

Full Length Research Paper

Characterization of de Bruijn graphs homomorphisms

Akinwande Mufutau Babatunde O.

Department of Mathematics, Lagos State University, P. M. B. 01, Ojo, Lagos State, Nigeria.
 E-mail: mboakinwande@yahoo.com.

Accepted 30 August, 2011

We study homomorphisms between de Bruijn digraphs of different orders. A main theme of this paper is to characterize de Bruijn graph homomorphisms such that the inverse of a factor in the lower order digraph is also a factor in the higher order one, where a factor is a collection of cycles that partition the digraph. We generalize Lempel's homomorphism by describing and characterizing a class of homomorphisms between two de Bruijn digraphs of arbitrarily different orders but with the same alphabet, the direction of these functions being of course from the higher order digraph to the lower order one. Finally, we single out the binary case, which due to its simplicity admits a more concise characterization.

Key words: Graph homomorphism, Lempel homomorphism, de Bruijn sequence, de Bruijn graphs.

INTRODUCTION

The main graphical tool used in the study of de Bruijn sequences are *de Bruijn digraphs*. Besides its use in the context of the de Bruijn sequences, they are also used as models for transportation networks, DNA algorithms and computer networks to mention a few. The main literature about properties of de Bruijn digraphs can be founded in (Bryant and Fredricksen, 1991; Lu et al., 2000). Our terminology follows that of Lempel (1970). First, let the de Bruijn graph (or Good diagram) B_n be a directed graph on 2^n vertices, where each vertex is labeled with a distinct binary n -tuple. A directed edge is drawn from vertex $(a_1 a_2 \dots a_n)$ to vertex $(b_1 b_2 \dots b_n)$ if and only if $b_k = a_{k+1}$ for $k = 1, 2, n-1$. The de Bruijn graphs B_n for $1 \leq n \leq 4$ are illustrated in Figure 1.

Now to obtain a de Bruijn sequence of span n , construct the de Bruijn digraph B_{n-1} and label the edge from $(x_1 x_2 \dots x_{n-1})$ to $(x_2 x_3 \dots x_{n-1} a)$ with the binary n -tuple $(x_1 x_2 \dots x_{n-1} a)$. The resulting graph has 2^n edges labeled with the 2^n distinct n -tuples. Denote the *in-degree* – that is, the number of directed edges entering a node – by $d_{in}(\vec{x})$ and the *out-degree* (defined analogously) by $d_{out}(\vec{x})$. The de Bruijn graph evidently satisfies

$d_{in}(\vec{x}) = d_{out}(\vec{x}) = 2$ for each node \vec{x} , and an elementary result in graph theory implies that B_n has an Euler circuit – that is, a closed path which traverses each edge exactly once. Since consecutive edges of such an Euler circuit are of the form $x_1 x_2 \dots x_n$ to $x_2 x_3 \dots x_{n+1}$, it is apparent that there exists a de Bruijn sequence of span n .

This paper describes and characterizes homomorphism between $B_{n+k}(q)$ and $B_n(q)$ for any integer $k \geq 1$ that perform like the *D*-morphism in the sense that taking the inverse by such a homomorphism of a vertex disjoint cycle in $B_n(q)$ produces vertex disjoint cycles in $B_{n+k}(q)$. And, we finally single out the binary case which, due to its simplicity, admits a more concise characterization.

TERMINOLOGY

For positive integers n and q greater than one, let Z_q^n be the set of all q^n vectors of length n with entries in the group Z_q of residues modulo q . When the group structure

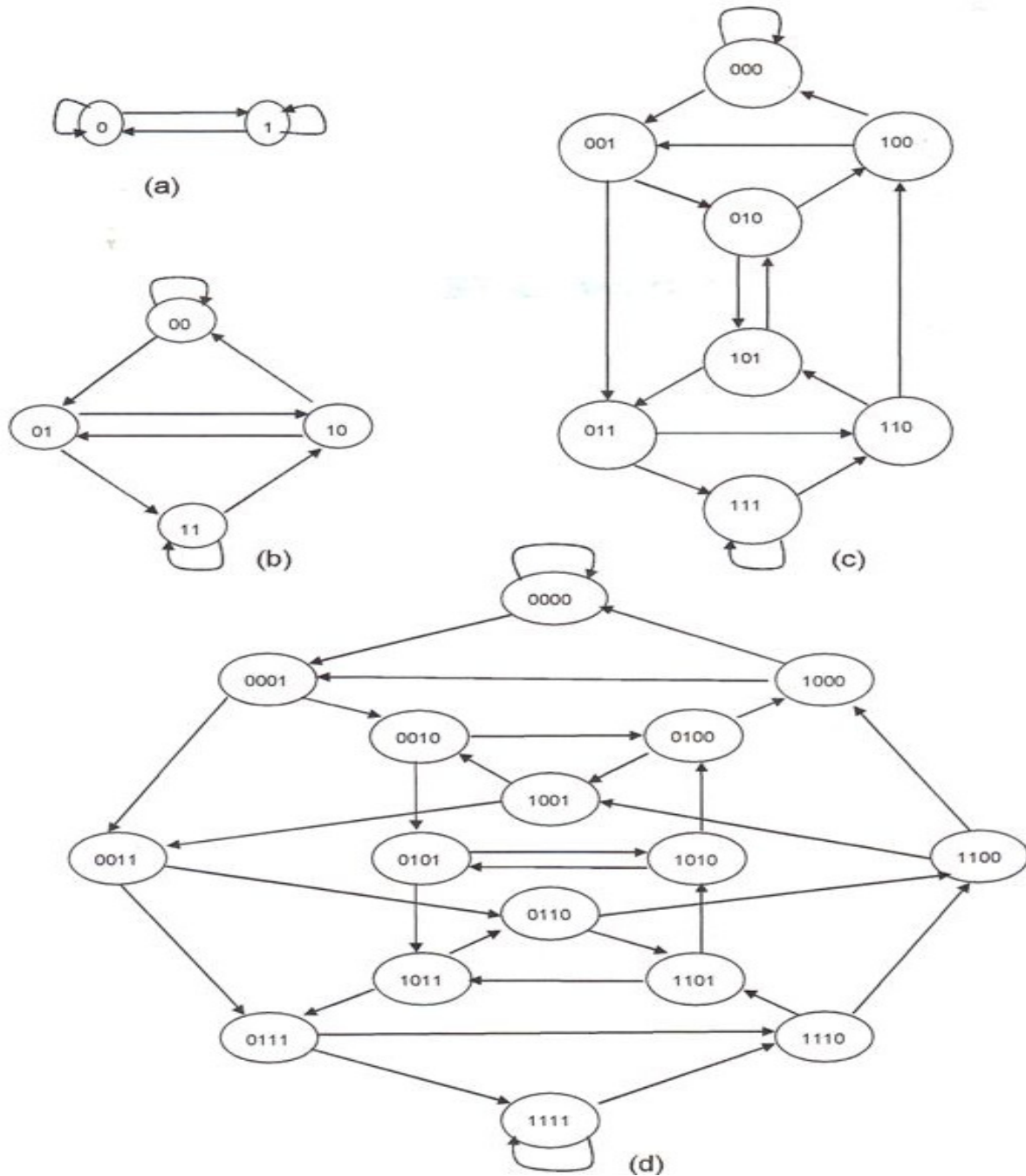


Figure 1. The de Bruijn digraphs B_n , $1 \leq n \leq 4$.

is not needed we will sometimes refer to elements of this group as symbols. For the rest of this paper, the elements of Z_q will often be used as symbols. However, the distinction should be clear from the context, as sometimes elements will be added in Z_q or multiplied in Z_q^* while in many times they will be simply concatenated.

Also, we will use the string notation to denote a vector, so a vector $(x_1 x_2 \dots x_n)$ will often be denoted as $x_1 x_2 \dots x_n$. Likewise, the terms *word*, *string* and *vector* will be used interchangeably to name the same object.

An *order n de Bruijn sequence* with alphabet in Z_q is a sequence that includes every possible string of size n as a subsequence of consecutive symbols. For example,

0011221020 is a de Bruijn sequence of order 2 with alphabet Z_3 .

An order n de Bruijn digraph, $B_n(q)$, is a directed graph with Z_q^n as its vertex set and for two vectors $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$, (X, Y) is an edge if and only if $y_i = x_{i+1}$; $i = 1, \dots, n-1$. We then say X is a predecessor of Y and Y is a successor of X . Evidently, every vertex has exactly q successors and q predecessors. Furthermore, two vertices are conjugate if they have the same successors. For example, 2120 and 0120 are conjugates in $B_4(3)$ with the common successors 1200, 1201 and 1202.

A cycle in $B_n(q)$ is a path that starts and ends at the same vertex. It is called vertex disjoint if it does not cross itself. Two cycles or two paths in the digraph are vertex disjoint if they do not have a common vertex.

Following Lempel's notation in 1970, a convenient representation of a vertex disjoint cycle $(X^{(1)}; \dots; X^{(l)})$ is via the ring sequence $(x_1 \dots x_l)$ of symbols from Z_q in such a way that the i^{th} vertex in the cycle starts with the symbol x_i . For example, the cycles (2222; 2222), (1212; 2121; 1212) and (0121; 1210; 2101; 1012; 0121) of $B_4(3)$ with respective lengths 1, 2 and 4 are represented by the ring sequences [2], [12] and [0121].

A translate of a word $w = (x_1, x_2, \dots, x_n)$ is a word $w + \lambda = (x_1 + \lambda, x_2 + \lambda, \dots, x_n + \lambda)$ where λ is any nonzero element in Z_q and the addition is performed in

Z_q . We also define a translate of a cycle as the cycle obtained by a translate of the ring sequence that defines this cycle. For example, one translate of the cycle 0121 mentioned above is defined by the ring sequence 2010, that is, (2010; 0102; 1020; 0201; 2010).

A cycle is primitive in $B_n(q)$ if it does not simultaneously contain a word and any of its translates. A function $d : Z_q^n \rightarrow Z_q$ is said to be translation invariant if

$d(w + \lambda) = d(w)$ for all $w \in Z_q^n$ and all $\lambda \in Z_q$. The weight $W(w)$ of a word or sequence w is the sum of all elements in w (not taken modulo q). Similarly, the weight of a cycle is the weight of the ring sequence that represents it.

Obviously a de Bruijn sequence of order n defines a Hamiltonian cycle in $B_n(q)$, that is, a cycle that visits each vertex exactly once and which we denote as a de Bruijn cycle. For example, the de Bruijn sequence 0011221020 yields the corresponding de Bruijn cycle is

(00; 01; 11; 12; 22; 21; 10; 02; 20; 00).

Finally for the significance and many known algebraic, combinatorial and graph-theoretical methods of construction of de Bruijn cycles, see (Akinwande, 2010). Cycles in the de Bruijn graph with various properties have been investigated by (Fredricksen, 1992; Mykkeltveit, 1972; Van Lantschoot, 1973) and some uses for de Bruijn sequences in Cryptography were proposed by (Gunther, 1988).

Jansen (1989) bases a new measure of sequence complexity on properties related to the de Bruijn graph, while applications to convolutional coding are studied in Dolinar (1992) and communication networks based on the graph are the subject of Sridhar and Raghavendra (1991).

HOMOMORPHISMS BETWEEN DE BRUIJN GRAPHS

Before we characterize graph homomorphisms between de Bruijn digraphs of different orders, we first discuss Lempel's valuable contribution to the theory of the de Bruijn graph and periodic binary sequences.

Lempel's homomorphism

Let G_1 and G_2 be two digraphs and v, v' be two arbitrary nodes in G_1 . A function H with domain G_1 and codomain G_2 is said to be a graph homomorphism if (Hv, Hv') is an edge in G_2 whenever (v, v') is an edge in G_1 . Define a map $D : B_{n+1}(2) \rightarrow B_n(2)$ by

$$D(a_1, a_2, \dots, a_{n+1}) = (a_1 + a_2, a_2 + a_3, \dots, a_n + a_{n+1})$$

where addition is modulo 2. This function defines a graph homomorphism and it is known as Lempel's D -morphism due to the fact that it was studied in Lempel (1970), although it can be traced back to Leach (1960).

Note that $D(x) = D(x+1)$ for all $x \in Z_2^{n+1}$. We define the dual of a cycle C , to be its bitwise complement \bar{C} . C is called self-dual if it is a rotation of \bar{C} . The following facts are proved in Lempel (1970).

Fact 1

A cycle of length p in $B_n(2)$ is the D -morphic image of two primitive, vertex disjoint cycles of length p in $B_{n+1}(2)$ if and only if it has an even number of ones in the ring sequence representation.

Fact 2

A cycle of length p in $B_n(2)$ is the D -morphic image of a self-dual cycle of length $2p$ in $B_{n+1}(2)$ if and only if it has an odd number of ones in the ring sequence representation. Two cycles are called *adjacent* if a vertex on one cycle has a conjugate on the other cycle. Swapping the successors of these two conjugate words joins the two cycles into one larger cycle. This is why any pair of conjugate words is called a *cross-join pair*, (a similar concept of a cross-join tuple can be defined for $q > 2$.) By Fact 1, if c_n is a Hamiltonian cycle in $B_{n-1}(2)$, then every word in Z_2^n is either on c_n or on \bar{c}_{n+1} , the two primitive pre-images of c_n by D . Lempel used this idea to construct de Bruijn cycles recursively by rejoining c_n and \bar{c}_n . The most obvious cross-join pair is the two *alternating strings* of size n $z_n = 010\dots$ and its complement \bar{z}_n which cannot be on the same cycle.

Recently, this method was implemented in Annexstein (1997) with an efficient, linear code, and more recently done even more efficiently in Akinwande (2010) with a jump from a given de Bruijn cycle in $B_n(2)$ to a higher order $B_{n+k}(2)$, for some integer k that is a power of 2, by pre-computing the effect of an iterative application of the D -morphism.

HOMOMORPHISM CHARACTERIZATION

The following proposition characterizes graph homomorphisms between de Bruijn digraphs of different orders.

Proposition 1

A necessary and sufficient condition for a map $D_{n,k} : B_{n+k}(q) \rightarrow B_n(q)$ to be a graph homomorphism is that:

$$D_{n,k}(X) = (d_k(x_1, \dots, x_{k+1}), d_k(x_2, \dots, x_{k+2}), \dots, d_k(x_n, \dots, x_{n+k}))$$

where $X = (x_1, \dots, x_{n+k})$ and d_k is any fixed function of $k+1$ variables.

Proof

Sufficiency is quite obvious so we will only prove the necessity.

Let $D_{n,k}(x_1, \dots, x_{n+k})$ be $(\tilde{x}_1, \dots, \tilde{x}_n)$ where $\tilde{x}_i = d_i(x_1, \dots, x_{n+k})$ is a function from Z_q^n to Z_q for all $i = 1, \dots, n$. For all values of $x_1, \dots, x_{n+k}, x_{n+k+1}$; (x_1, \dots, x_{n+k}) is a predecessor of (x_2, \dots, x_{n+k+1}) . Hence, since $D_{n,k}$ is a graph homomorphism the diagram below commutes, where the horizontal arrows indicate an edge in the de Bruijn digraph. That is,

$$\begin{array}{ccc} (x_1, \dots, x_{n+k}) & \xrightarrow{B_{n+k}} & (x_2, \dots, x_{n+k}, x_{n+k+1}) \\ D_{n,k} \downarrow & & \downarrow D_{n,k} \\ (\tilde{x}_1, \dots, \tilde{x}_n) & \xrightarrow{B_n} & (\tilde{x}_2, \dots, \tilde{x}_n, \tilde{x}_{n+1}) \end{array}$$

$$d_i(x_1, \dots, x_{n+k}) = d_{i-1}(x_2, \dots, x_{n+k+1}); \quad i = 2, \dots, n \quad (1)$$

To finish the proof we need to establish that

- (i) $d_i(x_1, \dots, x_{n+k}) = d_j(x_1, \dots, x_{n+k})$ for all $i \neq j$, and (ii) d_i depends at most on x_i, \dots, x_{i+k} . We establish this by iterating Equations (1) for $i = 2, \dots, n$.

To avoid confusion, we will denote d_i by d_i^L and d_i^R when it is applied to (x_1, \dots, x_{n+k}) and (x_2, \dots, x_{n+k+1}) respectively (the left and right sides of the diagram above). This is meant to remind us that, e.g., the first variable of the input of d_i^R is x_2 .

First $\tilde{x}_2 = d_2^L(x_1, \dots, x_{n+k}) = d_1^R(x_2, \dots, x_{n+k+1})$, so that d_1 does not depend on its $(n+k)^{th}$ variable and d_2 does not depend on its first variable.

Next, $\tilde{x}_3 = d_3^L(x_1, \dots, x_{n+k}) = d_2^R(x_3, \dots, x_{n+k+1})$ (noting that by the former result d_2^R does not depend on its first variable x_2). It follows that d_2 does not depend on its $(n+k)^{th}$ variable and d_3 does not depend on its first and second variables.

Continuing this way we see that for $i = 2, \dots, n$; d_{i-1} does not depend on the last variable and d_i does not depend on its first $(i-1)$ variables. In particular, d_n depends on at most x_n, \dots, x_{n+k} and d_{n-1} depends on at most $x_{n-1}, \dots, x_{n+k-1}$.

Next $d_{n-1}^L(x_{n-1}, \dots, x_{n+k-1}) = d_{n-2}^R(x_{n-1}, \dots, x_{n+k})$ implies that d_{n-2} does not depend on its $(n+k-1)^{st}$ variable. Continuing with Equations (1) iteratively and backwards this time we establish requirement (ii) above. But then Equations (1) reads

$$d_i^L(x_i, \dots, x_{i+k}) = d_{i-1}^R(x_i, \dots, x_{i+k}).$$

Hence, for all i , d_i is a fixed function of $k+1$ variables which establishes (i).

HOMOMORPHISMS WITH PROPERTY D

By Fact 1, a vertex disjoint cycle in B_n is the D -morphic image of two vertex disjoint cycles in B_{n+1} starting respectively with zero and one. The following definition will be seen to generalize this D -morphism to homomorphisms between de Bruijn digraphs of different orders.

Definition

A homomorphism $D_{n,k}$ from $B_{n+k}(q)$ to $B_n(q)$ is said to have *property D* if each vertex disjoint path in $B_n(q)$ is the image of exactly q^k non-overlapping vertex disjoint paths in $B_{n+k}(q)$, one for each starting string of size k . The function d_k corresponding to $D_{n,k}$ will also be said to have property D .

Note that, in particular, each vertex in $B_n(q)$ has q^k inverse image vertices by $D_{n,k}$. We will illustrate this property as well as the absence of this property with some examples before we state the theorem that characterizes functions d_k that have property D . In fact, a direct inspection of the sixteen Boolean functions of two variables shows that the only homomorphisms with property D from B_{n+1} to B_n are the D -morphism of Lempel and its bitwise complement. The next example concerns the case when $k = 2$.

Examples

(a) Consider the mapping $D_{1,2}$ from $B_3(2)$ to $B_1(2)$ that uses the function $d(x_1, x_2, x_3) = x_1 + x_2$. The inverse sets of 0 and 1 are respectively $\{000, 001, 110, 111\}$ and

$\{010, 011, 100, 101\}$. The edge $(0,1)$ of $B_1(2)$ is mapped back to the four edges $\{(001; 010), (001; 011), (110; 100), (110; 101)\}$. Note that, even though each edge in $B_1(2)$ is the image of four edges in $B_3(2)$, it is not possible to construct an edge starting with arbitrary strings of size two that is mapped to a given edge of $B_1(2)$. For instance, there is no edge in $B_3(2)$ that starts with either 01 or 10 and whose image is the edge $(0,1)$.

(b) The function $H_{n,k}$ from $B_{n+k}(q)$ to $B_n(q)$ for $k \geq 0$ and $n \geq 1$ was defined in (Chen and Chen, 1995) as $H_{n,k}(x_1, \dots, x_{n+k}) = (x_{k+1}, \dots, x_{n+k})$. In other words, this function trims the k leftmost symbols of a word so as to make it a word of size n . Obviously, this is a homomorphism having, according to the notation of Proposition (1), $d(x_1, \dots, x_{k+1}) = x_{k+1}$, hence the theorem below shows that it does not enjoy property D . In fact the q^k inverses of any cycle in $B_n(q)$ by $H_{n,k}$ disagree only in their first k terms while the body of the sequences are all *equal* to the original cycle.

(c) Using the function $d^{(1)}(x_1, x_2, x_3) = x_1 + x_3$ however, the edges $(0;0)$, $(0;1)$, $(1;0)$, $(1;1)$ of $B_1(2)$ are respectively mapped back to the following sets whose union constitutes the edge set of $B_3(2)$, each edge appearing exactly once.

$\{(000; 000), (010; 101), (101; 010), (111; 111)\}$
 $\{(000; 001), (010; 100), (101; 011), (111; 110)\}$
 $\{(001; 010), (011; 111), (100; 000), (110; 101)\}$
 $\{(001; 011), (011; 110), (100; 001), (110; 100)\}$

Hence $d^{(1)}$ enjoys property D while d does not.

Theorem 1

(a) A homomorphism from $B_{n+k}(q)$ to $B_n(q)$ that is induced by $d_k(x_1, \dots, x_{k+1})$ enjoys property D if and only if d_k is one to one in each of the variables x_1 and x_{k+1} when all the other variables are kept fixed.
 (b) The total number of homomorphisms with property D is $(A_q)^{q^{k-2}}$ where A_q is the number of $q \times q$ Latin squares.

Proof

Part (b) follows from (a) since $d_k(x_1, \dots, x_{k+1})$ defines a

Latin square for each set of fixed values of x_2, \dots, x_k given the condition in (a) (Sloane, 2003) for more about the sequence A_q . To prove part (a), first let d_k be a function with property D and $C = [c_1, \dots, c_l]$ be an arbitrary vertex disjoint path in $B_n(q)$. By definition of property D each word (x_1, \dots, x_k) in Z_q^k can be appended by a symbol x_{k+1} so that $d_k(x_1, \dots, x_k, x_{k+1}) = c_1$. This says that d_k is surjective from Z_q^k to Z_q (hence injective) with respect to the last variable.

Now let x'_1 be such that $d_k(x'_1, x_2, \dots, x_k, x_{k+1})$ and $d_k(x_1, x_2, \dots, x_k, x_{k+1})$ are equal to c_1 . Since d_k is bijective with respect to the last variable, there exist unique values $x_{k+2}, \dots, x_{n+k+1}$ such that $D_{n,k}(x_2, \dots, x_{n+k+1}) = (c_2, \dots, c_{n+1})$. If $x'_1 \neq x_1$ then the two distinct inverse edges $(x_1, \dots, x_{n+k}; x_2, \dots, x_{n+k+1})$, $(x'_1, x_2, \dots, x_{n+k}; x_2, \dots, x_{n+k+1})$ share a common vertex, contradicting property D . Hence d_k is one-to-one in the first variable. This establishes the necessary condition.

Conversely, let d_k have the claimed form and let $C = [c_1, \dots, c_l]$ be a vertex disjoint path in $B_n(q)$. Given any string x_1, \dots, x_k it is possible to find a value $b \in Z_q$ so that $d_k(x_1, \dots, x_k, b) = c_1$, since d_k is surjective with respect to the last variable. Hence the value c_1 has a set of q^k inverse images that includes all possible strings of size k as prefixes. The same argument can be repeated to show that the path C has exactly q^k inverse images. To show property D we need to show that the paths in $B_{n+k}(q)$ that form the inverse images of C are vertex disjoint, i.e., that no substring of size $n+k$ occurs more than once in the collection of these inverse images. Write the pre-images of C as a rectangular array $(x_{ij}); i=1, \dots, q^k, j=1, \dots, k+l$ where the set of prefixes of size k coincides with the q^k distinct words of this size. This means, of course, that $D_k(x_{i,j}, \dots, x_{i,n+k+j-1}) = c_j$. Let us denote by $\omega_{ij}(u)$ the substring of size u on the i^{th} row of (x_{ij}) that starts with (x_{ij}) i.e. $\omega_{ij}(u) = (x_{ij}, x_{i(j+1)}, \dots, x_{i(j+u-1)})$. Assume there exist integers $i_1, i_2, j_1, j_2, 1 \leq i_1, i_2 \leq q^k, 1 \leq j_1, j_2 \leq l-n+1$ such that $\omega_{i_1, j_1}(n+k)$ coincides with $\omega_{i_2, j_2}(n+k)$. Obviously

$j_1 \neq j_2$ implies that a string of size n occurs twice in C , thus contradicting the assumption that C is vertex disjoint. Assume then that j_0 is the smallest integer with $\omega_{i_1, j_0}(n+k) = \omega_{i_2, j_0}(n+k)$, which in particular means that $\omega_{i_1, j_0}(k) = \omega_{i_2, j_0}(k)$. By construction of (x_{ij}) , it is immediate that $j_0 > 1$. Hence $d_k(\omega_{i_1, j_0-1}(k+1)) = d_k(\omega_{i_2, j_0-1}(k+1)) = c_{j_0-1}$. Since the last k components of $\omega_{i_1, j_0-1}(k+1)$ and $\omega_{i_2, j_0-1}(k+1)$ are the same, the one-to-one property of d_k with respect to the first variable implies that $x_{i_1, j_0-1} = x_{i_2, j_0-1}$. Therefore $\omega_{i_1, j_0-1}(n+k) = \omega_{i_2, j_0-1}(n+k)$, which contradicts the minimality of j_0 . This establishes the theorem.

In general we see that applying the inverse of a homomorphism to a vertex disjoint cycle in $B_n(q)$ creates multiple cycles in $B_{n+k}(q)$. If $B_n(q)$ is partitioned into vertex disjoint cycles then the inverse homomorphism naturally induces a partition of $B_{n+k}(q)$ into vertex disjoint cycles.

THE BINARY CASE

We treat here the binary case separately because its simplicity allows for a more concise characterization of the shape of homomorphisms with property D .

Theorem 2

A necessary and sufficient condition for a homomorphism $D_{n,k}$ from $B_{n+k}(2)$ to $B_n(2)$ to have property D is that $d_k(x_1, \dots, x_{k+1}) = x_1 + h(x_2, \dots, x_k) + x_{k+1}$, where $h(x_2, \dots, x_k)$ is any Boolean function of $k-1$ variables.

Proof

By Theorem (1) we only need to show that a binary function d_k is bijective with respect to the first and last variables if and only if it has the form claimed in this Theorem. In effect, if $d_k(x_1, \dots, x_{k+1})$ is bijective in x_1 and in x_{k+1} then it satisfies the equations

$$d_k(\bar{x}_1, x_2, \dots, x_{k+1}) = 1 - d_k(x_1, x_2, \dots, x_{k+1}) = d_k(x_1, x_2, \dots, \bar{x}_{k+1})$$

So that $d_k(\bar{x}_1, x_2, \dots, x_k, \bar{x}_{k+1}) = d_k(x_1, x_2, \dots, x_k, x_{k+1})$. Therefore for each fixed set of values for x_2, \dots, x_k , $d_k(x_1, x_2, \dots, x_{k+1}) = d_{x_2, \dots, x_k}(x_1, x_{k+1})$ is either $x_1 + x_{k+1}$ or $x_1 + x_{k+1} + 1$. This can be rephrased to establish the necessity. The converse is obvious because d_k is linear in the first and last variables.

This elegant form of d_k is mainly due to the "lack" of terms in Z_2 . While Theorem (1) shows that $d_k = \alpha x_1 + h(x_2, \dots, x_k) + \beta x_{k+1}$ is sufficient for property D , the following example illustrates why property D homomorphisms cannot be all written in such a simple form even for $q = 3$. In fact, all the twelve 3×3 Latin squares can be written in function form as $f(b_1, b_2) = \alpha_1 b_1 + \alpha_2 b_2 + \alpha_3$ where $\alpha_i, b_i \in Z_3$, $\alpha_1 \neq 0$ and $\alpha_2 \neq 0$. From the values 0, 1, 2 of x_2 let $d_k(x_1, x_2, x_3)$ be respectively $x_1 + x_2 + x_3$, $2x_1 + x_3$, and $x_1 + 2x_2 + 2x_3$. Then d_k has property D by Theorem (1) but it is not linear in either x_1 or x_{k+1} , despite the simple form of Latin squares. Notice that when $q > 3$ most Latin squares are already nonlinear.

While the only binary homomorphism for $k = 1$ is Lempel's D -morphism (and its bitwise complement), there are essentially two homomorphisms for $k = 2$ that are induced by the functions $d^{(1)} = x_1 + x_3$ and $d^{(2)} = x_1 + x_2 + x_3$. Note that the former is just the D -morphism iterated twice. The only other two homomorphisms are bitwise complements of $d^{(1)}$ and $d^{(2)}$. The cases $k \geq 3$ allow for nonlinear homomorphisms such as $d(x_1, \dots, x_4) = x_1 + x_2 x_3 + x_4$. Let $C = [c_1, \dots, c_l]$ be an arbitrary but fixed cycle in B_n , started at a fixed word, say, 0...0. Then for each homomorphism $D_{n,k}$ with property D , C defines a map φ_C on the set Z_2^k as follows: $\varphi_C(z_1 \dots z_k)$ is the suffix of length k of $\varphi_C^{-1}C$ started at the string $z_1 \dots z_k$. The inverse image is generated by the recursive equation:

$$z_i = c_{i-k} + z_{i-k} + h(z_{i-k+1}, \dots, z_{i-1}); \quad i = k+1, \dots, k+l,$$

where h is as in Theorem (2) and z_1, \dots, z_k are the required initial conditions. It can be seen that property D implies that φ_C is a bijection. When the D -morphism is used, any de Bruijn cycle b_n yields the identity permutation on the set $\{0, 1\}$. This is a restatement of the

fact that the inverse image of any de Bruijn cycle b_n under the D -morphism makes two dual cycles in B_{n+1} . Since a binary de Bruijn cycle necessarily has an even number of ones, this follows immediately from Fact (1) above. The next proposition concerns the function $d^{(2)}$ defined above.

Proposition 2

For any integer $n \geq 1$ and any de Bruijn cycle $\mathbf{b}_n = [b_1 \dots b_{2^n}]$, the homomorphism induced by the Boolean function $d^{(2)}(x_1, x_2, x_3) = x_1 + x_2 + x_3$ defines a permutation of the set of seeds $\{00, 01, 10, 11\}$ with exactly one fixed point $z_1 z_2$ obtained by $z_1 = a_0 + \delta_{\tilde{n},0} a_1 + \delta_{\tilde{n},1} a_2$, $z_2 = a_1 + \delta_{\tilde{n},0} a_2 + \delta_{\tilde{n},1} a_0$ where $\tilde{n} = n \bmod 2$, $a_j = a_j^n := \sum_i b_{3i+j} \bmod 2$; $j = 0, 1, 2$, the sum is taken over the range of indices of \mathbf{b}_n ($1 \leq 3i + j \leq 2^n$), and addition in the index of a_j is taken modulo 3.

In other words, exactly one of the four sequences that form the pre-image of \mathbf{b}_n is a closed cycle in B_{n+2} . As a result, the other sequences together form one cycle of length $3 \cdot 2^n$.

Example

Let $\mathbf{b}_3 = [00011101]$. We see that $\tilde{n} = 1$, $a_0 = 1$, $a_1 = 1$ and $a_2 = 0$ so that the fixed point is $z_1 z_2 = 10$. Indeed the inverse image by $d^{(2)}$ gives the following four sequences:

$$\begin{array}{cc} \underline{0000010001} & \underline{0110100111} \\ \underline{1011001010} & \underline{1101111100} \end{array}$$

So the fixed point gives the only cycle of length 8 while the other three sequences make the following cycle of length 24: $\underline{000001000110100111011111}$.

Proof of proposition 2

Let $\bar{i} = i \bmod 3$. Iterating the relation $z_i = b_{i-2} + z_{i-1} + z_{i-2}$, which is satisfied by the sequence $\{z_i\}_{i=3}^{2^n+2}$, we get (for all i in the range of the latter sequence):

$$z_i = \sum_{j=0}^{\lfloor i/3 \rfloor - 1} (b_{3j+i} + b_{3j+i+1}) + \delta_{i,1} z_1 + \delta_{i,2} z_2 + \delta_{i,0} (z_1 + z_2)$$

where we define b_0 to be zero. Note that $2^n \bmod 3 = 1$ or 2 when n is even or odd respectively. For each of these two cases, using the above recursive equation and the requirement $z_{2^n+j} = z_j$; $j = 1, 2$ yields two linearly independent equations whose unique solution is as claimed.

Shifting \mathbf{b}_n by a number that is not a multiple of 3 permutes the numbers a_j ; $j = 0, 1, 2$. So it changes the permutation but still keeps one fixed point. This result is interesting because it is independent of the de Bruijn cycle used. The permutation induced by $d^{(1)}$ may or may not have a fixed point, depending on \mathbf{b}_n . As a result, the function $d^{(2)}$ can be used to generate de Bruijn cycles recursively by joining the shorter cycle (the one started at the fixed point) to the long cycle made of the other three starting values. Proposition 2 describes the way to identify the two starting digits of the short cycle.

There is no simple way to identify a pair of conjugate words to perform this cross-join operation a priori, for example the alternating strings may or may not be on the same cycle. Note that the existence of a word without a conjugate on the short cycle is guaranteed because otherwise the cycle must be a de Bruijn cycle (Fredricksen, 1982). Consequently, one can find a cross-join pair by only searching the small cycle for a word without a conjugate there. This search takes $O(N^2)$ in the worst case, where $N = 2^n$ is the length of the short cycle. This is manageable for small to medium word size n .

Conclusion

In this paper, we described a detailed characterization of a class of homomorphisms between de Bruijn digraphs of different orders with a property D that can be used to construct de Bruijn cycles recursively. For two positive integers n and k , property D allows a recursive construction of de Bruijn cycles that the inverse of a factor in $B_n(q)$ is a factor in $B_{n+k}(q)$ which generalized a well-known binary construction of Lempel.

REFERENCES

- Akinwande MBO (2010). Homomorphisms of nonbinary de Bruijn graphs with Applications. Ph.D. Dissertation, Clarkson University, New York, USA.
- Annexstein FS (1997). Generating de Bruijn Sequences: An Efficient Implementation. *IEEE Trans. Comput.*, 46(2): 198-200.
- Bryant RC, Fredricksen H (1991). Covering the de Bruijn graph. *Discrete Math.*, 89(2): 133-148.
- Chen C, Chen J (1995). A homomorphism of the de Bruijn graphs and its applications. *IEEE First International Conference on Algorithms and Architectures for Parallel Processing*, pp. 465-470.
- Dolinar S, Ko TM, McEliece R (1992). Some VLSI decompositions of the de Bruijn graph. *Discrete Math.*, 106/107: 189-198.
- Fredricksen H (1982). A Survey of Full Length Nonlinear Shift Register Cycle Algorithms. *SIAM Rev.*, 24(2): 195-221.
- Fredricksen H (1992). A new look at the de Bruijn graph. *Discrete Appl. Math.*, 37/38: 193-203.
- Gunther CG (1988). Alternating step generators controlled by de Bruijn sequences, W. L. Price, editors, *Processings EUROCRYPT'87*, Lecture Notes in Computer Science 304, Springer-Verlag, Berlin, pp. 5-14.
- Jansen CJA (1989). Investigations on nonlinear stream cipher systems: construction and evaluation methods. Ph.D. Thesis, Technical University, Delft.
- Leach EB (1960). Regular Sequences and Frequency Distributions. *Proc. Am. Math. Soc.*, 11: 566-574.
- Lempel A (1970). On a Homomorphism of the de Bruijn Graph and Its Applications to the Design of Feedback Shift Registers. *IEEE Trans. Comput.* C-19(12): 1204-1209.
- Lu C, Xu J, Zhang K (2000). On $(d, 2)$ -dominating numbers of binary undirected de Bruijn graphs. *Discrete Appl. Math.*, 105(1-3): 137-145.
- Mykkeltveit J (1972). A proof of Golomb's conjecture for the de Bruijn graph. *J. Combinatorial Theory Ser. B*(13): 41-45.
- Sloane NJA (2003). The On-Line Encyclopedia of Integer Sequences, published electronically at <http://www.research.att.com/~njas/sequences/>.
- Sridhar MA, Raghavendra CS (1991). Fault-tolerant networks based on the de Bruijn graph. *IEEE Trans. Comput.*, C-40. pp 1167-1174.
- Van Lantschoot EJ (1973). Double adjacencies between cycles of a circulating shift register. *IEEE Trans. Comput.*, C-22. pp. 944-955.