

Evolutionary Search for Low Autocorrelated Binary Sequences

Burkhard Militzer[†], Michele Zamparelli[‡] and Dieter Beule[‡]

Abstract— The search for low autocorrelated binary sequences is a classical example of a discrete frustrated optimization problem. We demonstrate the efficiency of a class of evolutionary algorithms to tackle the problem. A suitable mutation operator using a preselection scheme is constructed and the optimal parameters of the strategy are determined.

Keywords— Evolution Strategies, Frustrated Problem, Low Autocorrelated Binary Sequences

I. INTRODUCTION

THE interest in binary sequences with low autocorrelation emerged initially from several communication engineering problems in radar applications [1], [2], [3]. Several authors have provided empirical evidence of the complexity of the optimization task, essentially based on the low quality of the results obtained with diverse optimization strategies [2], [4], [5], [7]. Furthermore, ultrametricity analysis and a comparison of these results with the few analytical results available confirmed that the search for the *least autocorrelated binary sequence* (LABS) resembles the search for a needle in the haystack. LABS also provides a nice example of a frustrated problem. For this class of problems, for example the TSP or spin-glass systems, several contradictory requirements need to be optimized. In fact the search for LABS is equivalent to the search for the ground state of an Ising-spin system with long-range 4-spin interaction [2]. For low temperatures the escape time from local extrema grows dramatically and the system shows a spin-glass behavior, as explained in [8]. Although the quick improvement of modern computers has lately allowed an exhaustive search for relatively high dimensions [7], [9], recent efforts have concentrated on still higher values thus enlarging the search space to more than 10^{30} , for which an exhaustive search is impractical.

II. LABS

The optimization problem can be introduced as follows: let us consider a binary sequence of length L

$$s = \{s_1, s_2, \dots, s_L\}, \quad s_i = \pm 1. \quad (1)$$

[†]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. E-mail: militzer@ncsa.uiuc.edu

[‡]Institut für Physik, Humboldt-Universität zu Berlin, Invalidenstraße 110, D-10115 Berlin, Germany. E-mail: zamparelli@physik.hu-berlin.de, beule@physik.hu-berlin.de

The correlation of symbols with distance k is measured by the autocorrelation function

$$R_k = \sum_{i=1}^{L-k} s_i s_{i+k}. \quad (2)$$

The quadratic sum of all autocorrelation functions

$$E = \sum_{k=1}^{L-1} R_k^2 \quad (3)$$

is called energy because of the relation to an Ising-spin system with long-range 4-spin interaction (cf. [2]). The aim of finding sequences with low autocorrelation can then be formulated in two different ways: one either searches for low values of $\max\{|R_k|\}$ or maximizes the so called *merit factor*

$$F = \frac{L^2}{2E}. \quad (4)$$

The merit factor, which will be used as the fitness function throughout this work, was introduced by Golay [4] as a qualitative measure for a sequence on the grounds that it lends itself to better analytical treatment. It is also closely related to the signal-to-noise ratio used in signal processing applications [1].

All sequences can be divided into equivalence classes, if one considers that the operations of inversion ($s_i := -s_i$) and reversal ($s_i := s_{L-i}$) leave the R_k , and therefore F , unchanged. For odd L , only the particular subclass of the *skew-symmetric* sequences is usually considered. Fulfilling

$$s_{n+i} = (-1)^i s_{n-i}, \quad n = (L+1)/2 \quad (5)$$

for $i = 1, \dots, n-1$, these sequences have the property that $R_k = 0$ for all odd k . Since in the skew-symmetric case the right half of the sequence is determined by the left half, the problem of finding the best skew-symmetric sequence for length L is only as tedious as finding the general LABS for $(L+1)/2$. For a given L , the energies E of any pair of sequences differ by multiples of 4. For two skew-symmetric sequences, the energy difference is a multiple of 8. This leads to a further discretization of the merit factor F . Though most authors have concentrated on odd lengths using the skew-symmetric sequences as a sieve, there is no guarantee that the best sequences for odd L are skew-symmetric. Known exceptions are, to date, $L = 19, 23, 25, 31, 33, 35$, and 37 (cf. [7]). The advantage of restricting the search to skew-symmetric sequences can be seen by comparing the density of states for skew-symmetric and normal sequences (Figure 1). For large fitness, the number

of local maxima is decreasing exponentially, and numerical tests discussed in section IV show that maxima are well separated, making the search landscape similar to a golf course with its isolated holes as optima.

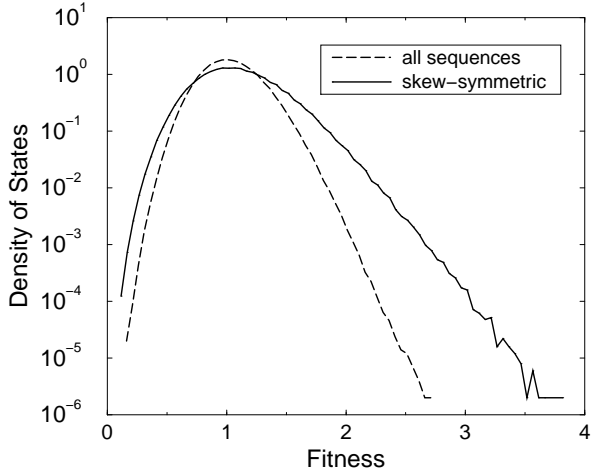


Fig. 1. Comparison of the density of states for the skew-symmetric (dashed line) and the complete search space (solid line), sampled with 10^7 randomly generated sequences for $L = 101$.

Because each autocorrelation function R_k contributes quadratically to the energy E (3), a single large R_k can reduce the fitness of a sequence drastically. The R_k are not independent and therefore each change in the sequence leading to an improvement of a certain R_k will modify the other R_k as well. This is the feature of what is normally called a *frustrated* problem, characterized by a rough landscape where local maxima are many, steep, and narrow.

The fitness fluctuations in a Hamming-one neighborhood around a local maximum are shown in the right graph of Figure 2. The left graph shows the fitness of sequences obtained by inversion of up to all bits, starting from the locally optimal one. The fitness decreases rapidly with the Hamming distance from the maximum, hinting at the narrowness of the local optimum.

The LABS problem gained international scientific recognition thanks to the works of Golay [1], [4], [10], [11], where the assumption that the R_k are normally distributed (the so called *ergodicity hypothesis*) brought about an analytical approximation for the partition sum, thus leading to an estimate for the probability distribution of F and the largest possible F for given L . The probability distribution decays exponentially for large F hinting again at the difficulty of the problem [12]. This approach was later improved in [2], where the analytical predictions and the results from simulated annealing were compared, and doubts were cast about the possibility of reaching fitness values substantially higher than 6 for $L \geq 100$ with any stochastic iterative improvement method. In [9], the statistical analysis of Golay was applied to the skew-symmetric subspace in order to compare the theoretical estimates with the data from an exhaustive search for $L = 55$. The authors concluded that the ergodicity hypothesis provides a reasonable estimate for high F .

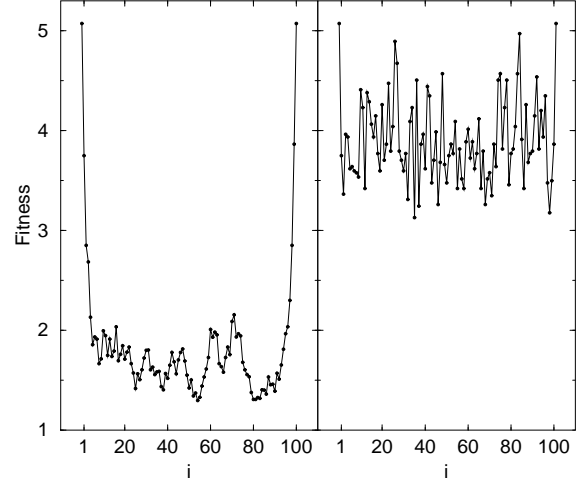


Fig. 2. Structure of the landscape. Fitness of sequences generated by inversion of a given locally maximal sequence up to position i (left graph) for $L = 100$. Fitness of all sequences in the neighborhood of the same sequence, generated by changing only bit i (right graph).

A serious attempt to apply evolutionary optimization techniques to LABS was made in [6] and in [7] in particular, where an advanced kind of parallel genetic algorithm was employed which made heavy use of heuristics as well as especially suited mutation operators and local hill-climbing. Recently the LABS problem was also used as a test bench for a new kind of optimization strategy (*democratic optimization* [5]) and a new trial with simulated annealing was made for $L = 100$ in [13].

III. STRATEGY

A. Evolution Strategies

Evolutionary algorithms (EA) have been successfully applied to the optimization of various frustrated, discrete problems [14], [15], [16]. Their independence from the analytical features of the fitness function makes them suitable for *black box* problems, in which no knowledge is available regarding the fitness production mechanism.

Among the diversity of EAs, we will concentrate on the so-called *evolution strategies* (ES) (μ, λ) and $(\mu + \lambda)$ [15]. Starting from a parent population of μ individuals, λ children are created by mutation. Each child is generated by mutating one parent that is chosen at random out of the parent population. After that a selection process follows. In case of the *comma* strategy, the μ children with highest fitness form the next generation. In the case of the *plus* strategy, the μ best individuals out of the set of the μ parents and the λ children form the next generation. In the case of the *plus* strategy, the best individual is always taken into the next generation and therefore the fitness of the best individual is non-decreasing.

Given the large number of local maxima, *comma* strategies seem to be more appropriate to the LABS problem [17]. There is no general rule for setting the optimal values for μ and λ . For the LABS problem, we will give estimates in section III-C.

Because of the long range correlations and the quadratic structure of the fitness function, we decided not to use any recombination operator to generate an offspring out of parts of fit sequences. Only a complex and computationally expensive procedure might prevent changes in long correlations to outweigh gains with short ones.

The absence of any recombination strategy makes our ES resemble another class of EA, *evolutionary programming* (EP), initially conceived to optimize machine intelligence [21] and later accepted as an all-purpose optimization method [22], [23]. In EP applications for continuous search spaces, individuals are typically mutated by adding normally distributed random numbers, with a mechanism making mutations more likely where they are less disruptive to the parenting genotype. It differs from ES in that all individuals generate an offspring, and in the usage of a probabilistic, tournament-like selection scheme which exerts a smaller degree of selective pressure. In this work, we will concentrate on ES.

B. Mutation Operator

The (μ, λ) strategies do not require any specific genetic representation nor any particular mutation operator (MO). By MO we mean any procedure which modifies a parent in order to create an offspring. MOs can be more or less disruptive and more or less independent from the nature of the task being optimized. For several problems though (cf. [17]) it is desirable to exploit all the knowledge available for contriving a MO, which is not as blind as mutation in nature but goes instead in a specific direction.

The simplest MO for LABS generates the new sequence by flipping one bit, chosen at random, in the old sequence. We call this the 1-flip MO. This operator can be improved in several ways. We found it advantageous to flip more than one bit at once and to make a *preselection* of the created individuals. As the energy of a sequence is determined by the square of the R_k it is especially advantageous to eliminate large R_k . Therefore we do not consider the full fitness function F but we demand that the elements of a certain subset of m especially large R_k be diminished. We select each of these R_k by taking the greatest out of t randomly chosen R_k . After the set of R_k to be reduced has been determined, we flip $n > 1$ bits at random and check whether all R_k in the subset have been reduced. In this case, we accept this sequence. Otherwise, we reject it and mutate the original sequence again. If s_{max} trials are unsuccessful we accept such a sequence as well. The strategy has the advantage that a certain directionality is imposed on the mutation process as compared to the complete randomness of the 1-flip operator.

We obtained the best results by setting $n = 2$, $m = 2$, $t = L/4$, and $s_{max} \approx L$. We refer to this setting as the preselective mutation operator (PMO).

The preselection scheme tackles the frustration by fulfilling some requirements first without concerning others that are only given attention in the final selection step. In this sense it shares some features of the democratic optimization method recently introduced in [5].

C. Measures for Population Diversity

The efficiency of an ES depends on the MO but also on the choice of μ and λ . A high value for λ can become computationally expensive, whereas setting it too low would not guarantee an effective scanning of the neighborhood. Using comma strategies in combination with the PMO, we studied the number of different sequences N_F with a fitness above a certain threshold F . To do a reasonable comparison for different λ at given μ , we determined this number using a fixed limit of fitness evaluations (number of generations times λ).

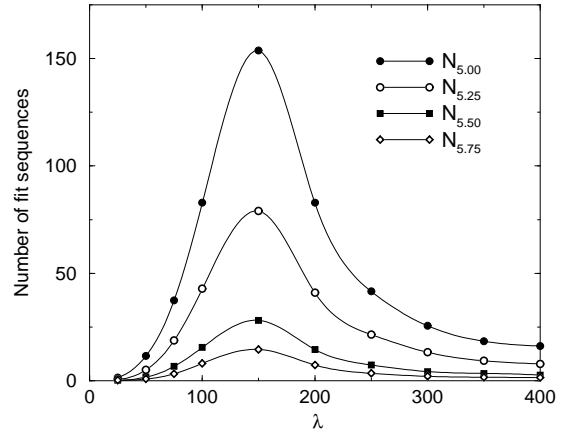


Fig. 3. The number of different sequences with a fitness above a certain threshold, as function of λ for $\mu = 10$ and skew-symmetric sequences of $L = 101$. Every data point was obtained by averaging over 200 separate runs with $2 \cdot 10^6$ fitness evaluations (number of generations times λ) each.

The results shown in Figure 3 as a function of λ exhibit a clear maximum at $\lambda \approx 150$. By choosing λ smaller and larger than this value, two rather different dynamics are obtained, as can be seen from Figure 4 and 5, where the fitness of the best individual in the population is plotted against the generation number. For small λ , one observes a stochastic pattern (Figure 4). For large λ (Figure 5), one finds a highly correlated pattern showing the repeated occurrence of certain fitness levels.

The repeated creation of the same locally optimal sequences is common to several EAs in discrete spaces [17], when reproduction chances are increased. To verify that this can explain the quantized dynamic shown in Figure 5 and to give a quantitative description of this phenomenon, we consider the probability p_g that the fittest individual in the population already occurred as such in any of the previous g generations. The upper plot in Figure 6 shows measurements of p_g for different values of λ and g .

The reoccurrence probability is significantly increased for high λ . Note by studying the slope of the curves that the biggest contributions are generated at $g = 1$ and $g = 2$. Therefore, for λ large enough, the chances that the fittest individual in the actual population will be found again in at most two generations are considerably high, thus explaining the quantized fitness of Figure 5. The slope of the curves decreases with g indicating that longer cycles of

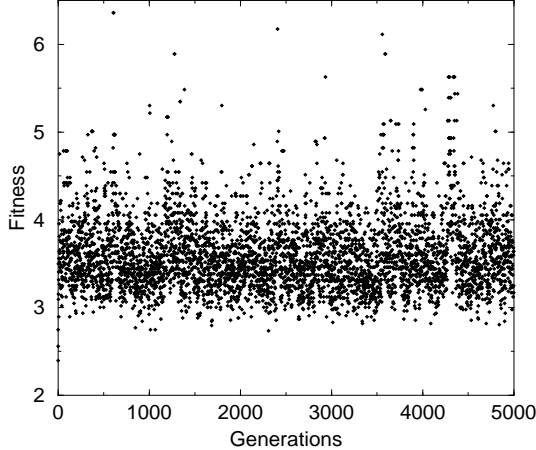


Fig. 4. Evolution of the fitness of the best individual in each generation for the preselective mutation operator in the case of $L = 101$, $\mu = 10$, and $\lambda = 100$. The dynamic shows no quantization. Notice that *comma* strategies do not guarantee any monotonicity.

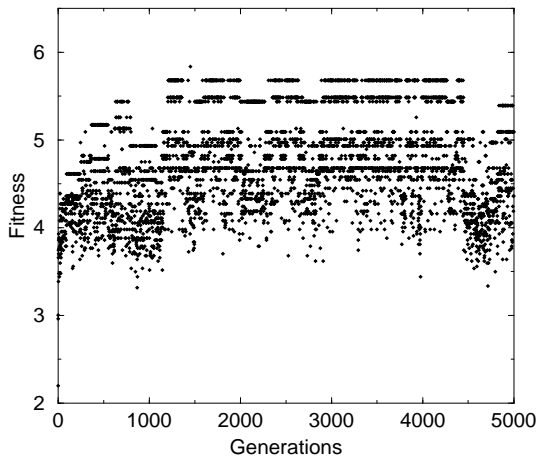


Fig. 5. Evolution of the fitness of the best individual in each generation for the preselective mutation operator in the case of $L = 101$, $\mu = 10$, and $\lambda = 250$. For these values of the parameters the fitness is clearly quantized.

fittest sequences become more and more unlikely.

In the second graph of Figure 6, we plotted p_{200} as a function of λ in order to use it as an estimate for the total probability for the reoccurrence of the fittest individual. If $\lambda \gtrsim 150$ the reoccurrence probability increases significantly and the above mentioned effect becomes dominant causing a decrease in efficiency.

Furthermore, additional investigations of the population diversity showed that the number of different sequences within the same generation N_D decreases only slowly, starting almost at the upper limit of $N_D = \mu = 10$ for small λ and then gradually decreasing to 8.0 at $\lambda = 400$ even if measured over a relatively large number of generations, 10^5 . Therefore, we do not consider the modest increase in number of duplicate individuals as the main cause for the performance decrease shown in Figure 3.

We use the peak in the same figure to define the optimal

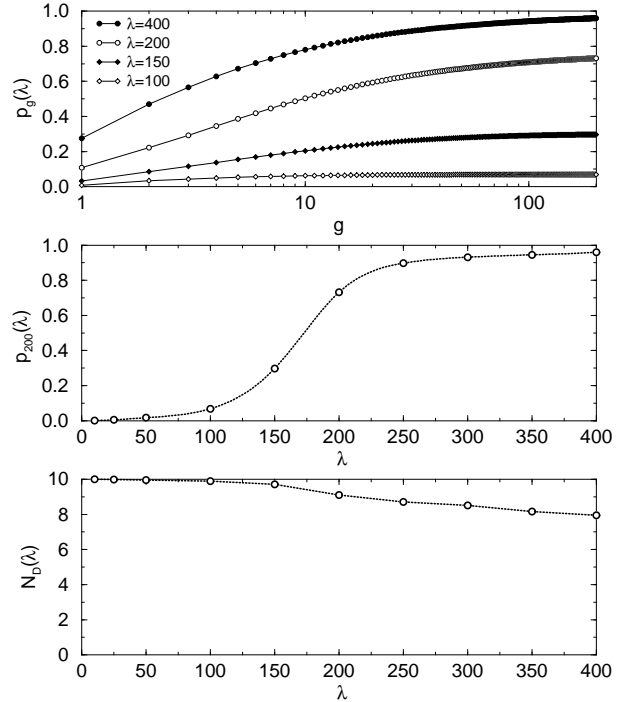


Fig. 6. The upper graph shows the probability $p_g(\lambda)$ that the fittest individual had already occurred as such in the g previous generations. All values were averaged over the last 5000 generations of 100 separate simulations, using the PMO for $L = 101$ and $\mu = 10$. In the middle graph, $p_{200}(\lambda)$ is plotted as function of λ in order to estimate the total reoccurrence probability. The significant increase shows that raising λ does not necessarily lead to a more successful strategy, whereas the average number of different sequences within the population N_D shown in the lower graph changes only slightly with increasing λ .

value of λ . This λ^{opt} can be determined for other values of μ and for different MOs with the same technique, as shown in Figure 7 for the PMO and for operators flipping randomly 1 and 2 bits. The slope of the graphs is related to the average size of the neighborhood that can be reached in one mutation, and for all three MOs, one finds a nearly linear dependence. The biggest neighborhood is generated by the 2-flip operator, whereas in case of one flip, the much smaller neighborhood corresponds to a reduced slope. The PMO also fits in this scheme since the preselection procedure leads to a reduction of the effective neighborhood but still allows comparatively large jumps in Hamming space.

IV. RESULTS

A comparison of N_F between the PMO and the other two MOs taken into consideration, shows that the PMO performs remarkably better. Setting $\mu = 10$ and $\lambda = \lambda^{opt}$ for the 1-bit flip, 2-bit flip, and the PMO and considering $2 \cdot 10^6$ fitness evaluations, one obtains $N_{5,0} = 38$, 6, and 153 respectively. The greater number of fit sequences found by using the PMO, confirms the advantage of employing a preselection scheme.

The results obtained with it ($\mu = 10$ and $\lambda \approx L$) up to $2 \cdot 10^6$ generations are shown in Table I. Due to the non-deterministic character of the ES, the size of the search space and roughness of the fitness function, the results

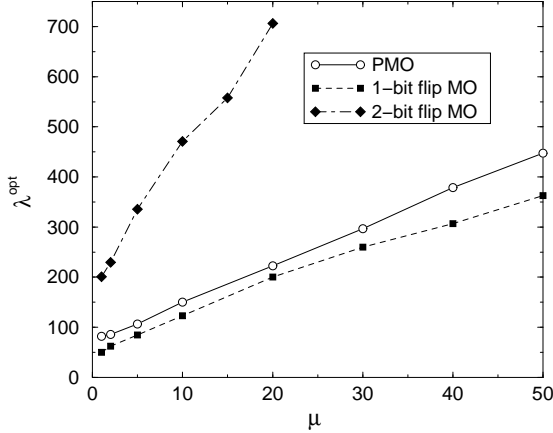


Fig. 7. λ^{opt} as a function of μ for the preselective mutation operator (PMO), the 1-bit random flip operator and 2-bit random flip operator applied skew-symmetric sequences of length $L = 101$.

TABLE I

HIGHEST MERIT FACTORS FOR SKEW-SYMMETRIC SEQUENCES AS PUBLISHED IN [18], [19], [20], [9], [7], AND BY OUR METHOD.

L	[18]	[19]	[20]	[9]	[7]	this work
81	7.32	8.20		8.04	8.20	8.20
101	6.06	8.36	6.91	8.36	8.36	8.82
103	5.90	9.56	7.76		9.56	9.56
105	6.07	8.25	7.61		8.25	8.78
107	6.53	8.46			8.46	8.46
109	6.15	8.97			8.97	8.97
111	6.02	8.97			8.97	8.97
113	6.33	8.49			8.49	8.49
115	6.40				8.60	8.88
117	6.42				8.12	8.71
119	6.01				7.67	8.02
121	6.61			6.75	8.67	8.67
141	6.01			6.48	7.45	8.83
161	6.02			6.02	6.89	8.39
181	5.70			6.02	6.77	7.75
201				5.92	6.29	7.46

strongly depend on the initial conditions, making it unlikely to reobtain the best values in every run. For all values of L fitness values to date were improved or reproduced by our strategy. In Figure 8 the results are compared with an approximation for largest possible fitness [10] of skew-symmetric and general sequences.

The values found for odd L from 101 to 121 are close to these predictions. For L over 121 momentous improvements were achieved. According to the approximation for the distribution of fitness values given in [2], for $L = 201$ the sequence with $F = 7.46$ is $2 \cdot 10^6$ times rarer than the one with $F = 6.29$ found in [7]. For $L = 100$, which is as tedious as $L = 201$ when exploiting skew-symmetry, we found the sequence shown in Table II with $F = 7.84$, which

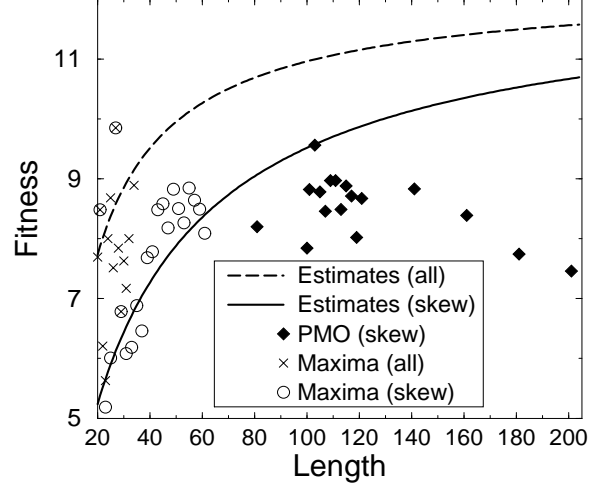


Fig. 8. Fitness of the best sequences generated by the preselective mutation operator and results for small L of the exhaustive search compared with the estimated optima given in [10].

is a significant improvement compared to $F = 6.39$ in [13]. According to the approximation for the distribution of fitness values there are only 4000 sequences with fitness equal or higher within the $2^{100} \approx 1.3 \cdot 10^{30}$ possible ones. A purely stochastic sampling procedure with the same amount of fitness evaluations ($2 \cdot 10^{10}$) would find fewer than 10^{-16} of such sequences.

Flipping two bits at a time is a compromise between the two aims of generating fit offspring and sampling the configuration space efficiently. When flipping too many bits at once only mediocre offspring are generated and the method degenerates to a random search. The probability of escaping from the many local maxima characterizing this search space increases dramatically by flipping several bits at a time. The impossibility of finding better sequences when exhaustively searching the neighborhood of higher-than-average sequences for up to five bits ($F > 7$, $L = 100$) also suggests the same conclusion.

TABLE II

BEST SEQUENCE GENERATED FOR $L = 100$ WITH A FITNESS OF 7.84.

1	- - + - + - - + + - + + + - + + + - + + + + - +	25
26	+ + - + - + + - - - + - + + - + - - - + - + + -	50
51	- - + + + - + - + - - - + - - - - + + + - - + + +	75
76	- + + + + - - - - - + + - + + + - - + - + -	100

V. SUMMARY

We have shown how ES can be successfully applied to the LABS problem. Checking the diversity between successive populations allowed the empirical determination of the most suitable values for the optimization parameters of the strategy. The particular form of the fitness was exploited to construct a mutation method more likely to

bridge the large distances between different local maxima. The selection procedure used in the ES is based on a computationally cheap preselection scheme that helps handle the frustration of the problem.

ACKNOWLEDGMENTS

We want to thank Werner Ebeling for triggering our interest in the LABS problem. We thank Torsten Asselmeyer, Frank M. Dittes, Helge Rosé, Oliver Rudzick, and Thorsten Pöschel for valuable hints and discussion. This work was supported by the BMFT Projekt EVOALG.

REFERENCES

- [1] M.J.E. Golay, "A Class of Finite Binary Sequences with Alternative Autocorrelation Values Equal to Zero", *IEEE Trans. on Information Theory*, vol. 18, pp. 449-450, May 1972.
- [2] J. Bernasconi, "Low autocorrelated binary sequences: statistical mechanics and configuration space analysis," *J. Physique*, vol. 48, pp. 559-567, April 1987.
- [3] K. Ohue and T. Aibara, "Low Autocorrelation-Sidelobe Skew Symmetric Binary Sequences Based on Complex Exponential Functions," *Elec. and Comm. in Japan*, vol. 74, pp. 24-31, March 1991.
- [4] M.J.E. Golay, "Sieves for Low Autocorrelation Binary Sequences," *IEEE Trans. on Information Theory*, vol. 23, pp. 43-51, January 1977.
- [5] F.M. Dittes, "Optimization on rugged landscapes: a new general purpose monte carlo approach," *Phys. Rev. Let.*, vol. 76, pp. 4651-4655, June 1996.
- [6] H. Mühlenbein, "Asynchronous parallel search by the parallel genetic algorithm", *Proceedings of the 3rd IEEE Symposium of Parallel and Distributed Processing*, V. Ramachandran Ed., Heidelberg: Springer 1991, pp. 526-533
- [7] A. Reinholz, *Ein paralleler genetischer Algorithmus zur Optimierung der binären Autokorrelations-Funktion*, Diplom Thesis, Universität Bonn, October 1993.
- [8] W. Krauth and M. Mézard, "Aging without disorder on long time scales," LPTENS preprint 94/20, 1994.
- [9] C. De Groot and D. Würtz, "Statistical Mechanics of Low Autocorrelation Skew-symmetric Binary Sequences," *Helvetica Physica Acta*, vol. 64, pp. 86-91, November 1991.
- [10] M.J.E. Golay, "The Merit Factor of Long Low Autocorrelation Binary Sequences," *IEEE Trans. on Information Theory*, vol. 28, pp. 543-549, May 1982.
- [11] M.J.E. Golay, "The Merit Factor of Legendre Sequences," *IEEE Trans. on Information Theory*, vol. 29, pp. 934-936, November 1983.
- [12] H. Rosé, W. Ebeling and T. Asselmeyer, "The Density of States - a Measure of the Difficulty of Optimisation Problems", in *Parallel Problem Solving from Nature IV. Proceedings of the International Conference on Evolutionary Computation*, H.-M. Voigt, W. Ebeling, I. Rechenberg and H.-P. Schwefel, Eds., Heidelberg: Springer 1996, pp. 208-217
- [13] O. Rudzick, *Evolutionäre Suche nach Sequenzen mit kleinen Autokorrelationen (LABS): Bestimmung der Entartung der Autokorrelationfunktion*, Report BMFT-Project EVOALG, 1994.
- [14] W. Ebeling and R. Feistel, *Physik der Selbstorganisation und Evolution*, Berlin: Akademie Verlag, 1982.
- [15] I. Rechenberg, *Evolutionstrategien - Optimierung technischer Systeme nach Prinzipien der biologischen Information*, Stuttgart-Bad Cannstatt: Friedrich Frommann Verlag, 1973.
- [16] H.-P. Schwefel, *Numerical Optimization of Computer Models*, New York: Wiley, 1981.
- [17] T. Bäck, *Evolutionary Algorithms in Theory and Practice*, Ph.D. Thesis Informatik Centrum Dortmund, Germany, 1995.
- [18] G.F.M. Beenker, T.A.C.M. Claasen and P.W.C. Hermens, "Binary Sequences with a Maximally Flat Amplitude Spectrum", *Philips J. of Research* vol. 40 pp. 289-304. 1985.
- [19] M.J.E. Golay and D.B. Harris "A new search for skewsymmetric binary sequences with optimal merit factors" *IEEE Trans. Inf. Theory* (USA), vol. 36, no. 5, pp. 1163-1166, September 1990.
- [20] Q. Wang, "Optimization by Simulating Molecular Evolution", *Biol. Cybern.* vol. 57, 1987, pp. 95-101.
- [21] L.J. Fogel, A.J. Owens, M.J. Walsh, *Artificial Intelligence through Simulated Evolution*, New York: Wiley & Sons, 1966.
- [22] W.M. Spears, K.A. De Jong, T. Bäck, D.B. Fogel, H. de Garis, "An Overview of Evolutionary Computation", in *European Conference on Machine Learning*, P.B. Brazil, Ed. New York: Springer 1993, pp. 442-459.
- [23] T. Bäck, G. Rudolph, H.-P. Schwefel "Evolutionary Programming and Evolution Strategies: Similarities and Differences", in *Proceedings of the Second Annual Conference on Evolutionary Programming (EP'93)*, D. B. Fogel and W. Atmar, Eds. San Diego CA: Evolutionary Programming Society 1993, pp. 11-22.