Open Problems on the Cross-correlation of m-Sequences

Tor Helleseth

Abstract Pseudorandom sequences are important for many applications in communication systems, in coding theory, and in the design of stream ciphers. Maximum-length linear sequences (or m-sequences) are popular in sequence designs due to their long period and excellent pseudorandom properties. In code-division multiple-access (CDMA) applications, there is a demand for large families of sequences having good correlation properties. The best families of sequences in these applications frequently use m-sequences in their constructions. Therefore, the problem of determining the correlation properties of m-sequences has received a lot of attention since the 1960s, and many interesting theoretical results of practical interest have been obtained. The cross-correlation of m-sequences is also related to other important problems, such as almost perfect nonlinear functions (APN) and almost bent functions (AB), and to the nonlinearity of S-boxes in many block ciphers including AES. This chapter gives an updated survey of the cross-correlation of m-sequences and describes some of the most important open problems that still remain in this area.

1 Introduction

Let $\{u_t\}$ and $\{u_t\}$ be sequences of period ε with symbols from the finite field GF(p) with p elements. Let ω be a primitive complex pth root of unity. The cross-correlation between the two sequences at shift τ is defined to be

$$C_{u,v}(\tau) = \sum_{t=0}^{\varepsilon-1} \omega^{u_{t+\tau}-v_t}.$$

If the two sequences are cyclically equivalent (i.e., only differ by a cyclic shift), the correlation is denoted autocorrelation instead of cross-correlation.

T. Helleseth (✉)
Department of Informatics, The Selmer Center, University of Bergen, Thormøhlensgate 55, 5008 Bergen, Norway
e-mail: Tor.Helleseth@ii.uib.no

Ç.K. Koç (ed.), *Open Problems in Mathematics and Computational Science*,
DOI 10.1007/978-3-319-10683-0_8

In a code-division multiple-access (CDMA) system, each user is assigned a sequence from a family of sequences. The quality of the communication depends on the selection of a family of sequences with good parameters.

Let $\mathscr{F}$ be a family of M cyclically distinct sequences of the same period ε:

$$\mathscr{F} = \{\{s_t^{(i)}\} \mid 1 \leq i \leq M\}.$$

The most important parameters for evaluating the quality of the family are $(M, \varepsilon, \theta_{\max})$ where $\theta_{\max}$ is the maximum value of the absolute magnitude of the (nontrivial) auto- and cross-correlation between any two sequences in the family, i.e.,

$$\theta_{\max} = \max\{|C_{s^{(i)},s^{(j)}}(\tau)| \mid i \neq j \text{ or } \tau \neq 0\}.$$

Many of the best sequence families can be constructed from linear recursions. To generate a sequence $\{s_t\}$ with symbols from GF(p), one can use a linear recursion of degree n and generate each symbol from the previous n symbols such that

$$s_{t+n} + c_{n-1}s_{t+n-1} + \cdots + c_0 s_t = 0, \quad c_i \in \mathrm{GF(p)}, c_0 \neq 0.$$

The initial state $(s_0, s_1, \ldots, s_{n-1})$ and the linear recursion uniquely determine the sequence $\{s_t\}$. Thus, the linear recursion generates p^n distinct sequences corresponding to the p^n initial states $(s_0, s_1, \ldots, s_{n-1})$. Clearly, some of these generated sequences may be cyclically equivalent.

The characteristic polynomial of the linear recursion is defined to be

$$f(x) = \sum_{i=0}^{n} c_i x^i.$$

The period of the sequences generated by the recursion with characteristic polynomial $f(x)$ is completely determined by the polynomial. It is a well-known fact that all these sequences will have period e where e is the smallest positive integer such that $f(x) \mid x^e - 1$. Furthermore, at least one of these sequences will have e as its smallest period.

Let $f(x)$ be a primitive polynomial, i.e., an irreducible polynomial with a zero α being a generator for the multiplicative group of GF(p$^{\mathrm{n}}$). Then the factorization of the primitive polynomial $f(x)$ is given by

$$f(x) = \prod_{i=0}^{n-1} (x - \alpha^{p^i}).$$

Since the generator α has order $p^n - 1$, then $f(x) \mid x^{p^n-1} - 1$ and any nonzero sequence generated by the recursion with characteristic polynomial $f(x)$ has period $p^n - 1$. This is maximum possible for a linear recursion of degree n, and any such sequence is therefore called a maximum-length sequence (or m-sequence).

During a period of the m-sequence, each nonzero consecutive n-tuple occurs exactly once during its period. In particular this implies that the m-sequence is as balanced as it can be for a sequence of period $p^n - 1$ since all nonzero symbols occur p^{n-1} times, while the 0 element occurs $p^{n-1} - 1$ time.

The trace function Tr_n from GF(p$^{\text{n}}$) to GF(p) is defined by

$$Tr_n(x) = \sum_{i=0}^{n-1} x^{p^i}.$$

The m-sequence can be written as

$$s_t = Tr_n(c\alpha^t),$$

where the p^n-1 different nonzero values of $c \in \mathrm{GF(p^n)}^* = \mathrm{GF(p^n)}\setminus\{0\}$ correspond to all possible shifts of the m-sequence.

Starting with one m-sequence of period $p^n - 1$, all other m-sequences of the same period can be obtained by decimating the sequence. The decimated sequence of $\{s_t\}$ is the sequence $\{s_{dt}\}$ which is an m-sequence if and only if $\gcd(d, p^n - 1) = 1$. The sequence and its decimated sequence are cyclically distinct if and only if $d \not\equiv p^i \pmod{p^n - 1}$ for $i = 0, 1, \ldots, n - 1$. The number of cyclically distinct m-sequences is $\phi(p^n - 1)/n$, where ϕ is Euler's ϕ function, and equals the number of primitive polynomials of degree n. For further results on linear recursions, the reader is referred to the classical book by Golomb [10].

Example 1 Let $p = 3$ and consider the linear recursion

$$s_{t+3} + 2s_{t+2} + s_t = 0.$$

The characteristic polynomial of the recursion is $f(x) = x^3 + 2x^2 + 1$. This is a primitive polynomial, and using the initial state (011), the recursion generates the m-sequence (01110211210100222012212020...) of period $\varepsilon = 26$. The recursion clearly generates all cyclic shifts of this sequence since all nonzero initial states are present in the m-sequence. In addition the recursion generates the all-zero sequence using the initial state (000). It is easily verified that decimating the sequence above by $d \equiv 3^i \pmod{26}$ gives the same sequence, while decimation by any $d \not\equiv 3^i \pmod{26}$ with $\gcd(d, 26) = 1$ gives a cyclically distinct m-sequence of period 26.

2 Correlation of m-Sequences

The cross-correlation at shift τ between two m-sequences that differ by a decimation d will be denoted by $C_d(\tau)$. The problem to determine the values and the number of occurrences of each value of the cross-correlation $C_d(\tau)$ between two m-sequences when τ runs through all $p^n - 1$ shifts has been studied for almost 50 years.

The simplest case to consider is the autocorrelation function of an m-sequence. One reason for the popularity of m-sequences is due to their two-valued autocorrelation and their importance in synchronization applications.

Theorem 1 *The autocorrelation function $C_1(\tau)$ of an m-sequence having period $\varepsilon = p^n - 1$ takes the value -1 for any shift $\tau \not\equiv 0 \pmod{p^n - 1}$ and the value $p^n - 1$ for any shift $\tau \equiv 0 \pmod{p^n - 1}$.*

Proof Let $\tau \not\equiv 0 \pmod{p^n - 1}$, then since the characteristic polynomial of the m-sequence $\{s_t\}$ also generates the sequence $\{s_{t+\tau} - s_t\}$, it follows that this is some shift of the m-sequence. Hence,

$$C_1(\tau) = \sum_{t=0}^{p^n-2} \omega^{s_{t+\tau}-s_t} = \sum_{t=0}^{p^n-2} \omega^{s_{t+\delta}} = -1$$

since $s_t - s_{t+\tau} = s_{t+\delta}$ for some δ depending on τ and the m-sequence is balanced (except for a "missing" 0) having p^{n-1} of each nonzero element and $p^{n-1} - 1$ zeros during a period of the sequence.

Some basic results useful for the analysis of $C_d(\tau)$ can be found in Helleseth [12].

Lemma 1 *The following properties hold for the cross-correlation $C_d(\tau)$:*

1. *If $dd' \equiv 1 \pmod{p^n - 1}$ or $d' \equiv dp^i$ for some integer i, then $C_d(\tau)$ and $C_{d'}(\tau)$ have the same correlation values with the same number of occurrences.*
2. *The value of $C_d(\tau)$ is a real number.*
3. *The sum of the cross-correlation values is determined by*

$$\sum_{\tau=0}^{p^n-2} (C_d(\tau) + 1) = p^n.$$

4. *The square sum of the cross-correlation values is determined by*

$$\sum_{\tau=0}^{p^n-2} (C_d(\tau) + 1)^2 = p^{2n}.$$

5. *The higher-order power sums of the cross-correlation are given by*

$$\sum_{\tau=0}^{p^n-2} (C_d(\tau))^r = -(p^n-1)^{r-1} + 2(-1)^{r-1} + a_r p^{2n}$$

where a_r is the number of solutions of the equations

$$x_1 + x_2 + \cdots + x_{r-1} + 1 = 0$$
$$x_1^d + x_2^d + \cdots + x_{r-1}^d + 1 = 0$$

and $x_i \in \mathrm{GF}(p^n)^*$ for $i = 1, 2, \ldots, r-1$.

The lemma above is useful to determine the number of occurrences of each value in $C_d(\tau)$ when there are rather few, say r, values that have already been determined. Then one can determine the complete distribution of the cross-correlation if one can find a_i for $2 < i < r$.

In Helleseth [12], the previous lemma was applied to prove a result first mentioned without proof in Golomb [11].

Theorem 2 *If $d \notin \{1, p, \ldots, p^{n-1}\}$ then $C_d(\tau)$ takes on at least three different values when $\tau = 0, 1, \ldots, p^n - 2$.*

Proof Suppose that $C_d(\tau)$ takes on only the two values x and y that occur r and $p^n - 1 - r$ times, respectively, in $C_d(\tau)$ when τ runs through all shifts $\tau = 0, 1, \ldots, p^n - 2$. Then using (3) and (4) in Lemma 1 leads to two equations in three unknowns x, y, and r. Eliminating r leads to the equation

$$(p^n x - (x+1))(p^n y - (y+1)) = p^{2n}(2 - p^n).$$

For $p = 2$ this is a Diophantine equation that can be shown to have no valid integer solutions (i.e., except $x = -1$ and $y = p^n - 1$ or $x = p^n - 1$ and $y = -1$ corresponding to the autocorrelation). For the nonbinary case when $p > 2$, similar divisibility properties in the ring $Q[\omega]$, where Q denotes the rational number field, imply that two-valued cross-correlation is impossible except in the autocorrelation case, i.e., when $d \equiv p^i \pmod{p^n - 1}$.

The cross-correlation between any two m-sequences $\{s_t\}$ and $\{s_{dt}\}$ with symbols from GF(p) of the same period $\varepsilon = p^n - 1$ can be written as an exponential sum. After a suitable shift, we can assume without loss of generality that $s_t = Tr_n(\alpha^t)$, and we therefore obtain

$$C_d(\tau) = \sum_{t=0}^{\varepsilon-1} \omega^{s_{t+\tau} - s_{dt}} = \sum_{t=0}^{\varepsilon-1} \omega^{Tr_n(\alpha^{t+\tau} - \alpha^{dt})} = \sum_{x \in \mathrm{GF(p^n)}^*} \omega^{Tr_n(cx - x^d)}$$

where $c = \alpha^{\tau}$. Finding the values and the number of occurrences of each value in the cross-correlation function $C_d(\tau)$ for τ in $\{0, 1, \ldots, p^n - 2\}$ is equivalent to determine the distribution of this exponential sum for any $c \neq 0$.

Since a two-valued cross-correlation is only possible when $d \equiv p^i \pmod{p^n - 1}$, it was natural that the early research on the cross-correlation had a strong focus on finding decimations leading to three-valued cross-correlation. The following sections will survey known cases where the cross-correlation takes on three or four values.

The mathematical techniques used to prove these results are rather different for different decimations and give interesting connections between the cross-correlation, exponential sums, and the solutions of special equations over finite fields.

Note that when we in the following find a decimation d with a correlation distribution, then, due to (1) in Lemma 1, the correlation distribution is the same for the decimations $dp^i \pmod{p^n - 1}$ for any i and for the inverse decimation by $d^{-1} \pmod{p^n - 1}$.

3 Three-Valued Cross-Correlation

3.1 Binary Sequences

There are more decimations leading to three-valued cross-correlation when $p = 2$ than in the case $p > 2$. First we consider three-valued cross-correlation in the case of binary sequences.

The pioneering result on three-valued cross-correlation was due to Gold [9] in 1968. Gold considered binary sequences and showed that $d = 2^k + 1$ for n odd and $\gcd(n, k) = 1$ gave a three-valued cross-correlation. Note that the condition n odd was later relaxed to $n/\gcd(n, k)$ odd which still implies that $\gcd(d, 2^n - 1) = 1$.

In 1968 Golomb [11] was the first to conjecture that $d = 2^{2k} - 2^k + 1$ leads to a three-valued cross-correlation when $n/\gcd(n, k)$ is odd, and he mentioned in this paper that this result was first proved by Welch, who never published his proof. Later in 1971 Kasami [19] published a proof in his famous paper on the weight distribution of several subcodes of the second-order Reed–Muller code.

Theorem 3 *Let $e = \gcd(n, k)$ and let n/e be odd. Let $d = 2^k + 1$ or $d = 2^{2k} - 2^k + 1$. Then $C_d(\tau)$ has the following distribution:*

$$\begin{array}{llll} -1 + 2^{\frac{n+e}{2}} & \textit{occurs} & 2^{n-e-1} + 2^{\frac{n-e-2}{2}} & \textit{times.} \\ -1 & \textit{occurs} & 2^n - 2^{n-e} - 1 & \textit{times.} \\ -1 - 2^{\frac{n+e}{2}} & \textit{occurs} & 2^{n-e-1} - 2^{\frac{n-e-2}{2}} & \textit{times.} \end{array}$$

The proof of these decimations use, a simple squaring technique combined with arguments to determine the number of solutions of some linearized polynomial. In the case $d = 2^k + 1$, one can compute rather directly, using simple properties of the trace function, that

$$(C_d(\tau) + 1)^2 = 2^n(1 + (-1)^{Tr_n(c+1)}).$$

It follows that $C_d(\tau)$ can only take the values $-1, -1 \pm 2^{\frac{n+1}{2}}$, and the distribution can be determined from (3) and (4) in Lemma 1.

In the case $d = 2^{2k} - 2^k + 1$ when n is odd and $\gcd(n, k) = 1$, a similar squaring argument gives

$$(C_d(\tau) + 1)^2 = 2^n N$$

where N is either 0 or equal to the number of zeros in GF(2^{n}) of the linearized polynomial

$$L(z) = z^{2^{6k}} + c^{2^{3k}} z^{2^{4k}} + c^{2^{2k}} z^{2^{2k}} + z.$$

In this case a more detailed argument shows that there is only 1 or 2 solutions in GF(2^{n}) of $L(z) = 0$ and that $C_d(\tau)$ can only take on the values $-1, -1 \pm 2^{\frac{n+1}{2}}$.

The generalization to the general case when $\gcd(n, k) = e > 1$ and n/e is odd is rather straightforward but more cumbersome. An elegant method for counting the solutions of the equation above is given by Bracken [1].

The following theorem provides a list of all decimations known to give three-valued cross-correlation in the binary case.

Theorem 4 *The cross-correlation $C_d(\tau)$ is three-valued and the correlation distribution is known for the following values of d:*

1. $d = 2^k + 1$, *where* $n/\gcd(n, k)$ *is odd.*
 2 $d = 2^{2k} - 2^k + 1$, *where* $n/\gcd(n, k)$ *is odd.*
3. $d = 2^{\frac{n}{2}} + 2^{\frac{n+2}{4}} + 1$, *where* $n \equiv 2 \pmod 4$.
4. $d = 2^{\frac{n+2}{2}} + 3$, *where* $n \equiv 2 \pmod 4$.
5. $d = 2^{\frac{n-1}{2}} + 3$, *where n is odd.*
6.

$$d = \begin{cases} 2^{\frac{n-1}{2}} + 2^{\frac{n-1}{4}} - 1, & \textit{when } n \equiv 1 \pmod 4 \\ 2^{\frac{n-1}{2}} + 2^{\frac{3n-1}{4}} - 1, & \textit{when } n \equiv 3 \pmod 4. \end{cases}$$

Comments Case (1) is the celebrated result proved by Gold [9]. Case (2) is the result first proved by Welch (see Golomb [11] and Kasami [19]). Cases (3) and (4) were proved by Cusick and Dobbertin [4] in 1996. Case (5) was a long-standing conjecture by Welch (see Golomb [11]) that was proved 30 years later by Canteaut et al. [2]. Case (6) is a consequence of the results by Dobbertin [6] and Hollmann and Xiang [16]. Cases (3), (4), and (6) were all conjectured in 1972 by Niho [23].

Since the conjectures (3), (4), and (6) by Niho [23] in 1972, which all have been proved, no new decimations of m-sequences have been found to give a three-valued cross-correlation, and it is widely believed that the list of decimations of binary m-sequences in the theorem above is complete.

Open Problem 1 Show that Theorem 4 contains all decimations with three-valued correlation between binary m-sequences.

This appears to be a very hard open problem. All decimations known to have only three values have their three values of the form $-1, -1 \pm 2^r$ for some r. Even to show that any three-valued decimation must have three such values is not known.

3.2 Nonbinary Sequences

For nonbinary sequences there are three-valued decimations that are analogous to the Gold as well as to the Kasami and Welch decimations. These are given in the following result due to Trachtenberg [25], for n odd, in his Ph.D. thesis from 1970. The result is generalized by Helleseth [12] (or actually in his master thesis from 1971) to the case when $n/\gcd(n,k)$ is odd. The generalization is rather straightforward using the properties of the subfield $\mathrm{GF(p^k)}$ of $\mathrm{GF(p^n)}$.

Theorem 5 *Let p be an odd prime. Then the following decimations have three-valued cross-correlation.*

1. $d = \frac{p^{2k}+1}{2}$ *where* $n/\gcd(n,k)$ *is odd.*
2. $d = p^{2k} - p^k + 1$ *where* $n/\gcd(n,k)$ *is odd.*

These are the only decimations for $p > 3$ that are known to give three-valued cross-correlation.

Open Problem 2 Show that Theorem 5 contains all decimations with three-valued cross-correlation between p-ary m-sequences when $p > 3$ is an odd prime.

For the ternary case there is an additional decimation with three-valued cross-correlation given in the following result by Dobbertin et al. [8].

Theorem 6 *Let $p = 3$ and $d = 2 \cdot 3^{\frac{n-1}{2}} + 1$ where n is an odd positive integer. Then the cross-correlation $C_d(\tau)$ is three-valued and has the following distribution:*

$$\begin{array}{lll} -1 + 3^{\frac{n+1}{2}} & \textit{occurs } \frac{1}{2}(3^{n-1} + 3^{\frac{n-1}{2}}) & \textit{times.} \\ -1 & \textit{occurs } 3^n - 3^{n-1} - 1 & \textit{times.} \\ -1 - 3^{\frac{n+1}{2}} & \textit{occurs } \frac{1}{2}(3^{n-1} - 3^{\frac{n-1}{2}}) & \textit{times.} \end{array}$$

There are numerical observations of the cross-correlation between ternary sequences that give decimations with three-valued cross-correlation and that have not yet been proved. If the following open problem, conjectured by Dobbertin et al. [8], is settled, this would explain all the known decimations of ternary

m-sequences with three-valued cross-correlation. Actually, a solution of the problem would complete the explanation of all currently known three-valued cross-correlation decimations for any p.

Open Problem 3 Let $p = 3$ and $d = 2 \cdot 3^r + 1$ where n is odd and

$$r = \begin{cases} \frac{n-1}{4} & \text{if } n \equiv 1 \pmod 4, \\ \frac{n-1}{4} & \text{if } n \equiv 3 \pmod 4. \end{cases}$$

Show that $C_d(\tau)$ has three-valued cross-correlation.

If $n/\gcd(n,k)$ is odd, Theorems 4 and 5 imply the existence of decimations with three-valued cross-correlation.

In the remaining cases when $n = 2^i$ for some positive integer $i \geq 2$, there are no known decimations having three-valued cross-correlation. It was conjectured by Helleseth [12] that in these cases, any decimation gives at least four cross-correlation values. This conjecture has recently been proved in the binary case by Katz [20]. The general case to settle the conjecture for all other values of p is still open.

Open Problem 4 Show that $C_d(\tau)$ is at least four-valued when $n = 2^i$ for all values of the prime p.

4 Four-Valued Cross-Correlation

One of the main contributions leading to new decimations with four-valued cross-correlation is due to Niho [23]. For his method to be applicable, then $n = 2k$ has to be even and d must be of the special form $d = s(2^k - 1) + 1$.

The main idea is to reduce the problem to compute the number of solutions of some special equations that depend on s.

The next theorem provides a list, in historical order, of all the decimations that have been proved to give four-valued cross-correlation. An important observation is that all the results in (1)–(4) are covered by the last case (5).

Theorem 7 *Let $v_2(i)$ be the highest power of 2 dividing the integer i. The cross-correlation $C_d(\tau)$ is four-valued and the correlation distribution is known for the following values of d:*

1. $d = 2^{\frac{n}{2}+1} - 1$, *where* $n \equiv 0 \pmod 4$.
2. $d = (2^{\frac{n}{2}} + 1)(2^{\frac{n}{4}} - 1) + 2$, *where* $n \equiv 0 \pmod 4$.
3. $d = \frac{2^{(n/2+1)r}-1}{2^r-1}$ $(0 < r < n/2, \gcd(n,r) = 1)$ *for* $n \equiv 0 \pmod 4$.
4. $d = \frac{2^{2k}+2^{s+1}-2^{k+1}-1}{2^s-1}$, *where* $n = 2k$ *and* $2s|k$.
5. $d = (2^k - 1)s + 1$, $s \equiv 2^r(2^r \pm 1)^{-1} \pmod{2^k + 1}$, *where* $v_2(r) < v_2(k)$.

Comments The first two cases in Theorem 7 are due to Niho [23] in his Ph.D. thesis. Case (3) is due to Dobbertin [5]. Case (4) was proved by Helleseth and Rosendahl [15]. The final case (5) is proved by Dobbertin et al. [7] and contains all the four previous cases.

Sketch of proof Since the Niho decimations have played a significant role in the cross-correlation of m-sequences, we will provide a short outline of the proof.

The main idea behind the proof of Theorem 7 is very simple and uses that any nonzero element $x \in \mathrm{GF}(2^{\mathrm{n}})$ can be written uniquely as $x = yz$ where $y \in \mathrm{GF}(2^{\mathrm{k}})$ and $z \in U$ where

$$U = \{z \in \mathrm{GF}(2^{\mathrm{n}}) \mid z^{2^k+1} = 1\}.$$

In particular, since $d = s(2^k - 1) + 1$, it follows that $d \equiv 1 \pmod{2^k - 1}$ and therefore $d \equiv -2s + 1 \pmod{2^k + 1}$. Hence, $y^d = y$ and $z^d = z^{-2s+1}$ and the cross-correlation can be written as

$$\begin{aligned}
C_d(\tau) &= \sum_{x \in \mathrm{GF}(2^{\mathrm{n}})^*} (-1)^{Tr_m(cx+x^d)} \\
&= \sum_{y \in \mathrm{GF}(2^{\mathrm{n}})^*, z \in U} (-1)^{Tr_n(cyz+yz^{-2s+1})} \\
&= \sum_{y \in \mathrm{GF}(2^{\mathrm{n}})^*, z \in U} (-1)^{Tr_k(yh(z))} \\
&= (2^k - 1)N + (2^k + 1 - N)(-1) \\
&= -1 + (N - 1)2^k.
\end{aligned}$$

Here N is the number of solutions $z \in U$ of the equation $h(z) = 0$ where

$$h(z) = cz + z^{-2s+1} + c^{2^k} z^{-1} + z^{2s-1}$$

which is equivalent to N being the number of solutions $z \in U$ to

$$p(z) = z^{2s-1} + c^{1/2} z^s + c^{2^{k-1}} z^{s-1} + 1 = 0.$$

Case (1) is one of the two decimations in Niho's thesis shown to be four-valued. In this case $s = 2$, i.e., $d = 2(2^k - 1) + 1$ and

$$p(z) = z^3 + c^{1/2} z^2 + c^{2^{k-1}} z + 1 = 0$$

which has at most three solutions for z. Hence, $N = 0, 1, 2, 3$ are the only possibilities leading to at most a four-valued cross-correlation with values in the set

$$\{-1 - 2^k, -1, -1 + 2^k, -1 + 2^{k+1}\}.$$

The correlation distribution follows from Lemma 1 using (3)–(5) and finding a_3.

In the other cases (2)–(5) in Theorem 7, we have

$$p(z) = z^{2^r+1} + az^{2^r} + bz + 1 = 0$$

which is known to have $0, 1, 2$ or $2^{\gcd(r,n)} + 1$ solutions in $\mathrm{GF}(2^{\mathrm{n}})$. A more detailed analysis shows that the number of solutions in U has these four possibilities and the four-valued cross-correlation distribution can be found as above.

There are numerical results that give decimations with four values that are not explained by this list. However, one believes that case (5) in Theorem 7 contains all four-valued cases when d is of the Niho form $d = s(2^k - 1) + 1$. The following conjecture was stated in Dobbertin et al. [7].

Open Problem 5 Any binary decimation of Niho type $d = s(2^k - 1) + 1, n = 2k$ with four-valued cross-correlation is of the form $d = (2^k - 1)s + 1, s \equiv 2^r(2^r \pm 1)^{-1} \pmod{2^k + 1}$, where $v_2(r) < v_2(k)$.

In the nonbinary case there are a few families known with four-valued cross-correlation. The following decimation in Helleseth [12] is the only known four-valued decimation that works for any prime p.

Theorem 8 *Let p be an odd prime and $d = 2 \cdot p^{\frac{n}{2}+1} - 1$, where $n \equiv 0 \pmod 4$. Then the cross-correlation $C_d(\tau)$ is four-valued and the distribution is known.*

Recently new ternary decimations with four-valued cross-correlation have been found by Zhang et al. [26].

Theorem 9 *Let $p = 3, n = 3k$, and $\gcd(k, 3) = 1$. If $d = 3^k + 1$ or $d = 3^{2k} + 2$. Then if r is odd, the cross-correlation $C_d(\tau)$ is four-valued (and six-valued for r even) and the distribution is known. (The distribution is conjectured to be the same if $\gcd(k, 3) = 3$).*

5 The −1 Conjecture

For binary sequences the cross-correlation values are obviously always integers. For $p > 2$, this may not always be the case even though the values of $C_d(\tau)$ are always real numbers. This follows from the definition of the cross-correlation function and the fact that the second half of an m-sequence is the negative of the first half. It was shown in Helleseth [12] that $C_d(\tau)$ is an integer for all τ if and only if $d \equiv 1 \pmod{p - 1}$.

Numerical results reveal that for $p = 2$, the cross-correlation always has -1 as one of its values. For $p > 2$ this happens for all decimations d where $d \equiv 1 \pmod{p-1}$. This was conjectured by Helleseth [12] and this is still an open problem.

Open Problem 6 Show that if $d \equiv 1 \pmod{p-1}$, then -1 always occurs as a value in $C_d(\tau)$.

It is trivial to reformulate the conjecture as a result of the number of common solutions of a special equation system.

Lemma 2 *Let $q = p^n$ and α be a primitive element in* GF(q). *Let N be the number of solutions $x_i \in$* GF(q) *of the equation system:*

$$\begin{aligned} x_0 + \alpha x_1 + \alpha^2 x_2 + \cdots + \alpha^{q-2} x_{q-2} &= 0 \\ x_0^d + x_1^d + x_2^d + \cdots + x_{q-2}^d &= 0. \end{aligned}$$

The -1 conjecture holds if and only if $N = q^{q-3}$ for all d where $\gcd(d, p^n - 1) = 1$ *and $d \equiv 1 \pmod{p-1}$.*

Proof Let N denote the number of common solutions of the two equations above. Then N can be expressed by the following exponential sum.

$$\begin{aligned} q^2 N &= \sum_{x_0, x_1, \ldots, x_{q-2} \in \mathrm{GF(q)}} \sum_{z_1, z_2 \in \mathrm{GF(q)}} \omega^{Tr_n(z_1(x_0 + \alpha x_1 + \cdots + \alpha^{q-2} x_{q-2}) + z_2(x_0^d + x_1^d + \cdots + x_{q-2}^d))} \\ &= \sum_{z_1, z_2 \in \mathrm{GF(q)}} \prod_{i=0}^{q-2} \sum_{x \in \mathrm{GF(q)}} \omega^{Tr_n(z_1 \alpha^i x + z_2 x^d)} \\ &= q^{q-1} + (q-1) \prod_{c \in \mathrm{GF(q)}} \sum_{x \in \mathrm{GF(q)}} \omega^{Tr_n(cx + x^d)} \\ &= q^{q-1} + \prod_{\tau=0}^{q-2} (C_d(\tau) + 1) \end{aligned}$$

since the contribution from $z_1 = z_2 = 0$ is q^{q-2}, and $z_1 = 0$ or $z_2 = 0$ contributes 0 if not both are zero. Furthermore, $z_1 / z_2^{d^{-1}}$ runs through all nonzero elements in GF(q) $q - 1$ times when z_1 and z_2 run through all nonzero elements in the field. Hence, $C_d(\tau) = -1$ for some τ if and only of $N = q^{q-3}$.

Another old problem on the cross-correlation between m-sequences that is more than 30-year-old is the following conjecture due to Sarwate and Pursley [24].

Open Problem 7 Let n be even. Show that $|C_d(\tau) + 1| \geq 2^{\frac{n}{2}+1}$.

For related surveys on m-sequences the reader is referred to [13, 14]. Other interesting results and open problems on this topic can be found in [18, 22].

6 Relations to APN and AB Functions

The cross-correlation of m-sequences has some interesting relations to almost perfect nonlinear mappings (APN) and almost bent functions (AB).

An *almost perfect nonlinear* function f is a mapping $f : \mathrm{GF}(2^{\mathrm{n}}) \mapsto \mathrm{GF}(2^{\mathrm{n}})$ such that

$$f(x+a) + f(x) = b$$

has at most two solutions for any $a \neq 0$, $b \in \mathrm{GF}(2^{\mathrm{n}})$. The function is said to be Δ-uniform if the maximum number of solutions is Δ, such that an APN function is the same as being 2-uniform.

The *Walsh transform* of f is defined by

$$\lambda_f(a,b) = \sum_{x \in \mathrm{GF}(2^{\mathrm{n}})} (-1)^{Tr(af(x)+bx)},$$

where $a, b \in \mathrm{GF}(2^{\mathrm{n}})$.

A function f is *almost bent* (AB) if

$$\{\lambda_f(a,b) : a,b \in \mathrm{GF}(2^{\mathrm{n}})\} = \{0, \pm 2^{(n+1)/2}\}.$$

It has been shown by Chaubaud and Vaudenay [3] that AB implies APN. APN functions and AB functions are of significant importance in the design of S-boxes in block ciphers.

Monomial AB functions where $f(x) = x^d$ can be obtained from Gold sequences and several of the decimations with three-valued cross-correlation.

Theorem 10 *The known monomial AB functions $f(x) = x^d$ are*

1. *Gold:* $d = 2^k + 1$, *where* $\gcd(n,k) = 1$.
2. *Kasami:* $d = 2^{2k} - 2^k + 1$, *where* $\gcd(n,k) = 1$.
3. *Welch:* $d = 2^{\frac{n-1}{2}} + 3$, *where n is odd.*
4. *Niho:*

$$d = \begin{cases} 2^{\frac{n-1}{2}} + 2^{\frac{n-1}{4}} - 1, \ if\ n \equiv 1 \pmod 4 \\ 2^{\frac{n-1}{2}} + 2^{\frac{3n-1}{4}} - 1, if\ n \equiv 3 \pmod 4. \end{cases}$$

Note that each of these cases corresponds to decimations with three-valued cross-correlation where the values are restricted to the set $\{0, \pm 2^{(n+1)/2}\}$. Thus, each corresponding monomial function $f(x) = x^d$ is AB. Dobbertin [6] conjectured that these are the only monomial AB functions.

Open Problem 8 Show that Theorem 10 contains all monomial $f(x) = x^d$ AB functions.

Since a monomial AB function is an APN function, the monomial functions $f(x) = x^d$ with d in Theorem 10 are also APN functions. In addition there are two more decimations leading to APN functions and which are not AB. The known monomial APN functions are given in the following theorem.

Theorem 11 *The known monomial APN functions* $f(x) = x^d$ *are*

1. *Gold:* $d = 2^k + 1$, *where* $\gcd(n,k) = 1$.
2. *Kasami:* $d = 2^{2k} - 2^k + 1$, *where* $\gcd(n,k) = 1$.
3. *Welch:* $d = 2^{\frac{n-1}{2}} + 3$, *where n is odd.*
4. *Niho:*

$$d = \begin{cases} 2^{\frac{n-1}{2}} + 2^{\frac{n-1}{4}} - 1, & if\ n \equiv 1 \pmod 4 \\ 2^{\frac{n-1}{2}} + 2^{\frac{3n-1}{4}} - 1, & if\ n \equiv 3 \pmod 4). \end{cases}$$

5. *Inverse:* $d = 2^n - 2 \equiv -1 \pmod{2^n - 1}$, *where n is odd.*
6. *Dobbertin:* $d = 2^{4k} + 2^{3k} + 2^{2k} + 2^k - 1$, *where* $n = 5k$.

Dobbertin [6] conjectured that these are the only monomial APN functions.

Open Problem 9 Show that Theorem 11 contains all monomial $f(x) = x^d$ APN functions.

It is easy to show that the cross-correlation values between two m-sequence obey $C_d(\tau) \equiv -1 \pmod 4$. The cross-correlation between $\{s(t)\}$ and its reverse sequence $\{s(-t)\}$ corresponds to the famous Kloosterman sum defined by

$$C_{-1}(\tau) = \sum_{x \in \mathrm{GF(p^n)^*}} \omega^{Tr(ax+x^{-1})}.$$

A well-known bound for the Kloosterman sum is

$$|C_{-1}(\tau)) + 1| \leq 2p^{n/2}.$$

For $p = 2$ it was shown by Lachuad and Wolfmann [21] that $C_{-1}(\tau)$ takes on all possible values $\equiv -1 \pmod 4$ that obey this bound.

The S-box used in the Advanced Encryption Standard (AES) is a permutation based on $f(x) = x^{-1}$ for $n = 8$. The correlation between x^{-1} and all affine functions take on the same values as $|C_{-1}(\tau)|$ when $\tau = 0, 1, \ldots, p^n - 2$. The S-box is 4-uniform (not APN) which is the best known uniformity for $n = 8$. The S-box is not AB but the correlation (and nonlinearity) is the best known for $n = 8$.

Conclusion

The cross-correlation of m-sequences is a challenging mathematical problem that has many important applications in communication systems. This chapter presents an updated overview of this problem and presented some of the remaining open problems that still exist in this area. Finally, a few connections have been given to AB and APN functions that are important in the design and analysis of S-boxes in block ciphers.

Acknowledgements This research was supported by the Norwegian Research Council.

References

1. C. Bracken, Designs, Codes, *Spin Models and the Walsh Transform*, Ph.D. thesis, Department of Mathematics, National University Ireland (NUI), Maynooth, 2004
   In this Ph.D. thesis one can find a nice proof of the number of solutions of a linearized polynomial playing an important role in the proof of the 3-valued crosscorrelation with the Kasami–Welch exponent $d = 2^{2k} - 2^k + 1$.
2. A. Canteaut, P. Charpin, H. Dobbertin, Binary m-sequences with three-valued crosscorrelation: a proof of Welch's conjecture. IEEE Trans. Inf. Theory **46**(1), 4–8 (2000)
   The more than 30 year old conjecture by Welch on a decimation with 3-valued crosscorrelation between two m-sequences is proved in this paper.
3. F. Chabaud, S. Vaudenay, Links between differential and linear cryptanalysis, in *Advances in Cryptology-EUROCRYPT'94* (Springer, New York, 1995), pp. 356–365
   The paper gives important relations between differential and linear analysis and shows in particular that AB functions are APN functions.
4. T.W. Cusick, H. Dobbertin, Some new three-valued crosscorrelation functions for binary m-sequences. IEEE Trans. Inf. Theory **42**(4), 1238–1240 (1996)
   The authors prove two conjectures due to Niho on two decimation that (for n even) give 3-valued crosscorrelation.
5. H. Dobbertin, One-to-one highly nonlinear power functions on GF(2^n). Appl. Algebra Eng. Commun. Comput. **9**(2), 139–152 (1998)
   The author finds a new decimation with 4-valued crosscorrelation, the first new one since Niho's Ph.D. thesis from 1972.
6. H. Dobbertin, Almost perfect nonlinear power functions on GF(2^n): the Niho case. Inf. Comput. **151**(1–2), 57–72 (1999)
   The author shows that two decimations conjectured by Niho to have 3-valued crosscorrelation for odd m give almost perfect nonlinear functions. This was an important step in order to later complete the proof of these conjectures in [16].
7. H. Dobbertin, P. Felke, T. Helleseth, P. Rosendahl, Niho type cross-correlation functions via Dickson polynomials and Kloosterman sums. IEEE Trans. Inf. Theory **52**(2), 613–627 (2006)
   Dickson polynomials were used for the first time to find the crosscorrelation between m-sequences. The paper also settled the correlation distribution of many new decimations with 4-valued crosscorrelation.
8. H. Dobbertin, T. Helleseth, P. Vijay Kumar, H. Martinsen, Ternary m-sequences with three-valued crosscorrelation function: two new decimations of Welch and Niho type. IEEE Trans. Inf. Theory **47**(4), 1473–1481 (2001)

The importance of this paper is that is found the first new nonbinary decimations with three values since the constructions 30 years earlier by Trachtenberg.

9. R. Gold, Maximal recursive sequences with 3-valued recursive cross-correlation functions. IEEE Trans. Inf. Theory **14**(1), 154–156 (1968)
   This pioneering paper defined the Gold decimation and proved that it had a 3-valued crosscorrelation and determined the complete correlation distribution. This was the basis for the important Gold sequences.
10. S.W. Golomb, *Shift Register Sequences* (Holden-Day, San Francisco, 1967)
   This is a classical book on linear and nonlinear recursions.
11. S.W. Golomb, Theory of transformation groups of polynomials over GF(2) with applications to linear shift register sequences. Inf. Sci. **1**(1), 87–109 (1968)
   The author states (without proof) that the crosscorrelation between binary m-sequences takes on at least three values. The Welch conjecture that two special decimations have 3-valued crosscorrelation was published here for the first time.
12. T. Helleseth, Some results about the cross-correlation function between two maximal linear sequences. Discrete Math. **16**(3), 209–232 (1976)
   This paper contains many basic results on the crosscorrelations of m-sequences. The first nonbinary decimation is found giving a four-valued crosscorrelation between two m-sequences. The distributions of several decimations are completely settled. The -1 conjecture is stated in this paper.
13. T. Helleseth, Crosscorrelation of m-sequences, exponential sums and Dickson polynomials. IEICE Trans. Fundamentals **E93A**(11), 2212–2219 (2010)
   Presents a survey on the crosscorrelation between binary m-sequences having at most 5-valued crosscorrelation with a focus on the many connections between exponential sums and Dickson polynomials.
14. T. Helleseth, P.V. Kumar, Sequences with low correlation, in *Handbook in Coding Theory*, eds. by V.S. Pless, W.C. Huffman, ch. 21 (Elsevier Science B.V., Amsterdam, 1998), pp.1765–1853
   This is a survey of sequences with low correlation that contains constructions and analysis of many important sequence families and some of their relations to coding theory.
15. T. Helleseth, P. Rosendahl, New pairs of m-sequences with 4-level cross-correlation. Finite Fields Appl. **11**(4), 674–683 (2005)
   This paper introduced new decimations with 4-valued cross correlation.
16. H.D.L. Hollmann, Q. Xiang, A proof of the Welch and Niho conjectures on cross-correlations of binary m-sequences. Finite Fields Appl. **7**(2), 253–286 (2001)
   This paper completed the proof of two decimations, for odd m, that were conjectured by Niho to lead to 3-valued crosscorrelation.
17. A. Johansen, T. Helleseth, A family of m-sequences with five-valued cross correlation. IEEE Trans. Inf. Theory **55**(2), 880–887 (2009)
   The distribution of the crosscorrelation of pairs of m-sequences with decimations giving five-valued crosscorrelation was found using techniques involving Dickson polynomials.
18. A. Johansen, T. Helleseth, A. Kholosha, Further results on m-sequences with five-valued cross correlation. IEEE Trans. Inf. Theory **55**(12), 5792–5802 (2009)
   This paper extends the results in [17] to other decimations with five-valued crosscorrelation. Some results depend on open conjectures on some exponential sums.
19. T. Kasami, The weight enumerators for several classes of subcodes of the 2nd order binary Reed–Muller codes. Inf. Control **18**(4), 369–394 (1971)
   The author determined the weight enumerator of some subcodes of the 2nd order Reed–Muller. A consequence of these results is a proof of the Kasami–Welch decimation leading to 3-valued crosscorrelation. This decimation was also proved by Welch (unpublished).
20. D. Katz, Weil sums of binomials, three-level cross-correlation and a conjecture by Helleseth. J. Combin. Theory A **119**(8), 1644–1659 (2012)
   The paper gives a solution of the conjecture of Helleseth that for $n = 2^i$ and $p = 2$ the crosscorrelation takes on at least 4 values.

21. G. Lachaud, J. Wolfmann, The weights of the orthogonals of the extended quadratic binary Goppa codes. IEEE Trans. Inf. Theory **36**(3), 686–692 (1990)
The paper shows that the Kloosterman sums takes on all possible values $\equiv -1 \pmod 4$ within its bound.
22. J. Lahtonen, G. McGuire, H.N. Ward, Gold and Kasami–Welch functions, quadratic forms, and bent functions. Adv. Math. Commun. **1**(2), 243–250 (2007)
Provides a local result on $C_d(0)$ for the Kasami–Welch decimation.
23. Y. Niho, *Multi-valued Cross-Correlation Functions Between Two Maximal Linear Recursive Sequences*, Ph.D. thesis, University of Southern California, Los Angeles, 1972
This thesis gave the complete crosscorrelation distribution of several decimations with 4-valued cross correlation. Furthermore, many conjectures on the cross correlation distribution of sequences with few values in the crosscorrelation were given. This Ph.D. thesis had a significant influence on later research on the crosscorrelation.
24. D. Sarwate, M. Pursley, Crosscorrelation properties of pseudorandom and related sequences. Proc. IEEE, **68**(5), 593–619 (1980)
This is a classical and excellent survey of the crosscorrelation between m-sequences.
25. H.M. Trachtenberg, *On the Cross-Correlation Functions of Maximal Linear Recurring Sequences*, Ph.D. thesis, University of Southern California, Los Angeles, 1970
The main result is the two families of decimation giving three-valued crosscorrelation. These are the only decimations that work for all nonbinary m-sequences.
26. T. Zhang, S. Li, T. Feng, G. Ge, Some new results on the cross correlation of m-sequences. arXiv:1309.7734 [cs.IT]
This recent paper gives new ternary decimations with four-valued crosscorrelation.