Advances in Mathematics of Communications
Volume 00, No. 0, 2007, 1–

Web site: http://www.aimSciences.org

BOUNDS ON THE GROWTH RATE OF THE PEAK SIDELOBE LEVEL OF BINARY SEQUENCES

Denis Dmitriev

The D. E. Shaw Group
39th Floor, Tower 45, 120 West Forty-Fifth Street
New York, NY 10036, USA

Jonathan Jedwab

Department of Mathematics, Simon Fraser University
8888 University Drive, Burnaby BC, Canada V5A 1S6

(Communicated by Marcus Greferath)

Abstract. The peak sidelobe level (PSL) of a binary sequence is the largest absolute value of all its nontrivial aperiodic autocorrelations. A classical problem of digital sequence design is to determine how slowly the PSL of a length n binary sequence can grow, as n becomes large. Moon and Moser showed in 1968 that the growth rate of the PSL of almost all length n binary sequences lies between order $\sqrt{n \log n}$ and $\sqrt{n}$, but since then no theoretical improvement to these bounds has been found.

We present the first numerical evidence on the tightness of these bounds, showing that the PSL of almost all binary sequences of length n appears to grow exactly like order $\sqrt{n \log n}$, and that the PSL of almost all m-sequences of length n appears to grow exactly like order $\sqrt{n}$. In the case of m-sequences, a key algorithmic insight reveals behaviour that was previously well beyond the range of computation.

1. Introduction

One of the oldest problems of digital sequence design, dating from the 1950s, is to determine those binary sequences whose aperiodic autocorrelations are collectively small (see [12] and [14], for example). A *sequence* A of length n is an n-tuple $(a_0, a_1, \ldots, a_{n-1})$, and the sequence is *binary* if each a_i takes the value -1 or 1. The *aperiodic autocorrelation* of the binary sequence A at shift u is given by

$$C_A(u) := \sum_{i=0}^{n-u-1} a_i a_{i+u} \quad \text{for } u = 0, 1, \ldots, n-1.$$

The measure of smallness of the aperiodic autocorrelations considered in this paper is the *peak sidelobe level* (PSL), given by:

$$M(A) := \max_{0<u<n} |C_A(u)| \quad \text{for } n > 1.$$

2000 *Mathematics Subject Classification:* Primary: 94A55, 68P30; Secondary: 05B10.

Key words and phrases: Aperiodic autocorrelation, peak sidelobe level, binary sequence, maximal length shift register sequence, m-sequence, asymptotic.

The first author was supported by NSERC of Canada via a 2006 Undergraduate Student Research Award.

The second author is supported by NSERC of Canada.

An alternative measure of smallness is the *merit factor*, given by

$$F(A) := \frac{n^2}{2\sum_{u=1}^{n-1}[C_A(u)]^2} \quad \text{for } n > 1$$

(see [4] for a survey).

We define M_n to be the optimal value of the PSL over the set $\mathcal{A}_n$ of all binary sequences of length n:

$$M_n := \min_{A \in \mathcal{A}_n} M(A).$$

The numerical value of M_n is known for $n \leq 70$ by exhaustive computer search (see [5] for a summary of results). Our principal interest, however, is in understanding the behaviour of M_n as $n \to \infty$. A classical result shows that the growth rate of the PSL of almost all binary sequences lies between order $\sqrt{n \log n}$ and order $\sqrt{n}$:

Theorem 1 (Moon and Moser 1968 [10])**.**

(*i*) *For any fixed* $\epsilon > 0$*, the proportion of sequences* $A \in \mathcal{A}_n$ *such that* $M(A) \leq (2+\epsilon)\sqrt{n \log n}$ *approaches* 1 *as* $n \to \infty$*.*

(*ii*) *If* $K(n)$ *is any function of* n *such that* $K(n) = o(\sqrt{n})$*, then the proportion of sequences* $A \in \mathcal{A}_n$ *for which* $M(A) > K(n)$ *approaches* 1 *as* $n \to \infty$*.*

(We use the notation o, O, Ω and Θ to compare the growth rates of functions $f(n)$ and $g(n)$ from $\mathbb{N}$ to $\mathbb{R}^+$ in the following standard way: f is $o(g)$ means that $f(n)/g(n) \to 0$ as $n \to \infty$; f is $O(g)$ means that there is a constant c, independent of n, for which $f(n) \leq cg(n)$ for all sufficiently large n; f is $\Omega(g)$ means that g is $O(f)$; and f is $\Theta(g)$ means that f is $O(g)$ and $\Omega(g)$.)

The functions $\sqrt{n \log n}$ and $\sqrt{n}$ in Theorem 1 are upper and lower bounds on the order of the growth rate of the PSL of almost all binary sequences. This paper is concerned with the tightness of these bounds. It is straightforward to show that the upper bound $O(\sqrt{n \log n})$ does not apply to all binary sequences: for example, the PSL of the all-ones sequence of length n is $n-1$. However it is possible that the upper bound, applying to almost all binary sequences, can be improved. We therefore ask:

1. Does the PSL of almost all binary sequences grow like $o(\sqrt{n \log n})$?

Turning to the lower bound $\Omega(\sqrt{n})$, it is an open question as to whether this lower bound applies to all binary sequences. We know that if the PSL of a family of binary sequences were to grow more slowly than order $\sqrt{n}$, then the asymptotic merit factor of this family would be unbounded:

Proposition 2 (Jedwab and Yoshida 2006 [5])**.** *Let* $\mathcal{B}$ *be a family of binary sequences and let each* $A_n \in \mathcal{B}$ *have length* n*. If* $\liminf_{n\to\infty}(M(A_n)/\sqrt{n}) = 0$ *then* $\limsup_{n\to\infty} F(A_n) = \infty$*.*

However the existence of a family of binary sequences with unbounded asymptotic merit factor is considered very unlikely by most (although not all) authors [4]. Assuming there is no such family, the most testing question for the lower bound becomes:

2. Is there a family of binary sequences whose PSL grows like $\Theta(\sqrt{n})$?

There are currently no known methods to answer these two questions with certainty. More remarkably, nearly forty years after Theorem 1 appeared, there is still no proof that the PSL of any specific family of binary sequences grows like

$O(\sqrt{n \log n})$, even though this is true of almost all binary sequences! Indeed, the strongest result to date is that the PSL of m-sequences grows like $O(\sqrt{n} \cdot \log n)$ (see Theorem 4).

In this paper we present the first experimental evidence that the answers to the above two questions are "no" and "yes" respectively. Specifically, we show numerically that the PSL of almost all binary sequences appears to grow like $\Theta(\sqrt{n \log n})$, and that the PSL of almost all m-sequences appears to grow like $\Theta(\sqrt{n})$. We also show that the PSL of all m-sequences appears to grow like $O(\sqrt{n} \cdot \log \log n)$.

This rest of this paper is organised as follows. Section 2 examines the growth rate of the PSL of randomly-selected binary sequences numerically. Section 3 reviews the definition and properties of m-sequences. Section 4 gives an efficient calculation method for the maximum PSL over all cyclic shifts of a given m-sequence. Section 5 applies this method to study the growth rate of the PSL of m-sequences up to length $2^{25} - 1$ numerically. Section 6 summarises the results of the paper.

2. The growth rate of the PSL of randomly-selected binary sequences

In this section we investigate numerically the growth rate of the PSL of randomly-selected binary sequences. Rather surprisingly, such a study does not appear to have carried out previously (to our knowledge).

For each value of $m \in \{2, 2.5, 3, 3.5, \ldots, 24.5\}$, a randomly-selected subset $\mathcal{Z}_{2^m-1}$ of the binary sequences $\mathcal{A}_{2^m-1}$ of length $2^m - 1$ (rounded to the nearest integer) was chosen. (For integer values of m, these sequence lengths have the same form as those of the m-sequences studied in later sections.) The "cryptographically random number generator" `CryptGenRandom()`[1] was used to control the subset selection, in order to minimise any influence of the random number generation algorithm on the PSL properties of the resulting sequences.

For each length $n = 2^m - 1$, the PSL of each sequence $Z \in \mathcal{Z}_n$ was calculated and compared with the Moon and Moser upper bound $2\sqrt{n \log n}$ (see Theorem 1 (i)). Figure 1 shows the variation of $\text{mean}_{Z \in \mathcal{Z}_n} M(Z)/(2\sqrt{n \log n})$ with $\log n$. The error bars show one standard deviation (as estimated from the data) above and below the mean value. The number $|\mathcal{Z}_n|$ of binary sequences of length n selected, as given in Table 1, was chosen to be sufficient to make the trend of the graph clear. The graph appears to be a (broadly) increasing function (which is bounded above by 1, from Theorem 1 (i)). We conclude empirically that the mean PSL of binary sequences of length n grows like $\Omega(\sqrt{n \log n})$ and therefore, by Theorem 1 (i), that

the PSL of almost all binary sequences of length n grows like $\Theta(\sqrt{n \log n})$.

Assuming this to be true, the bounding function $\sqrt{n \log n}$ of Theorem 1 (i) cannot be improved, although a reduction in the growth constant $2 + \epsilon$ might be possible. (It is clear from Theorem 1 (i) that for any fixed $\epsilon > 0$, $M_n \leq (2 + \epsilon)\sqrt{n \log n}$ when n is sufficiently large. The constant in this latter bound was improved from 2 to $\sqrt{2}$ by Mercer [8], but his proof applies only to M_n and not to almost all binary sequences.)

[1] supplied as part of the Microsoft Strong Cryptographic Provider, and described at `http://msdn2.microsoft.com/en-us/library/aa379942.aspx`

3. m-sequences

In this section we review the definition and properties of m-sequences.

Let $f(x) = 1 + \sum_{i=1}^{m} c_i x^i$ be a primitive polynomial of degree $m > 1$ over GF(2). Let $(a_0, a_1, \dots, a_{2^m-2})$ be a 0/1 sequence of length $2^m - 1$ whose first m elements take arbitrary values (not all zeroes), and whose subsequent elements satisfy the linear recurrence relation

$$a_i := \left(\sum_{j=1}^{m} c_j a_{i-j} \right) \bmod 2 \quad \text{for } m \le i < 2^m - 1.$$

Then the binary sequence $Y = (y_0, y_1, \dots, y_{2^m-2})$ of length $2^m - 1$ defined by $y_i = (-1)^{a_i}$ for $0 \le i \le 2^m - 2$ is a *maximal length shift register sequence*, often abbreviated to *m-sequence* (and also called an *ML-sequence* or *pseudonoise sequence*). The period of Y is $2^m - 1$, and the sum $\sum_{i=0}^{2^m-2} y_i$ of all the elements of Y is -1. Write Y_f for the m-sequence generated by $f(x)$ whose first m elements equal a specified m-tuple, say the m-tuple of all -1's.

Given a sequence (a_i) of length n, regard any expression for the sequence subscript to be reduced modulo n, so that $a_{i+n} = a_i$ for all i. The kth *cyclic shift* of a length n sequence $A = (a_i)$ is the length n sequence

$$T^k(A) := (a_{i+k}).$$

All $2^m - 1$ cyclic shifts $\{T^k(Y_f) : 0 \le k < 2^m - 1\}$ of the m-sequence Y_f are m-sequences. The set $\mathcal{F}_m$ of primitive polynomials of degree m over GF(2) has order $\frac{\phi(2^m-1)}{m}$, and the set

$$\mathcal{Y}_m := \{T^k(Y_f) : f \in \mathcal{F}_m,\, 0 \le k < 2^m - 1\} \tag{1}$$

of all m-sequences of length $2^m - 1$ has order $\frac{\phi(2^m-1)}{m} \cdot (2^m - 1)$. Golomb and Gong [3], in an update to the classic reference [2], give details of these and many other properties of m-sequences, including alternative definitions using the trace function or a cyclic Singer difference set.

Since the asymptotic merit factor of all m-sequences is 3 [6], by Proposition 2 we have:

Corollary 3. *The PSL of all m-sequences of length n grows like $\Omega(\sqrt{n})$.*

In 1980 McEliece [7] established the strongest known upper bound on the growth rate of the PSL of m-sequences, namely $O(\sqrt{n} \cdot \log n)$, and the growth constant was later reduced from 1 to $2/\pi$:

Theorem 4 (Sarwate 1984 [11])**.** *Let Y be an m-sequence of length n. Then*

$$M(Y) < 1 + \frac{2}{\pi}\sqrt{n+1} \log\left(\frac{4n}{\pi}\right).$$

The method of [7] and [11] involves estimation of the maximum absolute value of an incomplete exponential sum, using results obtained in 1918 by Vinogradov and by Pólya (see Tietäväinen [13] for an overview of this method).

The only proven results for the PSL of m-sequences of length $n = 2^m - 1$ are, as above, that the growth rate is $\Omega(\sqrt{n})$ and $O(\sqrt{n} \cdot \log n)$. Jedwab and Yoshida [5] investigated widespread claims, dating from the 1960s, that the actual growth rate for some or all m-sequences is $O(\sqrt{n})$ (and therefore $\Theta(\sqrt{n})$), but concluded that there is no theoretical basis for these claims. Based on exhaustive results for $m \leq 15$ and partial results for $16 \leq m \leq 20$, they found no experimental basis either: the strongest empirical conclusion supported by these data, for the mean PSL over all m-sequences of length n, is a growth rate of $O(\sqrt{n \log n})$.

4. The maximum PSL over all cyclic shifts of an m-sequence

Rather than seeking to calculate the mean PSL of all m-sequences of a given length, in this section we consider how to calculate efficiently the maximum PSL over all cyclic shifts of a given m-sequence. This leads to our main theoretical result, Theorem 7, whose proof depends on the following two lemmas.

Lemma 5. *Let $Y = (y_i)$ be an m-sequence of length n, and let $u \in \{1, 2, \dots, n-1\}$. Then there is an integer $r = r(Y, u) \in \{1, 2, \dots, n-1\}$ such that*

$$y_i y_{i+u} = y_{i+r} \quad \textit{for all } i, \tag{2}$$

and

$$\{r(Y, u) \ : \ 1 \leq u \leq n-1\} = \{1, 2, \dots, n-1\}. \tag{3}$$

Proof. Write $(y_i) = ((-1)^{a_i})$, so that (a_i) is the 0/1 m-sequence corresponding to Y. Then (2) is equivalent to

$$a_i + a_{i+u} \equiv a_{i+r} \pmod 2 \quad \text{for all } i,$$

which holds for some r in the given range by the well-known "shift-and-add property" of 0/1 m-sequences (see [3, Theorem 5.3], for example).

We prove (3) by showing that $r(Y, u) = r(Y, u')$ for $u, u' \in \{1, 2, \dots, n-1\}$ implies $u = u'$. By (2), $r(Y, u) = r(Y, u')$ implies that

$$y_i y_{i+u} = y_i y_{i+u'} \quad \text{for all } i,$$

so that $y_{i+u} = y_{i+u'}$ for all i. Since Y has period n, and by assumption $u, u' \in \{1, 2, \dots, n-1\}$, we deduce that $u = u'$ as required. □

Given a length n sequence $A = (a_0, a_1, \dots, a_{n-1})$, define

$$S_A(j) := \sum_{i=0}^{j-1} a_i \quad \text{for } j = 0, 1, 2, \dots$$

to be the (running) sum of the first j elements of A (and take $S_A(0) := 0$), and define

$$W(Y) := \max_{0 \leq k < n} M(T^k(Y))$$

to be the maximum PSL over all cyclic shifts of A. We now use Lemma 5 to express $W(Y)$ for a given m-sequence Y in terms of the function S_Y. The reason for taking the maximum, rather than the mean or minimum, PSL over all cyclic shifts is that this maximum can be interchanged with the maximum in the definition of $M(A)$.

Lemma 6. *Let Y be an m-sequence of length n. Then*

$$W(Y) = \max_{0<u<n, \ 0 \leq c < n} \left| S_Y(c+u) - S_Y(c) \right|. \tag{4}$$

Proof. Write $Y = (y_i)$. By definition of $W(A)$, $M(A)$, $C_A(u)$, and $T^k(A)$,

$$\begin{aligned} W(Y) &= \max_{0 \le k < n} \max_{0 < u < n} \left| C_{T^k(Y)}(u) \right| \\ &= \max_{0<u<n,\ 0 \le k < n} \left| \sum_{i=0}^{n-u-1} y_{i+k} y_{i+k+u} \right| \\ &= \max_{0<u<n,\ 0 \le k < n} \left| \sum_{i=0}^{n-u-1} y_{i+k+r(Y,u)} \right|, \end{aligned}$$

by Lemma 5. Then by (3),

$$\begin{aligned} W(Y) &= \max_{0<u<n,\ 0 \le k < n} \left| \sum_{i=0}^{n-u-1} y_{i+k+u} \right| \\ &= \max_{0<u<n,\ 0 \le k < n} \left| \sum_{i=0}^{u-1} y_{i+k+n-u} \right|, \end{aligned}$$

replacing u by $n - u$. Put $c = k + n - u$, and then replace the resulting range for c by a more convenient range of n consecutive integers (using the fact that the sequence (y_i) is periodic, with period n), to obtain

$$\begin{aligned} W(Y) &= \max_{0<u<n,\ 0 \le c < n} \left| \sum_{i=0}^{u-1} y_{i+c} \right| \\ &= \max_{0<u<n,\ 0 \le c < n} \left| S_Y(c+u) - S_Y(c) \right| \end{aligned}$$

by definition of S_Y. □

We now show how to use Lemma 6 to determine $W(Y)$ for a given m-sequence Y of length n, by means of a single pass through the sequence. To illustrate the method, consider the m-sequence

$$Y = (+ - - - - + - + - - + + - + +),$$

(using $+$ to represent the sequence element 1, and $-$ to represent -1) of length $n = 15$, which is generated by the primitive polynomial $f(x) = 1 + x + x^4$. By direct calculation we find the maximum value $W(Y)$ of the PSL over all 15 cyclic shifts of Y to be 5 (attained by $C_{T^8(Y)}(6) = -5$). From the plot of $S_Y(j)$ for $0 \le j < 30$ shown in the upper graph of Figure 2, we see that the right hand side of (4) equals $|S_Y(1+9) - S_Y(1)| = |-4-1| = 5$, in accordance with Lemma 6. Now the right hand side of (4) (and therefore the value of $W(Y)$) could alternatively be evaluated as the difference between the maximum and minimum value of $S_Y(j)$ over the range $0 \le j < n$, namely $S_Y(1) - S_Y(10)$, because this maximum value occurs to the left of the minimum value ($1 < 10$). On the other hand, suppose that we instead choose a different cyclic shift of Y, namely

$$Y' = (- + + - + + + - - - - + - + -)$$

and consult only the resulting plot of $S_{Y'}(j)$ shown in the lower graph of Figure 2. The difference between the maximum and minimum value of $S_{Y'}(j)$ is $S_{Y'}(7) - S_{Y'}(1) = 4$, which does not give the correct value of $W(Y')$. The reason for this is that the maximum value of $S_{Y'}(j)$ in the range $0 \le j < n$ lies to the right of the smallest j at which the minimum occurs ($7 > 1$). But since we know that the sum of all elements of an m-sequence is -1, we can add one to this difference to obtain

$5 = S_{Y'}(7) - S_{Y'}(1) + 1 = S_{Y'}(7) - (S_{Y'}(16) + 1) - 1 = S_{Y'}(7) - S_{Y'}(7+9)$, which is the correct value of the right hand side of (4) for Y' (with corresponding values $u = 9$, $c = 7$).

In general, given an m-sequence Y of length n, we can use this method to determine the value of $W(Y)$, as either the difference between the maximum and minimum value of $S_Y(j)$ over $0 \le j < n$ or as one more than this difference, without having to consider all cyclic shifts of Y. We now give a formal statement and proof of this result.

Theorem 7. *Let Y be an m-sequence of length n. Let j_1 be the largest integer j in the range $0 \le j < n$ for which $\max_{0 \le j < n} S_Y(j)$ is attained, and let j_2 be the smallest integer j in the range $0 \le j < n$ for which $\min_{0 \le j < n} S_Y(j)$ is attained. Then $j_1 \ne j_2$ and*

$$W(Y) = \begin{cases} S_Y(j_1) - S_Y(j_2) & \text{if } j_1 < j_2, \\ S_Y(j_1) - S_Y(j_2) + 1 & \text{if } j_1 > j_2. \end{cases} \tag{5}$$

Proof. The values j_1 and j_2 are uniquely defined, and $j_1 \ne j_2$ (otherwise $n = 1$), so exactly one of the cases $j_1 < j_2$ and $j_1 > j_2$ holds. Writing $Y = (y_i)$, for all j we have

$$\begin{aligned} S_Y(j+n) &= S_Y(j) + \sum_{i=j}^{j+n-1} y_i \\ &= S_Y(j) - 1, \end{aligned} \tag{6}$$

since the sum of the n elements of Y is -1.

Now

$$\begin{aligned} \max_{0<u<n,\ 0 \le c<n} \left| S_Y(c+u) - S_Y(c) \right| &\le \max_{0 \le j < 2n} S_Y(j) - \min_{0 \le j < 2n} S_Y(j) \\ &= \max_{0 \le j < n} S_Y(j) - \min_{n \le j < 2n} S_Y(j), \end{aligned}$$

by (6), and using (6) again we obtain

$$\begin{aligned} \max_{0<u<n,\ 0 \le c<n} \left| S_Y(c+u) - S_Y(c) \right| &\le \max_{0 \le j < n} S_Y(j) - \left(\min_{0 \le j < n} S_Y(j) - 1 \right) \\ &= S_Y(j_1) - S_Y(j_2) + 1, \end{aligned}$$

by the definition of j_1 and j_2. By Lemma 6 it is therefore sufficient to exhibit integers c and u satisfying

$$0 \le c < n \text{ and } 0 < u < n \tag{7}$$

such that

$$\left| S_Y(c+u) - S_Y(c) \right| = \begin{cases} S_Y(j_1) - S_Y(j_2) & \text{if } j_1 < j_2, \\ S_Y(j_1) - S_Y(j_2) + 1 & \text{if } j_1 > j_2. \end{cases} \tag{8}$$

Case 1: $j_1 < j_2$.: Take $c = j_1$ and $u = j_2 - j_1$ so that the inequalities (7) are satisfied, and then

$$\left| S_Y(c+u) - S_Y(c) \right| = S_Y(j_1) - S_Y(j_2).$$

Case 2: $j_1 > j_2$.: Take $c = j_1$ and $u = j_2 - j_1 + n$ so that (7) is again satisfied, and by (6) we have

$$\begin{aligned} \left|S_Y(c+u) - S_Y(c)\right| &= \left|S_Y(j_2+n) - S_Y(j_1)\right| \\ &= \left|S_Y(j_2) - 1 - S_Y(j_1)\right| \\ &= S_Y(j_1) - S_Y(j_2) + 1. \end{aligned}$$

Combination of cases 1 and 2 gives (8), as required. □

Theorem 7 gives a key computational advantage over previous approaches. It allows us to determine the maximum PSL over all cyclic shifts of a given m-sequence Y in $O(n)$ operations. By contrast, the best previous algorithm involved an initial calculation of the aperiodic autocorrelations of Y, followed by an updating of all the autocorrelations in $O(n)$ operations for each of the n cyclic shifts, for a total of $O(n^2)$ operations (see [5, Section IV] and [4, Section 3.2]). The reduction in complexity from $O(n^2)$ to $O(n)$ operations, for each of the $\phi(2^m-1)/m$ cyclically inequivalent m-sequences of length $n = 2^m - 1$, allows us to extend significantly the range of computation. (In fact we need examine only half this number of cyclically inequivalent m-sequences for $m > 2$, corresponding to one member of each pair of reciprocal primitive polynomials, as in [5, Section V].) This in turn reveals trends in the growth rate of the PSL of m-sequences that were previously well beyond reach, as we shall see in Section 5.

5. The growth rate of the PSL of m-sequences

In this section we use Theorem 7 to study the growth rate of the PSL of m-sequences of length up to $2^{25} - 1$ numerically.

As in Section 3, let $\mathcal{F}_m$ be the set of primitive polynomials of degree m over GF(2), and let $\mathcal{Y}_m$ be the set of m-sequences of length $n = 2^m - 1$. Jedwab and Yoshida [5] calculated the minimum, mean, and maximum value of $M(Y)$ over all $Y \in \mathcal{Y}_m$ for $m \leq 15$, and concluded empirically that the mean PSL of m-sequences grows like $\Omega(\sqrt{n})$ and like $O(\sqrt{n \log n})$. They suggested [5, p. 2253] that "it would be challenging to collect sufficient computational data to settle [the question as to whether the PSL of m-sequences grows like $\Theta(\sqrt{n})$] with reasonable confidence." Nonetheless, we were able to show experimentally that in fact the PSL of almost all m-sequences appears to grow like $\Theta(\sqrt{n})$, as we now describe.

Table 2 lists the minimum, mean and maximum value of $M(Y)$ over all m-sequences $Y \in \mathcal{Y}_m$ for $m \leq 17$. These results extend to degrees 16 and 17 the exhaustive results given in [5, Table I], reflecting larger computational time rather than algorithmic improvement. Figure 3 shows the variation of the minimum, mean and maximum value of $M(Y)/\sqrt{n}$ with $\log n$ for $m \leq 17$. The additional data for $m = 16$ and 17 in Figure 3 do not change the conclusion reached in [5]: the PSL of m-sequences appears to grow like $\Omega(\sqrt{n})$, but we cannot tell whether it grows like $\Theta(\sqrt{n})$.

The data of central interest here, presented in the rest of Table 2 and in Figure 4, lead to a more powerful conclusion. These data are the minimum, mean and maximum value of $W(Y_f)$ over all primitive polynomials $f \in \mathcal{F}_m$ for $m \leq 25$. They were calculated from (5), using the online tables of $\mathcal{F}_m$ for $m \leq 25$ provided by Chabaud [1]. (While [1] does not guarantee that the listed polynomials are primitive, the properties of the sequence they generate can be used to verify primitivity.)

Now from (1), we have

$$\begin{aligned} \max_{Y \in \mathcal{Y}_m} M(Y) &= \max_{f \in \mathcal{F}_m} \max_{0 \le k < n} M(T^k(Y_f)) \\ &= \max_{f \in \mathcal{F}_m} W(Y_f) \qquad (9) \end{aligned}$$

by definition of $W(Y)$. Furthermore

$$\begin{aligned} \operatorname*{mean}_{Y \in \mathcal{Y}_m} M(Y) &= \operatorname*{mean}_{f \in \mathcal{F}_m} \operatorname*{mean}_{0 \le k < n} M(T^k(Y_f)) \\ &\le \operatorname*{mean}_{f \in \mathcal{F}_m} \max_{0 \le k < n} M(T^k(Y_f)) \\ &= \operatorname*{mean}_{f \in \mathcal{F}_m} W(Y_f), \qquad (10) \end{aligned}$$

and similarly

$$\min_{Y \in \mathcal{Y}_m} M(Y) \le \min_{f \in \mathcal{F}_m} W(Y_f). \qquad (11)$$

Each of (9), (10) and (11) can be verified for $m \le 17$ from Table 2.

Figure 5 shows that the minimum, mean and maximum value of $W(Y_f)/(\sqrt{n} \cdot \log\log n)$ over $f \in \mathcal{F}_m$ are all (broadly) non-increasing as $\log n$ grows. This suggests that $W(Y)$ grows like $O(\sqrt{n} \cdot \log\log n)$ for all m-sequences Y of length n, which by (9) implies that

(12) the PSL of all m-sequences of length n grows like $O(\sqrt{n} \cdot \log\log n)$.

This strengthens the empirical conclusion of $O(\sqrt{n \log n})$ growth reached in [5].

Figure 6 shows the variation with $\log n$ of the minimum, mean and maximum value of $W(Y_f)/\sqrt{n}$ over $f \in \mathcal{F}_m$ for $m \le 25$. (The maximum value of $M(Y)/\sqrt{n}$ for $m \le 17$ shown in Figure 3 appears in scaled form as part of the graph of the maximum value of $W(Y_f)/\sqrt{n}$ in Figure 6, in accordance with (9).) The most striking feature of Figure 6 is that the graph of $\operatorname{mean}_{f \in \mathcal{F}_m} W(Y_f)/\sqrt{n}$ levels out as m reaches 13. In fact, for m ranging from 13 to 25 (the limit of our computations), $\operatorname{mean}_{f \in \mathcal{F}_m} W(Y_f)/\sqrt{n}$ lies within less than 0.3% of 1.31. This suggests strongly that the growth of $\operatorname{mean}_{f \in \mathcal{F}_m} W(Y_f)$ is $\Theta(\sqrt{n})$. Assuming this to be true, (10) then implies that the growth of $\operatorname{mean}_{Y \in \mathcal{Y}_m} M(Y)$ is $O(\sqrt{n})$. By Corollary 3, we deduce empirically that

(13) the PSL of almost all m-sequences of length n grows like $\Theta(\sqrt{n})$

(in other words the proportion of m-sequences Y of length n for which $M(Y)$ grows like $\Theta(\sqrt{n})$ approaches 1 as $n \to \infty$). This is the first numerical evidence of $\Theta(\sqrt{n})$ growth in the PSL of any family of binary sequences. There is some irony involved in reaching the conclusion (13), since one of the principal aims of [5] was to demonstrate the lack of theoretical or empirical support for similar claims dating from the 1960s (see Section 3)! We emphasise, however, that currently there is no proof of (13), and that the experimental results presented here were obtained only by taking advantage of Theorem 7 and by using significant computational resources, neither of which were available to the originators of these claims.

While the initial reason for studying $W(Y)$ was simply the computational convenience provided by Theorem 7, the apparent levelling out of the middle graph in Figure 6 suggests that the quantity $\operatorname{mean}_{f \in \mathcal{F}_m} W(Y_f)$, namely the mean over all

m-sequences of length $2^m - 1$ of the maximum of the PSL over all cyclic shifts, has special significance and deserves further study.

It is natural to ask whether a similar result to Theorem 7 can be used to obtain data on the PSL of very long binary sequences other than m-sequences. But the proof of Theorem 7 depends on the shift-and-add property described in Lemma 5, and the only binary sequences having this property are m-sequences [3, Theorem 5.3], so this computational method does not appear to be directly applicable to other families of binary sequences.

The proof of Theorem 4 involves estimation of the maximum absolute value of an incomplete exponential sum (see Section 3), resulting in a growth rate of $O(\sqrt{n} \cdot \log n)$ for the PSL of m-sequences. Montgomery and Vaughan [9] showed, subject to the Riemann Hypothesis for Dirichlet functions, that the absolute value of another incomplete exponential sum is bounded by $O(\sqrt{n} \cdot \log\log n)$, suggesting a possible proof method for the empirical conclusion (12). However the incomplete exponential sum estimated in [9] is of a rather different type from that considered in [11] (compare [13, equation (3.2)] with [13, equation (8.4)]), and in particular does not correspond to a binary sequence having the shift-and-add property that is a key element of the proof of Theorem 4. Consequently it is not clear to us whether the similarity between (12) and the result of [9] carries any significance.

6. Conclusion

We have shown experimentally that:

1. the PSL of almost all binary sequences of length n appears to grow like $\Theta(\sqrt{n \log n})$
2. the PSL of all m-sequences of length n appears to grow like $O(\sqrt{n} \cdot \log\log n)$
3. the PSL of almost all m-sequences of length n appears to grow like $\Theta(\sqrt{n})$.

In particular, this answers empirically both of the questions posed in Section 1, although none of these conclusions has yet been proved. The best proven results on the growth rate of the PSL remain Theorem 1 for general binary sequences, and Corollary 3 and Theorem 4 for m-sequences.

Acknowledgements

We are grateful to Simon Litsyn for helpful discussions, and for drawing our attention to a result similar to Lemma 6. We are grateful to Florent Chabaud for providing the online tables [1] of primitive polynomials of degree up to 25 over GF(2), which were crucial to our numerical experiments. We thank the referees for their careful comments.

References

[1] F. Chabaud, *Primitive polynomials over* GF(2), online, <http://fchabaud.free.fr/English/default.php?COUNT=1&FILE0=Poly>, May 2007.

[2] S.W. Golomb, "Shift Register Sequences," Holden-Day, Inc., San Francisco, CA, 1967.

[3] S.W. Golomb and G. Gong, "Signal Design for Good Correlation: for Wireless Communication, Cryptography, and Radar," Cambridge University Press, New York, NY, 2005.

[4] J. Jedwab, *A survey of the merit factor problem for binary sequences*, in "Sequences and Their Applications — Proceedings of SETA 2004" (eds. T. Helleseth et al.), Lecture Notes in Comput. Sci. vol. 3486, Springer-Verlag, Berlin Heidelberg, (2005), 30–55.

[5] J. Jedwab and K. Yoshida, *The peak sidelobe level of families of binary sequences*, IEEE Trans. Inform. Theory, **52** (2006), 2247–2254.

[6] H.E. Jensen and T. Høholdt, *Binary sequences with good correlation properties*, in "Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, AAECC-5 Proceedings" (eds. L. Huguet and A. Poli), Lecture Notes in Comput. Sci. vol. 356, Springer-Verlag, Berlin, (1989), 306–320.

[7] R.J. McEliece, *Correlation properties of sets of sequences derived from irreducible cyclic codes*, Inform. Contr., **45** (1980), 18–25.

[8] I.D. Mercer, *Autocorrelations of random binary sequences*, Combin. Probab. Comput., **15** (2006), 663–671.

[9] H.L. Montgomery and R.C. Vaughan, *Exponential sums with multiplicative coefficients*, Invent. Math., **43** (1977), 69–82.

[10] J.W. Moon and L. Moser, *On the correlation function of random binary sequences*, SIAM J. Appl. Math., **16** (1968), 340–343.

[11] D.V. Sarwate, *An upper bound on the aperiodic autocorrelation function for a maximal-length sequence*, IEEE Trans. Inform. Theory, **IT-30** (1984), 685–687.

[12] H.D. Schotten and H.D. Lüke, *On the search for low correlated binary sequences*, AEU — Int. J. of Electronics and Communications, **59** (2005), 67–78.

[13] A. Tietäväinen, *Vinogradov's method and some applications*, in "Number Theory and its Applications" (eds. C.Y. Yildirim and S.A. Stepanov), Lect. Notes Pure Appl. Math. vol. 204, Marcel Dekker, New York, (1999), 261–282.

[14] R.J. Turyn, *Sequences with small correlation*, in Error Correcting Codes (ed. H.B. Mann), Wiley, New York, (1968), 195–228.

Received May 2007; revised October 2007.

E-mail address: dmitriev@deshaw.com
E-mail address: jed@sfu.ca

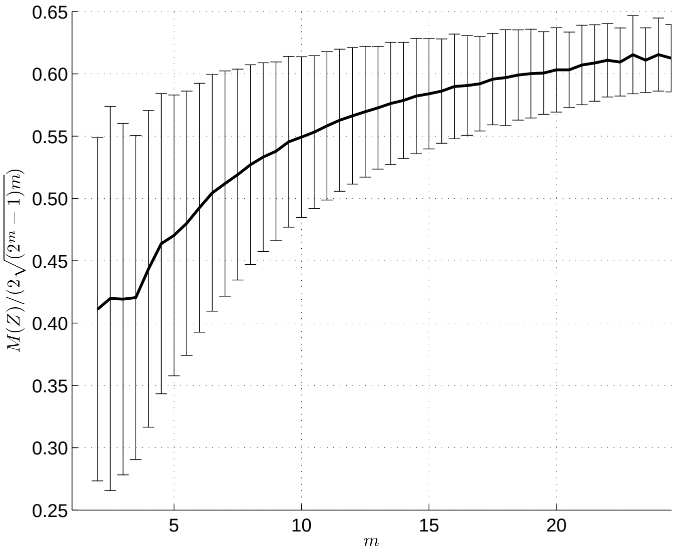

Figure 1. Comparison of the growth rate of the PSL of randomly-selected binary sequences Z of length $n = 2^m - 1$ with the Moon and Moser upper bound $2\sqrt{n \log n}$

lower value of m	upper value of m	# sequences
2	10.5	20000
11	12.5	10000
13	13.5	6000
14	14.5	5000
15	15.5	4000
16	16.5	3000
17	17.5	2000
18	18.5	1750
19	19.5	1500
20	20.5	1000
21	21.5	800
22	22.5	400
23	23.5	200
24	24.5	100

Table 1. Number of randomly-selected sequences of length $2^m - 1$ contributing to Figure 1

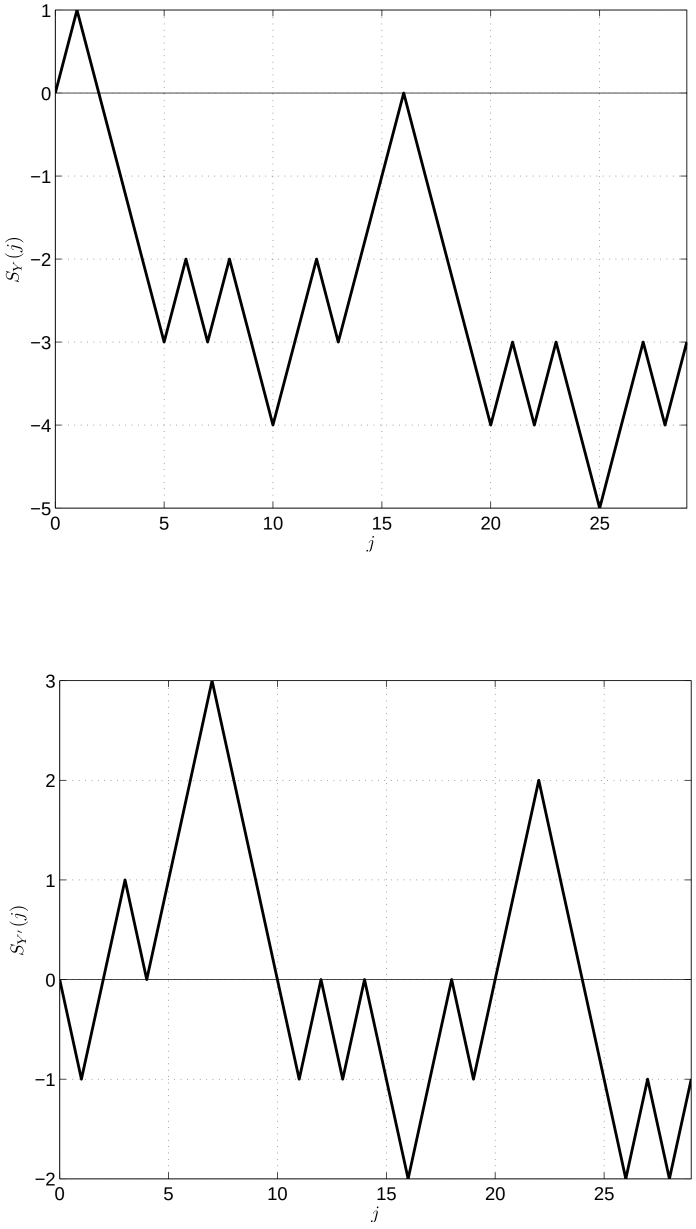

Figure 2. Running sum of two cyclic shifts of an m-sequence of length 15, for $0 \le j < 30$

m	$\min\limits_{Y\in\mathcal{Y}_m} M(Y)$	$\min\limits_{f\in\mathcal{F}_m} W(Y_f)$	$\operatorname*{mean}\limits_{Y\in\mathcal{Y}_m} M(Y)$	$\operatorname*{mean}\limits_{f\in\mathcal{F}_m} W(Y_f)$	$\max\limits_{Y\in\mathcal{Y}_m} M(Y)$	$\max\limits_{f\in\mathcal{F}_m} W(Y_f)$
2	1	2	1.33	2.00	2	2
3	1	3	2.14	3.00	3	3
4	3	5	3.60	5.00	5	5
5	4	6	5.16	6.67	7	7
6	6	10	7.84	10.33	11	11
7	8	13	11.71	14.22	16	16
8	13	19	16.88	20.38	22	22
9	19	25	24.89	29.04	34	34
10	29	36	35.93	40.77	46	46
11	42	51	52.20	58.39	68	68
12	61	72	76.45	84.51	107	107
13	85	97	108.74	118.13	144	144
14	125	141	156.08	167.35	207	207
15	175	201	222.28	236.14	295	295
16	258	281	318.80	335.22	433	433
17	363	391	453.87	473.63	626	626
18	—	544	—	669.45	—	860
19	—	775	—	947.95	—	1262
20	—	1066	—	1340.99	—	1842
21	—	1501	—	1896.53	—	2619
22	—	2128	—	2681.38	—	3635
23	—	3010	—	3793.22	—	5326
24	—	4237	—	5362.74	—	7546
25	—	5905	—	7586.37	—	11291

Table 2. Calculated values of $M(Y)$ and $W(Y)$ for m-sequences Y of length $n = 2^m - 1$

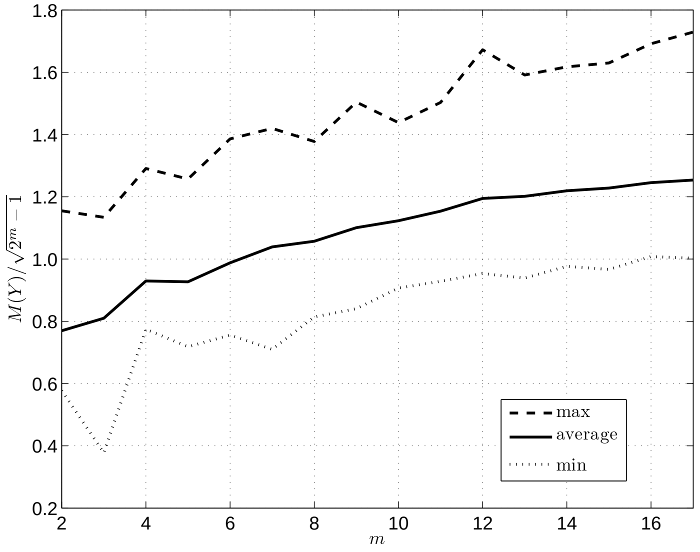

Figure 3. Comparison of the growth rate of $M(Y)$ with $\sqrt{n}$

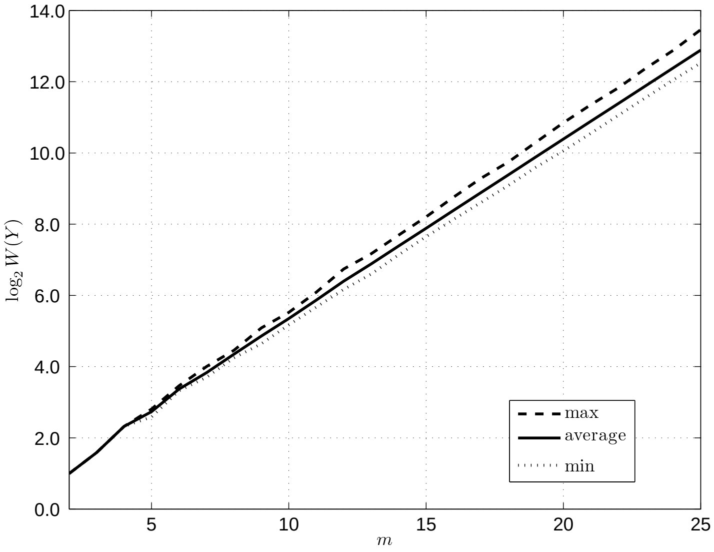

Figure 4. Minimum, mean and maximum values of $W(Y)$

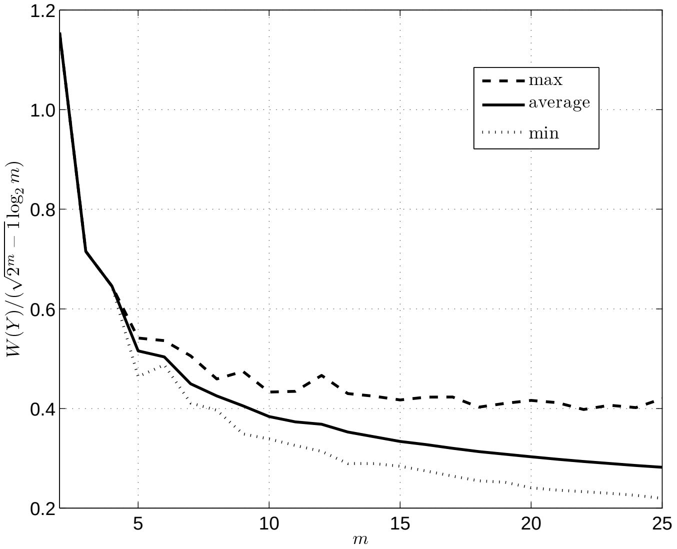

FIGURE 5. Comparison of the growth rate of $W(Y)$ with $\sqrt{n}\cdot\log\log n$

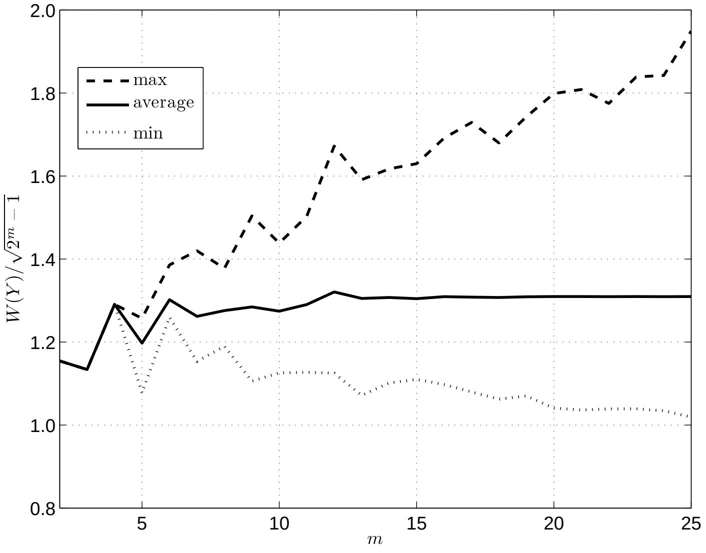

FIGURE 6. Comparison of the growth rate of $W(Y)$ with $\sqrt{n}$