THE MERIT FACTOR PROBLEM

PETER BORWEIN, RON FERGUSON, AND JOSHUA KNAUER

Abstract. The merit factor problem is of considerable practical interest to communications engineers and theoretical interest to number theorists. For binary sequences, although it is generally believed that the merit factor is bounded, it still has not been completely established that the number of even length Barker sequences, each with merit factor N, is bounded. In this paper, we present an overview of the problem and results of quite extensive searches we have conducted in lengths up to slightly beyond 200.

1. Introduction

For the sequence $A = [a_1, a_2, \ldots, a_N]$ the *kth acyclic autocorrelation coefficient*, or *kth shift sidelobe* $(0 \leq k \leq N-1)$, is given by

$$c_k = \sum_{j=1}^{N-k} a_j \overline{a_{j+k}}\,,$$

where the superimposed bar indicates the complex conjugate, in the case where the sequence takes complex values. Of particular interest are the polyphase sequences, where the modulus of each coefficient is 1. In these cases the 0th coefficient, or main lobe to engineers, is simply the length of the sequence. The other coefficients, or positive shift sidelobes, measure self-interference of a signal based on this sequence. This points, for example, to the use of signals based on sequences with low autocorrelation in radar detection. The energy in the kth shift sidelobe is defined as $|c_k|^2$ and higher sidelobe values correspond to energy inefficiencies in the signal. The base energy of the sequence is the total of the energies in these sidelobes, i.e.,

$$E = \sum_{i=1}^{N-1} |c_k|^2\,.$$

The merit factor, F, of the sequence relates energy in the sidelobes to energy in the main lobe,

$$F = \frac{N^2}{2E}\,.$$

2000 *Mathematics Subject Classification.* 11B83, 11Y55.

Key words and phrases. Integer sequences, autocorrelation, polynomials, merit factor, flatness.

The *merit factor* is a measure of the quality of the sequence in terms of engineering applications.

For number theorists, these coefficients arise in the expression for the modulus of a polynomial on the unit circle. For $p(z) = \sum_{j=1}^{N-k} a_j z^{j-1}$ the L_α norm of $p(z)$ on the unit circle C is given by

$$L_\alpha(p) = \left(\frac{1}{2\pi} \int_0^{2\pi} \left| p(e^{i\theta}) \right|^\alpha d\theta \right)^{1/\alpha} .$$

In particular, for polyphase sequences and $\alpha = 2$ and $z \in C$, we have

$$\left| p(z) \right|^2 = f(z)\overline{f(z)} = N + \sum_{k=1}^{N-1} \left(c_k z^{-k} + \overline{c_k} z^k \right),$$

so that

$$L_2(p) = \sqrt{N} . \tag{1}$$

For the L_4-norm we then obtain

$$\begin{aligned} L_4(p)^4 &= N^2 + 2 \sum_{k=1}^{N-1} c_k^2 \\ &= N^2 \left(1 + \frac{1}{F} \right), \end{aligned}$$

and

$$L_2(|p|^2 - N)^2 = L_4(p)^4 - N^2 = \frac{N^2}{F} . \tag{2}$$

This equation relates a higher merit factor with less deviation of $|p|$ from its L_2 average value of $\sqrt{N}$.

1.1. **Barker sequences.** Barker sequences are sequences for which $|c_k| \leq 1$ for $1 \leq k < N$, i.e., each autocorrelation coefficient has absolute value less than or equal one. For binary sequences, this implies that $|c_k| = 1$ for autocorrelation sums of odd length while $|c_k| = 0$ for sums of even length. Barker [1] was interested in their use for pulse compression of radar signals. They exist for lengths 2,3,4,5,7,11,13 and conjecturally for no longer length. Storer and Turin [45] proved that there are none for odd lengths greater than 13. For even lengths the conjecture has been proved for lengths up to 10^{22} by Leung and Schmidt [35].

For sequences consisting of 3rd, 4th, or 6th roots of unity, the condition $|c_k| = 0$ or 1 still applies. For sequences consisting of higher roots of unity or for more general polyphase sequences, we can have $0 < |c_k| < 1$. Borwein and Ferguson [8] have shown the existence of such sequences up to length 63.

1.2. **Skew-symmetric sequences.** A binary sequence $A = [a_1, a_2, \ldots, a_N]$ is *symmetric* (or *reciprocal*) if $a_j = a_{N+1-j}$ for each j and *antisymmetric* if $a_j = -a_{N+1-j}$ for each j. A skew-symmetric sequence is formed by interleaving an odd length symmetric sequence with an even antisymmetric sequence of length greater or less by 1. This means that for skew-symmetric $A = [a_1, a_2, \ldots, a_N]$, we have

$$a_j a_{j+k} + a_{N+1-j-k} a_{N+1-j} = 0$$

for odd k, implying that all even length autocorrelation sums are 0. Since half of the sidelobe energies are zero, it is natural to search for sequences of high merit factor among skew-symmetric sequences.

The equivalent condition with generalized or polyphase sequences would require a conjugate skew-symmetric sequence, i.e., a sequence which is conjugate reciprocal interleaved with a sequence which is the negative of its conjugate reciprocal. A difference here is that the odd length sequence may be either conjugate reciprocal or the negative of its conjugate reciprocal. In either case, we have

$$a_j \overline{a_{j+k}} + a_{N+1-j-k} \overline{a_{N+1-j}} = 0$$

for odd k so that again the even length autocorrelation sums are all 0. Thus we may expect to find high merit factor sequences among this class. In practice, however, we find that more optimal examples are obtained among reciprocal sequences of odd length. Then for any k,

$$a_j \overline{a_{j+k}} + a_{N+1-j-k} \overline{a_{N+1-j}} = a_j \overline{a_{j+k}} + a_{j+k} \overline{a_j}$$

is purely real. In these cases, the imaginary part of each autocorrelation disappears, so, in effect, half of the sidelobe energy expansion terms disappear for these sequences as well.

The square of the middle entry is either 1 or -1. For binary sequences, this implies that an odd length symmetric sequence is interleaved with an even length antisymmetric sequence, since $i = \sqrt{-1}$ cannot be an entry.

In all of these cases, specifying the entries up to and including the middle term is enough to determine the entire sequence, so searching unrestricted sequences at length N is comparable in complexity to searching symmetrics, antisymmetrics and skew-symmetrics at approximately double the length.

Thus, with binary sequences, we have evidence that, with high probability, we have found the optimal merit factor sequences for binary sequences up to length 85, and for skew-symmetrics up to length 165. For polyphase sequences, we believe the optimal results to be highly accurate to length 45. The best examples found at longer odd lengths are often from searches restricted to symmetric searches.

1.3. **Sequence equivalence.** There are a number of operations for which the sidelobe energies of a sequence remain unchanged and which generate a group under composition. These include:

1. Multiplication of all entries by a constant of modulus 1.
2. Taking the complex conjugate of all entries.
3. Sequence reversal.
4. Multiplication of successive entries by linearly increasing powers of a constant of modulus 1.

As operations on the space of binary sequences, the second is redundant while the remaining three generate a noncommutative group of order 8. Multiplying all entries by -1 or every second entry by -1 will give a new sequence, so these operations applied to a sequence of at length at least 2 will produce at least 4 different sequences. In most cases, sequence reversal will add 4 more. However, for symmetric, antisymmetric or skew-symmetric sequences the number produced remains at 4.

Using the first and fourth operations, any sequence can be transformed to a sequence with 1's as the first two entries. Using sequence reversal, followed again by transformation to a sequence with 1's as the first two entries, we obtain a second sequence, not necessarily different. Applying complex conjugation to these gives us a total of 4 from which to choose a canonical representative of the orbit class of the original sequence. One method is to compare successive entries and prefer a sequence for which the entry has a smaller argument. Normalization to this canonical form allows us to identify equivalence between sequences with the same base energy.

1.4. **Flat polynomials.** 'Flatness' of a polynomial on the unit circle is a term used to describe closeness of the modulus of its values to the average value over the whole circle. Equation 2 shows how the merit factor may be used to provide a measure for flatness of ±1 polynomials. This is further illustrated in Figures 1, 2, 3 below, showing the modulus of the polynomial $p(z) = \sum_{j=1}^{n} a_j z^{j-1}$ on the unit circle, where each $A = [a_1, a_2, \ldots, a_{63}]$ is a sequence of length 63.

Figure 1 arises from the randomly generated binary sequence

$$42F11C5DFFE24B8E$$

in hexadecimal notation with merit factor 1.4185. Here the leading 4 converts to 0100 in binary and $1, -1, -1$ in terms of ±1 coefficients (dropping the leading 0), while the following 2 converts to 0010 in binary and $-1, -1, 1, -1$ in ±1 coefficients.

Figure 2 represents the polynomial formed for the binary sequence

$$6C9B015052F14339$$

with merit factor $F = 9.5870$, which we believe is optimal for binary sequences of this length.

Figure 3 is a graph of the modulus of the polynomial formed with coefficients from the polyphase Barker sequence of length 63 with entries $a_j = e^{2\pi i \phi_j}$

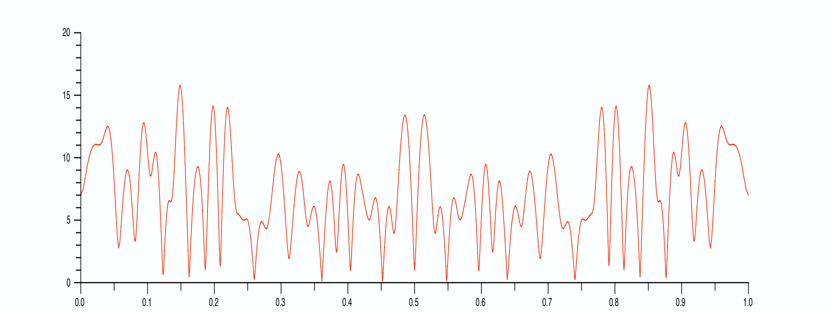

FIGURE 1. Modulus of $p(e^{2\pi it})$ for binary sequence with $F = 1.4185$ and $0 \le t \le 1$.

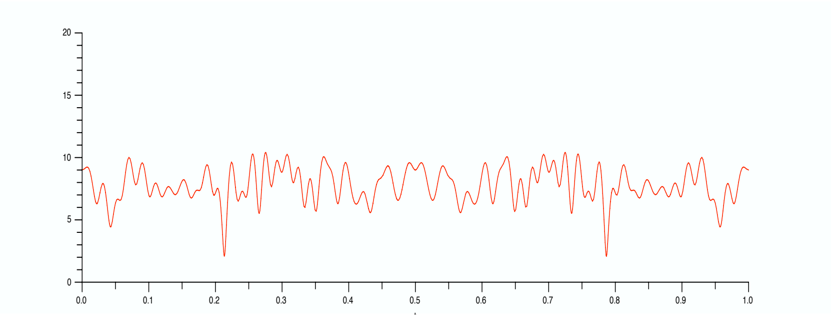

FIGURE 2. Modulus of $p(e^{2\pi it})$ for binary sequence with $F = 9.5870$.

with $\phi_1, \phi_2, \ldots, \phi_{63}$ having the values

0.000000, 0.000000, 0.044072, 0.100041, 0.124944, 0.044316, 0.915805,
0.834292, 0.896073, 0.072380, 0.153734, 0.145180, 0.264172, 0.409227,
0.678385, 0.779028, 0.703430, 0.582492, 0.464976, 0.434226, 0.137145,
0.048468, 0.004949, 0.928442, 0.365491, 0.394539, 0.867998, 0.074881,
0.666226, 0.614514, 0.194754, 0.471911, 0.761195, 0.956267, 0.323923,
0.119675, 0.556891, 0.854043, 0.099691, 0.332923, 0.935108, 0.561814,
0.731794, 0.132518, 0.422282, 0.875526, 0.519252, 0.026738, 0.368575,
0.879993, 0.399091, 0.939885, 0.425655, 0.919075, 0.551357, 0.209371,
0.855254, 0.577566, 0.272426, 0.992504, 0.662106, 0.376538, 0.022081,

which has $F = 37.5022$.

A high merit factor does not guarantee uniformity of closeness, since there may be narrow domains where spikes occur as illustrated in Figure 2. Still, the tendency toward a more uniform flatness with increasing merit factor is shown.

In his book [37], Littlewood introduced this notion of flatness and its expression in terms of the L_2 and L_4 norms. At that time, a plot of known optimal merit factor values would have had the appearance of Figure 4.

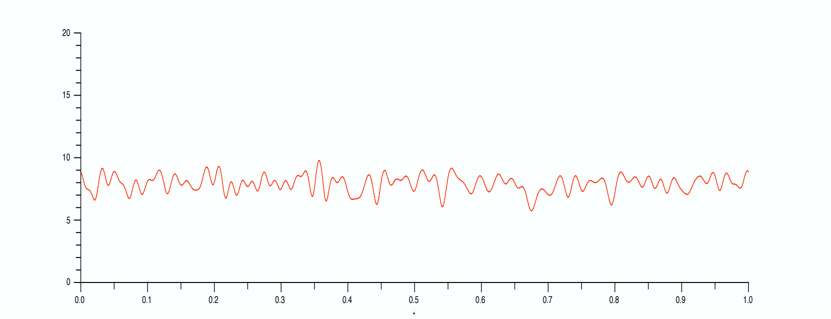

FIGURE 3. Modulus of $p(e^{2\pi it})$ for polyphase sequence with $F = 37.5022$.

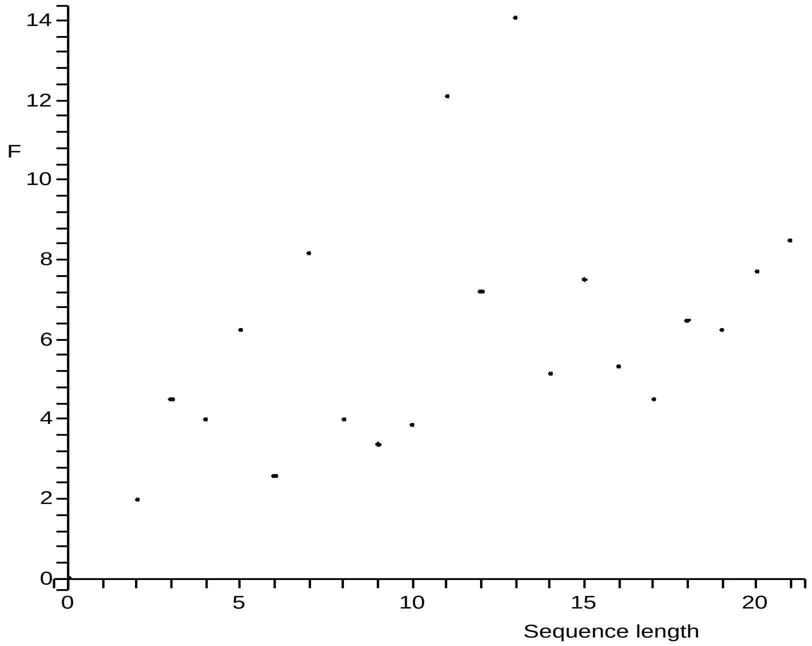

FIGURE 4. Optimal merit factor values for binary sequences to length 21.

The values for lengths 2,3,4,5,7,11, and 13 are in linear succession and correspond to Barker sequences. These lend considerable bias to the picture, the other points being more scattered. This may have been what led him to suggest the existence of an infinite sequence of polynomials p_{N_i} with merit factor of order $\sqrt{N_i}$ which would imply, in particular, that merit factors are unbounded. He further formulated the following conjecture:

Littlewood Conjecture: There exist constants C_1, C_2 such that we can find a sequence of polynomials p_N of increasing degree N with ±1-coefficients such that

$$C_1\sqrt{N} < \left|p_N(z)\right| < C_2\sqrt{N}$$

for all z on C.

The Rudin-Shapiro polynomials satisfy the upper bound of this conjecture. The existence of a sequence satisfying the lower bound, however, has not yet been confirmed. In the extension to polyphase sequences, Kahane [29] has both confirmed the conjecture and the existence of sequences with unbounded merit factors.

In contrast to Littlewood's suggestion, Golay [21], using a hypothesis that sidelobe energies move toward statistical independence as the length of binary sequences increase, developed an argument to show an asymptotic limit of 12.32... as maximal. This certainly does not settle the question, but gives an expression to what probably most researchers now believe, i.e., that an asymptotic upper bound exists.

It was noted by Turyn and later proved [26] that Legendre sequences rotated by 1/4 of their lengths will provide sequences with merit factors having an asymptotic limit of 6. A few authors suggested that this might be optimal. More recently, a further construction applied to Legendre sequences appears to produce sequences with asymptotic limit of approximately 6.34 for the merit factor [6], [33].

Computationally, this is still a very difficult problem. From results in the ranges for which adequate data can be collected, we can project that an asymptotic limit of $F > 7$ for sequences of increasing length is certainly expected. There is good evidence for $F > 8$ as well and even $F > 9$ appears likely. Finding another sequence with $F > 10$ beyond length 13, if indeed such exists, may be computationally out of range.

More comprehensive introductions to the history and applications of the binary merit factor problem are given in [25] and [28].

2. Search algorithms

Where a sequence has entries drawn from a finite alphabet, e.g., the Kth roots of unity, the number of sequences at a given length is finite. Thus, finding the optimal sequences of a given type up to a fixed length by checking the whole space is theoretically possible, but quickly becomes impractical as the length increases. More clever methods may be applied, which eliminate vast sections of the space from consideration as the search progresses in order to confirm optimal examples.

In contrast, at least beyond very short lengths, the search space for optimal polyphase sequences is not finite. Since the range of each coordinate is continuous, methods of calculus may be used to transform this into a finite problem. However, the number of local maxima and minima proliferates as the length increases, so this type of exhaustive search bogs down quite quickly.

Greater reach is achieved by using directed stochastic methods. Such algorithms include direct descent, simulated annealing, great deluge, genetic, and tabu search. Here optimality is not confirmed, but on the assumption that the number of locally optimal solutions within a given range is finite and the search method is not biased in locating these, statistical analysis using capture-recapture [40],[36] or the inverse collectors problem [34] can establish levels of confidence.

2.1. **Exhaustive searches.** For binary sequences, this is essentially a 2^{N-2} problem. Using a Gray code helps minimize recalculation through the iteration. To obtain precise information on base energy distributions we have conducted exhaustive searches up to length 44 for the general case and length 89 for skew-symmetrics.

2.2. **Branch and bound.** This is the method which has been used to confirm optimality of merit factor for general sequences up to length 60 [38], and for skew-symmetric sequences up to length 109 [9]. It uses the fact that the shorter and increasing length autocorrelation sums only involve terms progressing from the ends of a sequence to the middle. As a sequence is developed in this way, more sidelobe energy values become determined and better lower bounds for others are established. If this sum already exceeds some predetermined bound, then continued development can be aborted and the iteration passed to an earlier stage. The amount of truncation and thus the speed of the search space depends on this bound. A known sequence of low base energy supplies a good initial bound. This can be replaced when a better example is found.

2.3. **Directed stochastic searches.** A search starts either with a sequence in which all the coordinate values are randomly generated or a random selection of the coordinate values of an existing sequence are regenerated. A coordinate position is chosen either at random or by some directed method, a change in value is proposed and either accepted or rejected according to some selection criterion, which then returns either an altered or the same sequence accordingly. This second process repeats until either no further improvement is possible using this method or some other limit is reached. The process then either reverts to an earlier position or restarts. The decision to discontinue may be based on either time, computational resources, or an estimation that any improvement is unlikely.

For the descent method, changes are allowed only if there is an improvement in quality. Initially the coordinates for proposed changes are chosen at random, though a final stage may be added where coordinates are chosen iteratively until no further improvement is possible. Experience has shown, however, that better results are achievable if the demand for improvement at each stage is relaxed.

At the starting stage we have used the great deluge approach to locate a sequence of merit value above an initial bound. The method of descent was then used first by random and finally by consecutive (with wrap-around) choice of coordinates until either termination or a sequence with merit value exceeding a second bound was located.

Two levels of intensified search then followed for appropriate sequences in an approach derived from the genetic method.

1) For the same length sequence coordinates were chosen in order from the beginning of the sequence. Any sequence found exceeding another preset bound was chosen for further development. Otherwise the process backtracked to an earlier stage from which a further choice was possible.

2) An appropriate sequence from the first stage was then stripped back, dropping coordinate entries at both ends successively until either the merit factor dropped below an established bound or the last coordinate change was reached but not exceeded. This core sequence was then extended at either end iteratively while still retaining the merit factor bound.

These intensification methods ensure a comprehensive search of clusters of sequences of high merit factor at neighbouring lengths.

For binary sequences, there is a single option for changing a coordinate value or two choices for an extension of one unit.

For polyphase sequences, after a coordinate position was chosen for investigation, the base energy was expressed as a function of this coordinate variable, keeping other coordinates at their established values. This function was found to have up to two minimum values. The coordinate value at the lower of these was chosen for the proposed change.

2.4. Comparison of methods.

For determining optimal values for the merit factor of binary sequences, the resources required to conduct the branch and bound search at length 60 can be considered roughly the equivalent of those required to perform a fully exhaustive search at length 53, or a directed stochastic search at length 90 with a 99% confidence level of success. In terms of complexity, this translates to $O(2^N)$ for exhaustive methods, $O(1.84^N)$ for branch and bound and $O(1.5^N)$ for the directed stochastic algorithm used here.

3. Results

3.1. Growth trends for merit factor values.

Figures 5 and 6 show plots of best known merit factor values recorded for binary sequences and for the more general ployphase sequences. We believe Figure 6 shows probable optimal values for unrestricted binary sequences with lengths into the late 80's and for skew-symmetrics up to length 160. For polyphase sequences this confidence extends to length 45.

The horizontal line in Figure 5 is drawn for reference at $F = 9$. Up to length 90, there appears to be a trend for more of these values to approach or exceed 9. Beyond $N = 112$ all records are derived from skew-symmetrics, which appear to reflect trends for unrectricted sequences at half the lengths.

Figure 6 shows the much more rapid growth for polyphase sequences.

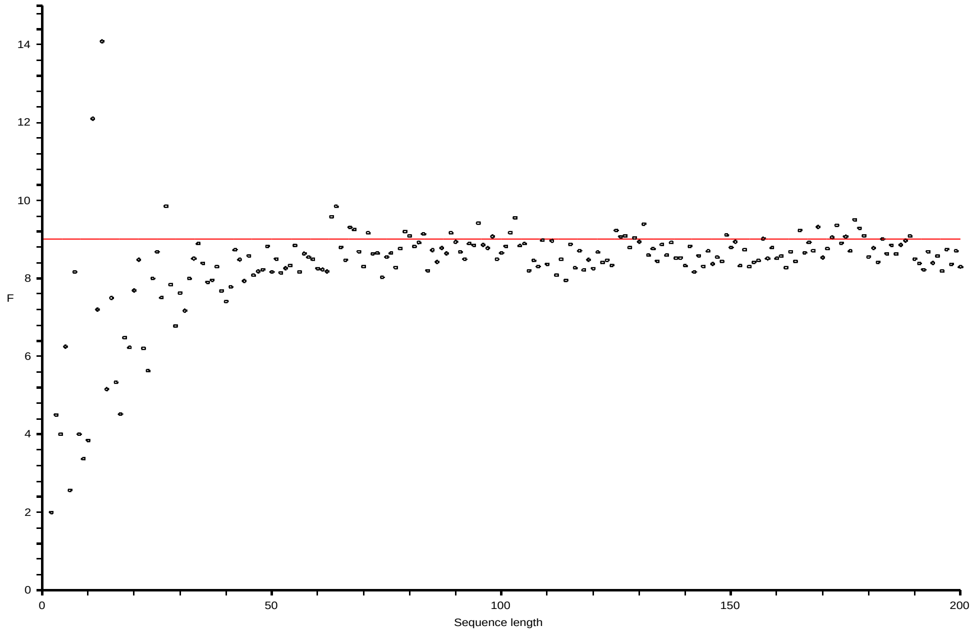

FIGURE 5. Largest merit factors recorded for binary sequences.

3.2. **Distribution of base energies for binary sequences.** Figure 7 shows histograms of all 2^{39} values of the reciprocal of the merit factor, $1/F = E/(2N^2)$, at length $N = 39$, normalized to have unit area. Each base energy value is congruent to 3 modulo 4. The graph on the left seems to be aligned with two smooth curves. A more careful analysis finds that it is better separated into four sections, each corresponding to a separate congruence class modulo sixteen for the base energy and each seeming to conform to a smooth curve. An alternative is to combine the congruences classes 3,7,11,15 modulo 16, giving the histogram on the right.

This shows a remarkable resemblance to the extreme value distribution. In fact, this was discovered through curve-fitting as illustrated in Figure 8.

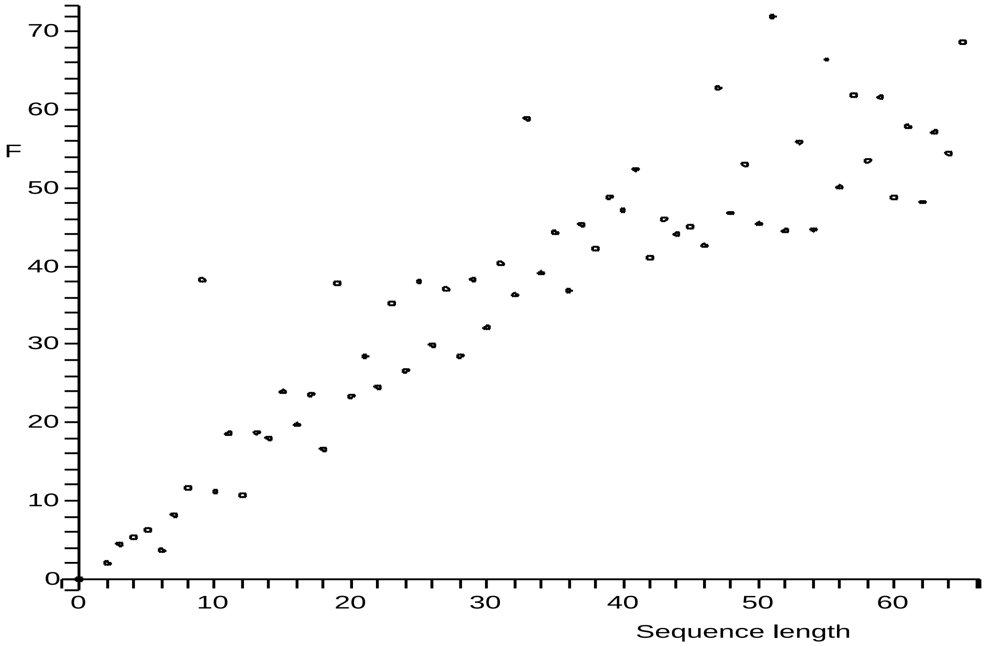

Figure 6. Largest merit factors recorded for polyphase sequences.

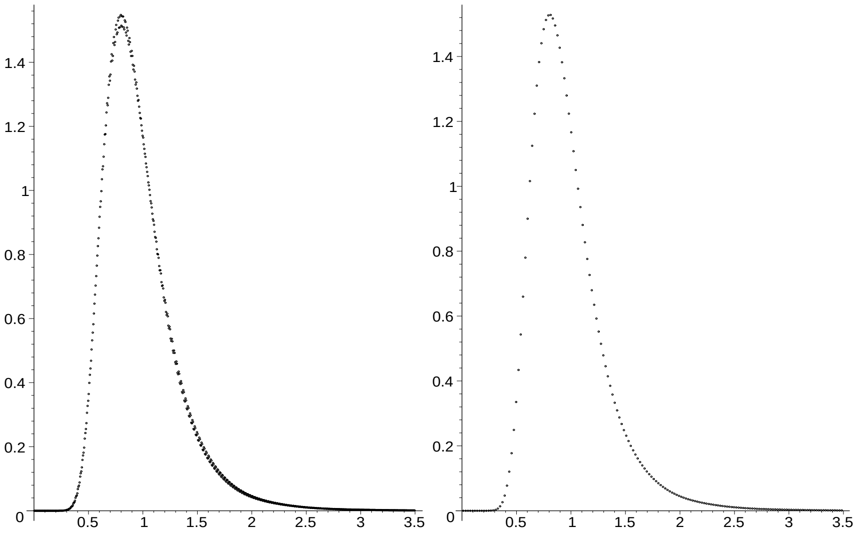

Figure 7. Full distribution of $1/F$ values at $N = 39$ and distribution combining counts from successive groups of 4.

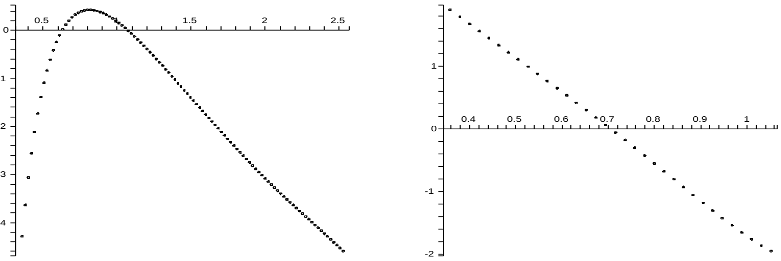

Figure 8. Plot of the logarithm of the values in the graph on the right in Figure 7 and and a plot of the logarithm of a linear fit to the right tail minus the values for this adjacent graph.

These graphs point to an exponential tail to the right but a doubly exponential growth on the left which are characteristic for the extreme value distribution. Figure 9 shows least squares fits of the extreme value distribution to histograms of $1/F$ values at lengths 39, 99, and 300. Since obtaining full distributions at lengths 39 and 300 is computationally out of range, these we obtained for 2^{37} and 2^{36} sequences generated by a random process.

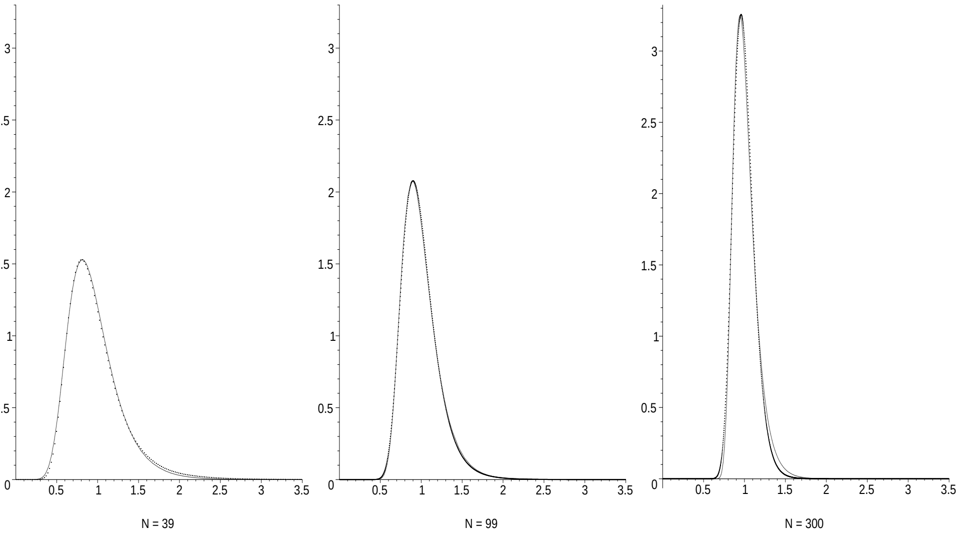

Figure 9. Least squares fit of extreme value distribution to $1/F$ distributions at lengths 39, 99 and 300.

Our primary interest is in the tail of the distributions on the left. The least squares fit is best at length 99. At 39, the curve overestimates the count of higher merit factor sequences while at 300 it gives an underestimate.

3.3. **High merit factor binary sequences.** Using our directed stochastic search method we have found approximately 2800 unequivalent sequences with merit factor greater than 8, of which probably more than 80% may be classified as new discoveries. This comprised two separate programs, the first designed to search through general unrestricted sequences and the second restricted to skew-symmetrics. The first found a sequence with merit factor greater than eight at length 115, while the second found such a sequence at length 233. The greatest length previously recorded was 161 by Militzer, Zamparelli and Beule[39]. The intensification methods described in the previous section were applied to high merit factor sequences found during the skew-symmetric search to find other non-skewsymmetric sequences of high merit in the surrounding cluster, finding sequences with $F \geq 8$ for all lengths between 100 and 200 except 112 and 114. A sequence with $F > 8$ at 112 was found using the first program.

We estimate the first program to have been close to exhaustive up to length 85, and the second to length 161 as outlined in the analysis below. Confidence that over 50% of the sequences with $F > 8$ have been found remains to lengths 90 and lengths 181 respectively. Comparing the percentage that the skew-symmetric sequences form of sequences found at odd lengths provides another rough yardstick. Between lengths 61 and 85 skew-symmetrics form 11% of sequences found at odd lengths with $F \geq 8$. These increase to 26% and 77% respectively for odd lengths from 91 and 99 and from 101 to 109. In fact from Figure 10 we expect this percentage to decrease from 11%, so we expect that we have found less than 40% and 12% of sequences with $F \geq 8$ in these respective ranges.

Figure 10 and Table 1 illustrate growth trends in numbers of high merit factor sequences. The first, Figure 10, is a log-linear graph of numbers of inequivalent sequences with $F \geq 7$. This suggests exponential growth in these numbers and gives what we suggest is the first empirical evidence that $\limsup F_{\max} > 7$.

Table 1 lists estimated numbers of sequences with $F > 8$ for general sequences and $F > 8$, $F > 8.5$ for skew-symmetrics at increasing lengths. The growth in numbers found in the general case is noticeable but still modest in lengths up to 85. For skew-symmetrics in lengths up to 181, where in depth searches were still possible, the growth in numbers with $F \geq 8$ is more apparent, and still apparent for $F > 8.5$.

Before our work, there were 15 inequivalent sequences with $F \geq 9$ known, one each at lengths 11, 13, 27, 64, 71, 83, 95, 105, 125, 127, 129, 131 and three at length 67. We have found 23 more which are listed in Table 2. There seems to be a trend for more of these to appear at longer lengths, but it is difficult to say more.

3.4. **The inverse collector's problem.** In collecting data on sequences with $F \geq 7$, we continued to run our programs at lengths up to 85 to the point that we could have good expectation that the best examples were collected. In advance, we do not know the size of the sample space of sequences

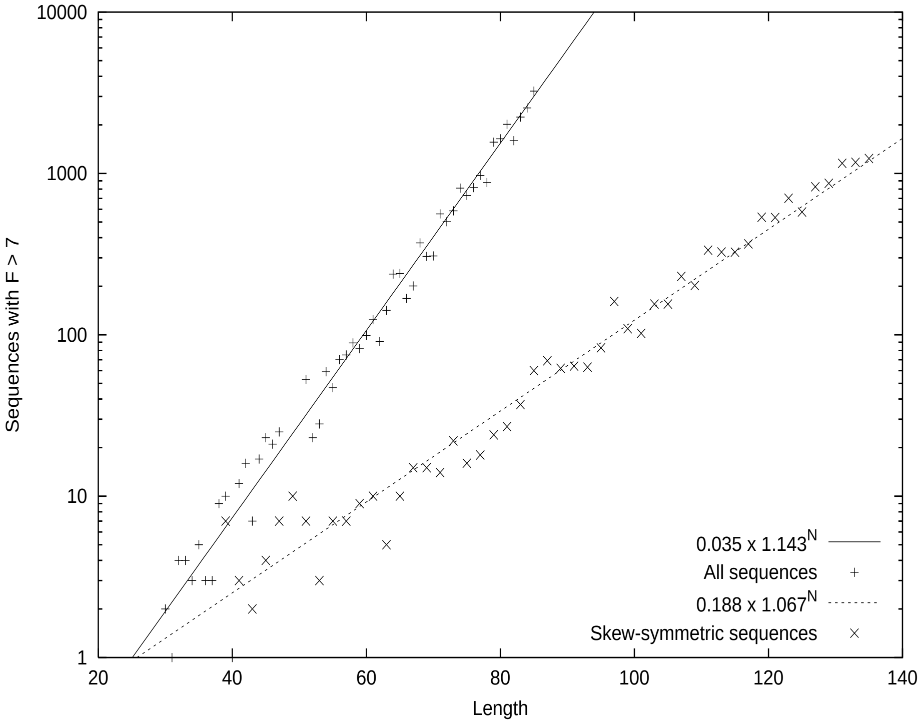

FIGURE 10. Growth in number of sequences with $F \geq 7$.

N	$F > 8$	$N^\dagger$	$F > 8$	$F > 8.5$
73	8	159	44	5
74	3	161	57	4
75	9	163	52	5
76	5	165	57	3
77	11	167	73	5
78	12	169	73	11
79	12	171	58	7
80	14	173	90	13
81	17	175	97	8
82	13	177	99	10
83	16	179	153	9
84	10	181	114	11
85	18	183	125	9

TABLE 1. Estimated numbers of inequivalent sequences with $F \geq 8$. Here † denotes skew-symmetric sequences.

N	F	Hexadecimal sequence
63	9.5870	64CBED0FAFEAC631
68	9.2480	FFD0B564E4D74798E
79	9.2050	7F36491D815A531AA871
80	9.0909	FFE81E89A8D1C665A9A5
89	9.1678	1FF924F246C19C2D4B8D454
95	9.3427	7FFC0FA333154B534DA71C69
98	9.0775	3FFAA55B45978719636C4F633
102	9.1746	383F0C38A4D5D6673480256A44
126	9.0720	3C7854315FE710B9990BB655FB2FE96D
149	9.0542	1C71C7AB46CDDABF9F82959501DCC6F016DB6D†
149	9.1137	1FE0003921C9CC3E4CBD0CE52CD8DA392AAA55†
157	9.0223	1F0600F83071FF993CC57ECD39955B6B294AA6B5†
165	9.2351	1D5B2B41689B1B24BAA6E846010E31B1887AF031FD†
169	9.3215	1C1C7C623B8EB1FD05DAFDD41DEBD5B0491226D6DAD†
172	9.0526	E03F9CF6030FF9EDBF293338351C5954B2A74D952A5
173	9.3179	18006FFE1FCF33F079C3D999D2D96B5334D5A5546AA9†
173	9.3645	1E03F9CF6030FF9EDBF293338351C5954B2A74D952A5†
175	9.0768	6AA32AF1A35998A5E530DAF8D30687D9983792FF37FE†
177	9.5052	1D3842C58FCB33401779175F7B977AAF330D49EC2E93D†
178	9.2915	3D3842C58FCB33401779175F7B977AAF330D49EC2E93D
179	9.0974	7A70858B1F9666802EF22EBEF72EF55E661A93D85D27A†
183	9.0073	6311C73B838E2A72BF958A85FD81ABF27F6DB5BB249136†
189	9.0847	1C39CE1FE1CBC67F3B7BF9002AB951713566D0DA55A4D92D†

TABLE 2. New sequences found with merit factor > 9. Here † indicates a skew-symmetric sequence.

with energies in this range. At some point, however, the continual repetition of previous examples suggests that we come close to exhausting this sample space. We use the statistical model described below to give substance to this observation. Part of this may be described as the inverse collector's problem as described in [11] and [34]. However, our problem is not specifically to establish the most probable size of the sample space, but to estimate the likelihood that we have found the example with the lowest base energy.

Let S be the size of the sample space, i.e., the total number of sequences, unique up to equivalence as described above, with $F \geq 7$. Let n be the number of trials in terms of sequences collected, and k the number of these which are different. Where M is the event that we have collected the optimal example, what we seek to evaluate is

$$P(M \mid n, k)\,,$$

the probability that we have the best example, given the values n and k arising from our data. We make assumptions:
(1) the occurrence of the different examples are equally likely;
(2) there is no specific bias on the actual size of the sample space in the range where this is significant.

Then we use

$$P(M \mid n,k) = \sum_{i\geq 0} P(M \mid S = k+i, k) P(S = k+i \mid n,k)\,.$$

From our assumptions, we have $P(M \mid S = k+i, k) = k/(k+i)$. Using Baysian probability theory, we then derive

$$P(S = k+i \mid n,k) = \frac{P(k \mid S = k+i, n)}{\sum_{j\geq 0} P(k \mid S = k+j, n)}\,.$$

Example: At length 80 we collected 9965 sequences of which 1636 were different. This calculates to a 0.998 probability that we have the optimal example [8], [9].

It remains, at present, computationally hard to verify optimality of merit factor values beyond length 60 for unrestricted binary sequences and about double this for skew-symmetrics. Other authors have published examples with 'high' merit factor without any further assessment of quality. This method of statistical analysis offers a way to estimate quality.

4. Comments

Determining the maximal merit factor for binary sequences of length N is widely regarded as a difficult task in combinatorial optimization. Indeed it appears as the fifth problem on CSPLib [24], a library of test problems for use in benchmarking constraint solvers.

We have found good evidence that the upper limit for $\max_N F > 8$ and even > 8.5. These maximal values may routinely exceed 9 in lengths over 200, but it would be difficult to establish this computationally. With the information we have collected, it is difficult to project whether values will continue to grow slowly or level off.

The match of the distribution of $1/F$ to the extreme value distribution for lengths under 100 is intriguing. If this continued, it would suggest a sharp cutoff for merit value as lengths increased. However, this fit becomes less good as we go further, giving an underestimate for the tail containing high merit factor values. Perhaps a match to a gaussian distribution is a better choice for longer lengths, but this would require further investigation.

The upper limit for $\max_N F$ is known to be infinite for polyphase sequences. Is the growth rate linear as might be suggested by 6? Is it possible to find finite alphabets for which the upper limit can be established as finite as well?

References

[1] R.H. Barker, Group synchronizing of binary digital sequences, *Communication Theory* (1953),273–287.

[2] H. Bauke, S. Mertens, Ground states of the Bernasconi model with open boundary conditions, `http://itp.nat.uni-magdeburg.de/ mertens/bernasconi/open.dat`, (2003).

[3] G.F.M. Beenker, T.A.C.M. Claasen, P.W.C. Hermens, Binary sequences with a maximally flat amplitude spectrum, *Philips Journal of Research* **40** (1985), 289–304.

[4] J. Bernasconi, Low autocorrelation binary sequences: statistical mechanics and configuration space analysis, *J. Phys.* **48** (1987), no. 4, 559–567.

[5] P. Borwein, *Computational excursions in analysis and number theory*, CMS Books in Mathematics **10**, Springer Verlag, New York, 2002.

[6] P. Borwein, K.-K.S. Choi, J. Jedwab, Binary sequences with merit factor greater than 6.34, *IEEE Trans. Inform. Theory* **50** (2004), no. 12, 3234–3249.

[7] P. Borwein and R. Ferguson, Polyphase sequences with low autocorrelation, 2003 (preprint).

[8] ———, Polyphase sequences with low autocorrelation, *IEEE Trans. Inform. Theory* **51** (2005), no. 4, 1564–1567.

[9] P. Borwein, R. Ferguson, J. Knauer, The merit factor of binary sequences, (to appear).

[10] F. Brglez, M. Stallmann, X.Y. Li, B. Militzer, Reliable cost predictions for finding optimal solutions to LABS problem: evolutionary and alternative algorithms, *Proceedings of The Fifth International Workshop on Frontiers in Evolutionary Algorithms (FEA2003), Cary, NC, USA, September 26-30*, 2003.

[11] B. Dawkins, Siobhan's problem: the coupon collector revisited, *The American Statistician* **45** (1991), no. 1, 76–82.

[12] H. Deng, Synthesis of binary sequences with good autocorrelation and crosscorrelation properties by simulated annealing, *IEEE Transactions on Aerospace Electronics Systems* **32** (1996), no. 1, 98–107.

[13] X. Deng, P. Fan, New binary sequences with good aperiodic autocorrelations obtained by evolutionary algorithm, *IEEE Communications Letters* **3** (1999), no. 10, 288–290.

[14] G. Dueck, New optimization heuristics, The great deluge algorithm and the record-to-record travel, *Journal of Computation Physics* **104** (1993), 86–92.

[15] C. de Groot, D. Würtz, Statistical Mechanics of Low Autocorrelation Skew-symmetric Binary Sequences, *Helvetica Physica Acta* **64** (1991), 86–91.

[16] C. de Groot, D. Wurtz, K.H. Hoffman, Low autocorrelation binary sequences: exact enumeration and optimization by evolutionary algorithm, *Optimization* **23** (1992), 369–384.

[17] L. Eberhardt, A course in quantitative ecology, `http://nmml.afsc.noaa.gov/Accessibility/AccLibQuantita.htm`, 2003.

[18] F.F. Ferreira, J.F. Fontanari, P.F. Stadler, Landscape statistics of the low-autocorrelation binary string problem, *J. Phys. A* **33** (2000), no. 48, 8635–8647.

[19] M.J.E. Golay, A class of finite binary sequences with alternate autocorrelation values equal to zero, *IEEE Transactions on Information Theory* **IT-18** (1972), no. 3, 449–450.

[20] ———, Sieves for low autocorrelation binary sequences, *IEEE Transactions on Information Theory* **IT-23** (1977), no. 1, 43–51.

[21] ———, The merit factor of long low autocorrelation binary sequences, *IEEE Transactions on Information Theory* **IT-28** (1982), no. 3, 543–549.

[22] ———, The merit factor of Legendre sequences, *IEEE Transactions on Information Theory* **IT-29** (1983), no. 6, 934–936.

[23] M.J.E. Golay, D.B. Harris, A new search for skewsymmetric binary sequences with optimal merit factors, *IEEE Transactions on Information Theory* **36** (1990), no. 5, 1163–1166.

[24] I.P. Gent, T. Walsh, CSPLib: a benchmark library for constraints, Technical report APES-09-1999 (1999), available from `http://csplib.cs.strath.ac.uk/.` A shorter version appears in the Proceedings of the 5th International Conference on Principles and Practices of Constraint Programming (CP-99).

[25] T. Høholdt, The merit factor of binary sequences, *Difference sets,sequences and their correlation properties*, Series C: Mathematical and Physical Sciences **542**, 227–237, Kluwer, 1999.

[26] T. Høholdt, H.E. Jensen, Determination of the merit factor of Legendre sequences, *IEEE Transactions on Information Theory* **34** (1988), no. 1, 161–164.

[27] F. Hu, P.Z. Fan, M. Darnell, F. Jin, Binary sequences with good aperiodic autocorrelations obtained by evolutionary algorithm, *Electronics Letters* **33** (1997), no. 8, 688–690.

[28] J. Jedwab, A survey of the merit factor problem for binary sequences, *Sequences and Their Applications – Proceedings of SETA 2004*, Lecture Notes in Computer Science **3486**, 30–55, Springer-Verlag, Berlin, 2004.

[29] J.-P. Kahane, Sur les polynômes à coefficients unimodulaires, *Bull. London Math. Soc.* **12** (1980), no. 5, 321–342.

[30] A. Kirilusha, G. Narayanaswamy, Construction of new asymptotic classes of binary sequences based on existing asymptotic classes, Dept. Math. and Comput. Science, University of Richmond, Summer Science Program Technical Report, July 1999.

[31] Ş.E. Kocabaş, A. Atalar, Binary sequences with low aperiodic autocorrelation for synchronization purposes, *IEEE Communications Letters* **7** (2003), no. 1, 36–38.

[32] R. Kristiansen, M.G. Parker, Binary Ssequences with asymptotic aperiodic merit factor > 6.3, preprint (2003).

[33] ———, Binary Sequences with merit factor > 6.3, *IEEE Trans. Inform. Theory* **50** (2004), no. 12, 3385–3389.

[34] E. Langford, R. Langford, Solution of the inverse coupon collector's problem, *Mathematical Scientist* **27** (2002), no. 1, 32–35.

[35] K.H. Leung, B. Schmidt, The field descent method, *Des. Codes Cryptogr.* **36** (2005), no. 2, 171–188.

[36] F.C. Lincoln, Calculating waterfowl abundance on the basis of banding returns, *U.S. Department of Agriculture Circulation* **18** (1930), 1–4.

[37] J.E. Littlewood, *Some problems in real and complex analysis*, D.C. Heath and Co. Raytheon Education Co., Lexington, Mass., 1968.

[38] S. Mertens, Exhaustive search for low-autocorrelation binary sequences, *Journal of Physics A: Mathematical and General* **29** (1996), no. 18, L473–L481. Updated results can be found at `http://itp.nat.uni-magdeburg.de/ mertens/bernasconi/open.dat.`

[39] B. Militzer, M. Zamparelli, D. Beule, Evolutionary Search for Low Autocorrelated Binary Sequences, *IEEE Transactions on Evolutionary Computation* **2** (1998), no. 1, 34–39.

[40] C.G.J. Petersen, The yearly immigration of young plaice into the Limfjord from the German Sea, *Report of Danish Biological Station* **6** (1896), 1–48.

[41] S. Prestwich, A hybrid search architecture applied to hard random 3-SAT and low-autocorrelation binary sequences, *The Sixth international conference on principles and practice of constraint programming*, Lecture Notes in Computer Science **1894**, 337–352, Springer-Verlag, Berlin, 2000.

[42] A. Reinholz, *Ein paralleler genetischer Algorithmus zur Optimierung der binären Autokorrelations-Funktion*, Masters Thesis, Universität Bonn, Oct. 1993.

[43] B. Schmidt, Cyclotomic integers and finite geometry, *J. Amer. Math. Soc.* **12** (1999), 929–952.

[44] M.R. Schroeder, *Number theory in science and communication*, 3rd edn., Springer-Verlag, New York, 1997.

[45] J. Storer, R. Turyn, On binary sequences, *Proc. Amer. Math. Soc.* **12** (1961), 394–399.