On the Aperiodic Autocorrelation of Binary Sequences

Raymond A. Kristiansen

March 7, 2003

Empty page

Abstract

This master thesis (hovedfag) looks at the aperiodic autocorrelation of binary sequences. We give an overview of search techniques and classes of sequences with low aperiodic autocorrelation sidelobes, and present two new classes. One of them is the extended Legendre semi-construction that appears to have a MF > 6.3 for large lengths. We also look at the multidimensional aperiodic autocorrelation, and present a construction for a new class of sequences with a very low multidimensional aperiodic autocorrelation.

Keywords: aperiodic autocorrelation, Golay-Rudin-Shapiro sequences, Legendre sequences, Merit Factor, multidimensional autocorrelation, sum-of-squares.

Acknowledgment

I would like to thank Matthew G. Parker for all his help and suggestions during the process. I am also grateful to my supervisor Tor Helleseth and the Coding Theory and Cryptography group at the Department of Informatics, University of Bergen.

I would also like to thank Kristine Jørgensen for correcting all my typos.

Contents

Chapter 1

Introduction

Binary sequences with low aperiodic autocorrelation play an important role in many communication engineering problems such as spread-spectrum transmission techniques. But finding these binary sequences has generally been recognised as a difficult problem, significantly more difficult than finding binary sequences with low periodic autocorrelation. The problem resembles the old problem of finding the needle in a haystack, since the number of sequences with high aperiodic autocorrelation outnumbers those of low aperiodic autocorrelation. In the last 50 years or so, researchers have developed techniques to find binary sequences with low aperiodic autocorrelation. The first part of this thesis looks at some of these techniques, and tries to develop new techniques. In particular, we develop a new semi-constuction for long binary sequences with aperiodic Merit Factor > 6.3, which for the long lengths constructed is a higher Merit Factor than for any other long binary sequence known in the literature.

A new field that has not been looked into before is that of multidimensional aperiodic autocorrelation. In fact, the periodic sum of squares measure, which is derived from multidimensional periodic autocorrelation, has been used as a measure of cryptographic strength for binary cryptosystems. However, the aperiodic sum of squares and its related multidimensional aperiodic autocorrelation have, to our knowledge, never been studied before. The second part of this thesis takes a closer look at the multidimensional aperiodic autocorrelation using algebraic normal form. We look at the multidimensional properties of Golay-Rudin-Shapiro sequences and present a construction for a new class of sequences with a low multidimensional aperiodic autocorrelation.

1.1 Definitions

An autocorrelation function (ACF) is a function that measures the self-similarity of a binary sequence. There are three common types of autocorrelation functions, periodic, negaperiodic and aperiodic, although the periodic case is the most studied. The periodic autocorrelation function (PACF) will measure the correlation of the sequence with a cyclic shift of itself. Let s be a binary sequence of length N, such that $s = \{s_0, s_1, ..., s_{N-1}\}$, $s_i \in Z_2$, and $s_i = 0$, $0 > i \geq N$.

Then the periodic autocorrelation function is defined as

$$\mathrm{PACF}_k(\mathbf{s}) : c_k = \sum_{i=0}^{N-1} (-1)^{s_i - s_{i+k}}, \qquad 0 \le k < N \tag{1.1}$$

The parameter N represents the period of the cyclic shift, k is the shift index at which the sequence is compared to itself, and the sequence indices, i, are taken mod N. Another measure of the autocorrelation is the negaperiodic autocorrelation (NACF). NACF will measure the correlation of the negacyclic shift of a sequence against itself. The function is defined as

$$\mathrm{NACF}_k(\mathbf{s}) : n_k = \sum_{i=0}^{N-1} (-1)^{s_i - s_{i+k} - \lfloor \frac{k+i}{N} \rfloor}, \qquad 0 \le k < N \tag{1.2}$$

N now represents half the period of the sequence. Once again the indices i are taken mod N. A third way to measure the autocorrelation is a combination of periodic and negaperiodic called the aperiodic autocorrelation function (AACF). When we combine the periodic and the negaperiodic shifts of the sequence we lose the circularity of the shifted sequence. Instead of shifting the sequence cyclically, we compare the window of indices where both the shifted sequence and sequence itself exist. The AACF is then defined as

$$\mathrm{AACF}_k(\mathbf{s}) : a_k = \sum_{i=0}^{N-k-1} (-1)^{s_i - s_{i+k}}, \qquad 1 \le k < N \tag{1.3}$$

We can also represent PACF, NACF and AACF as polynomial multiplication. A binary sequence $s = (s_0, s_1, s_2, ..., s_{N-1})$ can be associated with the polynomial $s(x) = s_0 + s_1 x + s_2 x^2 + ... + s_{N-1} x^{N-1}$. We can then define

$$PACF(\mathbf{s(x)}) = s(x)s(x^{-1}) \pmod{x^N - 1} \tag{1.4}$$

$$NACF(\mathbf{s(x)}) = s(x)s(x^{-1}) \pmod{x^N + 1} \tag{1.5}$$

$$AACF(\mathbf{s(x)}) = s(x)s(x^{-1}) \tag{1.6}$$

This representation includes all shifts $-N < k < N$, where the autocorrelation for the kth shift is the coefficient for x^k.

Example:

Let $N = 5$ and $s = 01011$ the PACF_k is $\{5, -3, 1, 1 -3\}$, NACF_k is $\{5, -1, 1, -1, 1\}$ and the AACF_k is $\{5, -1, 1, 0, -1\}$ as seen below :

k	Cyclic shift	PACF_k	Negacyclic shift	NACF_k	Acyclic shift	AACF_k
0	01011	5	01011	5	01011	5
1	10101	-3	00101	-1	0101	-2
2	11010	1	00010	1	010	1
3	01101	1	10001	-1	01	0
4	10110	-3	01000	1	0	-1

□

A common metric used to measure binary sequences with low autocorrelation is the Golay Merit Factor (MF) proposed by Golay [5].The periodic Merit Factor of a binary sequence s, of length N is given by

$$\mathrm{MF}_p(\mathbf{s}) = \frac{N^2}{2\sum_{k=1}^{N-1} c_k^2} \tag{1.7}$$

and the aperiodic Merit Factor of a binary sequence s of length N is given by

$$\mathrm{MF}_a(\mathbf{s}) = \frac{N^2}{2\sum_{k=1}^{N-1} a_k^2} \tag{1.8}$$

Unless otherwise noted, we will only look at the aperiodic autocorrelation a_k and the corresponding aperiodic Merit Factor. The higher the Merit Factor the lower the aperiodic values a_k , $1 \leq k < N$. Note that the trivial case where $a_0 = N$ is not used in the calculations of the Merit Factor. The optimal Merit Factor for a binary sequence s of odd length N is obtained when the AACF values are of the form $a_k = \{N, 0, \pm 1, ..., \pm 1, 0, \pm 1\}$, and for N even $a_k = \{N, \pm 1, 0, \pm, ..., \pm 1, 0, \pm 1\}$. This translates to a very loose upper bound on the aperiodic Merit Factor of a binary length N sequence of N or $\frac{N^2}{N-1}$ for N even or odd respectively.

We can also speak of the asymptotic aperiodic Merit Factor for a class of sequences. Let $\mathcal{C}$ be a class of sequences, and let $S_N \in \mathcal{C}$ be a sequence of length N. The asymptotic Merit Factor for the class $\mathcal{C}$ is then

$$\lim_{N\to\infty} \mathrm{MF}_a(\mathbf{S_N}) = \mathcal{F}_{\mathcal{C}} \tag{1.9}$$

Another metric measure for binary sequences with low autocorrelation is the sum-of-squares, σ_a, given by

$$\sigma_a(\mathbf{s}) = \sum_{k=1}^{N-1} |a_k|^2 \tag{1.10}$$

We see that σ_a is the major part of the Merit Factor function (1.8), and the Merit Factor function can be written with the use of σ_a

$$\mathrm{MF}(\mathbf{s}) = \frac{N^2}{2\sigma_a} \tag{1.11}$$

There is also a relation between the Merit Factor and the spectral properties of the signal corresponding to the sequence [13]

$$\sigma_a = \frac{1}{4\pi} \int_0^{2\pi} (|S(e^{i\omega})|^2 - N)^2 \, \mathrm{d}\omega, \tag{1.12}$$

where $S(e^{i\omega})$ is the Fourier Transform of the sequence s, and i is the root of -1. In other words, σ_a determines the mean-square derivation from the flat spectrum. (1.12) is significant because it shows how a complicated continuous integral can be computed in terms of a relatively simple discrete summation.

Example:
If $N = 5$ and $s = 01011$ the a_k-values are $\{5, -2, 1, 0, -1\}$ (see Example above). The sum of squares is then

$$\sigma_a = (-2)^2 + (1)^2 + (0)^2 + (-1)^2 = 4 + 1 + 1 \ = 6.$$

The Merit Factor for this sequence is then

$$\mathrm{MF}(01011) = \frac{25}{2(6)} = 2.08$$

□

In most of this thesis we will refer to the aperiodic sum-of-squares, σ_a, simply by the symbol σ.

1.2 Previous Work

Merit Factor

In 1977 Golay [5] introduced a criterion of goodness for low aperiodic autocorrelation binary sequences as an alternative to the minimal peak sidelobes, called the aperiodic Merit Factor (see equation 1.8). The Merit Factor is a way to measure the overall out-of-phase aperiodic autocorrelation for a binary sequence. Golay defined the Merit Factor such that the MF of a random binary sequence should be around 1 with a high probability, and this was also proven later by Høholdt [12]. Golay also established a conjecture for an upper bound for the Merit Factor

$$\mathrm{MF} \leq 12.32 \qquad \forall \ \ N \neq 13$$

that will be valid for all binary sequences except the Barker sequence of length $N = 13$ which has the highest known Merit Factor: 14.08. A binary sequence such that the aperiodic autocorrelation $a_k \in \{-1, 0, 1\}, \forall \ k \neq 0$, is called a Barker sequence and has a maximal Merit Factor. It is conjectured that Barker sequences exist only when N is prime and $N \leq 13$, and Storer and Turyn [25] proved this Barker conjecture for all odd N.

Apart from a few exceptions, exhaustive search has been the only way to find sequences with high Merit Factor. Since the complexity of an exhaustive search is $\mathrm{O}(2^N)$ the sequence length is limited by the current computation power. Turyn (see Golay [6]) did an exhaustive search up to length $N = 32$, and Lindner (see Luke [16]) up to length $N = 40$. In 1996 Mertens [18] did an exhaustive search for binary sequences for lengths up to 48. He used a new search algorithm that lowers the exponential configuration space from $\mathrm{O}(2^N)$ to $\mathrm{O}(1.85^N)$. In order to do this he used the symmetry of the aperiodic autocorrelation function and a branch and bound technique together with parallelization. With this algorithm he managed to do an exhaustive search for all lengths up to $N = 48$ using 313 hours of CPU time on a computer with 4 CPUs. He also estimated that the optimal MF for a large length sequence will be > 9.0. In 2002 Mertens [19] has extended his search up to length $N = 58$. The largest size ($N = 58$) took about two weeks on 156 CPUs (mostly PIII, 800 MHz). The result of his search can be found in Table 2.2 (page 13).

Skewsymmetric sequences

A skewsymmetric sequence of length $N = 2n + 1$ is found by interlacing a symmetric sequence $Ac\overline{A}$ of length $n + 1$ and an antisymmetric sequence $B\overline{B'}$ of length n, where the overbar indicates sequence reversal and the prime indicates that each sequence element is complemented. (see Section 2.8). Based on the observation that all odd Barker sequences are skewsymmetric, Golay [5] suggests a sieve to limit the search for binary sequences with high Merit Factor to skewsymmetric sequences. A search for all skewsymmetric sequences up to length $N = 59$ shows that skewsymmetric sequences have a high Merit Factor.

In 1990 Golay [8] extends his previous work on skewsymmetric binary sequences. Here he does a limited search for skewsymmetric sequences of length up to $N = 117$. Golay searches for symmetric and antisymmetric sequences with Merit Factor > 1.0, and the result of this search found MF > 8 for many sequences, and even some with MF > 9.

Another search for binary sequences with high MF where skewsymmetric sequences are used is described in [20]. In [20] the authors use an evolutionary search (somewhat similar to the directed search of this report in Section 2.4) on skewsymmetric sequences. They manage to find some good sequences of length up to 201 with MF > 7. The evolutionary searches take μ skewsymmetric starting sequences, and for generation/iteration find λ new skewsymmetric sequences by flipping $n > 1$ bits in the parents. The new sequence will be accepted if some of the highest a_k values have decreased, otherwise another $n > 1$ bits are flipped. The authors also show that the Merit Factor of a random skewsymmetric sequence is higher than that of a random sequence, which indicates that skewsymmetric sequences have high MF, but only $\frac{1}{3}$ of the optimal Merit Factors found are associated with skewsymmetric sequences.

Classes of Sequences

It has been shown that some of the known classes of sequences also have a high asymptotic Merit Factor. One class of sequences that can be constructed from the Hadamard difference set is the maximal length shift register sequences (m-sequences). In [13] Jensen, Jensen and Høholdt show that the asymptotic Merit Factor of any m-sequence of length $N = 2^n - 1$ is three.

Another class of sequences with a high Merit Factor is the modified Jacobi sequence. A modified Jacobi sequence is a binary sequence $S = (s_0, s_1, ..., s_{N-1})$ of length $N = pq$, where $p < q$ are distinct odd primes. A special case of the modified Jacobi sequences is when $q = p + 2$, which produces a Twin-Prime sequence. In [13] Jensen, Jensen and Høholdt show that a Twin-Prime sequence shifted by $\frac{1}{4}$ has an asymptotic Merit Factor of six. They have also shown that for any sufficiently large $N = pq$ it is possible to construct a Jacobi or modified Jacobi sequence of length N with asymptotic Merit Factor six, if p and q satisfy the condition

$$\frac{(p+q)^5 log^4 N}{N^3} \rightarrow 0, \quad \text{for } N \rightarrow \infty$$

A third class of sequences that can also be constructed from the Hadamard difference set is the Legendre sequence. In 1983 Golay [7] proved that a Legendre sequence offset by a fraction f of its length N has an asymptotic Merit Factor $\mathcal{F}$ that could be found by

$$\frac{1}{\mathcal{F}} = (\frac{2}{3} - 4|f| + 8f^2), \quad |f| \leq \frac{1}{2} \tag{1.13}$$

which gives an asymptotic MF of 6 when $|f| = \frac{1}{4}$. An offset sequence is one in which a fraction of f bits of the sequence is chopped off the end of the Legendre sequence and appended at the other, in other words a cyclic shift of fN places. Golay proved this using probability theory and an "external" assumption. This assumption was that, for the asymptotic case, one can consider the correlation values a_k to be independent random variables, which according to [13] they are not for a fixed N. In [11] Høholdt and Jensen give a proof that (1.13) is true, without the assumption Golay used. They also conjecture that the best possible asymptotic value of the Merit Factor is six for a given construction, and that the shifted Legendre sequences therefore are optimal.

Kirilusha and Narayanaswamy [14] used Legendre sequences in an attempt to find new constructions of binary sequences with high Merit Factor based on known constructions. They hoped to find a class of sequences with an asymptotic Merit Factor above 6.0, but they did not find such a class. What they did find was that adding ± 1 to the front of a shifted Legendre sequence did not change the asymptotic Merit Factor of 6.0. Extending this they showed that adding $u < O(N^{\frac{1}{2}})$ bits in front of a shifted Legendre sequence would not change the asymptotic Merit Factor if the u new bits were taken from the end of the shifted Legendre sequence itself. They also showed that by adding the last bits of any sequence to the front of the same sequence, the aperiodic autocorrelation of the new sequence is actually related to the periodic autocorrelation of the original sequence.

Another class of sequences that turns out to have high Merit Factor is an infinite family of sequences with low negaperiodic autocorrelation [21]. Two of these constuctions of even length binary sequences also have a high asymptotic Merit Factor of six. This class of sequences has length $N = 2p$ (see section 2.6).

A class of sequences that also has a high asymptotic Merit Factor was independently discovered by Golay [3] and Rudin-Shapiro [24]. The class of Golay Complementary Sequences (GCS) is known to exist for lengths $N = 2^\alpha 10^\beta 26^\gamma$, $\alpha, \beta, \gamma \geq 0$ [26]. Golay Complementary Pairs are binary sequences a and b that satisfy the condition

$$a_k(a) + a_k(b) = 0, \quad k \neq 0$$

where a_k is the aperiodic autocorrelation defined in (1.3). Generation of GCP of length $2N$ can be done from a GCP of length N using a recursion

$$(a, b) \rightarrow (a|b, a|b')$$

where '|' means concatenation and b' means the complement of sequence b. This is in fact the well-known Golay-Rudin-Shapiro recursion that gives the name for this class of sequences : Golay-Rudin-Shapiro sequences (GRS). In [1] Davis and Jedwab showed that GRS sequences can also be represented using Algebraic Normal Form (ANF) (see Section 3.3)

$$\mathrm{s(x)} = (\sum_{j=0}^{n-2} x_{\pi(j)} x_{\pi(j+1)}) + (\sum_{j=0}^{n-1} b_j x_j) + d, \qquad b_j, d \in Z_2 \tag{1.14}$$

where π is any permutation of Z_n, $x_{n-1}, x_{n-1}, ..., x_0$ are boolean variables and s is a length $N = 2^n$ binary sequence such that

$$s_i = s(x_{n-1} = i_{n-1}, x_{n-2} = i_{n-2}, ..., x_0 = i_0)$$

where $i_{n-1}, i_{n-2}, ..., i_0$ is the binary representation of the integer i. This representation defines a class of binary GRS sequences of length $N = 2^n$, $\mathcal{C}$ of size $2^{n+1} \cdot \frac{m!}{2}$. In [9] Jensen, Jensen and Høholdt show that the asymptotic Merit Factor of a GRS sequence of length $N = 2^n$ is three for a subset of the sequences in $\mathcal{C}$. This subset of sequences contains 2^{n+1} of the sequences in $\mathcal{C}$ and can be constructed using (1.14) without the permutation π. The authors also give a proof that the aperiodic autocorrelation for this subset of binary GRS sequences of length $N = 2^n$ follows the recursion

$$\sigma_n = 2\sigma_{n-1} + 8\sigma_{n-2}$$

where σ_n is the same as the sum-of-squares σ_a defined in (1.10).

Multidimensional autocorrelation

In addition to the 1-dimensional autocorrelation, one can also define a multidimensional two-point autocorrelation function for binary sequences of length $N = 2^n$. The multidimensional periodic sum-of-sqares, defined the same way as the 1-dimensional using the multidimensional periodic autocorrelation function, has been used in cryptography [17]. This multidimensional periodic autocorrelation is used to assess cryptographic strength of a boolean function.

Chapter 2

Sequences with good aperiodic Merit Factor

Finding binary sequences with good aperiodic Merit Factor (MF) is usually not an easy thing to do. There are two main ways of finding them. The first is to do an exhaustive search for sequences with good Merit Factor, and the second is to construct the sequences such that the MF is known to be high.

2.1 Random Distribution of the Merit Factor

When Golay defined the Merit Factor in [5] he constructed the measurement such that for large N the expected Merit Factor of a random sequence is 1. This was also proven by Høholdt in [12]. This means that if you take a random sequence of sufficient length, the Merit Factor for this sequence will be around 1 with a high probability. Based on this we would expect the random distribution of the Merit Factor to be centred around 1.

Fig. 2.1 show the distribution of 2^{18} random sequences of length $N = \{64, 256, 257, 1024\}$. From the diagrams it looks like the distribution is more concentrated around 1 when the sequence length increases. This would suggest that it will be much harder to find sequences with high Merit Factor of larger length, not only because there are more sequences, but also because the number of good sequences does not appear to increase with the sequence length. But the distribution seems to be independent when it comes to the number of prime factors in the sequence length. There is almost no difference in the distribution for length $N = 256$, an even number with factors $(2 \cdot 2 \cdot\ 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2)$, and length $N = 257$, an odd number with only one factor (257).

2.2 Exhaustive search for the best MF

We know that a random binary sequence is expected to have a Merit Factor of 1. The question is now how to find sequences with high MF, or even the sequences with the highest MF. But finding the best sequences from the 2^N different sequences of length N is a not an easy task. Because each autocor-

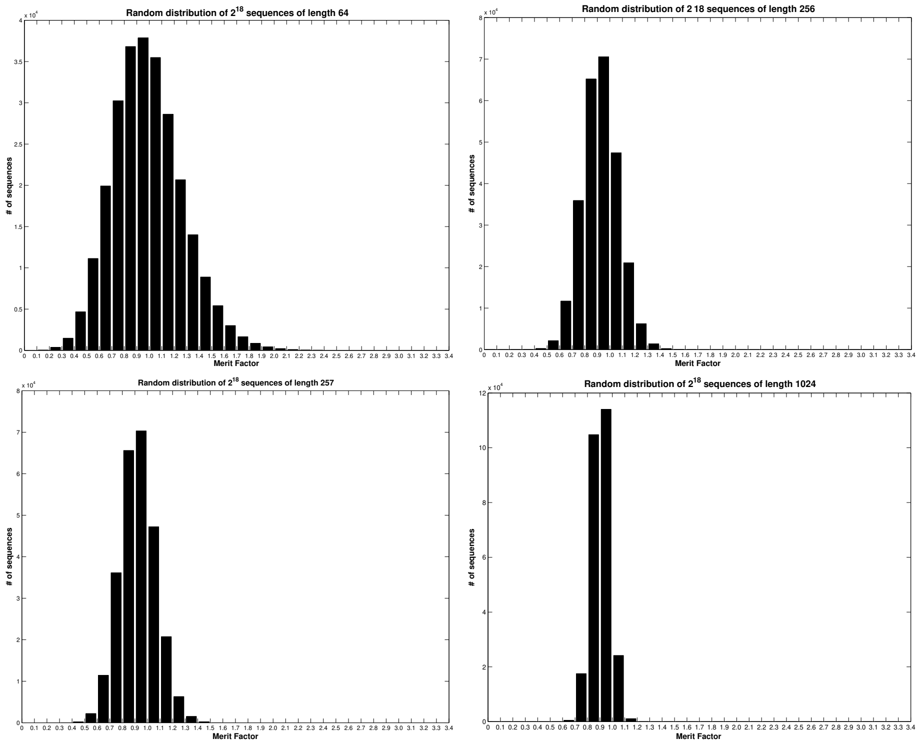

Figure 2.1: Random distribution of the Merit Factor

relation function a_k contributes quadratically to the sum-of-squares σ_a, (see equation 1.11) the existence of a single large a_k can reduce the Merit Factor of a sequence drastically. The a_k values are not independent and therefore each change in the sequence leading to an improvement of one a_k value will change the other a_k values as well. This means that the only way to find the sequence with the optimal MF appears to be by using an exhaustive search through all the different sequences. The complexity of this type of search is non-polynomial in N, specifically $O(N^2 \cdot 2^N)$, where for each of the 2^N sequences the MF computation requires $O(N^2)$.

This exponential complexity will prevent an exhaustive search for large N due to the computational resources needed for the search. Mertens has in [18] done a search up to $N = 48$ using a parallel algorithm with 4 computers, and by 2002 he has managed to get as high as $N = 58$ [19]. The exhaustive search for the best MF of length $N = 58$ took two weeks on 156 CPUs (mostly PIII, 800 MHz). The optimal sequences of length $27 \leq N \leq 58$ from this search can be found in Table 2.2. Sequences are written in run-length notation where each figure indicates the number of consecutive elements with the same sign.

Example:

Let 311 be the run-length notation for a sequence s. The binary form of s would then be $s = 00010$ if we start with the first elements in the alphabet, which is 0 for the binary case.

$\square$

Length, N	MF	σ_a	Sequence
27	9.85	37	34313131211211
28	7.84	50	34313131211212
29	6.78	62	212112131313431
30	7.63	59	551212111113231
31	7.17	67	7332212211112111
32	8.00	64	71112111133221221
33	8.51	64	742112111111122221
34	8.89	65	842112111111122221
35	8.39	73	7122122111121111332
36	7.90	82	3632311131212111211
37	7.96	86	844211211111122221
38	8.30	87	8442112111111122221
39	7.68	99	8212112123432111111
40	7.41	108	44412112131121313131
41	7.78	108	343111111222281211211
42	8.73	101	313131341343112112112
43	8.48	109	1132432111117212112213
44	7.93	122	525313113111222111211121
45	8.58	118	8212112123123432111111
46	8.08	131	823431231211212211111111
47	8.18	135	923431231211212211111111
48	8.23	140	3111111832143212221121121
49	8.83	136	215131311224112241141141
50	8.17	153	215131311224112241141142
51	8.50	153	23432111141313116212112121
52	8.14	166	51161212121111131223123332
53	8.26	170	4511311133251312221112111121
54	8.33	175	356225141212112222111111121
55	8.85	171	9212123212114321233211111111
56	8.17	192	7612231123241111132112122111
57	8.64	188	33232631111127121111221221211
58	8.54	197	1111131232138142121132432112

Table 2.1: Mertens [18][19] search for optimal MF for $27 \leq N \leq 58$

2.2.1 Modifications to speed up the search

There are a few ways to speed up the search a little, but there is no way to do anything about the non-polynomial complexity of this type of search. The first modification that can be done is to limit the search to the first half of the sequences, since the MF of two sequences related by negation will always be the same. For example the MF of 11001 will be the same as the MF of 00110. The proof for this comes from (1.3) : the aperiodic autocorrelation function only compares bits of the sequence pairwise, so if you compare the negation of the two bits, the result is the same :

$$\text{AACF}(\mathbf{s}) : a_k = \sum_{i=0}^{N-k-1} (-1)^{s_i + s_{i+k}} = \sum_{i=0}^{N-k-1} (-1)^{s'_i} + s'_{i+k}, \qquad 1 \leq k < N$$

where prime indicates that the element is complemented. For an exhaustive search with a lexicographic ordering it is straightforward to show that the last half of the sequences of length N will be a negation of the first half, based on the fact that the high order bit s_{N-1} will be 0 for the first half and 1 for the last half. Therefore in a search for sequences of length N it will be enough to search through the first 2^{N-1} sequences and still be able to find the highest MF.

The second modification that can be done to speed up the program is to stop the calculation of the MF if we can see that the new MF cannot meet a MF threshold. Let $\mathcal{F}$ be a threshold for the MF of sequences of length N, and let $\overline{\sigma}$ be the sum-of-squares that satisfies

$$\overline{\sigma} = \frac{N^2}{2\mathcal{F}}, \quad \mathcal{F} > 0$$

Then let s_i be the ith sequence of length N in the search. After calculation of $a_k(s_i)$ for $k \leq m$ the best possible aperiodic autocorrelation values for s_i will be $(a_0, a_1, ..., a_{m-1}, ..., \pm 1, 0, ..., \pm 1, 0, \pm 1)$ (see Section 1.1). Let σ_m be the sum-of-squares for the first m aperiodic autocorrelations of s_i

$$\sigma_m(s_i) = \sum_{k=1}^{m} |a_k|^2$$

Therefore if $\sigma_m(s_i) > \overline{\sigma} - \lceil \frac{N-m}{2} \rceil$ we know that

$$\text{MF}(s_i) < \mathcal{F}$$

and therefore we can drop the calculation of a_k for $k > m$. The question now is how to choose $\mathcal{F}$. Looking at Table 2.2 one could argue that $\mathcal{F} = 7.0$ is a good choice, but this could result in an inconclusive result as it is quite possible that a sequence with MF $> \mathcal{F}$ does not exist. A better choice for $\mathcal{F}$ is to choose the best MF found so far in the search. This way we will not calculate the MF for all the possible codewords, but it is nevertheless an exhaustive search over all possible sequences and we will find the highest MF.

The result of a speed comparison with the two modifications implemented can be seen in Table 2.2. As expected the first modification only takes half the time to compute since we are only searching through half the sequences. It looks like the second modification uses only $\frac{1}{4}$ of what the original search did at first, but when the sequence length is $N = 30$ it is about $\frac{1}{6}$. This would suggest that the second modification improves as the lengths gets longer. When we combine the two modifications the time used will be $< \frac{1}{8}$. Unfortunately $\frac{1}{8}$ of $O(2^N \cdot N^2)$ is stil non-polynomial complexity, but at least we will be able to extend our exhaustive search to $N+3$ compared to the original problem.

Length, N	Normal	with 1.mod.	with 2.mod.	with both mod.
15	0	0	1	0
16	0	0	0	0
17	0	0	0	1
18	0	1	0	0
19	1	1	1	0
20	3	2	1	0
21	8	3	2	1
22	16	8	3	2
23	35	18	8	4
24	76	38	14	7
25	165	82	28	14
26	354	177	65	33
27	762	382	116	60
28	1639	842	297	143
29	3524	1764	643	323
30	7547	3780	1216	612

Table 2.2: Speed comparisons for search space reduction (mod.1) and thresholding (mod.2)

2.3 m-sequence ordering

One problem with an exhaustive search is that it has to calculate the Merit Factor for each of the 2^N sequences of length N. The next step was therefore to find a way to use the a_k values of one sequence in the calculation for the next, in order to speed up the search. The normal approach is to go through the sequences in a lexicographic word order, where we start on the sequence representing 0 (00...00), up to the sequence representing $N-1$ (11...11). With this word ordering there was only a few reoccuring bits in the calculation of the a_k for two adjacent sequences.

An alternative to lexicographic word ordering is to use an maximal length shift register sequence (m-sequence) ordering of the sequences. An m-sequence of length $M = 2^N - 1$ will be an ordering of all different overlapping sequences of length N, except the all zero sequence. The Merit Factor of the m-sequence itself has also been shown to be good [13]. An m-sequence of length $M = 2^N - 1$ will have an asymptotic Merit Factor of three.

Maximal length shift register sequences can be generated with the use of primitive polynomials. Let $p(x)$ be a primitive polynomial of degree N. Then let

$$h_i(x) \equiv x^i \bmod p(x), \quad 0 \le i \le 2^N - 1 \tag{2.1}$$

and let $h_{i,0}, h_{i,1}, ..., h_{i,N-1}$ be the coefficents of $h_i(x)$ such that $h_i(x) = \sum_{t=0}^{N-1} h_{i,t}x^t$. We can then define a sequence of length $M = 2^N - 1$, by $m = (h_{0,t}, h_{1,t}, ..., h_{2^N-1,t})$, for an arbitary $0 \leq t \leq N-1$.

Example:

Let $p(x) = x^4 + x + 1$ be a primitive polynomial of degree 4. For $t = 0$ m is a length $M = 15$ m-sequence defined by $h_{i,0}$ for $0 \leq i \leq 14$

$$m = (h_{0,0}, h_{1,0}, ..., h_{14,0}) = 100011110101100$$

We can then use m to run through all the sequences of length 4 (except 0000). The first sequence will be bit 0 to 3 of m, the second will be bit 1 to 4, and so on. The order of the 15 sequences will be

$$1000, 0001, 0011, 0111, 1111, 1110, 1101, 1010, 0101, 1011, 0110, 1100, 1001, 0010, 0100.$$

Note that the last 3 sequences are the results of a wraparound of the m-sequence since the m-sequence is cyclic.

□

2.3.1 Update rule

Now that we have a new ordering of our search we need to find a way to find the aperiodic autocorrelation $a_k(s_i)$ based on $a_k(s_{i-1})$ for the same shift k. A closer look at an m-sequence of length $M = 2^N - 1$ will show that each subsequence of length N is a result of a cyclic or negacyclic shift of the subsequence before.

Example:

Let $m = 100011110101100$ be an m-sequence of length $M = 2^4 - 1 = 15$ and let $s_0 = 1000$ and $s_2 = 0011$ be two subsequences of length $N = 4$. A cyclic shift of s_0 would result in the sequence $s_1 = 0001$ and a negacyclic shift of s_2 result in the sequence $s_3 = 0111$.

□

This connection between adjacent sequences is something that can be used to speed up our calculation of each a_k for $1 \leq k \leq N$. A closer look at the a_k values for two adjacent sequences shows that most of the bit comparisons are the same. Take a look at an example :

Example:

Let $s_i = abcde$ be a binary sequence generated in our m-sequence ordering as described above. Let $s_{i+1} = bcdea'$ be the next sequence generated. This sequence can be viewed as a cyclic or negacyclic shift of s_i depending on whether $a' = a$ or $a' = \overline{a}$. Let us calculate the a_1 value for s_i and a'_1 for s_{i+1}, $a_1 = ab + bc + cd + de$ and $a'_1 = bc + cd + de + ea'$. Since $a_1 - a'_1 = (ab - ea')$ we can express a'_1 with a_1, as $a'_1 = a_1 + a'e - ab$. We can easily see that something similar will hold for the rest of the a_i-values.

□

As we can see in the example, we do not need to completely recalculate $a_k(s_{i+1})$ since if we know the values of $a_k(s_i)$, we only need to update these to calculate $a_k(s_{i+1}$. The algorithm for this update rule can be found in Algorithm 1.

INPUT :	$\mathbf{s}_i$	=	sequence
	N	=	sequence length
	a[ ]	=	an array of a_k for s_{i-1}
	last	=	old last bit
OUTPUT :	**a**[ ]	=	an updated array of a_k for s_i
	mf	=	merit factor of s

1 for k = 1 to k = N - 2 do
        if (s[N-1] == last) remove = 1
        else remove = -1
        if (s[0], s[k]) add = 1
        else add = -1
        a[k] = a[k] + *add* - *remove*
        $\sigma = \sigma + a[k]^2$
2 if (s[0] == s[N-1]) a[N-1] = 1
    else a[N-1] = - 1
3 $\sigma = \sigma + a[N-1]^2$
4 $mf = \frac{N^2}{2 \cdot \sigma}$

Algorithm 1: MF computation using the update rule

2.3.2 Modification to the m-sequence ordering

As with the lexicographic ordering, there are some modifications that could be done with the m-sequence ordering to speed up the search. Let us first look at the modifications done to the lexicographic ordering (see Section 2.2.1) and see if these can be applied here.

The first of the two modifications for lexicographical word ordering was to drop the last half of the sequences because they are just the negations of the first half. This is not possible for the m-sequence ordering because the last half is not the negation of the first. To show this it is enough to find an example where the last half is *not* a negation of the first :

Example:

Let $p(x) = x^5 + x^2 + 1$ be a primitive polynomial that generates the m-sequence $\mathcal{M}$ = 0000100101100111110001101110101 of length 31. Then $\mathcal{M}$ generates all the binary sequences of length $N = 5$, s_0 = 00000. As one can see s_4 = 01001 and s_8 = 10110. These two opposite sequences are found in the first half of the m-sequence ordering. Also in the last half s_{18} = 10001 and s_{24} = 01110 are two opposite sequences.

□

This shows that it is not possible to drop the last half of the sequences in an m-sequence ordering, and it therefore prevents us from using this type modification to the m-sequence ordered search.

The second modification for lexicographic ordering is to stop the calculation of the Merit Factor at a certain threshold Merit Factor. This type of threshold is not possible in the m-sequence ordering because the calculation of the $a_k(s_{i+1})$ is based on the $a_k(s_i)$ for the sequence before. Therefore all the a_k must be calculated for each of the sequences in the search thus giving a complexity of $O(N)$ for Alg. 1.

There is however a method to speed up the m-sequence ordering search. Instead of trying to speed up the calculation of the MF, one could try to generate the m-sequence faster. The time complexity to find each of the $2^N - 1$ different subsequences in an m-sequence was $O(N)$ in the first approach. But if we use a primitive trinomial instead of the primitive polynomial, one could lower this complexity to a constant $O(1)$ per sequence. The only problem with this modification is that for some lengths N trinomials do not exist. Time comparison between the two different generations of m-sequences of length $M = 2^N - 1$, for $15 \leq N \leq 29$ where a trinomial could be found, can be seen in Table 2.3

A closer look at the complexity for the m-sequence ordered exhaustive search shows that there is not very much to accomplish by reducing the m-sequence generation from $O(N)$ to $O(1)$. For each of the $2^N - 1$ sequences generated the MF computation of Alg. 1 still has $O(N)$ complexity both for the polynomial - and trinomial generation of the m-sequence. This would also explain the small reduction in time shown in Table 2.3. Also note that the computation time for the primitive polynomial case will depend on which polynomial we choose. The complexity of the whole search is $O(2^N \cdot N)$, where the exponential term dominates as it does for all exhaustive searches of this type.

Length, N	Polynomial	Trinomial	Time reduction
15	0	0	0%
17	0	0	0%
18	0	0	0%
20	1	1	0%
21	3	2	33%
22	8	6	25%
23	15	12	20%
25	66	55	17%
28	587	482	18%
29	1208	993	18%
30	2509	2047	18%

Table 2.3: Speed comparisons for the use of trinomials in m-sequence generation

2.3.3 Searching with different word ordering

One reason to use a word ordering other than the lexicographic order was to see if it was easier to predict the local MF maxima. If we look at Table 2.4 we can see the statistical difference between the successive MF values with the two different word orderings. As we can see, the worst-case-jump and the average jump are a little lower for the m-sequence ordering, but not enough to find a clear pattern. There exist however other word ordering schemes that might yield better results than the m-sequence ordering. One such ordering uses de Bruijn sequences to generate all the different sequences. A binary de Bruijn sequence is a sequence of length $D = 2^N$ that contains all possible subsequences of length N. The advantage of a de Bruijn sequence is that, as with the m-sequence ordering, one can use the update rule to compute successive subsequence MFs. For each length D the number of different binary de Bruijn sequences is shown to be [2]

$$M(2, D) = 2^{2^{D-1}-D} \tag{2.2}$$

One could then test all the different de Brujin sequence orderings, in hope of finding that one of them gives a clearer and more predictable sequence of MF values. If we found such a sequence ordering, we would be able to predict the local MF maxima, and therefore find the highest MFs without an exhaustive search. This approach involves a deeper theoretical understanding of de Bruijn sequences and we propose this idea as a future research project.

		Lex. ordering jumps		M-seq. ordering jumps	
Length, N	Merit Factor	Worst-Case	Average	Worst-Case	Average
10	3.85	3.26	0.83	2.74	0.53
11	12.10	10.96	0.82	9.68	0.53
12	7.20	6.32	0.77	5.31	0.47
13	14.08	12.80	0.72	11.27	0.45
14	5.16	4.24	0.66	3.50	0.40
15	7.50	6.61	0.61	5.45	0.37
16	5.33	4.51	0.57	3.05	0.33
17	4.52	3.66	0.53	2.80	0.31
18	6.48	5.56	0.50	4.26	0.29
19	6.22	5.13	0.47	3.75	0.27
20	7.69	6.72	0.45	4.99	0.26
21	8.48	7.15	0.43	4.92	0.24
22	6.21	5.16	0.41	3.86	0.23
23	5.63	4.59	0.39	3.28	0.22
24	8.00	6.76	0.38	5.33	0.21
25	8.68	7.44	0.36	6.08	0.20
26	7.51	6.42	0.35	4.41	0.19

Table 2.4: MF search with lexicographic and m-sequence ordering, length 10 - 26

2.4 Directed Search

The speed increase achieved by changing the word order does not help us very much in searching for binary sequences of greater length because we are still examining every one of 2^N sequences. For binary sequences of length $N > 60$, an exhaustive search for the sequence with the highest Merit Factor is not feasible with today's technology [1]. Instead we have to look at other limited ways to search for sequences with high MF. One alternative is to utilize the update rule used in the m-sequence ordered search (see Section 2.3.1), and try to make a more direct search for larger length N without searching through all the 2^N sequences. The hope here is to find a method to only search for sequences with high MF, and drop the majority of sequences with low MF. The criterion for using the update rule is to have an ordering that produces a sequence s_i that is either a result of a cyclic or negacyclic shift of the sequence before s_{i-1}.

Let $\mathcal{T}$ be a binary tree where each node is a sequence s_i and each node has two children, one that is a result of a cyclic shift of s_i, and another that is the result of a negacylic shift of s_i. Then assign the MF of the child to each edge between a parent and a child. A traversal of the tree $\mathcal{T}$ from the root $\mathbf{r}$ where we always choose the edge with the lowest weight (=MF) is then a directed search that can use the update rule.

Example:

Let $\mathcal{T}$ be a binary tree with 6 levels and each node is a binary sequence of length $N = 5$. Each node has a left child that is a cyclic shift of the node, and a right child that is a negacyclic shift of the node. Then let $\mathbf{r}$= 00000 be the root of $\mathcal{T}$. Fig. 2.2 then shows the best path from $\mathbf{r}$ to the bottom of $\mathcal{T}$ where the subtrees of the child with lowest MF are omitted.

$\square$

We can now use this method to define a decision rule to use in a directed search:

Decision Rule:

Let s_{i-1} be a binary sequence of length N. Then let s_i^1 be a cyclic shift of s_{i-1} and s_i^2 be a negacyclic shift of s_{i-1}. We can then define

$$s_i = \begin{cases} s_i^1 & \text{if } \mathrm{MF}(s_i^1) > \mathrm{MF}(s_i^2) \\ s_i^2 & \text{otherwise} \end{cases} \tag{2.3}$$

This ordering will most likely *not* generate all the different sequences of length N, and if it does generate all different sequences, the complexity will be the same as for the exhaustive search. But there is a possibility that the search will enter a circle and test the MF of the same sequences again and again. The reason for this is that the decision rule (2.3) does not take into consideration whether or not the new sequence s_i has appeared earlier in the search.

[1] Note that a naive search of a 2^{58} space is not feasible with today's technology, but Mertens [19] exponentially shrinks the search space by exploiting symmetry and branch and bound techniques to achieve an exhaustive search up to 2^{58}.

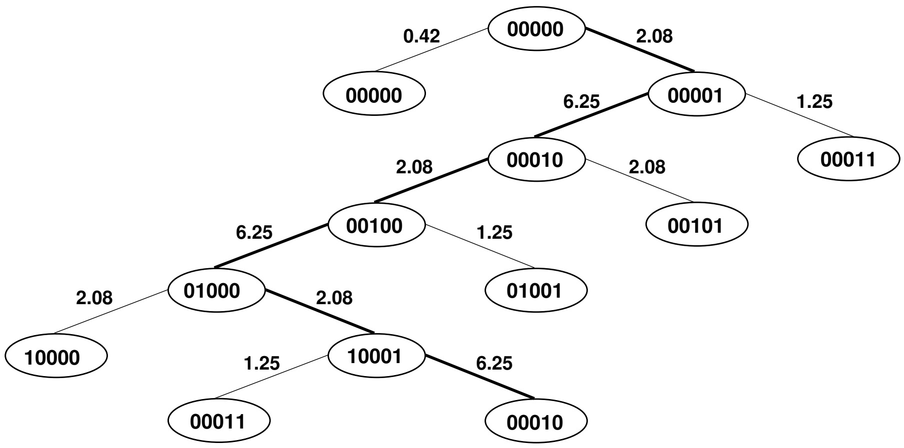

Figure 2.2: Tree traversal for a tree $\mathcal{T}$ with binary nodes of length $N = 5$

Example:

Let $s_i = 10101$ be the ith sequence in a directed search. Then assume that $s_{i+1} = s_{i+1}^2 = 01010$. Then the next sequence s_{i+2} must be s_i since $\mathrm{MF}(s_i^2) > \mathrm{MF}(s_i^1)$ based on the decision rule. We now have an endless circle.

□

Because of the possiblity of an endless circle and the intention to use the directed search for larger sequence length we will have to limit the number of sequences in the search. Based on the power of the computer the limit will be $\leq 2^{20}$ sequences. This limit is chosen based on tests that show that the MF of the best sequence found does not change much if we increase the number of sequences by a power of 2. But if we are to use this search for larger lengths $N > 1000$ this limit must also be raised.

In this type of directed search the result depends greatly on the starting sequence (or the root $\mathbf{r}$ in the tree traversal method). As for the other searches described above we are looking for the sequence with the highest Merit Factor. If we start with a sequence that already has a high Merit Factor the result of the search will be at least as high as the MF of the starting sequence, unless we keep the first sequence outside the search. Therefore to make this type of limited search useful we seek to find a Merit Factor that is higher than that of the starting sequence.

2.4.1 Zero starting sequence

The first binary starting sequence to use in our directed search was the zero sequence (00...00). The result of this search can be seen in Table 2.5. We can see that the MF of the starting sequence is low, because it consists of only zeros. But within the 2^{20} sequences that are tested, one will find at least one sequence with MF > 4.8 for all lengths $N \leq 60$. For example for length $N = 46$ this search finds a sequence with MF $= 7.20$ which is relatively close to the optimal

MF of 8.08 for this length (see Table 2.2). For larger values of N, the directed search does not find sequences with high MF after testing 2^{20} sequences. It appears that the the MF goes towards 1 as N becomes large.

		Zero start sequence		Random start sequence	
Length, N	# seq.	Start MF	Best MF	Start MF	Best MF
31	2^{20}	0.05	7.17	1.01	7.17
32	2^{20}	0.05	5.57	1.17	5.12
33	2^{20}	0.05	4.86	1.00	6.48
34	2^{20}	0.05	5.12	0.97	5.50
35	2^{20}	0.04	5.24	0.88	6.59
36	2^{20}	0.04	6.35	0.93	5.89
37	2^{20}	0.04	5.80	0.98	5.80
38	2^{20}	0.04	5.35	1.02	5.35
39	2^{20}	0.04	5.04	0.93	5.47
40	2^{20}	0.04	5.41	0.89	5.26
41	2^{20}	0.04	5.68	1.07	6.37
42	2^{20}	0.04	5.62	1.13	5.62
43	2^{20}	0.04	5.11	1.06	5.34
44	2^{20}	0.04	4.89	0.98	4.99
45	2^{20}	0.03	5.11	0.95	6.25
46	2^{20}	0.03	7.20	0.98	7.20
47	2^{20}	0.03	5.91	1.01	5.91
48	2^{20}	0.03	5.24	1.02	5.76
49	2^{20}	0.03	5.56	0.97	4.92
50	2^{20}	0.03	5.56	0.93	6.22
51	2^{20}	0.03	5.22	1.01	5.22
52	2^{20}	0.03	5.41	1.02	5.41
53	2^{20}	0.03	5.90	1.06	5.20
54	2^{20}	0.03	4.94	1.05	5.38
55	2^{20}	0.03	5.20	1.11	5.34
56	2^{20}	0.03	5.68	1.08	5.09
57	2^{20}	0.03	5.42	1.19	5.14
58	2^{20}	0.03	5.31	1.10	4.93
59	2^{20}	0.03	4.88	1.21	5.71
60	2^{20}	0.03	4.86	1.11	5.08
61	2^{20}	0.03	4.82	1.13	5.26
101	2^{20}	0.02	3.83	0.89	3.85
151	2^{20}	0.01	3.50	1.11	3.55
199	2^{20}	0.01	3.25	1.39	3.14
499	2^{20}	0.00	2.69	1.07	2.62
751	2^{20}	0.00	2.68	0.98	2.52
1499	2^{20}	0.00	2.38	0.97	2.34

Table 2.5: Directed Merit Factor search with zero and random start sequence

2.4.2 Random start sequence

Another choice for a starting sequence for the directed search is a binary sequence s_0 where the N bits of s_0 are randomly chosen to be 0 or 1. As we have shown in Section 2.1, this starting sequence will have a MF around 1, and thus be better than the zero starting sequence. The result of the directed search can be seen in Table 2.5. If we compare the result with the zero starting sequence, there does not appear to be a great difference between them.

Even though a random starting sequence has a higher MF than the start sequence of all zeros, it does not appear that this difference can be seen in the resulting best sequence. At some lengths the zero starting sequence results in a higher MF than the random starting sequence. This would suggest that the result of a directed search is not solely a result of the MF of the starting sequence. There might be other properties of a sequence that make it useful as a starting sequence in a directed search.

2.5 Shifted Legendre sequence

As an alternative to searching for binary sequences with good aperiodic Merit Factor, there are classes of constructed sequences with high asymptotic Merit Factor. One such class is the shifted Legendre sequence. The class of shifted Legendre sequences will have an asymptotic MF of six [7] [11]. As mentioned in Section 1.2 the class of shifted Legendre sequences is one of a few classes that have the highest known asymptotic Merit Factor.

The construction of a Legendre sequence of length $N = p$, p prime, can be achieved by finding a subset $\mathcal{S}$ of Z_p which specifies the positions of the 1s in the characteristic sequence $l(t)$ of $\mathcal{S}$:

$$l(t) = \begin{cases} 1 & \text{if } t \in \mathcal{S} \\ 0 & \text{if } t \notin \mathcal{S} \end{cases}, \tag{2.4}$$

when $0 \leq t \leq p-1$. The subset $\mathcal{S}$ is generated using a primitive generator α of GF(p),

$$\mathcal{S} = \left\{ \alpha^{2i} \bmod p \quad | \quad i = 0, ..., (\frac{p-1}{2}) - 1 \right\} \tag{2.5}$$

Høholdt and Jensen proved in [11] that the Merit Factor of an offset f of $l(t)$ is $\mathcal{F}$, and can be found by

$$\frac{1}{\mathcal{F}} = (\frac{2}{3} - 4|f| + 8f^2), \quad |f| \leq \frac{1}{2} \tag{2.6}$$

An offset sequence is one in which a fraction f bits of the sequence is chopped off the end of the Legendre sequence and appended at the other, in other words a cyclic shift of fN places. If we let $f = \frac{1}{4}$ then $l(t)$ will have an asymptotic Merit Factor of six.

Example:

Let $p = 19$ and $\alpha = 2$. Then $S = \{1, 4, 16, 7, 9, 17, 11, 6, 5\}$, and the sequence of length $N = 19$ is $l(t) = 0100111101010000110$. Then shift $l(t)$ by $\frac{1}{4}$ and let

$l' = 1111010100001100100$ be the new sequence. The Merit Factor for l' is then

$$\mathrm{MF}(l') = 6.22$$

□

2.5.1 Legendre sequence in directed search

A third alternative for a starting sequence for the directed search is to use a shifted Legendre sequence of length $N = p$, p prime. The result of the directed search with a shifted Legendre sequence as the starting sequence can be found in Table 2.6. The results show only a small increase in the Merit Factor from the starting sequence to the best sequence found. If we take a look at the best shift of a Legendre sequence in Table 2.6 we can see that when $N \geq 59$ the search does not find any higher MF than that of an optimal shift of a Legendre sequence. Also for $N < 59$ the best MF for this search is equal or below that of the search starting with a zero or random sequence (see Table 2.5).

Therefore it looks like the result achieved with a directed search using a shifted Legendre sequence as the start sequence would only find a MF that will be equal or below what can be found if we tested all shifts of a constructed Legendre sequence. But as will be shown later the Legendre sequence does have some properties that makes it useful in some other type of directed search.

2.6 Even length sequences with good MF

Another construction with an asymptotic Merit Factor of six is a class of sequences of length $N = 2p$, where p is prime [21]. Let $s(t)$ be a binary sequence of length N and $s'(t)$ be a binary sequence of length $2N$, where $s'(t) = s(t)$, for $0 \leq t < N$, and $s'(t) = s(t) + 1 \pmod 2$, for $N \leq t < 2N$. We can then use a subset C of Z_{2N} to define the characteristic sequence $s'(t)$:

$$s'(t) = \begin{cases} 1 & \text{if } t \in C \\ 0 & \text{if } t \notin C \end{cases},$$

where $C = (0, D_0) \cup (1, D_0) \cup (2, D_1) \cup (3, D_1) \cup F$. $F = \{p, 2 * p\}$ and D_i is defined by

$$D_i = \{\alpha^i, \alpha^{2+i}, \alpha^{4+i}, \alpha^{6+i}, ..., \alpha^{p-3+i}\}, \qquad 0 \leq i < 2$$

where α is a primitive generator over GF(p). The set (k, D_i), $0 \leq k < 4$, defines $k = t \bmod 4$, $r = t \bmod p$ for $r \in D_i$. t can then be recovered from (k, D_i) by means of the Chinese Remainder Theorem mod $4p$.

The sequence $s(t)$ from this construction will have an asympmtotic Merit Factor of six and ideal negaperiodic autocorrelation properties, while the whole sequence $s'(t)$ will have very good periodic autocorrelation properties, apart from one coefficient.

	Directed Search			Legendre sequences	
Length, N	Start MF	Best MF	# of seq.	Best shift	Best MF
31	4.04	6.41	2^{20}	21	6.41
37	4.23	5.80	2^{20}	9	4.23
41	3.69	5.84	2^{20}	11	4.78
43	5.34	5.34	2^{20}	9	5.89
47	4.95	5.78	2^{20}	10	5.34
53	3.76	5.62	2^{20}	18	4.31
59	6.19	6.19	2^{20}	15	6.19
61	5.57	5.85	2^{20}	17	5.85
67	5.77	5.77	2^{20}	15	6.29
71	4.64	6.07	2^{20}	55	6.07
73	4.76	5.01	2^{20}	19	5.01
79	5.43	5.58	2^{20}	56	5.75
83	5.81	5.93	2^{20}	23	5.93
89	4.67	5.76	2^{20}	26	5.76
97	5.25	5.25	2^{20}	24	5.25
101	4.99	5.39	2^{20}	26	5.39
151	5.84	5.84	2^{20}	37	5.92
199	5.63	5.84	2^{20}	54	5.84
499	5.82	5.85	2^{20}	382	5.98
751	5.91	5.93	2^{20}	192	5.93
1499	5.98	5.98	2^{20}	1131	6.00

Table 2.6: Directed MF search with a shifted Legendre start sequence of prime length

2.6.1 Directed search with even length sequences

These sequences can also be used as starting sequences for the directed search (see Section 2.4). The result of the directed search that uses these even length sequences as the starting sequence can be found in Table 2.7. We can see that the directed search yields a small improvement over the starting sequence, but it appears that the increase in the MF decreases as N grows.

2.7 Extended Directed Search

Each sequence in the directed search is either a cyclic or a negacylic shift of the sequence before. (see Section 2.4). Therefore it is possible to construct a sequence $\mathcal{S}$ that contains all the sequences in the search as subsequences. The sequence $\mathcal{S}$ contains all the subsequences in the search the in same way that a de Bruijn sequence of length $D = 2^N$ contains all the subsequences of length N. The length of $\mathcal{S}$ will depend on the limit we set for the directed search. Let l be the limit for the number of sequences in a directed search. The length of $\mathcal{S}$ will then be $t = N + l$.

Length, N	Start MF	Best MF	# of seq.
34	3.59	5.12	2^{20}
38	4.78	6.07	2^{20}
46	5.43	7.20	2^{20}
58	4.00	4.93	2^{20}
62	5.29	6.10	2^{20}
74	5.38	5.38	2^{20}
82	4.51	5.45	2^{20}
86	5.14	5.14	2^{20}
94	6.18	6.96	2^{20}
106	5.03	5.42	2^{20}
118	6.22	6.22	2^{20}
122	4.83	4.83	2^{20}
134	5.37	5.37	2^{20}
142	6.48	6.99	2^{20}
146	5.22	5.41	2^{20}
158	5.79	6.02	2^{20}
166	5.92	5.92	2^{20}
178	5.54	5.62	2^{20}
194	5.09	5.10	2^{20}
202	5.34	5.45	2^{20}
302	5.91	6.02	2^{20}
398	5.99	6.06	2^{20}
502	6.12	6.12	2^{20}
758	5.89	5.89	2^{20}
1502	5.98	5.99	2^{20}

Table 2.7: Directed MF search with an even length class of start sequences, $N = 2p$, p prime

Example :

Let $l = 6$ and let $s_0 = 01100$ be a binary sequence of length $N = 5$ and s_i, $1 \leq i \leq 6$, are generated by a directed search for l sequences such that

$$s_1 = 11001, s_2 = 10011, s_3 = 00111, s_4 = 01110, s_5 = 11101, s_6 = 11010$$

We can then construct a sequence $\mathcal{S} = 01100111010$ of length $t = N + l = 11$ that contains all the sequences s_i, $0 \leq i \leq 6$, as subsequences.

$\square$

Let $\mathcal{T}$ be a sequence of length $t = N + l$ generated from a directed search for sequences of length N as described above. We can now also use $\mathcal{T}$ to test the MF of all subsequences of length $N' = N + d$, $d > 0$. The search where we look both for good sequences of length N and sequences of length N' is called an extended directed search, since we now also look at the MF for an extended sequence of length $N' > N$.

Example :

Let $N = 5$ and let $t = 0110011101010011110110$. In the normal directed search we tested the MF of the sequence $s_i \in \{01100,\ 11001,10011,...\}$. For an extended directed search we can use the same sequence t and also test the MF for sequences s'_i of length $N' = 6$, $s'_i \in \{011001,\ 110011,\ 100111,\ ...\}$.

□

2.7.1 The complexity of the extended directed search

Alg. 2 shows an algorithm that uses the update rule in an extended directed search for both sequences of length N and sequences of length $N' = N+d$, $d > 0$. The function **MF()** in Alg. 2 is a normal $O(N^2)$ complexity calculation of the MF that stores the aperiodic autocorrelation values for use in **updateMF()**. **updateMF()** is the algorithm shown in Alg. 1 that finds the MF of a sequence s_i with a complexity of $O(N)$ when we know the a_k values for sequence s_{i-1}.

The complexity of a directed search through l sequences of length N is $O(l \cdot N + N^2)$. If we let l be a constant or a linear function of N we can write the complexity as $O(N^2)$. After we have incorporated a search for a second sequence length $N' = N+d$ in the directed search the complexity will be $O(l \cdot (N+d) + (N+d)^2)$ (see Alg. 2). We can also let d be a constant or a function of N, which makes the complexity $O(N^2)$, the same as for the directed search for a single length sequence. In practice the difference between a directed search and an extended directed search will only be a constant.

2.7.2 The starting sequence for an extended directed search

We now have an algorithm for an extended directed search for two sequence lengths. The next question is whether the search will find any sequences with high Merit Factor. In the sections above one can see that the directed search for sequences of length N was unable to find any new sequences with high Merit Factor. Therefore we can concentrate on the search results for sequences of length $N' = N + d$, $d > 0$. Since the key to the directed search is the starting sequence r, the first step was to do an extended directed search with some of the different starting sequences used in the directed search.

For most of the different starting sequences the extended directed search found some sequences with a higher MF than the previous directed search. But the most interesting results came from the unexpected difference in the highest MF when we used an unshifted Legendre sequence rather than a shifted one. It looks like the MF for each extended sequence of length $N' = N + d$, $0 < d \leq N$, is higher when we use an unshifted Legendre sequence instead of a shifted Legendre sequence as starting sequence. When we use a random sequence as the starting sequence the extended sequence will have about the same MF as for the basic directed search when $d = 0$ (see Table 2.5), for all small values of d, but as d increases the MF decreases. Fig. 2.3 shows the result of an extended directed search with $N = 499$, $0 \leq d \leq 151$, and $l = 2^{10}$. The left graph shows the highest MF for the extended sequence for both the Legendre starting sequences, and the right graph shows how much higher the MF of a search with

INPUT :	**N**	=	sequence length
	seq	=	any binary array of length N+1
	N'	=	the length of the second sequence
	seq'	=	empty binary array of length N'+1
	l	=	the number of sequence of length N to search for
OUTPUT :	**MF**	=	the MF for the best sequence of length N
	MF'	=	the MF for the best sequence of length N'

```
for i = 1 to N do
        seq'[i] = seq[i]
end = end' = N
MF = MF(seq, end, N)
for i = 1 to i = l do
        end = end + 1 (mod N + 2)
        tempMF = getNextMF(seq, end, N)
        if (tempMF > MF) MF = tempMF
        if (i + N ≥ N' )
                end' = end' + 1 (mod N' + 2)
                seq'[end'] = seq[end]
                if (N + i == N') tempMF' = MF(seq',end',N')
                else tempMF' = updateMF(seq',end',N')
                if (tempMF' > MF') MF' = tempMF'
        else
                end' = end' + 1 (mod N' + 1)
                seq'[end'] = seq[end]
        end if-else
end for-do
```

start function getNextMF(seq, end, N)

```
seq[end] = '1'
mf1 = updateMF(seq,end,N)
seq[end] = '0'
mf2 = updateMF(seq,end,N)
if (mf1 > mf2)
        seq[end] = '1'
        return mf1
else
        seq[end] = '0'
        return mf2
```

end function

Algorithm 2: Extended Directed Search the unshifted Legendre starting sequence is. Also note that for $d = 0$ the result of this extended directed search is the same as for a normal directed search. The result of this extended directed search would suggest that the unshifted Legendre sequence is better suited for this type of search, and therefore deserves a closer look.

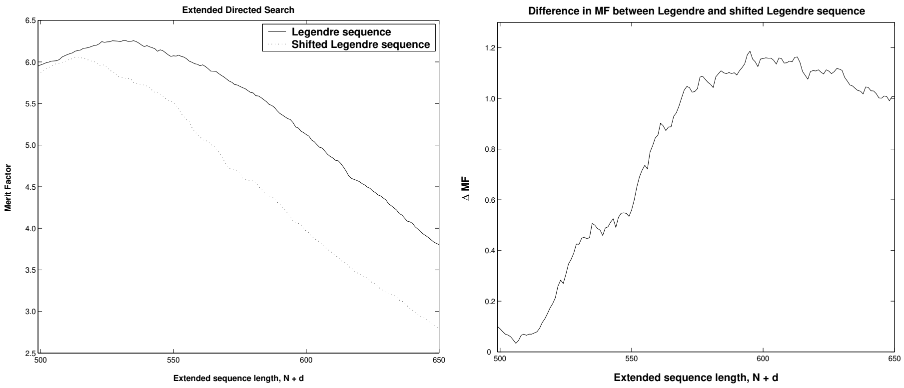

Figure 2.3: Extended Directed Search using a shifted Legendre and standard Legendre sequence

2.7.3 The Legendre Extended Directed Search

Let $b_{l,k}$ be the binary sequence of length k, $k > N$, with the highest MF from an extended directed search through l sequences, and let $b_{l,N''}$ be the extended sequence of length $N'' = N + d_{opt}$ such that

$$\mathrm{MF}(b_{l,N''}) > \mathrm{MF}(b_{l,N'}), \quad N' = N + d, \quad 0 \leq d \leq \frac{N}{10}$$

and let l be a large constant or a linear function of N. From Fig. 2.3 it appears that the MF started to decrease when $d > 40$. We therefore choose to concentrate the search on $0 \leq d \leq \frac{N}{10}$.

An extended directed search for good sequences of length $N'' = N + d_{opt}$ where $N = p$, p prime, $101 \leq N \leq 2477$ and $0 \leq d_{opt} \leq \frac{N}{10}$ was undertaken with an unshifted Legendre sequence of length N as the start sequence. The size of the search was $l = 2^{10}$. Fig. 2.4 shows the best MF for the optimal extension of N and Fig. 2.5 (left) shows d_{opt}. Fig. 2.5 (right) shows the minimum number of sequences $l_{min} \leq l$ that we can have without reducing the best MF found.

The data of Fig. 2.5 for both d_{opt} and l_{min} seems to fit a linear function. Using a MatLab polyfit function to find a function for the optimal extension $d(N)$ that could fit the data from the search gave us

$$d(N) = 0.059N + 0.77 \tag{2.7}$$

The same method was used to find a function $l(N)$ that made sure that $l(N) > l_{min}$

$$l(N) = 0.31N + 20 \tag{2.8}$$

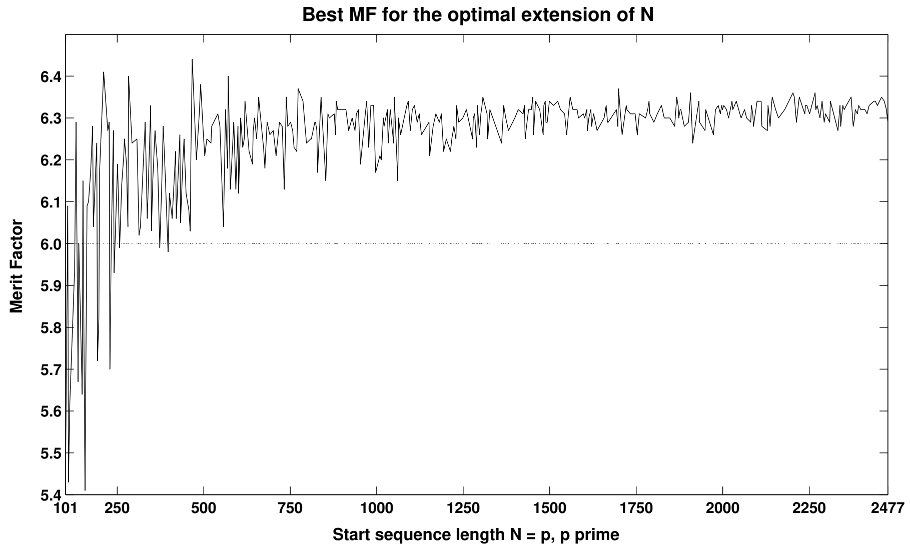

Figure 2.4: Extended Directed Search using an unshifted Legendre start sequence

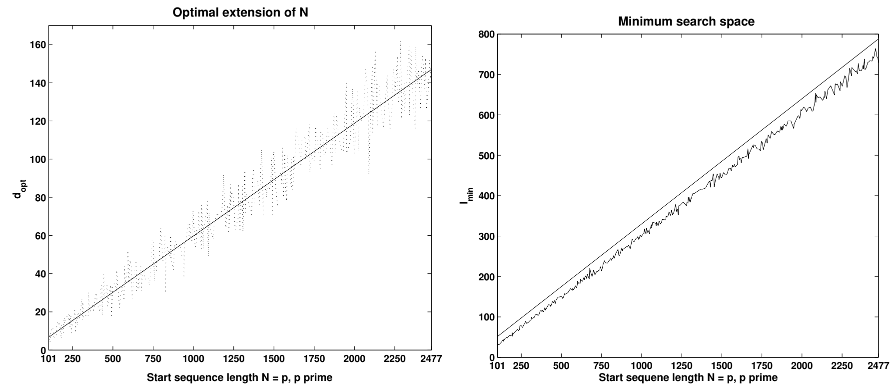

Figure 2.5: Optimal extension and minimum search space

Both the linear functions (2.7) and (2.8) are shown in Fig. 2.5.

The difference in the Merit Factor between the extended sequence of length $N'' = N + d_{opt}$ and $N'' = N + d(N)$ can be seen in Fig. 2.6. As expected the extension using an approximation function $d(N)$ is not as good as the optimal extension d_{opt}, but as N gets large the difference $\Delta\text{MF} \leq 0.02$.

Extended Legendre semi-construction

Now that we have an approximation to the length of the extension d based on the length of the starting sequenece N we can start another extended directed search. This time we will set the number of sequences to search for to $l = 0.31N + 20$, and the extended sequence length to $N' = N + d$, $d = 0.059N + 0.77$.

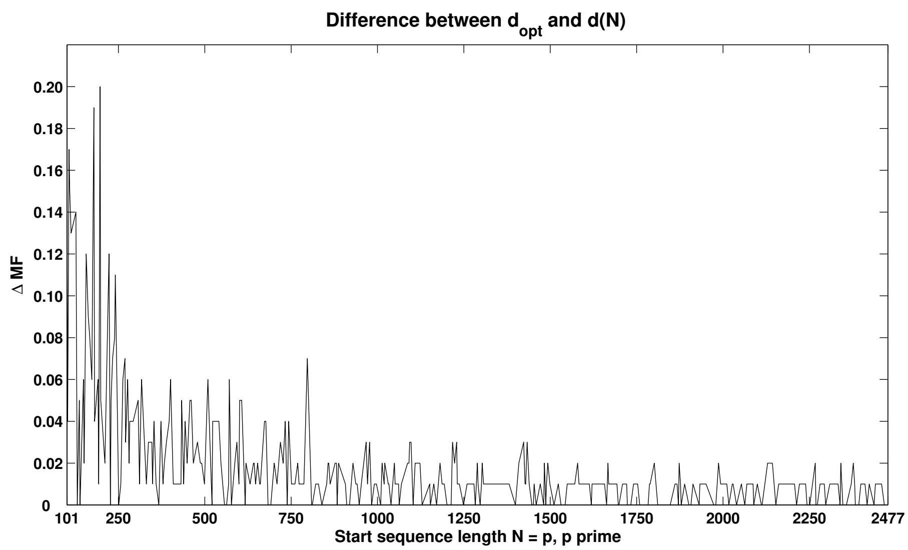

Figure 2.6: The difference between the extendsion d_{opt} and the approximation $d(N)$

The search will now be done for all prime N, $101 \leq N \leq 19997$. The MF from this search can be seen in Fig. 2.7. Based on this result of the extended directed search a conjecture is proposed :

Conjecture 1 *Let $\mathcal{S}$ be a binary sequence of length $p+l$, p prime, $l = 0.31p+20$, and let the first p bits of $\mathcal{S}$ be the Legendre sequence of length p. The last l bits of $\mathcal{S}$ can then be found by*

$$s_{p+k-1} = \begin{cases} 1 & \textit{if } MF(s_k, s_{k+1}, ..., s_{p+k-2}, 1) > MF(s_k, s_{k+1}, ..., s_{p+k-2}, 0) \\ 0 & \textit{otherwise} \end{cases},$$

for $1 \leq k \leq l$. It is then conjectured that there will exist a subsequence of length $N = p+d$, $d = \lfloor 0.059p + 0.77 \rfloor$, inside the sequence $\mathcal{S}$ with Merit Factor > 6.30 for any large p.

Complexity

As was shown above (see section 2.7.1) the complexity to find an extended sequence with high MF using a starting sequence of length N as described in Conjecture 1 will be $O(N^2)$ if both d and l are a contant or a linear function of N. This is a relatively low complexity compared to other searches, and any sequence construction will at least have a complexity of $O(N)$. For example the complexity to find the best shift of a modified Jacobi sequence will also be $O(N^2)$. This complexity of $O(N^2)$ means that if it takes t time to do a search with a start sequence of length N, it would take $4t$ for a search with a length $2N$ start sequence. Table 2.8 agrees with this, which would suggest that the practical complexity is also $O(N^2)$.

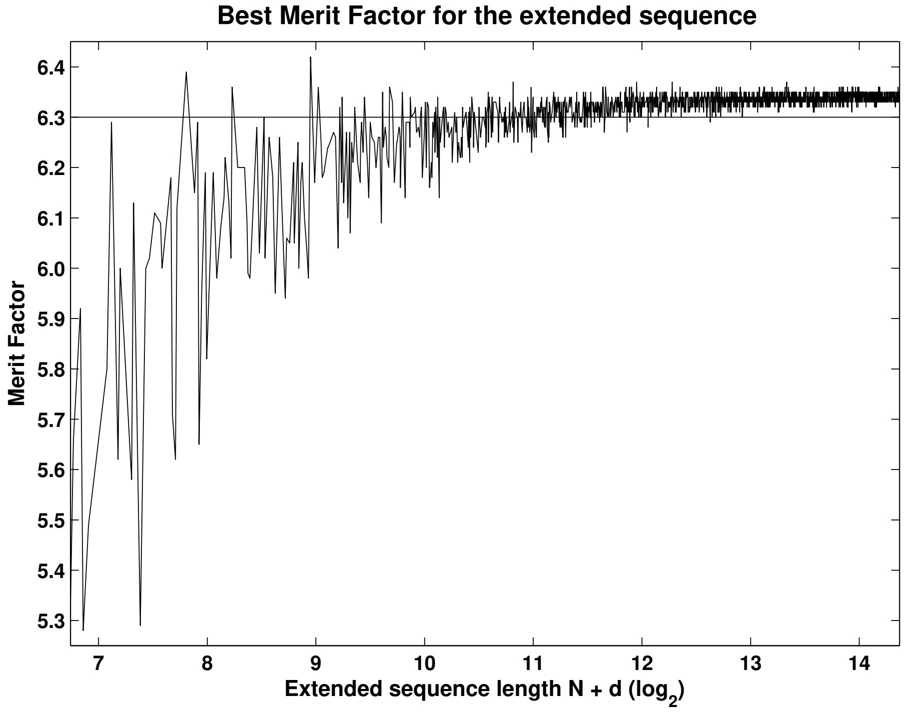

Figure 2.7: Extended directed search with $d = 0.059N + 0.77$ and $l = 0.31N + 20$, $101 \leq N \leq 19997$

Table 2.8: Computation time for an extended directed search

Length, N	Time (sec)
1997	0
4001	1
8009	5
16001	18
32003	81
64007	332

Other work with extended Legendre sequences

A. Kirilusha and G. Narayanaswamy [14] have done tests on extending shifted Legendre sequences by appending the last u bits of a Legendre sequence s to the front of s. They have shown that if $u < O(N^{\frac{1}{2}})$ the asymptotic Merit Factor of $u + s$ is six. A comparison between their construction with an optimal u_{opt} and the semi-construction described in Conjecture 1 have shown that $u_{opt} < d(N) = 0.059N + 0.77$, and that the MF of $u + s$ is lower than the MF of an extended directed search.

Examples

A few examples of long-length binary sequences with MF> 6.3, constructed using the Extended Legendre Semi-Construction can be downloaded from [15].

2.8 Skewsymmetric Merit Factor Search

If constructed correctly, skewsymmetric binary sequences are a type of sequence that will have high Merit Factor. Golay used this type of construction when he did his limited search for $N < 120$, where he found sequences with Merit Factor above 9 [8].

2.8.1 Construction of skewsymmetric sequences

Any skewsymmetric sequence s can then be considered as the interleaving of a symmetric sequence of the form

$$Ac\overline{A}$$

and an antisymmetric sequence

$$B\overline{B'}$$

where the overbar indicates sequence reversal and the prime indicates complementing of the sequences.

The construction of a skewsymmetric sequence of length N, N odd, consists of interleaving a symmetric sequence of length $\frac{N+1}{2}$ and an antisymmetric sequence of length $\frac{N-1}{2}$.

Example:

Let $s = abcba$ be a symmetric sequence of length 5, and let $s' = de\overline{e}\overline{d}$ be an antisymmetric sequence of length 4. Interleaving s and s' gives a new skewsymmetric sequence $t = adbec\overline{e}b\overline{d}a$ of length 9.

□

The reason skewsymmetric sequences have good Merit Factor lies in how they are constructed. All skewsymmetric sequences of length $N = 2n - 1$ satisfy

$$s_{n+l} = (-1)^l s_{n-l} \qquad l = 1, ..., n-1 \tag{2.9}$$

Theorem 1 *Let $s = (s_0, s_1, ..., s_{2n})$ be a binary sequence of length $N = 2n + 1$. Then $a_k = 0$ for all odd k.*

Proof

Let the elements of s be of the form $s_i = \{-1, +1\}$. We can then write the a_k in the form

$$a_k = \sum_{j=0, j \text{ even}}^{n} (s_j * s_{k+j}) + (s_{2n-j} * s_{2n-k-j})$$

Since all skewsymmetric sequences obey (2.9) we have that $s_j = s_{2n-j}$. This gives us

$$a_k = \sum_{j=0, j \text{ even}}^{n} s_j * (s_{k+j} + s_{2n-k-j})$$

For odd k, $s_{k+j} + s_{2n-k-j} = 0$ from (2.9), and $a_k = 0$.

Q.E.D.

from which it follows that all a_k with k odd vanish. To show this we can take a look at an example.

Example:

Let $t = adbec\bar{e}b\bar{d}a$ be a skewsymmetric sequence of length 9. The aperiodic autocorrelation can then be found by

a_k	a	d	b	e	c	$\bar{e}$	b	$\bar{d}$	a
1		a	d	b	e	c	$\bar{e}$	b	$\bar{d}$
2			a	d	b	e	c	$\bar{e}$	b
3				a	d	b	e	c	$\bar{e}$
4					a	d	b	e	c
5						a	d	b	e
6							a	d	b
7								a	d
8									a

As can be seen in the table above, all a_k vanish when k is odd.

□

Even though skewsymmetric constructions have found sequences with high Merit Factor [8] for odd length, we have no guarantee that the best sequences are skewsymmetric. Exhaustive search has shown that only $\frac{1}{3}$ of the best odd length sequences are skewsymmetric [18].

2.8.2 Skewsymmetric Legendre construction

To construct a skewsymmetric sequence with high Merit Factor, one usually needs a symmetric and an antisymmetric sequence with Merit Factor above 1. In [8] all the best skewsymmetric sequences are constructed from symmetric and antisymmetric sequences with MF above 1.2, and most of them above 1.4. Insted of doing a search for the symmetric and antisymmetric like Golay did, I was looking for a way to construct the symmetric and antisymmetric sequences. When $p = 4k+3$, p prime, the Legendre sequence of length p is an antisymmetric sequence if the first bit is dropped. It turned out that the Merit Factor for this sequence of length $n = p - 1$ is about 1.5. Also if $p = 4k + 1$ the Legendre sequence is a symmetric sequence if shifted by $\frac{p}{2}$. This symmetric sequence also has a high merit factor.

Let p_1 and p_2 be a twin prime pair where $p_2 = p_1 + 2$. Then we construct an antisymmetric sequence of length $n_1 = p_1 - 1$ and a symmetric sequence of length $n_2 = p_2$. To construct a skewsymmetric sequence we need a symmetric sequence of length $n_2 = n_1 + 1$, and our symmetric sequence is $n_2 = n_1 + 3$. If we drop the first and last bit of the symmetric sequence we have two sequences, hopefully with high Merit Factor. We can then construct the skewsymmetric sequence. The result of this construction can be found in Table 2.9. This skewsymmetric Legendre construction seems to have an asymptotic Merit Factor of about 3.0. This is lower than for the class of Legendre sequences (see Section 2.5), but this construction can be used for other lengths than the Legendre construction. We also get a small improvement for some of the lengths by shifting them around cyclically.

Length	MF symmetric	MF antisymmetric	MF skewsymmetric	Shift
61	1.3093	1.7928	3.5371	0
85	1.0993	1.4975	2.5476	0
117	1.0926	1.5617	2.6715	90
141	1.3217	1.8135	3.3157	0
205	1.4279	1.5598	3.4986	0
277	1.1757	1.5074	2.7628	3
357	1.4047	1.5173	2.9598	0
381	1.3989	1.6333	2.9100	0
397	1.4599	1.5539	2.9398	0
453	1.5390	1.5100	3.0586	0
477	1.4039	1.6119	3.1491	0
541	1.4467	1.5432	2.9969	0
565	1.3877	1.5004	3.0874	2
621	1.4454	1.6060	3.0772	0
693	1.4299	1.5035	3.0728	0
837	1.4613	1.5108	2.9033	0
925	1.5015	1.5046	3.0213	0
1045	1.4169	1.5010	3.0506	0
1141	1.4635	1.5008	3.0939	0
1197	1.4540	1.5474	3.1037	0
1237	1.4108	1.5006	2.7400	21
1317	1.4620	1.5064	3.1606	0
1621	1.4534	1.5011	3.1611	0
1653	1.4967	1.5010	3.0991	0
1717	1.4703	1.5009	2.8681	0
1765	1.4752	1.4995	2.9166	1760

Table 2.9: Legendre construction of skewsymmetric sequence

2.8.3 Alternative skewsymmetric Legendre construction

Let p_1 and p_2 be a pair of twin primes. If p_1 is of the form $4k+1$ and $p_1 < p_2$, we can construct a symmetric sequence by cyclically shifting the Legendre sequence of length p_1, by $\frac{p_1}{2}$. Then we construct an antisymmetric sequence of length $p_2 - 1$ by removing the first bit of the Legendre sequence of length p_2, where $p_2 = 4k+3$. We then interleave these two sequences to construct an almost skewsymmetric sequence (the symmetric and antisymmetric sequences have swapped length compared to Golay's construction). The Merit Factor of this construction is marked with (1) in Table 2.10.

If p_1 is of the form $4k+1$ and $p_1 > p_2$, we have a sequence of length $p_1 - 1$ by removing the first bit of the Legendre sequence of length p_1. We then construct an almost antisymmetric sequence of odd length by shifting cyclically the Legendre sequence of length p_2 by $\frac{p_2}{2}$. We can then interleave the two sequences and generate a skewsymmetric sequence. The merit factor of this construction is marked with (2) in Table 2.10.

As we can see the result is much better when p_1 is on the form $4k+1$ and $p_1 < p_2$. For this construction it appears that we have an asymptotic Merit Factor of 3. This means that this construction works best for half the prime twins, since the other half appears to have an asymptotic Merit Factor of 1.3.

Length, N	MF symmetric	MF antiSymmetric	MF skewsymmetric	Shift
59	0.96889	1.79283	2.87686 (1)	0
83	1.49025	1.49745	2.96684 (1)	0
119	1.48760	1.84180	1.22014 (2)	0
143	1.48966	1.62090	1.10667 (2)	0
203	1.33172	1.55982	3.31635 (1)	0
275	1.30849	1.50736	2.85917 (1)	0
359	1.49584	1.62069	1.31956 (2)	132
383	1.49610	1.54280	1.35065 (2)	298
395	1.41867	1.55386	2.96164 (1)	0
455	1.49672	1.56747	1.23957 (2)	93
479	1.49688	1.52640	1.23868 (2)	6
539	1.42163	1.54325	2.91658 (1)	0
563	1.48624	1.50040	3.19158 (1)	0
623	1.49760	1.51986	1.20556 (2)	494
695	1.49785	1.53704	1.36304 (2)	218
839	1.49822	1.56246	1.35756 (2)	766
923	1.50459	1.50797	1.25965 (2)	296
1043	1.47030	1.50104	3.06635 (1)	0
1139	1.51579	1.50076	3.13652 (1)	0
1199	1.49875	1.51180	1.32114 (2)	225
1235	1.45724	1.50055	2.77911 (1)	0
1319	1.49886	1.53890	1.33944 (2)	369
1619	1.48916	1.50109	3.16816 (1)	0
1655	1.49909	1.51418	1.28424 (2)	180
1715	1.50544	1.50092	2.88840 (1)	0
1763	1.51098	1.49946	2.94539 (1)	0

Table 2.10: Alternative Legendre construction of skewsymmetric sequence

2.9 Golay-Rudin-Shapiro Sequences

The class of Golay-Rudin-Shapiro (GRS) sequences was independently discovered by Golay [3] and Rudin-Shapiro [24]. The sequences in this class come in complementary pairs (CS pair) [22], which satisfy the useful property that their aperiodic autocorrelation coefficients sum to zero. Let $a = (a_0, a_1, ..., a_{N-1})$ and $b = (b_0, b_1, ..., b_{N-1})$ be binary sequences of length N and let their aperiodic autocorrelation be $a_k(a)$ and $a_k(b)$. The pair (a, b) is a complementary pair if

$$a_k(a) + a_k(b) = 0 \qquad 0 < k < N \tag{2.10}$$

The construction of the CS pairs can be done with a simple recursion

$$\begin{aligned} a_i &= a_{i-1}|b_{i-1} \\ b_i &= a_{i-1}|b'_{i-1} \end{aligned} \tag{2.11}$$

where "|" means concatenation of sequences and b' is the negation of the sequence b. The start of this recursion can be any complementary pair of sequences a_0 and b_0.

Example:

Let $a_0 = 1$ and $b_0 = 0$ be two binary sequences of length $N_0 = 1$. The recursion (2.11) then gives us

$N_0 = 1$	$N_1 = 2$	$N_2 = 4$	$N_3 = 8$
$a_0 = 1$	$a_1 = 10$	$a_2 = 1011$	$a_3 = 10111000$
$b_0 = 0$	$b_1 = 11$	$b_2 = 1000$	$b_3 = 10110111$

□

Because of the property that their aperiodic autocorrelation coefficients sum to zero, each of the sequences in the pair has also a high Merit Factor. It has been proved in [9] that the class of Golay-Rudin-Shapiro sequences have an asymptotic Merit Factor of three. It is also proven that the sum-of-squares indicator for this class satisfies

$$\sigma_n = 2\sigma_{n-1} + 8\sigma_{n-2} \tag{2.12}$$

In a later section we will show that this asymptotic MF and the sum-of-squares recursion only holds for a subclass of the whole GRS class (see Section 3.4.1).

2.9.1 GRS sequences in directed search

I have also tested the Golay-Rudin-Shapiro sequence as a starting sequence for the directed search. Since the length of GRS sequences is limited to $N = 2^n$, we have a limited number of sequence lengths to use in the search. Table 2.11 shows the result of a directed search with a Golay-Rudin-Shapiro sequence as the start sequence for length $N = 2^n$, for $3 \leq n \leq 10$. As we can see the increase in Merit Factor decreases as n grows, as it does for the Legendre and even length starting sequence.

Length, N	start MF	best MF	# of seq.
8	2.67	4.00	2^{20}
16	3.20	4.57	2^{20}
32	2.91	5.57	2^{20}
64	3.05	4.53	2^{20}
128	2.98	3.81	2^{20}
256	3.01	3.05	2^{20}
512	2.99	3.00	2^{20}
1024	3.00	3.00	2^{20}

Table 2.11: Directed Search with GRS start sequence

2.10 Overview

Here is an overview of the different searches and constructions presented in this chapter. Exhaustive search will find the optimal Merit Factor for a given length, but as the length gets larger the only way to find sequence with high MF is to use one of the known constructions. The best known asymptotic Merit Factor for a class of sequences is six, but with the new extended Legendre semi-construction it is possible to find sequences with MF > 6.3 for large length.

Exhaustive search

Two types of exhaustive search for sequences with good Merit Factor are presented. First using the lexigographic word ordering, and then with an m-sequence ordering using an update rule for faster computation of the aperiodic autocorrelation values. Without any modification the m-sequence ordering was about twice as fast as the lexicographic ordering (see Table 2.2 and 2.3). But after the modifications the lexicographic word ordering was almost three times as fast. Therefore it appears that the update rule could not beat the effectiveness of using a threshold and symmetries in the computation of the Merit Factor. Though we presented ways to speed up an exhaustive search, we had to limit the length to $N \leq 32$.

Even though Mertens [19] has done an exhaustive search up to $N = 58$ the computer resources needed for this type of search prevents us from finding any high Merit Factors for sequences of large length. This should indicate that exhaustive search is not the way to find sequences with high MF for sequences of large length.

Limited search

The method called directed search uses different starting sequences and the update rule from the m-sequence ordered exhaustive search. This search does not seem to give any new result. As the sequence length N grows large, it looks like the directed search with zero and random starting sequence will asymptote to 1, while the shifted Legendre and the even length cyclotomic construction asymptote to six.

Extended directed search

The directed search did have one more utilization. By extending the directed search we were able to find high Merit Factor for non-prime lengths N', using an unshifted Legendre sequence as the starting sequence. From the tests it looks like it is possible to find a Merit Factor > 6.3 for an extended sequence of length $N' = N + d$, $d = \lceil 0.059N + 0.77 \rceil$ for all large prime N. This semi-construction can be done in $\mathrm{O}(N^2)$ time, where N is the length of the Legendre start sequence.

Construction

There already exist several known constructions of binary sequences with high Merit Factor. The construction with the highest asymptotic Merit Factor are the shifted Legendre sequence [7] [11], the even length construction [21], shifted modified Jacobi sequence [13] and a special case of the modified Jacobi sequence the shifted Twin-Prime sequence [13], all with an asymptotic MF of six. Other constructions like the m-sequences [13] and Golay-Rudin-Shapiro sequences [9] have an asymptotic Merit Factor of three. The only new construction presented in this chapter is the skewsymmetric Legendre sequence. This construction using two Legendre sequences also appears to have an asymptotic Merit Factor of three.

Chapter 3

The Multidimensional Aperiodic Autocorrelation

In addition to the 1-dimensional autocorrelations defined in Section 1.1, one can also define a multidimensional two-point autocorrelation function for binary sequences of length $N = 2^n$. The multidimensional periodic autocorrelation has been used to assess cryptographic strength of a boolean function [17], but it appears that the aperiodic case has not been much studied. In this chapter we take a closer look at the multidimensional aperiodic autocorrelation and show that the class of Golay-Rudin-Shapiro sequences has a low sum-of-squares indicator. We also present a new class of sequences that has an even lower sum-of-squares indicator.

3.1 Definitions

Let $s \in Z_2^N$ be a binary sequence of length N, such that $N = 2^n$. We can then define the multidimensional two-point Periodic Autocorrelation Function (MPACF) of the sequence $s = (s_0, s_1, ..., s_{N-1})$ by,

$$c_k = \sum_{i=0}^{N-1} (-1)^{s_i + s_{i \oplus k}} \tag{3.1}$$

where $\oplus$ means 'bitwise' addition mod 2. In other words, for i and k each of bitlength n, $\oplus$ is defined by the equation $r = i \oplus k$, where $r_j = i_j + k_j \pmod 2$ and,

$$\begin{aligned} r &= \textstyle\sum_{j=0}^{n-1} r_j 2^j \\ i &= \textstyle\sum_{j=0}^{n-1} i_j 2^j \\ k &= \textstyle\sum_{j=0}^{n-1} k_j 2^j \end{aligned}$$

are the radix-2 decompositions of r, i, and k, respectively, where $r_j, i_j, k_j \in \{0, 1\}\ \forall\ j$.

We can say that i is multidimensionally greater or equal to k iff $i_j \geq k_j$, $\forall$ j, and we write this as $i \geq_m k$.

Example :
Let $0 \leq i < 16$, and $k = 5 = 0101$, then $i \geq_m k$ for $i \in \{0101, 0111, 1101, 1111\}$.

□

We define the Multidimensional Aperiodic Autocorrelation Functition (MAACF) for a sequence s by

$$a_k = \sum_{i \geq_m k} (-1)^{s_i + s_{i \oplus k}} \tag{3.2}$$

The multidimensional aperiodic sum-of-squares indicator for a sequence s is defined in the same way as the 1-Dimensional sum-of-squares indicator is defined (1.10),

$$\sigma_a(\mathbf{s}) = \sum_{k=1}^{N-1} |a_k|^2 \tag{3.3}$$

Example:

Let $s = abcdefgh$ be a binary sequence of length $N = 2^3 = 8$. The multidimesional aperiodic autocorrelation of s can then be found by

k	a	b	c	d	e	f	g	h	a$_k$
001		a		c		e		g	$ab + cd + ef + gh$
010			a	b			e	f	$ac + bd + eg + fh$
011				a				e	$ad + eh$
100					a	b	c	d	$ae + bf + cg + dh$
101						a		c	$af + ch$
110							a	b	$ag + bh$
111								a	ah

As can be seen in the table above a_{N-1} cannot be zero. The optimal sum-of-squares indicator is therefore lower-bound by $\sigma_a(\mathbf{s}) = 1$.

□

We can then define the multidimensional Aperiodic Merit Factor (MMF) the same way as for the 1-Dimensional Merit Factor

$$\text{MMF}(\mathbf{s}) = \frac{N^2}{2\sigma_a(\mathbf{s})} = \frac{2^{2n-1}}{\sigma_a(\mathbf{s})} \tag{3.4}$$

We know the 1-Dimensional Merit Factor of a random binary sequence is 1.0, but this is not true for this multidimensional Merit Factor. Looking at the average MMF for 2^{18} random sequences of random length 8 to 8192, one can see that the MMF increases as the length increases (see Fig. 3.1). This would suggest that we need to normalize the MMF in order to compare it to the 1-Dimensional MF. One such normalization is to use the random distribution of the multidimensional Merit Factor. We could find a function that fits the random distribution computed in Fig. 3.1, and use this function to normalize the MMF such that the MMF of a random sequence will be 1. As a result of the lack of a good motivation for this normalization we will mostly use the multidimensional sum-of-squares σ_n as a measurement of goodness for sequences of length $N = 2^n$.

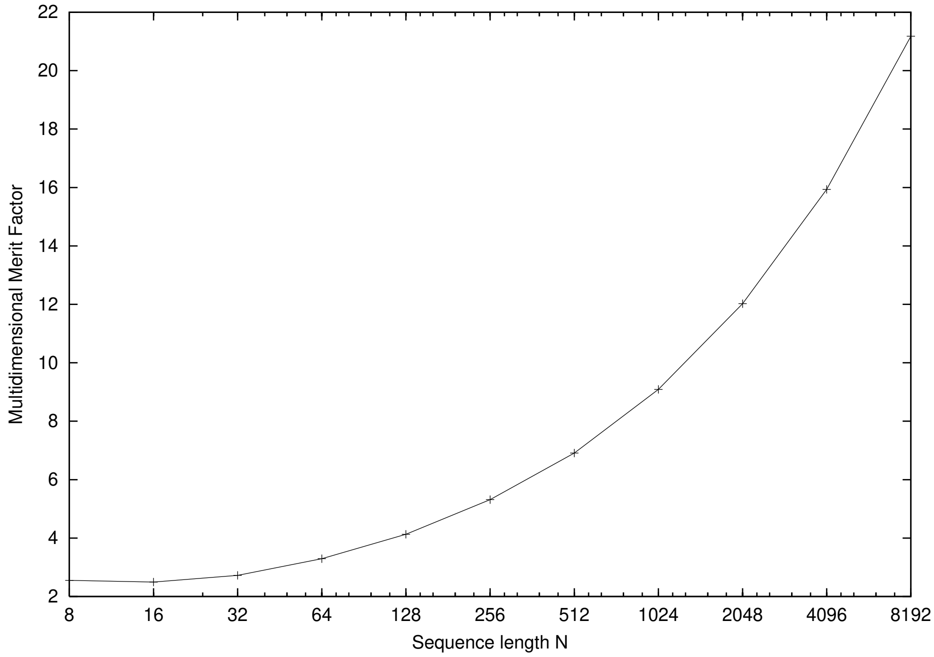

Figure 3.1: Average MMF for a random binary sequence (sample size 2^{18})

3.2 Exhaustive search for good MMF

As for the 1-dimensional MF the first approach to finding binary sequences with good multidimensional Merit Factor would be an exhaustive search for sequences with optimal MMF. In this search the word ordering was lexicographic, and the calculations of the MAACF were done with a recursive algorithm (see Alg. 3). Alg. 3 will return the value of a_k for a binary sequence seq of length N with the input $indexA = indexB = indexK = 0$, and K is the binary representation of k. The complexity of Alg. 3 will be the same as the number of nodes on a complete binary tree, $\mathrm{O}(2^n)$ or $\mathrm{O}(N)$, where $N = 2^n$ is the sequence length. Since there are $N - 1$ a_k-values the complexity to find the MMF of a sequence of length $N = 2^n$ is $\mathrm{O}(N^2)$.

The complexity of this search is the same as for the 1-Dimensional, $\mathrm{O}(2^N \cdot N^2)$, where N is the sequence length. This complexity limits naive searches to $N < 32$, but it is possible to use the same modifications to speed up the search as those described in Section 2.2.1. The results of this exhaustive search can be found in Table 3.1. The Algebraic Normal Form of the sequences as given in column 4 of Table 3.1 is described in Section 3.3. Also note that the MMFs for lengths 4 and 16, are optimal. This means the a_k values are zero for all values of k, except $k = 0$ and $k = N - 1$. For $N = 32$, I had to drop the general recursive algorithm (Alg. 3) in favor of a specialized one in order to do the search in reasonable time. This specialized one also only searches through half the sequences to speed up the search (see section 2.2.1). As of today there is no way to get a complete computer search for any $N > 32$.

```
INPUT :    (k        =  the k'th MAACF coefficient index)
           N         =  sequence length
           seq       =  binary sequence array of length N
           indexA    =  index for the start of block A
           indexB    =  index for the start of block B
           K         =  binary array representation of k
           indexK    =  index for K
OUTPUT :   a[k]      =  the k'th MAACF

calcMultiAAC(indexA, indexB, indexK, N, seq[ ], K[ ])

if (indexK is the last bit in K)
    if (K[indexK] == 1)
        return COMP(seq[indexA+1],seq[indexB])
    else
        return COMP(seq[indexA],seq[indexB]) +
                   COMP(seq[indexA+1],seq[indexB+1])
end if
if (K[indexK] == 1) return
    calcMultiAAC(indexA+N/2, indexB, indexK+1, N/2, seq[], K[])
else return
    calcMultiAAC(indexA, indexB, indexK+1, N/2, seq[], K[]) +
    calcMultiAAC(indexA+N/2, indexB+N/2, indexK+1, N/2, seq[], K[])
end function calcMultiAAC

function COMP(bit A, bit B)

if (A == B) return 1
else return -1

end function COMP
```

Algorithm 3: Recursive algorithm for computation of a_k

Length, N	Merit Factor	σ_n	CANF (see Section 3.3)
2	2.00	1	$0,$
4	8.00	1	$01,$
8	6.40	5	$02, 12,$
16	128.00	1	$012, 013, 023, 123$
32	24.38	21	$012, 013, 023, 123, 23, 24, 34,$

Table 3.1: Exhaustive search for binary sequences with best possible aperiodic MMF

3.3 Algebraic Normal Form

Expressions for binary sequences can be cumbersome when the length N starts to get larger. Algebraic Normal Form (ANF) is a way to express sequences multidimensionally in the form of boolean functions, and is especially suited to a multidimensional analysis of the sequence. For binary sequences of length $N = 2^n$, ANF can express the sequence as a boolean function, $f(X) : Z_2^N \rightarrow Z_2$, where $X = \{x_0, x_1, ..., x_{n-1}\}$ and $x_i = \{0, 1\}$. There is a transformation that can transform the binary sequence string to ANF and an inverse transformation that can transform back to the binary sequence again. Let us first look at all the possible outputs that can occur from $f(X)$. Since all the coefficients are modulo 2, we find that there are $2^n = N$ different outputs that can be constructed out of the n different variables x_i. The mapping between $f(X)$ and the length 2^N sequence, s, is given by

$$s_i = f(X = i) \tag{3.5}$$

where $i = \sum_{k=0}^{N-1} i_k 2^k$, $i_k \in \{0, 1\}$, is the usual 2-adic decomposition of i, and $X = i$ means $x_k = i_k$.

Example :

Let $n = 3$ and $f(X) = x_0x_1 + x_1x_2 + x_0$. The binary sequence $S = (S_0, S_1, ..., S_7)$ representing $f(X)$ is then given by

i	0	1	2	3	4	5	6	7
term	1	x_0	x_1	x_0x_1	x_2	x_0x_2	x_1x_2	$x_0x_1x_2$
S	0	1	0	1	0	0	1	0

□

Let $s = (s_0, s_1, ..., s_{N-1})$ be a binary sequence of length $N = 2^n$ where $s_i = f(X = i)$, and let $S = (S_0, S_1, ..., S_{N-1})$ be a binary representation of the ANF function $f(X)$ such that

$$f(X) = \sum_{i=0}^{N-1} (S_i \cdot \prod_{k=0}^{n-1} x_k^{i_k})$$

where $i = (i_0, i_1, ..., i_{n-1})$ is the binary representation of the integer i. We can then look at the transformation given by an N x N-matrix A, such that

$$A * s = S \mod 2 \tag{3.6}$$

$$A^{-1} * S = s \mod 2 \tag{3.7}$$

If we know the binary sequence which represents the ANF of $f(X)$, then the binary sequence string s can be determined. Therefore the function $f(X)$ generates the length N binary sequence s.

Example :

Let $f(X) = x_0x_1 + x_1x_2 + x_0$. The binary sequence s can then be found by setting X to all the binary sequences of length 3. This give us $s = 01000111$, as can be seen below

X	x_2	x_1	x_0	$f(X) = x_0x_1 + x_1x_2 + x_0$
0	0	0	0	0
1	0	0	1	1
2	0	1	0	0
3	0	1	1	0
4	1	0	0	0
5	1	0	1	1
6	1	1	0	1
7	1	1	1	1

The binary sequence s can also be found by using the transformation matrix A

$$S*A = [\,0\,1\,0\,1\,0\,0\,1\,0\,]* \begin{bmatrix} 1&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0\\ 1&0&1&0&0&0&0&0\\ 1&1&1&1&0&0&0&0\\ 1&0&0&0&1&0&0&0\\ 1&1&0&0&1&1&0&0\\ 1&0&1&0&1&0&1&0\\ 1&1&1&1&1&1&1&1 \end{bmatrix} = [\,0\,1\,0\,0\,0\,1\,1\,1\,] = s$$

□

3.3.1 Tensor Product

Premultiplying a length n vector by an N x N-matrix will have a time complexity of O(N^2). But for certain types of matrices there are faster ways to form the matrix-vector product. Let us define the Left Tensor Product as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \otimes \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ea & eb & fa & fb \\ ec & ed & fc & fd \\ ga & gb & ha & hb \\ gc & gd & hc & hd \end{pmatrix} \tag{3.8}$$

It can be shown that our transformation matrix $A = a_0 \otimes a_1 \otimes ... \otimes a_{n-1}$, where a_i is a 2 x 2-matrix, and therefore there exists an algorithm of complexity O($N * logN$), to compute the matrix-vector product. Instead of doing matrix multiplication row by row, we can split our binary sequence into vectors of length 2, and multiply each of these vectors by a_i. The way we split up the binary sequence is different for each of the n steps of the matrix multiplication. For our transformation the 2 x 2-matrices will all be

$$a_i = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{3.9}$$

If $A = a_0 \otimes a_1 \otimes ... \otimes a_{n-1}$, then $A^{-1} = a_0^{-1} \otimes a_1^{-1} \otimes ... \otimes a_{n-1}^{-1}$. Looking at a_i we can see that (3.9) is self inverse

$$a_i * a_i = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} * \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_{2\text{X}2} \quad \bmod 2 \tag{3.10}$$

We can therefore use the same algorithm to transform a binary sequence to ANF, and back again to the binary sequence.

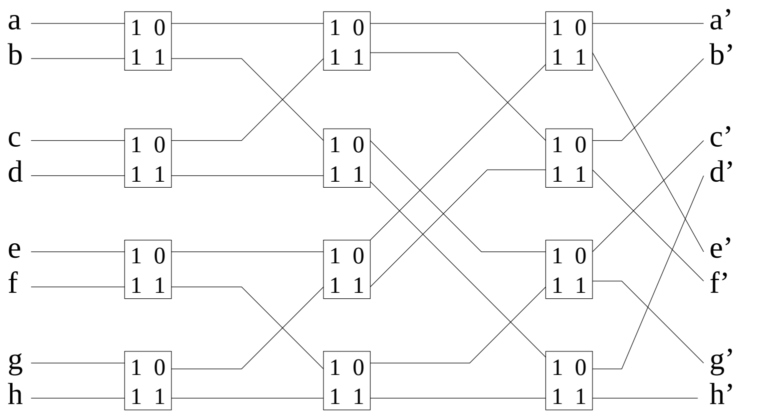

Figure 3.2: Algebraic Normal Form transformation

Example :

Let our binary sequence be $t = \{a, b, c, d, e, f, g, h\}$, of length $N = 2^3$, and let $A = a_0 \otimes a_1 \otimes a_2$. S can then be found by using the algorithm of Fig. 3.2, where $\{a', b', ..., h'\}$ is the output vector.

□

3.3.2 Condensed Algebraic Normal Form (CANF)

For notational convenience we sometimes omit the x’s from our ANF expansions. This condensed notation contains only the indeces of each term of the function $f(X)$, separated by a ’,’. To represent the constant term we can use an additional ’,’ at the end of the string.

Example :

Here are a few examples mapping ANF to CANF :

$x_0x_1 + x_1x_2 \rightarrow 01, 12,$
$x_0x_1x_2 \rightarrow 012,$
$x_0 + x_2 + 1 \rightarrow 0, 2, ,$
$x_0x_1x_2 + x_0x_1 + x_1x_2 + x_0x_2 + x_0 + x_1 + x_2 + 1 \rightarrow 012, 01, 12, 02, 0, 1, 2, ,$

□

I will often use this notation when describing boolean functions of n binary variables.

3.4 The MMF of Golay-Rudin-Shapiro sequences

Using ANF the complete set of binary Golay-Rudin-Shapiro (GRS) sequences can be defined by the following construction [1]

$$\mathrm{s(x)} = (\sum_{j=0}^{n-2} x_{\pi(j)}x_{\pi(j+1)}) + (\sum_{j=0}^{n-1} b_j x_j) + d, \qquad b_j, d \in Z_2 \tag{3.11}$$

where π is any permutation of Z_n, $x_{n-1}, x_{n-2}, ..., x_0$ are boolean variables and s is the length 2^n binary sequence such that

$$s_i = s(x_{n-1} = i_{n-1}, x_{n-2} = i_{n-2}, ..., x_0 = i_0) \tag{3.12}$$

where $i_{n-1}, i_{n-2}, ..., i_0$ is the binary representation of the integer i. Let the complete set of binary GRS sequence of length $N = 2^n$ be $\mathcal{C}_n$. The size of $\mathcal{C}_n$ has been shown to be [1]

$$|\mathcal{C}_n| = \frac{n!}{2} \cdot 2^{n+1} \tag{3.13}$$

It turns out that the Multidimensional Merit Factors for GRS sequences are very high. Computationally the sum-of-squares indicator, σ_n, for GRS sequences of length $N = 2^n$ appears to be given by,

$$\sigma_n(\mathbf{s}) = \sum_{\substack{i+k=n-1, \\ i \leq k}} 2^{2i} \binom{k}{i} \tag{3.14}$$

from which the MMF can be calculated using (3.4). Also the MMF appear to be invariant over the whole set of GRS sequences given by (3.11), and for a fixed length N. Therefore it will be sufficient to look at the special case where

$$s(x) = \sum_{j=0}^{n-2} x_j x_{j+1} \tag{3.15}$$

A list of the Multidimensional Merit Factors for binary GRS sequences of length 4 to 2048 can be found in Table 3.2. As can be seen from the table, the MMF and the corresponding σ_n values are much higher than the expected values for a random sequence (see Fig. 3.1). From Table 3.2 it is apparent that the sum-of-squares indicator σ_n follow a recursion:

Theorem 2 *The sum-of-squares σ_n for any Golay-Rudin-Shapiro sequence of the form*

$$s(x) = (\sum_{j=0}^{n-2} x_{\pi(j)}x_{\pi(j+1)}) + (\sum_{j=0}^{n-1} b_j x_j) + d, \qquad b_j, d \in Z_2$$

will follow the recursion $\sigma_n = 4\sigma_{n-2} + \sigma_{n-1}$.

Length, N	MMF	1-D MF	σ_n	CANF
4	8.0000	4.0000	1	$01,$
8	6.4000	2.6667	5	$01, 12,$
16	14.2222	3.2000	9	$01, 12, 23,$
32	17.6552	2.9091	29	$01, 12, 23, 34,$
64	31.5077	3.0476	65	$01, 12, 23, 34, 45,$
128	45.2597	2.9767	181	$01, 12, 23, 34, 45, 56,$
256	74.3039	3.0118	441	$01, 12, 23, 34, 45, 56, 67,$
512	112.5082	2.9942	1165	$01, 12, 23, 34, 45, 56, 67, 78,$
1024	178.9990	3.0029	2929	$01, 12, 23, 34, 45, 56, 67, 78, 89,$
2048	276.3410	2.9985	7589	$01, 12, 23, 34, 45, 56, 67, 78, 89, 9A,$

Table 3.2: Multidimensional Merit Factor for the GRS Construction

Proof of Theorem 2

The multidimensional aperiodic autocorrelation function a_k (3.2) can also be defined using a function $a_k(x)$ such that

$$a_k(x) = s(x)_{x_i=0} + s(x)_{x_i=1}, \qquad \forall\ i \text{ where } k_i = 1 \tag{3.16}$$

where k has a binary expansion as $(k_0, k_1, ..., k_{n-1})$ where $k_i \in Z_2 \forall i$, and k_0 represents the least significant bit of k. When $a_k(x)$ has degree 1, then the coefficients of $a_k(x)$ comprise an equal number of zeroes and ones and we say that the polynomial is balanced. In this case $a_k = 0$. When $a_k(x)$ has degree 0, then the coefficients of $a_k(x)$ are either all zero, or all one. In this case $a_k = 2^{n-wt(k)}$, where $wt(k)$ is the binary weight expansion of k. We can then define a_k based on $a_k(x)$ such that

$$a_k = \begin{cases} 0 & \text{if } \deg(a_k(x)) = 1 \\ 2^{n-wt(k)} & \text{if } \deg(a_k(x)) = 0 \end{cases} \tag{3.17}$$

where $wt(k)$ is the binary weight of the binary expansion of k. It follows from the way $a_k(x)$ is defined that for the GRS sequences $deg(a_k(x)) < 2$ for all k, and it is straightforward to show that (3.17) is the same as the general multidimensional aperiodic autocorrelation (3.2). It also follows that if (3.17) is valid for any single $s(x)$ of the form

$$s(x) = (\sum_{j=0}^{n-2} x_{\pi(j)} x_{\pi(j+1)}) \qquad , \tag{3.18}$$

then it is also true for any other sequence of the form in (3.11). Adding a linear term or a constant would not change the degree of $a_k(x)$ and therefore not a_k either. Therefore it will be sufficient to prove that the recursion holds for GRS sequences of the form (3.18). It will be easy to see that the degree of (3.16) is the same for all permutations π of (3.18), so we only give a proof for one permutation, $\pi(i) = i$.

All Golay-Rudin-Shapiro sequences can be represented as line graphs, where the nodes are the indices in the variable $x = (x_0, x_1, ..., x_{n-1})$, linear terms x_i

are represented as unconnected nodes i, and the constant can be represented by a '1' by the side of the figure.

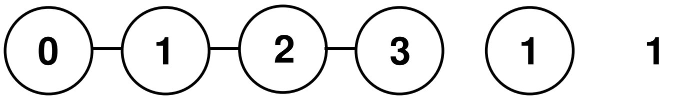

Figure 3.3: A line graph of a general Golay-Rudin-Shapiro sequence

Example:

The line graph for a GRS sequence of length $N = 2^4 = 16$ defined by the function $f(X) = x_0x_1 + x_1x_2 + x_2x_3 + x_1 + 1$ can be seen in Fig. 3.3.

□

Before I start on the proof let us take a look at an example.

Example :

Let $s(x) = x_0x_1 + x_1x_2$ be a GRS sequence of length $N = 2^3 = 8$. The table below shows the different $a_k(x)$ values and the corresponding a_k values.

k	$k_0k_1k_2$	$a_k(x)$	$\deg(a_k(x))$	a_k
1	100	$a_1(x) = x_1$	1	0
2	010	$a_2(x) = x_0 + x_2$	1	0
3	110	$a_3(x) = x_2 + 1$	1	0
4	001	$a_4(x) = x_1$	1	0
5	101	$a_5(x) = x_1 + x_1 = 0$	0	$2^{3-2} = 2$
6	011	$a_6(x) = x_0 + 1$	1	0
7	111	$a_7(x) = 1 + 1 = 0$	0	$2^{3-3} = 1$

We can also look at this example in terms of line graphs. Fig 3.4 shows how to find the graph of $a_k(x)$. If we color the nodes i where $k_i = 1$ black, we can based on (3.16) use the following rules to find the graph for $a_k(x)$:

$$\begin{array}{rl} \text{i)} & \text{if there is a connection between two black nodes, add 1} \\ \text{ii)} & \text{if a white node } j \text{ is connected to a black node, add } x_j \\ \text{iii)} & \text{if there is a connection between two white nodes, do nothing} \end{array} \tag{3.19}$$

These rules also work on binary strings where a black node equals 1 and a white node equals 0.

□

Now let us prove the general case by induction. Let

$$\sigma_{n-2} = \sum_{k=1}^{2^{n-2}-1} |a_{n-2,k}|^2 \qquad \text{and} \qquad \sigma_{n-1} = \sum_{k=1}^{2^{n-1}-1} |a_{n-1,k}|^2$$

be known values. We then split up all the $a_{n,k}(x)$ functions into four different sets based on the two highest bits k_{n-1} and k_{n-2} of $k = (k_0, k_1, ..., k_{n-1})$:

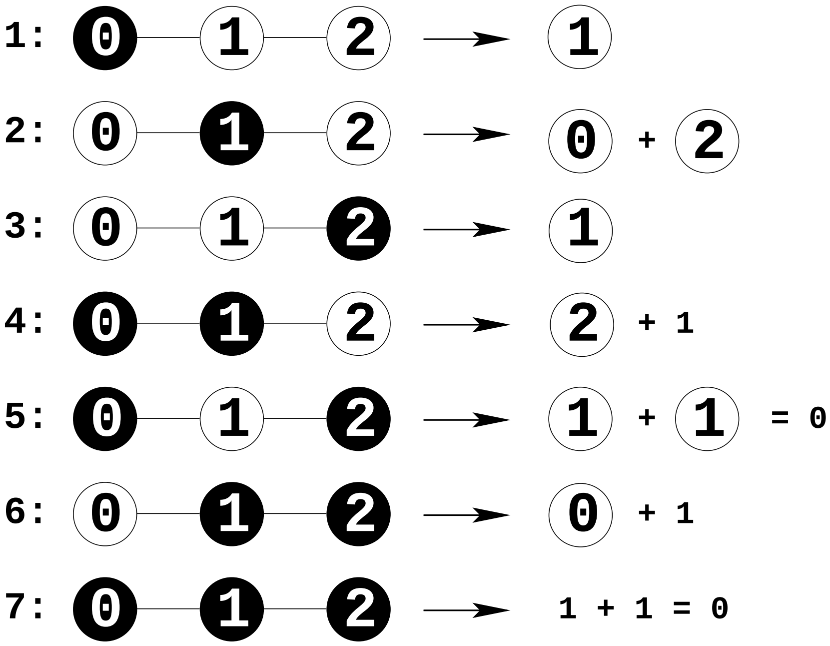

Figure 3.4: A line graph for the MAACF for $s(x) = x_0x_1 + x_1x_2$

set	k	$a_{n,k}(x)$	$\deg_(a_{n,k}(x))$	$a_{n,k}$
I	...00	$a_{n-1,k_{n-2}=0}(x)$	1	0
II	...10	$a_{n-2,k}(x) + x_{n-1}$	1	0
III	..001	$a_{n-2,k_{n-3}=0}(x) + x_{n-2}$	1	0
	..101	$a_{n-2,k_{n-3}=1}(x)$	≥ 0	$2^{n-wt(k)} = 2a_{n-2,k}$
IV	...11	$a_{n-1,k_{n-2}=1} + 1$	≥ 0	$2^{n-wt(k)} = a_{n-1,k}$

Proof for set I

For the first set $a_{n,k_{n-2}=0,k_{n-1}=0}(x)$ will be the same as $a_{n-1,k_{n-2}=0}$ because of rule *iii* in (3.19). Also $\deg(a_{n-1,k_{n-2}=0})$ must be 1 when $1 \leq k < 2^{n-2}$, because $k_j = 1$ is true for at least one value of $1 \leq j < n-2$, and therefore we have case i at least once.

Proof for set II

For the values where k ends with $k_{n-2} = 1$ and $k_{n-1} = 0$, we can split it in two. The first $n-2$ bits of k is the same as $a_{n-2,k}(x)$, and the last two bits gives an addition of x_{n-1}. It is then straightforward to show that if we add x_{n-1} to all the 2^{n-2} possible functions $a_{n-2,k}$ they will all have a degree of 1.

Proof for set III

When $k_{n-2} = 0$ and $k_{n-1} = 1$ the degree of $a_{n-k}(x)$ depends on k_{n-3}. If $k_{n-3} = 0$ the degree will be 1, for the same reasons as in set II. But if $k_{n-3} = 1$, the two x_{n-2} terms cancel each other, and $a_{n,k}(x) = a_{n-2,k_{n-3}=1}(x)$. We know that for half the values of $k = (k_0, k_1, ..., k_{n-3})$, $\deg(a_{n-2,k_{n-3}=1}(x))$ is 0, and for those values $a_{n,k} = 2^{n-wt(k)}$. If we let $n' = n-2$ and $w = wt(k_0, k_1, ..., k_{n-3}) = wt(k) - 1$, we can write this as

$$a_{n,k} = 2^{n-wt(k)} = 2^{n'-w+1} = 2a_{n-2,k}$$

This is also true when $k_{n-3} = 0$ as shown above, so we do not have to fix k_{n-3}.

Proof for set IV

The last case is when $k_{n-1} = k_{n-2} = 1$. Here we can just use $a_{n-1,k_{n-2}=1}$ with the addition of a constant because of rule i in (3.19). And for half the values of $k = (k_0, k_1, ..., k_{n-2})$ we know that $\deg(a_{n-1,k_{n-2}=1}) = 0$ since a GRS sequence of length $N = 2^{n-1}$ does not have perfect MAACF. Let then $n' = n-1$ and $w = wt(k_0, k_1, ..., k_{n-2}) = wt(k) - 1$. Then

$$a_{n,k} = 2^{n-wt(k)} = 2^{n'-w} = a_{n-1,k}$$

In summary we have the following relationship

$$a_{n,k} = \begin{cases} 2a_{n-2,k} & \text{if } k_{n-1} = 1 \text{ and } k_{n-2} = 0. \\ a_{n-1,k} & \text{if } k_{n-1} = 1 \text{ and } k_{n-2} = 1. \\ 0 & \text{otherwise} \end{cases} \tag{3.20}$$

and this gives us the recursion

$$\sigma_n = \sum_{k=1}^{2^n-1} |a_{n,k}|^2 = \sum_{k=1}^{2^{n-2}-1} |2a_{n-2,k}|^2 + \sum_{k=1}^{2^{n-1}-1} |a_{n-1,k}|^2 = 4\sigma_{n-2} + \sigma_{n-1} \tag{3.21}$$

Q.E.D.

It can be shown that 3.14 follows directly from 3.21, but we omit this proof.

3.4.1 A multidimensional look at the 1-dimensional MF

In section 2.9 we saw that the asymptotic 1-dimensional Merit Factor for a subset of the Golay-Rudin-Shapiro sequences is three and that the recursion for the 1-dimensional sum-of-squares is

$$\sigma_n = 2\sigma_{n-1} + 8\sigma_{n-2}$$

If we look at the multidimensional definition of GRS sequences (3.11) we can see that the proof in [9] only holds for the case

$$s(x) = \sum_{j=0}^{n-2} x_j x_{j+1} + \sum_{j=0}^{n-1} c_j x_j + d \tag{3.22}$$

However the complete set of GRS sequences is much bigger. The complete set of GRS sequences is of size

$$\frac{n!}{2} \cdot 2^{n+1}$$

while the subset proven in [9] has 2^{n+1} sequences. As $n \to \infty$ the complete set will be <u>much</u> larger.

For the complete set there seems to be a wider range on the 1-dimensional sum-of-squares values. Table 3.4.1 shows all the different sum-of-squares values for each $2 \leq n \leq 8$. The middle row of σ_n-values is the subclass following the recursion of [9]. The general case appears to follow another similar recursion

$$\sigma_n = 2\sigma_{n-1} + 8\sigma_{n-2} \pm 2^r \tag{3.23}$$

where $r \in \{-\infty, n, 2n, ...\}$.

n	σ_n			# of classes
2		2		1
3		12		1
4	24	40	56	3
5	112, 144	176	208, 240	5
6	416, 480, 544, 608	672	736, 800, 864, 928	9
7	1856, 1984, 2112, 2240, 2368, 2496, 2624	2752	2880, 3008, 3136, 3264, 3392, 3520	14
8	7296, 8064, 8320, 8576, 8832, 9088, 9344, 9600, 9856, 10112, 10368, 10624	10880	11136, 11392, 11648, 11904, 12160, 12416, 12672, 12928, 13184, 13440, 13696, 14208, 14464	26

Table 3.3: 1-dimensional σ_n for the Golay-Rudin-Shapiro sequences

3.5 A cubic Construction with good MMF

If one takes a look at the result of the exhaustive search in Table 3.1 we can see that the best possible binary sequences for length 16 and 32 are cubic. Using four cubic terms as a seed I was able to find a new construction with very high Multidimensional Merit Factor. The new construction can be defined as

$$s(x) = c(X) + x_{\pi(2)}x_{\pi(n-1)} + \sum_{j=2}^{n-2} x_{\pi(j)}x_{\pi(j+1)} + (\sum_{i=0}^{n-1} b_i x_i) + d, \quad b_i, d \in Z_2 \tag{3.24}$$

where $c(X) = x_{\pi(0)}x_{\pi(1)}x_{\pi(2)} + x_{\pi(0)}x_{\pi(1)}x_{\pi(3)} + x_{\pi(0)}x_{\pi(2)}x_{\pi(3)} + x_{\pi(1)}x_{\pi(2)}x_{\pi(3)}$, where π is any permutation of Z_n and where $x_{n-1}, x_{n-1}, ..., x_0$ are boolean variables and s is the length $N = 2^n$ binary sequence such that

$$s_i = s(x_{n-1} = i_{n-1}, x_{n-2} = i_{n-2}, ..., x_0 = i_0) \tag{3.25}$$

where $i_{n-1}, i_{n-2}, ..., i_0$ is the binary representation of the integer i.

As a result of the cubic seed, the construction only works for lengths from $N \geq 2^4 = 16$. The construction turned out to have an even higher MMF than the GRS sequences. A list of the MMF for sequences up to length $N = 4096$

$N=2^n$	MMF	σ_n	CANF
16	128.00	2	$012, 013, 023, 123, 23, 23,$
32	24.38	21	$012, 013, 023, 123, 23, 34, 24,$
64	81.92	25	$012, 013, 023, 123, 23, 34, 45, 25,$
128	75.16	109	$012, 013, 023, 123, 23, 34, 45, 56, 26,$
256	156.78	209	$012, 013, 023, 123, 23, 34, 45, 56, 67, 27,$
512	203.21	645	$012, 013, 023, 123, 23, 34, 45, 56, 67, 78, 28,$
1024	354.01	1481	$012, 013, 023, 123, 23, 34, 45, 56, 67, 78, 89, 29,$
2048	516.41	4061	$012, 013, 023, 123, 23, 34, 45, 56, 67, 78, 89, 9A, 2A,$
4096	840.12	9985	$012, 013, 023, 123, 23, 34, 45, 56, 67, 78, 89, 9A, AB, 2B,$

Table 3.4: MMF for a cubic Construction

can be found in Table 3.4.

Theorem 3 *The number of sequences of length $N = 2^n$ in the class of sequences $\mathcal{C}_n$ that satisfy (3.24) will be*

$$|\mathcal{C}_n| = \frac{n!}{4} \cdot 2^{n+1}$$

Proof of Theorem 3

Let there be n variables, $x_0, x_1, \ldots, x_{n-1}$. There are then $\binom{n}{4}$ ways of choosing the four cubic variables out of n variables. For the GRS part that are connected to the cubic seed, there are $\binom{4}{2} = 6$ ways of choosing two variables out of the four variables that form the cubic. These two variables are then used for the start and end links of the quadratic part. For the rest of the GRS part there are another $(n-4)$ variables that are used to form the quadratic section. There are $(n-4)!$ ways of ordering these variables. There are then 2^{n+1} possible linear terms and constant offsets. We then have

$$\begin{aligned} |\mathcal{C}_n| &= \tbinom{n}{4} \cdot \tbinom{4}{2} \cdot (n-4)! \cdot 2^{n+1} \\ &= \tfrac{n(n-1)(n-2)(n-3)}{4!} \cdot \tfrac{4!}{4} \cdot (n-4)! \cdot 2^{n+1} \\ &= \tfrac{n!}{4} 2^{n+1} \end{aligned}$$

Q.E.D.

Looking at the sum-of-squares σ_n for the aperiodic multidimensional autocorrelation we can see that this new cubic construction follows the same recursion as the Golay-Rudin-Shapiro sequences (see Section 3.4).

$$\sigma_n = 4\sigma_{n-2} + \sigma_{n-1}$$

But the relationship with the Golay-Rudin-Shapiro sequences is even closer, as the following theorem will show.

Theorem 4 *Let σ'_n be the sum-of-squares for any Golay-Rudin-Shapiro sequence of length $N = 2^n$. Then the sum-of-squares σ_n for any binary sequences that satisfy (3.24) follow the recursion*

$$\sigma_n = \sigma'_{n-2} + 16\sigma'_{n-4}.$$

Before we show the proof of Theorem 4, the computation of $n \leq 6$ is done by hand for the construction of (3.24). Table 3.5 lists the non-zero a_k values and the σ_n-values for $4 \leq n \leq 6$.

$n=4$ $k_3k_2k_1k_0$	$\lvert a_k\rvert$	$n=5$ $k_4k_3k_2k_1k_0$	$\lvert a_k\rvert$	$n=6$ $k_5k_4k_3k_2k_1k_0$	$\lvert a_k\rvert$
1111	1	11111	1	111111	1
		01111	2	101111	2
		10011	4	011111	2
				110011	4
$\sigma_4 = 1$		$\sigma_5 = 21$		$\sigma_6 = 25$	

Table 3.5: Non-zero $|a_k|$ values for the cubic construction

Proof of Theorem 4

Let us rewrite $\mathbf{s}(x)$ (3.24) as

$$s(x) = x_0x_1x_2 + x_0x_1x_3 + x_0x_2x_3 + x_1x_2x_3 + x_2x_3 + \mathrm{GRS}(x') \qquad (3.26)$$

where $x' = (x_3, x_4, ..., x_{n-1}, x_2)$ and $\mathrm{GRS}(x') = x_3x_4 + x_4x_5 + ... + x_{n-1}x_2$ is a general Golay-Rudin-Shapiro sequence (see equation 3.18) of length $N = 2^{n-2}$. Since the MMF of (3.24) is unchanged for any permutation, the proof of the case $\pi(k) = k$, $k \in Z_n$, also holds for all other π. Let the aperiodic multidimensional autocorrelation $a_{n,k}$ for the binary sequence s be represented as a function

$$a_{n,k}(x) = s(x)_{x_i=0} + s(x)_{x_i=1}, \qquad \forall\ i \text{ where } k_i = 1 \qquad (3.27)$$

where k has a binary representation $k_0, k_1, ..., k_{n-1}$, $k_i \in Z_2$. Then let $a'_{n-2,k}(x')$ be the multidimensional aperiodic autocorrelation for $\mathrm{GRS}(x')$ (see equation 3.16)

We can then split the calculation of $a_{n,k}$ into 16 different sets depending on the first 4 bits of k, and look at each set. Table 3.6 shows $a_{n,k}(x)$ for these 16 different sets. For this cubic construction we have this relationship between the degree of $a_{n,k}(x)$ and the value of $a_{n,k}$

$$a_{n,k} = \begin{cases} 2^{n-wt(k)} & \text{if } \deg(a_{n,k}(x)) = 0 \\ 0 & \text{if } deg(a_{n,k})(x) = 1 \\ 0 & \text{if } \deg(a_{n,k})(x) = 2 \end{cases} \qquad (3.28)$$

It is straightforward to compute the value of $a_{n,k}$ when the degree of $a_{n,k}(x) < 2$. Then let $q_i(x)$, $0 \leq i \leq 3$ be one of the 4 possible quadratics from table 3.6,

$$\begin{aligned} q_0(x) &= \ x_1x_2 + x_1x_3 + x_2x_3 \\ q_1(x) &= \ x_0x_2 + x_0x_3 + x_2x_3 \\ q_2(x) &= \ x_0x_1 + x_0x_3 + x_1x_3 \\ q_3(x) &= \ x_0x_1 + x_0x_2 + x_1s_2 \end{aligned}$$

and let $l_1(x) = \sum_{i=0}^{3}(c_ix_i)$ and $l_2(x) = \sum_{i\neq 0,1,2,3}^{n-1}(c_ix_i)$. From the definition of the aperiodic multidimensional autocorrelation we know that if $a_{n,k}(x)$ is balanced then $a_{n,k} = 0$. By a balanced boolean function we mean that the binary

set	$k_0k_1k_2k_3$	$a_{n,k}(x)$	$\deg(a_{n,k}(x))$	$a_{n,k}$
0	0000	$a'_{n-4,k_2=k_3=0}(x')$	1	0
1	1000	$x_1x_2 + x_1x_3 + x_2x_3 + a'_{n-4,k_2=k_3=0}(x')$	2	0
2	0100	$x_0x_2 + x_0x_3 + x_1x_3 + a'_{n-4,k_2=k_3=0}(x')$	2	0
3	**1100**	$x_2 + x_3 + a'_{n-4,k_2=k_3=0}(x)$	0	$4a'_{n-4,k}$
4	0010	$x_0x_1 + x_0x_3 + x_1x_3 + x_3 + a'_{n-2,k_2=1,k_3=0}(x')$	2	0
5	1010	$x_1 + a'_{n-2,k_2=1,k_3=0}(x')$	1	0
6	0110	$x_0 + a'_{n-2,k_2=1,k_3=0}(x')$	1	0
7	1110	$1 + a'_{n-2,k_2=1,k_3=0}(x')$	1	0
8	0001	$x_0x_1 + x_0x_2 + x_1x_2 + x_2 + a'_{n-2,k_2=0,k_3=1}(x')$	2	0
9	1001	$x_1 + a'_{n-2,k_2=0,k_3=1}(x')$	1	0
10	0101	$x_0 + a'_{n-2,k_2=0,k_3=1}(x')$	1	0
11	1101	$1 + a'_{n-2,k_2=0,k_3=1}(x')$	1	0
12	0011	$x_0 + x_1 + 1 + a'_{n-2,k_2=k_3=1}(x')$	1	0
13	1011	$x_1 + a'_{n-2,k_2=k_3=1}(x')$	1	0
14	0111	$x_0 + a'_{n-2,k_2=k_3=1}(x')$	1	0
15	**1111**	$1 + a'_{n-2,k_2=k_3=1}(x')$	0	$a'_{n-2,k}$

Table 3.6: $a_{n,k}(x)$ when we are fixing k_0, k_1, k_2, k_3

representation of $a_{n,k}(x)$ has the same number of 0s and 1s, or in other words one half of the autocorrelations is equal and the other half is not.

When $wt(k_0k_1k_2k_3) = 1$ we can write

$$a_{n,k}(x) = q_i(x) + l_1(x) + l_2(x) \tag{3.29}$$

Since it is straightforward to show that $a_{n,k}(x)$ is balanced when $\deg(l_2(x)) = 1$ (because $g_i(x)$ is balanced $\forall i$, and the variables in $l_2(x)$ are disjoint from the variables in $q_i(x)$), we let $l_2(x) = 0$ and compute all possible functions of the form (3.29). This can be done easily on a computer since we only have 4 variables x_i for the function. It turns out that $a_{n,k}(x)$ is always balanced except when $a_{n,k}(x) = q_0(x) + x_2$ and $a_{n,k}(x) = q_1(x) + x_2$. Looking at Table 3.6 (set 1 and 2) we see that this is only possible when $a'_{n-4,k_2=k_3=0}(x) = x_2$ but, for $n \geq 6$, $a'_{n-4,k_2=k_3=0}(x)$ can never be x_2 if we follow the rules from (3.19) for GRS sequences. This gives us a strong relationship between the degree of $a_{n,k}(x)$ and the value of $a_{n,k}$

$$\deg(a_{n,k}(x)) > 0 \Leftrightarrow a_{n,k} = 0$$

We now need to take a look at the two sets in Table 3.6 where the degree is 0.

A closer look at set 3

It looks like the $deg(a_{n,k}(x)) > 0$, but since we know $a'_{n-2,k_2=k_3=0}(x)$ is of degree 1 it is possible that $a'_{n-2,k_2=k_3=0}(x) = x_2+x_3$, thus making the $deg(a_{n,k}(x)) = 0$. This is only possible if $k_4 = k_{n-1} = 1$, which gives us this relationship

$$a_{n,k_0=k_1=1,k_2=k_3=0}(x) = a'_{n-4,k_4=k_{n-1}=1}(x)$$

Then let $n' = n - 4$, $w = wt(k_4, k_5, ..., k_{n-1}) = wt(k) - 2$ and $a'_{n-4,k} = 2^{n'-w}$. Then

$$a_{n,k_0=k_1=1,k_2=k_3=0} = 2^{n-wt(k)} = 2^2 * 2^{n'-w} = 4 * a'_{n-4,k} \tag{3.30}$$

A closer look at set 15

The other set where $deg(a_{n,k}(x)) = 0$ is when we fix $k_0 = k_1 = k_2 = k_3 = 1$. This is true whenever $deg(a_{n-2,k}(x)) = 0$. Let $n' = n-2$, $w = wt(k_2, k_3, ..., k_{n-1} = wt(k) - 2$ and $a'_{n-2,k} = 2^{n'-w}$. We then have

$$a_{n,k_0=k_1=k_2=k_3=1} = 2^{n-wt(k)} = 2^{n'-w} = a'_{n-2,k} \tag{3.31}$$

To summarise we have the following relationship

$$a_{n,k} = \begin{cases} 4a'_{n-4,k} & \text{if } k_0 = k_1 = 1 \text{ and } k_2 = k_3 = 0. \\ a'_{n-2,k} & \text{if } k_0 = k_1 = k_2 = k_3 = 1. \\ 0 & \text{otherwise} \end{cases} \tag{3.32}$$

and this gives us the recursion

$$\sigma_n = \sum_{k=1}^{2^n-1} |a_{n,k}|^2 = \sum_{k=1}^{2^{n-4}-1} |4a'_{n-4,k}|^2 + \sum_{k=1}^{2^{n-2}-1} |a'_{n-2,k}|^2 = \sigma'_{n-2} + 16\sigma'_{n-4} \tag{3.33}$$

This proof also holds if we add a linear term and/or a constant to (3.26), because the degree of $a_{n,k}$ is unchanged. It follows from (3.27) that the addition to the $a_{n,k}(x)$ will be 0 or 1 if we add a linear term and/or a constant to $s(x)$.

Q.E.D.

Corollary 1 *Let s be a binary sequence described by (3.24). Then the sum-of-squares σ_n for s follows the recursion*

$$\sigma_n = \sigma_{n-1} + 4\sigma_{n-2}$$

Proof

From Theorem 4 we know that

$$\sigma_n = \sigma'_{n-2} + 16\sigma'_{n-4}. \tag{3.34}$$

$$\sigma_{n-1} = \sigma'_{n-3} + 16\sigma'_{n-5} \tag{3.35}$$

$$\sigma_{n-2} = \sigma'_{n-4} + 16\sigma'_{n-6} \tag{3.36}$$

and from Theorem 2

$$\sigma'_{n-2} = \sigma'_{n-3} + 4\sigma'_{n-4} \tag{3.37}$$

$$\sigma'_{n-4} = \sigma'_{n-5} + 4\sigma'_{n-6} \tag{3.38}$$

Then we can then use (3.37) and (3.38) and substitute into (3.34)

$$\begin{aligned} \sigma_n &= \sigma'_{n-2} + 16\sigma'_{n-4} \\ &= \sigma'_{n-3} + 20\sigma'_{n-4} \\ &= \sigma'_{n-3} + 4\sigma'_{n-4} + 16\sigma'_{n-5} + 64\sigma'_{n-6} \\ &= (\sigma'_{n-3} + 16\sigma'_{n-5}) + 4(\sigma_{n-4} + 16\sigma'_{n-6}) \end{aligned}$$

and then substitute using (3.35) and (3.36) to get

$$\sigma_n = \sigma_{n-1} + 4\sigma_{n-2}$$

Q.E.D.

Chapter 4

Conclusion

In this thesis we have looked at new techniques to find binary sequences with low aperiodic autocorrelation. When it comes to exhaustive search, the algorithms proposed here, for both lexicographic and m-sequence ordering, are not fast enough to compete with the algorithm used by Mertens [18] [19]. But we have illustrated how hard it is to find the highest Merit Factor through an exhaustive search. As the problem appears to be NP-complete the only road to finding long sequences with high Merit Factor is the use of constructed classes of sequences.

Two new methods of construction are presented in this thesis. The first one is to interleave two Legendre sequences to construct a skewsymmetric sequence. This new class of sequences appears to have an asymptotic Merit Factor of three. The second class of sequences can be found by using a semi-construction. Combining the properties of Legendre sequences with a small limited search we were able to find new extended sequences with high Merit Factor. Using an algorithm with complexity $O(N^2)$ we can find new sequences with MF > 6.3 for large N, by extending a length N Legendre sequence.

In a new field of research we look at the multidimensional aperiodic autocorrelations (MAACF) for sequences of length $N = 2^n$. Here we show that the multidimensional autocorrelations for Golay-Rudin-Shapiro sequences have a low sum-of-squares for the aperiodic case. We also define a new cubic class of sequences and show that this has an even better sum-of-squares. The multidimensional sum-of-squares for both GRS and for the cubic construction can be computed recursively, without the need for explicit autocorrelation computation.

This thesis also proposes some ideas for further research. One of them is to use de Brujin sequences to order an exhaustive search. There is a chance that one of these sequence orderings will lead to the exposure of high MF sequences without search. Another idea for further research is to take a closer look at the Extended Legendre semi-construction. Some ideas might be to use the extended sequence in a new extended search, or to look for other constructions that will be improved using the extended directed search. It would also be interesting to take a closer look at the extension we get from the directed search. Is it possible to find the extension without the directed search ?

List of Tables

List of Figures

Bibliography

[1] J. A. Davis and J. Jedwab, "Peak-to-mean Power Control in OFDM, Golay Complementary Sequences, and Reed-Muller Codes"
IEEE Trans on Information Theory, Vol 45, No 7, pp 2397-2417, Nov 1999

[2] H. Fredricksen, "A Survey of full length nonlinear shift register cycle algorithms",
SIAM Review Vol 24, & No.2, pp 195-221, April 1982

[3] M.J.E.Golay, "Complementary series",
IEEE Trans on Information Theory, Vol 7, pp 82-87, Apr 1961

[4] M.J.E.Golay, "A Class of Finite Binary Sequences With Alternate Autocorrelation Values Equal to Zero",
IEEE Trans on Information Theory, Vol 18, No 3, pp 449-450, May 1972

[5] M.J.E.Golay, "Sieves for Low Autocorrelation Binary Sequences",
IEEE Trans on Information Theory, Vol 23, No 1, pp 43-51, Jan 1977

[6] M.J.E.Golay, "The merit factor of long low autocorrelation binary sequences",
IEEE Trans on Information Theory, Vol 28, No 3, pp 543-549, May 1982

[7] M.J.E.Golay, "The Merit Factor of Legendre sequences",
IEEE Trans on Information Theory, Vol 29, No 6, pp 934-936, Nov 1983

[8] M.J.E.Golay, "A New Search for Skewsymmetric Binary Sequences with Optimal Merit Factors",
IEEE Trans on Information Theory, Vol 36, No 5, pp 1163-1166, Sept 1990

[9] T. Høholdt, H. Jensen and J. Justesen, "Aperiodic Correlation and the Merit Factor of a Class of Binary Sequences",
IEEE Trans on Information Theory, Vol 31, No 4, pp 549-552, Jul 1985

[10] T. Høholdt, H. Jensen and J. Justesen, "Autocorrelation Properties of a Class of Infinite Binary Sequences",
IEEE Trans on Information Theory, Vol 32, No 3, pp 430-431, Jan 1986

[11] T. Høholdt and H. Jensen, "Determination of the Merit Factor of Legendre Sequences",
IEEE Trans on Information Theory, Vol 34, No 1, pp 161-164, Jan 1988

[12] T.Høholdt, "The Merit Factor of Binary Sequences",
Difference Sets, Sequences and Their Correlation Properties
Kluwer Academic Publishers, Dordrecht 1999, pp 227-237

[13] J. Jensen, H. Jensen and T. Høholdt, "The Merit Factor of Binary Sequences Related to Difference Sets",
IEEE Trans on Information Theory, Vol 37, No 3, pp 617-626, Jan 1991

[14] A. Kirilusha and G. Narayanaswamy, "Construction of New Asymptotic Classes of binary sequences based on existin asymptotic classes",
http://www.mathcs.richmond.edu/~jad/summerwork/extendedLegendre.pdf

[15] R.A. Kristiansen, "Examples of long-length binary sequences with MF> 6.3 constructed using the Extended Legendre Semi- Construction",
Web pages: http://www.ii.uib.no/~raymond/thesis/examples.html or
http://www.ii.uib.no/~matthew/Examples.html

[16] H. D. Luke, "Sequences and arrays with perfect periodic correlation"
IEEE Trans on Aerospace and Electronic Systems, vol 24, no 3, pp 287-294, May 1988

[17] S. Maitra, "Autocorrelation Properties of Correlation Immune Boolean Functions"
Springer-Verlag LNCS 2247, INDOCRYPT 2001, pp 242-253 , 2001

[18] Stephan Mertens, "Exhaustive search for low-autocorrelation binary sequences",
J. Phys. A, Vol. 29, pp L473-L481, 1996

[19] S. Mertens, The Bernasconi model
Web page: http://odysseus.nat.uni-magdeburg.de/~mertens/bernasconi/

[20] Burkhard Militzer, Michele Zamparelli and Dieter Beule, "Evolutionary Search for Low Autocorrelated Binary sequence",
IEEE Trans on Evolutionary Computation, Vol 2, No 1, pp 34-39, Apr 1998

[21] M. G. Parker, "Even Length Binary Sequence Families with Low Negaperiodic Autocorrelation"
Springer-Verlag LNCS 2227, AAECC-14 Pros., pp 200-210 , Nov 2001

[22] M.G.Parker, K.G.Paterson and C.Tellambura Golay Complementary Sequences
Wiley Encyclopedia of Telecommunications, Wiley Interscience, 2003

[23] A. Ralston, "De Bruijn Sequences - A model Example of the Interaction of Discrete Mathematics and Computer Science",
Mathematics Magazine, Vol 55, No 3, pp 131-143 , May 1982

[24] W. Rudin, "Some theorems on Fourier coefficients"
Proc. American Math. Soc., Vol 10, pp 855-859, 1959

[25] J. Storer and R. J. Turyn, "On Binary Sequences"
Proc. American Math. Soc., vol 12, pp 394-399, 1961

[26] R. J. Turyn, "Hadamard matrices, Baumert-Hall units, four-symbol sequences, pulse compression, and surface wave encoding" *J. Comb. Theory Ser. A*, vol 16, pp 313-333, 1974