De Bruijn Sequences from Nonlinear Feedback Shift Registers

Ming Li and Dongdai Lin

State Key Laboratory of Information Security,
Institute of Information Engineering,
Chinese Academy of Sciences, Beijing 100093, China
E-mail: {liming,ddlin}@iie.ac.cn

November 5, 2015

Abstract

We continue the research of Jansen et al. (IEEE Trans on Information Theory 1991) to construct De Bruijn sequences from feedback shift registers (FSRs) that contain only very short cycles. Firstly, we suggest another way to define the representative of a cycle. Compared with the definition made by Jansen et al., this definition can greatly improve the performance of the cycle joining algorithm. Then we construct a large class of nonlinear FSRs that contain only very short cycles. The lengths of the cycles in these n-stage FSRs are less than $2n$. Based on these FSRs, $O(2^{\frac{n}{2}-\log n})$ De Bruijn sequences of order n are constructed. To generate the next bit in the De Bruijn sequence from the current state, it requires only $2n$ bits of storage and less than $2n$ FSR shifts.

Keywords: De Bruijn sequence, feedback shift register, cycle joining method.

1 Introduction

A binary De Bruijn sequence of order n is a sequence of period 2^n in which each n-tuple occurs exactly once in one period [2]. These sequences have many applications in cryptography and modern communication systems. Numerous algorithms for generating these sequences are known, and a useful survey can be found in [5]. A classical method to construct De Bruijn sequences is to consider a feedback shift register (FSR) producing several cycles which are then joined together to form a full cycle, i.e., a De Bruijn cycle. Linear feedback shift registers (LFSRs) with simple cycle structures are often used for this purpose. The LFSRs that contain only very short cycles are good candidates, for example, the pure circulating registers and the pure summing registers [3,4]. The LFSRs that contain a very small number of cycles are also good candidates, for example, the LFSRs with characteristic polynomials of the form $(1+x)^m p(x)$ and $(1+x^m)p(x)$, where $p(x)$ is a primitive polynomial [9,10,13]. By joining the cycles in an LFSR, a large class of maximum-length FSRs can be constructed efficiently. However, this method requires the full knowledge of the cycle structure of the base FSR and the adjacency relations of the cycles in it. Hence, it seems hard to apply this method to a general FSR, especially, a nonlinear FSR.

Jansen et al. [8] proposed an algorithm for joining cycles of an arbitrary FSR. For a given FSR, they defined the representative of a cycle in this FSR as the numerically least state (regard a state as an integer) on this cycle. In the sequel we write least state, denoting the numerically least state. Then they showed that, interchanging the predecessors of the cycle representatives with the predecessors of their companions will result in a full cycle. For the application of their algorithm, one needs to test whether a state is the cycle representative of some cycle or not at every step. Therefore, the performance of their algorithm depends on the length of the longest cycle in the base FSR. FSRs that contain only very short cycles are needed. To find such FSRs, the authors turned to linear feedback shift registers. By conduct a large number of irreducible polynomials of the same degree, a polynomial whose period is very low (relative to its degree) is obtained. The LFSRs that take such polynomials as their characteristic polynomials contain only very short cycles, and they can be used to generate De Bruijn sequences.

The research of Jansen et al. [8] is continued in this paper. To improve the performance of the cycle joining algorithm, we suggest another way to define the representative of a cycle. Compared with the definition in [8], this definition doubles the efficiency of the cycle joining algorithm. Furthermore, the definition in this paper is flexible, which implies more choices of the cycle representative. We also find a class of FSRs that contains only very short cycles using a different method. These FSRs are not linear and they are easy to get. The size of the class of these FSRs is larger than those in [8]. It is shown that, $O(2^{\frac{n}{2}-\log n})$ De Bruijn sequences of order n can be constructed from these FSRs.

The paper is organized as follows. In Section 2, we introduce some necessary preliminaries. In Section 3, an algorithm for joining cycles of an arbitrary FSR is presented. In Section 4, a large class of nonlinear FSRs that contain only very short cycles are proposed. The number of De Bruijn sequences constructed from them is also given. In Section 5, we list some comparisons of the results in [8] and ours. In Section 6, we make a conclusion about our work.

2 Preliminaries

2.1 Boolean Functions

Let $\mathbb{F}_2 = \{0, 1\}$ be the finite field of two elements, and $\mathbb{F}_2^n$ be the vector space of dimension n over $\mathbb{F}_2$. A Boolean function $f(x_0, x_1, \ldots, x_{n-1})$ in n variables is a mapping from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. It is well known that it can be uniquely represented by its algebraic normal form (ANF), which is a multivariate polynomial. The order of f, denoted by $\mathrm{ord}(f)$, is the highest subscript i for which x_i occurs in the ANF of f. Note that the order of f is not equal to the number of variables in f. The Hamming weight of f is defined by $\mathrm{w}(f) = \sharp\{x : f(x) \neq 0\}$. The Hamming distance of two Boolean functions is defined by $\mathrm{d}(f, g) = \sharp\{x : f(x) \neq g(x)\}$. For two Boolean functions $f(x_0, x_1, \ldots, x_n)$ and $g(x_0, x_1, \ldots, x_m)$, we denote $f * g = f(g(x_0, x_1, \ldots, x_m),\ g(x_1, x_2, \ldots, x_{m+1}),\ \ldots, g(x_n, x_{n+1}, \ldots, x_{n+m}))$, which is a Boolean function of order $n + m$ [7].

Reed-Muller codes, named after Irving S. Reed and David E. Muller, are a family of linear error-correcting codes used in communications. The Reed-Muller code of order r and length $l = 2^m$, denoted by $RM(r, m)$, is the code that contains all the m-variable Boolean functions of degree no more than r. It was proved that, $RM(r, m)$ has minimum Hamming distance 2^{m-r} [12, 14]. Therefore, we have the following lemma.

Lemma 1. *[12, 14] Let g_1 and g_2 be two Boolean functions such that* $\mathrm{ord}(g_1) = \mathrm{ord}(g_2) = n$ *and* $\deg(g_1) = \deg(g_2) = r$*, then* $\mathrm{d}(g_1, g_2) \geq 2^{n+1-r}$.

2.2 Feedback Shift Registers

An n-stage feedback shift register (FSR) consists of n binary storage cells and a characteristic function f regulated by a single clock. In what follows, the characteristic function f is supposed to be nonsingular, i.e., of the form $f = x_0 + f_0(x_1, \ldots, x_{n-1}) + x_n$. The feedback function of this FSR is defined as $F(x_0, x_1, \ldots, x_{n-1}) = x_0 + f_0(x_1, \ldots, x_{n-1})$. The FSR with characteristic function f is denoted by $\mathrm{FSR}(f)$. At every clock pulse, the current state $(s_0, s_1, \ldots, s_{n-1})$ is updated by $(s_1, s_2, \ldots, s_{n-1}, F(s_0, s_1, \ldots, s_{n-1}))$. From an initial state $\mathbf{S}_0 = (s_0, s_1, \ldots, s_{n-1})$, after consecutive clock pulses, $\mathrm{FSR}(f)$ will generate a cycle $C = [\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_{l-1}]$, where $\mathbf{S}_{i+1}$ is the next state of $\mathbf{S}_i$ for $i = 0, 1, \ldots, l-2$ and $\mathbf{S}_0$ is the next state of $\mathbf{S}_{l-1}$. In this way, the set $\mathbb{F}_2^n$ is divided into cycles $C_1, C_2, \ldots, C_k$ by $\mathrm{FSR}(f)$, and reversely, it is easy to see, a partition of $\mathbb{F}_2^n$ into cycles determines an n-stage FSR. So we can treat $\mathrm{FSR}(f)$ as a set of cycles. The output sequences of $\mathrm{FSR}(f)$, denoted by $G(f)$, are the 2^n sequences $\mathbf{s} = s_0s_1\ldots$, such that $s_{t+n} = F(s_t, s_{t+1}, \ldots, s_{t+n-1})$ for $t \geq 0$. An FSR is called a linear feedback shift register (LFSR) if its characteristic function f is linear. For a linear Boolean function $f(x_0, x_1, \ldots, x_n) = a_0x_0 + a_1x_1 + \cdots + a_nx_n$, we can associate it with an univariate polynomial $c(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{F}_2[x]$. Most of the time, we do not discriminate between linear Boolean functions and univariate polynomials. For an n-stage FSR, the period of its output sequence is no more than 2^n. If this value is attained, we call the sequence De Bruijn sequence, and the FSR maximum-length FSR.

2.3 The Inverse of a Linear Characteristic Function

Let $f(x_0, x_1, \ldots, x_n)$ be a Boolean function and $\mathbf{a} = a_0a_1\ldots$, be a periodic sequence. Define $\theta(f)$ to be the mapping on the periodic sequences: $\theta(f)(\mathbf{a}) = \mathbf{b}$, where $\mathbf{b}$ is determined by $b_i = f(a_i, a_{i+1}, \ldots, a_{i+n})$. Let $\theta(f)^{-1}(\mathbf{a})$ be the set of sequences whose image is $\mathbf{a}$ under $\theta(f)$, i.e., $\theta(f)^{-1}(\mathbf{a}) = \{\mathbf{b} : \theta(f)(\mathbf{b}) = \mathbf{a}\}$. $\theta(f)^{-1}(\mathbf{a})$ contains 2^n sequences, and if f is linear, $\theta(f)^{-1}(\mathbf{a})$ is a linear space of dimension n over $\mathbb{F}_2$. It can be verified that, $\theta(f)^{-1}(\mathbf{0}) = G(f)$. Some properties of $\theta(f)^{-1}$ were given in [13]. Let g be the linear Boolean function with the least order such that $\theta(g)(\mathbf{a}) = \mathbf{0}$, then the linear complexity of $\mathbf{a}$ is defined to be the order of g. The minimal polynomial of $\mathbf{a}$, denoted by $m(\mathbf{a})$, is the univariate polynomial corresponding to g.

Lemma 2. *[13] Let $f(x)$ be a linear Boolean function and $\mathbf{a}$ be a periodic sequence.*

1. *If $\gcd(f, m(\mathbf{a})) = 1$, then $\theta(f)^{-1}(\mathbf{a}) = \mathbf{b} + G(f)$ for some $\mathbf{b} \in G(m(\mathbf{a}))$ with $m(\mathbf{b}) = m(\mathbf{a})$.*
2. *If f is irreducible and $m(\mathbf{a}) = hf^e$, $e \geq 1$ with $\gcd(h, f) = 1$, then $m(\mathbf{b}) = hf^{e+1}$ for all $\mathbf{b} \in \theta(f)^{-1}(\mathbf{a})$.*
3. *$\mathrm{lcm}\{m(\mathbf{b}) : \mathbf{b} \in \theta(f)^{-1}(\mathbf{a})\} = m(\mathbf{a})f$, where lcm is the least common multiple.*

3 Cycle Joining Algorithm

For a state $\mathbf{S} = (s_0, s_1, \ldots, s_{n-1})$, its companion is defined as $\widetilde{\mathbf{S}} = (s_0, s_1, \ldots, \bar{s}_{n-1})$, where $\bar{s}_{n-1}$ is the complement of s_{n-1}. Sometimes, we regard $\mathbf{S}$ as an integer, i.e., $\mathbf{S} = \Sigma_{i=0}^{n-1} s_i 2^{n-1-i}$. Two cycles C_1 and C_2 are said to be adjacent if they are state disjoint and there exists a state $\mathbf{S}$ on C_1 whose companion $\widetilde{\mathbf{S}}$ is on C_2. By interchanging the predecessors of $\mathbf{S}$ and $\widetilde{\mathbf{S}}$, the two cycles C_1 and C_2 are joined together. This is the basic idea of the cycle joining method. For the application of this method, we need to find the companion pairs shared by cycles. In [8], the cycle representative of a cycle is defined as the least state (regard states as integers) on this cycle, and they showed how to join the cycles in an arbitrary FSR into a full cycle with the help of cycle representatives. In the following, we suggest another way to define the cycle representative.

Definition 1. *Let C be a cycle such that the zero state $\mathbf{0}$ is not on C. The cycle representative of C is defined as the least state $\mathbf{S}$ on C such that: $\mathbf{S}$ contains the longest run of 0s and is of the form $(*, \ldots, *, \overbrace{0, \ldots, 0}^{t}, 1)$, where t is the length of the longest run of 0s.*

Note 1. *The cycle representative of C can also be defined as the greatest (or some other type, as long as it is uniquely defined) such state.*

It is easy to see that, for any cycle C such that $\mathbf{0} \notin C$, its representative is uniquely determined by C. For example, the cycle representative of $C = [001, 010, 100]$ is (001).

Theorem 1. *For a given FSR, let $C_0, C_1, \ldots, C_k$ be the cycles in it. Assume C_0 is the cycle that contains the zero state $\mathbf{0}$. Let $\mathbf{S}_i$ be the cycle representative of C_i for $i = 1, 2, \ldots, k$. If we interchange the predecessors of $\mathbf{S}_i$ and $\widetilde{\mathbf{S}}_i$ for $i = 1, 2, \ldots, k$, we get a full cycle.*

Proof. Let t_i be the length of the longest run of 0s in $\mathbf{S}_i$ for $i = 1, 2, \ldots, k$. Since $\mathbf{S}_i$ is of the form $\mathbf{S}_i = (*, \ldots, *, \overbrace{0, \ldots, 0}^{t_i}, 1)$, the length of the longest run of 0s in $\widetilde{\mathbf{S}}_i$ is $t_i + 1$. By the definition of the cycle representative, there is no state on C_i that contains more than t_i successive 0s. Therefore, $\widetilde{\mathbf{S}}_i$ is on some cycle other than C_i. Assume $\widetilde{\mathbf{S}}_i$ is on the cycle C_j with $j \neq i$. Since $\widetilde{\mathbf{S}}_i$ is of the form $\widetilde{\mathbf{S}}_i = (*, \ldots, *, \overbrace{0, \ldots, 0}^{t_i+1})$, $\widetilde{\mathbf{S}}_i$ is not the cycle representative of C_j, that is, $\widetilde{\mathbf{S}}_i \neq \mathbf{S}_j$. Again by the definition of the cycle representative, we get $t_j > t_i$. Let G be the directed graph that take $C_0, C_1, \ldots, C_k$ as its nodes, and there is a directed edge from C_i to C_j if and only if $\widetilde{\mathbf{S}}_i$ is on C_j. Then by the above discussion, G is a directed tree with root C_0. This tree represents a choice of companion pairs that repeatedly join two cycles into one ending with exactly one cycle, i.e., a full cycle. $\square$

By this theorem, we can join the cycles in an arbitrary FSR into a full cycle. For a given FSR, let $F(x_0, x_1, \ldots, x_{n-1})$ be the feedback function of this FSR. Let C be the full cycle determined by this FSR according to Theorem 1. From any state $\mathbf{U}_i$, the next state $\mathbf{U}_{i+1}$ in the full cycle C is calculated by the following algorithm.

This algorithm complements the value of the feedback function only if there is a cycle representative amongst the two possible successors $(u_{i+1}, \ldots, u_{i+n-1}, 0)$ and $(u_{i+1}, \ldots, u_{i+n-1}, 1)$. By the definition of the cycle representative in this paper, $(u_{i+1}, \ldots, u_{i+n-1}, 0)$ would never be a cycle representative, therefore, we only need to test whether $(u_{i+1}, \ldots, u_{i+n-1}, 1)$ is a cycle representative. While in [8], both of the two possible successors need to be tested. So the efficiency of the cycle joining algorithm is doubled. An obvious way to do the test is by

Algorithm 1 Generation of the next state in the full cycle

Input:

The feedback function $F(x_0, x_1, \ldots, x_{n-1})$ of the base FSR.

The current state $\mathbf{U}_i = (u_i, u_{i+1}, \ldots, u_{i+n-1})$.

Output: The next state $\mathbf{U}_{i+1}$.

if $(u_{i+1}, \ldots, u_{i+n-1}, 1)$ is a cycle representative **then**

$\mathbf{U}_{i+1} = (u_{i+1}, \ldots, u_{i+n-1}, F(u_{i+1}, u_{i+2}, \ldots, u_{i+n-1}) + 1)$

else

$\mathbf{U}_{i+1} = (u_{i+1}, \ldots, u_{i+n-1}, F(u_{i+1}, u_{i+2}, \ldots, u_{i+n-1}))$

end if

traversing the cycle that contains the state needing to be tested. Hence, it require $2n$ bits of storage and at most l FSR shifts for the generation of the next state in the full cycle, where l is the length of the longest cycle in the base FSR. However, for certain states it is immediately clear that they cannot be cycle representatives.

Theorem 2. *Let C be a cycle and $\mathbf{X} = (*, \ldots, *, 1, \overbrace{0, \ldots, 0}^{t}, 1)$ be a state in C, then none of the next t states would be a cycle representative.*

Proof. None of the next t states would be of the form $(*, \ldots, *, 1, \overbrace{0, \ldots, 0}^{u}, 1)$ with $u \geq t$. By the definition of the cycle representative, none of them is the cycle representative. $\square$

4 The Base FSRs

The performance of the cycle joining algorithm proposed in Section 3 depends heavily on the length of the longest cycle in the base FSR. Evidently, the FSRs that contain only very short cycles are needed. In [8], a class of such LFSRs are constructed using the theory of LFSRs. In this section, we propose another class of such FSRs which are nonlinear. Let $p(x) \in \mathbb{F}_2[x]$ be a polynomial. The period of $p(x)$, denoted by $\mathrm{per}(p(x))$, is the least integer k such that $p(x)|1 + x^k$. Some properties about the period of a polynomial can be found in [11].

Lemma 3. *Let $1 \leq t \leq 2^m$ be an integer, then* $\mathrm{per}((1 + x^t)(1 + x^{2^m})) = \mathrm{lcm}(t, 2^{m+1})$.

Proof. Let $t = 2^u v$, where v is an odd number. Then $(1 + x^t)(1 + x^{2^m}) = (1 + x^v)^{2^u}(1 + x)^{2^m} = (1 + x)^{2^u}(1 + x + \cdots + x^{v-1})^{2^u}(1 + x)^{2^m} = (1 + x + \cdots + x^{v-1})^{2^u}(1 + x)^{2^m+2^u}$. Since $\gcd((1+x+\cdots+x^{v-1})^{2^u}, (1+x)^{2^m+2^u}) = 1$, we have $\mathrm{per}((1+x+\cdots+x^{v-1})^{2^u}(1+x)^{2^m+2^u}) = \mathrm{lcm}(\mathrm{per}((1+x+\cdots+x^{v-1})^{2^u}), \mathrm{per}((1+x)^{2^m+2^u})$. It is easy to see, $\mathrm{per}(1+x+\cdots+x^{v-1}) = v$, hence, $\mathrm{per}((1+x+\cdots+x^{v-1})^{2^u}) = v2^u = t$. Since $1 \le 2^u \le 2^m$, we have $\mathrm{per}((1+x)^{2^m+2^u}) = 2^{m+1}$. So we get $\mathrm{per}((1+x^t)(1+x^{2^m})) = \mathrm{lcm}(t, 2^{m+1})$. $\square$

Theorem 3. *Let f be a Boolean function of order m, then for any sequence $\mathbf{s} \in G(f*(x_0+x_{2^m}))$ we have $\mathrm{per}(\mathbf{s}) \le 2^{2m+1}$.*

Proof. Let $\mathbf{s}$ be a sequence in $G(f*(x_0+x_{2^m}))$. Then we have $\theta(f*(x_0+x_{2^m}))(\mathbf{s}) = \theta(f)\theta(x_0+x_{2^m})(\mathbf{s}) = \mathbf{0}$. Therefore, $\theta(x_0+x_{2^m})(\mathbf{s}) \in \theta(f)^{-1}(\mathbf{0})$. Since $\theta(f)^{-1}(\mathbf{0}) = G(f)$, there exist some sequence $\mathbf{a} \in G(f)$ such that $\theta(x_0+x_{2^m})(\mathbf{s}) = \mathbf{a}$. This implies $\mathbf{s} \in \theta(x_0+x_{2^m})^{-1}(\mathbf{a})$. According to Case 3 of Lemma 2, we have $m(\mathbf{s})|m(\mathbf{a})(1+x^{2^m})$. Let $0 \le t \le 2^m$ be the period of $\mathbf{a}$, then $m(\mathbf{a})|(1+x^t)$. Thus $m(\mathbf{a})(1+x^{2^m})|(1+x^t)(1+x^{2^m})$. According to Lemma 3, $\mathrm{per}((1+x^t)(1+x^{2^m})) = \mathrm{lcm}(t, 2^{m+1})$. Consider that $m(\mathbf{s})|(1+x^t)(1+x^{2^m})$, we get $\mathrm{per}(m(\mathbf{s})) \le \mathrm{lcm}(t, 2^{m+1}) \le t2^{m+1} \le 2^{2m+1}$. $\square$

A more careful calculation shows that $\mathrm{per}(\mathbf{s}) \le (2^m-1)\,2^{m+1}$ for any sequence $\mathbf{s} \in G(f*(x_0+x_{2^m}))$. As a generalization of this theorem, we can prove that: let f be a Boolean function of order m and h be a linear Boolean function that corresponds to an irreducible polynomial of period k, then $\mathrm{per}(\mathbf{s}) \le k2^{2m+1}$ for any sequence $\mathbf{s} \in G(f*h^{2^m})$, where h^{2^m} means $\overbrace{h*h*\cdots*h}^{2^m}$.

Theorem 4. *Let $\mathrm{FSR}(f)$ be an m-stage maximum-length FSR. Then for any sequence $\mathbf{s} \in G(f*(x_0+x_{2^m}))$ we have $\mathrm{per}(\mathbf{s}) = 2^{m+1}$.*

Proof. Let $\mathbf{s}$ be a sequence in $G(f*(x_0+x_{2^m}))$, and $\mathbf{a}$ be the De Bruijn sequence in $G(f)$ such that $\mathbf{s} \in \theta(x_0+x_{2^m})^{-1}(\mathbf{a})$ (see the proof in Theorem 3). The minimal polynomial of $\mathbf{a}$ is of the form $m(\mathbf{a}) = (1+x)^C$, where $2^{m-1}+m \le C \le 2^m-1$ is the linear complexity of $\mathbf{a}$ (see [1]). According to Case 2 of Lemma 2, the minimal polynomial of any sequence in $\theta(x_0+x_1)^{-1}(\mathbf{a})$ is $(1+x)^{C+1}$. Since $\theta(x_0+x_2)^{-1}(\mathbf{a}) = \theta(x_0+x_1)^{-1}\theta(x_0+x_1)^{-1}(\mathbf{a})$, the minimal polynomial of any sequence in $\theta(x_0+x_2)^{-1}(\mathbf{a})$ is $(1+x)^{C+2}$. Repeat this process, we know that the minimal polynomial of any sequence in $\theta(x_0+x_{2^m})^{-1}(\mathbf{a})$ is $(1+x)^{C+2^m}$. Since $\mathrm{per}((1+x)^{C+2^m}) = 2^{m+1}$, we get $\mathrm{per}(\mathbf{s}) = \mathrm{per}(m(\mathbf{s})) = 2^{m+1}$. $\square$

Note 2. *Let f be the characteristic function of an m-stage maximum-length FSR. Using the same method as in Theorem 4 we can show that, for any sequence $\mathbf{s} \in G(f*(x_0+x_{2^{m+1}-C}))$ we have $\mathrm{per}(\mathbf{s}) = 2^{m+1}$, where C is the linear complexity of the De Bruijn sequences in $G(f)$.*

Denote the number 2^m+m by n. According to Theorem 3, the length of the cycles in $G(f*(x_0+x_{2^m}))$ are no more than $2n^2$ for any f of order m. Especially, if f is the characteristic function of a maximum-length FSR, then the length of the cycles in $G(f*(x_0+x_{2^m}))$ are no more than $2n$. These FSRs are good candidates for the cycle joining algorithm. In the following, we consider the number of full cycles constructed from them by the cycle joining algorithm. We use the fact about minimum Hamming distance of Reed-Muller codewords, which is suggested by Jansen et al. [8]. First, we need a lemma.

Lemma 4. *For any Boolean function f of order m we have* $\deg(f * (x_0 + x_{2^m})) = \deg(f)$.

Proof. For any term $x_{i_1}x_{i_2}\cdots x_{i_k}$, since $(x_{i_1}x_{i_2}\cdots x_{i_k}) * (x_0 + x_{2^m}) = (x_{i_1} + x_{i_1+2^m})(x_{i_2} + x_{i_2+2^m})\cdots(x_{i_k}+x_{i_k+2^m})$, we have $\deg((x_{i_1}x_{i_2}\cdots x_{i_k})*(x_0+x_{2^m})) \le \deg(x_{i_1}x_{i_2}\cdots x_{i_k})$. Hence, $\deg(f*(x_0+x_{2^m})) \le \deg(f)$. We associate each term of f with an integer: $N(x_{i_1}x_{i_2}\cdots x_{i_k}) = 2^{i_1}+2^{i_2}+\cdots+2^{i_k}$. Let $x_{j_1}x_{j_2}\cdots x_{j_d}$ be the term of f such that $N(x_{j_1}x_{j_2}\cdots x_{j_d})$ is the smallest among all the terms of f of degree d, where d is the degree of f. The lemma will be proved if we can show that $x_{j_1}x_{j_2}\cdots x_{j_d}$ is also a term of $f * (x_0 + x_{2^m})$. First, it is easy to see, $(x_{j_1}x_{j_2}\cdots x_{j_d}) * (x_0 + x_{2^m})$ contains the term $x_{j_1}x_{j_2}\cdots x_{j_d}$. Let $x_{i_1}x_{i_2}\cdots x_{i_k}$ be a term of f such that $x_{i_1}x_{i_2}\cdots x_{i_k} \ne x_{j_1}x_{j_2}\cdots x_{j_d}$. We need to show that $(x_{i_1}x_{i_2}\cdots x_{i_k})*(x_0+x_{2^m})$ does not contain the term $x_{j_1}x_{j_2}\cdots x_{j_d}$. If $k \ne d$, $(x_{i_1}x_{i_2}\cdots x_{i_k}) * (x_0 + x_{2^m})$ contains only terms of degree k, therefore, does not contain the term $x_{j_1}x_{j_2}\cdots x_{j_d}$. If $k = d$, by the definition of $x_{j_1}x_{j_2}\cdots x_{j_d}$, we have $2^{i_1} + 2^{i_2} + \cdots + 2^{i_k} > 2^{j_1} + 2^{j_2} + \cdots + 2^{j_d}$. $(x_{i_1}x_{i_2}\cdots x_{i_k}) * (x_0 + x_{2^m})$ contains only terms whose associated integers are no less than $2^{i_1} + 2^{i_2} + \cdots + 2^{i_k}$, therefore, does not contain the term $x_{j_1}x_{j_2}\cdots x_{j_d}$. □

Let f be a Boolean function of order m. Some necessary conditions for $\mathrm{FSR}(f)$ be a maximum-length FSR are given in [6]. One of these conditions is that $\deg(f) = m - 1$.

Theorem 5. *Let f_1 and f_2 be characteristic functions of two m-stage maximum-length FSRs with $f_1 \ne f_2$. If the cycle joining algorithm is applied to* $\mathrm{FSR}(f_1 * (x_0 + x_{2^m}))$ *and* $\mathrm{FSR}(f_2 * (x_0 + x_{2^m}))$ *respectively, the two resulting De Bruijn sequences are different.*

Proof. By Lemma 4, we have $\deg(f_1*(x_0+x_{2^m})) = \deg(f_1) = m-1$ and $\deg(f_2*(x_0+x_{2^m})) = \deg(f_2) = m - 1$. Since $\mathrm{ord}(f_1 * (x_0 + x_{2^m})) = \mathrm{ord}(f_2 * (x_0 + x_{2^m})) = 2^m + m$, according to Lemma 1, $\mathrm{d}(f_1 * (x_0 + x_{2^m}), f_2 * (x_0 + x_{2^m})) \ge 2^{2^m+2}$. Let $\mathrm{FSR}(h_1)$ and $\mathrm{FSR}(h_2)$ be the two maximum-length FSRs derived from $\mathrm{FSR}(f_1 * (x_0 + x_{2^m}))$ and $\mathrm{FSR}(f_2 * (x_0 + x_{2^m}))$ by the cycle joining algorithm. According to Theorem 4, $\mathrm{FSR}(f_1 * (x_0 + x_{2^m}))$ contains only cycles of length 2^{m+1}, hence, there are $\frac{2^{2^m+m}}{2^{m+1}} = 2^{2^m-1}$ cycles in this FSR. Every time two cycles in this FSR are joined together in the process of the cycle joining algorithm, the weight of the corresponding characteristic function is changed by 4. Therefore, we get $\mathrm{d}(f_1 * (x_0 + x_{2^m}), h_1) \le 4\left(2^{2^m-1} - 1\right) = 2^{2^m+1} - 4$. By the same reason, we have $\mathrm{d}(f_2 * (x_0 + x_{2^m}), h_2) \le 2^{2^m+1} - 4$. The proof of this theorem can be done as follows, $\mathrm{d}(h_1, h_2) \ge \mathrm{d}(f_1 * (x_0 + x_{2^m}), f_2 * (x_0 + x_{2^m})) - \mathrm{d}(f_1 * (x_0 + x_{2^m}), h_1) - \mathrm{d}(f_2 * (x_0 + x_{2^m}), h_2) \ge 2^{2^m+2} - \left(2^{2^m+1} - 4\right) - \left(2^{2^m+1} - 4\right) = 8$. □

The number of De Bruijn sequences of order m is $2^{2^{m-1}-m}$. According to Theorem 5, we have constructed $2^{2^{m-1}-m}$ De Bruijn sequences based on the FSRs with characteristic function of the form $f*(x_0+x_{2^m})$, where f is the characteristic function of an m-stage maximum-length FSR. Let $n=2^m+m$ be the order of $f*(x_0+x_{2^m})$, we have $2^{2^{m-1}-m}=O(2^{\frac{n}{2}-\log n})$. This implies, the size of the De Bruijn sequences we have constructed grows exponentially with the order.

When searching for FSRs that contain only very short cycles, we observed that, the n-stage pure circulating shift register, i.e. $\mathrm{FSR}(x_0+x_n)$, is the unique n-stage FSR that contains only cycles of length no more than n. The proof of this theorem is tedious, so we put it in the appendix.

Theorem 6. *Let* $\mathrm{FSR}(f)$ *be an* n*-stage FSR that contains only cycles of length no more than* n*, then* $f=x_0+x_n$*.*

5 Some Comparisons

In this section, we compare the results of Jansen et al. [8] with ours. At first, we present an example to illustrate that the two definitions of the cycle representative in [8] and in this paper are essentially different.

Example 1. *Let* $f=x_0+1+x_2$ *be the characteristic function of the unique* 2*-stage maximum-length FSR. Let* $g=f*(x_0+x_4)=x_0+x_2+x_4+1+x_6$*. According to Theorem 4, There are* 8 *cycles in* $\mathrm{FSR}(g)$*, all of them are of length* 8*.*

$$C_0=[000000,000001,000011,000110,001100,011000,110000,100000],$$

$$C_1=[000010,000100,001001,010010,100100,001000,010000,\underline{100001}],$$

$$C_2=[000101,001011,010111,101110,011100,111000,\underline{110001},100010],$$

$$C_3=[000111,001110,011101,111010,110100,101000,\underline{010001},100011],$$

$$C_4=[001010,010101,101011,010110,101100,\underline{011001},110010,100101],$$

$$C_5=[001101,011010,110101,101010,010100,\underline{101001},010011,100110],$$

$$C_6=[001111,011111,111111,111110,111100,\underline{111001},110011,100111],$$

$$C_7=[011011,110111,101111,011110,111101,111011,110110,\underline{101101}].$$

The first state in each cycle is the cycle representative defined in [8], and the underlined state in each cycle is the cycle representative defined in this paper. Let G *be the directed graph that take* $C_0,C_1,\ldots,C_7$ *as its nodes, and there is a directed edge from* C_i *to* C_j *if and only if* *the companion of the representative of C_i is located on C_j, then G is a directed tree with root C_0 (see the proof of Theorem 1). The two trees are shown below, where the left one is based on the definition of cycle representative in [8] and the right one is based on the definition in this paper.*

$$\begin{tikzcd}
 & C_0 & \\
C_1 \arrow[ur] & C_3 \arrow[u] & C_5 \arrow[ul] \\
C_2 \arrow[u] & C_6 \arrow[u] & C_7 \arrow[u] \\
C_4 \arrow[u] & &
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
 & C_0 & \\
C_1 \arrow[ur] & C_2 \arrow[u] & C_4 \arrow[ul] \\
C_3 \arrow[u] & C_6 \arrow[u] & C_7 \arrow[u] \\
C_5 \arrow[u] & &
\end{tikzcd}$$

We note that, the height of the directed tree base on the definition in [8] may achieve $Z(n)$, the maximum number of cycles in an n-stage FSR. While based on the definition in this paper, the hight does not exceed n.

Table 1 shows the performance of the cycle joining algorithm in the two papers. To generate the next state in the full cycle from the current state, it requires $3n$ bits of storage and $4n$ FSR shifts in [8]. While in this paper, it requires only $2n$ bits of storage and $2n$ FSR shifts. Since $O(2^{\frac{2n}{\log 2n}})/O(2^{\frac{n}{2}-\log n}) \to 0$ as $n \to \infty$, the number of De Bruijn sequences constructed in this paper is more than those in [8]. In fact, when $n > 30$ we have $2^{\frac{n}{2}-\log n} > 2^{\frac{2n}{\log 2n}}$ (see also Table 2).

Table 1: The performance of the cycle joining algorithm

	storage	FSR shifts	$\sharp$ De Bruijn sequences
Jansen et al. [8]	$3n$	$4n$	$O(2^{\frac{2n}{\log 2n}})$
Ours	$2n$	$2n$	$O(2^{\frac{n}{2}-\log n})$

In the following, we consider the base FSRs proposed in the two papers. In [8], a class of LFSRs that contain only very short cycles were constructed. Let $N_1(d)$ be the number of irreducible polynomials of degree d. According to the theory of finite fields, we have $N_1(d) = \sum_{t|d} \mu\left(\frac{d}{t}\right) 2^t$, where μ is the Möbius function. By conducting a half of these irreducible polynomials, a polynomial, denoted by $p(x)$, of degree $n_1(d) = d\left\lceil \frac{N_1(d)}{2} \right\rceil$ is obtained. The period of $p(x)$ is no more than $l_1(d) = 2^d - 1$, therefore, the LFSR with characteristic polynomial $p(x)$ contains cycles of length no more than $l_1(d)$. It is easy to see, there are

$\hat{N}_1(d) = \binom{N_1(d)}{\lceil \frac{N_1(d)}{2} \rceil}$ choices for $p(x)$. Based on these LFSRs, $\hat{N}_1(d)$ De Bruijn sequences of order $n_1(d)$ can be constructed. In Table 2(a), we list these numbers for $d = 2, 3, \ldots, 10$.

Let f be the characteristic function of an m-stage maximum-length FSR. The order of $f * (x_0 + x_{2^m})$ is $n_2(m) = m + 2^m$. According to Theorem 4, the length of the cycles in FSR($f * (x_0 + x_{2^m})$) is $l_2(m) = 2^{m+1}$. There are $\hat{N}_2(m) = 2^{2^{m-1}-m}$ choices for f. In Table 2(b), we list these numbers for $m = 2, 3, \ldots, 10$.

Table 2: The base FSRs

(a) Jansen et al. [8]

d	$N_1(d)$	$n_1(d)$	$l_1(d)$	$\hat{N}_1(d)$
2	1	2	3	1
3	2	3	7	2
4	3	8	15	3
5	6	15	31	$\geq 2^4$
6	9	30	63	$\geq 2^6$
7	18	63	127	$\geq 2^{15}$
8	30	120	255	$\geq 2^{27}$
9	56	252	511	$\geq 2^{52}$
10	99	500	1023	$\geq 2^{95}$

(b) Ours

m	$n_2(m)$	$l_2(m)$	$\hat{N}_2(m)$
2	6	8	2^0
3	11	16	2^1
4	20	32	2^4
5	37	64	2^{11}
6	70	128	2^{26}
7	135	256	2^{57}
8	264	512	2^{120}
9	521	1024	2^{247}
10	1034	2048	2^{502}

The orders of the De Bruijn sequences constructed in both papers do not cover all positive integers. In fact, the order of the De Bruijn sequences in [8] can only take the form of $n_1(d) = d \left\lceil \frac{\sum_{t|d} \mu\left(\frac{d}{t}\right) 2^t}{2} \right\rceil$ where d runs over the positive integers. While in this paper, the order of the De Bruijn sequences takes the form of $n_2(m) = m + 2^m$, where m runs over the positive integers. However, this problem can be partially solved according to Note 2. Some more solutions to this problem are needed and it will be studied further in the future.

6 Conclusion

The performance of the cycle joining algorithm proposed by Jansen et al. [8] is improved in this paper. A large class of nonlinear FSRs that contain only very short cycles are proposed. Based on these FSRs, $O(2^{\frac{n}{2}-\log n})$ De Bruijn sequences of order n are constructed, and it requires only $2n$ bits of storage and less than $2n$ FSR shifts to generate the next bit in the De Bruijn sequence. Since the orders of the De Bruijn sequences constructed in both [8] and this paper do not cover all positive integers, more research is needed.

Appendix: Proof of Theorem 6

For a state $\mathbf{S} = (s_0, s_1, \ldots, s_{n-1})$, the period of $\mathbf{S}$, denoted by $\mathrm{per}(\mathbf{S})$, is defined as the least positive integer $u \leq n$ such that $\mathbf{S} = (s_0, s_1, \ldots, s_{u-1}, \ldots, s_0, s_1, \ldots, s_{u-1})$. The period of any state is a divisor of its length. The conjugate of $\mathbf{S}$ is defined as $\widehat{\mathbf{S}} = (\bar{s}_0, s_1, \ldots, s_{n-1})$.

Lemma 5. *Let* $\mathbf{S}$ *be a state of length* n*. If* $\mathrm{per}(\mathbf{S}) < n$*, then* $\mathrm{per}(\widehat{\mathbf{S}}) = n$*.*

Proof. Let u and v be the period of $\mathbf{S}$ and $\widehat{\mathbf{S}}$ respectively. Since u is a divisor of n and $u < n$, we have $u \leq \frac{n}{2}$. Suppose $v < n$. Since v is a divisor of n, we have $v \leq \frac{n}{2}$. It is easy to see that $u + v \neq n$, therefore, we have $u + v < n$. Let $\mathbf{S} = (s_0, s_1, \ldots, s_{n-1})$ and $\widehat{\mathbf{S}} = (\bar{s}_0, s_1, \ldots, s_{n-1})$. By $\mathrm{per}(\mathbf{S}) = u$, we know $s_{u+v} = s_v$, then by $\mathrm{per}(\widehat{\mathbf{S}}) = v$, we know $s_v = \bar{s}_0$, therefore, we have $s_{u+v} = \bar{s}_0$. By $\mathrm{per}(\widehat{\mathbf{S}}) = v$, we know $s_{u+v} = s_u$, then by $\mathrm{per}(\mathbf{S}) = u$, we know $s_u = s_0$, therefore, we have $s_{u+v} = s_0$. This is a contradiction. □

Corollary 1. *Let* C_1 *and* C_2 *be two cycles in the* n*-stage pure circulating register of length less than* n*, then* C_1 *and* C_2 *are not adjacent.*

Proof. Let $\mathbf{S}$ be a state on C_1, then according to Lemma 5, $\widehat{\mathbf{S}}$ is not on C_2 ($\widehat{\mathbf{S}}$ is on some cycle of length n). This implies, C_1 and C_2 are not adjacent. □

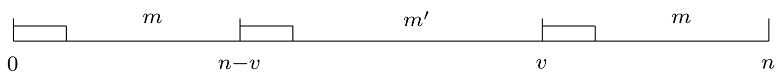
For a state $\mathbf{S} = (s_0, s_1, \ldots, s_{n-1})$, the extend-period of $\mathbf{S}$, denoted by $\mathrm{Eper}(\mathbf{S})$, is defined as the least positive integer u such that $(s_0, s_1, \ldots, s_{n-u-1}) = (s_u, s_{u+1}, \ldots, s_{n-1})$. If there is no such integer, then we say $\mathrm{Eper}(\mathbf{S}) = n$. It can be verified that, $\mathrm{Eper}(\mathbf{S})$ is the minimum period of the periodic sequences whose first n bits are $s_0 s_1 \ldots s_{n-1}$. The left shift operator L is defined as $L(s_0, s_1, \ldots, s_{n-1}) = (s_1, s_2, \ldots, s_{n-1}, s_0)$.

Lemma 6. *Let* $\mathbf{S} = (s_0, s_1, \ldots, s_{n-1})$ *be a state such that* $\mathrm{per}(\mathbf{S}) = n$*. Let* $\mathbf{R}$ *be the least state of the set* $\{\mathbf{S}, L(\mathbf{S}), \ldots, L^{n-1}(\mathbf{S})\}$*. Then we have* $\mathrm{Eper}(\mathbf{R}) = n$*.*

Proof. Since $\mathrm{per}(\mathbf{R}) = n$, we know $\mathbf{R} \neq (0, 0, \ldots, 0)$ and $\mathbf{R} \neq (1, 1, \ldots, 1)$. There are at least one 0 and at least one 1 in $\mathbf{R}$. Let t be the length of the longest run of 0s in $\mathbf{R}$. Then $\mathbf{R}$ is of the form $\mathbf{R} = (\overbrace{0, \ldots, 0}^{t}, 1, *, \ldots, *, 1)$. There may be more than one run of 0s of length t in $\mathbf{R}$. Let $\mathbf{R} = (r_0, r_1, \ldots, r_{n-1})$ where $r_0 = r_1 = \cdots = r_t = 0$ and $r_{n-1} = 1$. Suppose $\mathrm{Eper}(\mathbf{R}) < n$. By the definition of extended period, there exist some integer $0 < u < n$ such that, $(r_0, r_1, \ldots, r_{n-u-1}) = (r_u, r_{u+1}, \ldots, r_{n-1})$. Since the last element of $\mathbf{R}$ is 1, r_u is the beginning of some run of 0s of length t. We note that u is not a divisor of n, otherwise we would have $\mathrm{per}(\mathbf{R}) = u < n$. Therefore, we have $u \neq \frac{n}{2}$. If $u < \frac{n}{2}$, since u is not a divisor of n we can assume $n = qu + r$ such that $1 < r < u$. Define

$$v = \begin{cases} qu & \text{if } u < \frac{n}{2} \\ u & \text{otherwise} \end{cases}$$

It can be verified that, we always have 1) $\frac{n}{2} < v < n$; 2) r_v is the beginning of some run of 0s of length t; and 3) $(r_0, r_1, \ldots, r_{n-v-1}) = (r_v, r_{v+1}, \ldots, r_{n-1})$. Let m be the integer corresponding to $(r_0, r_1, \ldots, r_{n-v-1})$, i.e., $m = \sum_{i=0}^{n-v-1} r_i 2^{n-v-1-i}$, and m' be the integer corresponding to $(r_{n-v}, r_{n-v+1}, \ldots, r_{v-1})$. The following picture shows these numbers. The rectangles in the picture denote the longest run of 0s.

R:


Regarding states as integers, we have

$$\mathbf{R} = m + m' \cdot 2^{n-v} + m \cdot 2^v,$$
$$L^{n-v}(\mathbf{R}) = m + m \cdot 2^{n-v} + m' \cdot 2^{2n-2v},$$
$$L^{v}(\mathbf{R}) = m' + m \cdot 2^{2v-n} + m \cdot 2^v$$

Since all the three states belong to $\{\mathbf{S}, L(\mathbf{S}), \ldots, L^{n-1}(\mathbf{S})\}$ and $\mathbf{R}$ is the least state of this set, we have $L^{n-v}(\mathbf{R}) > \mathbf{R}$ and $L^v(\mathbf{R}) > \mathbf{R}$. However, by $(L^{n-v}(\mathbf{R}) - \mathbf{R})(L^v(\mathbf{R}) - \mathbf{R}) = [m(2^{n-v} - 2^v) + m'(2^{2n-2v} - 2^{n-v})][m(2^{2v-n} - 1) + m'(1 - 2^{n-v})] = -2^{n-v}[m(2^{2v-n} - 1) + m'(1 - 2^{n-v})]^2 < 0$, one of $L^{n-v}(\mathbf{R})$ and $L^v(\mathbf{R})$ is less than $\mathbf{R}$. This is a contradiction. □

Theorem 6. *Let* FSR(f) *be an n-stage FSR that contains only cycles of length no more than n, then $f = x_0 + x_n$.*

Proof. Let $\mathbf{S} = (s_0, s_1, \ldots, s_{n-1})$ be a state of length n. We call $\mathbf{S}$ a satisfied state if the next state of $\mathbf{S}$ in FSR(f) is $L(\mathbf{S}) = (s_1, \ldots, s_{n-1}, s_0)$. For the proof of this theorem we need to show that, all the states of length n are satisfied states. Our discussions are divided into two cases. Firstly, we consider the case per$(\mathbf{S}) = n$. Let $\mathbf{R}$ be the least state of the set $\{\mathbf{S}, L(\mathbf{S}), \ldots, L^{n-1}(\mathbf{S})\}$. According to Lemma 6, Eper$(\mathbf{R}) = n$. Consider the sequence $\mathbf{s}$ generated by FSR(f) with initial state $\mathbf{R}$. Since FSR(f) contains only cycles of length $\leq n$, we get per$(\mathbf{s}) \leq n$. By the definition of extended period, we know per$(\mathbf{s}) \geq$ Eper$(\mathbf{R}) = n$. Thus we have per$(\mathbf{s}) = n$. This implies $[\mathbf{S}, L(\mathbf{S}), \ldots, L^{n-1}(\mathbf{S})]$ is a cycle of FSR(f) and all the states on this cycle are satisfied states. Especially, $\mathbf{S}$ is a satisfied state. Then we consider the case per$(\mathbf{S}) < n$. According to Lemma 5, per$(\mathbf{S}) < n$ implies per$(\widehat{\mathbf{S}}) = n$. By the above discussion, $\widehat{\mathbf{S}}$ is a satisfied state. Therefore, the next state of $\widehat{\mathbf{S}}$ in FSR(f) is $(s_1, \ldots, s_{n-1}, \overline{s}_0)$. Since FSR$(f)$ is a nonsingular FSR, the two successors of a conjugate pair is a companion pair. Thus the next state of $\mathbf{S}$ in FSR(f) is $(s_1, \ldots, s_{n-1}, s_0)$ and $\mathbf{S}$ is a satisfied state. □

References

[1] A. H. Chan, R. A. Games, and E. L. Key, "On the complexities of De Bruijn sequences," J. Combin. Theory, ser. A, vol. 33, pp. 233-246, Nov. 1982.

[2] N. G. de Bruijn, “A combinatorial problem,” Proc. Kon. Ned. Akad. Wetensch, vol. 49, pp. 758-746, 1946.

[3] T. Etzion and A. Lempel, “Algorithms for the generation of full-length shift-register sequences,” IEEE Trans. Inf. Theory, vol. 30, no. 3, pp. 480-484, May. 1984.

[4] H. Fredricksen, “A class of nonlinear de Bruijn cycles,” J. Comb. Theory, Ser. A, vol. 19, no. 2, pp. 192-199, Sep. 1975.

[5] H. Fredricksen, “A survey of full length nonlinear shift register cycle algorithms,” SIAM Rev., vol.24, no. 2, pp. 195-221, Apr. 1982.

[6] S. W. Golomb, Shift Register Sequences, San Francisco, CA: Holden-Day, 1967.

[7] D. H. Green and K. R. Dimond, “Nonlinear product-feedback shift registers,” Proc. IEE, vol. 117, no. 4, pp. 681-686, Apr. 1970.

[8] C. J. A. Jansen, W. G. Franx and D. E. Boekee, “An efficient algorithm for the generation of deBruijn cycles,” IEEE Trans. Inf. Theory, vol. 37, no. 5, pp. 1475-1478, Sep. 1991.

[9] C. Y. Li, X. Y. Zeng, T. Helleseth, C. L. Li and L. Hu, “The properties of a class of linear FSRs and their applications to the construction of nonlinear FSRs,” IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 3052-3061, May. 2014.

[10] C. Y. Li, X. Y. Zeng, C. L. Li, and T. Helleseth, “A Class of De Bruijn Sequences,” IEEE Trans. Inf. Theory, vol. 60, no. 12, pp. 7955-7969, Dec. 2014.

[11] R. Lidl and H. Niederreiter, Finite Fields, in Encyclopedia of Mathematics and Its Applications, Reading, MA: Addison-Wesley, 1983, vol. 20.

[12] D. E. Muller, “Application of Boolean algebra to switching circuit design and to error correction,” IRE Trans. Electron. Comput., vol. EC-3, no. 9, pp. 6-12, Sep. 1954.

[13] J. Mykkeltveit, M. K. Siu and P. Tong, “On the cycle structure of some nonlinear shift register sequences,” Inf. Contr., vol. 43, no. 2, pp. 202-215, Nov. 1979.

[14] I. S. Reed, “A class of multiple-error-correcting codes and the decoding scheme,” IRE Trans. Inform. Theory, vol. IT-4, pp. 38-49, Sep. 1954.