Proceedings of the
International Conference on Complex Systems
Nashua, NH, 21–26 Sept. 1997.

Y. Bar-Yam (ed.), New England Complex Systems Institute (1997) .
Also on-line in the *InterJournal*.

Evolutionary Strategies of Optimization and the Complexity of Fitness Landscapes

Helge Rosé
Humboldt Universität zu Berlin,
Institut für Physik, Invalidenstr. 110, D-10115 Berlin, Germany

Evolutionary Algorithms have proved to be a powerful tool for solving complex optimization problems. The underlying physical and biological strategies can be described by the Master Equation formalism. Combination of both strategies creates a new basic class of Evolutionary Algorithms where robustness as well as performance of the optimization process improve substantially. To characterize the complexity of an optimization problem one may introduce a measure which remains invariant with regard to different schemes of representation: the density of states. It is the probability that an arbitrary chosen state has a certain fitness value. The knowledge of this probability makes it possible to estimate the optimal fitness value and the computational effort to find a better solution of the problem. A general method is presented which allows to approximate the density of states during the optimization process. This is demonstrated for frustrated sequences, road networks and especially for the secondary structures of RNA.

1 Introduction

The process of optimization is a fundamental principle which is reflected in the dynamical laws of physics, the progress of biological evolution and the technological development. The question about *optimal solutions* becomes more and more important in many spheres of our life. The fruitful confluence and interaction of various fields of science led to the idea of solving complex optimization problems by means of thermodynamical and biological principles. The last decade was marked by a great development of stochastic optimization algorithms which may be collected by the concept of *Evolutionary Algorithms* consisting of: Simulated Annealing [15], Evolutionsstrategien [19], Genetic Algorithms [14] and Evolutionary Programming [10]. The search for optimal solutions may be regarded by a motion $x(t)$ in a search space visiting global optima of a function $F(x)$. The motion $x(t)$ is not deterministic – there is a stochastic *generation* of *alternatives* for further motion. The realization of one of them takes place by an *assessment* related to the *fitness* $F(x)$ which denotes the quality of the state x of the optimization object. The class of all algorithms realizing the same probabilistic dynamics is called an *Evolutionary Strategy* [2].

Any information about the complexity of an optimization problem is very fruitful to design a good representation and algorithm for practical optimization. The general issue is the classification of the problems to understand the dynamics of optimization leading to optimal search algorithms.

An *optimization problem* provides the set X of all admitted states of the objects which has to be optimized and a fitness function F which quantifies the aim of the problem. But it tells us nothing about it how we should move from one state to an other. As part of the used strategy, we have to choose a mutation operator generating new states from the current one. At this point, the pure set of all possible states becomes a search space (X, N_Ξ) with a *neighborhood structure* N_Ξ consisting of all pairs of states connected by one mutation. The *fitness landscape* is the surface of F on (X, N_Ξ) and the dynamics of the optimization strategy reflects the motion of the searchers on this landscape. It is important to realize that the landscape depends on the problem by X and F as well as the used mutation operator generating N_Ξ.

To classify the optimization problem for any mutation operator and strategy we need measures depending on X and F but remains *invariant* related to the neighborhood structure. The *density of states* is the probability to find any state on a certain fitness level. Obviously this number is invariant with regard to N_Ξ and depends on X and F only. By means of the density it is possible to approximate the optimum of the problem and to assess the probability to find it by pure random search.

Using a Boltzmann strategy one may approximate the density of any optimization problem [1, 21]. In this article a improved method shall be presented which may approximate the density of states en passant during the optimization process.

First we consider the frustrated game proposed by Engel [6] and secondly we determine the density of graphs for the problem of road network optimization. Finally the method shall be applied to the folding problem of secondary RNA structures.

2 Evolutionary Strategies

The stochastic motion of searchers governed by the algorithm can be described by the dynamics of the probability $P(x, t)$ to find a searcher with state x at time t. Different algorithms may realize one and the same probabilistic dynamics. The class of all algorithms realizing the same dynamics is called an *Evolutionary Strategy* [2]. Evolutionary Algorithms may be divided according to their physical or biological origin into the classes of the Boltzmann strategy (BS) or Darwin strategy (DS). They can be described by the master equation of $P_x = P(x,t)$ [4, 1]

$$\frac{dP_x}{dt} = \gamma\left(\langle F\rangle - F\right)P_x + \sum_{y\in X}\left(W_{xy}P_y - W_{yx}P_x\right). \quad (1)$$

The first term represents the selection process and the second one the mutations of an Evolutionary strategy *minimizing* $F(x)$. In the case of a Simulated Annealing algorithm [17, 15, 16] we have to set $\gamma = 0$ and the transition probability to

$$W_{xy} = \begin{cases} S_{xy} & : \triangle F \le 0 \\ S_{xy}\, e^{-\beta\triangle F} & : \triangle F > 0 \end{cases} \quad (2)$$

with $\triangle F = F(x) - F(y)$ and the probability S_{xy} of a mutation of y to x. The dynamics (1) describes the Boltzmann strategy and its stationary distribution is the Boltzmann distribution

$$\lim_{t\to\infty} P(x,t) \sim e^{-\beta F(x)}.$$

Evolutionary Algorithms using a selection scheme adopted from the natural selection of biological systems may realize a Darwin strategy with $\gamma = 1$ and $W_{xy} = S_{xy}$. The first term of (1) was introduced by FISHER [9] and EIGEN [8] to explain a simple model of Darwinian selection with a reproduction rate related to the mean fitness of a population $\langle F\rangle = \sum F(x)P_x$. The combination of the strategies $\gamma > 0$ (2) generates a new class of Evolutionary Algorithms – the *Mixed Strategy* – which shows a improvement in robustness and velocity of optimization. For details of the simulation of the algorithms and the analytical investigation of the strategies see [4, 2, 22].

3 The Density of States

The concept of the density of states originates from statistical physics. It is widely used in solid state physics and also to describe thermodynamic optimization processes [18]. For the explanation, let us use a *minimization problem* defined by a finite set X of N possible states x of the optimization object and a fitness function $F : X \to \mathbb{R}$. The solutions of such a problem are the optimal states minimizing F.

If we do not know the explicit form of F all we can do is moving from state to state and looking for the optimal one. To do that we have to choose a group of mutation operators $\Xi_r : X \to X$ transforming one state in a neighboring one. The set of all pairs of neighbors builds the neighborhood structure N_Ξ of the search space (X, N_Ξ). The *fitness landscape* is a surface $F = F(x)$ in $X \times \mathbb{R}$ (see Fig. 1). An optimization problem can be represented by different fitness landscapes which depends on the choice of the mutation operator generating different neighborhood structures.

A general way to classify the landscape for any neighborhood structure is to build equivalence classes of the states which are invariant due to a change of N_Ξ. We have to choose an equivalence function yielding the same value for all states of a class. The only function we have given by the problem without further knowledge is the fitness function.

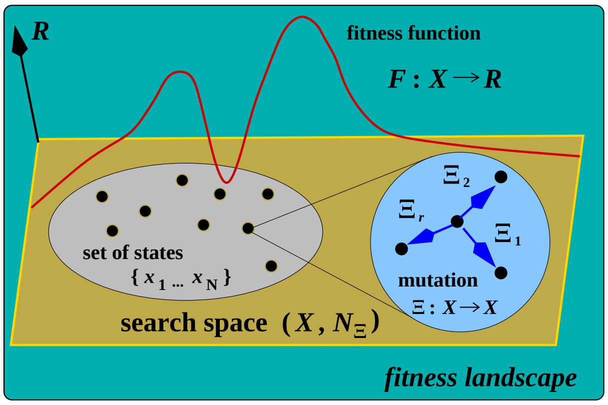

Figure 1: The concept of fitness landscapes encloses three entities: the set of states, a group of mutations and the fitness function.

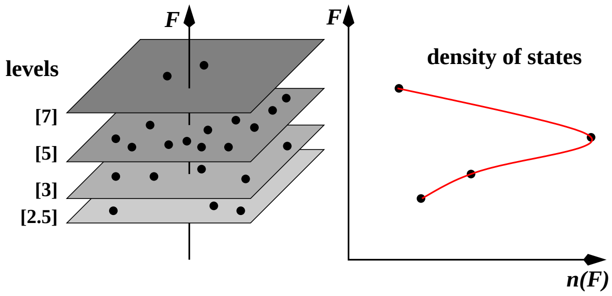

Figure 2: The numbers of states at the fitness levels of the problem form the density of states.

Using this fact we are able to define (see Fig. 2): The *equivalence classes* of the landscape are the fitness levels $[\epsilon] = \{x : x \in X,\, F(x) = \epsilon\}$. The number of states on the level $[\epsilon]$ is called *weight* W_ϵ and the probability

$$n(\epsilon) = \frac{W_\epsilon}{\sum_\epsilon W_\epsilon} \quad (3)$$

to find any state on the level $[\epsilon]$ is the *density of states* (DOS).

The DOS depends by definition on X and F but it does not depend on N_Ξ. In this sense we may say: The DOS is a classifying measure for the optimization problem. It tells us how sparsely states of a given quality are distributed. Most of the realistic problems possess only few optimal states which corresponds to a very small probability of the tail of the DOS. Thus, in the limit[1] $n(F) \to \frac{1}{N}$, the optimal fitness value can be estimated [21].

Another important fact is the possibility to derive informations about the computational effort of optimization. For problems characterized by a very steep decay in $n(F)$, one can almost be sure that an extensive enhancement of the computational effort does not necessarily result in better

[1] In case of one optimum F^* we have $n(F^*) = \frac{1}{N}$.

optimization values. Most interesting, however, are problems which display slow decay of the DOS in the region of optimal fitness values. Here, the number of better solutions remains sufficiently large, which means that more effort in the optimization has a good chance to yield better results. In case of pure random search, the effort needed to improve a solution can be derived from the DOS. This gives us an upper bound for all more effective optimization strategies. The time to find a solution by random search can be estimated by $\tau(F) \sim \frac{1}{n(F)}$, thus the effort to improve the solution by $\triangle F$ can be characterized by time ratio

$$\eta_r = \frac{\tau(F+\triangle F)}{\tau(F)} = \frac{n(F)}{n(F+\triangle)}.$$

If the time ratio η of the used optimization strategy remains smaller then η_r in the range of good fitness values, the enhancement of the computational effort has to be advised. Otherwise one should spend time to design a more suitable strategy for the problem.

3.1 Methods of determination of the density of states

It is a nice fact that the DOS can be approximated en passant during the optimization process. One method is to determine the DOS by simulating a ensemble of searchers according to the Boltzmann strategy [1, 21]. The probability distribution $P(F,t)$ of the ensemble , which is approximated by the frequency distribution of the searcher, tends towards the equilibrium density $P(F) \sim n(F)\exp(-\beta F)$. After a sufficiently long simulation time we obtain the DOS from the frequency distribution by multiplication with $\exp(\beta F)$. This method can be applied to any optimization problem but is restricted according the choice of the strategy.

To find a method without this restriction let us answer the question: How does the neighborhood structure influence the transition from one fitness level to another one? The neighborhood structure can be represented by its adjacency matrix which is $A_{xy} = 1$ for all pairs (x, y) connected by one mutation and otherwise $A_{xy} = 0$. With the assumption that all mutations have the same chance, the *transition probability of the states* with respect to one mutation is simply $S(x,y) = A_{xy}/\sum_y A_{xy}$, i.e. the transitions of the states are completely determined by the neighborhood structure.

To obtain the probability $T(\epsilon,\gamma)$ of a transition from a fitness level γ to the level ϵ by one mutation we introduce the *level structure* of the landscape by

$$\Phi_{\epsilon x} = \begin{cases} 1 & : F_\epsilon = F(x) \\ 0 & : F_\epsilon \neq F(x) \end{cases}, \; \Phi^\top = (\Phi_{x\epsilon}). \tag{4}$$

This matrix formalizes the affiliation of states to the fitness levels. Inserting the definition it is easy to prove the relations

$$\left(\Phi\Phi^\top\right)_{\epsilon\epsilon} = W_\epsilon,\ \mathrm{tr}\Phi\Phi^\top = \sum_\epsilon \mathrm{W}_\epsilon\,, \tag{5}$$

$$\sum_x S(x,y) = 1 \Longrightarrow \sum_\epsilon \left(\Phi S \Phi^\top\right)_{\epsilon\gamma} = W_\gamma\,, \tag{6}$$

leading to the *transition probability of the fitness levels*

$$T(\epsilon,\gamma) = \frac{\Phi S \Phi^\top}{\left(\Phi\Phi^\top\right)_{\gamma\gamma}},\ \sum_\epsilon T(\epsilon,\gamma) = 1\,. \tag{7}$$

Multiplying (7) with W and using (6) it becomes clear that the density of states is an eigenvector with eigenvalue one of the transition probability of the fitness levels.

$$TW = W,\ Tn = n,\ n(\epsilon) = \frac{\left(\Phi\Phi^\top\right)_{\epsilon\epsilon}}{\mathrm{tr}\Phi\Phi^\top} \tag{8}$$

We can see that the DOS depends only on the level structure Φ and is independent of S, e.g. it is invariant with respect to N_Ξ.

The eigenvector property (8) makes it possible to approximate $n(F)$ from the transition probability of the levels. We just have to count every transition caused by a mutation in a frequency matrix during the optimization and to calculate its eigenvectors numerically . This was already been mentioned in [18] in the context of Simulated Annealing. But in fact, the derivation of (8) shows that there is no reason to restrict the method to a particular strategy. Every strategy which is able to reach all points of the search space from any starting point, e.g. every ergodic strategy, may be used.

4 Examples

The first example is to compare the two methods described above. It was introduced by Engel [6] as a simple model of a frustrated optimization problems. A detailed description can be found in [1, 21].

The search space X consists of all sequences x of the length L with $\lambda = 4$ letters. For any sequence there are $(\lambda - 1)L$ point mutations generating the group of mutation operators. The number $\alpha(x)$ of letters with a alphabetic successor and the number $\pi(x)$ of letters which recur with a period $p = 5$ has to be maximized. These two requirements standing opposed to one another cannot be solved at the same time. Therefore this kind of problem is called frustrated. In order to formulate the problem as a minimization task the fitness function is chosen by $F = -\alpha(x) - b\pi(x)$, $b > 0$. It can be shown [20] that for $L = np$, $n = 1, 2, ...$ a bifurcation takes place if the parameter b becomes the critical value

$$b_c = \begin{cases} \frac{1}{p} & : \frac{L}{\lambda}\, integer \\ \frac{1}{p} - \frac{1}{L} & : \frac{L}{\lambda}\, rational \end{cases}. \tag{9}$$

For $b < b_c$ only the complete alphabetic sequence is the optimum. For $b \geq b_c$ optimal sequences are mostly periodic with a few breaks in the alphabetic order, e.g $L = 20$: **BBCDABBCDABBCDABBCDA**.

Fig. 3 shows the DOS of sequences with $L = 20$ which were obtained with help of both methods. The DOS for the Engel problem has a Gaussian like shape. This is not surprising since the fitness function is composed of a number of additive contributions.

The second method was used both with the Boltzmann strategy (BS) and the Mixed strategy (BDS). Whereas the first method is able to approximate the DOS up to a fitness value of $F = -17$ the second one reached the the optimum $F = -20$.

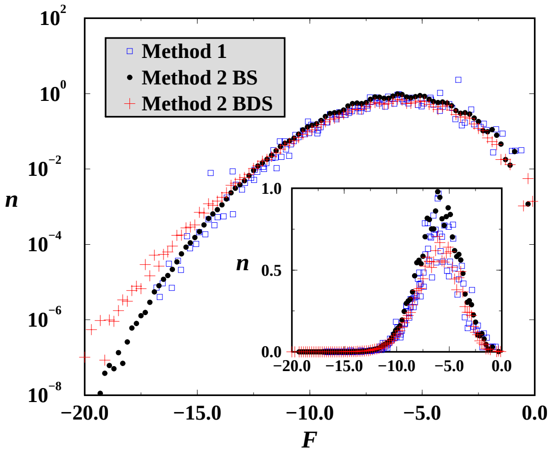

Figure 3: Density of states of the Engel problem of frustrated sequences with $L = 20$, $\lambda = 4$, $p = 5$, $b = 0.2$ obtained by the first method using the Boltzmann strategy and by the second method applying the Boltzmann strategy (BS) as well as the Mixed strategy (BDS).

The obtained results are very similar one to another in all of the three cases. One may conclude that the second method does not depend on the used optimization strategy. Thus the determination of the density of states is not confined by using the Boltzmann strategy anymore.

The second application is the optimization of road networks [7, 22]. As configuration 39 points of the street network of Martina Franca Fig. 4 were chosen.

The search space is the set of all connected graphs x of the point configuration. Deletion and insertion of a link are used as elementary mutations. There are two obvious factors to be minimized:

1. The distance from each point to each other point shall be as small as possible.
2. The network shall be as short as possible to minimize the cost of construction.

Thus the fitness function is given by

$$F(x) = (1 - \lambda)d(x) + \lambda l(x), \; 0 \leq \lambda \leq 1.$$

The first part d is the detour, i.e. the difference between the shortest and direct path from one point to an other averaged over all pairs. l is the length of the network. The two contradicting goals are weighted by the price factor λ, which allows us to choose the importance of shortness or construction cost, respectively.

The obtained DOS for the network optimization (see Fig. 5) is an example for a non Gaussian DOS (for details see [21, 22]) .

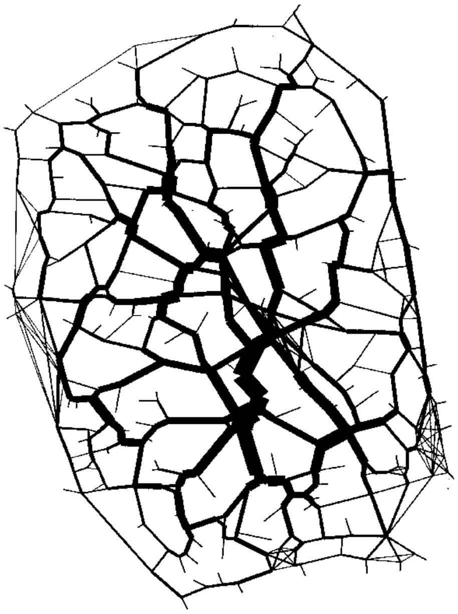

Figure 4: Road network of Martina Franca (Italy).

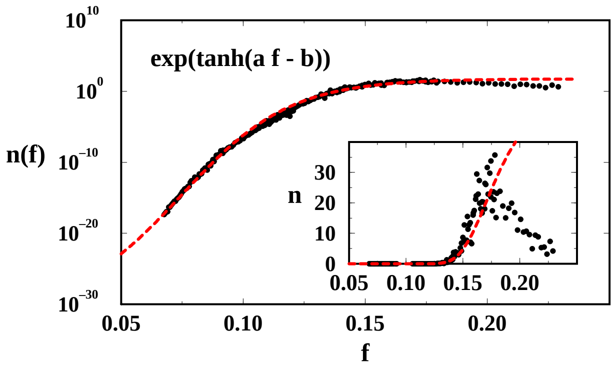

Figure 5: Density of states of the network optimization.

5 Secondary RNA structures

A central problem in biopolymer design and research is the folding of sequences into structures. In this field of research, the investigation of RNA sequences is very important because of their catalytic capacities and perfectly tailored properties [23, 3]. The first step of folding can be modelled by folding into secondary structures [11]. The DOS of the secondary RNA structures is important to understand the versatility of RNA molecules [12].

The number of structures increases exponentially with the sequence length L [11], therefore an exhaustive search becomes quickly infeasible for $L > 40$. The DOS can be determined by a specialized dynamic programming method [5] for longer sequences ($L = 76$), but this technique does not provide the explicit secondary structure.

The application of an Evolutionary Strategy overcomes this limitation. Using this optimization strategy the sub-optimal structures, which are relevant for the biological interpretation, as well as an approximation of the DOS are provided.

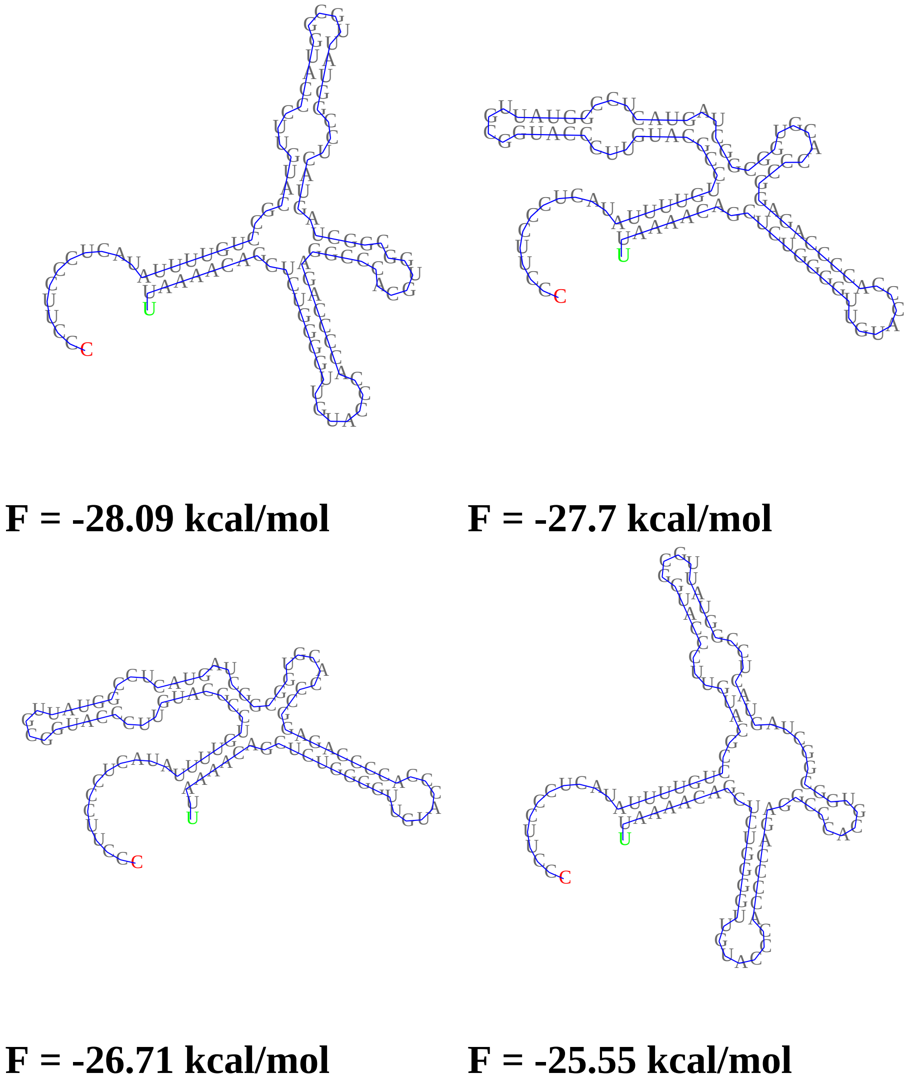

Figure 6: The secondary structures of a part ($L = 100$) of the polio virus type 1 Mahoney (AC V01148) optimized by the Boltzmann strategy. The strategy was able to find the minimal free energy structure as well as suboptimal structures.

The search space of the folding problem can be represented by the set of sequences consisting of balanced parentheses and dots. The pairs of parentheses has to be compatible with the possible Watson-Crick pairs (AU, GC) and GU of the given RNA sequence and the dots denotes unpaired nucleotides. For the optimization procedure the following mutations operators were chosen:

bind	.... $\rightarrow$.(.....)...
dissolve	.(.....)... $\rightarrow$ ....
move	.(.....)... $\rightarrow$ ..(....)...
pull tight	...(...)... $\rightarrow$.(((...))).
pull up	.(((...))). $\rightarrow$ ...(...)...

The fitness function is the free energy of the structure which was calculated by a subroutine of the VIENNA RNA package 1.1 beta [13] .

The minimization was carried out for the first part (5'-cloverleaf, $L = 100$) of the polio virus type 1 Mahoney (AC V01148) by the Boltzmann strategy. The Evolutionary strategy was able to obtain both the minimal free energy structure with $F = -28.09\,\text{kcal/mol}$ and the suboptimal structures of the RNA Fig. 6. The DOS was determined by the eigenvector method Fig. 7.

The obtained density for the Polio virus is overall Gaussian like the DOS for yeast obtained by the dynamic programming method in [5]. The slow decay of the DOS suggests that also the optimization of longer sequences may to be feasible by application of Evolutionary Strategies.

6 Conclusions

The density of states and in particular its tail at optimal fitness values is a classifying measure of the complexity of an optimization problem. It is invariant to any choice of the neighborhood structure of the fitness landscape and characterizes only the problem. The knowledge of this density allows the approximation of the optimal fitness value of the given problem and to assess the benefit of the further opti-

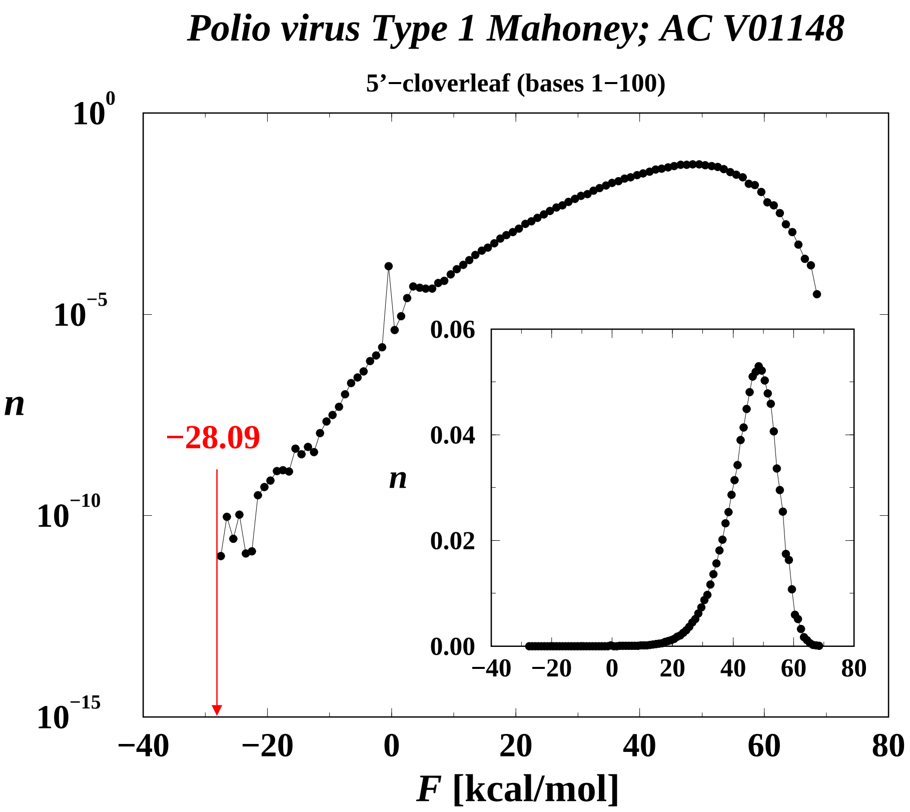

Figure 7: The density of states of the secondary structures of the polio virus.

mization process. This important fact makes it feasible to define a termination criterion which is relevant for practical optimizations. For wide a class of optimization problems the density of states can be approximated by the eigenvector method during the optimization process.

7 Acknowledgment

This work has been supported by the BMBF, *Verbundprojekt* EVOALG: FKZ 01 IB 403 B3. I thank Prof. W. Ebeling, T. Asselmeyer, U. Erdmann, C. Forst and J. Palinkas for ideas, help and discussions.

References

[1] Asselmeyer, T., W. Ebeling, and H. Rosé, "Smoothing representation of fitness landscapes - the genotype-phenotype map of evolution", *BioSystems* **39** (1996), 63–76.

[2] Asselmeyer, T., W. Ebeling, and H. Rosé, "Evolutionary strategies of optimization", *Phys. Rev. E* **56** (1997), 1171–1180.

[3] Beaudry, A., and G. Joyce, "Directed evolution of RNA enzyme", *Science* **257** (1992), 635–641.

[4] Boseniuk, T., W. Ebeling, and A. Engel, "Boltzmann and Darwin Strategies in Complex Optimization", *Phys. Lett.* **125** (1987), 307–310.

[5] Cupal, J., C. Flamm, A. Renner, and P. Stadler, "Density of states, metastable states, and saddle points. exploring the energy landscape of RNA molecule", *in preparation* (1997).

[6] Ebeling, W., A. Engel, and R. Feistel, *Physik der Evolutionsprozesse*, Akademie-Verlag Berlin (1990).

[7] Ebeling, W., H. Rosé, and J. Schuchhardt, "Evolutionary strategies for solving frustrated problems", *Proceedings of the First World Congress of Computational Intelligence* (Orlando,), IEEE (June 27-29 1994), 79–81.

[8] Eigen, M., "The selforganization of matter and the evolution of biological macromolecules", *Naturwiss.* **58** (1971), 465.

[9] Fisher, R.A., *The Genetical Theory of Natural Selection*, Oxford University Press Oxford (1930).

[10] Fogel, D.B., *Evolutionary computation – toward a new philosophy of machine intelligence*, IEEE Press Piscataway NJ (1995).

[11] Fontana, W., P.F. Stadler, E.G. Bornberg-Bauer, T. Griesmacher, I. Hofacher, M. Tacker, P. Tarazona, E.D. Weinberger,

and P. SCHUSTER, "Rna folding and combinatory landscapes", *Phys. Rev. E* **47**, 3 (1993), 2083–2099.

[12] HIGGS, P.G., "Thermodynamic properties of transfer RNA: A computational study", *J.Chem.Soc.Faraday Trans.* **91** (1995), 2531–2540.

[13] HOFACKER, I., W. FONTANA, P. STADLER, L. BONHOEFFER, M. TACKER, and P. SCHUSTER, "Fast folding and comparison of rna secondary structures", *Monatshefte f. Chemie* **125** (1994), 167–188.

[14] HOLLAND, J.H., *Adaptation in Natural and Artificial Sytstems*, University of Michigan Press Ann Arbor (1975).

[15] KIRKPATRICK, S., C.D. GELATT JR., and M.P. VECCHI, "Optimization by simulated annealing", *Science* **220** (1983), 671–680.

[16] LAARHOVEN, P.J.M., and E.H.C. AARTS, *Simulated Annealing: Theory and Applications*, Reidel Dordrecht (1987).

[17] METROPOLIS, N., A. ROSENBLUTH, M. ROSENBLUTH, A. TELLER, and E. TELLER, "Equation of state calculations by fast computing machines", *J. Chem. Phys.* **21** (1953), 1087–1092.

[18] NULTON, J.D., and P. SALAMON, "Statistical mechanics of combinatorial optimization", *Phys. Rev. A* **37**, 4 (1988), 1351–1356.

[19] RECHENBERG, I., *Evolutionsstrategien - Optimierung technischer Systeme nach Prinzipien der biologischen Information*, Friedrich Frommann Verlag (Günther Holzboog K.G.) Stuttgart - Bad Cannstatt (1973).

[20] ROSÉ, H., *Evolutionäre Strategien und Multitome Optimierung*, PhD thesis Humboldt Universität zu Berlin (in preparation 1997).

[21] ROSÉ, H., W. EBELING, and T. ASSELMEYER, "The density of states - a measure of the difficulty of optimisation problems", *Parallel Problem Solving from Nature IV. Proceedings of the International Conference on Evolutionary Computation* (Heidelberg,) (H.-M. VOIGT, W. EBELING, I. RECHENBERG, AND H.-P. SCHWEFEL eds.), Springer (1996).

[22] SCHWEITZER, F., W. EBELING, H. ROSÉ, and O. WEISS, "Optimization of road networks using evolutionary strategies", *subm. to Evolutionary Computation* (1997).

[23] SYMONS, R., "Small catalytic RNAs", *Ann.Rev.Biochem.* **61** (1992), 641–671.