Global Journal of Mathematical Sciences: Theory and Practical.
ISSN 0974-3200 Volume 4, Number 2 (2012), pp. 121–132

http://www.irphouse.com

Correlation of Pseudorandom Binary Sequences from de Bruijn Graphs

Mufutau B Akinwande
Department of Mathematics,
Lagos State University, Lagos - Nigeria.
E-mail: mboakinwande@yahoo.com

Abstract

Binary sequences with good correlation properties play an important role in many secure communication systems and testing of systems. In this paper, we describe and illustrate sets of pseudorandom sequences from de Bruijn graphs which have good correlation functions and critically analyze how we investigate all homomorphisms that give low correlation values between the binary sequences. We compute their correlation functions, which for certain nontrivial homomorphisms turn out to be asymptotically within a factor of 2.5 of the Welch bound.

AMS subject classification: 68R01, 68R10, 05C38.
Keywords: Binary sequences, Periodic sequences, Correlation, de Bruijn graphs.

1. Introduction

A (periodic) correlation function is an important measure to evaluate the effectiveness of pseudorandom sequences. In practice, the sequences are required to have the impulse-like autocorrelation function. Also, cross-correlation magnitudes of the distinct sequences must be as low as possible. Periodic sequences with good correlation properties are important for many applications in communication systems, and much effort has been expended on design techniques for such sequences.

Pseudorandom sequences with good correlation properties, large linear complexity, and balance statistics are widely used in modern communication and cryptology. In code division multiple access (CDMA) communication systems, low cross-correlation between the desired users and interfering users is important to suppress multiuser interference. Good autocorrelation properties are important for reliable initial synchronization and separation of the multipath components. Moreover, the number of available sequences should be sufficiently large so that the system can accommodate enough users.

The randomness and complexity properties of these sequences are important in some applications where security is an issue. Large linear span (also known as linear complexity) of the sequence is required to prevent it from being reconstructed from a portion of the sequence, for example, using the Berlekamp-Massey algorithm. With the exception of [1], the sequences in these sets are obtained from sequences of powers of primitive elements in fields of characteristic 2 by algebraic manipulations. As a consequence, the sequences all have periods of the form $2^n - 1$ (or, in the case of sequences obtained from the duals of nonprimitive BCH codes, period dividing $2^n - 1$) and the sizes of the sets are rather restricted. In literature, extensive research has been performed on how to generate sequences with these desired properties, some representative examples can be found in [2] and references therein.

De Bruijn sequences have been used in some communication systems whose performance depends on the correlation properties of de Bruijn sequences. And, in general, it is desired to design a set of sequences with an impulsive autocorrelation function and a zero cross-correlation function for many practical applications. However, as noted by Welch [3], Sarwate [4], Sidelnikov [5], and Massey [6]; it was impossible to construct such an ideal set of sequences. Therefore, searching large family of sequences with good autocorrelation function and cross-correlation function properties has been one of the most interesting topics in sequence design. It was first defined many years ago by Barker [7] and has attracted the interests of engineers, mathematicians, physicists and chemists. For evaluating the correlation properties, one good choice is to use the maximum magnitude of the autocorrelation function and the maximum magnitude of the cross-correlation function.

In section 2, we introduce some necessary concepts for sequence designs which will be used throughout the paper. The measures used to quantify the correlation properties of the sequences are discussed and illustrated with examples. Section 3 creates a novel method to generates pseudorandom binary sequences, this method relies on D-homomorphism between de Bruijn digraphs of different orders. Numerical computation of the maximum magnitude of the cross-correlation function with typical computer results are given in section 4. Section 5 offers conclusions and gives analysis comparing the correlations of the constructed parallel inverse images (sequences) in the context of known sequence families.

2. Background

Correlation is a measure of the similarity or relatedness between two phenomena. When properly normalized, the correlation measure is a real number between -1 and +1, where a correlation value of +1 denotes the two phenomena are identical, a correlation value of -1 means that they are diametrically opposite, and a correlation value of 0 means that they agree exactly as much as they disagree. Statistically, correlation between two sets of data is called their *covariance* while the correlation between two vectors is their (normalized) *dot product* in linear algebra.

Suppose that $\alpha = (a_1, a_2, \ldots, a_n)$ and $\beta = (b_1, b_2, \ldots, b_n)$ are two n-dimensional

vectors of real numbers, which could represent two sets of experimental data. The magnitudes of these vectors are $|\alpha| = \left(\sum_{i=1}^{T} a_i^2\right)^{1/2}$ and $|\beta| = \left(\sum_{i=1}^{T} b_i^2\right)^{1/2}$, the normalized vectors are $\alpha' = \dfrac{\alpha}{|\alpha|}$ and $\beta' = \dfrac{\beta}{|\beta|}$ and so their correlation is

$$C(\alpha, \beta) = \frac{< \alpha, \beta >}{|\alpha||\beta|} = \frac{\sum_{i=1}^{T} a_i b_i}{(\sum_{i=1}^{T} a_i^2)^{1/2} (\sum_{i=1}^{T} b_i^2)^{1/2}}.$$

Now, suppose that $\alpha = (a_1, a_2, \ldots, a_n)$ and $\beta = (b_1, b_2, \ldots, b_n)$ are both binary vectors and specifically that the $a_i s$ and $b_i s$ are restricted to the two values +1 and -1 (for technical reasons, we switch from bits to ± 1). Then both

$$|\alpha| = \left(\sum_{i=1}^{T} a_i^2\right)^{1/2} = \sqrt{T} = \left(\sum_{i=1}^{T} b_i^2\right)^{1/2} = |\beta|$$

from which

$$C(\alpha, \beta) = \frac{1}{T} \sum_{i=1}^{T} a_i b_i = \frac{1}{T}(A - D) = \frac{A - D}{A + D},$$

where A is the number of times, for i from 1 to n, that a_i and b_i agree, and D is the number of times that a_i and b_i disagree. Clearly, $A + D = n$, since a_i and b_i either agree or disagree at each value of i. Because $(+1)(+1) = (-1)(-1) = +1$, whereas $(+1)(-1) = (-1)(+1) = -1$, each agreement between a_i and b_i contributes -1, to the sum $\sum_{i=1}^{T} a_i b_i$.

Remark 2.1. If a_i and b_i agree completely, then $A = T$, $D = 0$, and $C(\alpha, \beta) = \dfrac{A - 0}{A + 0} = \dfrac{T}{T} = +1$. If a_i and b_i disagree everywhere, then $A = 0$, $D = T$, and $C(\alpha, \beta) = \dfrac{0 - D}{0 + D} = \dfrac{-T}{T} = -1$. If agreements and disagreements occur equally often, then $A = D = \dfrac{T}{2}$, $A - D = 0$, and $C(\alpha, \beta) = \dfrac{A - D}{A + D} = \dfrac{0}{T} = 0$.

Definition 2.2. (Equivalence Relation of Sequences) Let $\mathbf{a} = \{a_t\}$ and $\mathbf{b} = \{b_t\}$ be two periodic sequences. Then, they are called *cyclically equivalent* [2] if there exists an integer k such that $a_t = b_{t+k}$ for all $t \geq 0$ denoted by $\mathbf{a} = L^k(\mathbf{b})$. Otherwise, they are called *cyclically distinct*.

Definition 2.3. (Balanced and Almost Balanced Properties) Let $\mathbf{a} = \{a_t\}$ be a binary sequence of period T. Then, $\mathbf{a}$ is called *balanced* if the number of zeros is nearly equal to the number of ones in a period, i.e.

$$S = |\sum_{t=0}^{T-1}(-1)^{a_t}| \leq 1$$

where S denotes a difference between the numbers of zeros and ones of a binary sequence in a period. For odd T, $\mathbf{a}$ is balanced if and only if $S = 1$, and for even T, it is balanced if and only if $S = 0$. Otherwise, if T is even and $S = 2$, then $\mathbf{a}$ is called *almost balanced*.

Definition 2.4. (Trace Function) Let m and n be positive integers such that $m|n$. A trace function from $\mathbb{F}_{2^n}$ to $\mathbb{F}_{2^m}$, denoted by $Tr_m^n(x)$, is defined as

$$Tr_m^n(x) = x + x^{2^m} + \ldots + x^{2^{m(\frac{n}{m}-1)}}, \quad x \in \mathbb{F}_{2^n},$$

or simply as $Tr(x)$ if $m = 1$ and the context is clear.

Definition 2.5. (Decimation of Periodic Sequences) Let $\mathbf{a} = \{a_t\}$ be a binary sequence of period $T|(2^n - 1)$ and $f(x)$ be trace representation of $\mathbf{a}$. Let $0 < s < T$. Then a sequence $\mathbf{b} = \{b_t\}$, whose elements are given by $b_t = a_{st}, t = 0, 1, \ldots,$ is said to be $s -$ ***decimation*** of $\mathbf{a}$, denoted by $\mathbf{a}^{(s)}$. The trace representation of $\mathbf{a}^{(s)}$ is $f(x^s)$. That is,

$$\mathbf{a} \leftrightarrow f(x)\ ,\ \mathbf{a}^{(s)} \leftrightarrow f(x^s).$$

Definition 2.6. (Autocorrelation of Binary sequences) A (periodic) autocorrelation function of a binary sequence $\mathbf{a} = \{a_t\}, t = 0, 1, \ldots, T-1$ of period T is defined by

$$C_{\mathbf{a}}(\tau) = \Sigma_{t=0}^{T-1}(-1)^{a_{t+\tau}+a_t}, 0 \leq \tau \leq T-1$$

where τ is a phase shift of a and the indices are computed modulo T. For $\tau \neq 0$ (or any multiple of T), the values of $C_{\mathbf{a}}(\tau)$ are called *sidelobes*.

For a sequence $\mathbf{a}$ of period T, it is well known that $C_{\mathbf{a}}(\tau) = T$ occurs only at $\tau \equiv 0 \pmod{T}$. In particular, if $\mathbf{a}$ has the autocorrelation function of

$$C_{\mathbf{a}}(\tau) = \begin{cases} -1, & \text{if } \tau \neq 0 \bmod T \\ T, & \text{if } \tau \equiv 0 \bmod T, \end{cases}$$

then the sequence is said to have the *(ideal) two-level autocorrelation function*. A binary sequence with ideal two-level autocorrelation function corresponds to the *cyclic Hadamard difference set* [8].

Definition 2.7. (Perfect sequences) Let $\mathbf{a}$ be a binary sequence of period T. If its autocorrelation $C_{\mathbf{a}}(\tau)$ equal 0 for all $\tau \neq 0 \bmod T$, i.e.,

$$C_{\mathbf{a}}(\tau) = \begin{cases} 0, & \text{if } \tau \neq 0 \bmod T \\ T, & \text{if } \tau \equiv 0 \bmod T, \end{cases}$$

then **a** is called a *perfect sequence*.

However, the only known perfect binary sequence is $\mathbf{a} = (0, 1, 1, 1)$ or its complement [8]. For a period of $4 < T < 108900$, no perfect binary sequences are found [9], and it is conjectured in [10] that no other perfect binary sequences exist except for $T = 4$.

Definition 2.8. (Cross-correlation of Binary sequences) A (periodic) cross-correlation function of binary sequences **a** and **b** of period T is defined by

$$C_{\mathbf{a},\mathbf{b}}(\tau) = \Sigma_{t=0}^{T-1}(-1)^{a_{t+\tau}+b_t}, 0 \leq \tau \leq T-1$$

where τ is a phase shift of the sequence **a** and the indices are computed modulo T. The values of $C_{\mathbf{a},\mathbf{b}}(\tau)$ are referred to as sidelobes.

Let $T = 2^n - 1$. If $f(x)$ and $g(x)$ are trace representations of **a** and **b** with $f(0) = g(0) = 0$, respectively, then the cross-correlation is given by

$$\begin{aligned} C_{\mathbf{a},\mathbf{b}}(\tau) &= \Sigma_{t=0}^{2^n-2}(-1)^{a_{t+\tau}+b_t} \\ &= -1 + \sum_{x \in \mathbb{F}_{2^n}} (-1)^{f(\lambda x)+g(x)} = -1 + C_{f,g}(\lambda) \end{aligned}$$

where $\lambda = \alpha^\tau$ with $0 \leq \tau \leq 2^n - 2$, τ is a phase shift of the sequence **a**, and α is a primitive element of F_{2^n}. So, the cross-correlation function $C_{f,g}(\lambda)$ of trace representations has the one-to-one correspondence with the cross-correlation function $C_{\mathbf{a},\mathbf{b}}(\tau)$ of the corresponding sequences.

2.1. General Properties of Periodic Correlations

The following are well-known properties of correlations of periodic sequences:

- The autocorrelation is an even function of τ, i.e. $C_{\mathbf{a}}(-\tau) = C_{\mathbf{a}}(\tau)$
- The peak autocorrelation occurs at zero delay $C_{\mathbf{a}}(0) \geq C_{\mathbf{a}}(\tau), \quad \tau \neq 0$.
- The crosscorrelation functions have the following symmetrical property $C_{\mathbf{a},\mathbf{b}}(\tau) = C_{\mathbf{b},\mathbf{a}}(-\tau)$
- Crosscorrelation functions are not, in general, even functions and their peak value is not necessarily at $\tau = 0$.
- If a and b are uncorrelated, i.e. $C_{\mathbf{a},\mathbf{b}}(\tau) = C_{\mathbf{b},\mathbf{a}}(\tau) = 0$, then $C_{\mathbf{c}}(\tau) = C_{\mathbf{a}}(\tau) + C_{\mathbf{b}}(\tau)$, for $c = a \pm b$.

- If a, b, c, d are sequences of period T, then

$$\sum_{\tau=0}^{T-1} C_{\mathbf{a},\mathbf{c}}(\tau)[C_{\mathbf{b},\mathbf{d}}(\tau+m)] = \sum_{\tau=0}^{T-1} C_{\mathbf{a},\mathbf{b}}(\tau)[C_{\mathbf{c},\mathbf{d}}(\tau+m)]$$

  Setting $c = d$ and $a = b$, we obtain

$$\sum_{\tau=0}^{T-1} C_{\mathbf{b},\mathbf{c}}(\tau)[C_{\mathbf{b},\mathbf{c}}(\tau+m)] = \sum_{\tau=0}^{T-1} C_{\mathbf{b}}(\tau)[C_{\mathbf{c}}(\tau+m)]$$

  and setting $m = 0$

$$\sum_{\tau=0}^{T-1} |C_{\mathbf{b},\mathbf{c}}(\tau)|^2 = \sum_{\tau=0}^{T-1} C_{\mathbf{b}}(\tau)[C_{\mathbf{c}}(\tau)]$$

- The periodic autocorrelation and cross-correlation functions of the product of periodic sequences of relatively prime lengths are themselves the products of the individual correlation functions. For simplicity, we only give the expression for autocorrelation. Let $\{a_n\}$ and $\{b_n\}$ be periodic sequences of periods T_1 and T_2 respectively, $gcd(T_1, T_2) = 1$, by repeating the sequence $\{a_n\}$ T_1 times and $\{b_n\}$ T_2 times, we can form a product sequence $\{c_n\}$ of period $T = T_1T_2$, where $c_n = a_nb_n$ and $C_{\mathbf{c}}(\tau) = C_{\mathbf{a}}(\tau)C_{\mathbf{b}}(\tau)$.

Example 2.9. Consider a finite field $\mathbb{F}_{2^3}$ generated by a primitive polynomial x^3+x+1. Then, the primitive element α is a root of the primitive polynomial, and $\alpha^7 = 1,\ \alpha^3 + \alpha + 1 = 0$. Let $Tr(x)$ be a trace function from $\mathbb{F}_{2^3}$ to $\mathbb{F}_2$. Then,

$$Tr(1) = 1 + 1 + 1 = 1,$$

$$Tr(\alpha) = \alpha + \alpha^2 + \alpha^4 = 0,\ Tr(\alpha^2) = Tr(\alpha^4) = Tr(\alpha) = 0,$$

$$Tr(\alpha^3) = \alpha^3 + \alpha^6 + \alpha^5 = 1,\ Tr(\alpha^6) = Tr(\alpha^5) = Tr(\alpha^3) = 1.$$

Hence, a binary sequence represented by $a_t = Tr(\alpha^t)$, $t = 0, 1, \ldots, 6$, is $\mathbf{a} = (1, 0, 0, 1, 0, 1, 1)$, which is a binary m-sequence of period 7. Consequently, $Tr(x)$ is a trace representation of a binary m-sequence $\mathbf{a}$ of period 7.

Example 2.10. In Example 2.9, let $\mathbf{a} = (1, 0, 0, 1, 0, 1, 1)$ and $\mathbf{b} = \mathbf{a}^{(3)} = (1, 1, 1, 0, 1, 0, 0)$. Then, their trace representations are given by $f(x) = Tr(x)$, $g(x) = Tr(x^3)$, respectively. By computation, the cross-correlations $C_{\mathbf{a},\mathbf{b}}(\tau)$, $\tau = 0, 1, \ldots, 6$ and $C_{f,g}(\lambda)$, $\lambda = \alpha^\tau$ are shown in Table 1.

Table 1: Cross-correlations $C_{\mathbf{a},\mathbf{b}}(\tau)$ and $C_{f,g}(\lambda)$

τ	**b**	$L^\tau(\mathbf{a})$	$C_{\mathbf{a},\mathbf{b}}(\tau)$	λ	$C_{f,g}(\lambda)$
0	1110100	1001011	-5	1	-4
1	1110100	0010111	-1	α	0
2	1110100	0101110	-1	α^2	0
3	1110100	1011100	+3	α^3	+4
4	1110100	0111001	-1	α^4	0
5	1110100	1110010	+3	α^5	+4
6	1110100	1100101	+3	α^6	+4

3. Homomorphisms between de Bruijn digraphs

A homomorphism H between two digraphs G_1 and G_2 is a function that preserves the structure of the digraph. That is, if (x_1, x_2) is an edge in G_1 then (Hx_1, Hx_2) is also an edge in G_2. For two positive integers n and k, [11] characterizes such homomorphisms and describes a family of homomorphisms from $B_{n+k}(q)$ to $B_n(q)$ whose inverse assigns to an arbitrary vertex disjoint path in $B_n(q)$ a constant number, q^k, of non-overlapping preimage paths in $B_{n+k}(q)$. We will say that such a homomorphism enjoys property D or simply is a D-homomorphism. The following theorem is proved in [12] and they characterize D-homomorphisms between de Bruijn digraphs.

Theorem 3.1. A map $D_{n,k} : B_{n+k}(q) \to B_n(q)$ has property D if

$$D_{n,k}(x_1, \ldots, x_{n+k}) = (d(x_1, \ldots, x_{k+1}), d(x_2, \ldots, x_{k+2}), \ldots, d(x_n, \ldots, x_{n+k}))$$

where $d(y_1, \ldots, y_{k+1})$ is any function that is one-to-one with respect to each of the variables y_1 and y_{k+1} when all other variables are kept fixed.

Thus, for the binary case, its simplicity allows for a more concise characterization of the shape of homomorphisms with property D.

Theorem 3.2. 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

$$d_k(x_1, \ldots, x_{k+1}) = x_1 + h(x_2, \ldots, x_k) + x_{k+1},$$

where $h(x_2, \ldots, x_k)$ is any Boolean function of $k-1$ variables.

Proof. By Theorem 3.2.4 in [12], 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 $d_k(x_1, \ldots, x_{k+1})$ is bijective in x_1 and in x_{k+1} then it satisfies the equations

$$\begin{aligned} d_k(\bar{x}_1, x_2, \ldots, x_{k+1}) &= 1 - d_k(x_1, x_2, \ldots, x_{k+1}) \\ &= d_k(x_1, x_2, \ldots, \bar{x}_{k+1}). \end{aligned}$$

So that $d_k(\bar{x}_1, x_2, \ldots, x_k, \bar{x}_{k+1}) = d_k(x_1, x_2, \ldots, x_k, x_{k+1})$. Therefore for each fixed set of values for $x_2, \ldots, x_k$, $d_k(x_1, \ldots, x_{k+1}) = d_{x_2,\ldots,x_k}(x_1, x_{k+1})$ 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. ■

We observe that this elegant form of d_k is mainly due to the "lack" of terms in $\mathbb{Z}_2$. For $k = 1$ there is exactly one homomorphism from $B_{n+1}(2)$ to $B_n(2)$ (up to bitwise complement) that has been used to construct de Bruijn sequences recursively, namely, $d(x_1, x_2) = x_1 + x_2$, which is known as Lempel's D-morphism. [12] investigates recursions of higher order ($k \geq 2$) as well as nonbinary de Bruijn digraphs.

The method relies on D-homomorphisms between de Bruijn digraphs of different orders that were defined above. Thus we treat the backbone generator as a cycle in a de Bruijn digraph $B_n(q)$ of an appropriate order and alphabet size. The inverse of the backbone generator by a D-homomorphism makes a large number of inverse sequences that all have the same size as the original cycle, where the former are regarded as paths in the higher order de Bruijn digraph. In other words, parallel streams are constructed so that the image by the D-homomorphism of each stream is the backbone generator sequence itself, thanks to property D, whose definition also allows to reformulate Theorem 3.1 as follows.

Corollary 3.3. Given a cycle c in $B_n(q)$ and a D-homomorphism $D_{n,k}$ from $B_{n+k}(q)$ to $B_n(q)$, there is a one-to-one correspondence between the set of ordered arrays of k numbers in base m and the set $\mathcal{P}(c)$ of pre-image sequences c' in $B_{n+k}(q)$ of c by $D_{n,k}$.

In terms of producing parallel streams, this means that for a given backbone generator that can be considered as a cycle in an appropriate de Bruijn digraph and an a priori fixed D-homomorphism D, a stream can be produced from each *seed* of k numbers as follows. The current state of a parallel stream is an ordered array of k values $s_1, \ldots, s_k$ consisting of the k most recent values generated. When the current backbone generator produces a new random number z in base m, the next random number in the stream is the number s where $D(s_1, \ldots, s_k, s) = z$. Note that since $s_1, \ldots, s_k$ and z are given, property D ensures that s is uniquely defined. In the next section we show how to select the function D to obtain efficient implementation. We will now illustrate the method with a toy binary example.

Example 3.4. Consider the function $d(x_1, x_2, x_3, x_4) = x_1 + x_2x_3 + x_4$ - which is a D-homomorphism by Theorem 3.2- and the sequence $\mathbf{b}_4 := [0000111101100101]$ which is a de Bruijn cycle of order 4 (wrapping it around, each word of size 4 occurs exactly once in $\mathbf{b}_4$). The inverse images of $\mathbf{b}_4$ started with all possible words of size 3 are given below, where the underlined prefixes are all the possible seeds of size 3.

<u>000</u>0000110011011010	<u>100</u>1001100001110011
<u>001</u>0010101000100001	<u>101</u>1101010110010110
<u>010</u>0100011010111100	<u>110</u>1110001111101000
<u>011</u>1011000101001111	<u>111</u>0111111100000101

Table 2: Homomorphisms with Minimum Peak Sidelobes for Parallel Inverse Images when $k = 4$.

Sequence Length (bits)	Minimum Peak Sidelobe	Homomorphisms
64	14	0001101; 0110010; 0111110
128	20	1000001
256	34	0110111
512	60	0001101; 0001110; 0101001; 0111101; 1110111
1024	102	0111110

Remark 3.5. It is worth noting that property D is essential for this method to work, and not just the homomorphism property. To illustrate this, consider the homomorphism $H_{n,k}$ from $B_{n+k}(q)$ to $B_n(q)$ for $k \geq 0$ and $n \geq 1$ introduced in [13], where

$$H_{n,k}(x_1, \ldots, x_{n+k}) = (x_{k+1}, \ldots, x_{n+k}).$$

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 Theorem 3.2, $d(y_1, \ldots, y_{k+1}) = y_{k+1}$, hence it does not enjoy property D. The q^k inverses of any cycle in $B_2(q)$ by $H_{2,k}$ disagree only in their seeds while the body of the sequences are all *equal* to the original cycle.

Corollary 3.3 suggests that we use for backbone generator a sequence that forms a de Bruijn cycle or at least a vertex disjoint cycle in a de Bruijn digraph of some order.

4. Sidelobe Analysis

In this section, we describe and analyze how we investigate all homomorphisms and then determine the non-trivial homomorphisms that give low correlation values between the 2^k inverse paths by the homomorphism $D_{n,k}$ of a binary de Bruijn cycle for small values of k defined in Section 3. We determine the upper bound on the magnitudes of the sidelobes for the cross-correlation of the parallel inverse images. The C++ program written by the author finds the maximum sidelobes for all the homomorphisms of fixed values of n and k and all de Bruijn cycles as backbone images.

4.1. Implementation

Given the current value z_i of the backbone sequence, we need to initialize an arbitrary binary vectors of k seeds $\{x_0, \ldots, x_{k-1}\}$ for a particular pre-image and constants $\{A_1, \ldots, A_{2^{k-1}-1}\}$ where $A_i \in \{0, 1\}$, thus the homomorphism is given by $A_1 A_2 A_3 \ldots A_{2^{k-1}-1}$.

Recall that

$$d(x_1, \ldots, x_{k+1}) = x_1 + h(x_2, \ldots, x_k) + x_{k+1},$$

Table 3: Homomorphisms with Minimum Peak Sidelobes for Parallel Inverse Images when $k = 5$.

Sequence Length (bits)	Minimum Peak Sidelobe	Homomorphisms
64	16	000000110011010; 000011011100110; 011110011000111; etc.
128	22	000001000001010; 000101111000110; 001000101010111
256	36	000001000001010; 000010001001010; 000011111110010; etc.
512	54	000100110111100; 001111110100101; 011101110000001; etc.
1024	80	010111110001000

Table 4: Homomorphisms with Minimum Peak Sidelobes for Parallel Inverse Images when $k = 6$.

Sequence Length (bits)	Minimum Peak Sidelobe	Homomorphisms
64	16	00000000000000000000000010000011
128	26	00000000000000000011001100110010
256	40	00000000000000000010010001100001; 00000000000000000011010100110100
512	58	00000000000000000000100001111001; 00000000000000000000100011001101
1024	88	00000000000000000011010010101001

so we need to encode $h(x_2, \ldots, x_k)$. For all $i = 1, \ldots, 2^{k-1} - 1$ and $j = 1, \ldots, k$; x_j is multiplied by the coefficient A_i if and only if $(j-1)^{st}$ entry in the binary expansion of i is 1.

For example, for $k = 4 : x_3$ and x_4 are multiplied by A_6 because the second and third places in the binary expansion 110_2 of 6 are both 1. The remaining products are obtained in a similar manner. Thus, we have

$$d(x_1, \ldots, x_5) = x_1 + A_1x_2 + A_2x_3 + A_3x_2x_3 + A_4x_4 + A_5x_5 + A_6x_3x_4 + A_7x_2x_3x_4 + x_5.$$

Hence, when $k = 4$, the next value of the pre-image is given by

$$\begin{aligned} x_{i+4} = z_i + x_i &+ A_1x_{i+1} + A_2x_{i+2} + A_3x_{i+1}x_{i+2} \\ &+ A_4x_{i+3} + A_5x_{i+1}x_{i+3} \\ &+ A_6x_{i+2}x_{i+3} + A_7x_{i+1}x_{i+2}x_{i+3} \end{aligned}$$

Table 5: Table of Good Binary Sequence Families

Name	**Length** N	**Family Size**	**Maximum Sidelobe**
Gold (odd)	2^n-1, n odd	$N+2$	$1+\sqrt{2}\sqrt{N+1}$
Gold (even)	2^n-1, $n=4k+2$	$N+2$	$1+2\sqrt{N+1}$
Kasami (small)	2^n-1, n even	$\sqrt{N+1}$	$1+\sqrt{N+1}$
Kasami (large)	2^n-1, $n=4k+2$	$(N+2)\sqrt{N+1}$	$1+2\sqrt{N+1}$
Bent	2^n-1, $n=4k$	$\sqrt{N+1}$	$1+\sqrt{N+1}$
No	2^n-1, $n=2k$	$\sqrt{N+1}$	$1+\sqrt{N+1}$
Gong	$(2^n-1)^2$	$\sqrt{N}$	$3+2\sqrt{N}$
Paterson, Gong	p^2, p prime 3 mod 4	$\sqrt{N}+1$	$3+2\sqrt{N}$
Paterson	p^2, p prime 3 mod 4	N	$5+4\sqrt{N}$
$\mathbb{Z}_4-linear$, family I	$2(2^n-1)$, n odd	$N/2+1$	$2+\sqrt{N+2}$
$\mathbb{Z}_4-linear$, family II	$2(2^n-1)$, n odd	$(N+2)^2/4$	$2+2\sqrt{N+2}$
Weil	p, prime	$(N-1)/2$	$5+2\sqrt{N}$

and total number of possible homomorphisms is equal to $2^7-1=127$.

We obtained computationally (using the C++ program) the maximum sidelobe for six different backbone sequences each for sequence lengths of 64bits, 128bits, 256bits, 512bits and 1024bits. The minimum ("best") of the maximum sidelobes were summarized in Tables (2, 3, 4) for values of $k=4, 5, 6$.

5. Conclusions

Since the Welch bound [3] for a family of M length-N sequences says that the maximum sidelobe magnitude is bounded below by

$$\sqrt{N}\sqrt{\frac{MN-N}{MN-1}}.$$

In particular, we see that $\sqrt{N}$ is the best that can be expected asymptotically.

Table 5 gives a summary of some of the better-know binary sequence families[14]. The pseudorandom binary sequences obtained by inverting one cycle in a de Bruijn digraph into many distinct sequences (parallel inverse images) in a higher order de Bruijn graph via an appropriate graph homomorphism are well-balanced with cross-correlation sidelobes that are bounded by $2.5\sqrt{N}$, where its family size 2^k for $k=5$. Hence, this correlation bound is comparable to the performance of the better known sequence families, and thus make the sequences potentially applicable.

References

[1] Udaya, P., and Siddiqi, M. U., 1996, "Optimal biphase sequences with large linear complexiy derived from sequences over Z_4," *IEEE Trans. Inform. Theory*, 42:206–216.

[2] Golomb, S. W., and Gong, G., 2005, "Signal Designs for Good Correlation: for wireless communications, cryptography and radar applications," Cambridge University Press.
[3] Welch, L. R., 1974, Lower bounds on the maximum cross correlation of signals, *IEEE Trans. Inform. Theory*, IT-20:397–399.
[4] Sarwate, D. V., 1979, Bounds on Crosscorrelation and Autocorrelation of Sequences, *IEEE Trans. Inform. Theory*, IT-25:720–724.
[5] Sidelnikov, V. M., 1991, On mutual correlation of sequences, *Soviet Math. Doklady*, IT-12(1):197–201.
[6] Massey, J. L., 1990, "On Welch's bound for the crosscorrelation of a sequence set," IEEE ISIT'90, San Diego, CA, USA, pp. 385.
[7] Barker, R., 1953, "Group Synchronizing of Binary Digital Systems," in Communication Theory, Jackson W., ed., Butterworth, London.
[8] Baumert, L. D., 1971, "Cyclic Difference Sets," ser. Lecture Notes in Mathematics, Springer-Verlag.
[9] Schmidt, B., 1999, "Cyclotomic integers and finite geometry," *J. Amer. Math. Soc.*, 12:929–952.
[10] Jungnickel, D., and Pott, A., 1999, Difference sets: An introduction, in Difference Sets, Sequences and their Correlation Properties, A. Pott, P. V. Kumar, T. Helleseth and D. Jungnickel, Eds. NATO Science Series C, pp. 259–296.
[11] Alhakim, A., and Akinwande, M., 2009, "A mutiple stream generator based on de Bruijn digraph homomorphisms," *J. Stat. Comp. Simul.*, 79(11):1371–1380.
[12] Akinwande, M. B. O., 2010, "Homomorphisms of Nonbinary de Bruijn Graphs with Applications," Ph.D. dissertation, Clarkson University, New York.
[13] Chen, C., and 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.
[14] Rushanan, J., 2006, "Weil Sequences: A Family of Binary Sequences with Good Correlation Properties," ISIT, Seattle, USA, July 9-14, pp. 1648–1652.