reveal further insights into the relation between $nlc(\mathbf{s}_n^\infty)$ and $nlc(\mathbf{a}_n)$. Only a few results have been reported on nonlinear complexity of periodic sequences so far. Rizomiliotis [26] exploited the power series representation of binary sequences and proposed two constructions of binary sequences of period 2^m-1 with maximum nonlinear complexity for given linear complexity. In 2017 Sun et al. completely characterized the structure of periodic sequences $\mathbf{s}_n^\infty$ of maximum nonlinear complexity $n-1$ and proposed a recursive algorithm to generate all such sequences [27]. Later Xiao et al. [28] determined all binary sequences of period n having nonlinear complexity $n-2$ in a similar way. Motivated by recent work on nonlinear complexities of binary n-length sequences [25], this paper aims to reveal more explicit relations between $nlc(\mathbf{s}_n^\infty)$ and $nlc(\mathbf{a}_n)$, where $\mathbf{a}_n$ is a certain cyclic shift of $\mathbf{s}_n$. To this end, we are concerned with a particular set $\mathcal{B}(n,c)$ (in Def. 2) of sequences with a specific structure. We start with investigating how circular shift operators affect the nonlinear complexity of binary sequences $\mathbf{s}_n$ in $\mathcal{B}(n,c)$. Then we study the relation between the companion pairs of $\mathbf{s}_n$ and the companion pairs of $\mathbf{s}_n^\infty$, which enables us to establish a one-to-one correspondence between periodic sequences $\mathbf{s}_n^\infty$ with given nonlinear complexities and certain sequences $\mathbf{s}_n$ in $\mathcal{B}(n,c)$ (see Thm. 1 and Thm. 2). Based on the correspondence, all periodic binary sequences $\mathbf{s}_n^\infty$ with any prescribed nonlinear complexity ω can be completely generated (in Thm. 3). The generation process is summarized in two algorithms (Alg.1 and Alg.3), depending on the relation between $\omega = nlc(\mathbf{s}_n^\infty)$ and $\frac{n}{2}$. To the best of our knowledge, this is the first report of theoretical results on the subject that can be used to efficiently generate all periodic binary sequences with any prescribed nonlinear complexity.

The remainder of this paper is organized as follows. In Section 2, we introduce some basics of nonlinear complexity. Section 3 presents notations, definitions, and auxiliary results for the nonlinear complexity of n-length binary sequences. Section 4 is dedicated

to revealing more explicit relations between $nlc(\mathbf{s}_n^\infty)$ and $nlc(\mathbf{a}_n)$, where $\mathbf{a}_n$ is a certain cyclic shift of $\mathbf{s}_n$. In Section 5, we propose two algorithms to generate periodic binary sequences with any prescribed nonlinear complexity. Finally, Section 6 concludes the work of this paper.

2 Preliminaries

In Section 2 and Section 3, we will introduce basic notations, definitions and auxiliary lemmas. For readers' convenience we summarise important notations in Table 1 at the end of Section 3.

For a positive integer m, an m-stage *feedback shift register* (FSR) is a clock-controlled circuit consisting of m consecutive storage units and a feedback function f as displayed in Figure 1. Starting with an initial state $\underline{\mathbf{s}}_0 = (s_0, s_1, \ldots, s_{m-1})$, the states in the FSR will be updated by a clock-controlled transformation as follows:

$$\mathcal{F} : \underline{\mathbf{s}}_i = (s_i, s_{i+1}, \ldots, s_{i+m-1}) \longmapsto \underline{\mathbf{s}}_{i+1} = (s_{i+1}, \ldots, s_{i+m-1}, s_{i+m}),\ i \geq 0,$$

where $s_{i+m} = f(s_i, s_{i+1}, \ldots, s_{i+m-1})$, and the leftmost symbol for each state $\underline{\mathbf{s}}_i$ will be output. In this way an FSR produces a sequence $\mathbf{s} = (s_0, s_1, s_2, \ldots)$ based on each initial state $\underline{\mathbf{s}}_0$ and its feedback function f. The shift register sequence can be equivalently expressed as a sequence of states, $(\underline{\mathbf{s}}_0, \underline{\mathbf{s}}_1, \underline{\mathbf{s}}_2, \ldots)$, with the relation $\underline{\mathbf{s}}_i = \mathcal{F}(\underline{\mathbf{s}}_{i-1}) = \cdots = \mathcal{F}^i(\underline{\mathbf{s}}_0)$ for $i \geq 0$. When $\underline{\mathbf{s}}_p = \mathcal{F}^p(\underline{\mathbf{s}}_0) = \underline{\mathbf{s}}_0$ for the least integer $p \geq 1$, we obtain a cycle of states $\underline{\mathbf{s}}_0, \ldots \underline{\mathbf{s}}_{p-1}$, or equivalently a sequence $(s_0, \ldots, s_{p-1}, \ldots)$ of period p.

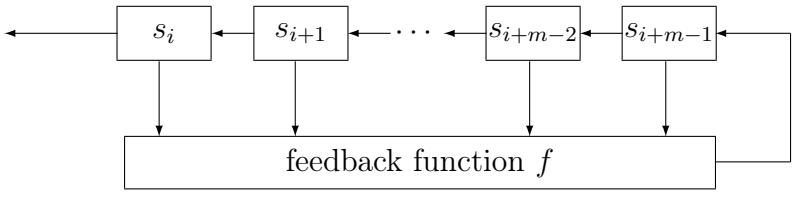

Figure 1: An m-stage FSR with feedback function f

In his influential book [1], Golomb intensively studied the relation between the feedback function of an FSR and its output sequences/cycles. He showed that an FSR generates disjoint cycles if and only if it uses nonsingular feedback functions of the form

$f(x_0, x_1, \ldots, x_{m-1}) = x_0 + g(x_1, \ldots, x_{m-1})$. Nonsingular FSRs are of practical interest as their output sequences have simpler structure and often exhibit desirable randomness properties [1]. For an m-stage nonsingular binary FSR, when its feedback function f is linear, the output sequence can have the longest period $2^m - 1$, which contains all nonzero m-tuples exactly once and is known as a *maximum-length* sequence (m-sequence for short); when f is nonlinear, its output sequence can have the longest period 2^m, containing all binary m-tuples exactly once, and is known as a *binary de Bruijn sequence of order* m [29]. Both m-sequences and de Bruijn sequences exhibit the *span property* [30, 31]. They are of significant interest in research and applications [32, 33]. While the theory of m-sequences is well explored, many problems about de Bruijn sequences remain unsolved, for instance, the necessary or sufficient properties of feedback functions for generating de Bruijn sequences, more explicit combinatorial structure of de Bruijn sequences and efficient generations of all de Bruijn sequences of modest lengths, etc.

2.1 Nonlinear complexity

Linear complexity profile has been an important measure of randomness of sequences for cryptography [9]. As an additional figure of merit to judge the randomness of sequences, Jansen and Boekee proposed the maximum-order complexity, later known as nonlinear complexity, of sequences [14, 17].

Definition 1. *([17]) The nonlinear complexity of a sequence $\boldsymbol{s}$ over an alphabet $\mathcal{A}$, denoted by $nlc(\boldsymbol{s})$, is the length of the shortest feedback shift registers that can generate the sequence $\boldsymbol{s}$.*

For a sequence $\mathbf{s} = (s_0, s_1, \ldots)$ over $\mathcal{A}$, the term s_{i+k} is deemed as the *successor* of the subsequence $\mathbf{s}_{[i:i+k]} = (s_i, \ldots, s_{i+k-1})$ for certain positive integers i and k. Some properties of the nonlinear complexity of sequences are recalled below.

Lemma 1. *([17]) The nonlinear complexity of a sequence $\boldsymbol{s}$ equals one plus the length of its longest identical subsequences that occur at least twice with different successors.*

The metric of nonlinear complexity is well defined for both finite-length sequences and infinite-length periodic sequences, which exhibit apparent difference in spite of the close connection. Below we recall some basics for finite-length sequences and periodic sequences, respectively.

Lemma 2. *([17]) For a finite-length sequence $(s_0, s_1, \ldots, s_{n-1})$ over an alphabet $\mathcal{A}$,*

(i) *its nonlinear complexity takes values ranging from 0 to $n-1$;*

(ii) *if its nonlinear complexity $c \geq \frac{n}{2}$, then the sequence $(s_0, s_1 \ldots, s_{n-1}, s_n)$ for any new symbol s_n in $\mathcal{A}$ has the same nonlinear complexity c;*

(iii) *if its nonlinear complexity $c \geq \lfloor \frac{n}{2} \rfloor$, then it cannot be written as $(s_0, s_1, \ldots, s_{k-1})^e$ for any proper divisor k of n.*

Throughout what follows, we will use $\mathbf{s}_n$ to denote a sequence $(s_0, s_1, \ldots, s_{n-1})$ that is not a repetition of any shorter subsequence and deem it as an aperiodic finite-length sequence. Given a sequence $\mathbf{s}_n = (s_0, \ldots, s_{n-1})$, the left circular shift operators $L^i(\mathbf{s}_n)$ on $\mathbf{s}_n$ are defined as $L^0(\mathbf{s}_n) = \mathbf{s}_n$, $L(\mathbf{s}_n) = (s_1, \ldots, s_{n-1}, s_0)$ and $L^i(\mathbf{s}_n) = L(L^{i-1}(\mathbf{s}_n)) = (s_i, s_{i+1}, \ldots, s_{n-1}, s_0 \ldots, s_{i-1})$ for $i \geq 1$. The right circular shift operators are defined by $R^i(\mathbf{s}_n) = L^{n-i}(\mathbf{s}_n)$. Under circular shift operators, we can derive from $\mathbf{s}_n$ a *shift equivalence class* $\{\mathbf{s}_n, L(\mathbf{s}_n), \ldots, L^{n-1}(\mathbf{s}_n)\}$. Notice that the nonlinear complexity of $\mathbf{s}_n$ is usually not an invariant under circular shift operators. For instance, while the sequence $\mathbf{s}_n = (1, 0, \ldots, 0)$ has nonlinear complexity 1, its shift sequence $L(\mathbf{s}_n) = (0, \ldots, 0, 1)$ has maximum nonlinear complexity $n-1$.

Shift operators can be similarly defined for periodic sequences. The shift equivalence class of a periodic sequence $\mathbf{s}_n^\infty$ is given by $\{\mathbf{s}_n^\infty, (L(\mathbf{s}_n))^\infty, \ldots, (L^{n-1}(\mathbf{s}_n))^\infty\}$. As recalled in the following lemma, the nonlinear complexity of a periodic sequence is an invariant under circular shift operators.

Lemma 3. *([17]) Let $\boldsymbol{s}_n^\infty$ be a sequence of period n over an alphabet $\mathcal{A}$ of size q. Then,*

(i) *the nonlinear complexity of $\boldsymbol{s}_n^\infty$ satisfies $\lceil \log_q(n) \rceil \leq nlc(\boldsymbol{s}_n^\infty) \leq n-1$;*

(ii) *all sequences in $\{\boldsymbol{s}_n^\infty, (L(\boldsymbol{s}_n))^\infty, \ldots, (L^{n-1}(\boldsymbol{s}_n))^\infty\}$ have the same nonlinear complexity.*

For finite-length sequences $\mathbf{s}_n$ and periodic sequences $\mathbf{s}_n^\infty$, despite their close connection, they can behave differently in some aspects. As indicated in Lemma 2 and Lemma 3, the nonlinear complexities of $\mathbf{s}_n$ and $\mathbf{s}_n^\infty$ take values from different ranges, and they behave differently under circular shift operators. For a finite-length sequence $\mathbf{s}_n$, since its corresponding periodic sequence $\mathbf{s}_n^\infty$ contains all shift equivalent sequences $\mathbf{s}_n, L(\mathbf{s}_n), \ldots, L^{n-1}(\mathbf{s}_n)$, it is clear that $\mathbf{s}_n^\infty$ has nonlinear complexity $nlc(\mathbf{s}_n^\infty) \geq nlc(L^i(\mathbf{s}_n))$ for any $0 \leq i < n$, i.e.,

$$nlc(\mathbf{s}_n^\infty) \geq \max_{0 \leq i < n} nlc(L^i(\mathbf{s}_n)).$$

While for some sequences $\mathbf{s}_n$ the equality of the above inequality can be achieved, the equality can not be reached for many other sequences $\mathbf{s}_n$, i.e., $nlc(\mathbf{s}_n^\infty) > \max\limits_{0\le i<n} nlc(L^i(\mathbf{s}_n))$. For instance, for the sequence $\mathbf{s}_{10} = (0010010010)$, we have

$$\big(\, nlc(L^i(\mathbf{s}_{10})) \,:\, i = 0, 1, 2, \ldots, 9 \,\big) = (\, 2, 2, 7, 6, 5, 4, 6, 5, 4, 3 \,) \text{ and } nlc(\mathbf{s}_{10}^\infty) = 9.$$

In this paper we will further investigate the varying behaviour of nonlinear complexity of binary sequences $\mathbf{s}_n$ under circular shift operators, thereby characterizing periodic binary sequences $\mathbf{s}_n^\infty$ with a prescribed nonlinear complexity. As a result, we are enabled to reveal more explicit relations between $nlc(\mathbf{s}_n^\infty)$ and $nlc(\mathbf{a}_n)$, where $\mathbf{a}_n$ is a certain n-length subsequence of $\mathbf{s}_n^\infty$. Throughout this paper, we will denote $c = nlc(\mathbf{s}_n)$ and $\omega = nlc(\mathbf{s}_n^\infty)$ for a finite-length sequence $\mathbf{s}_n$ and the corresponding periodic sequence $\mathbf{s}_n^\infty$, respectively.

Suppose $\mathbf{a}_n^\infty$ is a binary sequence with period n and nonlinear complexity ω generated by an ω-stage FSR. Recall from Figure 1 that $\mathbf{a}_n^\infty = (a_0, a_1, a_2, \ldots)$ can be equivalently expressed as a cycle of states $(\underline{\mathbf{a}}_0, \underline{\mathbf{a}}_1, \ldots, \underline{\mathbf{a}}_{n-1})$, where $\underline{\mathbf{a}}_t = \mathcal{F}^t(\underline{\mathbf{a}}_0)$ for $t \ge 0$. Given a state $\underline{\mathbf{a}}_i$, its *companion state* is defined by $\widehat{\underline{\mathbf{a}}_i} = (a_i, \ldots, a_{i+\omega-2}, \overline{a}_{i+\omega-1})$ with $\overline{a}_{i+\omega-1} = a_{i+\omega-1} \oplus 1$. A companion pair $(\underline{\mathbf{a}}_i, \widehat{\underline{\mathbf{a}}_i}) = (\underline{\mathbf{a}}_i, \underline{\mathbf{a}}_{i+d})$ for certain $d \ge 1$ can be denoted as $(\underline{\mathbf{a}}_i, \mathcal{F}^d(\underline{\mathbf{a}}_i))$. The structure of $(\underline{\mathbf{a}}_i, \mathcal{F}^d(\underline{\mathbf{a}}_i))$ will be frequently used in our discussion. Without loss of generality, we assume $d \le \lfloor \frac{n}{2} \rfloor$ (since for the case that $d > \lfloor \frac{n}{2} \rfloor$ we can consider the companion pair $(\underline{\mathbf{a}}_{i+d}, \mathcal{F}^{n-d}(\underline{\mathbf{a}}_{i+d}))$). We shall consider its left shift sequence $\mathbf{s}_n^\infty = (L^i(\mathbf{a}_n))^\infty$, which has a companion pair $(\underline{\mathbf{s}}_0, \mathcal{F}^d(\underline{\mathbf{s}}_0))$ with $d \le \lfloor \frac{n}{2} \rfloor$. Here we consider $\mathbf{s}_n^\infty$ instead of $\mathbf{a}_n^\infty$ for simplifying the notation in subsequent sections.

3 Characterizations of finite-length sequences

In this section, we consider finite-length sequences with certain structure and discuss the change of nonlinear complexities of those sequences under circular shift operators.

We first recall recent results on the nonlinear complexity of binary sequences.

Lemma 4. *([25]) For a binary finite-length sequence $(s_0, s_1, \ldots, s_{n-1})$, if it has nonlinear complexity $c \ge \frac{n}{2}$, then there exists exactly one pair of identical subsequences of length $c-1$ with different successors in $(s_0, s_1, \ldots, s_n)$.*

Lemma 5. *([25]) Let c and d be two positive integers with $c + d \ge 3$, and a binary*

sequence given by

$$\boldsymbol{s}_{c+d} = \overbrace{(s_0, \ldots, s_{d-1}) \ldots (s_0, \ldots, s_{d-1})}^{q \text{ repetitions}}(s_0, \ldots, s_{r-1}, \overline{s}_r) = \boldsymbol{s}_d^q\,(s_0, \ldots, s_{r-1}, \overline{s}_r) \qquad (1)$$

where $q = \lfloor \frac{c+d-1}{d} \rfloor$, $0 \le r = (c+d-1) - qd < d$, $\boldsymbol{s}_d$ *is aperiodic and* $\boldsymbol{s}_d^q = \overbrace{\boldsymbol{s}_d \ldots \boldsymbol{s}_d}^{q \text{ repetitions}}$, *and when* $r = 0$, $(s_0, \ldots, s_{r-1}, \overline{s}_r)$ *is deemed as* $\overline{s}_0$. *Then when* $c \ge d$, *the sequence* $\boldsymbol{s}_{c+d}$ *in* (1) *has nonlinear complexity* c.

The binary sequences of the form in (1) will be heavily used in subsequent discussions. Note that $\mathbf{s}_{c+d}$ can be equivalently expressed as

$$(s_0, s_1, \ldots, s_{c-2}) = (s_d, s_{d+1}, \ldots, s_{c+d-2}) \text{ and } s_{c-1} \ne s_{c+d-1}.$$

Given a sequence $\mathbf{s}_{c+d} = (s_0, s_1, \ldots s_{c+d-1})$, its two subsequences $\mathbf{s}_{[0:c]}$ and $\mathbf{s}_{[d:c+d]}$ are overlapped at components $s_d, \ldots, s_{c-1}$ when $c > d$ and are exactly next to each other when $c = d$. Liang et al. [25] showed that when $c \ge d$ the sequence $\mathbf{s}_{c+d}$ has nonlinear complexity c and spacing d between its companion pair $(\underline{\mathbf{s}}_0, \mathcal{F}^d(\underline{\mathbf{s}}_0))$ if and only if it has the form as in (1). When $c < d$, the statement is neither true for sufficiency nor for necessity. The result in Lemma 5 helps us calculate the nonlinear complexity of a binary sequence $\mathbf{s}_n$ starting with $\mathbf{s}_{c+d}$ in a direct way.

Corollary 1. *Suppose a binary sequence* $\boldsymbol{s}_n = (s_0, \ldots, s_{n-1})$ *has its subsequence* $\boldsymbol{s}_{c+d}$ *of the form in* (1) *for certain positive integers* c *and* d *with* $1 \le d \le \min\{n-c, \lfloor \frac{n}{2} \rfloor\}$. *Then we have* $nlc(\boldsymbol{s}_n) \ge c$, *where the equality is achieved when* $c \ge \lfloor \frac{n}{2} \rfloor$.

Proof. When $c \ge d$, it follows from Lemma 5 that $nlc(\mathbf{s}_{c+d}) = c$. Then $nlc(\mathbf{s}_n) \ge nlc(\mathbf{s}_{c+d}) = c$. When $c < d$, the subsequence $\mathbf{s}_{c+d}$ is of the form

$$\mathbf{s}_{c+d} = (s_0, \ldots, s_{c-2}, s_{c-1}, \ldots, s_{d-1})(s_0, \ldots, s_{c-2}, \overline{s}_{c-1}),$$

where the subsequences $(s_0, \ldots, s_{c-2})$ have different successors. By Lemma 1 we have $nlc(\mathbf{s}_{c+d}) \ge c$. So it is clear that $nlc(\mathbf{s}_n) \ge c$.

If $c \ge \lfloor \frac{n}{2} \rfloor$, then by Lemma 2 (ii) recursively, we have $nlc(\mathbf{s}_n) = nlc(\mathbf{s}_{c+d})$. Note that $nlc(\mathbf{s}_{c+d}) = c$ since $c \ge \lfloor \frac{n}{2} \rfloor \ge d$. Thus $nlc(\mathbf{s}_n) = c$ if $c \ge \lfloor \frac{n}{2} \rfloor$. □

Below we define a set $\mathcal{B}(n, c)$ that helps us establish a connection between nonlinear complexities of $\mathbf{s}_n^\infty$ and $\mathbf{s}_n$.

Definition 2. *For $1 \le c < n$ and $1 \le d \le \min\{n-c, \lfloor \frac{n}{2} \rfloor\}$, we denote by $\mathcal{B}(n,c,d)$ the set of aperiodic binary sequences $\boldsymbol{s}_n$ starting with $\boldsymbol{s}_{c+d}$ in (1), namely,*

$$\mathcal{B}(n,c,d) = \{\, \boldsymbol{s}_n = \boldsymbol{s}_{c+d}\, \boldsymbol{s}_{[c+d:n]} = \underbrace{((s_0,\ldots,s_{d-1})^q\,(s_0,\ldots,s_{r-1},\overline{s}_r)}_{length=c+d}\, \boldsymbol{s}_{[c+d:n]})\,\} \tag{2}$$

where $\boldsymbol{s}_d$ is aperiodic, $\overline{s}_r = s_r \oplus 1$, and the subsequence $\boldsymbol{s}_{[c+d:n]}$ is chosen from $\mathbb{Z}_2^{n-c-d}$ such that $\boldsymbol{s}_n$ is aperiodic. We call the parameter d the spacing of $\boldsymbol{s}_n$, denoted by $spac(\boldsymbol{s}_n)$, and define

$$\mathcal{B}(n,c) = \bigcup_{d=1}^{\min\{n-c,\lfloor \frac{n}{2} \rfloor\}} \mathcal{B}(n,c,d). \tag{3}$$

By Corollary 1, for the case of $c \ge \lfloor \frac{n}{2} \rfloor$, any $\mathbf{s}_n \in \mathcal{B}(n,c)$ has nonlinear complexity c. When $c \ge \frac{n}{2}$, Lemma 4 shows that $\mathbf{s}_{c+d}$ actually contains the unique companion pair in $\mathbf{s}_n$; moreover, Lemma 2 (ii) indicates that the subsequence $\mathbf{s}_{[c+d:n]}$ in $\mathbf{s}_n$ can be arbitrarily chosen. It is to be noted that for each given n, c and d, the set $\mathcal{B}(n,c,d)$ is non-empty. As a matter of fact, as the subsequence $\mathbf{s}_{c+d}$ is aperiodic and $\mathbf{s}_{[c+d:n]}$ can be arbitrarily chosen. The existence of aperiodic sequences in $\mathcal{B}(n,c,d)$ can be easily confirmed.

The following lemma characterizes the change of nonlinear complexities of sequences in $\mathcal{B}(n,c)$ under the left circular shift operators.

Lemma 6. *For $\boldsymbol{s}_n \in \mathcal{B}(n,c,d)$ and a positive integer $t < c$, its left shift sequence $L^t(\boldsymbol{s}_n)$ belongs to $\mathcal{B}(n,c-t,d)$. In particular, when $c - t \ge \lfloor \frac{n}{2} \rfloor$, we have*

$$nlc(L^t(\boldsymbol{s}_n)) = nlc(\boldsymbol{s}_n) - t = c - t.$$

Proof. According to Definition 2, the sequence $\mathbf{s}_n \in \mathcal{B}(n,c,d)$ has the form

$$\mathbf{s}_n = \mathbf{s}_{c+d}\mathbf{s}_{[c+d:n]} = \underbrace{(s_0,\ldots,s_{d-1})^q(s_0,\ldots,s_{r-1},\overline{s}_r)}_{\text{length}=c+d}\, \mathbf{s}_{[c+d:n]}, \tag{4}$$

which implies $(s_0,s_1,\ldots,s_{c-2}) = (s_d,s_{d+1},\ldots,s_{c+d-2})$ and $s_{c-1} \ne s_{c+d-1}$. Let $\mathbf{a}_n = L(\mathbf{s}_n)$. Then $(a_0,\ldots,a_{c-3}) = (s_1,\ldots,s_{c-2}) = (s_{d+1},\ldots,s_{c+d-2}) = (a_d,\ldots,a_{c+d-3})$ and $a_{c-2} = s_{c-1} \ne s_{c+d-1} = a_{c+d-2}$. Thus we can write $\mathbf{a}_n$ as

$$L(\mathbf{s}_n) = \mathbf{a}_n = \mathbf{a}_{c+d-1}\mathbf{a}_{[c+d-1:n]} = \underbrace{(a_0,\ldots,a_{d-1})^{q_1}(a_0,\ldots,a_{r_1-1},\overline{a}_{r_1})}_{\text{length}=c+d-1}\, \mathbf{a}_{[c+d-1:n]}, \tag{5}$$

where $\mathbf{a}_d = (a_0, a_1, \ldots, a_{d-1}) = (s_1, s_2, \ldots, s_{d-1}, s_0)$, $q_1 = \lfloor \frac{(c+d-1)-1}{d} \rfloor = \lfloor \frac{c+d-2}{d} \rfloor$, $r_1 = (c+d-2) - q_1 d$ and $(a_0, \ldots, a_{r_1-1}, \overline{a}_{r_1})$ reduces to $\overline{a}_{r_1}$ when $r_1 = 0$. It is clear that $\mathbf{a}_d = L(\mathbf{s}_d)$ is aperiodic. Thus we have $L(\mathbf{s}_n) \in \mathcal{B}(n, c-1, d)$. Furthermore, it follows from Corollary 1 that $L(\mathbf{s}_n)$ has nonlinear complexity $c-1$ when $c-1 \geq \lfloor \frac{n}{2} \rfloor$. By repeatedly applying the above process, the desired statement follows. □

Lemma 6 can be interpreted alternatively: given a sequence $\mathbf{a}_n \in \mathcal{B}(n, c-1, d)$ in (5) with $d \leq n-c$ and $c-1 \geq \lfloor \frac{n}{2} \rfloor$, if $a_{n-1} = a_{d-1}$, then its right cyclic shift sequence $R(\mathbf{a}_n)$ has the form in (4), implying that $\mathbf{s}_n = R(\mathbf{a}_n)$ belongs to $\mathcal{B}(n, c, d)$ and has nonlinear complexity $nlc(\mathbf{a}_n) + 1$. With this observation, we introduce the following parameter of a sequence in $\mathcal{B}(n, c, d)$ with $c \geq \lfloor \frac{n}{2} \rfloor$, which indicates the potential increment of the nonlinear complexity of sequences under the right circular shift operators.

Definition 3. *Given a certain positive integer t, if a sequence $\boldsymbol{s}_n$ in $\mathcal{B}(n, c, d)$ satisfies*

$$s_{n-1-i} = s_{(d-1-i) \bmod d} \quad \text{for } 0 \leq i < t, \text{ and } s_{n-1-t} \neq s_{(d-1-t) \bmod d},$$

then we call $s_{n-t}, \ldots, s_{n-1}$ the added terms of $\boldsymbol{s}_n$ and denote by $add(\boldsymbol{s}_n)$ the number t of the added terms of $\boldsymbol{s}_n$.

Definition 3 indicates that $(s_{n-t}, \ldots, s_{n-1}) = (s_{(d-t) \bmod d}, \ldots, s_{(d-1) \bmod d})$. As shown in Figure 2, for the case that $t < d$, the subsequences $\mathbf{s}_{[d-t:d]}$ and $\mathbf{s}_{[n-t:n]}$ in gray are identical and $s_{d-1-t} \neq s_{n-1-t}$. In this case $s_{n-t}, \ldots, s_{n-1}$ in gray are the added terms of $\mathbf{s}_n$. For some sequences $\mathbf{s}_n$, we may have $add(\mathbf{s}_n) = t \geq d$. In such a case, it follows that $(s_{n-t}, \ldots, s_{n-d-1}, s_{n-d}, \ldots, s_{n-1}) = (s_{n-(t-d)}, \ldots, s_{n-1}, s_0, \ldots, s_{d-1})$.

$$\begin{aligned} \mathbf{s}_n &= (s_0, \ldots, s_{d-1})^q (s_0, \ldots, s_{r-1}, \overline{s}_r)\, (s_{c+d}, \ldots, s_{n-1}) \\ &= (s_0, \ldots, s_{d-1-t}, \underleftarrow{s_{d-t}, \ldots, s_{d-1}})(s_d, \ldots, s_{n-t-1}, \underleftarrow{s_{n-t}, \ldots, s_{n-1}}). \end{aligned}$$

Figure 2: Added terms in Definition 3 for $t < d$

Remark 1. *Note that each $\boldsymbol{s}_n$ in $\mathcal{B}(n, c)$ with $c \geq \lfloor \frac{n}{2} \rfloor$ has $add(\boldsymbol{s}_n) \leq n-c-1$. For $\boldsymbol{s}_n$ in $\mathcal{B}(n, c, d)$ with $c \geq \lfloor \frac{n}{2} \rfloor$, assume $add(\boldsymbol{s}_n) = t$, we now consider the sequence $\boldsymbol{s}_n^2 = \boldsymbol{s}_n \boldsymbol{s}_n$. From Definition 3, in the subsequence $(s_{n-t}, \ldots, s_{n-1}, s_0, \ldots, s_{c+d-1})$ of $\boldsymbol{s}_n^2$ we have*

$$(s_{n-t}, \ldots, s_{n-1}, s_0, \ldots, s_{c-2}) = (s_{n-t+d}, \ldots, s_{d-1}, s_d, \ldots, s_{c+d-2}) \quad \text{and} \quad s_{c-1} \neq s_{c+d-1}.$$

Let $\boldsymbol{v}_d = (v_0, \ldots, v_{d-1}) = (s_{n-t}, \ldots, s_{n-t+d-1})$. *Then*

$$(s_{n-t}, \ldots, s_{n-1}, s_0, \ldots, s_{c+d-1}) = \overbrace{(v_0, \ldots, v_{d-1}) \cdots (v_0, \ldots, v_{d-1})}^{q_2 \text{ repetitions}}(v_0, \ldots, v_{r_2-1}, \overline{v}_{r_2}),$$

where $q_2 = \lfloor \frac{c+d+t-1}{d} \rfloor$ *and* $r_2 = (c+d+t-1) - q_2 d$. *Thus the above subsequence satisfies* $(s_{n-t}, \ldots, s_{n-1}, s_0, \ldots, s_{c+d-1}) \in \mathcal{B}(c+d+t, c+t, d)$ *with nonlinear complexity* $c+t$. *Then* $c + t \leq nlc(\boldsymbol{s}_n^2) \leq n-1$, *implying* $t \leq n-c-1$.

For a binary sequence $\mathbf{s}_n \in \mathcal{B}(n, c)$, the following proposition shows that $add(\mathbf{s}_n)$ plays an important role in the varying behaviour of the nonlinear complexity of sequences under the right circular shift operators.

Proposition 1. *For* $\boldsymbol{s}_n \in \mathcal{B}(n, c, d)$ *with* $c \geq \lfloor \frac{n}{2} \rfloor$, $d \leq \min\{n-c, \lfloor \frac{n}{2} \rfloor\}$ *and* $add(\boldsymbol{s}_n) = t$, *the nonlinear complexity values of its shifted sequences have the following properties,*

(i) *for any* $1 \leq k \leq \min\{t, n-c-d\}$, $R^k(\boldsymbol{s}_n) \in \mathcal{B}(n, c+k, d)$ *and* $nlc(R^k(\boldsymbol{s}_n)) = c+k$;

(ii) *for any* $t < k \leq n-c-d$, *we have* $nlc(R^k(\boldsymbol{s}_n)) = c+t$.

Proof. (i) For $\mathbf{s}_n \in \mathcal{B}(n, c, d)$ with $add(\mathbf{s}_n) = t$, we have $(s_{n-t}, \ldots, s_{n-1}) = (s_{(d-t) \bmod d}, \ldots, s_{(d-1) \bmod d})$, thus

$$(s_{n-t}, \ldots, s_{n-1})(s_0, \ldots, s_{c-2}) = (s_{(d-t) \bmod d}, \ldots, s_{(d-1) \bmod d})(s_d, s_{d+1}, \ldots, s_{c+d-2}),$$

and $s_{c-1} \neq s_{c+d-1}$. When $n-c-d \geq 1$, let $\mathbf{a}_n = R(\mathbf{s}_n)$, thus

$$(a_0, a_1, \ldots, a_{c-1}) \ = (s_{n-1}, s_0, \ldots, s_{c-2}) = (s_{d-1}, s_d, \ldots, s_{c+d-2}) = (a_d, a_{d+1}, \ldots, a_{c+d-1})$$

and $a_c \neq a_{c+d}$. That is to say, we can write $\mathbf{a}_n$ as

$$R(\mathbf{s}_n) = \mathbf{a}_n = \mathbf{a}_{c+d+1}\mathbf{a}_{[c+d+1:n]} = \underbrace{(a_0, \ldots, a_{d-1})^{q_2}(a_0, \ldots, a_{r_2-1}, \overline{a}_{r_2})}_{\text{length}=c+d+1} \mathbf{a}_{[c+d+1:n]},$$

where $q_2 = \lfloor \frac{(c+d+1)-1}{d} \rfloor = \lfloor \frac{c+d}{d} \rfloor$ and $r_2 = (c+d+1) - 1 - q_2 d$. So we see that the sequence $\mathbf{a}_n = R(\mathbf{s}_n)$ belongs to $\mathcal{B}(n, c+1, d)$ and has nonlinear complexity $c+1$.

When $t \leq n-c-d$, by induction on t, we see that the sequence $\mathbf{a}_n = R^k(\mathbf{s}_n)$ with $k \leq t$ belongs to $\mathcal{B}(n, c+k, d)$ and has nonlinear complexity $c+k$. When $t > n-c-d$ and $k \leq n-c-d$, the statement holds similarly. However, when $t > n-c-d$ and $n-c-d < k \leq t$, the sequence $\mathbf{a}_n = R^k(\mathbf{s}_n)$ satisfies $a_i = a_{i+d}$ with $0 \leq i < n-d$ and

$d \le \min\{n-c, \lfloor \frac{n}{2} \rfloor\} \le \lfloor \frac{n}{2} \rfloor$, implying that $\mathbf{a}_n$ is contained in the periodic sequence $\mathbf{a}_d^\infty$ and then $nlc(\mathbf{a}_n) \le nlc(\mathbf{a}_d^\infty) \le d-1 \le \lfloor \frac{n}{2} \rfloor - 1$, that is to say, $\mathbf{a}_n = R^k(\mathbf{s}_n)$ no longer belongs to $\mathcal{B}(n, c+k, d)$. The difference between cases of $t \le n-c-d$ and $t > n-c-d$ are visualized in Figure 3, where a copy of $\mathbf{s}_n$ is added on its left side and the gray area covers the sequence $\mathbf{s}_{c+d}$ as in (1). Therefore, for any $1 \le k \le \min\{t, n-c-d\}$, $nlc(R^k(\mathbf{s}_n)) = nlc(\mathbf{s}_n) + k = c + k$.

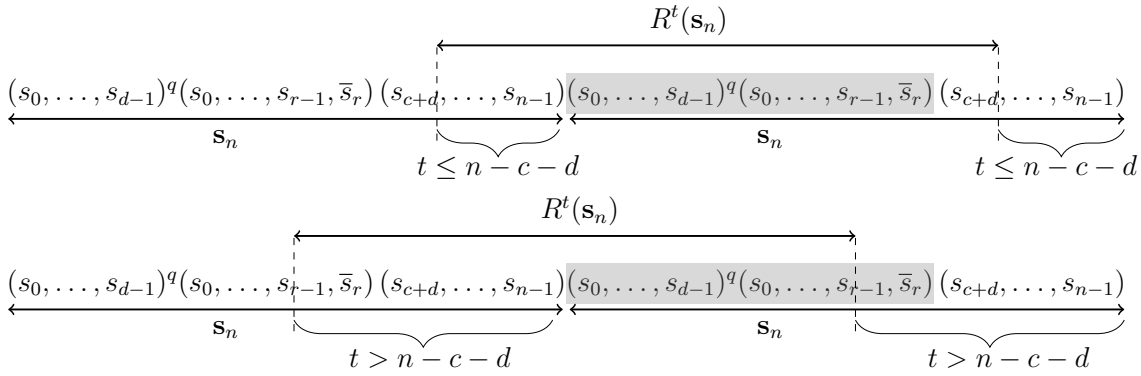

Figure 3: The visualized description of Proposition 1 (i)

(ii) Let $R^t(\mathbf{s}_n) = \mathbf{a}_n$. It follows from Proposition 1 (i) and $t < n-c-d$ that

$$R^t(\mathbf{s}_n) = \mathbf{a}_n \in \mathcal{B}(n, c+t, d) \text{ with } a_{n-1} \ne a_{d-1} \tag{6}$$

and $nlc(\mathbf{a}_n) = c+t$. Then $\mathbf{a}_{c+t+d}$ is the subsequence formed exactly by the companion pair $(\underline{\mathbf{a}}_0, \mathcal{A}^d(\underline{\mathbf{a}}_0))$ of $\mathbf{a}_n$. For any $t < k \le n-c-d$, we have $R^k(\mathbf{s}_n) = R^{k-t}(\mathbf{a}_n)$ contains $\mathbf{a}_{c+t+d}$, which implies that $nlc(R^k(\mathbf{s}_n)) \ge c+t$. We shall prove the statement by induction on k. Suppose that the nonlinear complexity of $R^{t+1}(\mathbf{s}_n) = R(\mathbf{a}_n)$ is larger than $c+t$, that is to say, $nlc(R(\mathbf{a}_n)) = c' > c+t$. Note that it follows from Lemma 4 and $c' > \lfloor \frac{n}{2} \rfloor$ that $R^k(\mathbf{s}_n)$ only has a unique companion pair. If the unique companion pair of $R(\mathbf{a}_n) = (a_{n-1}, a_0, \ldots, a_{n-2})$ does not begin with a_{n-1}, then $L(R(\mathbf{a}_n)) = \mathbf{a}_n = (a_0, \ldots, a_{n-2}, a_{n-1})$ contains the unique companion pair of $R(\mathbf{a}_n)$, which implies that $nlc(\mathbf{a}_n) \ge nlc(R(\mathbf{a}_n)) = c'$. It contradicts that $nlc(\mathbf{a}_n) = c+t < c' = nlc(R(\mathbf{a}_n))$. Thus the unique companion pair of $R(\mathbf{a}_n) = (a_{n-1}, a_0, \ldots, a_{n-2})$ begins with a_{n-1}, which yields that $R(\mathbf{a}_n) \in \mathcal{B}(n, c', d')$. Hence $a_{n-1} = a_{d'-1}$. Together with (6), we have $a_{d'-1} = a_{n-1} \ne a_{d-1}$, implying that $d' \ne d$. According to Lemma 6, $L(R(\mathbf{a}_n)) = \mathbf{a}_n \in \mathcal{B}(n, c'-1, d')$ with $nlc(\mathbf{a}_n) = c'-1$ and $d' \ne d$. Moreover $\mathbf{a}_n \in \mathcal{B}(n, c+t, d)$ with $nlc(\mathbf{a}_n) = c+t$, thus $c' = c+t+1$. That is to

say, $\mathbf{a}_n$ satisfies $\mathbf{a}_n \in \mathcal{B}(n, c+t, d)$ and $\mathbf{a}_n \in \mathcal{B}(n, c+t, d')$ with $d' \neq d$. From the structure of $\mathbf{a}_n$ in (1), it is clear that

$$a_i = a_{i+d},\ a_i = a_{i+d'},\ 0 \leq i \leq c-2 \quad \text{and} \quad a_{c-1} \neq a_{c+d-1},\ a_{c-1} \neq a_{c+d'-1}. \tag{7}$$

In the case of $c+t \geq n/2$, it follows from Lemma 4 that $\mathbf{a}_n$ only has a unique companion pair, which contradicts that $d' \neq d$. In the case of $c = \lfloor \frac{n}{2} \rfloor$ with odd n and $t = 0$, i.e. $c+t = \lfloor \frac{n}{2} \rfloor$, without loss of generality, suppose $d' < d$ then $1 \leq d' < d \leq \lfloor \frac{n}{2} \rfloor$. If $d < \lfloor \frac{n}{2} \rfloor$, then according to (7), one has $a_{c-1-d} = a_{c-1}$ and $a_{c-1-d} = a_{c+d'-1-d}$ implying that $a_{c-1} = a_{c+d'-1-d}$, while $a_{c+d'-1-d} = a_{c+d'-1}$ and $a_{c-1} \neq a_{c+d'-1}$ follows $a_{c-1} \neq a_{c+d'-1-d}$. It is a contradiction. If $d = \lfloor \frac{n}{2} \rfloor$, then from (7) we can see that

$$(a_c, a_{c+1}, \ldots, a_{c+d'-1}) = (a_{c-d}, a_{c+1-d}, \ldots, a_{c+d'-1-d}) = (a_0, a_1, \ldots, a_{d'-1}) = \mathbf{a}_{d'},$$

thus its Hamming weight $\mathrm{wt}(a_c, a_{c+1}, \ldots, a_{c+d'-1}) = \mathrm{wt}(\mathbf{a}_{d'})$. While again by (7), one has

$$(a_c, a_{c+1}, \ldots, a_{c+d'-2}, a_{c+d'-1}) = (a_{c \bmod d'}, \ldots, a_{(c+d'-2) \bmod d'}, a_{(c+d'-1) \bmod d'} \oplus 1),$$

where $(a_{c \bmod d'}, \ldots, a_{(c+d'-2) \bmod d'}, a_{(c+d'-1) \bmod d'})$ is a shifted version of $\mathbf{a}_{d'}$. Hence $\mathrm{wt}(\mathbf{a}_{d'}) \neq \mathrm{wt}(a_c, a_{c+1}, \ldots, a_{c+d'-1})$, which is a contradiction. Thus

$$nlc(R(\mathbf{a}_n)) = nlc(R^{t+1}(\mathbf{s}_n)) = c' = c + t.$$

By induction on k ranging from $t+1$ to $n-c-d$, the proof follows. □

Given a sequence $\mathbf{s}_n \in \mathcal{B}(n, c, d)$, Proposition 1 reveals the changing pattern of $nlc(R^k(\mathbf{s}_n))$ for $k \leq n-c-d$. However, when $k > n-c-d$, the change of $nlc(R^k(\mathbf{s}_n))$ does not indicate a strong pattern. Consider a sequence $\mathbf{s}_n \in \mathcal{B}(n, c)$ with $c \geq \lfloor \frac{n}{2} \rfloor$. The varying behaviour of $nlc(R^k(\mathbf{s}_n))$ for $0 \leq k \leq n-1$ is discussed in Remark 2.

Remark 2. *Without loss of generality, we assume $c = \lfloor \frac{n}{2} \rfloor$. For $\boldsymbol{s}_n \in \mathcal{B}(n, \lfloor \frac{n}{2} \rfloor)$, among the right shift sequences $R^k(\boldsymbol{s}_n)$: $k = 0, 1, \ldots, n-1$, there may exist positive integers $0 < j_1 < \cdots < j_r \leq n-1$ such that $R^{j_1}(\boldsymbol{s}_n), \ldots, R^{j_r}(\boldsymbol{s}_n)$ also belong to $\mathcal{B}(n, \lfloor \frac{n}{2} \rfloor)$.*

For $0 \leq k < j_1$, assume $\boldsymbol{s}_n \in \mathcal{B}(n, \lfloor \frac{n}{2} \rfloor, d)$ for certain $d \leq \lfloor \frac{n}{2} \rfloor$ has $add(\boldsymbol{s}_n) = t$ and define an integer-valued sequence $\boldsymbol{c}_0 = (nlc(\boldsymbol{s}_n), \ldots, nlc(R^{j_1-1}(\boldsymbol{s}_n)))$. Then the values in the sequence $\mathbf{c}_0$ vary in this way: $\boldsymbol{c}_0$ starts with $\lfloor \frac{n}{2} \rfloor$, when $k \leq \min\{t, n-c-d\}$, the values in $\boldsymbol{c}_0$ increase one by one as in Proposition 1 (i); when $t < k \leq n-c-d$, the

values in $\boldsymbol{c}_0$ remain unchanged as in Proposition 1 (ii); and when $n-c-d<k<j_1$, the change of values in $\boldsymbol{c}_0$ does not exhibit a strong pattern.

For $k=j_1$, assume $\boldsymbol{a}_n=R^{j_1}(\boldsymbol{s}_n)\in\mathcal{B}(n,c,d_1)$ has $add(\boldsymbol{a}_n)=t_1$ and take

$$\boldsymbol{c}_1=\big(nlc(R^{j_1}(\boldsymbol{s}_n)),\ldots,nlc(R^{j_2-1}(\boldsymbol{s}_n))\big)=\Big(nlc(\boldsymbol{a}_n),\ldots,nlc(R^{j_1'-1}(\boldsymbol{a}_n))\Big),$$

where $j_1'=j_2-j_1$. As k increases from j_1 to j_2-1, the values in the sequence $\boldsymbol{c}_1$ vary in a similar manner as the above analysis for $\boldsymbol{c}_0$. As k ranges from j_2 to $n-1$, we can similarly define the sequences $\boldsymbol{c}_i$ for $i=2,\ldots,r$ as above. The values in these sequences vary similarly to that of $\boldsymbol{c}_0$.

To sum up, the above analysis shows that

$$\big(nlc(\boldsymbol{s}_n),nlc(R(\boldsymbol{s}_n),\ldots,nlc(R^{n-1}(\boldsymbol{s}_n))\big)=\boldsymbol{c}_0\cup\boldsymbol{c}_1\cup\cdots\cup\boldsymbol{c}_r,$$

where the values in each $\boldsymbol{c}_i$, $i=0,1,\ldots,r$, vary according to Proposition 1 first and then change in unclear patterns.

For sequences $\mathbf{s}_n\in\mathcal{B}(n,c)$, below we consider a subset $E(\mathbf{s}_n)$ of its shift equivalence class and the representative of $E(\mathbf{s}_n)$, which will be used in our subsequent discussions.

Definition 4. *Let $\boldsymbol{s}_n$ be a sequence in $\mathcal{B}(n,c)$ and $E(\boldsymbol{s}_n)=\{R^k(\boldsymbol{s}_n):0\le k<n\}\cap\mathcal{B}(n,c)$. A sequence $\widetilde{\boldsymbol{s}}_n\in E(\boldsymbol{s}_n)$ satisfying*

$$add(\widetilde{\boldsymbol{s}}_n)\ge add(\boldsymbol{a}_n),\ \forall\,\boldsymbol{a}_n\in E(\boldsymbol{s}_n)$$

is said to be a representative sequence of $\boldsymbol{s}_n$. Furthermore, we define $\mathcal{R}(n,c)$ as the set of all sequence representatives in $\mathcal{B}(n,c)$, i.e.,

$$\mathcal{R}(n,c)=\bigcup_{\boldsymbol{s}_n\in\mathcal{B}(n,c)}\big\{\widetilde{\boldsymbol{s}}_n\in E(\boldsymbol{s}_n):add(\widetilde{\boldsymbol{s}}_n)\ge add(\boldsymbol{a}_n),\forall\,\boldsymbol{a}_n\in E(\boldsymbol{s}_n)\big\}.\tag{8}$$

The following example illustrates some definitions and results in this section.

Example 1. *Consider $n=9$ and a binary finite-length sequence $\boldsymbol{s}_9=(000010010)$. It is clear that $\boldsymbol{s}_9=\boldsymbol{s}_5\boldsymbol{s}_{[5:9]}=(\underline{\mathbf{0}0001}0010)\in\mathcal{B}(9,4,1)$ as in Definition 2, where $\boldsymbol{s}_{c+d}$ is underlined and $\boldsymbol{s}_d$ is in bold for $d=1$ and $c=4$. By Definition 3 we have $add(\boldsymbol{s}_9)=1$ since $s_8=0=s_0$ (which is displayed in bold) and $s_7\ne s_0$. Since $add(\boldsymbol{s}_9)=t=1<n-c-d=4$, from Proposition 1 we have $nlc(R(\boldsymbol{s}_n))=c+1=5$ and $nlc(R^k(\boldsymbol{s}_9))=c+t=5$ for $k=2,3,4$.*

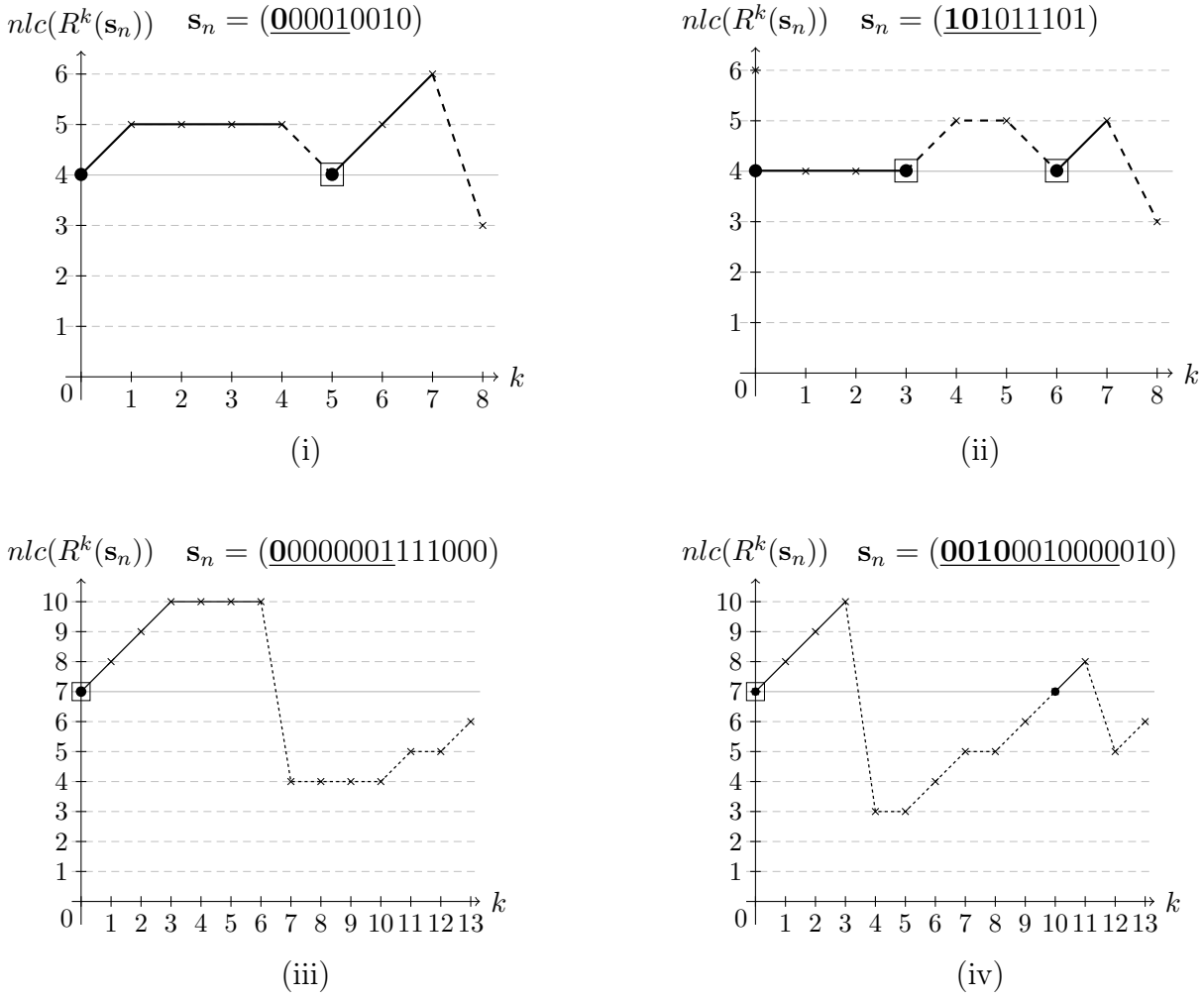

Figure 4: The visualized description of Example 1

Furthermore, among the shift sequences $R^k(\boldsymbol{s}_9)$ *for* $k = 0, 1, \ldots, 8$*, only the sequences* $\boldsymbol{s}_9$ *and* $R^5(\boldsymbol{s}_9)$ *belong to* $\mathcal{B}(9, 4)$*, indicating that* $E(\boldsymbol{s}_9) = \{\boldsymbol{s}_9, R^5(\boldsymbol{s}_9)\}$ *from Definition 4. Observe that*

$$\big(nlc(R^k(\boldsymbol{s}_9)) : k = 0, 1, \ldots, 8\,\big) = (4, 5, 5, 5, 5) \cup (4, 5, 6, 3\,) = \mathbf{c}_0 \cup \mathbf{c}_1$$

where the values in $\boldsymbol{c}_0 = \big(nlc(R^k(\boldsymbol{s}_9)) : k = 0, 1, 2, 3, 4\,\big)$ *and* $\mathbf{c}_1 = (nlc(R^k(\boldsymbol{s}_9)) : k = 5, 6, 7, 8)$ *vary as in Proposition 1 and the analysis is as in Remark 2. In addition, since* $add(R^5(\boldsymbol{s}_9)) = 2 > add(\boldsymbol{s}_9) = 1$*, the representative of* $E(\boldsymbol{s}_9)$ *is* $R^5(\boldsymbol{s}_9)$*, which belongs to* $\mathcal{R}(9, 4)$ *as in Definition 4.*

The above discussion for the sequence $\boldsymbol{s}_9$ is visualized in Figure 4 (i), where the axises represent the values k and the corresponding $nlc(R^k(\boldsymbol{s}_9))$, respectively. In Figure 4 (i), the sequences in $E(\boldsymbol{s}_9)$ are displayed as black solid dots, and the representative sequence of $E(\boldsymbol{s}_9)$ is marked with a rectangle. In addition, the solid line shows the varying behavior of $nlc(R^k(\boldsymbol{s}_9))$ that can be explained by Proposition 1 and the dotted line shows the varying behavior of $nlc(R^k(\boldsymbol{s}_9))$ that does not exhibit a strong pattern.

Sequences $\boldsymbol{s}_n \in \mathcal{B}(n, \lfloor \frac{n}{2} \rfloor)$ may behave quite differently. To illustrate this, we further consider three sequences $\boldsymbol{s}_9 = (\underline{\mathbf{10}1011}101) \in \mathcal{B}(9,4,2)$, $\boldsymbol{s}_{14} = (\underline{\mathbf{0}0000001}111000) \in \mathcal{B}(14,7,1)$ and $\boldsymbol{s}_{14} = (\underline{\mathbf{0010}0010000}010) \in \mathcal{B}(14,7,4)$ where $\boldsymbol{s}_{c+d}$ is underlined and $\boldsymbol{s}_d$ is in bold. For these sequences, Figure 4 (ii), (iii), (iv) displays the sequences in $E(\boldsymbol{s}_n)$, the representative sequences in $E(\boldsymbol{s}_n)$, and the varying behavior of $nlc(R^k(\boldsymbol{s}_n))$ for $k = 0, 1, \ldots, n-1$, respectively.

At the end of this section, we summarize important notations in Table 1 for readers' convenience.

4 The nonlinear complexity of $\mathbf{s}_n$ and $\mathbf{s}_n^\infty$

This section investigates the relation between the nonlinear complexity of periodic sequences and that of finite-length sequences in $\mathcal{B}(n,c)$. Let $\mathcal{P}(n,\omega)$ denote the set of n-periodic binary sequences with nonlinear complexity ω, where $\lceil \log_2(n) \rceil \le \omega \le n-1$ as indicated by Lemma 3 (i). Below we first discuss some properties of subsequences of a periodic binary sequence $\mathbf{s}_n^\infty$.

Lemma 7. *Given a sequence $\boldsymbol{a}_n^\infty$ in $\mathcal{P}(n,\omega)$, suppose its left shift sequence $\boldsymbol{s}_n^\infty = L^i(\boldsymbol{a}_n^\infty)$ has a companion pair $(\underline{\boldsymbol{s}}_0, \mathcal{F}^d(\underline{\boldsymbol{s}}_0))$ with $d \le \lfloor \frac{n}{2} \rfloor$, where $\underline{\boldsymbol{s}}_0 = (s_0, \ldots, s_{\omega-1})$. Then,*

(i) *$\boldsymbol{s}_{\omega+d} = (s_0, \ldots, s_{\omega+d-1}) \in \mathcal{B}(\omega+d, \omega, d)$ and $nlc(\boldsymbol{s}_{\omega+d}) = \omega$ if $\omega \ge d$;*

(ii) *when $\omega + d \le n$, one has $\boldsymbol{s}_n = (s_0, \ldots, s_{n-1}) \in \mathcal{B}(n, \omega, d)$ and $nlc(\boldsymbol{s}_n) = \omega$ if $\omega \ge d$;*

(iii) *when $\omega + d > n$, for any $0 \le j < \omega - 1$ and $n_1 = (\omega + d) - j$, the subsequence $\boldsymbol{s}_{[j:\omega+d]} = (s_j, \ldots, s_{\omega+d-1}) \in \mathcal{B}(n_1, n_1 - d, d)$.*

Proof. (i) For the subsequence $\mathbf{s}_{\omega+d}$ that starts from $\underline{\mathbf{s}}_0$ and ends with $\underline{\mathbf{s}}_d = \mathcal{F}^d(\underline{\mathbf{s}}_0) = \widehat{\underline{\mathbf{s}}}_0$, we have $(s_0, \ldots, s_{\omega-2}) = (s_d, \ldots, s_{\omega-2+d})$ and $s_{\omega-1} = s_{\omega-1+d} \oplus 1$. Thus we can write $\mathbf{s}_{\omega+d}$

Table 1: Some necessary notations of this paper.

notation	description
$\mathbf{s}_n$	an aperiodic sequence $(s_0, s_1, \ldots, s_{n-1})$
$\mathbf{s}_{[i:i+k]}$	the subsequence $(s_i, \ldots, s_{i+k-1})$
$\mathbf{s}_{c+d}$	$\mathbf{s}_d^q(s_0, \ldots, s_{r-1}, \overline{s}_r)$ as in (1)
$\mathcal{B}(n, c, d)$	the set $\{\mathbf{s}_{c+d}\,\mathbf{s}_{[c+d:n]}\}$ with $1 \le c < n$, $d \le \min\{n-c, \lfloor \frac{n}{2} \rfloor\}$
$\mathcal{B}(n, c)$	$\bigcup_{d=1}^{\min\{n-c, \lfloor \frac{n}{2} \rfloor\}} \mathcal{B}(n, c, d)$
$add(\mathbf{s}_n)$	for $\mathbf{s}_n \in \mathcal{B}(n, c, d)$, the integer t such that $s_{n-1-i} = s_{(d-1-i) \bmod d}, \forall\, 0 \le i < t$, and $s_{n-1-t} \ne s_{(d-1-t) \bmod d}$
$\mathcal{B}_0(n, c)$	$\mathcal{B}_0(n, c) = \{\mathbf{s}_n \in \mathcal{B}(n, c) : add(\mathbf{s}_n) = 0\}$
$E(\mathbf{s}_n)$	$\{R^k(\mathbf{s}_n) : 0 \le k < n\} \cap \mathcal{B}(n, c)$ for $\mathbf{s}_n \in \mathcal{B}(n, c)$
$\mathcal{R}(n, c)$	$\bigcup_{\mathbf{s}_n \in \mathcal{B}(n,c)} \{\widetilde{\mathbf{s}}_n \in E(\mathbf{s}_n) : add(\widetilde{\mathbf{s}}_n) \ge add(\mathbf{a}_n), \forall\, \mathbf{a}_n \in E(\mathbf{s}_n)\}$
$\mathbf{s}_n^\infty$	the periodic sequence $(\mathbf{s}_n)^\infty$ from $\mathbf{s}_n$
$\mathcal{P}(n, \omega)$	the set of sequences with period n and nonlinear complexity ω
$\overline{S}$	$\overline{S} = \{L^k(\mathbf{s}_n^\infty) : \mathbf{s}_n^\infty \in S, 0 \le k < n\}$
$S_1 \cong S_2$	$\overline{S}_1 = \overline{S}_2$ for two sets S_1 and S_2 of periodic sequences

as

$$\mathbf{s}_{\omega+d} = (s_0, \ldots, s_{d-1}, s_d, \ldots, s_{\omega+d-2}, \overline{s}_{\omega-1}) = (\mathbf{s}_d^q(s_0, \ldots, s_{r-1}, \overline{s}_r)).$$

The statement directly follows from Definition 2 and Lemma 5.

(ii) If $\omega + d \le n$, then $\mathbf{s}_{\omega+d}$ is contained in $\mathbf{s}_n$, namely,

$$\mathbf{s}_n = (s_0, \ldots, s_{\omega+d-1}, s_{\omega+d}, \ldots, s_{n-1}) = \mathbf{s}_{\omega+d}\,\mathbf{s}_{[\omega+d:n]}.$$

If $\omega \ge d$, then it implies $nlc(\mathbf{s}_n) \ge nlc(\mathbf{s}_{\omega+d}) = \omega$. Since $nlc(\mathbf{s}_n) \le nlc(\mathbf{s}_n^\infty) = \omega$, we have $nlc(\mathbf{s}_n) = \omega$ and $\mathbf{s}_n \in \mathcal{B}(n, \omega, d)$.

(iii) For the case of $\omega + d > n$, since the subsequence $\mathbf{s}_{\omega+d}$ belongs to $\mathcal{B}(\omega + d, \omega, d)$, it follows that $s_{i_1} = s_{i_1+d}$ for $0 \le i_1 < \omega - 1$ and $s_{\omega-1} = \overline{s}_{\omega-1+d}$. Hence for any $j \ge 0$, the

subsequence $\mathbf{s}_{[j:\omega+d]}$ also satisfies the relation

$$s_{i_1} = s_{i_1+d} \ \text{ for } \ j \le i_1 < \omega - 1 \ \text{ and } \ s_{\omega-1} = \overline{s}_{\omega-1+d}. \tag{9}$$

This implies that for any j with $0 \le j < \omega - 1$, the subsequence $\mathbf{s}_{[j:\omega+d]}$ has the form

$$((s_j, \ldots, s_{j+d-1})^q (s_j, \ldots, s_{j+r-1}, \overline{s}_{j+r}))$$

where $q = \lfloor \frac{\omega+d-j-1}{d} \rfloor$, $r = (\omega + d - j - 1) - qd$. According to Definition 2, the subsequence $\mathbf{s}_{[j:\omega+d]}$ belongs to $\mathcal{B}(n_1, n_1 - d, d)$ with $n_1 = \omega + d - j$. □

With the introduced definitions, we present the main theorems of this paper below.

Theorem 1. *For any $\boldsymbol{s}_n^\infty$ in $\mathcal{P}(n, \omega)$ and any integer c with $\lceil \log_2(n) \rceil \le c \le \min\{\omega, \lfloor \frac{n}{2} \rfloor + 1\}$, there exists an integer k such that $L^k(\boldsymbol{s}_n) \in \mathcal{B}(n, c)$. Furthermore,*

$$\bigcup_{\omega=c}^{n-1} \mathcal{P}(n, \omega) = \left\{ (L^k(\boldsymbol{s}_n))^\infty \ : \ \boldsymbol{s}_n \in \mathcal{B}(n, c),\ 0 \le k < n \right\}.$$

On the other hand, given a sequence $\mathbf{s}_n$ in $\mathcal{B}(n, c, d)$ with $c \ge \lfloor \frac{n}{2} \rfloor$ and $add(\mathbf{s}_n) = t$, its companion pairs are of the form $\mathbf{s}_{c+d}$ as in (1). This property leads to the statement in Proposition 1 (i), which implies

$$add(R^k(\mathbf{s}_n)) = add(\mathbf{s}_n) - k \ \text{ and } \ nlc(R^k(\mathbf{s}_n)) = nlc(\mathbf{s}_n) + k$$

for any $1 \le k \le \min\{t, n - c - d\}$. This observation results in a more explicit expression of the nonlinear complexity of $\mathbf{s}_n^\infty$ from its n-length subsequences in $\mathcal{R}(n, c)$.

Theorem 2. *For $\boldsymbol{s}_n$ in $\mathcal{R}(n, c)$ with $c \ge \lfloor \frac{n}{2} \rfloor$, one has $nlc(\boldsymbol{s}_n^\infty) = nlc(\boldsymbol{s}_n) + add(\boldsymbol{s}_n)$.*

4.1 Proof of Theorem 1 with $c \le \lfloor \frac{n}{2} \rfloor + 1$

Proof of Theorem 1. For the periodic sequence $\mathbf{a}_n^\infty$ in $\mathcal{P}(n, \omega)$, suppose its left shift sequence $\mathbf{s}_n^\infty = L^i(\mathbf{a}_n^\infty)$ has a companion pair $(\underline{\mathbf{s}}_0, \mathcal{F}^d(\underline{\mathbf{s}}_0))$ with $d \le \lfloor \frac{n}{2} \rfloor$, we shall show that $L^{\omega-c}(\mathbf{s}_n)$ belongs to $\mathcal{B}(n, c)$. According to Lemma 7 (ii), if $\omega + d \le n$ then the subsequence $\mathbf{s}_n$ belongs to $\mathcal{B}(n, \omega, d)$. This together with Lemma 6 implies that $L^{\omega-c}(\mathbf{s}_n) \in \mathcal{B}(n, c, d)$; if $\omega + d > n$, then letting $j = \omega + d - n$ we have $L^j(\mathbf{s}_n) \in \mathcal{B}(n, n - d, d)$, where $n - d \ge \lceil \frac{n}{2} \rceil$. We need to consider two cases: $n - d \ge c$ and $n - d < c$. For the case that $n - d \ge c$,

by applying Lemma 6, one has $L^{j+(n-d-c)}(\mathbf{s}_n) = L^{\omega-c}(\mathbf{s}_n) \in \mathcal{B}(n,c,d)$; for the case that $n-d<c$, since $n-d \geq \lceil \frac{n}{2} \rceil$, we have $c = \frac{n}{2}+1$ and $d = \frac{n}{2}$, indicating $L^j(\mathbf{s}_n) \in \mathcal{B}(n, \frac{n}{2}, \frac{n}{2})$. That is to say,

$$L^j(\mathbf{s}_n) = (s_j, \ldots, s_{j+n-1}) = \left((s_j, \ldots, s_{j+\frac{n}{2}-2}, s_{j+\frac{n}{2}-1})(s_j, \ldots, s_{j+\frac{n}{2}-2}, \overline{s}_{j+\frac{n}{2}-1})\right).$$

Due to $nlc(\mathbf{s}_n^\infty) = \omega \geq c = \frac{n}{2}+1$, the subsequence $\mathbf{s}_{[j-1:\omega+d]}$ belongs to $\mathcal{B}(n+1, n+1-d, d)$ with $d = \frac{n}{2}$. Thus it follows from (9) that $s_{j-1} = s_{j-1+d}$ and $s_{j-1+(n-d)} \neq s_{j-1+(n-d)+d} = s_{j-1+n} = s_{j-1}$. This is a contradiction since $s_{j-1+d} = s_{j-1+(n-d)}$ for $d = \frac{n}{2}$. Hence one always has $L^{\omega-c}(\mathbf{s}_n) \in \mathcal{B}(n,c)$, the first statement follows.

Therefore, for any sequence $\mathbf{a}_n^\infty$ in $\mathcal{P}(n,\omega)$ and any integer c with $\lceil \log_2(n) \rceil \leq c \leq \min\{\omega, \lfloor \frac{n}{2} \rfloor + 1\}$, we have

$$\mathbf{a}_n^\infty \in \{(L^k(\mathbf{s}_n))^\infty : \mathbf{s}_n \in \mathcal{B}(n,c), 0 \leq k < n\}.$$

In addition, it is clear that $\left\{(L^k(\mathbf{s}_n))^\infty : \mathbf{s}_n \in \mathcal{B}(n,c),\ 0 \leq k < n\right\} \subset \bigcup_{\omega=c}^{n-1} \mathcal{P}(n,\omega)$. Thus for $\lceil \log_2(n) \rceil \leq c \leq \min\{\omega, \lfloor \frac{n}{2} \rfloor + 1\}$, one has

$$\bigcup_{\omega=c}^{n-1} \mathcal{P}(n,\omega) = \left\{(L^k(\mathbf{s}_n))^\infty : \mathbf{s}_n \in \mathcal{B}(n,c),\ 0 \leq k < n\right\}. \qquad \square$$

Theorem 1 presents a one-to-one correspondence between the finite-length sequences set and a set of all periodic sequences with nonlinear complexity not less than c. For each periodic sequence $\mathbf{s}_n^\infty$ with nonlinear complexity $\omega \geq c$, Theorem 1 indicates the structure of an its n-length subsequence $\mathbf{a}_n$, which is in the form of (2). According to the structure of finite-length sequences $\mathbf{a}_n$, we can obtain all periodic sequences with nonlinear complexities not more than $\frac{n}{2}$ in Subsection 5.1.

When $c = \lfloor \frac{n}{2} \rfloor$ or $\lfloor \frac{n}{2} \rfloor + 1$, from Corollary 1 the nonlinear complexity of the n-length subsequence $\mathbf{a}_n$ can be determined, i.e. $nlc(\mathbf{a}_n) = c$. In what follows, we shall determine the exact value of nonlinear complexity ω of periodic sequences in $\{(L^k(\mathbf{s}_n))^\infty : \mathbf{s}_n \in \mathcal{B}(n,c),\ 0 \leq k < n\}$, where ω belongs to the set $[c, n-1]$ and $c \geq \lfloor \frac{n}{2} \rfloor$. With the further analysis, we can obtain all periodic sequences with nonlinear complexities not less than $\frac{n}{2}$ in Subsection 5.2.

4.2 Proof of Theorem 2 with $c \geq \lfloor \frac{n}{2} \rfloor$

Recall from (8) that

$$\mathcal{R}(n,c) = \bigcup_{\mathbf{s}_n \in \mathcal{B}(n,c)} \{\widetilde{\mathbf{s}}_n \in E(\mathbf{s}_n) : add(\widetilde{\mathbf{s}}_n) \geq add(\mathbf{a}_n), \forall\, \mathbf{a}_n \in E(\mathbf{s}_n)\},$$

where $E(\mathbf{s}_n) = \{L^k(\mathbf{s}_n) : 0 \leq k < n\} \cap \mathcal{B}(n,c)$. Note that $\mathcal{R}(n,c)$ contains all cyclic shift inequivalent sequences in $\mathcal{B}(n,c)$. From Lemma 3 (ii), it suffices to investigate the nonlinear complexity of the periodic sequence $\mathbf{s}_n^\infty$ derived from $\mathbf{s}_n$ in $\mathcal{R}(n,c)$.

We are now ready to prove Theorem 2, namely, for $\mathbf{s}_n$ in $\mathcal{R}(n,c)$ with $c \geq \lfloor \frac{n}{2} \rfloor$, one has $nlc(\mathbf{s}_n^\infty) = nlc(\mathbf{s}_n) + add(\mathbf{s}_n)$.

Proof of Theorem 2. Suppose $nlc(\mathbf{s}_n^\infty) = \omega$ and $(\underline{\mathbf{s}}_i, \mathcal{F}^d(\underline{\mathbf{s}}_i))$ is a companion pair of $\mathbf{s}_n^\infty$ with $1 \leq d \leq \lfloor \frac{n}{2} \rfloor$. For the convenience of readers, without loss of generality we will simplify i into 0 in the following proof. It allows us to consider the companion pair $(\underline{\mathbf{s}}_0, \mathcal{F}^d(\underline{\mathbf{s}}_0))$ of $\mathbf{s}_n^\infty$. Below we will show that ω can be represented as $\omega = nlc(\mathbf{a}_n) + add(\mathbf{a}_n)$ for a concrete n-length subsequence $\mathbf{a}_n$ of $\mathbf{s}_n^\infty$ derived from $(\underline{\mathbf{s}}_0, \mathcal{F}^d(\underline{\mathbf{s}}_0))$, and then prove that $\omega = nlc(\mathbf{s}_n) + add(\mathbf{s}_n)$ for $\mathbf{s}_n \in \mathcal{R}(n,c)$. We divide the discussion according to the value of $\omega + d$.

Case (1): $\omega + d < n$. In this case, Lemma 7 (ii) implies $\mathbf{s}_n \in \mathcal{B}(n,\omega,d)$. Note that $add(\mathbf{s}_n) = 0$, otherwise it follows from Proposition 1 (i) and $n - \omega - d \geq 1$ that the nonlinear complexity of $R(\mathbf{s}_n)$ is $\omega + 1$, which is larger than $nlc(\mathbf{s}_n^\infty) = \omega$, a contradiction. By Definition 3 and Lemma 6, we see that $\mathbf{a}_n = L^{\omega - c}(\mathbf{s}_n) \in \mathcal{B}(n,c,d)$ and $add(\mathbf{a}_n) = \omega - c$, which implies

$$\omega = c + add(\mathbf{a}_n) = nlc(\mathbf{a}_n) + add(\mathbf{a}_n).$$

From the fact that $\mathbf{s}_n \in \mathcal{R}(n,c)$, we have $add(\mathbf{s}_n) \geq add(\mathbf{a}_n) = \omega - c$. On the other hand, taking $add(\mathbf{s}_n) = t$, from Remark 1, there is a pair of identical $(c + t - 1)$-tuple subsequences with different successors in $\mathbf{s}_n^\infty$. This implies $c + t \leq \omega$. Thus in this case we have $add(\mathbf{s}_n) = add(\mathbf{a}_n) = \omega - c$.

$$\mathbf{s}_{\omega+d} = (s_0, s_1, \ldots, s_{j-1}) \underbrace{(s_j, s_{j+1}, \ldots, s_n, \ldots, s_{\omega+d-1})}_{\text{length}=n}$$

Figure 5: The description of Case (2) in Theorem 2

Case (2): $\omega + d \geq n$. In this case, we know the subsequence $\mathbf{s}_{\omega+d}$ satisfies

$$s_i = s_{i+d} \text{ for } 0 \leq i \leq \omega - 2 \text{ and } s_i = s_{i+n} \text{ for } i \geq 0.$$

Take $j = \omega+d-n$ and let $\mathbf{a}_n = L^j(\mathbf{s}_n)$. Then $\mathbf{a}_n = (a_0, a_1, \ldots, a_{n-1}) = (s_j, s_{j+1}, \ldots, s_{j+n-1})$ $= (s_j, s_{j+1}, \ldots, s_{\omega+d-1})$ in Figure 5. It follows from Lemma 7 (iii) that $\mathbf{a}_n \in \mathcal{B}(n, n-d, d)$. Since $n - d \geq \lceil \frac{n}{2} \rceil$, $\mathbf{a}_n$ has nonlinear complexity $n - d$ by Corollary 1. From the above equalities, the sequence $\mathbf{a}_n$ satisfies that for $0 \leq j_1 < j$,

$$a_{(n-1)-j_1} = s_{(j+n-1)-j_1} = s_{(j-1)-j_1} = s_{(j-1)-j_1+d} = a_{(d-1)-j_1},$$

where $j = \omega + d - n < d$ and $j - j_1 \leq d - 1 \leq \omega - 1$. Since $s_{n-1} = s_{-1} \neq s_{d-1}$, according to Definition 3, we have

$$add(L^j(\mathbf{s}_n)) = add(\mathbf{a}_n) = j = \omega + d - n = \omega - (n - d).$$

This implies $\omega = nlc(\mathbf{a}_n) + add(\mathbf{a}_n)$.

Consider the following set

$$S = \left\{ L^k(\mathbf{s}_n) \,:\, L^k(\mathbf{s}_n) \in \mathcal{B}(n, c'),\ c' \geq \frac{n}{2} \text{ and } 0 \leq k < n \right\}.$$

We see that ω can be represented as the form $nlc(\mathbf{a}_n) + add(\mathbf{a}_n)$ for $\mathbf{a}_n = L^j(\mathbf{s}_n)$, $j = \omega+d-n$, in S. By Remark 1 we know that for any $\mathbf{u}_n \in S$, $\mathbf{u}_n^\infty$ contains a pair of identical subsequences with length $nlc(\mathbf{u}_n) + add(\mathbf{u}_n) - 1$ with different successors. Thus we have $\omega = \max\limits_{\mathbf{u}_n \in S}(nlc(\mathbf{u}_n) + add(\mathbf{u}_n))$. In the following we shall show $\omega = c + t$ for three subcases.

Subcase (2.1): Consider the set

$$S_0 = \{L^k(\mathbf{s}_n) \,:\, 0 \leq k < n\} \cap \mathcal{B}(n, c) = E(\mathbf{s}_n).$$

It follows that $\max\limits_{\mathbf{u}_n \in S_0}(nlc(\mathbf{u}_n) + add(\mathbf{u}_n)) = c + \max\limits_{\mathbf{u}_n \in S_0}(add(\mathbf{u}_n)) = c + add(\mathbf{s}_n) = c + t$ since $\mathbf{s}_n$ is a sequence representative of $E(\mathbf{s}_n)$ with maximal $add(\mathbf{s}_n) = t$ in the set S_0.

Subcase (2.2): Consider $S_1 = \{L^k(\mathbf{s}_n) \,:\, L^k(\mathbf{s}_n) \in \mathcal{B}(n, c_1), 0 \leq k < n\}$ with $c_1 = c + t_1$ and $t_1 \geq 1$. It follows that $c_1 + \max\limits_{\mathbf{u}_n \in S_1}(add(\mathbf{u}_n)) = c + t_1 + \max\limits_{\mathbf{u}_n \in S_1}(add(\mathbf{u}_n))$. As shown in Lemma 6, for each $\mathbf{u}_n \in S_1$, one can get $L^{t_1}(\mathbf{u}_n) \in S_0$. In addition, since

$$\begin{aligned} S_1 &= \{L^k(\mathbf{s}_n) \,:\, L^k(\mathbf{s}_n) \in \mathcal{B}(n, c + t_1),\ 0 \leq k < n\} \\ &= \{L^k(\mathbf{s}_n) \,:\, L^{k+t_1}(\mathbf{s}_n) \in \mathcal{B}(n, c),\ add(L^{k+t_1}(\mathbf{s}_n)) \geq t_1,\ 0 \leq k < n\} \\ &= \{R^{t_1}(L^k(\mathbf{s}_n)) \,:\, L^k(\mathbf{s}_n) \in \mathcal{B}(n, c),\ add(L^k(\mathbf{s}_n)) \geq t_1,\ 0 \leq k < n\}, \end{aligned}$$

it follows that $\max_{\mathbf{u}_n\in S_1}(add(\mathbf{u}_n)) = \max_{\mathbf{u}_n\in S_0}(add(\mathbf{u}_n)) - t_1 = t - t_1$. This yields $\max_{\mathbf{u}_n\in S_1}(nlc(\mathbf{u}_n) + add(\mathbf{u}_n)) = c + t_1 + (t - t_1) = c + t$.

Subcase (2.3): Consider $S_2 = \{L^k(\mathbf{s}_n) : L^k(\mathbf{s}_n) \in \mathcal{B}(n, c_2),\ 0 \le k < n\}$ with $\frac{n}{2} \le c_2 < c$. Take $c_2 = c - t_2$ with $t_2 \ge 1$. We shall investigate $c - t_2 + \max_{\mathbf{u}_n\in S_2}(add(\mathbf{u}_n))$ for S_2. For each $\mathbf{u}_n \in S_0$, it follows from Lemma 6 that $L^{t_2}(\mathbf{u}_n) \in S_2$. In a similar manner, one has

$$\begin{aligned} S_0 &= \{L^k(\mathbf{s}_n) : L^k(\mathbf{s}_n) \in \mathcal{B}(n, c),\ 0 \le k < n\} \\ &= \{L^k(\mathbf{s}_n) : L^{k+t_2}(\mathbf{s}_n) \in \mathcal{B}(n, c_2),\ add(L^{k+t_2}(\mathbf{s}_n)) \ge t_2,\ 0 \le k < n\} \\ &= \{R^{t_2}(L^k(\mathbf{s}_n)) : L^k(\mathbf{s}_n) \in \mathcal{B}(n, c_2),\ add(L^k(\mathbf{s}_n)) \ge t_2,\ 0 \le k < n\}. \end{aligned}$$

Then $\max_{\mathbf{u}_n\in S_2}(add(\mathbf{u}_n)) - t_2 = \max_{\mathbf{u}_n\in S_0}(add(\mathbf{u}_n))$. This yields $\max_{\mathbf{u}_n\in S_2}(nlc(\mathbf{u}_n) + add(\mathbf{u}_n)) = c - t_2 + \max_{\mathbf{u}_n\in S_2}(add(\mathbf{u}_n)) = c - t_2 + (t + t_2) = c + t$.

Combining the above three subcases, we have $\omega = \max_{\mathbf{u}_n\in S}(nlc(\mathbf{u}_n) + add(\mathbf{u}_n)) = c + t$. The desired conclusion thus follows. □

Remark 3. *With the condition in Theorem 2, it follows that each sequence $\boldsymbol{a}_n$ in the set $E(\boldsymbol{s}_n)$ have $nlc(\boldsymbol{a}_n^\infty) = nlc(\boldsymbol{s}_n^\infty) = nlc(\boldsymbol{s}_n) + add(\boldsymbol{s}_n)$ by Lemma 3 (ii). Equivalently, for any sequence a_n with nonlinear complexity $c \ge \frac{n}{2}$, by Lemma 4 it contains a unique companion pair $(\underline{\boldsymbol{a}}_i, \widehat{\underline{\boldsymbol{a}}}_i)$. This implies $L^i(\boldsymbol{a}_n) \in \mathcal{B}(n, c)$. Hence we have $nlc(\boldsymbol{a}_n^\infty) = nlc(\boldsymbol{s}_n^\infty) = nlc(\boldsymbol{s}_n) + add(\boldsymbol{s}_n)$, where $\boldsymbol{s}_n$ is a representative of $E(\boldsymbol{a}_n)$.*

Theorem 2 gives a method to determine the value of the nonlinear complexity of periodic sequences from that of finite-length sequences. Based on Theorem 2, below we give a corollary of the value of $\max_{0\le i<n} nlc(R^i(\mathbf{s}_n))$.

Corollary 2. *For $\boldsymbol{s}_n$ in $\mathcal{R}(n, c, d)$ with $c \ge \lfloor \frac{n}{2} \rfloor$ and $add(\boldsymbol{s}_n) = t$, one has*

$$\max_{0\le i<n} nlc(R^i(\boldsymbol{s}_n)) \begin{cases} = c + t, & \text{if } t \le n - c - d, \\ \ge n - d, & \text{if } t > n - c - d. \end{cases}$$

Proof. It follows from Theorem 2 that for $\mathbf{s}_n \in \mathcal{R}(n, c, d)$ with $add(\mathbf{s}_n) = t$, we have $nlc(\mathbf{s}_n^\infty) = c + t$. Hence $\max_{0\le i<n} nlc(R^i(\mathbf{s}_n)) \le nlc(\mathbf{s}_n^\infty) = c + t$. According to Proposition 1 (i), if $t \le n - c - d$, i.e. $c + d + t \le n$, then $R^t(\mathbf{s}_n) \in \mathcal{B}(n, c + t, d)$ with nonlinear complexity $c + t$, implying $\max_{0\le i<n} nlc(R^i(\mathbf{s}_n)) = c + t = nlc(\mathbf{s}_n^\infty)$; if $t > n - c - d$, i.e.

$c+d+t>n$, then one has $R^{n-c-d}(\mathbf{s}_n) \in \mathcal{B}(n, n-d, d)$ with nonlinear complexity $n-d$, thus $\max\limits_{0\le i<n} nlc(R^i(\mathbf{s}_n)) \ge n-d$. □

Based on Corollary 2, we can see that for $\mathbf{s}_n$ in $\mathcal{R}(n,c,d)$ with $c \ge \lfloor \frac{n}{2} \rfloor$ and $add(\mathbf{s}_n) = t$, if $t \le n-c-d$, then $\max\limits_{0\le i<n} nlc(R^i(\mathbf{s}_n)) = nlc(\mathbf{s}_n^\infty)$. That is to say, Corollary 2 gives a sufficient condition of $\max\limits_{0\le i<n} nlc(R^i(\mathbf{s}_n)) = nlc(\mathbf{s}_n^\infty)$. When $t > n-c-d$, numerical results for $n \le 30$ indicate that $\max\limits_{0\le i<n} nlc(R^i(\mathbf{s}_n)) = n-d$. However, the previous technique is not sufficient to prove or disprove this equality. If the conjecture is true, then we can obtain a sufficient and necessary condition of $\max\limits_{0\le i<n} nlc(R^i(\mathbf{s}_n)) = nlc(\mathbf{s}_n^\infty)$. Here we propose an open problem below on this observation and cordially invite interested readers to attack this problem.

Open Problem 1. *For $\boldsymbol{s}_n$ in $\mathcal{R}(n,c,d)$ with $c \ge \lfloor \frac{n}{2} \rfloor$ and $add(\boldsymbol{s}_n) = t$, is it true that $\max\limits_{0\le i<n} nlc(R^i(\boldsymbol{s}_n)) = n-d$ when $t > n-c-d$?*

5 Periodic sequences with prescribed nonlinear complexity

Based on new theoretical results in Theorem 1 and Theorem 2, we shall give two algorithms to generate all periodic binary sequences in $\mathcal{P}(n,\omega)$ with any prescribed nonlinear complexity. For every shift equivalence class $\{\mathbf{s}_n^\infty, (L(\mathbf{s}_n))^\infty, \ldots, (L^{n-1}(\mathbf{s}_n))^\infty\}$, since each sequence in the class can generate the whole class, it suffices to consider one sequence with certain property when generating the shift equivalence class. Given a set S of sequences with period n, we denote by $\overline{S}$ the union of all cyclic shift equivalence classes of sequences in S. Then two sets S_1 and S_2 are deemed to be cyclic shift equivalent, denoted as $S_1 \cong S_2$, if $\overline{S_1} = \overline{S_2}$.

Theorem 3. *Let n and ω be two positive integers with $\lceil \log_2(n) \rceil \le \omega \le n-1$. Let $\mathcal{B}(n,c)$ be given in (3), $\mathcal{B}_0(n,c) = \{\boldsymbol{s}_n \in \mathcal{B}(n,c) : add(\boldsymbol{s}_n) = 0\}$ and $\mathcal{R}(n,c)$ be defined by (8), respectively. Then we have*

$$\mathcal{P}(n,\omega) \cong \{\boldsymbol{s}_n^\infty : \boldsymbol{s}_n \in S(n,\omega)\}$$

where $S(n,\omega)$ is given as follows,

(i) *if* $\omega \le \frac{n}{2}$, *then* $S(n,\omega) = \{\boldsymbol{s}_n \in \mathcal{B}_0(n,\omega) : L^k(\boldsymbol{s}_n) \notin \mathcal{B}(n,\omega+1), 0 < k < n\}$;

(ii) *if* $\omega \ge \frac{n}{2}$, *then* $S(n,\omega) = \{\boldsymbol{s}_n \in \mathcal{R}(n,c) : add(\boldsymbol{s}_n) = \omega - c\}$, *where* $c = \lceil \frac{n}{2} \rceil$.

Proof. (i) According to Theorem 1, for any integer c with $\lceil \log_2(n) \rceil \le c \le \min\{\omega, \lfloor \frac{n}{2} \rfloor + 1\}$, we have

$$\bigcup_{\omega=c}^{n-1} \mathcal{P}(n,\omega) = \left\{ (L^k(\mathbf{s}_n))^\infty \, : \, \mathbf{s}_n \in \mathcal{B}(n,c),\, 0 \le k < n \right\}.$$

Then for $\lceil \log_2 n \rceil \le \omega \le \lfloor \frac{n}{2} \rfloor$, we can express

$$\begin{aligned}
&\mathcal{P}(n,\omega) \\
&= \bigcup_{\omega_1=\omega}^{n-1} \mathcal{P}(n,\omega_1) \Big\backslash \bigcup_{\omega_2=\omega+1}^{n-1} \mathcal{P}(n,\omega_2) \\
&= \left\{ (L^k(\mathbf{s}_n))^\infty \, : \, \mathbf{s}_n \in \mathcal{B}(n,\omega),\, 0 \le k < n \right\} \Big\backslash \left\{ (L^k(\mathbf{s}_n))^\infty \, : \, \mathbf{s}_n \in \mathcal{B}(n,\omega+1),\, 0 \le k < n \right\} \\
&\cong \left\{ \mathbf{s}_n^\infty \, : \mathbf{s}_n \in \mathcal{B}(n,\omega), L^k(\mathbf{s}_n) \notin \mathcal{B}(n,\omega+1), 0 < k < n \right\}.
\end{aligned}$$

For $\mathbf{s}_n$ in the set $\left\{ \mathbf{s}_n \in \mathcal{B}(n,\omega) : L^k(\mathbf{s}_n) \notin \mathcal{B}(n,\omega+1), 0 < k < n \right\}$, we have $s_{n-1} \ne s_{d-1}$, that is to say $add(\mathbf{s}_n) = 0$, which implies $\mathbf{s}_n \in \mathcal{B}_0(n,\omega)$. Thus for $\lceil \log_2(n) \rceil \le \omega \le \lfloor \frac{n}{2} \rfloor$,

$$\mathcal{P}(n,\omega) \cong \{ \mathbf{s}_n^\infty \, : \mathbf{s}_n \in \mathcal{S}(n,\omega) \},$$

where $\mathcal{S}(n,\omega) = \{\mathbf{s}_n \in \mathcal{B}_0(n,\omega) : L^k(\mathbf{s}_n) \notin \mathcal{B}(n,\omega+1), 0 < k < n\}$. As a result, the desired statement of Theorem 3 (i) follows.

(ii) Again by Theorem 1, let $c = \lceil \frac{n}{2} \rceil$ and $\omega \ge c$, we have

$$\bigcup_{\omega=c}^{n-1} \mathcal{P}(n,\omega) = \left\{ (L^k(\mathbf{s}_n))^\infty \, : \, \mathbf{s}_n \in \mathcal{B}(n,c),\, 0 \le k < n \right\} \cong \{ \mathbf{s}_n^\infty : \mathbf{s}_n \in \mathcal{R}(n,c) \}$$

since $\mathcal{R}(n,c)$ contains all cyclic shift inequivalent sequences in $\mathcal{B}(n,c)$. Then by Theorem 2 we have $\mathcal{P}(n,\omega) \cong \{\mathbf{s}_n^\infty : \mathbf{s}_n \in \mathcal{R}(n,c) : add(\mathbf{s}_n) = \omega - c\} = \{\mathbf{s}_n^\infty \, : \, \mathbf{s}_n \in S(n,\omega)\}$. □

Based on Theorem 3 (i) and (ii), the following two subsections are dedicated to the generation of periodic binary sequences with any given nonlinear complexity.

Algorithm 1 Generation of all periodic binary sequences in $\mathcal{P}(n,\omega)$ with $\omega \leq \frac{n}{2}$

INPUT: Given two global variables n and ω with $\lceil \log_2(n) \rceil \leq \omega \leq \lfloor \frac{n}{2} \rfloor$.
OUTPUT: The set $\mathcal{P}(n,\omega)$.
Main Algorithm
$\mathcal{B}_0(n,\omega) \leftarrow \texttt{genB}(n,\omega)$
$\mathcal{B}(n,\omega+1) \leftarrow \texttt{genB}(n,\omega+1)$
$\mathcal{S}(n,\omega) = \{\mathbf{s}_n \in \mathcal{B}_0(n,\omega) : L^k(\mathbf{s}_n) \notin \mathcal{B}(n,\omega+1), 0 < k < n\}$
$\mathcal{P}(n,\omega) = \left\{(L^k(\mathbf{s}_n))^\infty : \mathbf{s}_n \in \mathcal{S}(n,\omega), 0 \leq k < n\right\}$
function $\texttt{genB}(n,c)$ // Generate $\mathcal{B}_0(n,\omega)$ and $\mathcal{B}(n,\omega+1)$
    for $d = 1$ to $\lfloor \frac{n}{2} \rfloor$ **do**
        $\mathcal{B}(n,c,d) \leftarrow \emptyset$
        while $(a_0, a_1, \ldots, a_{d-1}) \in \mathbb{Z}_2^d$ is aperiodic **do**
            for $i = 0$ to $c+d-2$ **do**
                $s_i \leftarrow a_{i \bmod d}$
            end for
            $s_{c+d-1} \leftarrow a_{(c+d-1) \bmod d} \oplus 1$
            for $(s_{c+d}, \ldots, s_{n-2}) \in \mathbb{Z}_2^{n-c-d-1}$ **do**
                $\mathbf{s}_{n-1} \leftarrow (s_0, \ldots, s_{c+d-1}, s_{c+d}, \ldots, s_{n-2})$
                if $c = \omega$ **then**
                    $\mathbf{s}_n \leftarrow (\mathbf{s}_{n-1}, \overline{s}_{d-1})$
                else
                    $\mathbf{s}_n \leftarrow (\mathbf{s}_{n-1}, s_{n-1})$ with $s_{n-1} \in \mathbb{Z}_2$
                end if
                if $c \geq \lfloor \frac{n}{2} \rfloor$ or n is prime **then** // Conditions that $\mathbf{s}_n$ is aperiodic
                    Add the sequence $\mathbf{s}_n$ to $\mathcal{B}(n,c,d)$
                else if $\mathbf{s}_n$ is aperiodic **then**
                    Add the sequence $\mathbf{s}_n$ to $\mathcal{B}(n,c,d)$
                end if
            end for
        end while
    end for
    return $\bigcup_{1 \leq d \leq \lfloor \frac{n}{2} \rfloor} \mathcal{B}(n,c,d)$
end function

5.1 Periodic sequences with nonlinear complexity $\leq \frac{n}{2}$

In this subsection, all periodic sequences in $\mathcal{P}(n,\omega)$ with $\lceil \log_2(n) \rceil \leq \omega \leq \lfloor \frac{n}{2} \rfloor$ can be generated. By summarizing up the preceding analysis, we propose an algorithm to generate all periodic binary sequences in $\mathcal{P}(n,\omega)$ for any $\lceil \log_2(n) \rceil \leq \omega \leq \lfloor \frac{n}{2} \rfloor$.

In Algorithm 1, we provide detailed steps to generate the set $\mathcal{B}(n,c)$ for any $1 \leq c < n$. The function $\texttt{genB}(n,\omega+1)$ in Algorithm 1 for $\omega+1 \geq \frac{n}{2}$ is similar to the algorithm in [25]. For $\lceil \log_2(n) \rceil \leq \omega \leq \lfloor \frac{n}{2} \rfloor$, Algorithm 1 generates all periodic sequences in $\mathcal{P}(n,\omega)$ based on Theorem 3 (i).

We now consider the complexity of Algorithm 1, where Steps 4, 5, 6 dominate the complexity. In generating $\mathcal{B}_0(n,\omega)$, for each $1 \leq d \leq \lfloor \frac{n}{2} \rfloor$ the loops on $(a_0,\ldots,a_{d-1})$ and $(s_{\omega+d},\ldots,s_{n-2})$ contribute the time and memory complexity $O(2^{n-\omega-1})$, implying Step 4 has complexity $O(n2^{n-\omega-2})$ as d ranges from 1 to $\lfloor \frac{n}{2} \rfloor$. Similarly Step 5 has complexity $O(n2^{n-\omega-2})$. Step 6 generates the set

$$\begin{aligned}\mathcal{S}(n,\omega) =& \{\mathbf{s}_n \in \mathcal{B}_0(n,\omega) : L^k(\mathbf{s}_n) \notin \mathcal{B}(n,\omega+1), 0<k<n\} \\ =& \left\{L^k(\mathbf{s}_n) \,:\, \mathbf{s}_n \in \mathcal{B}_0(n,\omega),\, 0 \leq k < n\right\} \Big\backslash \mathcal{B}(n,\omega+1).\end{aligned}$$

By using the data structure of hash table, this step will have time and memory complexity $O(\min\{n|\mathcal{B}_0(n,\omega)|, |\mathcal{B}(n,\omega+1)|\})$, where $|S|$ is denoted as the size of a set S. Since $|\mathcal{B}_0(n,\omega)| < 2^{n-\omega-1}$, the complexity of Algorithm 1 can be obtained as $O(n2^{n-\omega-2}) + O(n2^{n-\omega-2}) + O(n|\mathcal{B}_0(n,\omega)|) \approx O(n2^{n-\omega-1})$.

On the other hand, when one uses a brute-force approach to generating sequences of given nonlinear complexities, for each n-periodic binary sequence, it is required to calculate its nonlinear complexity with algorithms like the ones proposed in [17, 19], which have complexity $O(n^2\log_2(n))$. On average, we may consider the complexity of exhaustive search for all periodic sequences in $\mathcal{P}(n,\omega)$ is $\frac{2^n O(n^2\log_2(n))}{n} \approx O(n\log_2(n)2^n)$. This indicates that Algorithm 1 has an advantage of factor $2^{\omega+1}\log_2(n)$ over an exhaustive search for periodic binary sequences in $\mathcal{P}(n,\omega)$. It is apparent that such an advantage becomes significant as ω increases.

Below we provide an example to illustrate the process and result of Algorithm 1.

Example 2. *Take $n = 7$ and $\omega = 3$. By the function $\texttt{genB}(n,c)$ in Algorithm 1, we generate the set $\mathcal{B}_0(n,\omega)$ consisting of 18 binary sequences of length 7. Among these sequences, these twelve sequences* (0001101), (0110100),(1010001), (0001011),(0101100),

(1011000), (1110100), (1010011), (0100111), (1110010),(1001011),(0101110), *satisfy the property that all their cyclic shift equivalent sequences (except for themselves) are not contained in* $\mathcal{B}(n, \omega+1)$. *That is to say, the above twelve sequences form the set* $\mathcal{S}(n, \omega)$. *On the other hand, by exhaustive search we obtain the following sequences in* $\mathcal{P}(n, \omega)$*:*

$$
\begin{array}{c}
(0001101)(0011010)(0110100)(1101000)(1010001)(0100011)(1000110) \\
(0001011)(0010110)(0101100)(1011000)(0110001)(1100010)(1000101) \\
(1110100)(1101001)(1010011)(0100111)(1001110)(0011101)(0111010) \\
(1110010)(1100101)(1001011)(0010111)(0101110)(1011100)(0111001)
\end{array}
$$

It is easily seen that $\mathcal{P}(n, \omega)$ *can be obtained by applying circular shift operations on sequences in* $\mathcal{S}(n, \omega)$, *i.e.,* $\mathcal{P}(n, \omega) = \{(L^k(\boldsymbol{s}_n))^\infty : \boldsymbol{s}_n \in \mathcal{S}(n, \omega),\ 0 \le k < n\}$.

5.1.1 Generation of binary de Bruijn sequences

Here we consider a particular case of $n = 2^m$ and $\omega = m$, which corresponds to the generation of binary de Bruijn sequences of order m. In the literature, graphical, algorithmic and algebraic approaches have been proposed to generate binary de Bruijn sequences [32, 33]. The graphical approach, known as the BEST theorem for de Bruijn, Ehrenfest, Smith and Tutte, showed that there are in total $2^{2^{m-1}-m}$ binary de Bruijn sequences of order m and they can be derived in an inductive manner [33]. Algorithmic and algebraic approaches can be used to generate some de Bruijn sequences. However, to the best of our knowledge, only the graphical approach can constructively generate all de Bruijn sequences of order m.

For the case of $n = 2^m$ and $\omega = m$, we can adjust some steps in Algorithm 1 for better performance. For a binary periodic sequence, a run of 0's of length k is a string of consecutive k 0's flanked by 1 and a run of 1's of length k is a string of consecutive k 1's flanked by 0. It is well-known that any binary de Bruijn sequence of order m satisfies run properties as below [30]:

$$
\begin{cases}
\text{(i) } 2^{m-2-i} \text{ runs of zeros, } 2^{m-2-i} \text{ runs of ones of length } i, \text{ for } 1 \le i < m-1; \\
\text{(ii) no run of zeros nor ones of length } m-1; \\
\text{(iii) a single run of } m \text{ zeros and a single run of } m \text{ ones.}
\end{cases} \tag{10}
$$

From the above three properties, it is easily seen that any binary de Bruijn sequence of order m has Hamming weight 2^{m-1} and contains the subsequence $(1, \overbrace{0, \ldots, 0}^{m}, 1)$, which

has the companion pair $(0^m, 0^{m-1}1)$ with spacing $d = 1$. With this observation, we can fix the sequence $\mathbf{s}_{c+d}$ in Algorithm 1 as $\mathbf{s}_{c+d} = \mathbf{s}_{m+1} = (\overbrace{0, \ldots 0}^{m}, 1)$ and $s_{n-1} = 1$, and focus on only the binary sequences $(s_{m+1}, \ldots, s_{n-2})$ of Hamming weight $2^{m-1} - 2$. Thus we can reduce $\mathcal{B}_0(n, \omega)$ to a smaller set

$$\widetilde{\mathcal{B}}_0(n, m) = \{\mathbf{s}_n = (\overbrace{0, \ldots, 0}^{m}, 1, s_{m+1}, \ldots, s_{n-2}, 1) \,:\, \mathbf{s}_n \text{ satisfies (10)}\}, \tag{11}$$

which is generated in Algorithm 2.

Algorithm 2 Generation of binary sequences in $\widetilde{\mathcal{B}}_0(2^m, m)$

```
1: function genB̃₀(n, m)                                  // Generate B̃₀(2^m, m)
2:     s_{m+1} ← (0, 0, ..., 0, 1) and B̃₀(n, m) ← ∅
3:     for (s_{m+1}, ..., s_{n-2}) ∈ Z_2^{n-m-2} with Hamming weight 2^{m-1} − 2 do
4:         if s_n = (0^m 1, s_{m+1}, ..., s_{n-2}, 1) satisfies the run properties (10) then
5:             Add the sequence s_n to B̃₀(n, m)
6:         end if
7:     end for
8: end function
```

Corollary 3. *Let $n = 2^m$ with a positive integer m and let $\widetilde{\mathcal{B}}_0(n, m)$ be as in (11). Then, we obtain de Bruijn sequences of order m as follows,*

$$\mathcal{P}(n, m) = \left\{(L^k(\boldsymbol{s}_n))^\infty : \boldsymbol{s}_n \in \mathcal{S}(n, m), 0 \le k < n\right\}$$

where $\mathcal{S}(n, m) = \{\boldsymbol{s}_n \in \widetilde{\mathcal{B}}_0(n, m) : L^k(\boldsymbol{s}_n) \notin \mathcal{B}(n, m+1), 0 < k < n\}$ and $|\widetilde{\mathcal{B}}_0(n, m)| = \frac{1}{2^{2m-2}}\binom{2^{m-2}}{2^{m-3}, 2^{m-4}, \ldots, 2^1, 2^0}^2$ with $\binom{M}{m_1, m_2, \ldots, m_k} = \frac{M!}{m_1! m_2! \ldots m_k!}$.

Proof. According to the above analysis, we can reduce $\mathcal{B}_0(n, \omega)$ to $\widetilde{\mathcal{B}}_0(n, m)$ in (11). Thus $\mathcal{P}(n, m)$ can be obtained immediately from Theorem 3 (i). In the following we determine the size of $\widetilde{\mathcal{B}}_0(n, m)$. A binary sequence of period $2^m - 1$ having the same run distribution as m-sequences of order m is called as a run sequence. Note that sequences in $\widetilde{\mathcal{B}}_0(n, m)$ are in one-to-one correspondence with all binary run sequences of period $2^m - 1$. For each sequence in $\widetilde{\mathcal{B}}_0(n, m)$, by deleting one 0 in the longest m run of 0's, we obtain a run sequence of period $2^m - 1$. Hence, the number of $\widetilde{\mathcal{B}}_0(n, m)$ is the same as the number of shift inequivalent run sequences, from [31] which is equal to $\frac{1}{2^{2m-2}}\binom{2^{m-2}}{2^{m-3}, 2^{m-4}, \ldots, 2^1, 2^0}^2$. □

Example 3. *When $m = 4$, Algorithm 2 generates $\widetilde{\mathcal{B}}_0(16, 4)$ with size 36, then we filter these 36 shift inequivalent sequences by $\mathcal{B}(16, 5)$, thus obtaining all 16 shift inequivalent sequences in $\mathcal{P}(16, 4)$. And when $m = 5$, Algorithm 2 generates $\widetilde{\mathcal{B}}_0(32, 5)$ with size 88200, then we filter these shift inequivalent sequences by $\mathcal{B}(32, 6)$, which yields all 2048 shift inequivalent sequences in $\mathcal{P}(32, 5)$.*

5.2 Periodic sequences with nonlinear complexity $\geq \frac{n}{2}$

Recall from Theorem 3 (ii) that, for $\omega \geq \frac{n}{2}$,

$$\mathcal{P}(n, \omega) \cong \{\mathbf{s}_n^\infty : \mathbf{s}_n \in \mathcal{S}(n, \omega)\},$$

where $\mathcal{S}(n, \omega)$ for $c = \lceil \frac{n}{2} \rceil$ is given by $S(n, \omega) = \{\mathbf{s}_n \in \mathcal{R}(n, c) : add(\mathbf{s}_n) = \omega - c\}$. This result also holds for $c = \lfloor \frac{n}{2} \rfloor$ and can be proved similarly. Moreover, the characterizations on sequences $\mathcal{R}(n, \lceil \frac{n}{2} \rceil)$ in Subsection 5.2.1 can be similarly made for $c = \lfloor \frac{n}{2} \rfloor$ but with several tedious cases to be discussed. We therefore only present relevant results for the case of $c = \lceil \frac{n}{2} \rceil$.

To illustrate the result of Theorem 3 (ii), we first present a toy example below.

Example 4. *Take an example for the case $n = 8$ and $c = 4$. All finite-length sequences in $\mathcal{B}(8, 4)$ can be obtained from the function $\texttt{genB}(n, c)$ in Algorithm 1, or from the algorithm in [25] by letting $k_1 = 0$. We first group $\mathcal{B}(8, 4)$ into shift equivalence classes. For each shift equivalence class in $\mathcal{B}(8, 4)$, we determine the sequence representatives according to Definition 4, thereby obtain the set $\mathcal{R}(8, 4)$. In this way we get $\bigcup_{\omega=4}^{7} \mathcal{P}(8, \omega) \cong \{\boldsymbol{s}_n^\infty : \boldsymbol{s}_n \in \mathcal{R}(8, 4)\}$. By computing the number of added terms of each sequence in $\mathcal{R}(8, 4)$, we can determine the nonlinear complexity of the corresponding periodic sequences from Theorem 2. Table 2 lists all binary sequences with period 8 and nonlinear complexity ω, $\omega \geq 4$, up to shift equivalence obtained in this way. The result is consistent with the exhaustive search presented in [17, Table 3.2].*

We carry out experiments for n up to 40, and they all confirm that Theorem 3 (ii) is consistent with exhaustive search.

The previous section discussed the generation of binary de Bruijn sequences, which is a particular case for $\omega \leq \frac{n}{2}$. Below we discuss another particular case where the prescribed nonlinear complexity ω achieves the maximum value $n - 1$.

Table 2: Binary sequences of period 8 with nonlinear complexity ω.

$\mathcal{B}(8,4)$	$\mathcal{R}(8,4)$	$add(\mathbf{s}_n)$	$\cong \mathcal{P}(8, 4+add(\mathbf{s}_n))$
(00100011), (10010001), (00110010)	(00100011), (10010001), (00110010)	0	$(00100011)^\infty$
(10011000), (00100110), (10001001)	(10011000), (00100110), (10001001)		$(10011000)^\infty$
(01100111), (11011001), (01110110)	(01100111), (11011001), (01110110)		$(01100111)^\infty$
(11011100), (01101110), (11001101)	(11011100), (11011001), (01110110)		$(11011100)^\infty$
(11110000), (00001111)	(11110000), (00001111)		$(11110000)^\infty$
(10110100), (01001011)	(10110100), (01001011)		$(10110100)^\infty$
(00001101)	(00001101)		$(00001101)^\infty$
(00001011)	(00001011)		$(00001011)^\infty$
(11110100)	(11110100)		$(11110100)^\infty$
(11110100)	(11110100)		$(11110100)^\infty$
(01010000), (00001010)	(00001010)	1	$(00001010)^\infty$
(10101111), (11110101)	(11110101)		$(11110101)^\infty$
(00001110)	(00001110)		$(00001110)^\infty$
(11110001)	(11110001)		$(11110001)^\infty$
(10101100)	(10101100)		$(10101100)^\infty$
(01010011)	(01010011)		$(01010011)^\infty$
(11011000)	(11011000)		$(11011000)^\infty$
(00100111)	(00100111)		$(00100111)^\infty$
(00001001), (10010000)	(10010000)	2	$(10010000)^\infty$
(01000101), (01010001)	(01010001)		$(01010001)^\infty$
(10111010), (10101110)	(10101110)		$(10101110)^\infty$
(11110110), (01101111)	(01101111)		$(01101111)^\infty$
(00001100)	(00001100)		$(00001100)^\infty$
(11110011)	(11110011)		$(11110011)^\infty$
(00010000), (00001000)	(00001000)	3	$(00001000)^\infty$
(01010010), (01001010)	(01001010)		$(01001010)^\infty$
(10101101), (10110101)	(10110101)		$(10110101)^\infty$
(11101111), (11110111)	(11110111)		$(11110111)^\infty$

Remark 4. *It was shown in [27] that sequences in $\mathcal{P}(n, n-1)$ can be generated from a recursive approach by applying the Euclidean algorithm on n and certain positive integers. Similarly, the set $\mathcal{P}(n, n-2)$ was later completely characterized in [28]. In this paper, Theorem 3 (ii) can produce all sequences in $\mathcal{P}(n, \omega)$ with $\frac{n}{2} \leq \omega \leq n-1$. Below we discuss the connection between Theorem 3 (ii) and the work in [27]. Its connection with the work in [28] is similar and thus not included here.*

For the case $\omega = n-1$ and a sequence $\boldsymbol{s}_n \in \mathcal{R}(n, \lceil \frac{n}{2} \rceil)$ with $spac(\boldsymbol{s}_n) = d$, namely,

$$\boldsymbol{s}_n = (\overbrace{\boldsymbol{s}_d \ldots \boldsymbol{s}_d}^{q}(s_0, \ldots, s_{r-1}, \overline{s}_r)\, \boldsymbol{s}_{[c+d:n]}) = (\boldsymbol{s}_d^q(s_0, \ldots, s_{r-1}, \overline{s}_r)\, \boldsymbol{s}_{[c+d:n]}),$$

suppose $add(\boldsymbol{s}_n) = t = \omega - c = \lfloor \frac{n}{2} \rfloor - 1$, then $s_{n-1-i} = s_{(d-1-i) \bmod d}$ for $0 \leq i < \lfloor \frac{n}{2} \rfloor - 1$.

and $s_{n-\lfloor \frac{n}{2} \rfloor} \neq s_{d-\lfloor \frac{n}{2} \rfloor}$. *When* $\omega + d \leq n$, *i.e.,* $d = 1$, *the sequence* $\boldsymbol{s}_n = (s_0^c \, \overline{s}_0 \, s_0^t)$. *Then*

$$\boldsymbol{a}_n = R^t(\boldsymbol{s}_n) = (s_0^{n-1} \, \overline{s}_0),$$

which corresponds to the sequence in [27, Theorem 1 (i)]. When $\omega + d > n$, *suppose* $\boldsymbol{s}_{[i:i+\omega+d]}$ *is the subsequence formed by the companion pair* $(\underline{\boldsymbol{s}}_i, \underline{\boldsymbol{s}}_{i+d})$. *Then* $\boldsymbol{a}'_n = L^i(\boldsymbol{s}_n)$ *has the companion pair* $(\underline{\boldsymbol{s}}_0, \underline{\boldsymbol{s}}_d)$. *As discussed in the proof of Theorem 2, we know the sequence* $\boldsymbol{a}_n = L^j(\boldsymbol{a}'_n)$ *with* $j = (\omega + d) - n$ *belongs to* $\mathcal{B}(n, n-d, d)$ *and has* $add(\boldsymbol{a}_n) = \omega - (n-d) = d - 1$. *This implies that* $\boldsymbol{b}_n = R^{d-1}(\boldsymbol{a}_n)$ *satisfies* $b_i = b_{i+d}$ *for* $i = 0, 1, \ldots, n-d-1$ *and has the form*

$$\boldsymbol{b}_n = (\overbrace{\boldsymbol{b}_d \ldots \boldsymbol{b}_d}^{q}(b_0, \ldots, b_{r-1})) = (\boldsymbol{b}_d^q(b_0, \ldots, b_{r-1}))$$

where $r = n \bmod d$ *and* $q = \frac{n-r}{d}$. *This corresponds to the sequences characterized in [27, Theorem 1 (ii)]. For instance, given a sequence* $\boldsymbol{s}_n = (0101010010101010) \in \mathcal{B}(16, 8, 7)$ *with* $add(\boldsymbol{s}_n) = 7$, *where the added terms are in bold, we have* $\boldsymbol{a}_n = L^{15}(\boldsymbol{s}_n) = (0010101001\mathbf{010101}) \in \mathcal{B}(16, 9, 7)$ *with* $add(\boldsymbol{a}_n) = 6$. *Then the sequence* $\boldsymbol{b}_n = R^{d-1}(\boldsymbol{a}_n) = R^6(\boldsymbol{a}_n) = (0101010010101001)$ *has the form* $(0101010)^2(01) = ((01)^3 0)^2(01)$, *which is consistent with the instance given in [27, Example 1].*

According to Theorem 3 (ii), each periodic sequence with a prescribed nonlinear complexity $\omega \geq \frac{n}{2}$ can be derived from its n-length subsequence $\mathbf{s}_n$ in $\mathcal{R}(n, c)$ with a desired $add(\mathbf{s}_n) = \omega - c$ with $c = \lceil \frac{n}{2} \rceil$. It is thus of significant interest to further explore the structure of sequences in $\mathcal{R}(n, c)$. The following subsection further discusses the sequence representatives in $\mathcal{R}(n, c)$. After the discussion, we will propose Algorithm 3 to generate all binary sequences in $\mathcal{P}(n, \omega)$ with $\omega \geq \frac{n}{2}$.

5.2.1 Characterizations on sequence representatives

For $\mathbf{s}_n \in \mathcal{B}(n, c)$ with $c = \lceil \frac{n}{2} \rceil$ given in (2), this section further investigates the set $E(\mathbf{s}_n) = \{L^k(\mathbf{s}_n) : 0 \leq k < n\} \cap \mathcal{B}(n, c)$, which helps us generate sequence representatives in $\mathcal{B}(n, c)$ more efficiently. In what follows, we shall first consider the necessary conditions such that both $\mathbf{s}_n$ and $R^h(\mathbf{s}_n)$ are contained in $\mathcal{B}(n, c)$.

Lemma 8. *For* $\boldsymbol{s}_n \in \mathcal{B}(n, c, d)$ *with* $c \geq \frac{n}{2}$, *if* $\boldsymbol{s}_n$ *has a shift equivalent sequence* $\boldsymbol{v}_n = R^h(\boldsymbol{s}_n) \in \mathcal{B}(n, c, d')$, *then* h *satisfies at least one of inequalities:* $n - (c+d) + d' \leq h < c + d'$ *and* $n - (c + d) < h \leq c + d' - d$.

Proof. For a sequence $\mathbf{s}_n \in \mathcal{B}(n,c,d)$, we shall first prove that if $R^h(\mathbf{s}_n)$ belongs to $\mathcal{B}(n,c,d')$, then the value h satisfies $n-(c+d) < h < c+d'$. Suppose that $h \le n-c-d$. According to (2), $\mathbf{s}_n \in \mathcal{B}(n,c,d)$ has the form

$$\mathbf{s}_n = (\mathbf{s}_d^q(s_0,\ldots,s_{r-1},\overline{s}_r)\,\mathbf{s}_{[c+d:n]}),$$

where $(q-1)d+r+1 = c$. It is clear that the c-tuples $(\mathbf{s}_d^{q-1}(s_0,\ldots,s_{r-1},s_r))$ and $(\mathbf{s}_d^{q-1}(s_0,\ldots,s_{r-1},\overline{s}_r))$ form a companion pair in $\mathbf{s}_n$. Due to $h \le n-(c+d)$, this companion pair is also contained in the right cyclic shift sequence $R^h(\mathbf{s}_n)$. Since $nlc(R^h(\mathbf{s}_n)) = c \ge \frac{n}{2}$, Lemma 4 shows that this companion pair is the unique one of $R^h(\mathbf{s}_n)$. It contradicts the assumption that $R^h(\mathbf{s}_n)$ belongs to $\mathcal{B}(n,c)$. Thus we have $h > n-(c+d)$. For the sequence $\mathbf{v}_n = R^h(\mathbf{s}_n)$ and $h' = n-h$, similarly we have $h' = n-h > n-(c+d')$, implying $h < c+d'$. Thus the statement $n-(c+d) < h < c+d'$ holds.

Furthermore, by setting $b = h-(n-(c+d)))$ and $b' = h'-(n-(c+d'))$, we have $b+b' = (h+h')-(n-(c+d)+n-(c+d')) = (2c-n)+d+d' \ge d+d'$ since $c \ge \frac{n}{2}$. Hence at least one of $b \ge d'$ and $b' \ge d$ holds. If $b \ge d'$, then $n-(c+d)+d' \le h < c+d'$; if $b' \ge d$, then $n-(c+d')+d \le h' = n-h < c+d$ and thus $n-(c+d) < h \le (c+d')-d$. $\square$

In the proof of Lemma 8, we see the symmetric relation between $\mathbf{s}_n$ and $\mathbf{v}_n = R^h(\mathbf{s}_n)$ in $\mathcal{B}(n,c)$. If a pair of shift equivalent sequences $(\mathbf{s}_n, R^h(\mathbf{s}_n))$ in $\mathcal{B}(n,c)$ satisfies $h \le (c+d')-d$, then $(\mathbf{v}_n = R^h(\mathbf{s}_n), R^{n-h}(\mathbf{v}_n))$ satisfies $h' = n-h \ge (n-(c+d'))+d$ where $spac(\mathbf{s}_n) = d$ and $spac(\mathbf{v}_n) = d'$. Therefore, it suffices to consider the shift equivalent sequence $R^h(\mathbf{s}_n)$ of each $\mathbf{s}_n$ in $\mathcal{B}(n,c)$ with $(n-(c+d))+d' \le h \le c+d'-1$.

In the following, for the pair of shift equivalent sequences $(\mathbf{s}_n, \mathbf{v}_n)$ in $\mathcal{B}(n,c)$, Proposition 2 characterizes certain structure of the sequence $\mathbf{s}_n$ in (i) and determines $\mathbf{v}_n$ and $add(\mathbf{v}_n)$ in (ii), which will help us find shift equivalent sequences and then determine the sequence representatives by deleting sequences that are not sequence representatives in Algorithm 3.

Proposition 2. *Suppose a sequence $\boldsymbol{s}_n \in \mathcal{B}(n,c,d)$ with $c = \lceil \frac{n}{2} \rceil$ has a shift equivalent sequence $\boldsymbol{v}_n = R^h(\boldsymbol{s}_n) \in \mathcal{B}(n,c,d')$, where $(n-(c+d))+d' \le h \le c+d'-1$. Then*

(i) *the subsequence $\boldsymbol{s}_{[g:n]}$ of $\boldsymbol{s}_n$ satisfies*

$$s_i = s_{i+d'} \quad for \quad g \le i < n-d', \tag{12}$$

where $g = (c+d)-d'$ in the case of $d < n-c$, and $g = 2d-d'-1$ in the case of $d = n-c$;

(ii) *let* $r_1 \le g$ *and* $r_2 \ge n - d'$. *Suppose* $s_i = s_{i+d'}$ *holds for any integer* $r_1 \le i < r_2$ *and* $s_i \ne s_{i+d'}$ *for* $i = r_1 - 1, r_2$, *where the subscripts are taken modulo* n. *Take* $c' = r_2 - r_1 + 1$. *If* $c' \ge c$, *then* $h = c + (n - r_2 - 1)$ *and* $add(\boldsymbol{v}_n) = c' - c$.

Proof. (i) Since $n - (c + d) + d' \le h < c + d'$, $\mathbf{s}_{[(c+d)-d':n]} = (s_{(c+d)-d'}, \ldots, s_{n-1})$ is a subsequence of $\mathbf{s}_{[n-h:n]}$, and $\mathbf{s}_{[n-h:n]}$ is contained in the subsequence $\mathbf{v}_{c+d'-1}$ of $\mathbf{v}_n = R^h(\mathbf{s}_n) = \mathbf{s}_{[n-h:n]}\mathbf{s}_{n-h}$ in $\mathcal{B}(n, c, d')$. According to the definition of $\mathcal{B}(n, c, d')$ in (2), the subsequence $\mathbf{v}_{c+d'-1}$ satisfies $v_i = v_{i+d'}$ for $i = 0, 1, \ldots, c-2$. Thus it follows from $\mathbf{s}_{[(c+d)-d':n]} \subseteq \mathbf{s}_{[n-h:n]} \subseteq \mathbf{v}_{c+d'-1}$ that $s_i = s_{i+d'}$ for $i = (c + d) - d', \ldots, n - d' - 1$. Here it requires $d < n - c$ from $(c + d) - d' \le n - d' - 1$. That is, in the case of $d < n - c$, the subsequence $\mathbf{s}_{[(c+d)-d':n]}$ of $\mathbf{s}_n$ satisfies

$$s_i = s_{i+d'} \quad \text{for} \quad (c + d) - d' \le i < n - d'. \tag{13}$$

In the case of $d = n - c$: consider $\mathbf{v}_n \in \mathcal{B}(n, c, d')$ and $R^{h'}(\mathbf{v}_n) = \mathbf{s}_n$ with $h' = n - h$. We divide the discussion into two subcases according to the value of h'. For the subcase of $h' \ge [n - (c + d')] + d$, then $h \le (c + d') - d$. When $d' < n - c$, it follows from (13) that $\mathbf{v}_n \in \mathcal{B}(n, c, d')$ satisfies $v_i = v_{i+d}$ with $(c + d') - d \le i < n - d$. Thus the pair of shift equivalent sequences $\mathbf{v}_n$ and $R^{h'}(\mathbf{v}_n) = \mathbf{s}_n$ has been considered in (13). That is, when $d' < n - c$ there is no need to consider $\mathbf{s}_n$ and $R^h(\mathbf{s}_n) = \mathbf{v}_n$. When $d' = n - c$, $(n - (c + d)) + d' \le h \le (c + d') - d$ implies that $n - c \le h \le c$. If $c = \frac{n}{2}$ with even n, we have $\mathbf{s}_n = (\mathbf{s}_d\,(s_0, \ldots, s_{d-2}, \overline{s}_{d-1}))$ and $h = \frac{n}{2}$. If $(s_0, \ldots, s_{d-2}, \overline{s}_{d-1})$ is aperiodic, then $\mathbf{s}_n$ and $R^{\frac{n}{2}}(\mathbf{s}_n)$ are shift equivalent in $\mathcal{B}(n, \frac{n}{2})$ with the same number of added terms. In this time, $\mathbf{s}_n$ can be kept as a candidate for the sequence representatives. If $c = k+1$ with odd $n = 2k + 1$, we have $\mathbf{s}_n = (\mathbf{s}_k^2\,\overline{s}_0)$ and $k \le h \le k + 1$, while $R^k(\mathbf{s}_n)$ and $R^{k+1}(\mathbf{s}_n)$ do not belong to $\mathcal{B}(2k+1, k)$. That is to say, when $d = n-c$ the subcase of $h' \ge [n-(c+d')]+d$ does not need to be considered.

For the subcase of $h' < [n - (c + d')] + d$, then $h > (c + d') - d$, implying that $(c + d') + 1 - d \le h \le c + d' - 1$. Note that $n - ((c + d') + 1 - d) = 2d - d' - 1$. Thus $\mathbf{s}_{[2d-d'-1:n]} = (s_{2d-d'-1}, \ldots, s_{n-1})$ is a subsequence of $\mathbf{s}_{[n-h:n]}$, and $\mathbf{s}_{[n-h:n]}$ is contained in the subsequence $\mathbf{v}_{c+d'-1}$ of $\mathbf{v}_n = R^h(\mathbf{s}_n)$ in $\mathcal{B}(n, c, d')$. According to the definition of $\mathcal{B}(n, c, d')$ in (2), we have the subsequence $\mathbf{s}_{[2d-d'-1:n]}$ of $\mathbf{s}_n$ satisfies

$$s_i = s_{i+d'} \quad \text{for} \quad 2d - d' - 1 \le i < n - d'.$$

The desired conclusion in (12) thus follows.

(ii) With the assumption that $s_i = s_{i+d'}$ holds for $r_1 \le i < r_2$, the sequence $\mathbf{u}_n^\infty = (u_0, u_1, \ldots, u_{n-1})^\infty = L^{r_1}(\mathbf{s}_n^\infty)$ satisfies

$$u_i = u_{i+d'} \text{ for } 0 \le i < r_2 - r_1 \text{ and } u_i \ne u_{i+d'} \text{ for } i = -1, r_2 - r_1. \tag{14}$$

This implies that $\mathbf{u}_{c'+d'}$ belongs to $\mathcal{B}(c'+d', c', d')$ by $c' = r_2 - r_1 + 1$. When $c' + d' \le n$, from (14) we have $\mathbf{u}_n \in \mathcal{B}(n, c', d')$ with $add(\mathbf{u}_n) = 0$. Thus, if $c' \ge c$, then by Definition 3 and Lemma 6 we have the sequence $\mathbf{v}_n = L^{c'-c}(\mathbf{u}_n) \in \mathcal{B}(n, c, d')$ with $add(\mathbf{v}_n) = c' - c$. When $c' + d' > n$, let $\mathbf{u}'_n = \mathbf{u}_{[c'+d'-n, c'+d']} = L^{c'+d'-n}(\mathbf{u}_n)$. It follows from Lemma 7 (iii) and (14) that $\mathbf{u}'_n \in \mathcal{B}(n, n-d', d')$ with $add(\mathbf{u}'_n) = c' + d' - n$. Due to $c + d' \le n$, again by Definition 3 and Lemma 6 it implies that $L^{n-d'-c}(\mathbf{u}'_n) = L^{c'-c}(\mathbf{u}_n) = \mathbf{v}_n$ belongs to $\mathcal{B}(n, c, d')$ with $add(\mathbf{v}_n) = (c'+d'-n)+(n-d'-c) = c'-c$. Together the two cases, we both have $\mathbf{v}_n = L^{c'-c}(\mathbf{u}_n) = L^{r_1+(c'-c)}(\mathbf{s}_n) = R^h(\mathbf{s}_n)$ with $h = n - (r_1 + c' - c) = c + (n - r_2 - 1)$ and $add(\mathbf{v}_n) = c' - c$. □

In order to generate the sequence representatives in the set $\mathcal{R}(n, \lceil \frac{n}{2} \rceil)$, from Proposition 2 (i), we only need to investigate sequences $\mathbf{s}_n$ in $\mathcal{B}(n, \lceil \frac{n}{2} \rceil)$ satisfying (12) for any $1 \le d' \le \lfloor \frac{n}{2} \rfloor$, since only these sequences may have shift equivalent sequences in $\mathcal{B}(n, \lceil \frac{n}{2} \rceil)$. Thus we can generate the set $\mathcal{R}(n, c)$ with $c = \lceil \frac{n}{2} \rceil$ as follows. For each $1 \le d \le n - c$ and each $\mathbf{s}_{c+d}$, consider $\mathbf{s}_n$ satisfying (12) for every aperiodic subsequence $(s_{c+d-d'}, \ldots, s_{c+d-1})$ with $1 \le d' \le n - c$. By Proposition 2 (ii), suppose that $\mathbf{s}_n$ satisfies $c' \ge c$, then we obtain directly $\mathbf{v}_n = R^h(\mathbf{s}_n) \in \mathcal{B}(n, c)$ and $add(\mathbf{v}_n) = c' - c$. When $\mathbf{s}_n$ and $R^h(\mathbf{s}_n) = \mathbf{v}_n$ are considered as a pair of shift equivalent sequences in $E(\mathbf{s}_n)$, the one with less added terms will be deleted. As we let all subsequences $\mathbf{s}_{c+d}$ go through the above process, every pair of shift equivalent sequences in $E(\mathbf{s}_n)$ are considered. Thus all sequences $\mathbf{a}_n$ with maximal $add(\mathbf{a}_n)$ are kept. This allows us to obtain sequence representatives more efficiently, which in turn constitute the set $\mathcal{R}(n, c)$. Below we give an example to illustrate the above analysis to delete sequences which are not representatives by Proposition 2.

Example 5. *Take an example for $n = 12$, $c = 6$ and $d = 2$. Consider $\boldsymbol{s}_d = \boldsymbol{s}_2 = (01)$ and $\boldsymbol{s}_{c+d} = (\mathbf{01}010100)$. Run through d' with $1 \le d' \le 6$. For $d' = 1$, by (12) only the subsequence $\boldsymbol{s}_{[8:12]} = (0000)$ needs to be considered, thus $\boldsymbol{s}_{12} = (\underline{0}10101\underline{\mathbf{0}}00000)$. By Proposition 2 (ii), we can see that $r_1 = 6$ (the corresponding terms given in bold and underlined), and $r_2 = 12$ (the corresponding terms are underlined), implying $c' = r_2 - r_1 + 1 = 7$. Since $c' > c$, we get $h = n - (r_2 + 1 - c) = 5$. So $\boldsymbol{v}_{12} = R^h(\boldsymbol{s}_{12}) = R^5(\boldsymbol{s}_{12}) \in \mathcal{B}(12, 6, 1)$ with $add(\boldsymbol{v}_{12}) = c' - c = 1$. The pair of shift equivalent sequences $\boldsymbol{s}_{12}$ and $\boldsymbol{v}_{12}$*

Algorithm 3 Generation of all periodic binary sequences in $\mathcal{P}(n,\omega), \omega \geq \frac{n}{2}$

Main Algorithm
INPUT: A positive integer n
OUTPUT: The set $\mathcal{P}(n,\omega)$ with $\lceil \frac{n}{2} \rceil \leq \omega \leq n-1$.
$\mathcal{R}(n, \lceil \frac{n}{2} \rceil) \leftarrow \texttt{genR}(n, \lceil \frac{n}{2} \rceil)$
$\mathcal{P}(n,\omega) = \left\{ (L^k(\mathbf{s}_n))^\infty : \mathbf{s}_n \in \mathcal{R}(n, \lceil \frac{n}{2} \rceil),\ add(\mathbf{s}_n) = \omega - \lceil \frac{n}{2} \rceil, 0 \leq k < n \right\}$
function $\texttt{genR}(n, \lceil \frac{n}{2} \rceil)$ // Generate $\mathcal{R}(n, \lceil \frac{n}{2} \rceil)$
    $\mathcal{R}(n,c) \leftarrow \emptyset$, $U \leftarrow \emptyset$ and $c \leftarrow \lceil \frac{n}{2} \rceil$
    for $d = 1$ to $n - c$ **do**
        while $(v_0, v_1, \ldots, v_{d-1}) \in \mathbb{Z}_2^d$ is aperiodic **do**
            $s_i \leftarrow v_{i \bmod d},\ 0 \leq i \leq c+d-2$, $s_{c+d-1} \leftarrow v_{(c+d-1) \bmod d} \oplus 1$ and $V \leftarrow \emptyset$
            if $d < n - c$ **then**
                for $d' = 1$ to $n - c$ **do**
                    if $(s_{c+d-d'}, \ldots, s_{c+d-1})$ is aperiodic **then**
                        $s_i \leftarrow s_{i-d'},\ c+d \leq i \leq n-1$ and $r_1 \leftarrow c+d-d'$
                        if $\mathbf{s}_n = (s_0, s_1, \ldots, s_{n-1}) \notin U$ **then**
                            while $s_{r_1-1} = s_{r_1+d'-1}$ **do**
                                $r_1 = r_1 - 1$
                            end while
                            Similarly obtain r_2 such that $s_i = s_{i+d'}$ holds for $n - d' - 1 \leq i < r_2$
                            if $\Delta = (r_2 - r_1 + 1) - c \geq 0$ **then** //Exclude sequences from $\mathcal{R}(n,c)$
                                If $add(\mathbf{s}_n) < \Delta$, then add $(s_{c+d}, \ldots, s_{n-1})$ to V
                                Otherwise set $h = c + (n - r_2 - 1)$ and add $R^h(\mathbf{s}_n)$ to U
                            end if
                        end if
                    end if
                end for
                Add $\mathbf{u}_n \leftarrow (\mathbf{s}_{c+d}, u_{c+d}, \ldots, u_{n-1})$ with $(u_{c+d}, \ldots, u_{n-1}) \in \mathbb{Z}_2^{n-c-d} \backslash V$ to $\mathcal{R}(n,c)$
            else if $\mathbf{s}_n = \mathbf{s}_{c+d} = (s_0, s_1, \ldots, s_{c+d-1}) \notin U$ **then**
                Add $\mathbf{s}_n$ to $\mathcal{R}(n,c)$.
                for $d' = 1$ to $n - c$ **do**
                    if $(s_{n-d'}, \ldots, s_{n-1})$ is aperiodic, $s_i = s_{i+d'}$ with $i \in [2d-1-d', n-d')$ **then**
                        Repeat Steps 16−23 while revising Step 21 as
                        If $add(\mathbf{s}_n) < \Delta$ then add $\mathbf{s}_n$ to U.
                    end if
                end for
            end if
        end while
    end for
    return $\mathcal{R}(n,c) \leftarrow \mathcal{R}(n,c) \backslash U$
end function

Table 3: The shift equivalent sequences for $n = 12$, $c = 6$ and $d = 2$ with $\mathbf{s}_d = (01)$.

$\mathbf{s}_{c+d} = (01010100)$ and $\mathbf{v}_n = R^h(\mathbf{s}_n) \in \mathcal{B}(n, c, d')$							
d'	$\mathbf{s}_n$	r_1	r_2	$add(\mathbf{v}_n) = c' - c$	h	$\mathbf{v}_n$	Delete
1	($\underline{0}10101\underline{\mathbf{0}}00000$)	6	12	1	5	(000000101010)	$\mathbf{s}_n$
2	–						
3	(0101$\underline{\mathbf{0}}$1001001)	4	10	1	7	(100100101010)	$\mathbf{v}_n$
4	(0101$\underline{\mathbf{0}}$100010$\underline{0}$)	4	11	2	6	(000100010101)	$\mathbf{s}_n$
5	($\underline{0}1\underline{\mathbf{0}}$101001010)	2	12	5	5	(010100101010)	$\mathbf{s}_n$
6	(01$\underline{\mathbf{0}}$10100$\underline{0}$0101)	2	7	0	10	(010100010101)	$\mathbf{v}_n$

is found, and $\boldsymbol{s}_{12}$ *should be deleted since* $add(\boldsymbol{s}_{12}) = 0 < add(\boldsymbol{v}_{12}) = 1$. *Since the last two terms of* $\boldsymbol{s}_{c+d} = (01010100)$ *is* (00), *it is impossible for* $d' = 2$ *by the definition of aperiodic sequences. The sequences which are not representatives can be deleted similarly for* $d' = 3, 4, 5, 6$, *which are given in Table 3.*

Remark 5. *In Algorithm 3, from Proposition 2 we give the detailed steps to generate the set* $\mathcal{R}(n, c)$ *for any* $c = \lceil \frac{n}{2} \rceil$. *For* $\lceil \frac{n}{2} \rceil \leq \omega \leq n - 1$, *Algorithm 3 generates all periodic sequences in* $\mathcal{P}(n, \omega)$ *based on Theorem 3 (ii). From the steps in generating* $\mathcal{R}(n, c)$, *we see that the loops on* $(v_0, \ldots, v_{d-1})$ *and* $(u_{c+d}, \ldots, u_{n-1})$ *contribute the dominating time and memory complexity, roughly,* $O(2^{n/2})$. *This indicates that Algorithm 3 has an advantage of factor* $2^{n/2} n \log_2(n)$ *compared to the exhaustive search for sequences* $\boldsymbol{s}_n^\infty$ *with nonlinear complexity* ω.

6 Conclusion

Our contributions in this paper are twofold: the first contribution is to investigate the varying behavior of the nonlinear complexity of finite-length sequences under circular shift operators, and the second contribution is the establishment of a one-to-one correspondence between the set of periodic sequences with certain nonlinear complexities and the set of certain finite-length sequences with a particular structure. As an application of the correspondence, we present two efficient algorithms to generate all periodic sequences with any prescribed nonlinear complexity.

Acknowledgments

The authors would like to thank the editor and the anonymous referees for their detailed reading and valuable comments that greatly improved the presentation and quality of the article.

References

[1] S. W. Golomb, *Shift Register Sequences: Secure and Limited-Access Code Generators, Efficiency Code Generators, Prescribed Property Generators, Mathematical Models*, 3rd ed. Singapore: World Scientific, 2017.

[2] P. Martin-Löf, "The definition of random sequences," *Inf. Control.*, vol. 9, no. 6, pp. 602–619, Dec. 1966.

[3] M. Wang, "Cryptographic aspects of sequence complexity measures," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, 1988.

[4] T. Beth and Z.-D. Dai, "On the complexity of pseudo-random sequences - or: if you can describe a sequence it can't be random," in *Advances in Cryptology - EUROCRYPT '89* (Lecture Notes in Computer Science), vol. 434, Berlin, Germany: Springer-Verlag, 1990, pp. 533–543.

[5] H. Niederreiter, "Some computable complexity measures for binary sequences," in *Sequences and their Applications '98*, London, U.K.: Springer-Verlag, 1999, pp. 67–78.

[6] J. L. Massey and S. Serconek, "Linear complexity of periodic sequences: a general theory," in *Advances in Cryptology - CRYPTO '96* (Lecture Notes in Computer Science), vol. 1109, Berlin, Germany: Springer-Verlag, 1996, pp. 358–371.

[7] H. Niederreiter, "Linear complexity and related complexity measures for sequences," in *Progress in Cryptology - INDOCRYPT 2003* (Lecture Notes in Computer Science), vol. 2904, Berlin, Germany: Springer-Verlag, 2003, pp. 1–17.

[8] J. L. Massey, "Shift-register synthesis and BCH decoding," *IEEE Trans. Inf. Theory*, vol. 15, no. 1, pp. 122–127, Jan. 1969.

[9] R. A. Rueppel, *Analysis and Design of Stream Ciphers*. Berlin, Germany: Springer-Verlag, 1986.

[10] W. Meidl and H. Niederreiter, "On the expected value of the linear complexity and the k-error linear complexity of periodic sequences," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2817–2825, Nov. 2002.

[11] T. Helleseth, S.-H. Kim, and J.-S. No, "Linear complexity over $\mathbb{F}_p$ and trace representation of Lempel-Cohn-Eastman sequences," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1548–1552, Jun. 2003.

[12] C. Ding, T. Helleseth, and W. Shan, "On the linear complexity of Legendre sequences," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1276–1278, May 1998.

[13] P. Rizomiliotis, N. Kolokotronis, and N. Kalouptsidis, "On the quadratic span of binary sequences," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1840–1848, May 2005.

[14] C. J. A. Jansen and D. E. Boekee, "The shortest feedback shift register that can generate a given sequence," in *Advances in Cryptology - CRYPTO'89* (Lecture Notes in Computer Science), vol. 435, Berlin, Germany: Springer-Verlag, 1990, pp. 90–99.

[15] L. Işık and A. Winterhof, "Maximum-order complexity and correlation measures," *Cryptography*, vol. 1, no. 1, pp. 1–7, May 2017.

[16] Z. Chen, A. I. Gómez, D. Gómez-Pérez, and A. Tirkel, "Correlation measure, linear complexity and maximum order complexity for families of binary sequences," *Finite Fields Appl.*, vol. 78, no. 101977, Feb. 2022.

[17] C. J. A. Jansen, "Investigations on nonlinear streamcipher systems: construction and evaluation methods," Ph.D. dissertation, Technical University of Delft, Delft, The Netherlands, 1989.

[18] P. Rizomiliotis and N. Kalouptsidis, "Results on the nonlinear span of binary sequences," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1555–1563, Apr. 2005.

[19] K. Limniotis, N. Kolokotronis, and N. Kalouptsidis, "On the nonlinear complexity and Lempel-Ziv complexity of finite length sequences," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4293–4302, Nov. 2007.

[20] H. Niederreiter and C. Xing, "Sequences with high nonlinear complexity," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6696–6701, Oct. 2014.

[21] Y. Luo, C. Xing, and L. You, "Construction of sequences with high nonlinear complexity from function fields," *IEEE Trans. Inf. Theory*, vol. 63, no. 12, pp. 7646–7650, Dec. 2017.

[22] A. S. Castellanos, L. Quoos, and G. Tizziotti, "Construction of sequences with high nonlinear complexity from a generalization of the hermitian function field," *J. Algebra Appl.*, vol. 23, no. 2, 2450037, Feb. 2024.

[23] J. Peng, X. Zeng, and Z. Sun, "Finite length sequences with large nonlinear complexity," *Adv. Math. Commun.*, vol. 12, no. 1, pp. 215–230, Feb. 2018.

[24] L. Yi, X. Zeng, and Z. Sun, "On finite length nonbinary sequences with large nonlinear complexity over the residue ring $\mathbb{Z}_m$," *Adv. Math. Commun.*, vol. 15, no. 4, pp. 701–720, Nov. 2021.

[25] S. Liang, X. Zeng, Z. Xiao, and Z. Sun, "Binary sequences with length n and nonlinear complexity not less than $n/2$," *IEEE Trans. Inf. Theory*, vol. 69, no. 12, pp. 8116–8125, Dec. 2023.

[26] P. Rizomiliotis, "Constructing periodic binary sequences with maximum nonlinear span," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4257–4261, Sep. 2006.

[27] Z. Sun, X. Zeng, C. Li, and T. Helleseth, "Investigations on periodic sequences with maximum nonlinear complexity," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6188–6198, Oct. 2017.

[28] Z. Xiao, X. Zeng, C. Li, and Y. Jiang, "Binary sequences with period N and nonlinear complexity $N-2$," *Cryptogr. Commun.*, vol. 11, no. 4, pp. 735–757, Jul. 2019.

[29] N. G. de Bruijn, "A combinatorial problem," *Proceedings of the Section of Sciences of the Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam*, vol. 49, no. 7, pp. 758–764, Jun. 1946.

[30] S. Golomb, "On the classification of balanced binary sequences of period 2^n-1," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 730–732, Nov. 1980.

[31] G. Kim and H.-Y. Song, "Statistical span property of binary run sequences," *IEEE Trans. Inf. Theory*, vol. 69, no. 4, pp. 2713–2721, Apr. 2023.

[32] T. Helleseth, and C. Li, "Pseudo-noise sequences," in Concise Encyclopedia of Coding Theory, New York: Chapman and Hall/CRC Press, 2021, vol. 25, pp. 613–644.

[33] H. Fredricksen, "A survey of full length nonlinear shift register cycle algorithms," *SIAM Review*, vol. 24, no. 2, pp. 195–221, Apr. 1982.