FORSCHUNGSZENTRUM
ROSSENDORF e.V.

FZR

FZR-92
October 1995
Preprint

Frank-Michael Dittes

Optimization on Rugged Landscapes: A Democratic Approach

Forschungszentrum Rossendorf e.V.
Postfach 51 01 19 · D-01314 Dresden
Bundesrepublik Deutschland

Telefon (0351) 260 3273
Telefax (0351) 260 3700

Optimization on Rugged Landscapes: A Democratic Approach

Frank-Michael Dittes*

Research Center Rossendorf, Institute for Nuclear and Hadronic Physics, D-01314 Dresden, Germany

Abstract

A new strategy for finding optimal solutions to complex problems with many competing requirements is proposed. It consists in a simultaneous optimization of the energy, cost or fitness function of the system itself, and of sub-systems of all sizes with an appropriate weight function. For various spin glasses (the $N-k$ model, the low autocorrelation binary sequence model and the Coulomb glass) and for travelling salesman problems the corresponding Monte Carlo algorithm is shown to yield results superior to those obtained by previous optimization techniques.

PACS numbers: 02.60.Pn, 02.70.Lq, 05.50.+q, 87.10.+e

The description of complex systems with many competing requirements generically leads to multi-peaked or "rugged" functions: the energy, cost or fitness landscape (over the space of all possible configurations of the system) is characterized by a complicated structure with many local optima. In a physical context typical examples are the energy functions of many-body systems with a high amount of frustration: glasses and spin glasses, antiferromagnetics on certain lattices and many random systems. In a broader setting rugged landscapes describe, e. g., the fitness of biological individuals (as a function of the pheno- or genotype structure), the "payoff" of economical or social actions (as a function of the sequence of undertaken decisions in the course of time) or the cost or time required to achieve a certain aim. Finding the global optimum of such systems is a highly non-trivial problem since fulfilling one requirement often leads to an increase in the mismatch of others. In the notion of energy or fitness landscapes such situations correspond to *local* optima, i. e., to situations from which no way out by changing a single degree of freedom is possible.

Different strategies have been proposed to tackle this problem, adopting — in view of the general validity of the problem — ideas from different areas of science. Most prominent examples are simulated annealing and its recent developments [1, 2, 3, 4] (based on the thermodynamically motivated Metropolis algorithm [5]), evolutionary and genetic algorithms imitating adaptive processes in biological systems [6, 7] and neural network techniques [8]. A common feature of these approaches is that they consider exclusively the energy or cost of the *total* system — trying to find configurations improving this quantity without completely excluding "steps into the wrong direction" (in order to avoid trapping in local optima).

It is the aim of the present Letter to formulate an opposite approach. Based on the assumption that an optimal (or near-optimal) solution to complex problems can be found by an interplay of *individual* interests, synchronized to a certain extent by group and system regulations, I propose an efficient and robust Monte Carlo optimization algorithm, often performing better than previously used techniques. Apparently, there is some support for the above assumption from one of the most complex systems — human society – relying, at least partially, on similar principles. Vice versa, the mathematical formulation presented below allows a more detailed insight into

why *democratic* structures, allowing individuals and groups of them to follow their own interests — taking into account those of other individuals as well as of the system as a whole —, are most capable of solving complex problems.

In order to be specific I use a physical terminology and try to find a global minimum ("ground state") of the energy E of a system with N degrees of freedom (called spins or sites). I suppose that each of them can take only a *finite* number M of values; the generalization to continuous problems can be formulated in a straightforward way. Every configuration of the system can be specified by a vector $\mathrm{s} = \{s_1, \ldots, s_N\}, s_i \in \{1, \ldots, M\}$, and the energy landscape is given by a map from the set of all possible configurations into the real numbers: $\mathrm{s} \rightarrow E(\mathrm{s})$. The aim of a Monte Carlo algorithm is to evolve a random initial configuration into the ground state (or close to it) within the smallest possible amount of configuration changes by generating the latter randomly, usually considering only single-spin flips. Than, the optimization strategy is determined by the *criterion* according to which a configuration change is accepted or discarded.

For finite N, the total energy function $E(\mathrm{s})$ may be represented as a sum of k-spin interaction terms, $0 \leq k \leq N$:

$$E(\mathrm{s}) = \sum_{k=0}^{N} E_k(\mathrm{s}) \,, \tag{1}$$

where E_0 is configuration independent, and the general form of E_k reads

$$E_k(\mathrm{s}) = \sum_{i_1=1}^{N} \ldots \sum_{i_k=1}^{N} e_k(s_{i_1}, \ldots, s_{i_k}) \quad , k = 1, \ldots, N \,. \tag{2}$$

The basic objects of the algorithm are the single-spin energies $E^{[1]}_{\{n\}}$ defined as the sum over all interaction terms in which the spin with label n occurs — taking the orientation of all other spins as given:

$$E^{[1]}_{\{n\}}(\mathrm{s}) = \frac{1}{N} E_0 + \sum_{k=1}^{N} \sum_{i_1, \ldots, i_k=1}^{N} e_k(s_{i_1}, \ldots, s_{i_k}) \, \frac{1}{k} \sum_{j=1}^{k} \delta_{n i_j} \,, n = 1, \ldots, N \,, \tag{3}$$

where $\delta_{nm} = 1$ for $n = m$ and 0 otherwise; the weight factor $1/k$ is necessary to insure the additivity relation (4) (see below). From expression (3), one

can construct the set of all multiple-spin energies by defining $E^{[2]}_{\{n_1,n_2\}}(\mathrm{s}) = E^{[1]}_{\{n_1\}}(\mathrm{s}) + E^{[1]}_{\{n_2\}}(\mathrm{s})$, etc. Obviously, the number of different K-spin energy functions generated in this way is given by the binomial $\binom{N}{K}$, so that the total number of energy landscapes equals $\sum_{K=1}^{N}\binom{N}{K} = 2^N - 1$.

The key idea of the proposed optimization strategy is to search for minima not on the single landscape E, but on the *ensemble* of landscapes defined above, i. e., to perform a random walk on a multivalued landscape — looking at each moment of time for steps diminishing the value of a randomly chosen member of this landscape ensemble. There are two properties of the ensemble that make this approach work.
1. Different landscapes are not independent of each other but synchronized in the sense that, for all K and any of the system's configurations, the average value of K-energies is proportional to the system energy:

$$E(\mathrm{s}) = \binom{N}{K}^{-1} \sum_{1\le i_1<\ldots<i_K\le N} E^{[K]}_{\{i_1,\ldots,i_K\}} \,. \tag{4}$$

Therefore, *on average*, an optimization of group energies brings the system closer to its global energy minimum.
2. The structure of *local* minima is, at least for sufficiently random systems, different for different landscapes. As a consequence, switching from the optimization of E to those of its sub-systems, and vice versa, avoids the danger of getting trapped into local minima.

The Monte Carlo algorithm realizing this approach for a system $\mathcal{S}$ of N spins looks as follows.
1. For each site n construct the set of all sub-systems of $\mathcal{S}$ affected by a flip of spin n, i. e., the "neighborhoods" of n. If the number of different neighborhoods is too large, e. g., for spin glasses with medium or long range interactions, it is impossible to consider all of them. In this case, I selected a specific sequence of "K-neighborhoods", $K = 1,\ldots,N$, defined by site n, and the $K-1$ sites most adjacent to it, correspondingly. Note that in application to travelling salesman problems (TSP) instead of single-site flips an elementary configuration flip has to be considered (see below).
2. Define a weight function P over the set of chosen neighborhoods. This

function determines the probability of selecting a given neighborhood to be active for optimization. The structure of P is of central importance for the performance of the algorithm: If P is concentrated on large landscapes, covering almost the whole system S, the algorithm spends most of its time in local minima of E. On the other hand, peaking P on small landscapes, the amount of purely "egoistic" moves prevails, and the total energy strongly fluctuates at a comparatively high level. The best choice should be somehow in between: I obtained good results by generating a random number x uniformly distributed between 0 and 1, and determining the neighborhood size K according to

$$K \equiv P^{-1}(x) = K_{\max} + 1 - [(1 - x + x/(K_{\max} + 1)^{\alpha})^{-\frac{1}{\alpha}}] \,, \qquad (5)$$

where $K_{\max}$ is the interaction radius, i. e., the maximum number of sites affected by a single-spin flip, and α is a small positive number being roughly proportional to $1/K_{\max}$; [] denotes the integer part. Note that the parameter α characterizes, loosely speaking, the "amount of democracy" in the algorithm. Large positive α almost completely exclude spin flips preferable for individuals only, corresponding to a "totalitarian", "centralized" evolution, whereas large negative α lead to an "anarchic" behavior — in both cases the performance of the optimization algorithm turns out to be poor. The fact that the optimal α decreases with increasing $K_{\max}$ indicates, in particular, that — contrary to naive expectation — with raising complexity of a system the rôle of centralism should be reduced.
3. Randomly choose a site n and one of its neighborhoods, for any moment of discrete time. Accept a flip of spin n if it lowers the corresponding neighborhood energy ("random updating"). If the acceptance rate of trial flips is smaller than the inverse number of possible flips a more effective prescription is the following: Choose a sub-system size K according to $P(K)$, examine all possible single-spin flips and accept a randomly chosen one which leads to lower energy of its K-neighborhood ("fitter updating") [9, 10]. I used the first variant for TSP and the latter one for spin glasses (SG).

The algorithm has been tested, first, on three SG.
1. Kauffman's $N-k$ model [11]. This is a system of N spins, each inter-

acting with k others. The system energy is given by

$$E(\mathrm{s}) = \frac{1}{N}\sum_{i=1}^{N} E_i(s_i; s_{i_1}, \ldots, s_{i_k}) \; . \tag{6}$$

I considered the k sites $i_1, \ldots, i_k$ to be adjacent to i on a twodimensional lattice and the E_i to be random numbers uniformly distributed between 0 and 1. The structure of the energy landscape (6) has been studied in great detail. The number of local optima changes from 1, for $k = 0$, to $\mathcal{O}(2^N/N)$ for $k = N - 1$, and their energies are shown to be normally distributed for large k [12]. A numerical test on 200 randomly chosen local minima for $N = 24^2$ and $k = 12$ yields a mean energy per spin m of 0.314, with a standard deviation σ of about 0.007. The $N{-}k$ model served as a testing ground for a first attempt to use the minimization of sub-system energies for the optimization of the total E. The approach followed by Kauffman et al. [9] assumed a static division of the system into "patches" of equal size selfishly trying to minimize their energy. It has been found that an optimal choice of the patch size allows the system to reach energies better than m by 2σ [see the dashed line in Fig. 1]. In the same Figure, the results of 25 runs of the algorithm formulated above are shown. I used $N = 24^2$ and $k = 12$ (correspondingly, $K_{\max} = 13$) and the fitter updating; a good choice for the parameter α turned out to be $\alpha = 1$. After 5000 time steps, i. e., less than 10 executed flips per spin, an average energy per spin of ≈ 0.266 has been reached, corresponding to almost 7 standard deviations below the mean local optimum (solid line). One clearly sees the advantage of the algorithm proposed — demonstrating the importance of using sub-systems of *all* sizes and trying to find a permanent balance between individual, group and system energies.

2. The low autocorrelation binary sequence (LABS) model [13]. This optimization problem originating from communication engineering consists in finding the low lying states of an ordered onedimensional spin glass with long-range 4-spin interaction given by the energy function [14, 10]:

$$E(\mathrm{s}) = \frac{1}{N}\sum_{k=1}^{N-1}\Big(\sum_{i=1}^{N-k} s_i s_{i+k}\Big)^2 \; . \tag{7}$$

The problem is very hard, the deep minima of $E(\mathrm{s})$ being presumably of "golf-hole like" character [14]. The exact ground states are known only for

$N \leq 32$, within a restricted class of configurations up to $N = 59$. For larger system sizes the best results obtained by algorithms based on simulated annealing (SA), differ by a factor of almost 2 from the ground state energies expected on the basis of ergodicity arguments [14, 13, 15] as well as from an extrapolation of the small-N energies. With $\alpha = 0$ (corresponding to the long-range character of the problem, $K_{\max} = N$) the algorithm presented above easily recovers the exact results for small N and makes a considerable step beyond SA results for larger N: In a series of 25 runs each over 100000 time steps, it found for $N = 101$ an *average* minimum energy per spin of 0.079 [see Fig. 2], the best configuration given by $\mathrm{s}_{\min} = 6122122212171112112344131121121245111113322323111 11$. Here, each digit denotes the number of consecutive spins with the same orientation. The energy of $\mathrm{s}_{\min}$ is $E_{\min}/N = 0.059798...$, corresponding to 6 σ below the mean local energy minimum ($m = 0.137, \sigma \approx 0.013$). Note that the number of Monte Carlo steps necessary to reach the above mentioned energies within SA can be estimated to be 10^{19} for $E/N = 0.08$ and 10^{200} for $E/N = 0.06$ [14]!

3. As another example with long-range interaction I treated a random Coulomb system. Specifically, I considered 500 points randomly distributed within the unit cube by applying the NAG-routine g05caf after an initialization of the random number generator by a call of g05cbf(11) [16]. In the spirit of the Coulomb glass problem [17] I looked for the distribution of positive and negative unit charges realizing the global minimum of the energy function

$$E(\mathrm{s}) = 2 \sum_{i=2}^{N} \sum_{j=1}^{i} \frac{s_i s_j}{|\mathbf{r}_i - \mathbf{r}_j|} \,. \tag{8}$$

The positions of 25 randomly chosen local minima for each value of the neighborhood size K are plotted in Fig. 3. One can see the existence of an optimal patch size in Kauffman's sense of the order of 3, followed by a counterreaction between sizes of 20 – 40, and a slow relaxation towards the local minima of $E(\mathrm{s})$. The latter have a mean value of $m = -8490$, with a standard deviation σ of about 80. Fig. 3 also shows a typical time evolution generated by the optimization algorithm (with $\alpha = 0$), reaching energies below -9000 after $\approx 3 \times 10^4$ fitter steps. The best value found by the algorithm, was -9094, again more than 7 σ below m.

I conclude that for spin glasses with < 500 sites the algorithm is capable of finding, after a few 10^4 time steps, energies that are about 7 standard deviations below the mean local energy minimum.

Another extensively investigated class of hard combinatorial problems is that of TSP. The aim is to find the shortest cyclic tour connecting a given number of points ("cities"). Characterizing a tour by the order the cities are visited, for a problem with N cities each tour may be represented as a permutations of $\{1,\ldots,N\}$, the total number of possible configurations being equal to $N!$. Given any configuration, another one can be obtained by cutting the tour at two segments (connecting, say, sites i and j, and sites k and l) and inverting one of the parts, i. e., replacing the configuration $\ldots,i-1,i,j,j+1,\ldots,k-1,k,l,l+1,\ldots$ by $\ldots,i-1,i,k,k-1,\ldots,j+1,j,l,l+1,\ldots$ [18]. The "energy" to be minimized is the length of the tour defined as the sum of distances d_{ij} between the cities:

$$E=\sum_{n=1}^{N} d_{nn+1}=\frac{1}{2}\sum_{n=1}^{N}(d_{nn-1}+d_{nn+1})\,, \tag{9}$$

where the numbering is assumed to go along the tour, and all indices are to be taken modulo N. In line with the algorithm, single-site energies $E^{[1]}_{\{n\}}$ are defined by $E^{[1]}_{\{n\}}=\frac{1}{2}(d_{nn-1}+d_{nn+1})$, so that $E=\sum_{n=1}^{N}E^{[1]}_{\{n\}}$. Obviously, any configuration change of the type mentioned above affects only four sites (in the above representation, those with indices i,j,k, and l) so that, for any given flip, the energy changes on 15 landscapes have to be considered (one 4-site, four 3-site, six 2-site and four 1-site energies). One has, therefore, $K_{\max}=4$, and I considered all 15 landscapes in parallel; a good choice for α turned out to be 2.5. On the other hand, at each moment of time the number of possible flips is large $(N\times(N-2)/2)$ so that I preferred the trial flip evolution to the fitter one.

In order to test the algorithm, I took three problems for which the optimal configuration is known [19]:
1. The 100 city tour kroA100. In a series of 25 runs the algorithm found the true optimum within 4×10^5 to 3×10^7 trial flips. These numbers can be improved by adopting the heuristic rule [20] that configuration changes are promising only if all affected cities are rather close to each other in the plane.

Restricting myself to trial flips for which the site with index k is one of the nb cities located closest to site i I obtained good results already for $nb = 7$: The algorithm finds extremely fast configurations close to the optimal one at the 1 % level and needs between 1×10^5 and 12×10^6 trial flips (with a mean value of about 3×10^6) in order to find the global minimum. This is illustrated in Fig. 4 where the resulting curves are compared to one obtained by another recently proposed Monte Carlo technique [4].
2. Grötschel's 442-problem (pcb442). The exact optimum is of length 50,783 (in the notation of Ref. [19]). With the same parameters as above, the best tour found by the algorithm within a series of 10 runs, each continuing over 2×10^7 trial flips was of length 50,936, slightly improving the results reported in Ref. [20].
3. The 532-city problem of Padberg and Rinaldi [21] (att532) where the optimal solution is of length 27,686. The algorithm finds 27,802, less than 0.4% above the true minimum.

In summary, a new optimization strategy for finding optimal or near-optimal solutions to problems with many competing requirements has been proposed. It is based on the *simultaneous* optimization of the fitness, cost or energy landscape of the system and of the individual sites forming it — with a weight function realizing a balance between interests of different scales. On its basis, a Monte Carlo algorithm has been formulated and tested for a number of SG and of TSP. The proposed algorithm turns out to be superior to other Monte Carlo techniques in quality and speed, reaching its most significant advantage for very hard problems. Coming back to the consideration of social systems the algorithm illustrates why democratic structures are effective in finding optimal solutions to complex problems and how an optimal relation between individual and global interests could look like.

Acknowledgment: I am indebted to Stuart Kauffman for drawing my attention to the N–k model and for the stimulating discussion of Ref. [9].

References

* Electronic address: dittes@fz-rossendorf.de

[1] S. Kirkpatrick, D. C. Gellat, and M. P. Vecchi, Science **220**, 671 (1983).

[2] H. Szu and R. Hartley, Phys. Lett. A **122**, 157 (1987).

[3] T. J. P. Penna, Phys. Rev. E **51**, R1 (1995).

[4] B. Hesselbo and R. B. Stinchcombe, Phys. Rev. Lett. **74**, 2151 (1995).

[5] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[6] I. Rechenberg, *Evolutionsstrategie* (Friedrich Frommann Verlag, Stuttgart, 1973).

[7] J. Holland, *Adaption in natural and artificial systems* (MIT Press, Cambridge, Mass., 1992).

[8] J. J. Hopfield and D. W. Tank, Biol. Cybern. **5**, 141 (1985).

[9] S. Kauffman, W. G. Macready, and E. Dickinson, Preprint SFI-94-06-031, Santa Fe, 1994; S. Kauffman, *At Home in the Universe* (Oxford University Press, Oxford, England, 1995).

[10] W. Krauth and M. Mézard, Z. Phys. B **97**, 127 (1995).

[11] S. Kauffman, *The Origins of Order* (Oxford University Press, Oxford, England, 1993).

[12] E. D. Weinberger, Phys. Rev. A **44**, 6399 (1991).

[13] M. J. E. Golay, IEEE Trans. Inform. Theory **IT-23**, 43 (1977); **IT-28**, 543 (1982).

[14] J. Bernasconi, J. Physique **48**, 559 (1987).

[15] J. P. Bouchaud and M. Mézard, Preprint LPTENS 94/14.

[16] The Numerical Algorithms Group Fortran Library, Mark 16.

[17] M. Pollak and M. Ortuño, in *Electron-electron interactions in disordered systems*, edited by A. L. Efros and M. Pollak (North Holland, Amsterdam, 1985), p. 287; B. I. Shklovskii and A. L. Efros, *Electronic properties of doped semiconductors* (Springer, Berlin, 1984).

[18] S. Lin and B. W. Kernighan, Oper. Res. **21**, 498 (1973).

[19] The TSP library may be obtained by anonymous ftp to titan.cs.rice.edu as the file 'tsplib.tar' in the directory 'public'.

[20] G. Dueck and T. Scheuer, J. Comput. Phys. **90**, 161 (1990).

[21] M. Padberg and G. Rinaldi, Oper. Res. Lett. **6**, 1 (1987).

Figure captions

Fig. 1
Best-so-far energies of 25 runs for the $N-k$ model (6) with $N=24^2$ and $k=12$. The full straight line indicates the value of the mean local minimum, the dashed one the best result of Ref. [11].

Fig. 2
Maximum, mean (over 25 runs) and minimum values of the best-so-far energies for the LABS model (7) with $N=101$. The straight lines give (from above to below) the value of the mean local minimum, the best result achieved by simulated annealing [14, 10] and the expected position of the global minimum for large N [14, 13, 15].

Fig. 3
Points: Values of 25 local minima of the Coulomb glass (8) as a function of the neighborhood size K. Curve: Time evolution of a single run of the algorithm for this energy function. Note the typical "Manhattan" panorama.

Fig. 4
Best-so-far length as a function of time for the TSP problem kroA100. A total of 25 runs is displayed and compared to the curve published in Ref. [4].

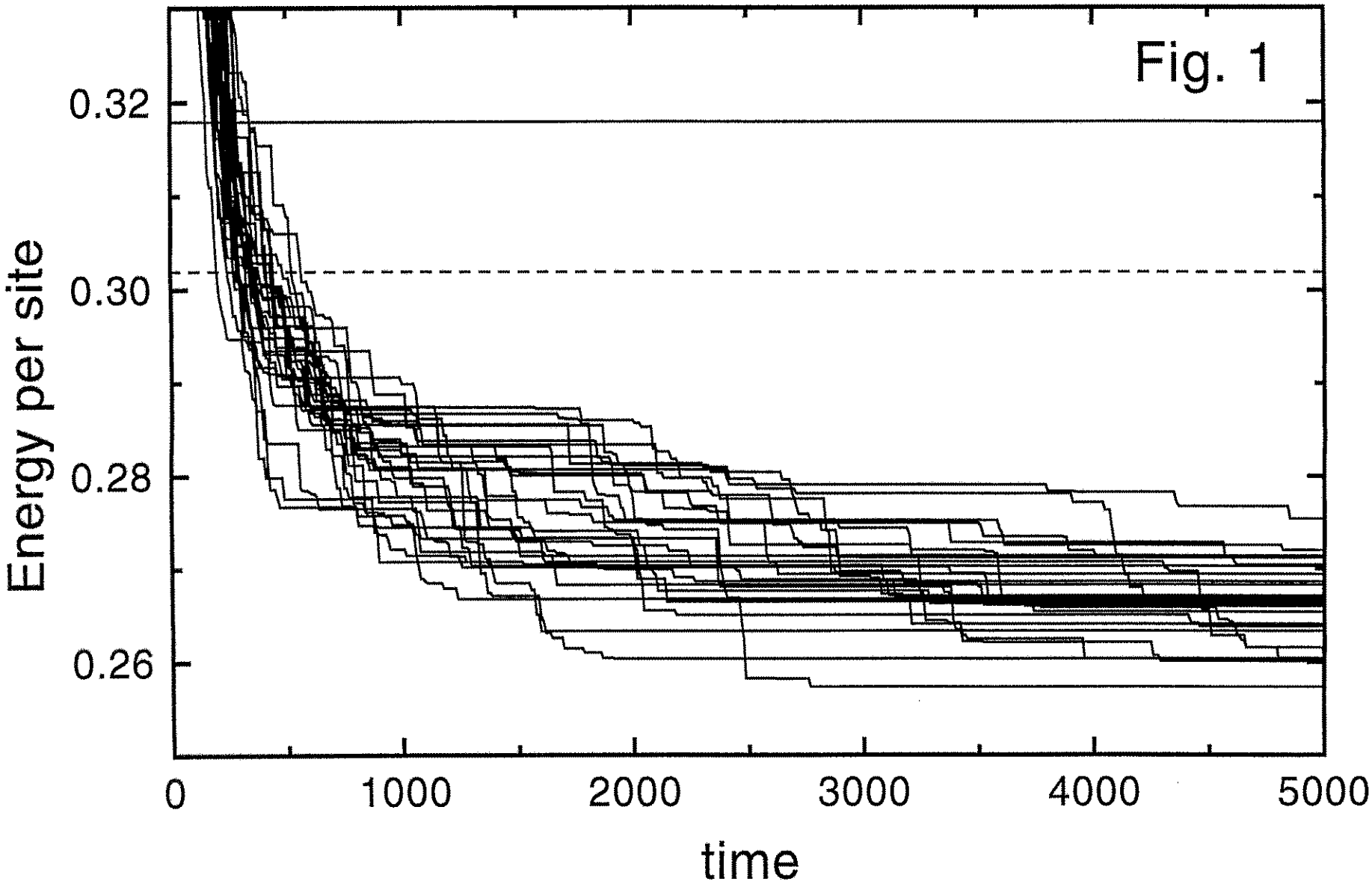


Fig. 1
0.32
0.30
0.28
0.26
Energy per site
0
1000
2000
3000
4000
5000
time

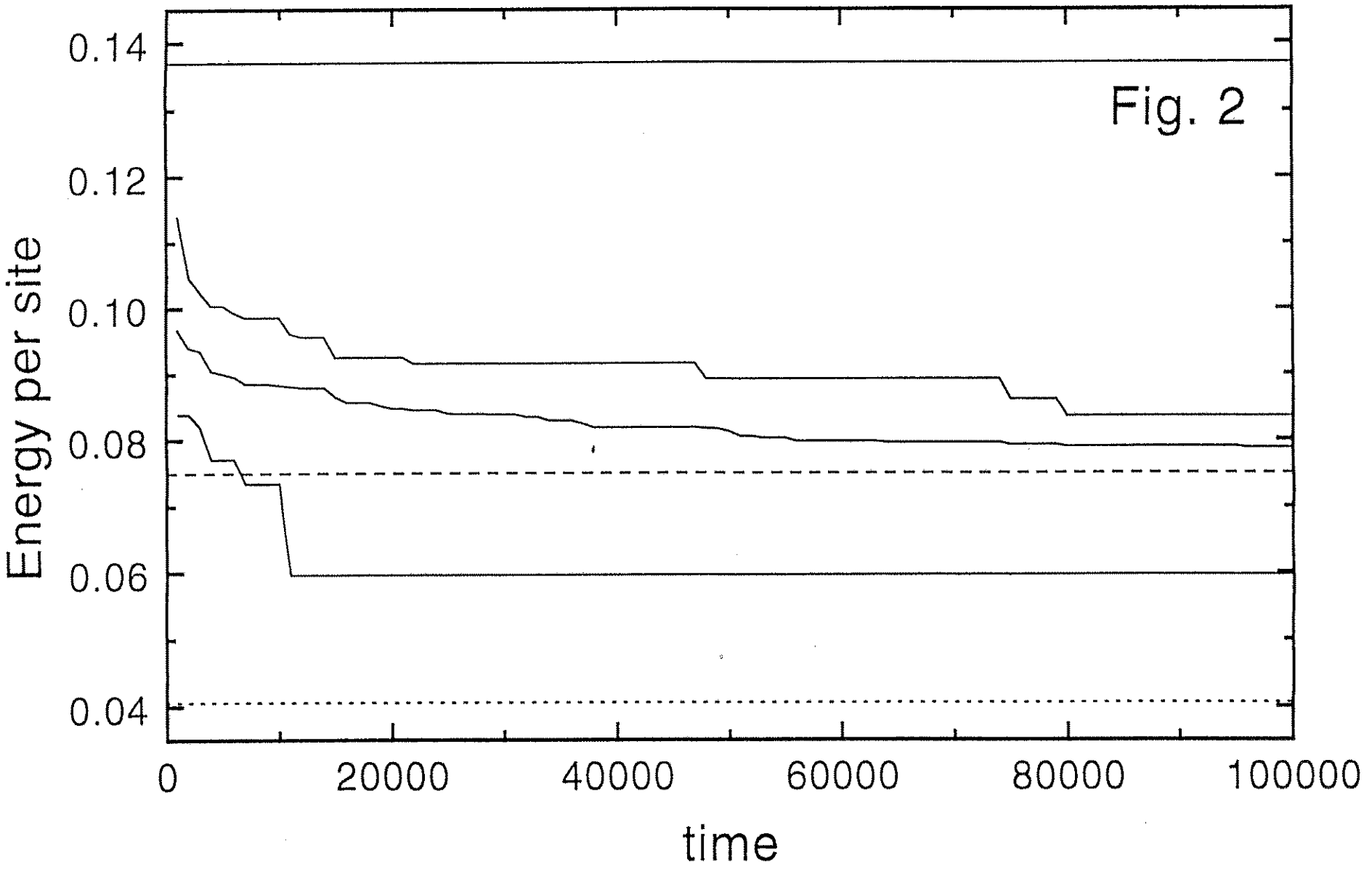
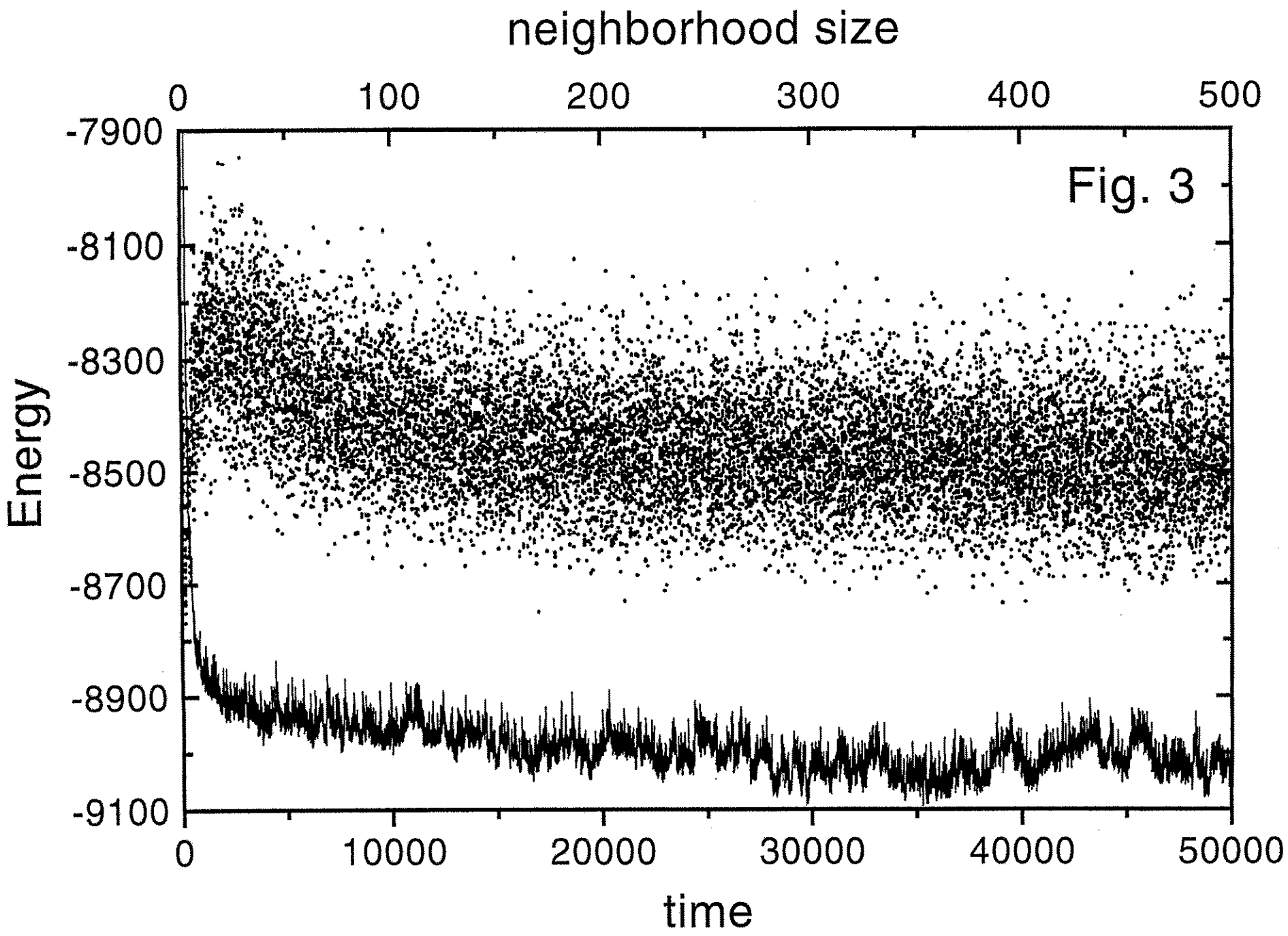


Fig. 2
Energy per site
0.14
0.12
0.10
0.08
0.06
0.04
0
20000
40000
60000
80000
100000
time


neighborhood size
0
100
200
300
400
500
-7900
-8100
-8300
-8500
-8700
-8900
-9100
Energy
Fig. 3
0
10000
20000
30000
40000
50000
time

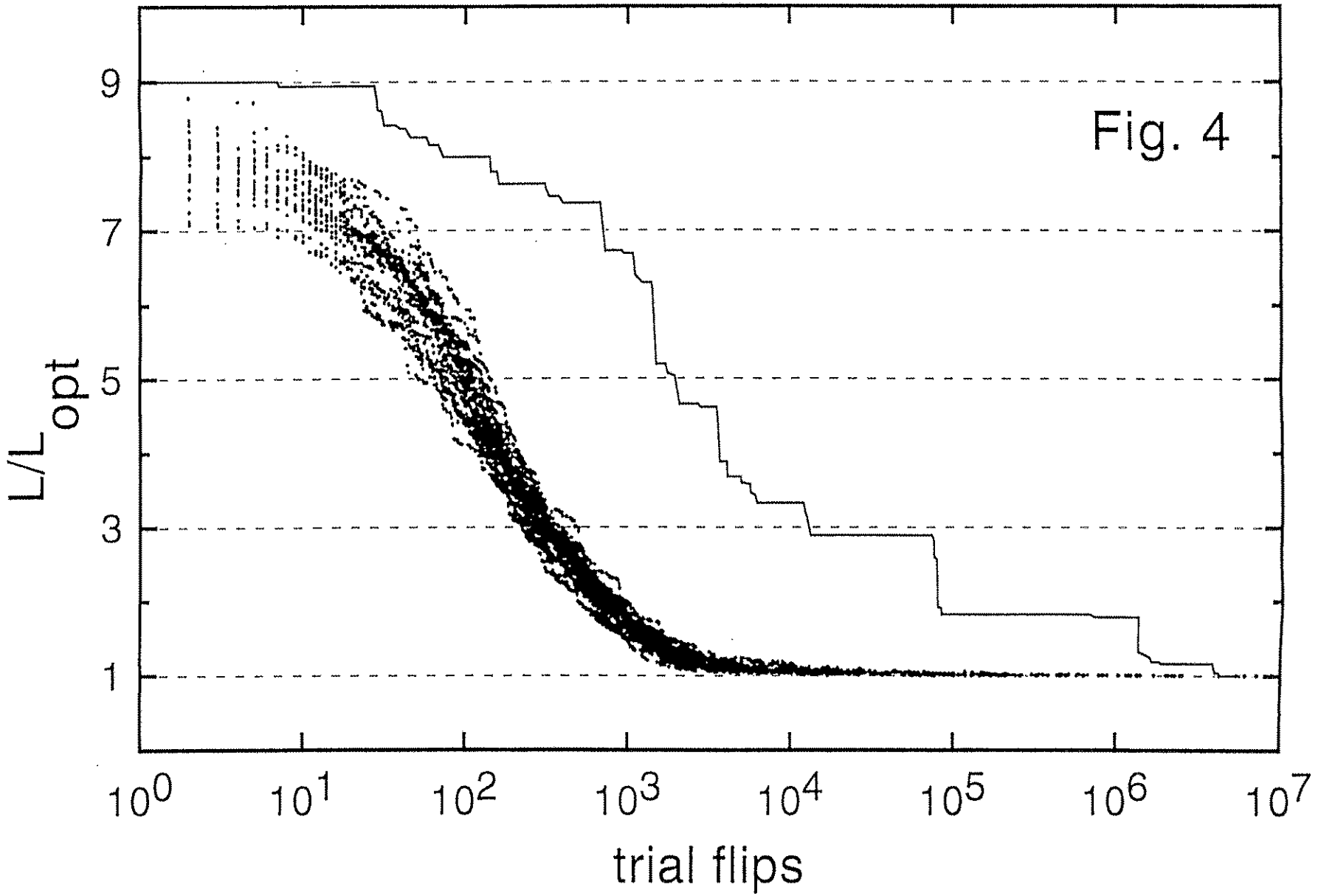


Fig. 4
9
7
5
3
1
L/L_opt
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
trial flips