Evolution in a rugged fitness landscape

Henrik Flyvbjerg and Benny Lautrup

Computational Neural Network Center, The Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

(Received 3 June 1992)

Kauffman's NK model for genetic evolution and adaption is analyzed for $K = N - 1$. In this case it describes adaptive walks on random fitness landscapes, and its dynamics is equivalent to the Metropolis algorithm for Derrida's random-energy model at zero temperature. We derive analytical expressions for the average length and duration of adaptive walks, and for the variance about these averages. The results are exact to leading order in N, the number of genes. We also find that the lengths of walks are Poisson distributed to leading order in $1/\ln N$, and that the duration of walks essentially is exponentially distributed to leading order in $1/N$.

PACS number(s): 87.10.+c, 05.40.+j, 05.50.+q

I. INTRODUCTION

Energy landscapes with many local minima are by now a well-studied subject in the statistical mechanics of spin glasses, and are currently studied in the statistical mechanics of protein folding. Similar landscapes of *fitness* with many local maxima instead of minima have also attracted attention in evolutionary biology and in computer science. Theoretical biologists use them in models of evolution. Computer scientists must deal with them, as they appear in combinatorial optimization problems, and in the training of neural networks — and use algorithms mimicking evolution to try to solve these hard optimization problems [1]. In the present paper we consider a proto-typical model suggested by Kauffman [2, 3]: the evolution of a general haploid organism with a single copy of chromosomes. Evolution is driven by random mutations of individual genes and takes place in a random fitness landscape. We choose to study this particular system, because we believe it is not so particular, after all: we expect it appears as the universal result of a coarse-grained description of a wide class of systems (Sec. II). Averaging over landscapes and evolutionary histories, we find probability distributions for the length (Secs. III–V) and duration (Secs. VI and VII) of evolutions to local fitness maxima.

Some of the analytical results presented below have been seen in numerical studies [4, 5], and derived in [6]. We have included them to make the presentation self-contained. Different but related results have been obtained for the NK model with general $K \gg 1$ in [7]. We developed our analytical understanding of the evolution of *isolated* species of the kind described below in order to obtain an analytical understanding of their *co*evolution in an "ecosystem" of *interacting* species, like the ecosystem described in [8]. Such an analytical understanding is much desirable, both by itself and as a guide for computer simulations, because simulations of an entire ecosystem are very demanding, and difficult to do with precision. An analytical understanding of one model can be used to understand what to expect, and where to look for it, in simulations of related models that cannot be solved analytically.

Using our analytical understanding of the evolution of species in isolation, we have been able to obtain several analytical results for coevolving species, as they are modeled with Kauffman's NKC model. Our results include the existence and location of a phase transition between chaotic and frozen dynamics, a natural order parameter for this transition, and the relaxation time of the system anywhere in the two phases. We have collected all necessary results on evolution in isolation in the present paper, and collected our results on coevolution in the accompanying paper [9]. The reader looking for results on the NKC model should be able to start with the second paper, and return to the present paper on the NK model only when referred to it for an explanation.

II. THE SYSTEM

We consider a very simple organism, whose properties are entirely given by N distinct genes. We do not distinguish between genotype and phenotype, and ignore genetic diversity within a population. So an entire species of this simple organism is represented by the N genes. Truly, diversity, maintained by mutation and trimmed by selection, provides the mechanism of evolution. But if selection proceeds much faster than mutations occur, then we may represent a population by a dominant type, as suggested by Gillespie [10], and ignore diversity, except as a reason to choose the particular dynamics we choose. We do not describe extinction of species, nor their proliferation through branching, with this simple model, though we could do so.

We consider an evolution driven by mutations of individual, randomly chosen genes. One mutation occurs per unit of time. If it leads to higher fitness, it is accepted, and the population changes accordingly, and is said to have evolved one step. If a mutation leads to lower fitness, it is rejected, and the population remains unchanged during that time step. Tie situations, with two genetic configurations having the same fitness, do not occur (have measure zero), due to the way we assign fitness to genetic configurations.

Each gene can occur in A different versions, or *alleles*. So the set of genes representing a species can occur in A^N different configurations. We call the set of these configurations *configuration space*. In the case of $A = 2$, the two possible alleles of each gene may be labeled 0 and 1, and configuration space may be visualized as the corners of the unit cube in N dimensions, each corner corresponding to a different possible organism, and corners separated by one edge being one-mutant neighbors. We shall assume $A = 2$ for the remainder of this paper with no essential loss of generality. We expect very similar results for any other value of A, as long as $A \ll N$. We shall often assume N is a large number, and expand in $1/N$ or take the limit $N \to \infty$ in order to simplify calculations and resulting expressions.

We assign a fixed fitness to each point in configuration space. This assignment defines a fitness *landscape*. We assume that this fitness landscape is maximally rugged: the fitness f of any point in configuration space is random, drawn from a continuous distribution $p(f)$, the same distribution p being used at all points. The particular distribution used does not matter; we shall not even bother to introduce it in our considerations below, because it turns out that it disappears again by a transformation of variables to $F = \int_{-\infty}^{f} df' \, p(f')$. In the case where p is uniform on the interval $0 \leq f \leq 1$, we have $f = F$. So for convenience we shall refer to F as the fitness, although F in the general case really denotes the probability for fitness less than f. The elimination of $p(f)$ in equations expresses that the value f of the fitness is irrelevant; only the probability F of being less fit matters.

We have two reasons to consider random fitness landscapes; the first reason is a conjecture, the second is proven correct in this and the accompanying paper:

(1) Evolution in any fitness landscape having an effectively finite correlation length will, when viewed at sufficiently coarse-grained scales of time and space (configuration space, i.e.), look like evolution in a random fitness landscape. So evolution in a random fitness landscape describes the large-scale behavior of evolution in a large class of landscapes. Consequently, with this choice of landscape we are avoiding the particular, while treating a quite general case.

(2) It is technically convenient: the absence of correlations allows us to derive a number of analytical results.

Notice that from a mathematical point of view, N might as well be the number of positions in the primary sequence of a protein, with $A = 20$ denoting the 20 amino acids that potentially could occur at each position. Or $A = 4$ could denote the 4 nucleotides possible at each site in a DNA sequence of length N.

Alternatively, we may think of the N genes and their A alleles as N Potts spins and their A possible values in an A-state Potts model. With $V = -f$ denoting the *energy* of a spin configuration, we recognize Derrida's random-energy model [11, 12]. In this language, the dynamics of mutations described above is the random-site Metropolis algorithm at zero temperature.

Viewed as a model for evolution by mutation and selection in a fitness landscape, the system just described

is a special case of Kauffman's NK model, the case of $K = N - 1$ [2, 3].

III. ESTIMATING THE LENGTH OF WALKS

Evolution traces out a path in configuration space. At each time step, the path is either extended one step from its current end point to a nearest neighbor — when a mutation leading to higher fitness is offered to and accepted by evolution — or the path is *not* extended — because a mutation leading to lower fitness is offered and rejected. This path is often referred to as an *adaptive walk*.

In this and the next two sections, we are not concerned with the temporal aspects of evolution, but only with the length ℓ of adaptive walks. This limitation simplifies the description a good deal. In subsequent sections temporal aspects are treated.

Before we get involved with mathematics, let us estimate the average length of adaptive walks, and the average fitness they lead to. The qualitative picture thus obtained is confirmed by rigorous calculations in the next section.

We assume N is large. The dimension of configuration space is N. We assume the length of adaptive walks is much smaller than $\sqrt{N}$, and find this assumption consistent with the results it leads to. Since the walk proceeds by random mutations, it proceeds in random directions in configuration space. There are many more directions than there are steps in the walk, by assumption. So each step in the walk has a different direction. In each step of the adaptive walk, the fitness F is increased. The value it increases to is uncorrelated (to leading order in $1/N$; see the Appendix for details) with its previous value, except it is larger, of course. Consequently, in each step $1 - F$ is halved, on the average. Thus, starting the walk with $F = 0$, after ℓ steps the average fitness is $1 - 2^{-\ell}$. An adaptive walk stops when all neighbor positions have lower fitness than the current position. Since fitnesses are random and uncorrelated, this happens when N independent random numbers happen to be smaller than F. On the average, this occurs when $1 - F \sim 1/N$. This is our estimate for the average final fitness, and, setting $1 - F \sim 2^{-\ell}$, we have an estimate for the average length of an adaptive walk:

$$\bar{\ell} \simeq \ln N / \ln 2. \tag{1}$$

In the derivation of this result, we neglected correlations between fluctuations around the averages that we worked with. They do not change the logarithmic dependence on N in Eq. (1), but do change the coefficient of $\ln N$.

In addition to a more precise result for the average length of adaptive walks, we want to know the probability distribution Q_ℓ for ℓ. In [5], "long upper tails containing little probability" were seen in numerical results for Q_ℓ. So one may wonder whether Q_ℓ decreases as a power of ℓ at large ℓ, or faster. The next sections fills in these gaps in our picture of adaptive walks.

IV. THE LENGTH OF WALKS

When duration is not of interest, but length is, the simplest quantity to work with is the probability density $p_\ell(F)$ that an adaptive walk contains (at least) ℓ steps, and has fitness F after these ℓ steps. Evolution by random mutations through fitter one-mutant neighbors can be described approximately by a recursion relation:

$$p_{\ell+1}(F) = \int_0^F dF' \frac{1 - F'^{N-1}}{1 - F'} p_\ell(F') \quad \text{for} \quad \ell = 1, 2, \ldots.$$

(2)

This recursion relation expresses that fitness F is acquired in $\ell + 1$ evolutionary steps by acquiring any lower fitness F' in ℓ steps, and taking one more step to fitness F. Taking the last step requires that not all N one-mutant neighbors in configuration space are less fit. One is — the one that was reached after $\ell - 1$ steps. The remaining $N - 1$ neighbors have fitness less than F', each with probability F', since their fitness is random. Here we assume that they were not probed previously by the path of evolution. This assumption is only approximately true, so Eq. (2) is an approximation. Within this approximation, the probability that not all neighbors are less fit is $1 - F'^{N-1}$. When this is the case, the $(\ell + 1)$th evolutionary step will be taken, and leads to any fitness above F' with equal probability, hence to fitness F in the interval dF' with probability $dF'/(1 - F')$.

The approximation we have made with Eq. (2) relies on N being large. While the power $N - 1$ on F' excludes evolutionary back-tracking, Eq. (2) does not exclude that the path of evolution *intersects* itself or visits other points in configuration space that it probed and rejected at an earlier time. Such intersections are forbidden by the dynamics, which forces the path to always higher degrees of fitness in a fixed landscape, or to stop at a local maximum. However, in Eq. (2), the $N-1$ one-mutant neighbors which are not a state's immediate predecessor in evolution are all treated as if they were never probed before by the evolutionary process. But some of them may have been, in which case we know that their fitness is lower than the current one. So Eq. (2) yields an upper bound for the true value of $p_\ell(F)$, because the exact relation has a power lower than or equal to $N - 1$, where Eq. (2) has $N-1$. This exact power depends on the entire path of evolution up to the currents state, so the approximation made with Eq. (2) causes a vast simplification of the problem. In the appendix we give arguments that this approximation is correct to leading order in an expansion in $1/N$.

In view of the further approximation considered below, all we really need are results to leading order in N. But since we can solve Eq. (2) as it stands — i.e., with backtracking forbidden, and self-intersection permitted — we shall do that for definiteness.

Introducing the monotonic function

$$H_N(F) = \sum_{k=1}^N \frac{1}{k} F^k,$$

(3)

a change of variable to $H = H_{N-1}(F)$ in Eq. (2) gives

$$p_{\ell+1}(H) = \int_0^H dH' \, p_\ell(H') \quad \text{for} \quad \ell = 1, 2, \ldots,$$

(4)

which is easily iterated to give

$$p_\ell(H) = \frac{1}{(\ell - 1)!} \int_0^H dH' \, (H - H')^{\ell-1} p_1(H').$$

(5)

For definiteness and notational convenience, we let all adaptive walks begin in the least-fit state, characterized by $F = 0$. Since there is zero probability for this state being a local maximum of fitness, the first step of the adaptive walk is always taken. For notational convenience, we let ℓ denote the number of steps taken *in excess* of this first step. Then the initial condition reads as

$$p_{-1}(F) = \delta(F).$$

(6)

This rather eccentric choice of initial condition assures that the walk has a predecessor for all values of $\ell \geq 0$. This makes formulas look simpler, and makes Eqs. (2) and (4) valid also for $\ell = -1$. They have the unique solution

$$p_\ell(F) = \frac{1}{\ell!} H_{N-1}(F)^\ell \quad \text{for} \quad \ell = 0, 1, 2, \ldots.$$

(7)

Obviously, for fixed $F < 1$

$$H_N(F) \to -\ln(1 - F) \quad \text{for} \quad N \to \infty,$$

(8)

while for $F = 1$, $H_N(1)$ are the *harmonic numbers* discussed by Knuth in [13],

$$H_N(1) = \sum_{k=1}^N \frac{1}{k} = \psi(N+1) + \gamma_E = \ln N + \gamma_E + O(N^{-1}),$$

(9)

where $\psi(x) = d \ln \Gamma(x)/dx$, and $\gamma_E = 0.577\,215\,66\ldots$ is Euler's constant. For general F we note that

$$H_N(F) = \text{li}(F^{N+1}) - \ln[-\ln(F)] + O(N^{-1}),$$

(10)

where li is the logarithmic integral. We shall need that

$$H_N(1 - x/N) = \ln N + \gamma_E - \text{Ein}(x) + O(N^{-1})$$

$$\text{for} \quad x \sim O(1),$$

(11)

where $\text{Ein}(x)$ is an entire function related to the exponential integral [14]:

$$\text{Ein}(x) = \int_0^x dt \frac{1 - e^{-t}}{t} = \text{E}_1(x) + \ln x + \gamma_E.$$

(12)

As stated above, Eq. (2) is the simplest relation we can write down for a probability describing the length of the adaptive walks considered here, in the approximation specified. The probability that a walk contains (at least) ℓ steps is obtained from $p_\ell(F)$ by integration over F:

$$P_\ell = \int_0^1 dF\, p_\ell(F) = \frac{1}{\ell!} \int_0^1 dF\, H_{N-1}(F)^\ell$$

$$= \frac{1}{\ell!} \sum_{k_1,\ldots,k_\ell=1}^{N-1} \frac{1}{k_1 \cdots k_\ell (k_1 + \cdots + k_\ell + 1)}$$

$$\text{for} \quad \ell = 0, 1, 2, \ldots. \quad (13)$$

Integration over F on both sides in Eq. (2) gives

$$P_{\ell+1} = P_\ell - \int_0^1 dF\, F^{N-1} p_\ell(F), \quad (14)$$

which obviously cannot be made into a closed equation for P_ℓ. The remaining integral in Eq. (14) is the probability that an adaptive walk contains *exactly* ℓ steps. This is a quantity of interest. We introduce the notation Q_ℓ for it, and $q_\ell(F)$ for the corresponding probability density that a walk stops with fitness F after exactly ℓ steps:

$$q_\ell(F) = F^{N-1} p_\ell(F) = F^{N-1} \frac{1}{\ell!} H_{N-1}(F)^\ell$$

$$\text{for} \quad \ell = 0, 1, 2, \ldots, \quad (15)$$

$$Q_\ell = \int_0^1 dF\, q_\ell(F) = \frac{1}{\ell!} \int_0^1 dF\, F^{N-1} H_{N-1}(F)^\ell$$

$$= \frac{1}{\ell!} \sum_{k_1,\ldots,k_\ell=1}^{N-1} \frac{1}{k_1 \cdots k_\ell (k_1 + \cdots + k_\ell + N)}$$

$$\text{for} \quad \ell = 0, 1, 2, \ldots, \quad (16)$$

From Eq. (14) follows

$$Q_\ell = P_\ell - P_{\ell+1}. \quad (17)$$

Since Eq. (6) implies

$$Q_{-1} = 0, \quad P_0 = 1, \quad (18)$$

normalization of $q_\ell(F)$ and Q_ℓ follows trivially from Eq. (17):

$$\sum_{\ell=0}^\infty \int_0^1 dF\, q_\ell(F) = \sum_{\ell=0}^\infty Q_\ell = \sum_{\ell=0}^\infty (P_\ell - P_{\ell+1}) = P_0 = 1. \quad (19)$$

Here we have used $\lim_{\ell \to \infty} P_\ell = 0$, and we have set the upper limit on the sum over ℓ to infinity for convenience. Strictly speaking, this upper limit is A^N, the number of points in configuration space. We shall see below that typical values for ℓ are of order $\ln N$, and much larger values of ℓ occur with probabilities that are more than exponentially suppressed. So the effect of this change in the upper limit is truly negligible.

Inserting Eq. (15) in Eq. (19) and summing over ℓ, we see that normalization means

$$\int_0^1 dF\, F^{N-1} \exp[H_{N-1}(F)] = 1 \quad (20)$$

for any positive integer N. This identity is exact, and may also be proven directly; we leave that for the reader's entertainment.

The generating function for the probabilities Q_ℓ reads as

$$\tilde{Q}(\lambda) = \sum_{\ell=0}^\infty \lambda^\ell Q_\ell = \int_0^1 dF\, F^{N-1} \exp[\lambda H_{N-1}(F)]. \quad (21)$$

Despite our ability to evaluate the integral in Eq. (20) we have not been able to evaluate the integral in Eq. (21) for general λ. But as we have already neglected terms of subleading order in $1/N$, we may continue to do so with no further loss of generality. To this end we write $F = 1 - x/N$, and observe that $F^N = \exp(-x) + O(x^2/N)$. Consequently, the integrand in Eq. (21) is negligible unless $x \sim 1$, and, to leading order in $1/N$, we have for $\tilde{Q}(\lambda)$, Q_ℓ, and its first moment μ_1:

$$\tilde{Q}(\lambda) = N^{\lambda-1} \int_0^\infty dx\, e^{-x - \lambda(\text{Ein}(x) - \gamma_E)}, \quad (22)$$

$$\tilde{Q}(1) = 1, \quad (23)$$

$$Q_\ell = \frac{1}{N\ell!} \int_0^\infty dx\, e^{-x} [\ln N + \gamma_E - \text{Ein}(x)]^\ell$$

$$= \frac{1}{N\ell!} \left\{ (\ln N)^\ell + O[(\ln N)^{\ell-1}] \right\}, \quad (24)$$

$$\mu_1 = \sum_{\ell=0}^\infty \ell Q_\ell = \frac{dQ}{d\lambda}(1) = \ln N + \mu_1^{(\text{finite})} + O(N^{-1}). \quad (25)$$

Here

$$\mu_1^{(\text{finite})} = \int_0^\infty dx\, [\gamma_E - \text{Ein}(x)]\, e^{-x + \gamma_E - \text{Ein}(x)}$$

$$= 0.099\,13... \quad (26)$$

is a constant that we have not been able to express in terms of known constants. Our result for μ_1 — and for μ_2 given in Eq. (34), agree with the two-digit numerical results given in [5].

Equation (24) shows that to leading order in $\ln N$, Q_ℓ is a Poisson distribution. This simple results has a simple explanation: the Poisson distribution is obtained because all adaptive walks terminate with essentially the same fitness F. F belongs to an interval of width $\sim 1/N$ at $F = 1$. This is seen from our rewriting $\int^1 dF\, F^{N-1}$ as $(1/N) \int_0 dx \exp(-x)$. Thus, in the interval $[0,1]$, NF^{N-1} is almost a δ function with support at $F \simeq 1$. If we replace it with that in the formulas above, we arrive at a Poisson distribution.

V. Q_ℓ'S POISSON BEHAVIOR

In this section we elaborate on Q_ℓ's similarity with a Poisson distribution, and compare it with such distributions for various values of N.

With the notation

$$\langle \cdots \rangle = \int_0^1 dF\, q(F) \cdots, \quad (27)$$

where

$$q(F) = \sum_{\ell=0}^{\infty} q_\ell(F) = F^{N-1} e^{H_{N-1}(F)} \qquad (28)$$

is the probability density that an adaptive walk terminates at a local fitness maximum with fitness F, we have a positive measure on the fitness interval $[0, 1]$. Equation (20) shows that this measure is normalized. We write the integral in Eq. (21) in terms of this measure and cumulant-expand it:

$$
\begin{aligned}
\tilde{Q}(\lambda) &= \langle e^{(\lambda-1)H_{N-1}} \rangle \\
&= \exp\left(\langle e^{(\lambda-1)H_{N-1}} - 1 \rangle_c \right) \\
&= \exp\left((\lambda-1)\mu_1 + \frac{1}{2}(\lambda-1)^2 \langle H_{N-1}^2 \rangle_c \right. \\
&\qquad \left. + \frac{1}{3!}(\lambda-1)^3 \langle H_{N-1}^3 \rangle_c \cdots \right),
\end{aligned}
\qquad (29)
$$

where the first cumulants are

$$\langle H_{N-1} \rangle_c = \langle H_{N-1} \rangle = \mu_1 = \ln N + \mu_1^{(\text{finite})} + O(N^{-1}), \qquad (30)$$

$$\langle H_{N-1}^2 \rangle_c = \langle (H_{N-1} - \mu_1)^2 \rangle = 0.167\,33... + O(N^{-1}), \qquad (31)$$

$$\langle H_{N-1}^3 \rangle_c = \langle (H_{N-1} - \mu_1)^3 \rangle = -0.083\,70... + O(N^{-1}), \qquad (32)$$

$$
\begin{aligned}
\langle H_{N-1}^4 \rangle_c &= \langle (H_{N-1} - \mu_1)^4 \rangle - 3\langle (H_{N-1} - \mu_1)^2 \rangle^2 \\
&= 0.038\,15... + O(N^{-1}),
\end{aligned}
\qquad (33)
$$

$$\vdots$$

Here we have used that to leading order in $1/N$ these expectation values receive contributions only from values of F obeying $F = 1 - x/N$ with $x \sim 1$, i.e., where $H_{N-1}(F) - \mu_1 = \gamma_E - \text{Ein}(x) - \mu_1^{(\text{finite})}$. Consequently, all cumulants beyond the first are ~ 1, while the first, μ_1, is $\sim \ln N$. Neglecting cumulants higher than the first in Eq. (29), we arrive at the generating function for a Poisson distribution with the same mean, μ_1, as Q_ℓ has:

$$\tilde{Q}(\lambda) = \exp[(\lambda - 1)\mu_1]. \qquad (34)$$

Figure 1 shows Q_ℓ against ℓ for $N = 10, 100, 1000$, and $10\,000$ as open symbols connected by lines. The lines are only meant to guide the eye. Q_ℓ was found by numerical integration of the expressions for Q_ℓ deriving from Eq. (21). The values for the Poisson distributions with the same mean values are shown as solid circles, which in most cases fall within the open symbols. This agreement is rather striking. It is *not* just due to the central limit theorem making both Q_ℓ and the Poisson distribution well approximated by the same Gaussian distribution, and hence by each other. This is illustrated in Fig. 1 for the case of $N = 10$: the dashed line shows the Gaussian distribution with the same mean and variance as Q_ℓ has. Clearly, it does not approximate Q_ℓ, shown as open circles, as well as the Poisson distribution with the same mean as Q_ℓ, shown as solid circles. In addition to that it has non-negligible support for negative values of ℓ.

We can also compare Q_ℓ's moments, μ_n, with the moments of the Poisson distribution with the same mean, μ_1:

$$\mu_1 = \langle H_{N-1} \rangle = \ln N + \mu_1^{(\text{finite})}, \qquad (35)$$

$$\mu_2 = \mu_1 + \langle H_{N-1}^2 \rangle_c = \mu_1 + 0.167\,33..., \qquad (36)$$

$$\mu_3 = \mu_1 + \langle H_{N-1}^2 \rangle_c + \langle H_{N-1}^3 \rangle_c = \mu_1 + 0.083\,63..., \qquad (37)$$

$$
\begin{aligned}
\mu_4 &= \mu_1 + \langle H_{N-1}^2 \rangle_c + 3(\mu_1 + \langle H_{N-1}^2 \rangle_c)^2 + \langle H_{N-1}^3 \rangle_c \\
&\quad + \langle H_{N-1}^4 \rangle_c \\
&= \mu_1 + 0.167\,33... + 3(\mu_1 + 0.167\,33...)^2 + 0.122\,15... \,.
\end{aligned}
\qquad (38)
$$

As expected from Eq. (29), we see that when we neglect cumulants beyond the first, the nth moment, μ_n, depends on the first moment, μ_1, as the nth moment of a Poisson distribution does. We also see that this neglect introduces an error of just a few percent in the moments shown for $N \geq 100$. We expect this error to increase with the order n of the moment μ_n, and know that it decreases as $1/\ln N$.
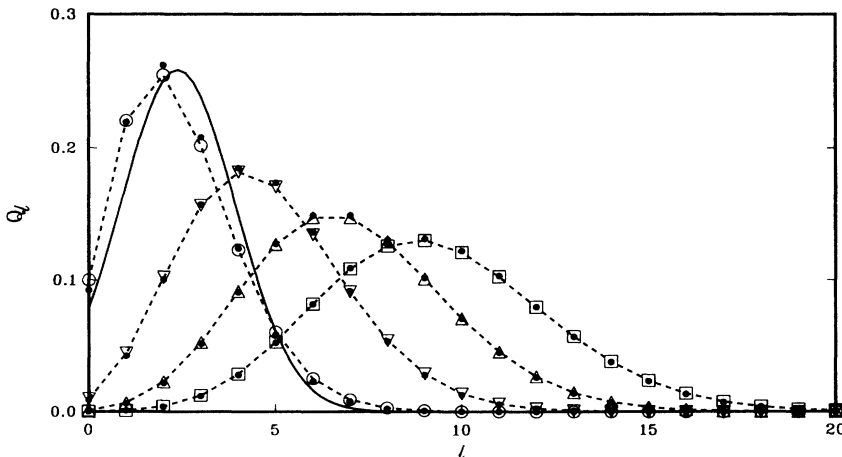

FIG. 1. Q_ℓ vs ℓ for $N = 10$ (○), 100 (▽), 1000 (△), and $10\,000$ (□). The connecting dashed lines are only meant to guide the eye. Poisson distributions with the same mean values are shown with the symbol ●. In the case of $N = 10$, the Gaussian distribution with the same mean and variance as Q_ℓ is shown as a solid line.

VI. ESTIMATING THE DURATION OF WALKS

Since we let the adaptive walk start out with fitness $F = 0$, the probability Q_0 that it is at a local fitness maximum at time $t = 0$ after the first (uncounted) step is

$$Q_0 = 1/N. \tag{39}$$

This is a rigorous result.

On the average and to leading order in $1/N$, each step taken, including the first, reduces $1 - F$ by a factor 2. Each step thereby doubles the probability that the ensuing step will be the last, while it halves the probability per unit of time that the next step is taken. Consequently, the probability per unit of time for the walk to terminate is constant during the walk. This means

$$Q_t = \frac{1}{\bar{t}} \exp(-t/\bar{t}). \tag{40}$$

Using the exact result in Eq. (39), we have the estimates

$$\bar{t} = N \tag{41}$$

and

$$Q_t = \frac{1}{N} \exp(-t/N). \tag{42}$$

This last equation shows that NQ_t remains a finite function of t/N in the limit $N \to \infty$, and its kth moment is proportional to N^k. In particular, we see that the standard deviation

$$\sigma(t) = N \tag{43}$$

scales like the average $\bar{t}$. This is in contrast to the scaling laws found for the average *length* of walks and *its* standard deviation.

In the next section we show how this section's estimates are modified when we account properly for fluctuations and their correlations.

VII. THE DURATION OF WALKS

Let $p_{\ell,M;t}(F)$ denote the probability that an adaptive walk at time t has proceeded ℓ steps, thereby reaching a point in configuration space having fitness F and M less-fit neighbors. The time evolution of $p_{\ell,M;t}(F)$ is found as follows. As above, we neglect the fact that an adaptive walk cannot intersect itself or any site that was previously probed by its evolution and discarded for being less fit. As explained in the Appendix, this is a leading order approximation in an expansion in $1/N$. Within this approximation, but explicitly forbidding back-tracking, $p_{\ell,M;t}(F)$'s development in time is given by

$$p_{\ell,M;t+1}(F) = \frac{M}{N} p_{\ell,M;t}(f) + B_{M-1,N-1}(F)\Phi_{\ell-1;t}(F), \tag{44}$$

where

$$\Phi_{\ell;t}(F) = \int_0^F dF' \, \phi_{\ell;t}(F') \tag{45}$$

and

$$\phi_{\ell;t}(F) = \frac{1}{1-F} \sum_{M=0}^{N} \left(1 - \frac{M}{N}\right) p_{\ell,M;t}(F). \tag{46}$$

Equation (44) expresses that an adaptive walk has length ℓ, fitness F, and number of less-fit neighbors M at time $t+1$ for one of two mutually exclusive reasons: it was either characterized by these values at time t, and took no step between time t and time $t + 1$ — this happens with the probability given as the first term on the right-hand side in Eq. (44) — *or* a step *was* taken between time t and time $t + 1$, and the adaptive walk arrived at values (ℓ, F, M) with that step — this happens with the probability given as the second term on the right-hand-side of Eq. (44). $\Phi_{\ell-1;t}(F)$ is the transition probability density at time t to fitness F from less-fit one-mutant neighbor configurations arrived at in $\ell - 1$ steps. It is an integral over $F' < F$ of $\phi_{\ell-1;t}(F')$, the transition probability density at time t from fitness F' arrived at in $\ell - 1$ steps to any more-fit one-mutant neighbor configuration.

A configuration with fitness F, arrived at from a less-fit configuration, will have a total of M less-fit neighbor configurations, when $M - 1$ of the $N - 1$ new neighbor configurations are less fit. This happens with binomially distributed probability,

$$B_{M-1,N-1}(F) = \binom{N-1}{M-1} F^{M-1}(1 - F)^{N-M}, \tag{47}$$

when we treat the landscape's quenched randomness as if the one-mutant neighborhood of any configuration arrived at is "annealed," thereby allowing the adaptive walk to self-intersect, with the exception that backtracking remains forbidden.

As initial condition for Eq. (44) we choose as before, with no essential loss of generality, to let the adaptive walk start out in the least-fit configuration, at a time that is chosen to be -1 for notational convenience. We let ℓ denote the number of steps taken *in excess* to the first step, which is always taken. Then the initial condition reads as

$$p_{\ell,M;-1}(F) = \delta_{\ell,-1} \, \delta_{M,0} \, \delta(F). \tag{48}$$

Inserted in Eq. (44), this initial condition gives the equivalent initial condition

$$p_{\ell,M;0}(F) = B_{M-1,N-1}(F) \, \delta_{\ell,0}. \tag{49}$$

Again our rather eccentric choice of initial condition assures that the walk has a predecessor at $t = 0$, as at all later times. This makes formulas look simpler.

The simpler Eq. (2) is contained in Eq. (44): the probability that an adaptive walk reaches length ℓ and fitness F at time t is $\Phi_{\ell-1;t-1}(F)$. Consequently, the probability that it reaches length ℓ and fitness F at all, denoted $p_\ell(F)$ in Sec. IV, is

$$p_\ell(F) = \sum_{t=0}^{\infty} \Phi_{\ell-1;t-1}(F) \quad \text{for} \quad \ell \geq 0. \tag{50}$$

Using this with Eq. (44), one obtains an equation for

$p_\ell(F)$, Eq. (2).

Equation (44) is a linear integro-difference equation. The fact that it is nonlocal in F does not prevent its solution, since it can be made local by appropriate differentiation after F. Introducing the generating function

$$\tilde{p}_M(\lambda, F; \tau) = \sum_{\ell=0}^{\infty} \lambda^\ell \sum_{t=0}^{\infty} \tau^t \, p_{\ell, M; t}(f) \tag{51}$$

and the corresponding generating functions for transition probability densities

$$\tilde{\phi}(\lambda, F; \tau) = \frac{1}{1 - F} \sum_{M=0}^{N} (1 - M/N) \tilde{p}_M(\lambda, F; \tau) \tag{52}$$

and

$$\tilde{\Phi}(\lambda, F; \tau) = \int_0^F dF' \, \tilde{\phi}(\lambda, F'; \tau), \tag{53}$$

the initial condition, Eq. (49), reads as

$$\tilde{p}_M(\lambda, F; 0) = B_{M-1, N-1}(F) \tag{54}$$

and Eq. (44) itself, after a minor rearrangement, reads

$$\tilde{p}_M(\lambda, F; \tau) = \frac{N}{N - \tau M} B_{M-1, N-1}(F)$$
$$\times \left[1 + \lambda \tau \tilde{\Phi}(\lambda, F; \tau) \right]. \tag{55}$$

Consequently,

$$\tilde{\phi}(\lambda, F; \tau) = h_{N-1}(F; \tau)[1 + \lambda \tau \tilde{\Phi}(\lambda, F; \tau)], \tag{56}$$

where we have introduced

$$h_{N-1}(F; \tau) = \frac{1}{1 - F} \sum_{M=0}^{N-1} \frac{N - M}{N - \tau M} B_{M-1, N-1}(F)$$
$$= \sum_{M=1}^{N-1} \frac{N - 1}{N - \tau M} B_{M-1, N-2}(F)$$
$$= \frac{N - 1}{N - \tau} \frac{1}{1 - F} \, {}_2F_1(2 - N, 1; 2 - N/\tau; F). \tag{57}$$

Here ${}_2F_1$ is Gauss's hypergeometric function. For later use, we note that

$$\tau h_{N-1}(1 - x/N; \tau) = (N - 1)x^{-1 - N(\tau^{-1} - 1)} e^{-x} \gamma$$
$$\times (1 + N(\tau^{-1} - 1); x) + O(N^{-1}), \tag{58}$$

where γ is the incomplete γ function. We shall also need the function

$$H_N(F; \tau) = \int_0^F dF' h_N(F'; \tau) \tag{59}$$

and make contact with Sec. IV by noting that

$$H_N(F; 1) = H_N(F). \tag{60}$$

For later use, we note that

$$\tau H_{N-1}(1 - x/N; (1 + z/N)^{-1})$$
$$= \ln N - \psi(1 + z) - \gamma_E - \mathcal{I}(x; z) + O(N^{-1}), \tag{61}$$

when $x \sim 1$ and $z \sim 1$, and we have introduced

$$\mathcal{I}(x; z) = \int_0^1 dy \, y^z \frac{1 - e^{-x(1-y)}}{1 - y} - \gamma_E, \tag{62}$$

$$\mathcal{I}(x; 0) = \text{Ein}(x) - \gamma_E. \tag{63}$$

Equation (56) is solved by

$$\tilde{\phi}(\lambda, F; \tau) = h_{N-1}(F; \tau) \exp[\lambda \tau H_{N-1}(F, \tau)], \tag{64}$$

and consequently

$$\tilde{p}_M(\lambda, F; \tau) = \frac{N B_{M-1, N-1}(F)}{N - \tau M} \exp[\lambda \tau H_{N-1}(F, \tau)]. \tag{65}$$

In this result λ only occurs multiplied by τ. This is because in the series expansion of this result each power of λ represents one step taken in configuration space by the adaptive walk, and each such step takes one unit of time, represented by one power of τ. Powers of τ not occurring in conjunction with λ, on the other hand, represent time steps during which the adaptive walk did not progress.

The relation between length and duration of adaptive walks is contained in

$$p_{\ell; t} = \int_0^1 dF \sum_{M=0}^{N} p_{\ell, M; t}(f) \tag{66}$$

and therefore in

$$\tilde{p}(\lambda; \tau) = \int_0^1 dF \sum_{M=0}^{N} \tilde{p}_M(\lambda, F; \tau)$$
$$= \frac{1}{\lambda \tau} \left(\frac{1}{(1 - \tau)^\lambda} - 1 \right) + O(N^{-1}). \tag{67}$$

The generating function at time $t \geq 0$,

$$\tilde{p}_t(\lambda) = \sum_{\ell=0}^{\infty} \lambda^\ell p_{\ell; t}, \tag{68}$$

is obtained from $\tilde{p}(\lambda; \tau)$ via the relation

$$\tilde{p}_t(\lambda) = \frac{1}{2\pi i} \oint \frac{d\tau}{\tau^{t+1}} \tilde{p}(\lambda; \tau)$$
$$= \frac{1}{2\pi i \lambda} \oint \frac{d\tau}{\tau^{t+2}} \left(\frac{1}{(1 - \tau)^\lambda} - 1 \right) + O(N^{-1})$$
$$= \frac{\sin(\pi \lambda)}{\pi \lambda} B(1 - \lambda, \lambda + t + 1) + O(N^{-1}), \tag{69}$$

where the closed path of integration in the complex τ plane encircles $\tau = 0$ once in the positive direction. Using Cauchy's theorem, the last identity was established by moving the path to lie along the integrand's branch cut on the real axis, $\tau \geq 1$. The function $B(x, y)$ is the beta

function, Euler's integral of the first kind. Notice that the normalization condition

$$\tilde{p}_t(1) = \sum_{\ell=0}^{\infty} p_{\ell;t} = 1 \quad \text{for all } t \geq 0 \tag{70}$$

is satisfied by the result in Eq. (69). The same result gives, to leading order in $1/N$, that

$$(\bar{\ell})_t = \sum_{\ell=0}^{\infty} \ell\, p_{\ell;t} = \frac{d\tilde{p}_t(1)}{d\lambda} = \psi(t+2) + \gamma_E - 1$$
$$= \ln t + \gamma_E - 1 + O(t^{-1}) \tag{71}$$

and

$$\sigma_t^2(\ell) = (\bar{\ell}^2)_t - (\bar{\ell})_t^2 = \psi(t+2) + \gamma_E + \sum_{k=1}^{t+1} \frac{1}{k^2}$$
$$= \ln t + \gamma_E + \pi^2/6 + O(t^{-1}). \tag{72}$$

Thus we see our estimate confirmed: the average length of an adaptive walk grows logarithmically with time. Furthermore, we see that the variance of the length grows like the average length, like for a biased random walk. This similarity is no coincidence, since the adaptive walk in many respects resembles a simple, biased random walk.

In the last identity in Eq. (67) it was tacitly assumed that N itself was the only quantity of order N. Consequently, the time dependence found from this identity is reliable only when t is far from being of order N. This restriction needs not prevent t from being large and the asymptotic forms in Eqs. (71) and (72) from being valid.

When $t \sim N$, walks reach local maxima and terminate, according to our estimate for their duration. This, of course, is an average result. For example, there is a probability $\sim 1/N$ that an adaptive walk terminates already after its first step. Now let us substantiate the estimate: the probability that a walk terminates with length ℓ and fitness F at time t is

$$q_{\ell;t}(F) = p_{\ell,N;t}(F) - p_{\ell,N;t-1}(F). \tag{73}$$

Contact is made with Sec. IV by observing

$$q_\ell(F) = \sum_{t=0}^{\infty} q_{\ell;t}(F) = \lim_{t\to\infty} p_{\ell,N;t}(F). \tag{74}$$

We introduce

$$Q_{\ell;t} = \int_0^1 dF\, q_{\ell;t}(F) \tag{75}$$

and

$$\tilde{Q}(\lambda;\tau) = \sum_{\ell=0}^{\infty} \lambda^\ell \sum_{t=0}^{\infty} \tau^t Q_{\ell;t} \tag{76}$$

and have

$$\tilde{Q}(\lambda;\tau) = (1-\tau) \int_0^1 dF\, \tilde{p}_N(\lambda, F; \tau)$$
$$= \int_0^1 dF\, F^{N-1} \exp[\lambda \tau H_{N-1}(F;\tau)]$$
$$= e^{-\lambda(\psi(1+z)+\gamma_E)} \mathcal{F}(\lambda;z) + O(N^{-1}), \tag{77}$$

where Eq. (65) was used in the second identity, and $F = 1 - x/N$, $\tau = (1 + z/N)^{-1}$, $x, z \sim 1$, in the third. We have introduced the N-independent function

$$\mathcal{F}(\lambda;z) = \int_0^\infty dx\, e^{-x-\lambda\mathcal{I}(x;z)}. \tag{78}$$

Equation (77) is the time-dependent extension of Eq. (21). From the generating function in Eq. (77) we derive the average time it takes for an adaptive walk to reach a local maximum:

$$\bar{t} = \sum_{\ell,t=0}^{\infty} t Q_{\ell;t} = \frac{\partial \tilde{Q}}{\partial \tau}(1;1)$$
$$= \left\langle H_{N-1}(F;1) + \frac{\partial H_{N-1}}{\partial \tau}(F;1) \right\rangle$$
$$= \left\langle \int_0^F dF' \sum_{M=1}^{N-1} \frac{N(N-1)}{(N-M)^2} B_{M-1,N-2}(F') \right\rangle$$
$$= N\left(\frac{\pi^2}{6} - \frac{\partial\mathcal{F}}{\partial z}(1;0)\right) + O(1)$$
$$= 1.223\,98\ldots N + O(1), \tag{79}$$

$$\bar{t^2} - \bar{t}^2 = N^2\left(2\zeta(3) + \frac{\partial^2\mathcal{F}}{\partial z^2}(1;0) - \left(\frac{\partial\mathcal{F}}{\partial z}\right)^2(1;0)\right)$$
$$+ O(N) \tag{80}$$
$$= 1.717\,88\ldots N^2 + O(N), \tag{81}$$

where ζ is Riemann's zeta function. We have not been able to relate the derivatives of $\mathcal{F}$ in these equations to known mathematical constants.

Comparing this section's results with those of Sec. IV, we notice a big difference between the length and the duration of adaptive walks in a random fitness landscape: while typical lengths are relatively closer to the average length, the larger the system size N is, typical durations can differ from the average by an amount the size of this average. This picture is confirmed by the following expression for Q_t, the probability that a walk has duration t:

$$Q_t = \frac{1}{2\pi i} \oint \frac{d\tau}{\tau^{1+t}} \tilde{Q}(1;\tau)$$
$$= \frac{1}{2\pi i N} \oint dz\, e^{\frac{t}{N}z - \psi(1+z) - \gamma_E} \mathcal{F}(1;z). \tag{82}$$

Here the closed path of integration in the complex τ plane encircles $\tau = 0$ once in the positive direction, while a similar path of integration in the complex z plane, obtained by the substitution $z = N(\tau^{-1} - 1)$, has been moved to lie along the negative real axis. That is, the only place in the z plane where $\mathcal{F}(1;z)$ is not analytic. We have not found a more closed analytical expression for Q_t in the large-N limit than Eq. (82). Equation (82) suffices, however, since it shows that for $t/N \sim 1$ we have $Q_t \sim N^{-1}$. Hence, in the limit $N \to \infty$, $N Q_t$ is a finite function of the variable t/N. We have found this function numerically. Its graph is shown in Fig. 2 as the fully drawn line. The dashed line shows the graph for the estimate in Eq. (40) with the exact value in Eq. (79) used for $\bar{t}$. From the figure it seems that for $t/N \geq 1$, Q_t is essen-
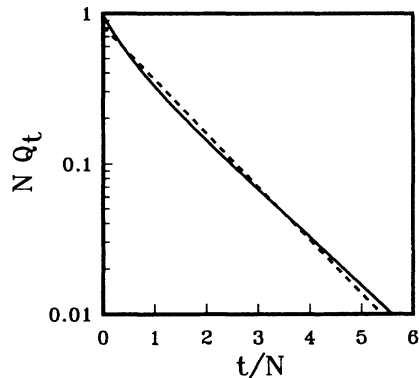
FIG. 2. NQ_t vs t/N for $N = \infty$. Fully drawn curve: exact result from Eq. (82). Dashed curve: estimate from Eq. (40) with the exact value for $\bar{t}$ taken from Eq. (79).

tially an exponential function, or at least exponentially bounded, though other possibilities cannot be eliminated on the basis of the figure.

VIII. CONCLUSIONS

We have obtained rigorous results to leading order in $1/N$ for the length and duration of adaptive walks in an N-dimensional binary genome space equipped with a random fitness landscape, a special case of Kauffman's NK model. We found the average length scales as $\ln N$ and so does the variance of the distribution of lengths. We have also obtained analytical expressions for the prefactors in these scaling laws, and found that to leading order in $1/\ln N$, lengths are Poisson distributed.

For the duration of adaptive walk, we found qualitatively different results. While the average duration is proportional to N with a constant of proportionality we have found analytically, the variance of the duration is proportional to N^2, again with analytically known coefficient. So while typical *lengths* of adaptive walks are relatively close to their average, typical *durations* vary over a range with magnitude equal to their average. We extended this result by showing analytically that in the limit $N \to \infty$, t/N has a finite distribution. Numerically, we found this distribution falls off exponentially for $t/N \geq 1$.

This quantitative, analytical understanding of the stochastic, adaptive dynamics of isolated species in a random fitness landscape permit us to find the dynamical phase structure of coevolving species of the same kind. This is the subject of the accompanying article.

ACKNOWLEDGMENTS

APPENDIX

In this appendix we argue that if the dimension N of configuration space is sufficiently large compared to the length of a finite path in that space, we cannot distinguish, to leading order in N, between the path of a random walk and the path of an adaptive walk in a random fitness landscape.

Assume that the dimension N of configuration space is much larger than the length of adaptive walks in that space. Then we can neglect the fact that the adaptive walk avoids itself and all configurations previously probed by it. The reasoning goes as follows: Since mutations occur on random genes, a step is added to the walk by probing random directions in configuration space, until one leading to higher fitness is found. Then the walk is extended one step in that direction, and the procedure repeated from the new position in configuration space. By this algorithm, correlations between successive directions chosen by the walk are of order $1/N$. So to leading order in an expansion in $1/N$ successive directions are uncorrelated, and we have a *random* walk at hand. Successive directions are also *different* to leading order. Consequently, if the length of the walk is much smaller than $\sqrt{N}$, all directions chosen by it are different, and it obviously does not self-intersect [15].

By assuming that the adaptive walk never probes a site in configuration space that it has probed before, we found, in Sec. IV, that walks have length $\sim \ln N$, which *is* much smaller than $\sqrt{N}$ for N large. We conclude that our assumption that the walk is short compared to N is self-consistent correct.

We may ask whether we can find all subleading terms in an expansion in $1/N$ without knowing the entire history of an adaptive walk. The answer is negative.

An adaptive walk does not back-track, while a random walk does with probability $\sim 1/N$ per step. We can handle a random walk without back-tracking analytically. But back-tracking is not the only $1/N$ effect distinguishing an adaptive walk from a random one, however. An adaptive walk also forms no closed loops, and does not visit sites in configuration space that were probed previously, but not visited for lack of fitness. Thus an adaptive walk is not only self-avoiding, but also avoids many one-mutant neighbors to itself. A short random walk visits such sites with probability $\sim 1/N$. This is seen as follows.

Self-intersection by a random walk requires the formation of a closed loop by the walk, i.e., at least two steps, of opposite orientation, must be taken in each dimension in which the loop extends. So the probability for the formation of a closed loop of length ℓ' on a random walk of length ℓ is, to leading order, suppressed by a factor $(\ell - \ell')/N^{\ell'/2}$, where $\ell' \geq 4$. Nearest neighbors to the walk can be visited in one step less, i.e., with probability $(\ell - \ell')/N^{\ell'/2-1}$. For $\ell' = 4$ this probability is $\sim \ell/N$. So to leading order in $1/N$ we can treat the adaptive walk as a random walk. We can also treat it as a random walk without back-tracking, thereby describing some of the $1/N$ effects at play. But a full description of $1/N$ effects requires more information than the walk's current position in configuration space.

In summary, to leading order in $1/N$ we may add a step to the adaptive walk by treating the one-mutant neighbors to the current configurations as if they had never been visited or probed before. Consequently, the probability that M of these N neighbors are less fit than the current configuration is binomial, $B_{M,N}(F)$, where F is the fitness of the current configuration. If we take into account that the previous configuration is known to be less fit, the probability is $B_{M-1,N-1}(F)$, as given in Eq. (47).

When we forbid back-tracking, our treatment is exact for a configuration space which is a Cayley tree with co-ordination number N. It should not be confused with an "annealed" fitness landscape, as an approximation to the "quenched" landscape we start out with. Not if "annealed" means rechoosing the fitness of a configuration every time it is probed by the adaptive walk. If we did that we would have no maxima, since a higher fitness could always be attained by sufficiently many trials. The picture of an "annealed" fitness landscape applies only in the sense that the fitnesses of all N (or $N-1$) neighbors to a configuration are rechosen every time that configuration is visited, *and kept fixed during the visit,* thereby possibly making the visit permanent.

[1] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley, New York, 1989).

[2] S. A. Kauffman and S. Levin, J. Theor. Biol. **128**, 11 (1987).

[3] S. A. Kauffman, *Origins of Order: Self-Organization and Selection in Evolution* (Oxford University, Oxford, 1990).

[4] E. Weinberger, J. Theor. Biol. **134**, 125 (1988).

[5] C. A. Macken and A. S. Perelson, Proc. Natl. Acad. Sci. U.S.A. **86**, 6191 (1989).

[6] C. A. Macken, P.S. Hagan, and A. S. Perelson, SIAM J. Appl. Math. **51**, 799 (1991).

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

[8] S. A. Kauffman and S. Johnsen, J. Theor. Biol. **149**, 467 (1991).

[9] P. Bak, H. Flyvbjerg, and B. Lautrup, following paper, Phys. Rev. A **46**, 6724 (1992).

[10] J. Gillespie, Evolution **38**, 1116 (1984).

[11] B. Derrida, Phys. Rev. B **24**, 2613 (1981).

[12] E. Gardner and B. Derrida, J. Phys. A **22**, 1975 (1989).

[13] D. E. Knuth, *The Art of Computer Programming* (Addison-Wesley, New York, 1968), Vol. 1.

[14] *Handbook of Mathematical Functions,* edited by M. Abramowitz and I. A. Stegun, 9th ed (Dover, New York, 1972), p. 228, footnote 3.

[15] B. Derrida and H. Flyvbjerg, J. Phys. (Paris) **48**, 971 (1987).