# Low autocorrelation binary sequences: exact enumeration and optimization by evolutionary strategies

## C. De Groot , D. Würtz & K.H. Hoffmann

# LOW AUTOCORRELATION BINARY SEQUENCES: EXACT ENUMERATION AND OPTIMIZATION BY EVOLUTIONARY STRATEGIES

C. DE GROOT,* D. WÜRTZ** and K. H. HOFFMANN***

*Universität Heidelberg, Inst. für Theor. Physik, Heidelberg*
**Interdisziplinäres Projektzentrum für Supercomputing Eidgenössische Technische Hochschule Zürich, ETH-Zentrum*
***Universität Heidelberg, Inst. für Theor. Physik, Heidelberg*

We investigate skew-symmetric sequences with chain lengths up to $N = 71$, giving a complete table of all merit factors $F \geq 7$ and their associated configurations. We also calculate the exact thermodynamical properties of shorter chains ($N \leq 55$). We then introduce an evolutionary strategy, describing the properties of our search algorithm and comparing our results to those of other heuristic methods such as simulated annealing. We find the highest merit factors ever reached for chains of length $81 \leq N \leq 201$.

## 1. INTRODUCTION

The investigation of the properties of low autocorrelation binary sequences has a long history, and is of great interest for technical as well as theoretical reasons [1–6]. Autocorrelation sequences play an important role in several engineering applications, as a convenient method of synchronization in digital communication systems, or in their use for modulating pulses for radar and sonar ranging [4]. From the theoretical point of view, it seems to be amongst the most difficult optimization problems, and its cost function seems to show low lying states of extremely irregular structure.

One is searching for a configuration with minimal off-peak autocorrelation coefficients. The autocorrelation function is defined on a Boolean hypercube with configuration space $\bar{S} = \{\pm 1\}^N$. The elements of the configuration space are binary sequences $S \in \bar{S}$, $S = (s_1, \ldots, s_N)$ with length $N$. To measure the *quality* of such autocorrelation binary sequences the following standard criterion, called *merit factor F*, was introduced by Golay [1].

$$F = \frac{N^2}{2 \sum_{k=1}^{N-1} R_k^2},$$

(1)

where

$$R_k = \sum_{i=1}^{N-k} s_i s_{i+k}; \quad 1 \leq k \leq N - 1. \tag{2}$$

The optimal autocorrelation binary sequence can then be defined as $F \rightarrow$ *maximum* for fixed $N$ or alternatively

$$E(S) = \frac{N^2}{2F} = \sum_{k=1}^{N-1} R_k^2 \rightarrow \text{minimum}. \tag{3}$$

We see from equation (3) that the cost function $E$ becomes equivalent to the energy of a one dimensional spin system with long range four-spin interactions. One can therefore apply known methods from statistical mechanics. *Skew-symmetric* sequences of odd length $N = 2n - 1$ are defined by

$$s_{n+l} = (-1)^l s_{n-l}, \quad 1 \leq l \leq n - 1. \tag{4}$$

They are of special interest as all $R_k$ with odd $k$ vanish, and thus these sequences seem to be good candidates for high merit factors even in the complete phase space. Furthermore, the configuration space is essentially reduced, having now $2^{(N+1)/2}$ elements instead of $2^N$. Finally, we like to mention that, to our knowledge, it has not been shown whether this problem is NP-complete or not, but from previous numerical investigations it seems that this optimization problem is amongst the most difficult ones.

First attempts to find the autocorrelation binary sequence with the highest merit factor $F$ for a given length $N$ go back to 1977, and were undertaken by an exhaustive search in the complete configuration space (exact enumeration [2]). However, the number of elements in the configuration space grows exponentially with increasing chain length $N$. Therefore, in the literature one finds exact results only up to chain lengths $N \leq 32$ (Turyn, ref. in [3]) for the full configuration space and $N \leq 59$ [2] for the skew-symmetric configuration space.

With increasing $N$, one rapidly reaches the limit of available computer power. Exhaustive searches become impractical and one needs *sieves* to find those binary sequences with nearly optimal merit factors. Beenker *et al.* [4] started a very intensive investigation with different heuristic methods ranging from simple time domain toggling algorithms to the more sophisticated simulated annealing approach in 1985. These authors went up to chain lengths $N \leq 199$, however all the merit factors they found between $N = 101$ and 199 were about $F \sim 6$. In 1987 Bernasconi [5, 6] reinvestigated this problem very carefully from a statistical mechanics point of view, using simulated annealing techniques [7–9]. Within a "mean field" approach, he also evaluated the partition function for finite chains and from this analytical expression he derived the thermodynamic quantities in the thermodynamic limit $N \rightarrow \infty$. From this he rederived Golay's conjecture [3] that the largest possible merit factor $F$ in the limit $N \rightarrow \infty$ should be smaller than 12.3248. Bernasconi's comparison of his "mean field" results to data obtained from simulating annealing gave satisfactory agreement. However, investigating the energy landscape of the phase space, Bernasconi concluded that stochastic search procedures will normally not lead to merit factors larger than $F \sim 5$ for very long sequences, although such sequences with $5 \lesssim F \lesssim 12$ should exist. These numerical findings are in agreement with zero temperature scaling arguments applied by Moore [10] to the problem. His renormalization group transformation

shows that the problem is especially difficult! From another point of view, Wang [11] claimed that optimization by algorithms based on evolutionary strategies can give much better results than other heuristic search algorithms, including simulated annealing. As examples, he gave for sequences with $N = 101$, 103 and 105, the following optimal merit factors: $F = 6.911, 7.766$ and 7.614. However, considering Wang's results, one must also notice that Bernasconi [6] has found better sequences by simulated annealing than Beenker et al. [4], in particular one with $F = 9.56$ for a skew-symmetric chain of length $N = 103$.

We reinvestigate the problem from two different points of view. In the first part, we present many new results mainly for skew-symmetric sequences of length up to $N = 71$ obtained by exact enumeration. In the second part, we investigate an heuristic search algorithm based on evolutionary strategies, and compare the results with those obtained from simulated annealing methods.

## 2. EXACT ENUMERATION

The availability of more powerful computers now allows to reinvestigate the problem by exact enumeration. We have done this for skew-symmetric binary sequences, going far beyond the existing limit of $N = 59$. For chain lengths up to $N = 71$ we calculated not only the best merit factor but also all configurations with $F \geq 7$, or at least the five best values. For shorter chains ($N \leq 55$) we have determined all energy values giving us the distribution functions $\rho(E)$ or $\rho(F)$ and their exact thermodynamic properties.

The exact enumeration was done on the Cray X-MP at the ETH Zürich. Here we followed ideas of Golay [2] who showed that any skew-symmetric sequences can be considered as the interleaving of a symmetric sequence of the form $Ac\bar{A}$ and an anti-symmetric sequence of the form $BB'$. The bar indicates sequence reversal and the prime denotes complemention of the sequence. It is apparent that the autocorrelation $R_{2k}$ of the complete interleaved sequence is the sum of the autocorrelations of the two detwined components for $k$ shifts. It is also evident that all contributions to $R_{2k}$ come in pairs, except for a single $+1$ from the symmetric component for every even $k$, or a single $-1$ from the anti-symmetric component for every odd $k$. We have implemented this algorithm on the Cray using only logical and bit manipulation operations, coding the elements of the binary sequences $A$ and $B$ as 0 (for $-1$) and 1 (for $+1$) in two individual 64-bit Cray words.

The results are given in Table 1. All merit factors $F$, energies $E$, number of configurations, and the configurations themselves with $F \geq 7$ are presented for $N \leq 71$. Note that every configuration is fourfold degenerate due to spin inversion and sequence reversal. The configurations are encoded, starting with the central spin to the left, by their decimal counterpart, and can be easily decoded by going to their dual representation. For $N = 11$, for example, the maximum merit factor is given by $F = 12.1$ for the sequence $- - - + + + - + + - +$, where the signs are those of the corresponding $s_i$. Due to skew-symmetry, the whole sequence can be regained from the left part and the central spin, i.e. $s_1, \ldots, s_n$. For the sequence considered, this is 000111. Read as a dual representation, it corresponds to $000111_2 \hat{=} 7_{10}$. The decoding proceeds in the reverse manner. All but one

**Table 1** All merit factors $F$, energies $E$, number of configurations (#), and decoded configurations with $F \geq 7.0$ (At least 5 largest values are given). Note that every configuration is fourfold degenerate: spin inversion and sequence reversal.

| N | n | F | E | (#) | CONFIGURATION |
|---|---|---|---|---|---|
| 5 | 3 | 6.25000 | 2 | (1) | 0 |
|   |   | 1.25000 | 10 | (1) | 1 |
| 7 | 4 | 8.16667 | 3 | (1) | 1 |
|   |   | 2.22727 | 11 | (2) | 0,2 |
|   |   | 1.28947 | 19 | (0) | |
|   |   | 0.87037 | 27 | (0) | |
|   |   | 0.70000 | 35 | (1) | 3 |
| 9 | 5 | 3.37500 | 12 | (2) | 1,3 |
|   |   | 2.02500 | 20 | (3) | 4,5,7 |
|   |   | 1.44643 | 28 | (2) | 0,2 |
|   |   | 1.12500 | 36 | (0) | |
|   |   | 0.92045 | 44 | (0) | |
| 11 | 6 | 12.10000 | 5 | (1) | 7 |
|   |   | 4.65385 | 13 | (2) | 1,3 |
|   |   | 2.88095 | 21 | (0) | |
|   |   | 2.08621 | 29 | (2) | 2,10 |
|   |   | 1.63514 | 37 | (5) | 4,5,6,14,15 |
| 13 | 7 | 14.08333 | 6 | (1) | 3 |
|   |   | 6.03571 | 14 | (0) | |
|   |   | 3.84091 | 22 | (7) | 4,6,14,16,18,23,31 |
|   |   | 2.81667 | 30 | (2) | 7,20 |
|   |   | 2.22368 | 38 | (2) | 1,11 |
| 15 | 8 | 7.50000 | 15 | (2) | 6,29 |
|   |   | 4.89130 | 23 | (0) | |
|   |   | 3.62903 | 31 | (4) | 3,9,33,41 |
|   |   | 2.88462 | 39 | (6) | 7,13,22,28,37,62 |
|   |   | 2.39362 | 47 | (15) | 4,14,16,18,23,12,24,31, 36,46,53,56,58,61,63 |
| 17 | 9 | 4.51563 | 32 | (1) | 124 |
|   |   | 3.61250 | 40 | (4) | 28,29,82,83 |
|   |   | 3.01042 | 48 | (18) | 6,12,14,16,18,24,36,44, 56,58,66,72,74,93,106, 117,125,127 |
|   |   | 2.58036 | 56 | (3) | 3,9,33 |
|   |   | 2.25781 | 64 | (14) | 13,17,19,27,37,49,57,59, 67,84,86,92,107,116, |
| 19 | 10 | 5.46970 | 33 | (2) | 125,186 |
|   |   | 4.40244 | 41 | (2) | 28,249 |
|   |   | 3.68367 | 49 | (4) | 27,57,124,254 |
|   |   | 3.16667 | 57 | (2) | 129,169 |
|   |   | 2.77692 | 65 | (22) | 6,14,18,24,36,44,56,74, 104,119,132,144,154,166, 192,213,223,226,232,234, 253,247 |
| 21 | 11 | 8.48077 | 26 | (1) | 508 |
|   |   | 6.48529 | 34 | (0) | |
|   |   | 5.25000 | 42 | (4) | 28,54,328,448 |
|   |   | 4.41000 | 50 | (3) | 79,249,269 |
|   |   | 3.80172 | 58 | (4) | 57,372,427,468 |
| 23 | 12 | 5.18627 | 51 | (4) | 28,448,482,1017 |
|   |   | 4.48305 | 59 | (9) | 36,56,132,226,232,503, 528,576,744 |
|   |   | 3.94776 | 67 | (9) | 27,57,297,372,502,508, 769,854,937 |
|   |   | 3.52667 | 75 | (16) | 79,89,109,113,121,124, 197,222,231,249,269, 271,334,452,581,593 |
|   |   | 3.18675 | 83 | (21) | 30,54,60,78,108,125, 247,287,317,328,499,532, 668,733,832,837,849,864, 917,994,1000 |
| 25 | 13 | 6.00962 | 52 | (1) | 113 |
|   |   | 5.20833 | 60 | (2) | 1173,1207 |
|   |   | 4.59559 | 68 | (11) | 503,952,1320,1393,1469, 1489,1792,1813,1834,1855 1928 |
|   |   | 4.11184 | 76 | (12) | 27,49,57,73,193,227,233, 577,745,769,1526,2044 |
|   |   | 3.72024 | 84 | (7) | 109,243,249,269,502,595, 1061 |
| 27 | 14 | 9.85135 | 37 | (1) | 1928 |
|   |   | 8.10000 | 45 | (0) | |
|   |   | 6.87736 | 53 | (1) | 219 |
|   |   | 5.97541 | 61 | (1) | 227 |
|   |   | 5.28261 | 69 | (3) | 1855,3199,3970 |
| 29 | 15 | 6.78226 | 62 | (2) | 3199,4829 |
|   |   | 6.00714 | 70 | (5) | 1270,1809,3198,4247,4318 |
|   |   | 5.39103 | 78 | (3) | 399,4133,4615 |
|   |   | 4.88953 | 86 | (8) | 398,774,5282,5583,5879, 5999,7210,7295 |
|   |   | 4.47340 | 94 | (12) | 356,438,454,462,1808, 1855,2838,4544,4928, 5955,6871,7172 |
| 31 | 16 | 6.08228 | 79 | (1) | 869 |
|   |   | 5.52299 | 87 | (4) | 436,462,5262,6357 |
|   |   | 5.05789 | 95 | (4) | 7714,15813,16249,11172 |
|   |   | 4.66505 | 103 | (7) | 483,1167,1549,6157,9349, 14991,16286 |
|   |   | 4.32883 | 111 | (13) | 463,487,1270,3198,4318, 4724,5518,6356,9310, 11707,11758,12319,14909 |
| 33 | 17 | 6.18750 | 88 | (2) | 484,24636 |
|   |   | 5.67188 | 96 | (1) | 27181 |
|   |   | 5.23558 | 104 | (5) | 16284,16334,16603,19713, 30907 |
|   |   | 4.86161 | 112 | (2) | 1948,17589 |
|   |   | 4.53750 | 120 | (28) | 966,974,2414,3346,4324, 7234,8526,8644,9306, 13416,14406,14574,15813, 15976,16249,16602,17052, 18815,19319,19840,19872, 21213,27316,28576,29568, 29600,32497,32499 |
| 35 | 18 | 6.88202 | 89 | (2) | 7990,16284 |
|   |   | 6.31443 | 97 | (0) | |
|   |   | 5.83333 | 105 | (5) | 2412,42066,42264,49182, 61446 |
|   |   | 5.42035 | 113 | (1) | 1948 |
|   |   | 5.06198 | 121 | (2) | 40884,54554 |
| 37 | 19 | 6.45755 | 106 | (1) | 122892 |
|   |   | 6.00439 | 114 | (0) | |
|   |   | 5.61066 | 122 | (5) | 6989,7990,16284,17372, 86729 |
|   |   | 5.26538 | 130 | (10) | 1819,4825,47693,61321, 66727,74757,75764,101381, 114737,116080 |
|   |   | 4.96014 | 138 | (18) | 1932,2412,3492,9038, 10148,17052,25103,31624, 33732,37962,56224,68124, 79370,79520,92716, 102304,112256,123564 |
| 39 | 20 | 7.68182 | 99 | (2) | 3492,237184 |
|   |   | 7.10748 | 107 | (5) | 3505,101381,159649, 198149,245785 |
|   |   | 6.61304 | 115 | (6) | 7397,6989,7964,7990, 127238,261656 |
|   |   | 6.18293 | 123 | (1) | 71925 |
|   |   | 5.80534 | 131 | (6) | 1932,3609,57317,123564, 230664,247129 |
| 41 | 21 | 7.78241 | 108 | (1) | 247129 |
|   |   | 7.24569 | 116 | (2) | 15929,18864 |
|   |   | 6.77823 | 124 | (0) | |
|   |   | 6.36742 | 132 | (2) | 7964,459541 |
|   |   | 6.00357 | 140 | (9) | 225265,247155,303029, 316031,475925,508183, 508189,508215,517055 |
| 43 | 22 | 8.48165 | 109 | (1) | 316031 |
|   |   | 7.90171 | 117 | (1) | 31859 |
|   |   | 7.39600 | 125 | (0) | |
|   |   | 6.95113 | 133 | (6) | 13851,34606,136377, 598127,638474,900954 |
|   |   | 6.55674 | 141 | (4) | 147948,300911,463032, 508189 |

**Table 1** (*Continued*)

| N | n | F | E | (#) | CONFIGURATION |
|---|---|---|---|---|---|
| 45 | 23 | 8.58051 | 118 | (1) | 27943 |
|  |  | 8.03571 | 126 | (2) | 25540,147948 |
|  |  | 7.55597 | 134 | (0) |  |
|  |  | 7.13028 | 142 | (1) | 31859 |
|  |  | 6.75000 | 150 | (0) |  |
| 47 | 24 | 8.18148 | 135 | (5) | 25540,786918,2089933,<br>2621019,3677248 |
|  |  | 7.72378 | 143 | (1) | 27943 |
|  |  | 7.31457 | 151 | (1) | 459581 |
|  |  | 6.94654 | 159 | (3) | 61926,96719,3730132 |
|  |  | 6.61377 | 167 | (19) | 55886,158750,509448,<br>542270,788376,1508989,<br>1524589,1849996,1865019,<br>1953905,2041599,2378398,<br>2470152,2483208,2616360,<br>3686114,3735074,3795799,<br>4190694 |
| 49 | 25 | 8.82721 | 136 | (1) | 3735075 |
|  |  | 8.33681 | 144 | (1) | 459581 |
|  |  | 7.89803 | 152 | (0) |  |
|  |  | 7.50313 | 160 | (2) | 6099819,7537141 |
|  |  | 7.14583 | 168 | (6) | 2378398,3686114,3797340,<br>5722745,8158135,8384755 |
| 51 | 26 | 8.50000 | 153 | (1) | 14797559 |
|  |  | 8.07764 | 161 | (1) | 6493451 |
|  |  | 7.69527 | 169 | (1) | 3735075 |
|  |  | 7.34746 | 177 | (3) | 406000,7569877,16360319 |
|  |  | 7.02973 | 185 | (1) | 12981437 |
| 53 | 27 | 8.26176 | 170 | (1) | 25169676 |
|  |  | 7.89045 | 178 | (2) | 4165170,5195874 |
|  |  | 7.55108 | 186 | (0) |  |
|  |  | 7.23969 | 194 | (0) |  |
|  |  | 6.95297 | 202 | (4) | 4169526,6493451,15694328,<br>31510518 |
| 55 | 28 | 8.84503 | 171 | (2) | 448730,25952500 |
|  |  | 8.44972 | 179 | (1) | 448731 |
|  |  | 8.08824 | 187 | (0) |  |
|  |  | 7.75641 | 195 | (0) |  |
|  |  | 7.45074 | 203 | (4) | 2084717,8680018,16697760,<br>25973920 |
|  |  | 7.16825 | 211 | (0) |  |
| 57 | 29 | 8.64096 | 188 | (1) | 60619860 |
|  |  | 8.28827 | 196 | (0) |  |
|  |  | 7.96324 | 204 | (0) |  |
|  |  | 7.66274 | 212 | (1) | 86151572 |
|  |  | 7.38409 | 220 | (5) | 448730,14653044,16697760,<br>16697761,98373872 |
|  |  | 7.12500 | 228 | (0) |  |
| 59 | 32 | 8.49024 | 205 | (1) | 8338846 |
|  |  | 8.17136 | 213 | (0) |  |
|  |  | 7.87557 | 221 | (0) |  |
|  |  | 7.60043 | 229 | (0) |  |
|  |  | 7.34388 | 237 | (6) | 14653044,16697761,<br>24988966,166543342,<br>242482010,260526397 |
|  |  | 7.10408 | 245 | (2) | 42027443,127860936 |
| 61 | 31 | 8.08913 | 230 | (1) | 313712297 |
|  |  | 7.81723 | 238 | (0) |  |
|  |  | 7.56301 | 246 | (1) | 118779405 |
|  |  | 7.32480 | 254 | (2) | 6549711,109036492 |
|  |  | 7.10115 | 262 | (6) | 208141855,302148871,<br>468850391,486304766,<br>513981540,520476434 |

| N | n | F | E | (#) | CONFIGURATION |
|---|---|---|---|---|---|
| 63 | 32 | 7.32288 | 271 | (3) | 2976147,663838027,<br>982181639 |
|  |  | 7.11290 | 279 | (2) | 218096102,634672048 |
|  |  | 6.91463 | 287 | (2) | 613505756,767476903 |
|  |  | 6.72712 | 295 | (4) | 276172021,308061437,<br>520499268,810014746 |
|  |  | 6.54950 | 303 | (27) | 34765969,69142962,<br>96280776,117963975,<br>226499390,284524834,<br>284664200,302148871,<br>386804624,473713318,<br>476294017,556630105,<br>612979105,615968385,<br>712216491,718844101,<br>732532088,781920348,<br>849718969,936401182,<br>941126659,946904575,<br>1012684206,1031805006,<br>1067443734 |
| 65 | 33 | 7.76654 | 272 | (4) | 284524834,1523793913,<br>1792713494,1954075916 |
|  |  | 7.54464 | 280 | (0) |  |
|  |  | 7.33507 | 288 | (0) |  |
|  |  | 7.13682 | 296 | (6) | 162662151,520499268,<br>532877671,712216491,<br>1044374028,1246913808 |
|  |  | 6.94901 | 304 | (13) | 7879878,8067312,<br>473815991,521141651,<br>958169409,983814374,<br>1067443734,1420587485,<br>1421178057,1439109814,<br>1879280719,1887504101,<br>2078226701 |
| 67 | 34 | 9.31328 | 241 | (3) | 8067312,234579911,<br>4034499923 |
|  |  | 9.01406 | 249 | (0) |  |
|  |  | 8.73346 | 257 | (0) |  |
|  |  | 8.46981 | 265 | (0) |  |
|  |  | 8.22161 | 273 | (1) | 3644817931 |
|  |  | 7.98754 | 281 | (0) |  |
|  |  | 7.76644 | 289 | (0) |  |
|  |  | 7.55724 | 297 | (1) | 1891678487 |
|  |  | 7.35902 | 305 | (8) | 276309709,280078184,<br>418459655,477232198,<br>1077426072,2959643923,<br>3061538542,3941640606 |
|  |  | 7.17093 | 313 | (2) | 133811995,1065652422 |
| 69 | 35 | 8.44149 | 282 | (1) | 5724065189 |
|  |  | 8.20862 | 290 | (0) |  |
|  |  | 7.98823 | 298 | (0) |  |
|  |  | 7.77941 | 306 | (3) | 1891678487,2216934712,<br>5782778514 |
|  |  | 7.58121 | 314 | (4) | 1065652422,2233740913,<br>6221308582,6445661715 |
|  |  | 7.39286 | 322 | (0) |  |
|  |  | 7.21364 | 330 | (0) |  |
|  |  | 7.04290 | 338 | (7) | 1077426072,1152810842,<br>1885379349,5730772784,<br>6652100897,7859333228,<br>8068999846 |
| 71 | 36 | 9.16545 | 275 | (1) | 16516843532 |
|  |  | 8.90636 | 283 | (0) |  |
|  |  | 8.66151 | 291 | (0) |  |
|  |  | 8.42977 | 299 | (0) |  |
|  |  | 8.21010 | 307 | (0) |  |
|  |  | 8.00159 | 315 | (2) | 6581002042,17167158682 |
|  |  | 7.80341 | 323 | (0) |  |
|  |  | 7.61480 | 331 | (1) | 11719464641 |
|  |  | 7.43510 | 339 | (2) | 1072143793,8056782919 |
|  |  | 7.26369 | 347 | (7) | 1088807513,1885379349,<br>2752816076,3816846617,<br>8378187679,9779016863,<br>12254319549 |
|  |  | 7.10000 | 355 | (1) | 1243182895 |
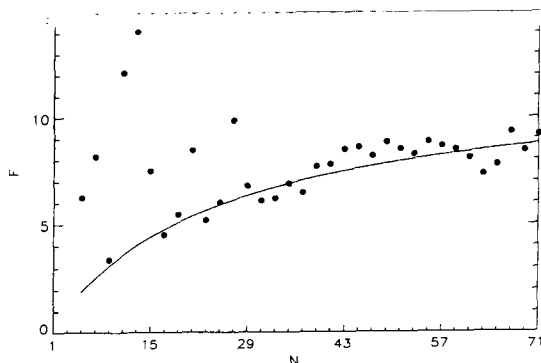
**Figure 1a**  Exact results for the optimal merit factor $F$ (dots) vs. chain length $N$. The solid line is the approximation given by eq. (5).

$(N = 61)$ of the maximum merit factors found had been already obtained by Beenker *et al.* [4] using their heuristic search algorithms. As this work was being done, we received a draft of Golay and Harris [16] from Bernasconi, where they evaluated the maximum merit factors for $N \leq 69$ in agreement with our results. Figure 1 summarizes our results for skew-symmetric sequences, showing the optimal merit factor as a function of chain length $N$. The exact results in Figure 1a are also compared with the approximate value

$$F^{\text{opt}} \sim \frac{12.3248}{(\pi^3 N^3/4)^{1/N} + 12.3248/N} \xrightarrow[N \to \infty]{} 12.3248 \tag{5}$$

derived by Golay [3] based on the so-called *ergodicity postulate*. This postulate supposes that the autocorrelation values $s_i s_{i+k}$ of the binary sequence are equally likely. This means that the above equation was derived under the assumption that the $R_k$ are random variables. The gap widths are given in Fig. 1b and compared with the approximate expression

$$\Delta F^{\text{opt}} \sim \frac{(F^{\text{opt}})^2}{F^{\text{opt}} + (N/4)^2}, \tag{6}$$
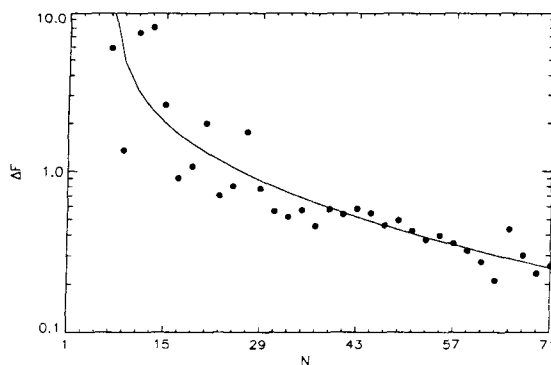


**Figure 1b**  Exact results for the gap widths $\Delta F$ (dots) compared with the approximate expression eq. (6).
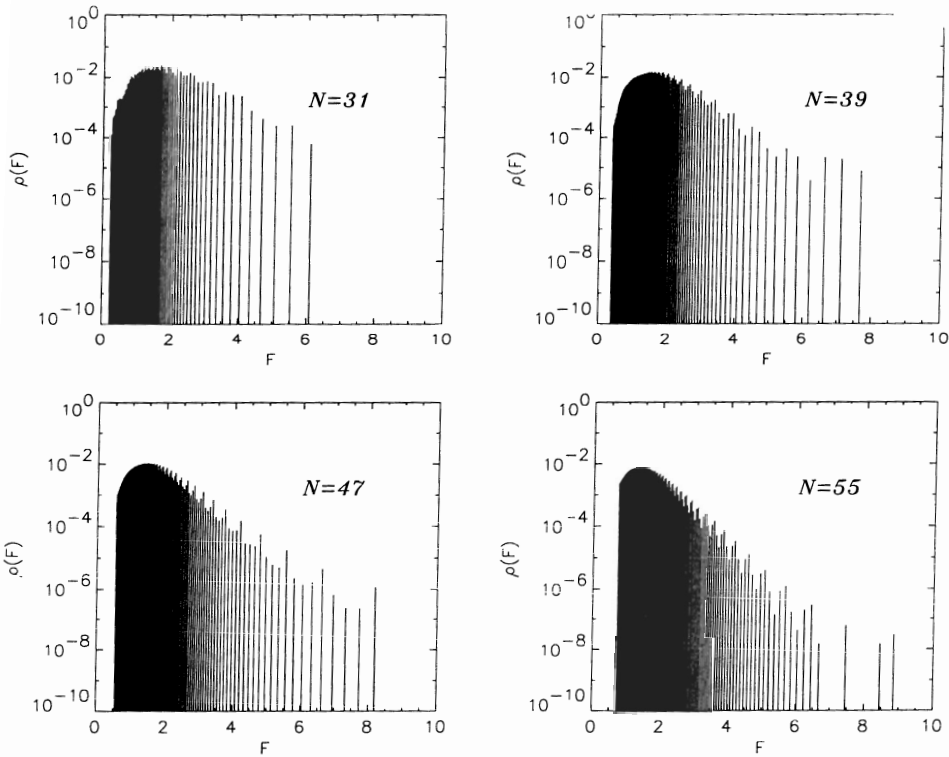
**Figure 2**   Normalized distribution of the merit factors $\rho(F)$ of skew-symmetric sequences for lengths $N = 31$, 39, 47 and 55. Note that the histograms are truncated for small values of $F$.

where we assumed that the energy gap is $\Delta E = 8$. Both figures indicate that the ergodicity postulate leads to a reasonable approximation in the limit of infinitely long chains.

We have evaluated all energy values for chain lengths $N \leq 55$, giving us the distribution functions for the energies $\rho(E)$, as well for the merit factors $\rho(F)$. As examples, the results for $N = 31$, 39, 47 and 55 are shown as histograms in Figure 2. Here we have normalized $\rho(F)$ by the number of possible configurations $2^{(N+1)/2}$. The histograms demonstrate again in an impressive way the large gaps for high merit factors $F$.

We have also investigated the thermodynamic properties of the associated spin chain with four spin interactions. If we consider $E(S)$ given by equation (3) as an energy function, the optimization problem is equivalent to finding low-energy configurations of the corresponding physical system. The distribution of the configurations is determined in statistical mechanics via the Boltzmann distribution

$$P_T(S) = \frac{1}{Z} e^{-E(S)/T}, \tag{7}$$

where $Z = \sum_{\bar{S}} \exp[-E(S)/T]$ is the partition function at given temperature $T$. From $Z$ we can calculate thermodynamic averages, such as the energy and specific

heat. Based on Golay's ergodicity postulate, Bernasconi [5, 6] has evaluated the partition function, which is given in the case of skew-symmetric sequences by

$$Z = 2^{(N+1)/2} \prod_{l=1}^{(N-1)/2} \frac{1}{(1 + 4l/T)^{1/2}}. \tag{8}$$

In the limit of infinite chain length, this yields

$$\frac{4}{N^2} \langle E \rangle \sim \frac{1}{2\bar{\beta}^2} (2\bar{\beta} - \ln(1 + 2\bar{\beta})) \tag{9}$$

for the energy and

$$\frac{4}{N} C \sim 1 + \frac{1}{1 + 2\bar{\beta}} - \frac{1}{\bar{\beta}} \ln(1 + 2\bar{\beta}) \tag{10}$$

for the specific heat, where $\bar{\beta} = 2N/T$. These predictions of the ergodicity postulate are compared with the exact results in Figure 3 and Figure 4 for chains of length $N = 31$, 39, 47, and 55. There is a clear discrepancy between the approximate analytical results given by equations (9, 10) and the exact data points. The occurence of this *ergodicity breaking* was also observed by Bernasconi [5] in a comparison with data obtained from simulated annealing.

## 3. OPTIMIZATION BY EVOLUTIONARY STRATEGIES

Beenker *et al.* [4] and Bernasconi [5, 6] have done much work on the optimization of binary sequences with simulated annealing programs. Recently, Wang [11] claimed that one should get better results if one optimizes binary sequences by evolutionary strategies based on the analogy between mathematical optimization and molecular evolution. He tested his hypothesis on skew-symmetric chains with lengths $N = 101$, 103 and 105, and found better results than Beenker *et al.*: 6.911, 7.766 and 7.614 instead of 6.058, 5.900 and 6.071, respectively. Indeed, Bernasconi recently found [6] a skew-symmetric sequence of chain length $N = 103$ with a merit factor $F = 9.56$ by simulated annealing.
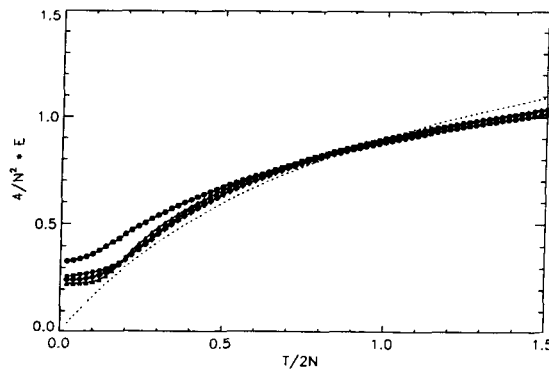


**Figure 3** Normalized averaged energy $4/N^2\langle E \rangle$ vs. normalized temperature $T/2N$, obtained from exact enumeration of finite skew-symmetric sequences of lengths $N = 31$ (dots), 39 (squares), 47 (diamonds) and 55 (triangles). The dashed curve is the approximate result given by eq. (9).
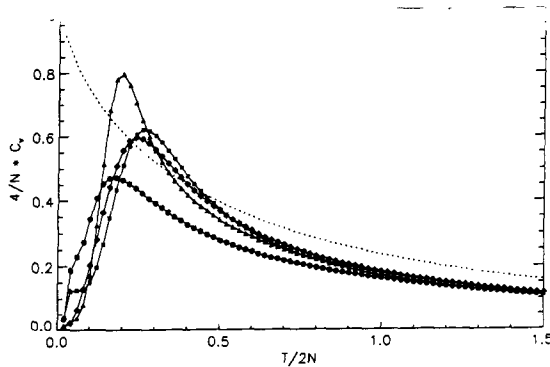
**Figure 4**  Normalized specific heat $4C/N$ vs. normalized temperature $T/2N$, obtained from exact enumeration of finite skew-symmetric sequences of lengths $N = 31$ (dots), 39 (squares), 47 (diamonds) and 55 (triangles). The dashed curve is the approximate result given by eq. (10).

We have investigated the optimization of binary sequences by evolutionary strategies in much more detail than Wang to see how far this approach can be taken, also comparing it to simulated annealing. First, we present some basic ideas of evolutionary algorithms. Since evolutionary methods are much more diverse than simulated annealing algorithms, we report in detail on our implementation of the algorithm. Finally, we present results obtained for low-autocorrelation binary sequences.

An optimization process making use of evolutionary strategies consists of the following elements [11–13]:

- Start from an arbitrary number of configurations (templates) in the phase space. Then generate a given number of new configurations (mutants) according to a given scheme (mutation dynamics).

- Take out a specified number of configurations with the best values (survivor set) according to a given decision rule (selection).

- If the survivor set does not change during consecutive iterations, stop the evolution process. Otherwise use the mutants in the survivor set as new templates and continue the iteration.

To be more specific, we give a short description of our evolutionary strategy for the optimization of binary sequences. We start with a given population size of templates (PSZ) which have a mutual minimum Hamming distance (DMIN). For each template in the population we generate a given number of new mutants (SW). So after one generation the population size is increased and consists of PSZ(SW + 1) members. We sort the increased population according to their merit factor and choose the PSZ sequences with the highest merit factors as the survivor set. However, in this selection process we take care that the members or vectors have at least Hamming distance DMIN to guarantee diversity in the population.

How do we generate the mutants? There are a great number of possibilities. Our generation scheme is characterized by the following rules. The starting point is kept unchanged. The program tries to create an optimal mutant with Hamming

distance one. To achieve this we generate a given number of random indices (STT), which give us different directions for trials in the first mutation step. This means that we go exactly one step on the Boolean hypercube in these directions and interchange the sign at this endpoint. We evaluate the cost function for this new vector. We select the one which gives the best improvement in energy from these STT "1-error-mutants". Selection means that this vector becomes a member of the new increased population. Now we try to find an optimal "2-error-mutant" by adding a new error to the "1-error-mutant". The process for finding this mutant is the same as in the previous step, except that the index which generated the change on the chain is now excluded from the random index generation. In other words, we do not go back on the Boolean hypercube. The generation of new mutants for the initially chosen starting vector is finished when we reach the maximum step size SW. The generation process is continued by taking the next template from the initial population. At the end, the increased population will be reduced to its original size as already mentioned above.

We now describe how our algorithm detects local minima. The program checks each new mutant against its template. If more than a given number represent no improvement the program knows that it is effectively stuck in a local minimum. A special routine is then used to move the template out of its local minimum. The procedure for leaving a local minimum is very similar to that for the generation of mutants. Instead of the generation of SW-error-mutants, we generate a larger set of (SW + SW)-error-mutants. We select one at random from the set of the best (SW + 1)- to (SW + SW)-error-mutants, and replace the template with this mutant. This guarantees that the new template does not return to the old minimum in the next step. The program stops when it exceeds a given number of local minima (LMNUM) for the best vector of the generation. The best value of the cost function ever reached during the simulation, and the configuration belonging to it, are the final results of the optimization process.

In a first step, we have compared the results from optimization with evolutionary strategies to those obtained by other heuristic methods. Figure 5
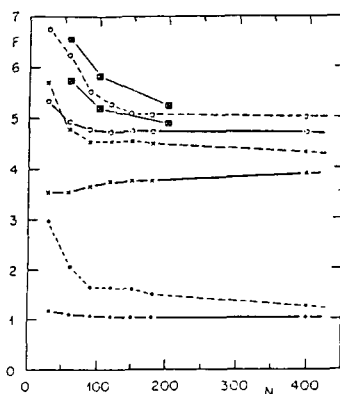


**Figure 5** Merit factors $F$ for some typical chain lengths, obtained from different algorithms; dots: random sequences, crosses: greedy algorithm, circles: simulated annealing (from ref. [5]), squares: optimization by our evolutionary strategy; upper curve: best, lower curve: average value out of 200 optimizations.
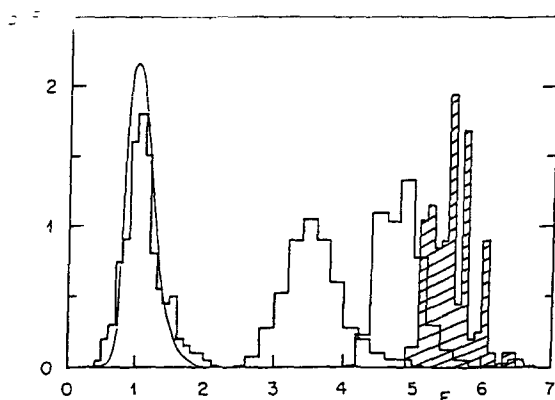
**Figure 6** Distribution of merit factors $\rho(F)$ for 200 generated sequences of length $N = 61$ obtained from different heuristic search algorithms. From left to right: random sequences, greedy algorithm, simulated annealing (from ref. [5]), and optimization by our evolutionary strategy (shaded).

shows the merit factors $F$ for some typical chain lengths up to $N = 201$. The average value and the best merit factors obtained from 200 optimizations are presented for four different heuristic algorithms: (i) random sequences, (ii) greedy algorithm, (iii) simulated annealing, and (iv) optimization by evolutionary strategies. The first three curves are taken from the work of Bernasconi [5]. Figure 5 indicates that our evolutionary algorithm gives merit factors 5% to 10% better than those obtained with simulated annealing in [5]. The distributions of these 200 sequences are given in Figure 6 showing left to right the (i) random, (ii) quenched, (iii) annealed, and (iv) evolutionary distribution functions. The sequences for the greedy algorithm are called quenched, since it is identical to simulated annealing at zero temperature. Again, the plot represents an improvement. Bernasconi has also investigated the dependence of the merit factor on the cooling rate in a simulated annealing process. For $N = 49$, he found that the energy of the annealed sequences decreases logarithmically with the cooling rate, which seems to be a characteristic feature of NP-complete problems [14, 15]. The curve in Figure 7 extrapolates to $F^{opt} \sim 5.3$, whereas the true optimal merit factor is at least 8.83 in the case of skew-symmetric binary sequences of length $N = 49$ [5]. The population size for evolutionary strategies seems to correspond to the cooling rate in the simulated annealing. An infinitely large population will find the optimal configuration with a high probability. We find also a logarithmic dependence of the merit factor on population size, which extrapolates to a maximum value of $F^{opt} \sim 7.7$. This value is still far away from the optimal $F$, however, better than that obtained by simulated annealing. These results are shown in Figure 7. We have also investigated the quality of the results as a function of CPU time. Instead of stopping the program after a given number of local minima (LMNUM), we stopped after a given CPU time limit. In Figure 8 the merit factor is plotted versus CPU time (IBM 3083 mainframe) for the best and averaged values of 50 optimizations. The *average quality* of the results increases logarithmically with CPU time.

We have also optimized skew-symmetric binary sequences with our evolutionary strategy. First, we have compared the optimization of the full configuration
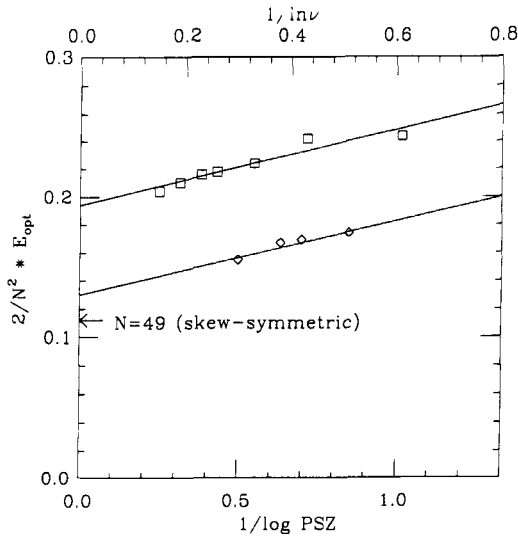
**Figure 7** Comparison of the normalized optimal energy $2/N^2\langle E\rangle$ as a function of the cooling rate $v$ for simulated annealing (squares) (from ref. [5]) and as function of the population size (diamonds) in the case of optimization by our evolutionary strategy for a chain of length $N = 49$. The arrow indicates the optimal value of the skew-symmetric binary sequence.

space with that of the skew-symmetric for $N = 61$. The results in Figure 9 show that in the skew-symmetric phase space all large merit factors were found, in contrast to the full phase space, where we achieved only merit factors of $F \leq 6.5$. As another special case, we have considered $N = 101$. The dependence of the inverse merit factors $(2E/N^2)$ as function of population size is shown in Figure 10, extrapolating to an optimal merit factor of $F \sim 9.0 \pm 0.7$. The best merit factor we detected was 8.36, much improved on than the results of Wang (7.85) and Beenker *et al.* (6.06).
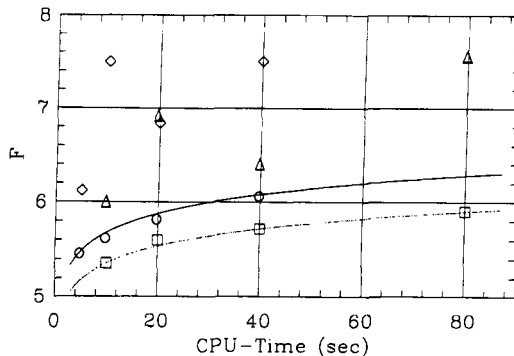


**Figure 8** Merit factor $F$ vs. CPU time for the best and averaged values out of 50 optimization runs, diamond: $N = 49$, triangle: $N = 61$, best values, circle: $N = 49$, square: $N = 61$, averaged values, respectively. The lines give a logarithmic least square fit to the averaged values.
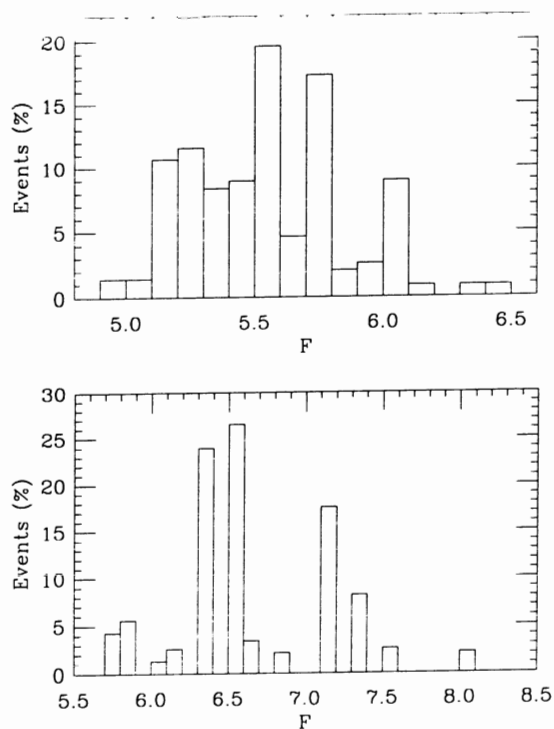
**Figure 9** Distribution of the best merit factors for chains of length $N = 61$ obtained from 200 runs of optimization with our evolutionary strategy. Top: full configuration space, bottom: skew-symmetric configuration space.
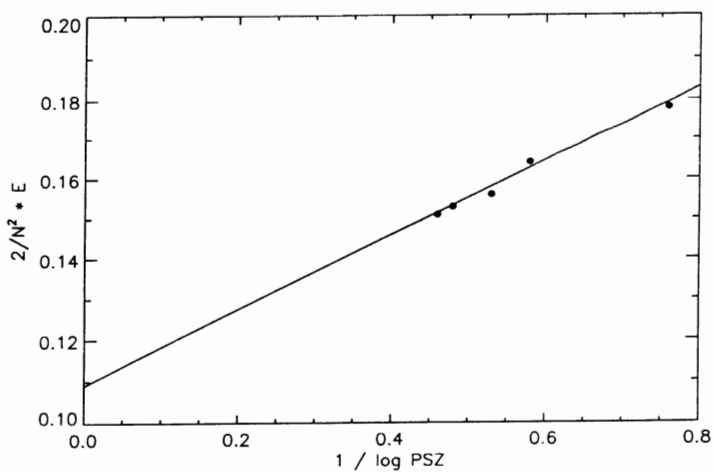


**Figure 10** Dependence of the inverse merit factor $F^{-1} = 2E/N^2$ as function of the population size for a skew-symmetric sequence of length $N = 101$.

**Table 2** Merit factors obtained from different search algorithms for long chains with $73 \le N \le 201$. The merit factor 9.56 marked by a star was also found by Bernasconi [9].

| N | Beenker | Our work | Golay | Wang |
|---|---------|----------|-------|------|
| 73 | 7.485 | | 7.66 | |
| 75 | 9.248 | | 8.25 | |
| 77 | 8.100 | | 8.28 | |
| 79 | 7.377 | | 7.67 | |
| 81 | 7.323 | 8.04 | 8.20 | |
| 83 | 7.811 | | 9.14 | |
| 85 | 7.028 | | 8.17 | |
| 87 | 7.464 | | 8.39 | |
| 89 | 7.558 | | 8.18 | |
| 91 | 7.127 | | 8.68 | |
| 93 | 7.232 | | 8.61 | |
| 95 | 7.151 | | 9.42 | |
| 97 | 7.351 | | 8.78 | |
| 99 | 7.282 | | 8.38 | |
| 101 | 6.058 | 8.36 | 8.36 | 6.911 |
| 103 | 5.900 | | 9.56* | 7.766 |
| 105 | 6.071 | | 8.89 | 7.614 |
| 107 | 6.527 | | 8.46 | |
| 109 | 6.150 | | 8.97 | |
| 111 | 6.022 | | 8.97 | |
| 113 | 6.334 | | 8.49 | |
| 115 | 6.401 | | | |
| 117 | 6.421 | | | |
| 119 | 6.006 | | | |
| 121 | 6.607 | | | |
| 123 | 6.538 | | | |
| 125 | 6.655 | | | |
| 127 | 6.509 | | | |
| 129 | 6.667 | | | |
| 131 | 6.055 | | | |
| 133 | 5.936 | | | |
| 135 | 6.458 | | | |

| N | Beenker | Our work |
|---|---------|----------|
| 137 | 6.428 | |
| 139 | 6.110 | |
| 141 | 6.010 | 6.48 |
| 143 | 5.920 | |
| 145 | 6.112 | |
| 147 | 6.053 | |
| 149 | 6.215 | |
| 151 | 5.953 | |
| 153 | 6.008 | |
| 155 | 6.039 | |
| 157 | 6.507 | |
| 159 | 5.877 | |
| 161 | 6.023 | |
| 163 | 5.783 | |
| 165 | 5.574 | |
| 167 | 5.921 | |
| 169 | 6.041 | |
| 171 | 6.310 | |
| 173 | 6.220 | |
| 175 | 6.021 | |
| 177 | 6.025 | |
| 179 | 5.958 | |
| 181 | 5.700 | 5.81 |
| 183 | 5.591 | |
| 185 | 5.852 | |
| 187 | 5.712 | |
| 189 | 5.864 | |
| 191 | 5.674 | |
| 193 | 6.078 | |
| 195 | 6.000 | |
| 197 | 5.758 | |
| 199 | 5.930 | |

Table 2 summarizes the merit factors we found with our evolutionary algorithm for chain lengths $81 \le N \le 201$ and compares them with results obtained by other heuristic algorithms. One observes that the results of Beenker *et al.* are lower than ours. The table also contains results from a heuristic recently used by Golay and Harris [16]. They gave a recipe for constructing a search algorithm, based on the observation that the individual merit factors of the symmetric $(A)$ and antisymmetric sequences $(B)$, mentioned in the previous section, are themselves unusually large for the complete skew-symmetric sequence. Golay's results based on this observation for sequence lengths $73 \le N \le 113$ are as good as our results obtained from optimization by evolutionary strategies.

## 4. SUMMARY

We have investigated the problem of finding high quality autocorrelation binary sequences by exact enumeration and by an heuristic optimization algorithm based on evolutionary strategies. Exact results for the ground state and thermodynamic properties were compared with a mean field like approach based on Golay's ergodicity postulate, and with other heuristic search algorithms. The results obtained from our stochastic search algorithm are similar to those of other authors using simulated annealing. It seems that our optimization method with evolutionary search strategies is superior to other heuristic algorithms used so far for long chains. This result must, however, be stated with some care, since we found that the results obtained from stochastic methods depend on the effort one undertakes for the implementation of the algorithms. For simulated annealing, for example, one must find an optimal cooling schedule, or, in the case of optimization with evolutionary strategies, an optimal set of parameters. We have also evaluated ground states of the two dimensional $\pm J$ and the Gaussian spin glass model [17] coming to the opposite conclusion. Our simulated annealing results were slightly better than those with our evolutionary algorithm.

### Acknowledgement

### References

[1] Golay, M. J. E. (1972) A Class of Finite Binary Sequences with Alternate Autocorrelation Values Equal to Zero, *IEEE Trans. on Inform. Theory* **IT-18**, 449
[2] Golay, M. J. E. (1977) Sieves for Low Autocorrelation Binary Sequences, *IEEE Trans. on Inform. Theory* **IT-23**, 43
[3] Turyn, R. (1982) (unpublished results) given in: M. J. E. Golay, The Merit Factor of Long Low Autocorrelation Binary Sequences, *IEEE Trans. on Inform. Theory* **IT-28**, 543
[4] 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. of Research* **40**, 289
[5] Bernasconi, J. (1987) Low Autocorrelation Binary Sequences: Statistical Mechanics and Configuration Space Analysis, *J. Physique* **48**, 559
[6] Bernasconi, J. (1988) Optimization Problems and Statistical Mechanics, in: Proceedings of the Workshop on *Chaos and Complexity,* Torino, World Scientific, Singapore.
[7] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953) Equation of State Calculations by Fast Computing Machines, *J. Chem. Phys.* **21**, 1087
[8] Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983) Optimization by Simulated Annealing, *Science* **220**, 671
[9] Morgenstern, I. and Würtz, D. (1987) Simulated Annealing for 'Spin-Glass-Like' Optimization Problems, *Z. für Physik* **B67**, 397
[10] Moore, M. A. (1987) Zero-Temperature Scaling and Combinatorial Optimization, NSF-ITP preprint 87-05i
[11] Wang, Q. (1987) Optimization by Simulating Molecular Evolution, *Biol. Cybern.* **57**, 95
[12] Rechenberg, I. (1973) Evolutionsstrategie, Frommann-Holzboog, Stuttgart
[13] Schwefel, H.-P. (1977) Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie, Birkhäuser, Basel

4 Huse. D A and Fischer, D S (1986) Residual Energies After Slow Cooling of Disordered Systems, *Phys. Rev. Lett.* **57,** 2203

[15] Grest, G. S., Soukoulis, C. M. and Levin, K. (1986) Cooling-Rate Dependence for the Spin-Glass Ground-State Energy: Implications for Optimization by Simulated Annealing, *Phys. Rev. Lett.* **56,** 1148

[16] Golay, M. J. E. and Harris, D. *A New Search for Skewsymmetric Binary Sequences with Optimal Merit Factors,* preprint 1989

[17] van Hemmen, J. L. and Morgenstern, I. (ed.) (1983) Heidelberg Colloquium on Spin Glasses, *Lecture Notes in Physics* **192,** Springer, Berlin