

# Stretching de Bruijn sequences

Abbas Alhakim<sup>1</sup> · Maher Nouiehed<sup>2</sup>

Received: 6 May 2016 / Revised: 16 November 2016 / Accepted: 23 November 2016 /  
Published online: 1 December 2016  
© Springer Science+Business Media New York 2016

**Abstract** We give a one-step construction of de Bruijn sequences of general alphabet size and with order  $n + k$ , given a de Bruijn sequence of order  $n$  and any integer  $k > 1$ . This is achieved by using an appropriate class of graph homomorphisms between de Bruijn digraphs whose orders differ by an integer  $k$ . The method starts with a lower order de Bruijn cycle, finds its inverse cycles in the higher order digraph, which are then cross-joined into one full cycle. Therefore, this generalizes the Lempel’s binary construction and the Alhakim–Akinwande construction for non-binary alphabets and a wide class of homomorphisms.

**Keywords** de Bruijn sequence · de Bruijn graph homomorphism · Lempel’s D-morphism · Recursive construction · Linear feedback shift register

**Mathematics Subject Classification** 05C45 · 05C38 · 05C60 · 68R01

## 1 Introduction

Given an integer  $n \geq 1$ , a de Bruijn sequence of order  $n$  is a string of letters from a finite alphabet such that each possible string of  $n$  letters occurs exactly once in the sequence as a subsequence of contiguous letters. These combinatorial objects have found their way to many engineering applications such as communication, stream ciphers and random number

---

Communicated by L. Teirlinck.

---

✉ Abbas Alhakim  
aa145@aub.edu.lb

Maher Nouiehed  
nouiehed@usc.edu

<sup>1</sup> Department of Mathematics, American University of Beirut, Beirut, Lebanon

<sup>2</sup> Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, USA

generation. They are known in the engineering community as shift register sequences because they are implemented in special hardware called shift registers, see Golomb [6].

We will assume that the alphabet is  $\Sigma = \{0, \dots, q - 1\}$ , where  $q$  is taken as a power of prime. The reader will notice that for several results below,  $q$  need not be a power of prime, but we simply keep this assumption throughout the paper because the main construction is based on a linear recurrence sequence of order  $n$  that admits the maximal possible period of  $q^n - 1$  (a maximal length LFSR sequence). The existence of the latter requires  $q$  to be a power of prime.

An equivalent definition of de Bruijn sequences is to identify them with Hamiltonian cycles on the de Bruijn digraph  $B(n, q)$ . The vertex set of the de Bruijn digraph is the set of all possible strings of size  $n$ , and an edge from  $\mathbf{x} = x_1, \dots, x_n$  to  $\mathbf{y} = y_1, \dots, y_n$  is on the digraph if and only if  $y_i = x_{i+1}$  for all  $i = 1, \dots, n - 1$ .

Lempel [9], studies a graph homomorphism  $D : B(n + 1, 2) \rightarrow B(n, 2)$  of the form:

$$D(x_1, x_2, \dots, x_{n+1}) = (x_1 + x_2, x_2 + x_3, \dots, x_n + x_{n+1}),$$

widely known as Lempel's D-morphism. This homomorphism is used to invert a de Bruijn cycle in  $B(n, 2)$  into two distinct vertex disjoint cycles  $B(n + 1, 2)$  that between them cover all the vertices of  $B(n + 1, 2)$ . His well-known recursive construction is thus based on connecting these two inverse cycles into one de Bruijn cycle via a well-known cycle joining method, sometimes referred to as the "cross-join surgery", see Fredricksen [4], that requires one to identify a pair of vertices on the two cycles, whose successors on their respective cycles can be swapped to make one new full cycle of order  $n + 1$ . Alhakim and Akinwande [2] perform this construction in the non-binary case by introducing a class of translation invariant homomorphisms that invert a cycle in  $B(n, q)$  into  $q$  cycles in  $B(n + 1, q)$ . The recursive step is completed via a generalized cross-join surgery which stitches all the inverse cycles together at once. Mandal and Gong [10] describes the  $k$ th order D-morphic construction of de Bruijn sequences. That is, their method uses a homomorphism  $H : B(n + k, 2) \rightarrow B(n, 2)$  which is obtained by iterating the D-morphism  $k$  times. The idea of iterating the D-morphism was used earlier by Chang et al. [3] with a number of iterations that is a power of 2. This homomorphism of order  $k$  can be used to construct higher order de Bruijn sequences using lower order ones, but incurs several complications when the order  $k$  is not a power of 2. Mykkeltveit et al. [11] study algebraically the cycle structure of cascading two shift registers. This connection of shift registers can be regarded as inverting the output of one shift register sequence using the recursive relation of the other. In this paper we consider a class of homomorphisms from  $B(n + k, q)$  to  $B(n, q)$  and investigate in a graphical framework the cycle structure of the corresponding inverse images of cycles in  $B(n, q)$ .

When  $k = 1$ , and for an appropriate homomorphism of [2], each of the possible  $q$  inverse images of a cycle  $C$  in  $B(n, q)$  is initiated by a distinct digit, and then one new digit is appended for each digit in  $C$ . The translation invariance of the homomorphism guarantees that the last digit to be placed is identical to the initial digit, thus the inverse image is indeed a cycle. When  $k > 1$ , an inverse image of  $C$  is initiated by a word  $x_1, \dots, x_k$  and, likewise, a new digit is appended for each digit in  $C$ . If the last  $k$  digits of this inverse image are denoted by  $y_1, \dots, y_k$ , then the inverse image is a cycle exactly when  $x_1, \dots, x_k = y_1 \dots y_k$ . When these two words are not equal, which turns out to be the more likely possibility, one can start a new inverse image of  $C$  with the initial word  $y_1, \dots, y_k$ . This puts two inverse images together. One can iterate this juxtaposition of inverse images until eventually, after  $m$  iterations, where  $m < q^k$  depends on the homomorphism and on the initial word  $(x_1, \dots, x_k)$ , an inverse image ends exactly with the initial word  $x_1, \dots, x_k$ . This finishes an inverse cycle whose length is  $m$  times the length of  $C$ .

For a fixed cycle  $C$  and a fixed homomorphism, if we consider the map  $F$  defined as  $F(x_1, \dots, x_k) = y_1, \dots, y_k$ , where  $x_1, \dots, x_k$  and  $y_1, \dots, y_k$  are as above, then  $F$  is a permutation on  $\Sigma^k$ . Evidently, the number of inverse cycles of  $C$  is the number of cycles of this permutation.

One of the main contributions of this paper is to identify a class of linear homomorphisms that invert one de Bruijn cycle of  $B(n, q)$  into exactly two vertex disjoint cycles in  $B(n+k, q)$  that, between them, include all the vertices on the digraph. The other main contribution is a step-by-step construction of a de Bruijn sequence which of course involves a search-free identification of a cross-join pair of vertices that allows us to join the two cycles into one de Bruijn sequence in  $B(n+k, q)$ .

While the homomorphisms used in Chang et al. [3] and Mykkeltveit et al. [10] are  $k$ th order iterates of the D-morphism and consider only binary sequences, the current construction is based on a wider class of functions that are applicable for any alphabet whose size is a power of prime. Despite the fact that the construction in Alhakim and Akinwande [2] applies to non-binary alphabets, they are based on homomorphisms between two de Bruijn digraphs of consecutive orders. So their procedure produces a de Bruijn sequence of order  $n + 1$  using one of order  $n$ . If a higher order sequence is desired the procedure has to be iterated several times, à la Lempel. Indeed, the homomorphisms they present can be seen as a direct generalization of that of Lempel's. For instance, every inverse of a cycle is a cycle. The first step in the construction is the computation of  $q$  inverse cycles of a de Bruijn cycle, where  $q$  is any alphabet size (that need not be a power of prime). The second step therefore calls for a method to stitch all these cycles together. As mentioned in the previous paragraph, the current construction allows us to construct a sequence of order  $n + k$  directly—i.e., in one iteration—from a sequence of order  $n$ ; and thanks to the choice of homomorphism, only two cycles have to be connected.

In Sect. 2 we give preliminary definitions and results that are required throughout the paper. In particular, we state and prove Theorem 1 which fundamentally establishes that constructing de Bruijn sequences of order  $n + k$  from one of order  $n$  is possible.

## 2 Preliminaries

A linear feedback shift register (LFSR) sequence is based on a linear recurrence relation over the Galois field  $GF(2)$  of the form  $x_{k+1} = G(x_1, x_2, \dots, x_k) = a_1x_1 + a_2x_2 + \dots + a_kx_k$  where  $G$  is called the feedback function and the coefficients  $a_1, \dots, a_k$  are carefully chosen to produce a sequence with good statistical properties. More generally, a feedback function produces a purely cyclic sequence if and only if the function is one-to-one in  $x_1$ ; in the binary case,  $G(x_1, x_2, \dots, x_k) = x_1 + g(x_2, x_3, \dots, x_k)$ . For a general base, the recurrence relation

$$ax_i + g(x_{i+1}, \dots, x_{i+k-1}) + bx_{i+k} = 0; \quad a \neq 0, \quad b \neq 0 \tag{1}$$

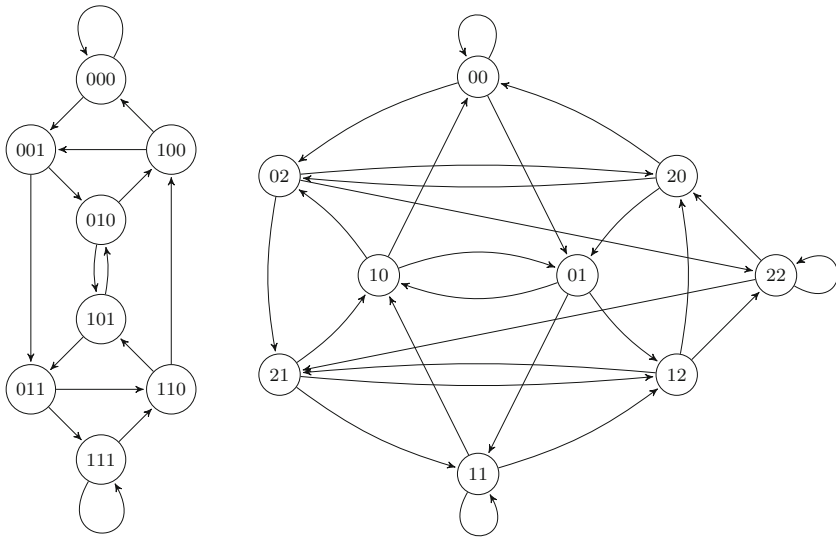
gives a purely cyclic sequence.

In this paper, we use such recurrence relations but with the addition of external terms, i.e., we study the recurrence relations of the form:

$$ax_i + g(x_{i+1}, \dots, x_{i+k-1}) + bx_{i+k} = b_i; \quad a \neq 0, \quad b \neq 0 \tag{2}$$

for some nonconstant sequence  $\{b_i\}$ .

For convenience, we will identify the alphabet  $\Sigma$  with the group  $Z_q$  of residues modulo  $q$ . For a de Bruijn digraph  $B(n, q)$  the vertex set is then  $V = Z_q^n$ . If  $(\mathbf{x}, \mathbf{y})$  is an edge from



**Fig. 1** de Bruijn digraphs  $B(3, 2)$  (left) and  $B(2, 3)$  (right)

the vertex  $\mathbf{x} = (x_1, \dots, x_n)$  to the vertex  $\mathbf{y} = (y_1, \dots, y_n)$ , we say  $\mathbf{x}$  is a predecessor of  $\mathbf{y}$ , and  $\mathbf{y}$  is a successor of  $\mathbf{x}$ . We note that each vertex has  $q$  successors and  $q$  predecessors. Two vertices  $yx_2, \dots, x_n$  and  $zx_2, \dots, x_n$  with  $y \neq z$  are said to be conjugate vertices, so that they have the same set of successors. As an example, a binary de Bruijn digraph and a ternary de Bruijn digraph are given in Fig. 1.

We will interchangeably use  $(x_1, \dots, x_n)$  as a vertex in the de Bruijn digraph  $B(n, q)$  and as a tuple in  $Z_q^n$ . We also denote by  $y^k = (y, \dots, y)$  a constant string of size  $k$ . A cycle in  $B(n, q)$  is a path that starts and ends at the same vertex. A path or a cycle is called *vertex disjoint* if it does not visit a vertex more than once. Two cycles or two paths in the digraph are vertex disjoint if they do not have a common vertex. A *factor* is a collection of vertex disjoint cycles that cover all the vertices of one digraph. A de Bruijn cycle is a Hamiltonian cycle on the de Bruijn digraph, i.e., a cycle that traverses each vertex exactly once. Thus, a de Bruijn cycle is trivially an instance of a factor. A de Bruijn sequence is identified with a de Bruijn cycle in an obvious way. For example, the cycle  $(00, 01, 11, 10, 00)$  of  $B(2, 2)$  is identified with  $00110$ . It is customary, see Lempel [9], to represent such a cycle as  $[0011]$ , with the first and last  $n - 1$  digits identified for a sequence of order  $n$ . This is referred to as the ring sequence representation of the cycle.

### 2.1 Homomorphisms with Property (D)

A graph homomorphism  $H$  from a digraph  $\mathcal{G}_1$  to a digraph  $\mathcal{G}_2$  is a function from the vertex set of  $\mathcal{G}_1$  to the vertex set of  $\mathcal{G}_2$  such that whenever  $(v, w)$  is an edge in  $\mathcal{G}_1$ , then  $(Hv, Hw)$  is an edge of  $\mathcal{G}_2$ . If  $C = (v_1, \dots, v_l)$  is a path in  $\mathcal{G}_1$ , we define the image of  $C$  by  $H$  as the path  $(Hv_1, \dots, Hv_l)$ .

In Akinwande [1], Proposition 3.2.1 states that a graph homomorphism  $H$  from  $B(n+k, q)$  to  $B(n, q)$  is characterized by the form

$$H(x_1, \dots, x_{n+k}) = (h(x_1, \dots, x_{k+1}), \dots, h(x_n, \dots, x_{n+k})). \tag{3}$$

for an arbitrary function  $h$  from  $Z_q^{k+1}$  to  $Z_q$ .

The following definition, also given in [1], is meant to characterize homomorphisms that can lift a factor in  $B(n, q)$  to a factor in  $B(n+k, q)$ , thus generalizing Lempel’s D-morphism, hence the name.

**Definition 1** A homomorphism  $H$  is said to have Property (D) if each vertex disjoint path in  $B(n, q)$  is the image of  $q^k$  non-overlapping vertex disjoint paths in  $B(n+k, q)$ .

Note that this implies that each vertex in  $B(n, q)$  has  $q^k$  inverse images in  $B(n+k, q)$ , yet the converse is not true. To see this consider, for example, the homomorphism defined by letting  $h(x_1, \dots, x_{k+1}) = x_1$  in Eq. (3).

**Lemma 1** Given a word  $x_1, \dots, x_k \in Z_q^k$ , an element  $b \in Z_q$ , and a homomorphism  $H$  with Property (D), there exist uniquely two elements  $a$  and  $c$  in  $Z_q$  such that

$$h(a, x_1, \dots, x_k) = h(x_1, \dots, x_k, c) = b, \text{ where } h \text{ is the function associated with } H \text{ in Eq. (3).}$$

*Proof* This lemma follows immediately from Theorem (3.2.4) in [1], which states that a homomorphism  $H$  enjoys Property (D) if and only if  $h(x_1, \dots, x_{k+1})$  is one-to-one in each of the variables  $x_1$  and  $x_{k+1}$ , when it is considered as a single variable function with all other variables kept fixed. □

Each vertex in  $B(n, q)$ , which can be considered a path, has  $q^k$  inverse images. Thus, the total number of vertices in the pre-images of an order  $n$  de Bruijn cycle is  $q^k \times q^n = q^{k+n}$ . Theorem 1 below establishes that all of these vertices are distinct, hence they are all the vertices of  $B(n+k, q)$ . Given a factor  $\mathcal{F}$  in  $B(n, q)$ , the collection of inverse images of all cycles in  $\mathcal{F}$  will be referred to as the inverse of  $\mathcal{F}$ .

**Theorem 1** Let  $H$  be a homomorphism with property (D) from  $B(n+k, q)$  to  $B(n, q)$ . The inverse image by  $H$  of a factor in  $B(n, q)$  is a factor in  $B(n+k, q)$ .

*Proof* Since two inverse images of two vertex disjoint cycles of  $B(n, q)$  cannot overlap, it is sufficient to show that all inverse images of a vertex disjoint cycle  $C$  have no vertex in common. Suppose that  $C_1, \dots, C_{q^k}$  are the inverse images of  $C$ , each being initialized by a distinct  $k$ -tuple. We note that the length of each  $C_i$  is  $k + |C|$  where  $|C|$  is the number of digits in the ring sequence of  $C$ . Let  $C_i(j)$ ;  $j = 1, \dots, k + |C|$  be the  $j$ th digit of  $C_i$ . Suppose first that  $(C_{i_1}(j_1), \dots, C_{i_1}(j_1 + n + k - 1)) = (C_{i_2}(j_2), \dots, C_{i_2}(j_2 + n + k - 1))$  for some  $i_1, i_2, j_1, j_2$ . If  $j_1 \neq j_2$ , their images by  $H$  would be equal but they would fall in two different locations in  $C$ , which cannot happen because  $C$  is vertex disjoint.

Suppose then that  $(C_{i_1}(j), \dots, C_{i_1}(j + n + k - 1)) = (C_{i_2}(j), \dots, C_{i_2}(j + n + k - 1))$  for some  $i_1 \neq i_2$ . It follows in particular that  $(C_{i_1}(j), \dots, C_{i_1}(j+k-1)) = (C_{i_2}(j), \dots, C_{i_2}(j+k-1))$ . Since all inverse paths have different seeds, it is evident that  $j > 1$ . Lemma 1 implies that  $C_{i_1}(j-1) = C_{i_2}(j-1)$ . Iterating this several times shows that  $(C_{i_1}(1), \dots, C_{i_1}(k)) = (C_{i_2}(1), \dots, C_{i_2}(k))$ , which contradicts the fact that all paths have different seeds. This establishes the theorem. □

Another proof that avoids the technical details was contributed by a referee and it goes as follows. A nonsingular factor may be identified with a subgraph of the de Bruijn digraph which contains all the vertices but only a subset of the edges, and with the additional property that exactly one edge enters each vertex and exactly one edge leaves each vertex. Consider the subgraph of  $B(n+k, q)$  defined as the inverse image by  $H$  of a factor in  $B(n, q)$ . When

$H$  has Property (D), this subgraph is also nonsingular, for two or more edges cannot enter a vertex as that would contradict Property (D). Also, there must be an edge that enters each vertex otherwise  $H$  would not be a homomorphism. That is, this inverse image is a factor.

It follows from the theorem that the  $q^k$  pre-images of a de Bruijn cycle form a factor with an arbitrary number of cycles, depending on the homomorphism. When  $k = 1$ , for example, using any of the homomorphisms provided in Alhakim and Akinwande [2], each of the  $q$  inverse images is itself a cycle. If also  $q = 2$ , it reduces to the case of Lempel’s homomorphism, which gives two inverse cycles.

In this paper, considering the left hand side of either Eq. (1) or (2) as a function  $h$  defined on  $Z_q^{k+1}$ , we will regard the sequence generated by Eq. (1) as one whose image by  $h$  is constantly zero, while the sequence generated by Eq. (2) as an inverse image of the given sequence  $\{b_i\}$ .

When the feedback function  $G$  is linear with maximal period,  $h$  corresponds to the well-known LFSR recurrence. We will refer to this as a maximal period LFSR homomorphism. If the inverted sequence in  $B(n, q)$  is a de Bruijn sequence, the inverse images form a factor in  $B_{n+k}(q)$ , by Theorem 1. We then formulate the problem of connecting the inverse paths as a non-homogeneous version of the well-known matrix method of generating pseudo-random vector generators [12], using a homomorphism which connects all but one inverse path into one long cycle. We also show how the last path forms a cycle which will be efficiently joined to construct a higher order de Bruijn sequence. This is done in Sect. 3, while in Sect. 5, we give a concise algorithm that summarizes all the steps needed to efficiently generate the inverse cycles and connect them into one full cycle. We conclude the paper in Sect. 6 with an overview of the results, giving an eye on potential applications and further development.

### 3 Inverse cycles

In this section, the construction of the inverse images of a vertex disjoint cycle in  $B_n(q)$  via a homomorphism with Property (D) is given. Let  $h : Z_q^{k+1} \rightarrow Z_q$  be of the form:

$$h(x_1, \dots, x_{k+1}) = ax_1 + g(x_2, x_3, \dots, x_k) + x_{k+1}; \quad a \neq 0 \tag{4}$$

Using  $h$ , define a homomorphism  $H : B_{n+k}(q) \rightarrow B_n(q)$  as in Eq. (3), so that  $H$  obviously enjoys Property (D). In the next subsections we will investigate the homomorphisms that result from LFSR recurrence, and in particular show that the class of maximal period LFSR homomorphisms provide a two-cycle structure in the higher order digraph that allows for an efficient construction of de Bruijn cycles.

#### 3.1 Constructing the inverse images

In this subsection, we will describe the recursive construction of the inverse paths of a vertex disjoint cycle in  $B(n, q)$ . Let  $\mathbf{b} = (b_1, b_2, \dots, b_p)$  be an arbitrary vertex disjoint cycle of period  $p$  in  $B(n, q)$  and for each  $i = 0, \dots, q^k - 1$  let  $(s_{i,1}, s_{i,2}, \dots, s_{i,k})$  be the  $q$ -ary representation of  $i$  with zeroes on the left as necessary. The inverse path that starts with the latter string will be called the  $i$ th inverse path. Applying  $h$ , each digit  $b_j$  in  $\mathbf{b}$  is the image of a string of  $k + 1$  digits. This is formulated by the equation:

$$s_{i,j+k} = b_j - as_{i,j} - g(s_{i,j+1}, \dots, s_{i,j+k-1}); \quad a \neq 0 \tag{5}$$

Using this equation, we can recursively construct the  $q^k$  inverse images. At each step, we use the previous  $k$  terms and the corresponding  $b_j$  to calculate the next term. We let  $\mathbf{S}$  be

a matrix holding the  $q^k$  inverse paths, each on a separate row, we now define it recursively. For  $l \geq 1$ , denote by  $S_i[j : j + l - 1]$  a string of  $l$  contiguous symbols in the  $i$ th row that starts with  $s_{i,j}$ . Using Theorem 1 and Property (D) of the function  $h$ , we see that the vectors  $S_i[j : j + n + k - 1]$  are distinct for all  $i, j$  and appear only once each. The rows of the matrix  $\mathbf{S}$  are initialized with distinct seeds  $S_i[1 : k]$ , for  $i = 0, \dots, q^k - 1$ . Iterating Eq. (5),  $\mathbf{S}$  ends up being of size  $q^k \times (p + k)$ , with the  $j$ th column in  $\mathbf{S}$  corresponding to the digit  $b_{j-k}$  of  $\mathbf{b}$ , for  $j = k + 1, \dots, p + k$ .

### 3.2 Connecting the inverse images into cycles

A step by step procedure that connects these paths into cycles in the higher order digraph is now described. We noted earlier that in both [2] and [9], each inverse path is itself a cycle. Moreover, these cycles are translates of one another, a fact that helped in finding a way to join the paths. Theorem 2 below shows a similar but more convolved relationship between the inverse paths when a general LFSR recurrence is used for the homomorphism. Theorem 3 in the next subsection shows that the class of maximum period LFSR homomorphisms, generates two inverse cycles: a long cycle that includes  $q^k - 1$  paths, and a short cycle containing the remaining path. For an LFSR, Eq. (4) takes the form:

$$h(x_1, \dots, x_{k+1}) = x_{k+1} - a_k x_k - \dots - a_1 x_1; \quad a_1 \neq 0, \tag{6}$$

so that the characteristic polynomial of the corresponding pure LFSR recurrence (when  $h$  is constantly set to zero) is

$$f(x) = x^k - a_k x^{k-1} - a_{k-1} x^{k-2} - \dots - a_1. \tag{7}$$

In turn, Eq. (5) becomes:

$$s_{i,j+k} = b_j + a_1 s_{i,j} + a_2 s_{i,j+1} + \dots + a_k s_{i,j+k-1} \quad a_1 \neq 0. \tag{8}$$

The next main result describes the nature of the pre-images of  $\{b_j\}$  as well as an efficient way to generate each pre-image. That is, we can use it to jump and generate any term in an inverse path using only the seeds of this path. First, let  $\mathbf{u} = (u_1, u_2, \dots, u_k) = (0, 0, \dots, 0, 1)$  be a string of size  $k$ . We define the following recurrence relation:

$$u_j = a_1 u_{j-k} + a_2 u_{j-k+1} + \dots + a_k u_{j-1}; \tag{9}$$

i.e., the generated sequence  $\{u_j\}$  is the LFSR sequence starting with  $0^{k-1}1$ . Recall that we symbolically represent the  $i$ th inverse image as the  $i$ th row  $S_i$  of the matrix  $S$ , which is initiated by the vector  $S_i[1 : k]$ .

**Theorem 2** *The inverse images of the sequence  $\{b_j\}$  satisfy the following:*

- (i) *The pre-image  $\{r_j\} := S_0$  of  $\{b_j\}$  that starts with  $0^k$  is a convolution of  $\{b_j\}$  with the sequence  $\{u_j\}$  given by (9). That is,  $r_j = \sum_{l=1}^{j-1} b_l u_{j-l}$ , for all  $j \geq 1$  (with an empty sum interpreted as zero).*
- (ii) *If  $S_i[1 : k] \neq 0^k$ , The pre-image  $S_i$  is a componentwise sum of  $\{r_j\}$  and  $\{s_j\}$ , where the latter is the ordinary LFSR sequence defined by the seed  $S_i[1 : k]$  and the recurrence (9).*

*Proof* To prove (i), we need to check that  $\{r_j\}$  defined by the convolution equation satisfies the recurrence relation (8) with the initial conditions  $r_1 = \dots = r_k = 0$ . By direct inspection, these initial conditions follow from the choice of the initial vector  $\mathbf{u}$  of  $\{u_j\}$ . Also in view of

the initial vector  $\mathbf{u}$ , we see that for all  $j \geq 1$ ,  $r_{k+j} = b_j + \sum_{l=1}^{j-1} b_l u_{k+j-l}$ . Using Eq. (9) for each  $u_{k+j-l}$  and then interchanging the order of summations we have

$$\begin{aligned} r_{k+j} &= b_j + \sum_{l=1}^{j-1} b_l \sum_{v=1}^k a_v u_{j+v-l-1} \\ &= b_j + \sum_{v=1}^k a_v \sum_{l=1}^{j-1} b_l u_{j+v-l-1} \\ &= b_j + \sum_{v=1}^k a_v \sum_{l=1}^{j+v-1} b_l u_{j+v-l-1} \\ &= b_j + \sum_{v=1}^k a_v r_{j+v-1}, \end{aligned}$$

where the third equality follows by noting that  $u_{j+v-1-l} = 0$  whenever  $j \leq l \leq j + v - 1$ .

For part (ii), we will consider the companion matrix of the linear recurrence that corresponds to the characteristic polynomial (7). Namely,

$$C = \begin{pmatrix} 0 & 0 & \dots & 0 & a_1 \\ 1 & 0 & \dots & 0 & a_2 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & a_{k-1} \\ 0 & \dots & 0 & 1 & a_k \end{pmatrix}.$$

We will also use the shorter notation  $\mathbf{s}_{i,j}$ ,  $\mathbf{r}_j$ ,  $\tilde{\mathbf{b}}_j$  for the  $k$ -dimensional row vectors  $S_i[j : k + j - 1]$ ,  $(r_j, \dots, r_{k+j-1})$ , and  $[0, \dots, 0, b_j]$  respectively, and we let  $\mathbf{a}$  be the last column of  $C$ . We will now prove inductively that, for all  $j \geq 1$ ,  $\mathbf{s}_{i,j} = \mathbf{s}_{i,1} C^{j-1} + \mathbf{r}_j$ .

The base case,  $j = 1$ , follows trivially because  $\mathbf{r}_1$  is the zero vector and  $C^0$  is the identity matrix. Assume then that for a certain  $j \geq 1$  the equation is true.

Using vector notation, Eq. (8) translates to  $s_{i,k+j} = \mathbf{s}_{i,j} \mathbf{a} + b_j$ . It follows by the inductive hypothesis that

$$\begin{aligned} s_{i,k+j} &= \left( \mathbf{s}_{i,1} C^{j-1} + \mathbf{r}_j \right) \mathbf{a} + b_j \\ &= \mathbf{s}_{i,1} C^{j-1} \mathbf{a} + \left( \mathbf{r}_j \mathbf{a} + b_j \right) \\ &= \mathbf{s}_{i,1} C^{j-1} \mathbf{a} + r_{k+j}, \end{aligned}$$

where the last line follows because  $\{r_j\}$  satisfies the recurrence (8). Since  $s_{i,k+j}$  is the  $k$ th component of  $\mathbf{s}_{i,j+1}$ , this completes the inductive argument. Also, since  $\mathbf{s}_{i,1} C^{j-1} \mathbf{a}$  is the  $(k + j)$ th term of the pure LFSR sequence that starts with  $S_i[1 : k]$ , this completes the proof of the theorem. □

It is worth mentioning that Theorem 2 is a direct generalization of Lempel’s binary results where the LFSR recurrence is reduced to  $h(x_1, x_2) = 0$ , and  $h$  is Lempel’s homomorphism. When starting with 0 and 1 respectively, this recurrence gives the constant sequences  $0, 0, \dots, 0$  and  $1, 1, \dots, 1$ . Because  $k = 1$ , the sequence  $u_j$  is initialized by  $u_1 = 1$  and thus is also constantly 1. It follows by the first part of Theorem 2 that  $r_j = \sum_{l=1}^j b_l$ . This sum was used in Lempel [9] as an efficient way to get the inverse image that starts with 0. The

other inverse image is obtained by interchanging 1s and 0s. That is, by adding the constant sequence of 1s.

For an inverse image of  $\{b_i\}$  that starts with  $(x_1, \dots, x_k)$  denote by  $(y_1, \dots, y_k)$  the last  $k$  terms of this inverse path. We define  $F : Z_q^k \rightarrow Z_q^k$  to be the function that maps  $(x_1, \dots, x_k)$  to  $(y_1, \dots, y_k)$ . From the definition of Property (D), it is easy to see that  $F$  is bijective. When using the homomorphisms provided in [2] and [9],  $F$  is the identity function on the set of digits  $0, 1, \dots, q - 1$ . The following corollary shows that the current case is quite different. Its proof follows immediately from that of Theorem 2, yet it is of prime importance for the theme of this paper.

**Corollary 1** *The function  $F$  is an affine transformation. That is, for some matrix  $U$  and constant vector  $c$  that does not depend on the input vector:*

$$F(s_{i,1}, s_{i,2}, \dots, s_{i,k}) = (s_{i,1}, s_{i,2}, \dots, s_{i,k}) U + c, \tag{10}$$

where  $U = C^p$  and  $c$  is defined as  $F(0, \dots, 0)$ .

### 4 Non-homogeneous matrix method

Using an arbitrary LFSR homomorphism, the inverse paths are grouped into cycles of generally different periods. We show here that  $q^k - 1$  inverse paths belong to a single cycle when a maximum period LFSR homomorphism is used. In effect, if  $\mathbf{z} = (s_{i,1}, \dots, s_{i,k})$  is the row vector given by the initial segment of the  $i$ th row of  $\mathbf{S}$ , then the terminal segment  $\mathbf{z}' = F(\mathbf{z})$  is indeed the initial segment  $(s_{i',1}, \dots, s_{i',k})$  for some  $i'$ . If  $i = i'$ , then the  $i^{th}$  inverse path is a cycle, otherwise repeating Eq. (10) appends two inverse paths together. This can be iterated until  $\mathbf{z}$  is encountered again. Thus, joining inverse paths can be formulated as a non-homogeneous version of the well-known matrix method of generating pseudo-random vectors, which cannot be found in the literature (based on a personal communication between one of the authors with Niederreiter, the author of [12]). It is stated as follows:

$$\mathbf{z}_{i+1} = \mathbf{z}_i U + \mathbf{c}. \tag{11}$$

The matrix method explained in [12] considers the homogeneous case and therefore assumes  $\mathbf{c}$  to be the zero vector. Thus, it is clear that the zero vector is a fixed vector that forms a separate cycle. However, in the non-homogeneous problem a fixed vector  $\mathbf{z}_f$  must satisfy  $\mathbf{z}_f = \mathbf{z}_f U + \mathbf{c}$ . In the theorem below, we show the necessary conditions needed to maximize the period of the output sequence of Eq. (11).

**Lemma 2** *If the characteristic polynomial of the matrix  $U$  is primitive over  $F_q$ , then  $(\mathbf{I} - U^i)$  is non-singular for any  $i = 1, \dots, q^k - 2$ .*

*Proof* Let  $\lambda_j, j = 1, \dots, q^k$  be the eigenvalues of  $U$ . We note that if  $(\mathbf{I} - U^i)$  is singular, then there exists some  $j$  with  $\lambda_j^i = 1$ . Since the characteristic polynomial of  $U$  is primitive then the roots  $(\lambda_j)$  of this equation are primitive elements of the field. Thus,  $\lambda_j^i = 1$  if and only if  $i = q^k - 1$  and  $(\mathbf{I} - U^i)$  is singular if and only if  $i = q^k - 1$  (Note that we are only considering  $i \leq q^k$ ). □

**Theorem 3** *Suppose  $U$  is an invertible matrix. The following two properties are equivalent:*

1. *The period of  $\mathbf{z}_n$ ,  $Per(\mathbf{z}_n)$ , is  $q^k - 1$  for all seeds  $\mathbf{z}_0 \neq \mathbf{z}_f$  in  $F_q$ .*
2. *The characteristic polynomial of  $U$  is primitive over  $F_q$ .*

*Proof* Suppose the characteristic polynomial of  $\mathbf{U}$  is a primitive. By the previous lemma,  $\mathbf{I} - \mathbf{U}$  is invertible and so there exists a unique fixed point given by

$$\mathbf{z}_f = \mathbf{c}(\mathbf{I} - \mathbf{U})^{-1} \tag{12}$$

By iterating Eq. (11), we get the following:

$$\mathbf{z}_{n+i} = \mathbf{z}_n \mathbf{U}^i + \mathbf{c}(\mathbf{I} - \mathbf{U})^{-1} \times (\mathbf{I} - \mathbf{U}^i) \tag{13}$$

The period of this recursive function is equal to the smallest integer  $i$  such that  $\mathbf{z}_{n+i} = \mathbf{z}_n$ . It follows that for an arbitrary vector  $\mathbf{z}_n \neq \mathbf{z}_f \in \mathbb{Z}_q^n$ , the period  $i$  of the sequence started with  $\mathbf{z}_n$  is at most  $q^k - 1$ . By Eq. (13), this period must satisfy the equation  $[\mathbf{z}_n - \mathbf{c}(\mathbf{I} - \mathbf{U})^{-1}] (\mathbf{I} - \mathbf{U}^i) = \mathbf{0}$ .

Since  $\mathbf{z}_n \neq \mathbf{z}_f$ ,  $\mathbf{z}_n - \mathbf{c}(\mathbf{I} - \mathbf{U})^{-1} \neq \mathbf{0}$ . Therefore,  $[\mathbf{z}_n - \mathbf{c}(\mathbf{I} - \mathbf{U})^{-1}] (\mathbf{I} - \mathbf{U}^i) = \mathbf{0}$  if and only if:

1.  $(\mathbf{I} - \mathbf{U}^i)$  is singular and  $[\mathbf{z}_n - \mathbf{c}(\mathbf{I} - \mathbf{U})^{-1}]$  is a left eigenvector, or
2.  $\mathbf{U}^i = \mathbf{I}$

However, by Lemma 2, for all  $i = 1, \dots, q^k - 2$ ,  $(\mathbf{I} - \mathbf{U}^i)$  is a non-singular matrix if and only if the characteristic polynomial of  $\mathbf{U}$  is primitive. □

### 5 De Bruijn sequence construction

In this section we summarize the steps of the previous section into an algorithm that outputs an order  $(n + k)$  de Bruijn sequence using one of order  $n$ .

---

#### Algorithm

Given:  $B = [b_1, b_2, \dots, b_{q^n}]$  is a de Bruijn sequence of order  $n$  written in ring sequence form with  $(b_1, \dots, b_n) = 0^n$ , and  $H$  is a maximal period LFSR homomorphism.

---

- (1) Construct  $S_0[1 : q^n + k]$  as an inverse by  $H$  of  $B$  that starts with  $S_0[1 : k] = 0^k$
- (2) For  $i = 1, \dots, q^k - 1$  do lines (3)–(5)
- (3) let  $S_i[1 : k]$  be the  $q$ -ary representation of  $i$  (with zeroes on the left as needed)
- (4)  $(d_1, \dots, d_k) \leftarrow S_i[1 : k]$
- (5) For  $j = k + 1$  to  $q^n + k$  do the following
  - (I)  $D = a_1 d_1 + \dots + a_k d_k$
  - (II)  $S_i[j] \leftarrow S_0[j] + D$
  - (III)  $(d_1, \dots, d_k) \leftarrow (d_2, \dots, d_k, D)$
- (6)  $\mathbf{c} \leftarrow S_0[q^n + 1 : q^n + k]$
- (7)  $(z_1, \dots, z_k) \leftarrow \mathbf{c}(\mathbf{I} - \mathbf{U})^{-1}$
- (8)  $f \leftarrow$  the decimal representation of  $(z_1, \dots, z_k)_q$  {regarded as an integer represented in base  $q$ }
- (9) Initialize  $dB$  to  $S_f[1 : n + k]$
- (10) Let  $m$  be the decimal representation of  $(S_f[2], \dots, S_f[k + 1])_q$
- (11) Append  $S_m[n + k : q^n + k]$  to  $dB$
- (12) Let  $row$  be the decimal representation of  $(S_m[q^n + 1], \dots, S_m[q^n + k])_q$
- (13) Repeat the following two steps  $q^k - 2$  times
  - Append  $S_{row}[k + 1 : q^n + k]$  to  $dB$
  - Update  $row$  to be the decimal representation of  $(S_{row}[q^n + 1], \dots, S_{row}[q^n + k])_q$
- (14) Append  $S_m[k + 1 : n + k - 1]$  to  $dB$
- (15) Append  $S_f[n + k + 1 : q^n]$  to  $dB$
- (16) Output  $dB$

We will next prove the correctness of the above algorithm, but we first need some definitions and lemmas. For ease of reference we will denote by  $B_j = (b_j, \dots, b_{j+n-1})$  the vertex in  $B$  starting at  $b_j$ . We note that  $H$  and  $h$  are as given in (3) and (6) respectively, and  $h$  has a primitive characteristic polynomial, corresponding to a maximum period LFSR of order  $k$ .

**Definition 2** For any  $j = 1, \dots, q^n$ , let  $S^j = \{S_i[j : j + n + k - 1]; i = 0, \dots, q^k - 1\}$  be the set of vertices that are mapped to  $B_j$  by  $H$ .

**Lemma 3** Let  $B_{j_1}$  and  $B_{j_2}$  be any two conjugate vertices in  $B$ . Then for each  $i = 0, \dots, q^k - 1$ , the vertex  $S_i[j_1 : j_1 + n + k - 1]$  has exactly one conjugate in  $S^{j_2}$ .

*Proof* Consider a vertex  $S_{i_1}[j_1 : j_1 + n + k - 1]$  in  $S^{j_1}$ . The  $q^k$  words  $S_i[j_2 + 1 : j_2 + k]$ ,  $i = 0, \dots, q^k - 1$  are all distinct, because the initial seeds of the matrix  $S$  are distinct, and by appealing to Lemma 1  $j_2$  times. Hence there is a unique  $i_2$  such that  $S_{i_2}[j_2 + 1 : j_2 + k] = S_{i_1}[j_1 + 1 : j_1 + k]$ . Also by the alignment property of  $S$ ,  $h(S_{i_1}[j_1 : j_1 + k]) = b_{j_1}$  and  $h(S_{i_2}[j_2 : j_2 + k]) = b_{j_2}$ . But  $B_{j_1}$  and  $B_{j_2}$  being conjugate vertices in  $\mathbf{b}$  implies that  $b_{j_1+j} = b_{j_2+j}$  for  $j = 1, \dots, n - 1$ . Since  $h$  is bijective in the rightmost variable, it follows that  $S_{i_1}[j_1 + 1 : j_1 + n + k - 1] = S_{i_2}[j_2 + 1 : j_2 + n + k - 1]$ . Therefore,  $S_{i_2}[j_2 : j_2 + n + k - 1]$  is the unique conjugate of  $S_{i_1}[j_1 : j_1 + n + k - 1]$  in  $S^{j_2}$ .  $\square$

The following lemma follows directly from the fact that the vector  $\mathbf{c}$  in Eq. (11) is the terminal string of size  $k$  in  $S_0$  and by Eq. (12).

**Lemma 4** The inverse path that starts with  $0^k$  cannot be the fixed cycle.

**Lemma 5** The first and last vertices of the fixed cycle,  $S_f[1 : n + k]$  and  $S_f[q^n : q^n + n + k - 1]$ , are not conjugate vertices, (where  $S_f[q^n + 1 : q^n + n + k - 1] = S_f[1 : n + k - 1]$ ).

*Proof* Suppose that  $S_f[1 : n + k]$  and  $S_f[q^n : q^n + k - 1]$  are conjugate. Since  $S_f$  is a cycle,  $S_f[1 : n + k]$  is the successor of  $S_f[q^n : q^n + k - 1]$ . However, a vertex can be a successor of its conjugate if and only if it is of the form  $y^{n+k}$  for some element  $y$ . Consequently the fixed vertex  $\mathbf{z}_f$  of Eq. (11) is  $y^k$ . Because  $\mathbf{c} \neq \mathbf{0}$  in that equation,  $y \neq 0$  either. Since  $B_1 = 0^n$ , the term  $s_{f,k+1}$  is a pure LFSR term of  $h$ . Thus, by Eq. (6),  $y - a_k y - \dots - a_1 y = 0$ , so that  $1 - a_1 - \dots - a_k = 0$ , and the characteristic polynomial of  $h$ , given in Eq. (7), is then divisible by  $x - 1$ . This contradicts the fact that  $f(x)$  is primitive.  $\square$

*Proof (of Correctness of the Algorithm)* Step (1) builds the pre-image that starts with  $0^k$ . Lines (3)–(5) construct the inverse image  $S_i$  for each  $i \neq 0$  in a way that the initial  $k$  seeds are the digits of the  $q$ -ary representation of  $i$ , say  $(i)_q$ . Each inverse image is constructed by the for loop in line (5), using part (ii) of Theorem 2. In particular, line (5) (I) produces the next digit of the pure LFSR cycle with the recurrence given by Eq. (9) but with the initial condition given in line (4) as  $S_i[1 : k]$ . It follows by Theorem 2 that line (5) (II) gives the pre-image that starts with  $(i)_q$ .

Since  $S_0[1 : k] = 0^n$ , Eq. (11) asserts that the vector  $\mathbf{c}$  is as given by line (6). Hence, line (7) gives the initial seed of the fixed cycle.

Evidently, by the choice of  $H$ , the inverse paths are grouped into two cycles one of which is  $S_f$ . The other cycle can be started by any initial seed other than  $S_f[1 : k]$ , in particular  $S_f[2 : k + 1]$  initiates the other, longer cycle with  $(q^k - 1)q^n$  vertices. Since  $B$  is trivially a factor of  $B(n, q)$ , Theorem (1) asserts that all vertices of  $B(n + k, q)$  are on these two cycles. We now verify that lines (8)–(15) cross-joins them into a de Bruijn cycle.

By the definition of  $B$ ,  $B_1 = 0^n$ , this vertex is the successor of  $B_{q^n} = (b_{q^n}, 0, \dots, 0)$ ; which is the last vertex in  $B$ . Since  $b_{q^n}$  is necessarily different from zero, these two vertices

form a conjugate pair in  $B$ . By Lemma 3, there exists a unique vertex in  $S^q$  that forms a conjugate pair with  $S_f[1 : n + k]$ . By Lemma 5, this conjugate vertex, which we will denote by  $\mathbf{v}_{last}$ , is not on the fixed cycle. Since  $\mathbf{v}_{last}$  is the last vertex on an inverse path, it must be followed by the first vertex of some other inverse path. It follows that there is a unique vertex,  $\mathbf{v}_{first}$ , in  $S^1$  that is also a successor of  $S_f[1 : n + k]$ . It is then evident that the inverse path that starts with  $\mathbf{v}_{first}$  has  $S_f[2 : k + 1]$  as seed. This justifies the choice of  $m$  in line (10). That is,  $S_m[1 : n + k]$  is a successor of  $S_f[1 : n + k]$ . This is why line (11) connects the digits beginning with  $(n + k)$ th location.

We have thus located a cross join pair for the short cycle (fixed cycle) and the long cycle which starts with  $\mathbf{v}_{first} = S_m[1 : n + k]$  and ends with some appropriate  $\mathbf{v}_{last}$  that is conjugate to  $S_f[1 : n + k]$ . Since  $H$  the homomorphism is based on a linear recurrence relation with maximum period, its characteristic polynomial is primitive and Theorem 3 guarantees that  $q^k - 1$  paths can be connected by the iterations in lines (12) and (13). Moreover, for each of the other  $q^k - 2$  rows note that the first  $k$  digits are a repetition of the last  $k$  digits on the previous path so they must not be appended.

Line (14) cycles back to the inverse path  $S_m$  but only puts the terms that were left out in line (11). At this point the  $n + k$  most recent digits form the vector  $\mathbf{v}_{last}$ , which is conjugate with  $S_f[1 : n + k]$ . Hence line (15) rounds up the cross-join operation by putting the rest of  $S_f$ . □

*Example 1* Consider the primitive polynomial  $x^3 + 2x^2 + 1$  over  $GF(3)$ , which corresponds to the recurrence equation  $x_n = x_{n-1} - x_{n-3}$ . It therefore induces the homomorphism  $h(x_1, x_2, x_3, x_4) = x_1 - x_3 + x_4$ . We will use this homomorphism to invert the sequence  $\mathbf{b} = [002212011]$ , which is a tertiary de Bruijn sequence of order 2. That is, we will find the 27 inverse paths in  $B(5, 3)$  using the corresponding inhomogeneous recurrence equation  $x_n = x_{n-1} - x_{n-3} + b_{n-3}$ . The inverse path that starts with  $0^3$  is 000002122102, ending with 102 which gives the row vector  $\mathbf{c} = (1, 0, 2)$ . The companion matrix is a  $3 \times 3$  matrix with  $a_1 = -1, a_2 = 0$  and  $a_3 = 1$ . Once  $\mathbf{U} = \mathbf{C}^9$  is calculated, one finds the matrices  $(\mathbf{I} - \mathbf{U})^{-1}$ . These are

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \mathbf{U} = \mathbf{C}^9 = \begin{pmatrix} 2 & 1 & -1 \\ 2 & 2 & 1 \\ -1 & 1 & 0 \end{pmatrix}, (\mathbf{I} - \mathbf{U})^{-1} = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix},$$

The unique fixed cycle therefore starts with the vector  $\mathbf{c}(\mathbf{I} - \mathbf{U})^{-1} = (1, 2, 0)$ . We can then compute the fixed cycle and get 120202200120. Having found  $S_f[2 : 4] = 202$  we get the inverse path 202000202010, whose first vertex is a successor of  $S_f[1 : 5]$ , the first vertex of the fixed cycle. A list of all 27 inverse paths, ordered by their seeds, is given below.

$\widehat{000}$ 002122102	$\widehat{100}$ 221101011	$\widehat{200}$ 110110220
001 112000012	101 001012221	201 220021100
002 222211222	102 111220101	202 000202010
010 021020020	110 210002202	210 102011111
011 101201200	111 020210112	211 212222021
012 211112110	112 100121022	212 022100201
020 010221211	120 202200120	220 121212002
021 120102121	121 012111000	221 201120212
022 200010001	122 122022210	222 011001122

Following the steps of the algorithm we build the following de Bruijn sequence of order 5, where different chunks of digits come from different steps of the algorithm.

12020 . 00202010 . 021020020 . 010221211 . 212222021 . 120102121 . 012111000  
 . 002122102 . 111220101 . 001012221 . 201120212 . 022100201 . 220021100  
 . 221101011 . 101201200 . 110110220 . 121212002 . 222211222 . 011001122  
 . 122022210 . 102011111 . 020210112 . 100121022 . 200010001  
 . 112000012 . 211112110 . 210002202 . 0 . 2200

As a matter of fact, the above algorithm was written with the objective of making the steps of construction transparent, rather than saving in storage or time complexity. In a real implementation there is room for improvement. We note specifically that even though our code stores all inverse paths in lines (2)–(5), only the first row  $S_0[1 : q^n + k]$  has to be computed in full while the rest can be calculated and appended to the constructed sequence at once. This is true because by part (ii) of Theorem 2 the remaining rows can be computed from  $S_0$  and the pure linear recurrence by updating the seeds. Moreover, we don't need to find the inverse of  $(\mathbf{I} - \mathbf{U})$ , we only need to solve the linear system  $(\mathbf{I} - \mathbf{U})\mathbf{z}_f = \mathbf{c}$ . This and the calculation of  $\mathbf{U}$  can be done efficiently, the latter using well-known techniques for raising a companion matrix to a high power, as the diagonalization of a companion matrix is straightforward, see for example Horn and Johnson [8].

Although the storage of the full sequence  $\mathbf{b}$  may be avoided by choosing one whose feedback function is known, the major difficulty is that the sequence  $S_0$ —which is a convolution of two sequences—must be calculated and stored. This is essential in order to compute the vector  $\mathbf{c}$ , which in turn is required to identify the fixed cycle and therefore the cross-join pair.

In short, besides the storage requirement of  $S_0$ , and the solution of the linear system, the amount of calculation needed to append a digit is quite limited, noting that in Step (5) (I) the number of terms can be a fixed number much smaller than  $k$  by choosing a primitive polynomial with few terms only. This would give  $O(q^{n+k})$  which is linear in the length of the sequence being constructed.

We note that if the objective is to ‘elongate’ the period of the de Bruijn sequence  $\mathbf{b}$ , Lemma 4 guarantees that using  $0^k$  is sufficient to reach the long cycle with period  $(q^k - 1)q^n$ , or more generally,  $(q^k - 1)p$  if  $\mathbf{b}$  is replaced with any periodic sequence of period  $p$  that forms a vertex disjoint cycle in  $B(n, q)$ .

*Remark 1* Grain, a well known cipher [7], can be seen within the framework of our construction. This cipher also uses a maximum period LFSR as well as a non-linear feedback shift register (NLFSR), but the former is used to mask the input of the latter. In our terminology, the NLFSR in Grain is used as a homomorphism that maps the LFSR cycle, which belongs to a lower order de Bruijn digraph, to several inverse paths in a higher order de Bruijn digraph. This nonlinear homomorphism doesn't guarantee a desirable cycle structure. This practically means that the period of Grain depends on the chosen seed. Guided by our current results, we suggest to swap the two shift registers in order to get a better and well defined cycle structure and thus obtain a guaranteed large period for the generated output sequence. Developing and testing this suggested idea will form the subject matter of future work.

## 6 Conclusion

We showed how to use a maximal period linear recurrence relation of order  $k$  to transform a de Bruijn cycle of order  $n$  into two vertex disjoint cycles in the de Bruijn digraph of order

$n + k$ . We then presented a method to identify a conjugate pair where the two cycles are cross-joined into one de Bruijn cycle of order  $n + k$ . The proposed construction thus serves as a successful illustration of getting a de Bruijn sequence of high order from one of low order without iterating the Lempel homomorphism—and for any base  $q$  that is a power of prime.

This construction is of course valid for all  $k$  as there are  $\frac{\phi(q^k - 1)}{k}$  primitive polynomials of degree  $k$  over  $GF(q)$ . It is to be stressed that the class of homomorphisms with Property (D) is large and any homomorphism can be used once the cycle structure in the inverse factor is well understood so as to identify conjugate pairs.

**Acknowledgements** We would like to thank the referees for their careful review and sharp remarks that helped us greatly improved the quality and the presentation of the paper.

## References

1. Akinwande M.B.O.: Homomorphisms of nonbinary de Bruijn graphs with applications. Ph.D. Dissertation, Clarkson University, New York (2010).
2. Alhakim A., Akinwande M.: A recursive construction of nonbinary de Bruijn sequences. *Des. Codes Cryptogr.* **60**, 155–169 (2010).
3. Chang T., Park B., Kim Y.H.: An efficient implementation of the D-homomorphism for generation of de Bruijn sequences. *IEEE Trans. Inf. Theory* **45**, 1280–1283 (1999).
4. Fredricksen H.: A survey of full length nonlinear shift register cycle algorithms. *SIAM Rev.* **24**, 195–221 (1982).
5. Geffe P.R.: How to protect data with ciphers that are really hard to break. *Electronics* **46**(1), 99–101 (1973).
6. Golomb S.: *Shift Register Sequences*. Holden-Day, San Francisco (1967).
7. Hell M., Johansson T., Meier W.: Grain: a stream cipher for constrained environments. ECRYPT (European Network of Excellence for Cryptology), eSTREAM Project (2005).
8. Horn R.A., Johnson C.R.: *Matrix Analysis*, 2nd edn. Cambridge University Press, New York (2013).
9. Lempel A.: On a homomorphism of the de Bruijn graph and its applications to the design of feedback shift registers. *IEEE Trans. Comput. C* **19**, 1204–1209 (1970).
10. Mandal K., Gong G.: Cryptographically strong de Bruijn sequences with large periods. In: Knudsen L.R., Wu H. (eds.) SAC 2012. LNCS, pp. 104–118. Springer, Heidelberg (2012).
11. Mykkeltveit J., Siu M.-K., Tong P.: On the cyclic structure of some nonlinear shift register sequences. *Inf. Control* **43**(2), 202–215 (1979).
12. Niederreiter H.: *Random Number Generation and Quasi Monte Carlo Methods*, pp. 205–206. SIAM, Philadelphia (1992).