Exhaustive search for low-autocorrelation binary sequences

1996 J. Phys. A: Math. Gen. 29 L473

(http://iopscience.iop.org/0305-4470/29/18/005)

J. Phys. A: Math. Gen. **29** (1996) L473–L481. Printed in the UK

LETTER TO THE EDITOR

Exhaustive search for low-autocorrelation binary sequences

S Mertens†

Institut für Theoretische Physik, Otto-von-Guericke Universität, Postfach 4120, D-39016 Magdeburg, Germany

Received 17 May 1996

Abstract. Binary sequences with low autocorrelations are important in communication engineering and in statistical mechanics as ground states of the Bernasconi model. Computer searches are the main tool in the construction of such sequences. Owing to the exponential size $O(2^N)$ of the configuration space, exhaustive searches are limited to short sequences. We discuss an exhaustive search algorithm with run-time characteristic $O(1.85^N)$ and apply it to compile a table of exact ground states of the Bernasconi model up to $N = 48$. The data suggest $F > 9$ for the optimal merit factor in the limit $N \to \infty$.

1. Introduction

Binary sequences $S = \{s_1 = \pm 1, \ldots, s_N\}$ with low off-peak autocorrelations

$$C_k(S) = \sum_{i=1}^{N-k} s_i s_{i+k} \tag{1}$$

have applications in many communication engineering problems [1]. One exciting example has been their use in high-precision interplanetary radar measurements to check out space–time curvature [2].

Physicists prefer to consider binary sequences as one-dimensional systems of Ising-spins. In this context, low-autocorrelation binary sequences appear as minima of the energy

$$E(S) = \sum_{k=1}^{N-1} C_k^2(S). \tag{2}$$

This is the Bernasconi model [3]. It has long-range 4-spin interactions and is completely deterministic, i.e. there is no explicit or quenched disorder as in spin glasses. Nevertheless the ground states are highly disordered, quasi by definition. This self-induced disorder very much resembles the situation in real glasses. In fact, the Bernasconi model exhibits features of a glass transition such as a jump in the specific heat [3] and slow dynamics and ageing [4].

A clever variation of the replica method allows an analytical treatment of the Bernasconi model in the high-temperature regime [5, 6]. For the low-temperature regime analytical results are rare, and the ground states, in particular, are not known. With periodic boundary conditions, i.e. with

$$C_k = \sum_{i=1}^{N} s_i s_{(i+k-1)(\mathrm{mod}N)+1} \tag{3}$$

† E-mail address: stephan.mertens@physik.uni-magdeburg.de

0305-4470/96/180473+09$19.50

instead of (1), the construction of ground states is possible for special values of N. For example, for $N = 4n + 3$ prime, the modified Legendre sequence

$$s_j = \begin{cases} j^{1/2(N-1)} \bmod N & 1 \leqslant i < N \\ \pm 1 & i = N \end{cases} \tag{4}$$

yields $C_k^2 = 1$, the minimum possible value for odd N. Other ground states can be constructed from linear shift register sequences based on primitive polynomials over Galois fields. This construction requires $N = 2^p - 1$ with p prime. See [1, 6] for details.

For the model with open boundary conditions (1) no construction of ground states is known, even for special values of N. The Legendre sequences are far from the true ground states [7]. The only exact results have been provided by exhaustive enumerations; however, these are restricted to systems smaller than $N = 32$ [8] owing to the exponential complexity of the problem. Partial enumerations allow larger values of N but are not guaranteed to yield true ground states. Promising candidates for partial enumerations are the skew-symmetric sequences of odd length $N = 2n - 1$. These sequences satisfy

$$s_{n+l} = (-1)^l s_{n-l} \qquad l = 1, \ldots, n-1 \tag{5}$$

from which it follows that all C_k with k odd vanish. The restriction to skew-symmetric sequences reduces the effective size of the problem by a factor of two; however, the true ground states are *not* skew-symmetric for several values of N, as we shall see below.

Finding the ground states of the Bernasconi model has turned out to be a hard mathematical problem. Golay [3, 8] has conjectured that the maximal merit factor

$$F = \frac{N^2}{2E} \tag{6}$$

should obey $F \lesssim 12.32$ for $N \gg 1$. However, heuristic searches among skew-symmetric sequences up to $N = 199$ suggest $F \approx 6$ for long sequences [9], a value consistent with results from simulated annealing [3]. This large discrepancy indicates that the ground states, i.e. the sequences with high merit factors $6 < F \lesssim 12$, must be extremely isolated energy minima in configuration space. Stochastic search procedures including simulated annealing are not well suited to finding these 'golf holes'. Exhaustive search seems to be the only approach, at least for small systems. The complete configuration space has been searched up to $N = 32$ [8], the skew-symmetric subspace up to $N = 71$ [10, 11]. Fifty days of CPU time on a special-purpose computer have been used for an exhaustive search for binary sequences up to $N = 40$ that minimize $\max_k |C_k|$ [12].

In this letter we discuss a fast algorithm for the exhaustive enumeration. It is fast enough to yield exact ground states of the Bernasconi model up to $N = 48$ and can easily be modified for partial enumerations. The data are used to estimate the optimal merit factor in the large N limit.

2. The algorithm

Any algorithm that performs an exhaustive search for the ground state of the Bernasconi model has to cope with the enormous size (2^N) of the configuration space. This exponential complexity very soon limits the accessible values of N and calls for methods to restrict the search to smaller subspaces without missing the true ground state. *Symmetries* are an obvious device for cutting out portions of the configuration space. We shall see below that the use of symmetries can reduce the size of the search space by a factor of about an eighth. A method borrowed from combinatorial optimization, *branch and bound*, will prove useful

for reducing the complexity from $O(2^N)$ to $O(b^N)$ with $b < 2$. We shall further see that the enumeration problem is almost perfectly suited for *parallelization*.

2.1. Symmetries

The correlations C_k (1) are unchanged when the sequence is complemented or reversed. When alternate elements of the sequence are complemented, the even-indexed correlations are not affected, the odd-indexed correlations only change sign. Hence, with the exception of a small number of symmetric sequences, the 2^N sequences will come in classes of eight which are equivalent. The total number of non-equivalent sequences is slightly larger than 2^{N-3}.

The m left-most and m right-most elements of the sequence can be used to parametrize the symmetry-classes. For $m = 3$ and N odd, this gives 12 classes:

$$\begin{array}{ll} --\cdots--- & --\cdots-++ \\ --\cdots--+ & --\cdots-+++ \\ --\cdots-+- & --+\cdots--+ \\ --\cdots-++ & --+\cdots-++ \\ --\cdots+-- & --+\cdots+-- \\ --\cdots+-+ & --+\cdots+++ \end{array} .$$

For N even there are 10 classes. In general, the number, c, of symmetry classes that can be distinguished by m left-most and m right-most elements reads

$$c(m) = 2^{2m-3} + 2^{m-2+(N \bmod 2)} \tag{7}$$

and the number of non-equivalent configurations reduces to a fraction

$$\frac{c(m)}{2^{2m}} = \frac{1}{8} + \frac{1}{2^{m+2-(N \bmod 2)}}. \tag{8}$$

The optimal value, $1/8$, is approached with increasing m.

2.2. Branch and bound

Branch and bound methods are commonly used in combinatorial optimization [13] and (less frequently) in statistical mechanics [14, 15]. They solve a discrete optimization problem by breaking up its feasible set into successively smaller subsets (*branch*), calculating bounds on the objective function value over each subset and using these to discard certain subsets from further consideration (*bound*). The procedure ends when each subset has either produced a feasible solution, or been shown to contain no better solution that already in hand. The best solution found during this procedure is a global optimum.

The idea is of course to discard many subsets as early as possible during the branching process, i.e. to discard most of the feasible solutions before actually evaluating them. The success of this approach, which depends upon the branching rule and (very heavily) upon the quality of the bounds, can be quite dramatic. Numerical investigations have shown, for example, that the n-city travelling salesman problem (TSP) can be solved exactly in time $O(n^\alpha)$ with $\alpha < 3$ using branch and bound methods [13]! This is no contradiction to the exponential complexity of the TSP since the latter is the guaranteed, i.e. worst-case complexity, while the former refers to the *typical* case, averaged over many instances of the TSP.

In accordance with our symmetry classes, we specify a set of feasible solutions by fixing the m left-most and m right-most elements of the sequence. The $N-2m$ centre elements are

not specified, i.e. the set contains 2^{N-2m} feasible solutions. Given a feasible set specified by the m border elements, four smaller sets are created by fixing the elements s_{m+1} and s_{N-m} to ± 1. This is the branching rule. It is applied recursively until all elements have been fixed. The energy of the resulting sequence is compared with the minimum energy found so far. If it is lower, the sequence is kept as the potential ground state. After all $c(m)2^{N-2m}$ sequences have been considered, the potential ground state has turned into a true one.

Lower bounds are usually obtained by replacing the original problem over a given subset with an easier (relaxed) problem such that the solution value of the latter bounds that of the former. A good relaxation is one that (i) is easy and fast to solve and (ii) yields strong lower bounds. Most often these are conflicting goals.

A relaxation of the LABS problem is given by adjusting the free elements (i.e. the centre elements $s_{m+1}, \ldots, s_{N-m}$) to minimize all values C_k^2 *separately*, i.e. we replace the original problem

$$E_{\min} = \min_{\text{subset}} \left(\sum_{k=1}^{N-1} C_k^2 \right) \tag{9}$$

by the relaxed version

$$E_{\min}^* = \sum_{k=1}^{N-1} \min_{\text{subset}} (C_k^2) \leqslant E_{\min}. \tag{10}$$

Unfortunately, $E_{\min}^*$ is still not easy to calculate; however, we can proceed with our relaxation by providing a lower bound $E_{\mathrm{b}} \leqslant E_{\min}^*$. For that purpose we choose an arbitrary sequence from the subset with correlations C_k. Complementing a free element $s_i \mapsto -s_i$ can decrement $|C_k|$ at most by 2. Let f_k denote the number of terms $s_i s_{i+k}$ in C_k that contain at least one free element. This leads to

$$E_{\mathrm{b}} = \sum_{k=1}^{N-k} \min\{b_k, (|C_k| - 2f_k)^2\} \leqslant E_{\min}^* \leqslant E_{\min} \tag{11}$$

where $b_k = (N-k) \bmod 2 \in \{0, 1\}$ is the minimum value $|C_k|$ can attain. The f_k are given by

$$f_k = \begin{cases} 0 & k \geqslant N - m \\ 2(N - m - k) & N/2 \leqslant k < N - m \\ N - 2m & k < N/2 \end{cases} \tag{12}$$

i.e. the long-range correlations are not affected by our relaxation. E_{b} is not the strongest bound on $E_{\min}$, but its calculation is very fast.

Now we have gathered all the ingredients for formulating the branch and bound procedure **search** (algorithm 1). This procedure is called with two parameters specifying the subset to search: a binary sequence $S = \{s_1, \ldots, s_N\}$ and an integer m. The subset consists of all 2^{N-2m} sequences that can be generated from S by varying the $N-2m$ centre elements. S_{opt} is a global variable that holds the sequence with the minimum energy found so far. On entry, the size of the subset is checked. If it contains more than two sequences, branch and bound (lines 3–10) is applied. Otherwise the sequences in the subset are evaluated (lines 11–23).

The procedure **search** is called from a driving procedure with $c(m_0)$ subsets, each representing a different symmetry class. In practice, we used $m_0 = 6$ with 528 (N even) and 544 (N odd) symmetry classes.

Algorithm 1. Procedure **search** (S, m) to search for the minimum-energy configuration S_{opt} within the subset (S, m) of all configurations.

```
1:  n ← N − 2m; {number of free elements s_i}
2:  if n > 2 then {> 2 sequences in subset}
3:          if E_b(S, m) ⩾ E(S_opt) then {bound}
4:                  return;
5:          else {branch}
6:                  search (S, m + 1);
7:                  s_{m+1} ← −s_{m+1}; search (S, m + 1);
8:                  s_{N−m} ← −s_{N−m}; search (S, m + 1);
9:                  s_{m+1} ← −s_{m+1}; search (S, m + 1);
10:         end if
11: else if n = 1 then {2 sequences in subset}
12:         if E(S) < E(S_opt) then
13:                 S_opt ← S;
14:         end if
15:         s_{m+1} ← −s_{m+1};
16:         if E(S) < E(S_opt) then
17:                 S_opt ← S;
18:         end if
19: else {1 sequence in subset}
20:         if E(S) < E(S_opt) then
21:                 S_opt ← S;
22:         end if
23: end if
```

To measure the impact of branch and bound, we initiated two runs on the same machine: one 'straight' enumeration (omitting lines 3–5) and the other with activated bound mechanism. Figure 1 displays the effect of branch and bound on the CPU time. The straight enumeration shows the expected $\mathrm{O}(2^N)$ behaviour. Branch and bound reduces the complexity to $\mathrm{O}(1.85^N)$. Although this is still exponential, the gain in speed is worth the small effort. The branch and bound enumeration for $N = 44$ took about two days on a Sun UltraSparc I 170 workstation. This compares well with the extrapolated 68 days for the straight enumeration!

2.3. Parallelization

The different symmetry classes can be searched independently. Hence, the straight enumeration is perfectly parallelizable into $c(m)$ threads of control. Branch and bound complicates the situation. Whenever a better sequence is found by one thread, it should be communicated immediately to all other threads to ensure that the best $E(S_{\mathrm{opt}})$ is always used in the bounding test (line 3). But $E(S_{\mathrm{opt}})$ is accessed very frequently, so the necessary synchronization would spoil the parallelization. Giving each thread its own local copy of S_{opt} preserves perfect parallelization but abandons most of the benefits of branch and bound!

A solution to this dilemma is provided by the work-pile paradigm [16]. The symmetry classes to be searched are put on a central work pile and a number of worker threads are launched together. Each worker thread requests an assignment of work from the work pile (i.e. a symmetry class), performs the search, and then asks for a new work assignment. This process is repeated until all symmetry classes have been considered.

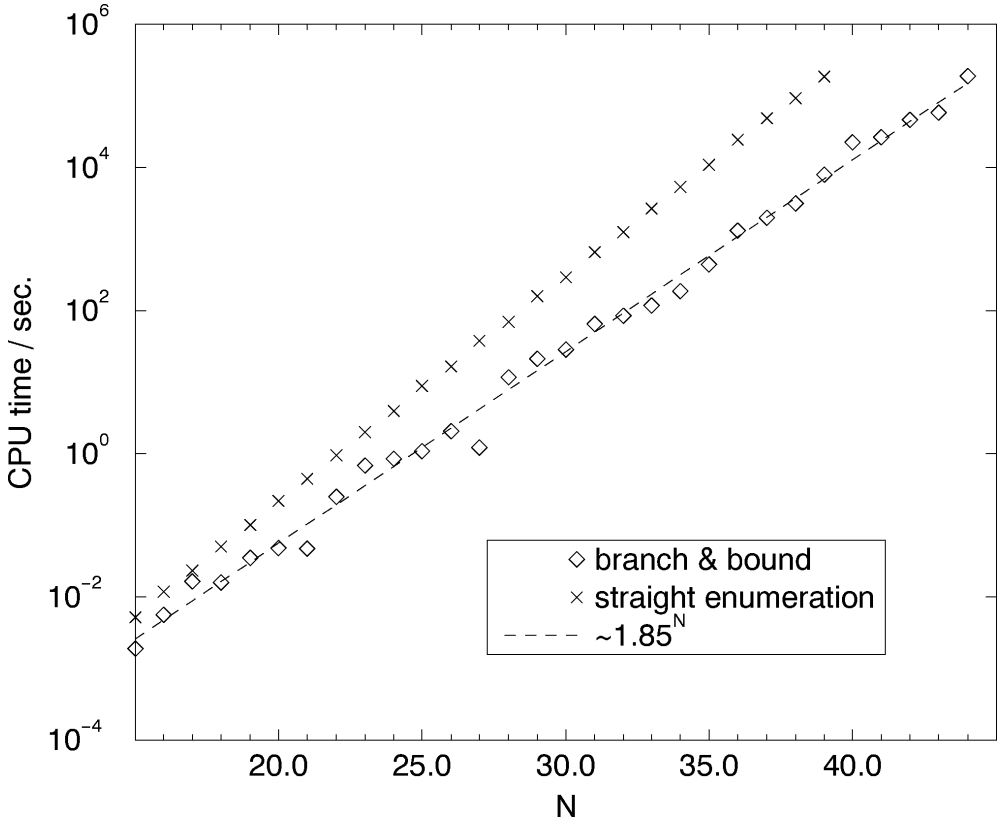

Figure 1. CPU time for exhaustive search algorithm versus N. Times are measured on a Sun UltraSparc I 170 workstation.

The access to the work pile has to be protected with a mutual exclusion lock, allowing only one thread at a time to read or modify data from the work pile. If each worker thread uses its own local copy of S_{opt}, this is the only synchronization needed. To propagate the best S_{opt} as quickly as possible among the workers, it is stored in the work pile. A worker that requests a new work assignment compares its own local S_{opt} with the global one and updates the one with the higher energy under the protection of the lock. This method limits the use of a suboptimal S_{opt} to the search within $n-1$ symmetry classes, where n is the number of worker threads. The delay in propagation of the optimal S_{opt} is minimized by choosing $c(m) \gg n$. In this case, the work-pile paradigm has the additional advantage of evenly distributing the load among all worker threads. Because of branch and bound, the enumerations in some symmetry classes may take considerably less time than in others. A worker that encounters these 'easy' classes simply gets more classes to search.

On a four-processor Sun SPARCstation 20, the number of worker threads ($\leqslant 4$) is much smaller than $c(m)$ for $m = 6$, so the work-pile paradigm should yield almost perfect parallelization. Figure 2 shows that this is indeed the case. The low speed-up factors for small N are due to the relative costs of thread generation and synchronization compared with the actual enumeration.

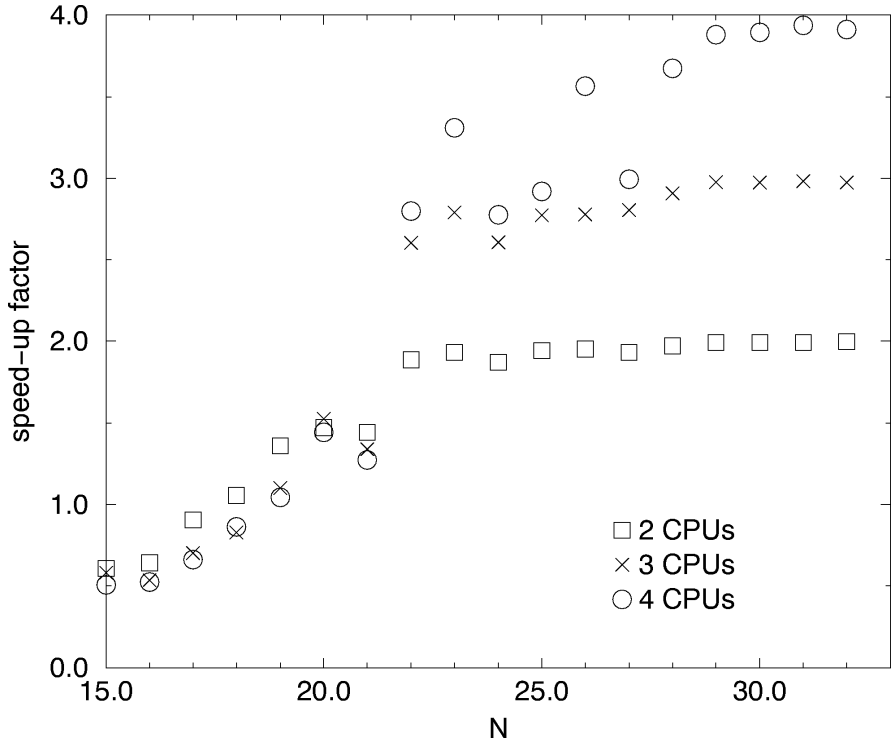

Figure 2. Speed-up factor of the branch and bound algorithm on a symmetric multiprocessor platform using two, three or four CPUs.

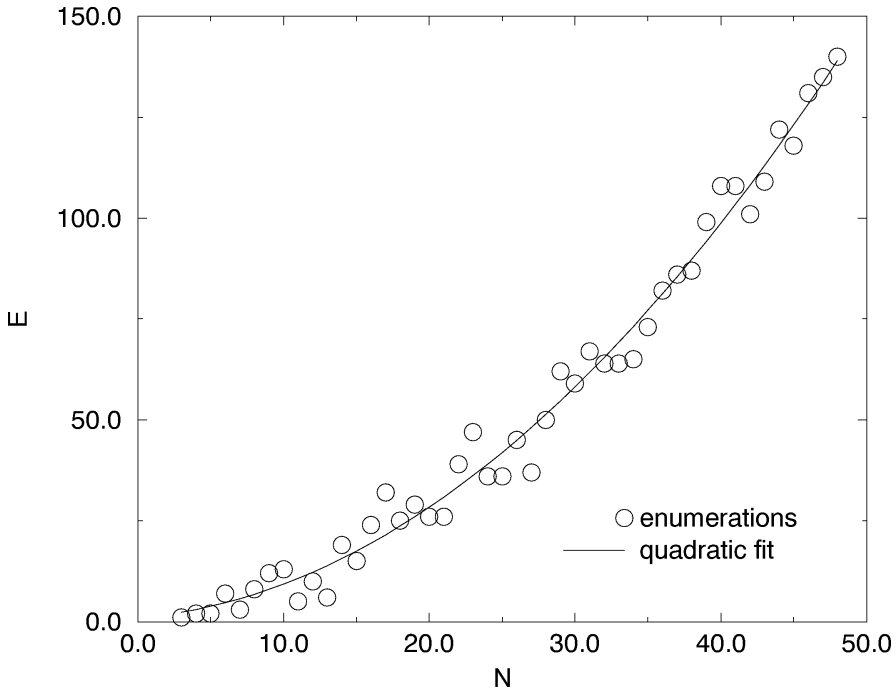

Figure 3. Ground state energy of the Bernasconi Hamiltonian versus N.

Table 1. Ground states of the Bernasconi model for $3 \leqslant N \leqslant 48$. Sequences are written in run-length notation: each figure indicates the number of consecutive elements with the same sign.

N	$E_{\min}$	Sequence
3	1	21
4	2	211
5	2	311
6	7	1113
7	3	1123
8	8	12113
9	12	42111
10	13	22114
11	5	112133
12	10	1221114
13	6	5221111
14	19	2221115
15	15	52221111
16	24	225111121
17	32	252211121
18	25	441112221
19	29	4111142212
20	26	5113112321
21	26	27221111121
22	39	51221111233
23	47	212121111632
24	36	2236111112121
25	36	337111121221
26	45	21212111116322
27	37	34313131211211
28	50	34313131211212
29	62	212112131313431
30	59	551212111113231
31	67	7332212211112111
32	64	71112111133221221
33	64	742112111111122221
34	65	842112111111122221
35	73	7122122111121111332
36	82	3632311131212111211
37	86	844211211111122221
38	87	8442112111111122221
39	99	82121121234321111111
40	108	44412112131121313131
41	108	343111111222281211211
42	101	313131341343112112112
43	109	11324321111717212112213
44	122	525313113111222111211121
45	118	821211212312343211111111
46	131	8234312312112122111111111
47	135	9234312312112122111111111
48	140	311111183214321222112112 1

3. Results

Using the multithreaded branch and bound algorithm and 313 hours of CPU time on a four-processor Sun SPARCstation 20, the ground states of the Bernasconi model have been

found up to $N = 48$ (table 1). The enumeration for $N = 32$ (the previous peak value) took only 80 seconds, $N = 39$ was performed in one hour. It is remarkable that of the 22 optimal skew-symmetric sequences in the range $5 \leqslant N \leqslant 47$ [10], seven (i.e. one third) have energies well above the true ground state energy. This should be kept in mind if one uses skew-symmetric sequences to estimate the ground state energy in the limit $N \to \infty$.

Figure 3 shows the ground state energies E versus N. In contrast to the model with periodic boundary conditions there are no visible regular patterns for special values of N [6]. The energies seem to follow $E \propto N^2$ for all values of N. A quadratic fit yields

$$F = \lim_{N\to\infty} \frac{N^2}{2E} = 9.3 \tag{13}$$

and leads us to the tentative conclusion that

$$F = \lim_{N\to\infty} \frac{N^2}{2E} > 9. \tag{14}$$

This estimate is in agreement with Golay's conjecture $F \lesssim 12.32$ and has to be compared with the value $F \approx 6.0$ found by heuristic searches for long skew-symmetric sequences [9] and by simulated annealing [3]. This indicates once more that heuristic and probabilistic methods fail to find the ground states of the Bernasconi model. Every algorithm of this kind should be judged by the percentage of values it finds from table 1.

Thanks are due to A Engel and J Richter for guiding the author's attention to the wonderful world of branch and bound and to S Kobe for providing helpful references.

References

[1] Schroeder M R 1984 *Number Theory in Science and Communication* (Berlin: Springer)
[2] Shapiro I I, Pettengill G H, Ash M E, Stone M L, Smith W B, Ingalls R P and Brockelman R A 1968 Fourth test of general relativity *Phys. Rev. Lett.* **20** 1265–9
[3] Bernasconi J 1987 Low autocorrelation binary sequences: statistical mechanics and configuration space analysis *J. Physique* **48** 559
[4] Krauth W and Mézard M 1995 Aging without disorder on long time scales *Z. Phys.* B **97** 127–31
[5] Bouchaud J P and Mézard M 1994 Self induced quenched disorder: a model for the glass transition *J. Phys. I France* **4** 1109–14
[6] Marinari E, Parisi G and Ritort F 1994 Replica field theory for deterministic models: I. binary sequences with low autocorrelation *J. Phys. A: Math. Gen.* **27** 7615–45
[7] Golay M J E 1983 The merit factor of Legendre sequences *IEEE Trans. Inf. Theory* **IT-29** 934–6
[8] Golay M J E 1982 The merit factor of long low autocorrelation binary sequences *IEEE Trans. Inf. Theory* **IT-28** 543
[9] Beenker G F M, Claasen T A C M and Hermens P W C 1985 Binary sequences with a maximally flat amplitude spectrum *Philips J. Res.* **40** 289–304
[10] Golay M J E 1977 Sieves for low autocorrelation binary sequences *IEEE Trans. Inf. Theory* **IT-23** 43–51
[11] de Groot C, Würtz D and Hoffmann K H 1992 Low autocorrelation binary sequences: exact enumeration and optimization by evolutionary strategies *Optimization* **23** 369–84
[12] Lindner J 1975 Binary sequences up to length 40 with best possible autocorrelation function *Electron. Lett.* **11** 507 1975
[13] Balas E and Toth P 1985 Branch and bound methods *The Traveling Salesman Problem* ed E L Lawler *et al* (Chichester: Wiley) pp 361–401
[14] Kobe S and Hartwig A 1978 Exact ground state of finite amorphous Ising systems *Comput. Phys. Commun.* **16** 1–4
[15] Hartwig A, Daske F and Kobe S 1984 A recursive branch-and-bound algorithm for the exact ground state of Ising spin-glass models *Comput. Phys. Commun.* **32** 133–8
[16] Kleiman S, Shah D and Smaalders B 1996 *Programming with Threads* (SunSoft Press/Prentice Hall)