PHYSICAL REVIEW A VOLUME 44, NUMBER 10 15 NOVEMBER 1991

Local properties of Kauffman's N-k model: A tunably rugged energy landscape

Edward D. Weinberger
Max Planck Institute for Biophysical Chemistry, Postfach 2841, D-3400 Göttingen, Federal Republic of Germany
(Received 24 September 1990; revised manuscript received 14 June 1991)

The N-k model is a dilute, k-ary spin glass in which the state of each of the N sites is affected by that site and k of its neighbors. As a function of k for large k, we explicitly compute the number of local minima of the Hamiltonian, the distribution of locally minimal energies and the first two moments of that distribution, and a number of statistical properties of "downhill" walks from random starting positions to local optima on these landscapes, including estimates for their length. We suggest some implications of these results for spin-glass physics and for approximating other landscapes that cannot be modeled using more conventional, quadratically coupled spin glasses.

PACS number(s): 64.60.Cn, 87.10.+e

I. INTRODUCTION AND SPECIFICATION OF THE MODEL

The properties of multipeaked "fitness" or "energy" landscapes have attracted attention, both in many areas of physics and in a wide variety of other fields, including evolutionary biology and computer science. Biologists [1–7] have embraced this paradigm in the hope that they might learn something about evolution, and computer scientists [8,9] hope to use evolutionary strategies in developing methods of solving combinatorial optimization problems. Similar issues have also attracted the attention of physicists [10,11], who speculate that the thermodynamics of glassy systems, such as polymers and other more or less random covalent networks, are intimately related to the complex structure of the barrier heights present in the potential surface. The common denominator in all of this work is a notion that has become known as a "rugged landscape." If one is a biologist, such a landscape can be interpreted as a fitness landscape; if one is a computer scientist, the landscape is the set of allowable configurations in some optimization problem; and if one is a physicist or chemist, it is a glass and/or spin glass.

The primary focus of the most well-developed theory of rugged landscapes, namely, the theory of spin glasses, has been the *global* properties of specific kinds of landscapes; that is, the properties of the partition function and, in particular, the existence and nature of phase transitions for landscapes with pairwise interactions (such as the Sherrington-Kirkpatrick model) [12]. Although there have been some results concerning metastable states of these landscapes [13], and there have been some investigations of the global properties of more complex landscapes (e.g., k-ary spin glasses), we believe that the previously unexplored *local properties* of these more complex landscapes determine, for example, evolutionary adaptation. Suppose we assume, following Gillespie [14], that selection often proceeds on a significantly faster time scale than mutation. In this regime, Gillespie argues that long periods of stasis are punctuated by the effectively instantaneous replacement of the wild type by an advantageous mutation, and that deleterious mutations will die out sufficiently rapidly to be invisible over a slow time scale. This leads to the metaphor of evolutionary adaptation as an "uphill walk on a rugged landscape."

There are also compelling reasons for workers in other disciplines to be interested in such walks. In computer science, the most obvious algorithm for solving combinatorial optimization problems, such as the well-known "traveling salesman problem," is that of "iterative improvement." This algorithm, which consists of making incremental improvements to an initial trial solution to the problem, is obviously another example of an adaptive walk, but this time in the "downhill" direction. Downhill walks in phase space are also relevant to the low-temperature, dynamical behavior of glassy materials, for which the waiting time to return to equilibrium may be effectively infinite. Yet another place in which the local properties of a landscape play an important role is in the economic concept of a Nash equilibrium (in which no economic actor can improve his lot by acting unilaterally), which is clearly a *local* optimum. In each of these cases, results concerning the lengths of these adaptive walks and the energy of a typical local optimum are clearly of interest, both for "greedy" or "gradient" walks, in which the walker steps to the best of its neighbors, and for "random adaptive" walks, in which the next step is chosen at random from the set of better neighbors.

The purpose of this paper is to discuss the local properties of a class of landscapes which has become known by the generic name, the N-k model [3]. N-k landscapes are representative of a wide class of models that stochastically assign energies to the vertices of the Boolean hypercube, provided that the energies have a roughly Gaussian distribution and that the sequence of energies generated by unbiased random walks on the landscape approximates a stationary Markov chain [15]. This class of landscapes, which includes a variety of spin-glass models, is characterized more completely below.

Heuristically, the N-k model is an attempt to understand the dependence of a system with N "parts," on a parameter k which measures how richly interconnected the parts of the system are. (In order to conform to the

original specification of the model, we take k to be the number of *other* parts interacting with a given part. Note that some other authors take this quantity to be the *total* number of interactions per site.) As we will see, tuning k from 0 to $N-1$ tunes the ruggedness of the corresponding energy landscape. The ruggedness of the landscape alters the character of downhill walks toward optima under any of a variety of optimization procedures, ranging from mutation and selection of fitter variants ("hill climbing") to annealing (either physical or simulated) to recombinational strategies. Thus understanding the structure of N-k landscapes is of importance in the wide variety of situations mentioned above. In order to fix terminology, we will use the word "energy" to describe the landscape parameter that is to be optimized; we choose it only to appeal to an audience of physicists. For the same reason, we adopt the convention that the optima of interest are *minima,* rather than maxima. However, as the foregoing makes clear, the issues we wish to consider are richly interdisciplinary. The words "fitness" or "payoff" could equally well have been chosen, and we could also be interested in maxima, rather than minima. Our methods apply to local maxima with only trivial modifications.

The simplest N-k landscapes are constructed on the N-dimensional hypercube whose vertices are the set $\mathcal{B}_N$ of all possible Boolean vectors of length N. (Although it is biologically more plausible to allow either four symbols per site, thus modeling the interactions of bases in a DNA molecule, or to allow a symbol per site for each of the 20 amino acids, and thus model a protein, we will show that the qualitative features of these generalizations are already present in the simple binary model.) To form the landscape, we assume that the "Hamiltonian" H of the entire N bit string is the average of contributions from each of the individual bits. We choose the contribution h_i from the ith bit as a function of the state of that bit and $k<N$ other bits. The individual h_i's can therefore have 2^{k+1} possible values, one for each of the possible states of the $k+1$ bits on which h_i depends. Each of these 2^{k+1} possible energy contributions is assigned by selecting an independent random variable from some specified probability distribution,

$$P(x)=\text{Prob}(h_i \leq x) ,$$

thus expressing our ignorance of the exact nature of the interaction between each site and the others. This set of assignments constitutes the "energy table" for the ith bit. There is a different, independently generated table for each of the N bits which, once chosen, is never reassigned. In other words, the model is a "quenched," rather than an "annealed" model. See Table I for a pictorial description of the model.

It remains to specify the "neighborhood" of each bit; that is, the k other bits upon which it depends. The simplest of these cases, at least when k is even, is to imagine that the neighborhoods consist of the $k/2$ bits that are the nearest neighbors on each side of the bit in question and to assume that the bits are arranged in a circle (periodic boundary conditions), which guarantees that each bit does, indeed, have k neighbors. The other extreme is to consider "neighborhoods" in which the k bits

TABLE I. Graphical representation of the N-k model for $N=8$, $k=2$, adjacent neighborhoods, showing the tables used for computing the fifth and sixth site energies, but not the others, because they are computed by a similar procedure. In the 8-bit string whose energy is to be computed, the substring consisting of bits 4, 5, and 6 is "101," so the fifth site energy is 0.73, as per the underlined entry in the table at left. Similarly, the substring consisting of bits, 5, 6, and 7 is "010," so the sixth site energy is 0.29. In general, there will be 2^{k+1} entries in these tables. The ith of the N tables is indexed by the $(k+1)$-bit substring formed by concatinating bit i with its k "neighbors."

Bit position	1	2	3	4	5	6	7	8
N bit string to be assigned an energy	1	0	1	1	0	1	0	1
Energy contribution of bit position	0.39	0.46	0.91	0.18	0.73	0.29	0.84	0.70

Energy of above string = Average of energy contributions = 0.56

Table for computing the contribution of fifth bit position

Bit 4	Bit 5	Bit 6	Value of indep. sample from $P(x)$
0	0	0	0.32
0	0	1	0.21
0	1	0	0.19
0	1	1	0.93
1	0	0	0.87
1	0	1	0.73
1	1	0	0.64
1	1	1	0.88

Table for computing the contribution of sixth bit position

Bit 5	Bit 6	Bit 7	Value of indep. sample from $P(x)$
0	0	0	0.99
0	0	1	0.10
0	1	0	0.29
0	1	1	0.22
1	0	0	0.86
1	0	1	0.39
1	1	0	0.48
1	1	1	0.61

are chosen at random. Note that these two extremes correspond to two important kinds of spin glasses: the adjacent neighborhoods correspond to a one-dimensional, short-range spin glass; the random neighborhoods correspond to a dilute, long-range spin glass. However, one notable feature of this model is the possibility of studying the effects of dimensionality of the lattice by "unraveling" a d-dimensional rectangular lattice with L sites on a side into a one-dimensional lattice with $N=L^d$ sites. When the adjacent neighborhood of site i in d dimensions is unraveled into a one-dimensional lattice and the sites are numbered consecutively, the sites in the d-dimensional neighborhood are now site numbers $i-1, i+1, \quad i-L, i+L, i-L^2, i+L^2, \ldots, i-L^d, i+L^d$. Another possibility of the model is that of generating "intermediate-range" spin glasses by allowing the neighborhood of each site i to depend randomly on some k bit subset of the N^γ bits of the string closest to i, for $\gamma < 1$.

To see that alterations in k do indeed "tune" the ruggedness of the landscape, consider the landscapes generated by the largest and smallest possible k values: For $k=0$, the energy contribution of each site depends only on the bit value at that site, and not on bit values at any of the other sites. If the probability distribution from which the energy contributions are sampled is continuous, the probability that two such samples are exactly equal is zero; hence, a single specific sequence comprised of the fitter bit value in each position is the single, global optimum in the energy landscape. This simple case corresponds to the mean-field approximation to the energy of a spin system in physics and the haploid, multilocus, two-allele additive genetic model in population genetics [16]. We can define the "local correlation" of an energy landscape as the average correlation of the energy of each configuration with the energy of its 1-mutant neighbors, that is, those configurations obtained by flipping any single bit to the opposite state. Since such a flip can only alter energy by $O(1/N)$ in $k=0$ landscapes with N sites, such landscapes are highly locally correlated. Furthermore, any downhill walk from a given configuration through less energetic 1-mutant neighbors must eventually terminate at the unique global minimum. Because the length of the walk is just the Hamming distance from the initial string to the global minimum and because roughly half the bits will be in their less energetic state for a randomly chosen initial string, the expected walk length is $N/2$.

The fully connected N-k model (that is, $k=N-1$) yields a completely random energy landscape. In this case, the energy contribution of each site depends on all of the other sites, so that the energy of each bit string is assigned an energy independent of (and therefore locally uncorrelated with) its neighbors. This limiting case of the N-k landscape is therefore the Derrida random energy model of spin-glass physics [17]. Such random landscapes have very many local optima $[O(2^N/N)]$, walks to optima are short $[O(\ln N)]$, and only a small fraction of local optima are accessible from any initial string [1,18,19]. In other words, important features of the landscape vary dramatically as k varies from 0 to $N-1$. However, such uncorrelated landscapes are unrealistically "rugged;" in most practical situations, there is almost certainly *some* correlation between neighboring points. We propose that a wide class of such landscapes are at least qualitatively described by N-k landscapes with intermediate values of k. In particular, N-k landscapes approximate generic members of the class of so-called "AR(1) landscapes," in which all correlations are completely determined by correlations between neighboring points [15]. AR(1) landscapes were so named because the sequence of energies obtained via an unbiased random walk on them must be, in the limit of large landscapes, an AR(1) (first-order autoregressive) process. The AR(1) process, characterized by the difference equation

$$H_{t+1}=\rho H_t+\Delta H_t \ , \tag{1.1}$$

where ΔH_t is a sequence of independent, identically distributed Gaussian random variables, and $0\le\rho\le 1$, is the most general stationary, Gaussian random process that is also (first-order) Markov. The Markov property is effectively a "maximum entropy condition" that guarantees that no additional information about the random variable H_t from a knowledge of H_{t-s} for $s>1$, and thus no additional information about the energy of a given point, can be obtained from points further away than its nearest neighbors. [For all stochastic processes, the conditional entropy of H_t, given H_{t-1} and H_{t-2}, is bounded above by the conditional entropy of H_t, given only H_{t-1}. The Markov property implies that this upper bound is actually achieved for AR(1) landscapes.] This description implies that AR(1) landscapes in general and N-k landscapes in particular are statistically iostropic, in the sense that their statistical properties are invariant with respect to "translation" from one point to another. AR(1) landscapes include the traveling salesman problem [20], and they are ubiquitous in optimization problems in computer design [21].

In this paper, we focus primarily on the special case in which k is large (but possibly much smaller than its upper bound of N), and in which the mean μ and the variance σ^2 of the probability distribution of site energies are finite. Under these assumptions, we show that the energies of local optima for both the random and adjacent neighborhood N-k model are asymptotically normally distributed with a mean of approximately

$$\mu-\sigma\left[\frac{2\ln(k+1)}{k+1}\right]^{1/2}$$

and a variance of approximately

$$\frac{(k+1)\sigma^2}{N[k+1+2(k+2)\ln(k+1)]} \ .$$

The same calculation also allows us to deduce that the average Hamming distance between optima—which is roughly twice the length of a typical gradient walk—is approximately $N\log_2(k+1)/(k+1)$. We also derive recurrence relations for the mean and variance of the distribution of site energies at a given step in terms of the mean and variance at the previous step for the random adaptive walk. Not only do these recurrence relations

characterize the approximately Gaussian distribution of energies as the walk progresses, but they can also be used to argue that the length of a random adaptive walk is roughly $2\ln 2 \approx 1.39$ times longer than a gradient walk on a landscape with the same large N and k values. The paper concludes by showing that these results and the methods used to derive them apply at least qualitatively to a variety of more general cases.

II. NUMBERS AND ENERGIES OF LOCAL OPTIMA IN THE GENERAL CASE

We note first that the statistical properties of the landscape are unchanged whenever the labels 1 and 0 are exchanged at a single site. We can therefore choose a particular locally optimal configuration and exchange the appropriate 1 and 0 labels to make that configuration have the labels $(0,0,\ldots,0)=\underline{0}$ without changing the quantities we wish to calculate. The truth of this claim is readily verified by means of a geometric argument: We imagine that an arrow is drawn from each vertex of the Boolean hypercube to each adjacent vertex with lower energy. Verticies corresponding to local energy minima will then have no arrows pointing away from them. If we view the label exchange as a rotation of the Boolean hypercube such that a specified corner of it lands on the origin, we see that the rotation does not disturb the number of arrows entering and leaving each vertex.

The condition for a configuration to be a local minimum is that its energy increases when we flip any one of its bits. This will be the case if

$$h_1(\underline{0})+h_2(\underline{0})+\cdots+h_N(\underline{0}) \leq h_1(\underline{0}_i)+h_2(\underline{0}_i)+\cdots+h_N(\underline{0}_i)\,, \quad (2.1)$$

where the configuration $\underline{0}_i$ is the $\underline{0}$ configuration with bit i flipped to 1. The details of what happens when bit i is flipped are seen most clearly by restricting attention to the $k=2$ case with adjacent neighborhoods. In this case, the site energies $h_{i-1}(\underline{0})$, $h_i(\underline{0})$, and $h_{i+1}(\underline{0})$ change to the site energies $h_{i-1}(\underline{0}_i)$, $h_i(\underline{0}_i)$, and $h_{i+1}(\underline{0}_i)$ respectively. No other site energies change. The condition that the configuration $\underline{0}$ is a local optimuum is thus

$$h_{i-1}(\underline{0})+h_i(\underline{0})+h_{i+1}(\underline{0}) \leq h_{i-1}(\underline{0}_i)+h_i(\underline{0}_i)+h_{i+1}(\underline{0}_i) \quad (2.2)$$

for $1 \leq i \leq N$. The random variables

$$h_{i-1}(\underline{0}_i)+h_i(\underline{0}_i)+h_{i+1}(\underline{0}_i)$$

are mutually independent for different values of i, and are also independent of

$$h_{i-1}(\underline{0})+h_i(\underline{0})+h_{i+1}(\underline{0})$$

because the energy tables for each bit i are constructed from *different, independent* random variables. It follows that the inequalities (2.2) are, in fact, probabilistically independent events. For arbitrary neighborhoods, the corresponding inequalities are

$$\sum_{j\in\nu_i} h_j(\underline{0}) \leq \sum_{j\in\nu_i} h_j(\underline{0}_i)\,, \quad (2.3)$$

where ν_i is the neighborhood of the ith site, and where all of the subscripts are interpreted modulo N, a convention to be followed throughout this paper.

Because of the independence of the h_i's, the joint probability of all of the inequalities (2.3) being satisfied simultaneously is the product of the probabilities that each inequality is satisfied in isolation. Furthermore, the probability that a single one of these inequalities is satisfied is just $1-P_{k+1}(\sum_{j\in\nu_i} h_j)$, where $P_{k+1}(x)$ is the distribution of the sum of $k+1$ independent samples from the distribution $P(x)$ of site energies. A basic result of the theory of probability states that P_{k+1} is the $(k+1)$-fold convolution of $P(x)$ with itself. The probability that a randomly chosen configuration is a local optimum is therefore

$$\Omega=\int_{-\infty}^{\infty} dP(h_1)\int_{-\infty}^{\infty} dP(h_2)\cdots\int_{-\infty}^{\infty} dP(h_N)\left\{\prod_{i=1}^{N}\left[1-P_{k+1}\left[\sum_{j\in\nu_i} h_j\right]\right]\right\}, \quad (2.4)$$

where we write $dP(x)$ rather than $p(x)dx$ to remind ourselves that the density of $P(x)$ may involve δ functions. Note that the only change that must be made if we are interested in local *maxima* is that the sense of the inequality in (2.1) is reversed. $1-P_{k+1}(\sum_{j\in\nu_i} h_j)$ must then be replaced by $P_{k+1}(\sum_{j\in\nu_i} h_j)$ in (2.4).

Similar expressions can be written down for the mth moment of the distribution of local energy optima, which is

$$\mathcal{E}[E^m|\text{ configuration is local optimum}]=\frac{\mathcal{E}[E^m]}{\text{Prob(configuration is local optimum)}}\,. \quad (2.5)$$

If we let $\underline{1}$ denote the vector with all unit entries, and we denote the vector of energy contributions as $\underline{h}$, so that

$$\frac{\underline{1}^T\underline{h}}{N}\equiv\frac{h_1+h_2+\cdots+h_N}{N}\,,$$

(2.5) is equivalent to

$$\frac{1}{\Omega}\int_{-\infty}^{\infty}dP(h_1)\int_{-\infty}^{\infty}dP(h_2)\cdots\int_{-\infty}^{\infty}dP(h_N)\left[\frac{\underline{1}^T\underline{h}}{N}\right]^m\left\{\prod_{i=1}^{N}\left[1-P_{k+1}\left[\sum_{j\in\nu_i}h_j\right]\right]\right\}. \tag{2.6}$$

These integrals can all be evaluated by first computing

$$Z(\beta)=\int_{-\infty}^{\infty}dP(h_1)\int_{-\infty}^{\infty}dP(h_2)\cdots\int_{-\infty}^{\infty}dP(h_N)e^{-\beta(\underline{1}^T\underline{h})/N}\left\{\prod_{i=1}^{N}\left[1-P_{k+1}\left[\sum_{j\in\nu_i}h_j\right]\right]\right\} \tag{2.7}$$

and then taking the absolute value of successive β derivatives when $\beta=0$. [The choices of the suggestive variable names β and Z is not accidental; the relationship between (2.7) and the average partition functions in spin-glass physics will be discussed in the last section of this paper.]

III. NUMBERS AND ENERGIES OF LOCAL OPTIMA FOR LARGE k

We now consider the special case mentioned in the Introduction; that is, we assume that k is large and the distribution of site energies has finite mean and variance. Because these assumptions will allow us to apply the central-limit theorem [22,23], we first investigate the way in which the statistics of local optima vary as a function of μ and σ^2 when $P(x)$ is a normal distribution with these parameters as mean and variance, respectively. $P_{k+1}(x)$ is then also normal with mean $(k+1)\mu$ and variance $(k+1)\sigma^2$, so we have the explicit formula

$$Z(\beta)=\frac{1}{(2\pi\sigma^2)^{N/2}}\times\int_{-\infty}^{\infty}d\underline{h}\exp\left[-(\underline{h}-\mu\underline{1})^2/2\sigma^2-\frac{\beta}{N}\underline{1}^T\underline{h}\right]\times\left[\frac{1}{2\pi(k+1)\sigma^2}\right]^{N/2}\times\left[\prod_i\int_{\mathcal{S}}^{\infty}e^{-[\xi_i-(k+1)\mu]^2/2(k+1)\sigma^2}d\xi_i\right], \tag{3.1}$$

where $\mathcal{S}=\sum_{j\in\nu_i}h_j$. Shifting the means of all of the Gaussians in $Z(\beta)$ from μ to zero corresponds to making the substitutions $h_i-\mu=y_i$ and $\xi_i-(k+1)\mu=\eta_i$ in the integrals above. It is thus clear that $Z(0)$, and therefore the distribution of the number of optima remains unchanged under a shift in the mean of $P(x)$. Note that changing the means of the Gaussians by μ will also shift $(1/Z)\partial Z/\partial\beta$— and thus the average energy of a local minimum—by the same amount. Similarly, via the substitutions $\sigma h_i=y_i$ and $\sigma\sqrt{k+1}\xi_i=\eta_i$, it is easy to check that the mean of the number of optima remains unchanged when the variance of $P(x)$ changes from unity to σ^2 and the mean of $P(x)$ remains zero. However, this change in P changes both the mean energy at a local minimum and its standard deviation by the factor σ.

It remains to evaluate (3.1) for $\mu=0$, $\sigma^2=1$, and $k\gg1$. An evaluation of (3.1) when $\beta=0$ will also give us a hint as to how to obtain the entire probability distribution of local optima in this case. To do this, we observe that for both the adjacent and random neighborhood models

$$\sum_{j\in\nu_i}h_j=(k+1)\bar{h}+O(\sqrt{k+1}),$$

where

$$\bar{h}=\frac{1}{N}\underline{1}^T\underline{h}=\frac{1}{N}(h_1+h_2+\cdots+h_N).$$

Thus, for large k we have

$$Z(0)=\left[\frac{N}{2\pi}\right]^{1/2}\int_{-\infty}^{\infty}d\bar{h}\exp\left[-\frac{N\bar{h}^2}{2}\right]\left[\left[\int_{\sqrt{k+1}\bar{h}}^{\infty}+\int_{\sqrt{k+1}\bar{h}}^{\sqrt{k+1}\bar{h}+O(1)}\right]\frac{e^{-\xi^2/2}}{\sqrt{2\pi}}d\xi\right]^N. \tag{3.2}$$

Because $\bar{h}<0$, the first integral in parentheses is $O(1)$. However, using the estimate $\bar{h}\approx-[2\ln(k+1)/(k+1)]^{1/2}$ derived below, it is easy to check that the second integral in parentheses is $O(1/k\ln k)$. Actually, the $O(1)$ terms in the limit of the second integral fluctuate essentially randomly in sign, so that the error involved in ignoring this integral is actually of much smaller order, the larger of $O(1/k\ln k\sqrt{N})$ and $O(1/(k\ln k)^2)$. The remaining expression can be written in the form

$$\left[\frac{N}{2\pi}\right]^{1/2}\int_{-\infty}^{\infty}d\bar{h}\exp\left[-\frac{N\bar{h}^2}{2}+N\ln\int_{\sqrt{k+1}\bar{h}}^{\infty}\frac{e^{-\xi^2/2}}{\sqrt{2\pi}}d\xi\right].$$

The integrand of the outer integral is of the form $e^{-Nf(\bar{h})}$, which has an extremely sharp maximum when the function

$$f(\bar{h})=\frac{\bar{h}^2}{2}-\left[\ln\int_{\sqrt{k+1}\bar{h}}^{\infty}\frac{e^{-\xi^2/2}}{\sqrt{2\pi}}d\xi\right]$$

has a minimum. The standard method for evaluating such integrals is the method of steepest descent, which involves approximating $f(\bar{h})$ by a truncation of its Taylor series, expanded about its minimum value, which we denote by h_m. The condition that h_m must satisfy is found via a straightforward differentiation of f to be

$$h_m=-\sqrt{k+1}\frac{e^{-(k+1)h_m^2/2}}{\int_{\sqrt{k+1}h_m}^{\infty}e^{-\xi^2/2}d\xi}. \tag{3.3}$$

Because we must have $h_m<0$, the denominator in (3.3) must be between $\sqrt{\pi/2}$ and $\sqrt{2\pi}$, and the right side of (3.3) is monotone decreasing. Equation (3.3) must therefore have a unique solution, which lies in the interval

$$-\left[\frac{(2-\delta)\ln(k+1)}{k+1}\right]^{1/2}>h_m>-\left[\frac{2\ln(k+1)}{k+1}\right]^{1/2}, \tag{3.4}$$

for arbitrarily small δ and sufficiently large k. It is easy to check (3.4) by noting that the inequality that results in substituting one of the bounds for h_m reverses its sense when the other bound is substituted.

A second differentiation, some algebra, and the use of (3.3) allows us to compute

$$\left.\frac{d^2f(\bar{h})}{d\bar{h}^2}\right|_{\bar{h}=h_m}=1+(k+2)h_m^2.$$

The required Taylor series is then

$$f(\bar{h})=f(h_m)+\frac{1+(k+2)h_m^2}{2}(\bar{h}-h_m)^2+\cdots,$$

and

$$Z(0)\approx\left[\frac{N}{2\pi}\right]^{1/2}e^{-Nf(h_m)}\int_{-\infty}^{\infty}e^{-N[1+(k+2)h_m^2](\bar{h}-h_m)^2/2}d\bar{h}$$

$$=\frac{e^{-Nf(h_m)}}{\sqrt{1+(k+2)h_m^2}}$$

$$\approx\left[1+2\left[\frac{k+2}{k+1}\right]\ln(k+1)\right]^{1/2}\left[\left[\frac{1}{k+1}\right]^{1/(k+1)}\int_{-\infty}^{\sqrt{2\ln(k+1)}}\frac{e^{-t^2/2}}{\sqrt{2\pi}}dt\right]^N. \tag{3.5}$$

In other words, the probability that a randomly chosen configuration is a local optimum if $N\to\infty$ and k is fixed is

$$Z(0)=O(\lambda^N),$$

where λ is the k-dependent quantity in the second set of square brackets above. If we define

$$\chi(\underline{e})=\begin{cases}1 & \text{if configuration } \underline{e} \text{ is a local optimum}\\ 0 & \text{otherwise},\end{cases}$$

then the expected number of local optima is

$$\mathcal{E}\left[\sum_{\underline{e}}\chi(\underline{e})\right]=O((2\lambda)^N),$$

because there are 2^N configurations $\underline{e}$, and each has the probability $O(\lambda^N)$ of being a local optimum.

With only slightly more effort, we can explicitly compute the probability $\phi(u)$ that $\bar{h}=u$. If $p(x)$ is Gaussian with zero mean and unit variance, the identity

$$\delta(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{izx}dz$$

allows us to write this distribution as

$$\phi(u)=\left[\frac{1}{2\pi}\right]^{N+1}\int_{-\infty}^{\infty}e^{-\underline{h}^2/2}d\underline{h}\,\delta(u-\underline{1}^T\underline{h}/N)\left[\prod_i\int_{\left[\sum_{j\in v_i}h_j\right]}^{\infty}\frac{e^{-\xi_i^2/2(k+1)}}{\sqrt{k+1}}d\xi_i\right]$$

$$=\left[\frac{1}{2\pi}\right]^{N+1}\int_{-\infty}^{\infty}e^{-\underline{h}^2/2}d\underline{h}\int_{-\infty}^{\infty}e^{iz(u-\underline{1}^T\underline{h}/N)}dz\left[\prod_i\int_{\left[\sum_{j\in v_i}h_j\right]}^{\infty}\frac{e^{-\xi^2/2(k+1)}}{\sqrt{k+1}}d\xi_i\right]$$

$$\approx\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\frac{N}{2\pi}\right]^{1/2}e^{-N\bar{h}^2/2}d\bar{h}\int_{-\infty}^{\infty}e^{iz(u-\bar{h})}dz\left[\int_{\sqrt{k+1}\bar{h}}^{\infty}\frac{e^{-\xi^2/2)}}{\sqrt{2\pi}}d\xi\right]^N$$

$$\approx\left[\frac{N}{2\pi}\right]^{1/2}e^{-Nf(h_m)}\int_{-\infty}^{\infty}e^{izu}dz\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-N[1+(k+1)h_m^2](\bar{h}-h_m)^2/2}e^{-iz\bar{h}}d\bar{h}\right].$$

The iterated integrals are just the forward and backward Fourier transform of the Gaussian in the inner integral. The expression above must then evaluate to

$$\phi(u)=\frac{e^{-Nf(h_m)}}{\sqrt{1+(k+2)h_m^2}}\times\left[\frac{N[1+(k+2)h_m^2]}{2\pi}\right]^{1/2}e^{-N[1+(k+2)h_m^2](u-h_m)^2/2}.$$

Note that $\phi(u)du$ is the probability that a given configuration is a local minimum *and* that it has energy between u and $u+du$. The probability that the configuration has energy between u and $u+du$, given that it is a local minimum is thus:

$$\text{Prob}\{u<\bar{h}<u+du\,|\,\text{local min}\}=\left[\frac{N[1+(k+2)h_m^2]}{2\pi}\right]^{1/2}e^{-N[1+(k+2)h_m^2](u-h_m)^2/2}du\ .$$

It follows that the expected energy of a local minimum is h_m, and the variance about this mean is $1/\{N[1+(k+2)h_m]^2\}$. Figure 1 shows a comparison between these analytical results and the computer simulations reported in Table III. (The data in this table is reprinted from Ref. [2], except that we report here the energy of local *minima,* instead of local *maxima,* as reported there). In all cases, the underlying site energy distribution is the uniform distribution on the unit interval, for which the mean is $\frac{1}{2}$ and the variance is $\frac{1}{12}$, and for which the energy of local minima is one minus the energy of local maxima. The statistics of downhill walks to local minima and uphill walks to local maxima are clearly identical. Local minima are consistently somewhat lower for landscapes with random neighborhoods for the same reason that walk lengths for random landscapes are longer: some sites in random neighborhood landscapes are relatively unconstrained, and the lower energies that these sites can attain will more than offset the high energies that are the best that can be attained by unusually constrained sites.

IV. STATISTICS OF DOWNHILL WALKS FOR LARGE-k LANDSCAPES

As we saw in the Introduction, many of the properties of N-k landscapes are determined by the (pair) correlations $R(d)$ between configurations separated by a given Hamming distance d.

$$R(d)=\left[1-\frac{d}{N}\right]\left[1-\frac{k}{N-1}\right]^d$$

for the random-neighbor model [24], and

$$R(d)=1-\frac{d(k+1)}{N}+\frac{d}{N\binom{N-1}{d-1}}\sum_{l=1}^{k}(k+1-l)\binom{N-l-1}{d-2}$$

for the adjacent-neighbor model [25]. If $k/N=O(1)$, these results suggest that N-k landscapes split apart into uncorrelated patches, so that the local optima should be more or less randomly distributed in the space, in contrast to the ultrametric distribution of local minima of spin-glass Hamiltonians. Precisely the former behavior was observed [26] in the free-energy landscape of RNA secondary structures, which recent studies show to be considerably less correlated than the spin-glass landscape [24]. When $k/N=O(1)$, therefore, it is meaningful to estimate the average distance between optima and the average lengths of gradient and random adaptive walks. As

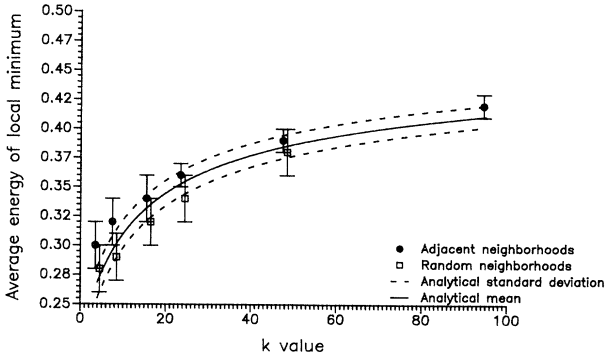

FIG. 1. Comparison of the energy of local optima for both the random and adjacent neighborhood models with the expressions given in the text for $N=96$ and the k values used in the tables. The $\mathbb{I}$ surrounding each data point shows the standard deviation of 100 simulated random adaptive walks; the solid and dotted lines show the analytical mean and standard deviation. The k values of the simulation data are shifted slightly to distinguish the two different sets of data points.

in Sec. III, we also assume that the variance of the site energy distribution is finite.

The analysis of Sec. III shows that the variance of the distribution of local minima is small, so that relatively high local minima are unlikely to trap the downhill walk prematurely. This explains why there is good agreement between the average energy of a local optimum as computed above and the simulation data, even though the average taken in the simulation data is biased by the number of starting points that can reach each optimum and the analytically computed average is not. The small variance in the distribution of local minima also explains why the energies of local optima reached by gradient and random adaptive walks are roughly the same (see Tables II–V).

The number of bits required to specify a particular local optimum is the base-2 logarithm of the number of local optima, or, asymptotically, $N(1-|\log_2\lambda|)$, where λ is the quantity in the second set of square brackets in (3.5). The asymptotic Hamming distance between neighboring local optima is therefore $N|\log_2\lambda|$, which must also be the average diameter of the basin of attraction of each optimum. We use the symbol D for their common value. As N increases, the probability of starting within $o(N)$ steps of the "edge" of one of these basins of attraction [i.e., $N|\log_2\lambda|/2+o(N)$ steps from an optimum] approaches unity.

This expression gives the mean length of gradient walks from random starts to local optima in the large-N limit. For finite N, the fact that a gradient walk will almost always step to the *nearest* local optimum must be taken into account. The probability that the random starting place for the walk is a distance d away from the nearest local optimum is the probability of having d of the D bits "wrong," given that one of the two nearest local optima has been chosen as a target, or

$$2^{1-D}\begin{bmatrix} D \\ d \end{bmatrix} \approx \left[\frac{8}{\pi D}\right]^{1/2} e^{-2(d-D/2)^2/D} .$$

The mean Hamming distance from the random start to the chosen optimum is then approximately

$$D-\left[\frac{8}{\pi D}\right]^{1/2}\int_{D/2}^{\infty} re^{-2(r-D/2)^2/D}dr=\frac{D}{2}-\left[\frac{D}{2\pi}\right]^{1/2} .$$

For even moderate k, the integral in the second set of square brackets in (3.5) is very close to unity, so that

$$D=|\log_2\lambda|\approx\frac{\log_2(k+1)}{k+1} . \tag{4.1}$$

For $k>4$, the agreement between the simulated local optima and the estimates suggests that the asymptotic regime has been attained. We note good agreement with (4.1) for longer walks for $k>4$, even with the finite-N correction ignored, and only moderate agreement for the shortest walks, even with the finite-N correction. Here, the Gaussian approximation to the binomial distribution breaks down.

The above result only provides a lower bound to the length of a random adaptive walk because the walk need not terminate at the nearest local optimum to its starting point. An alternative approach is to derive recursion re-

TABLE II. Mean energies of 100 simulated gradient walks from random starts to local optima for various N and k values. The numbers in parentheses are the standard deviations. For ease of comparison, $k=N-1$ entries appear where $k=N$, enclosed in square brackets. (a) Adjacent neighborhoods. (b) Random neighborhoods; results for $k=0$ and $k=N-1$, which would be the same as those in (a), are omitted.

k \ N	8	16	24	48	96
			(a) Adjacent neighborhoods		
0	0.34(0.08)	0.34(0.06)	0.34(0.05)	0.34(0.03)	0.34(0.02)
2	0.28(0.06)	0.29(0.04)	0.29(0.04)	0.29(0.02)	0.29(0.02)
4	0.29(0.06)	0.29(0.04)	0.29(0.03)	0.29(0.03)	0.29(0.02)
8	[0.32(0.04)]	0.31(0.04)	0.31(0.03)	0.31(0.02)	0.31(0.02)
16		[0.35(0.04)]	0.34(0.04)	0.34(0.03)	0.34(0.02)
24			[0.37(0.03)]	0.36(0.02)	0.35(0.02)
48				[0.39(0.02)]	0.38(0.02)
96					[0.42(0.01)]
			(b) Random neighborhoods		
0					
2	0.29(0.05)	0.29(0.04)	0.29(0.04)	0.29(0.03)	0.29(0.02)
4	0.29(0.06)	0.28(0.05)	0.29(0.04)	0.28(0.02)	0.29(0.02)
8		0.31(0.04)	0.30(0.03)	0.30(0.02)	0.29(0.02)
16			0.35(0.03)	0.33(0.02)	0.32(0.02)
24				0.35(0.02)	0.34(0.02)
48					0.38(0.02)
96					

TABLE III. Mean walk lengths of the walks used to compute (a) Table II(a) and (b) Table II(b).

k \ N	8	16	24	48	96
			(a)		
0	4.4(1.5)	8.4(1.9)	12.3(2.3)	24.1(3.7)	48.9(4.7)
2	2.9(1.2)	5.8(2.1)	8.4(2.5)	17.3(3.5)	34.3(5.1)
4	2.1(1.1)	4.3(1.7)	6.7(2.1)	12.5(3.4)	24.5(4.0)
8	[1.5(0.8)]	2.6(1.1)	4.2(1.7)	8.5(2.4)	16.5(3.1)
16		[1.8(1.0)]	2.5(1.1)	5.1(1.7)	9.5(2.1)
24			[1.7(1.0)]	3.2(1.2)	7.0(2.1)
48				[1.7(0.8)]	3.8(1.4)
96					[1.7(0.9)]
			(b)		
0					
2	3.0(1.1)	6.0(2.0)	8.9(2.4)	17.8(3.6)	35.1(5.6)
4	2.3(1.1)	4.6(1.9)	6.8(2.1)	13.7(3.1)	26.8(4.7)
8		2.9(1.2)	4.5(1.7)	9.1(2.6)	18.2(3.5)
16			2.4(1.1)	5.5(1.9)	11.3(3.2)
24				3.6(1.4)	7.9(2.3)
48					4.2(1.5)
96					

lations for the moments of the distribution of site energies after the $(n+1)$st step, given the values of these moments after the previous step. The derivation starts with the observation that, for the adjacent-neighbor model, each bit not only *affects* exactly $k+1$ sites but each site is also *affected by* $k+1$ bits. In the random-neighbor model, the number of sites that are affected by each bit is a random variable with mean $k+1$. We can therefore think of an adaptive step in both models as the replacement of $k+1$ old site energies in (2.1) by $k+1$ new site energies with a smaller average. We then have the recursion

$$\sum_{j=1}^{N} h_j^{(n+1)} = \sum_{j\notin\nu_i} h_j^{(n)} + \sum_{j\in\nu_i} r_j^{(n)} , \tag{4.2}$$

where $h_j^{(n)}$ is the energy of the jth site after the nth step and $r_j^{(n)}$ is the new site energy that results from replacement in the $(n+1)$st step. The neighborhood ν_i consists of site i, the bit that was flipped in the $(n+1)$st step, and the k neighboring sites. The variables $h_j^{(n)}$ are independent, identically distributed random variables, each with mean μ_n and variance σ_n^2. Because the r's are chosen subject to the condition that

TABLE IV. Mean energies of 100 simulated random adaptive walks from random starts to local optima for various N and k values. The numbers in parentheses are the standard deviations. For ease of comparison, $k=N-1$ entries appear where $k=N$, enclosed in square brackets. (a) Adjacent neighborhoods. (b) Random neighborhoods; results for $k=0$ and $k=N-1$, which would be the same as those in (a), are omitted.

k \ N	8	16	24	48	96
			(a) Adjacent neighborhoods		
0	0.35(0.08)	0.35(0.06)	0.34(0.04)	0.34(0.03)	0.34(0.02)
2	0.30(0.07)	0.30(0.04)	0.30(0.03)	0.30(0.02)	0.29(0.02)
4	0.30(0.06)	0.29(0.04)	0.30(0.04)	0.30(0.03)	0.30(0.02)
8	[0.34(0.06)]	0.32(0.04)	0.32(0.03)	0.31(0.02)	0.32(0.02)
16		[0.35(0.04)]	0.34(0.03)	0.34(0.02)	0.34(0.02)
24			[0.37(0.03)]	0.36(0.02)	0.36(0.02)
48				[0.40(0.02)]	0.39(0.02)
96					[0.42(0.01)]
			(b) Random neighborhoods		
0					
2	0.30(0.06)	0.29(0.04)	0.29(0.03)	0.29(0.03)	0.29(0.02)
4	0.32(0.05)	0.29(0.04)	0.29(0.04)	0.28(0.03)	0.28(0.02)
8		0.31(0.04)	0.31(0.04)	0.30(0.02)	0.29(0.02)
16			0.35(0.03)	0.33(0.03)	0.32(0.02)
24				0.35(0.02)	0.34(0.02)
48					0.38(0.02)
96					

TABLE V. Mean walk lengths of the walks used to compute (a) Table IV(a) and (b) Table IV(b).

k \ N	8	16	24	48	96
			(a)		
0	4.5(1.2)	8.6(1.9)	12.6(2.2)	24.3(3.4)	48.8(4.6)
2	4.1(1.9)	8.1(3.2)	11.2(3.1)	22.5(4.6)	45.2(6.6)
4	3.2(1.8)	6.6(2.5)	9.4(2.9)	19.3(3.9)	37.3(6.1)
8	[2.7(1.5)]	4.7(2.3)	7.7(3.0)	15.3(4.3)	27.7(5.3)
16		[3.3(1.7)]	4.8(2.1)	9.6(3.0)	19.3(4.2)
24			[3.5(1.4)]	7.4(3.0)	15.0(3.9)
48				[3.9(1.9)]	8.9(3.0)
96					[5.1(2.4)]
			(b)		
0					
2	4.4(1.8)	8.1(2.8)	12.5(3.8)	26.5(5.1)	46.9(6.1)
4	3.6(1.8)	7.3(2.9)	10.9(3.3)	22.9(5.6)	44.5(7.9)
8		5.3(2.5)	8.0(3.2)	17.0(4.3)	34.7(6.5)
16			4.8(2.1)	10.1(3.4)	21.6(4.8)
24				7.4(2.6)	16.0(4.3)
48					9.3(2.6)
96					

$$\sum_{j\in\nu_i} r_j^{(n)} < \sum_{j\in\nu_i} h_j^{(n)}, \tag{4.3}$$

we have the recursion relations

$$\mu_{n+1}=\rho\mu_n+(1-\rho)\mathcal{E}\left[\bar{r}^{(n)}\left|\sum_{j\in\nu_i} r_j^{(n)} < \sum_{j\in\nu_i} h_j^{(n)}\right.\right] \tag{4.4a}$$

and

$$\sigma_{n+1}^2=\rho\sigma_n^2+(1-\rho)(k+1)\mathrm{Var}\left[\bar{r}^{(n)}\left|\sum_{j\in\nu_i} r_j^{(n)} < \sum_{j\in\nu_i} h_j^{(n)}\right.\right]. \tag{4.4b}$$

Here, $\bar{r}^{(n)}$ is the average over the $k+1$ sites replaced at the $(n+1)$st step, and $\rho=1-[(k+1)/N]$ is the correlation coefficient in (1.1). At the beginning of the walk, the h's are simply samples from the underlying site energy distribution, and thus have the known mean μ_0, and variance σ_0^2. As the walk proceeds, site fitnesses (the h's) are replaced such that (4.3) is satisfied, and μ_n increases according to (4.4a). Note, however, that the replacement process does not change the statistical properties of the entries in the site energy tables, but merely selects some of them instead of others. In particular, the energy table entries *taken as a whole* continue to have mean μ_0 and variance σ_0^2, but only those sets of site energies satisfying (4.3) are replaced. For k sufficiently large, the sums in (4.3) and the average $\bar{r}^{(n)}$ are all asymptotically Gaussian, so that

$$\mathcal{E}\left[\bar{r}^{(n)}\left|\sum_{j\in\nu_i} r_j^{(n)} < \sum_{j\in\nu_i} h_j^{(n)}\right.\right]=\mu_{\bar{r}^{(n)}}=\left[\frac{k+1}{2\pi\sigma_n^2}\right]^{1/2}\int_{-\infty}^{\infty} e^{-(k+1)(r-\mu_n)^2/2\sigma_n^2}\left[\frac{\int_{-\infty}^{r} a e^{-(k+1)(a-\mu_0)^2/2\sigma_0^2}da}{\int_{-\infty}^{r} e^{-(k+1)(a-\mu_0)^2/2\sigma_0^2}da}\right]dr \tag{4.5a}$$

and

$$\mathrm{Var}\left[\bar{r}^{(n)}\left|\sum_{j\in\nu_i} r_j^{(n)} < \sum_{j\in\nu_i} h_j^{(n)}\right.\right]=\mathcal{E}\left[(\bar{r}^{(n)})^2\left|\sum_{j\in\nu_i} r_j^{(n)} < \sum_{j\in\nu_i} h_j^{(n)}\right.\right]-\mu_{\bar{r}^{(n)}}^2,$$

where

$$\mathcal{E}\left[(\bar{r}^{(n)})^2\left|\sum_{j\in\nu_i} r_j^{(n)} < \sum_{j\in\nu_i} h_j^{(n)}\right.\right]\left[\frac{k+1}{2\pi\sigma_n^2}\right]^{1/2}\int_{-\infty}^{\infty} e^{-(k+1)(r-\mu_n)^2/2\sigma_n^2}\left[\frac{\int_{-\infty}^{r} a^2 e^{-(k+1)(a-\mu_0)^2/2\sigma_0^2}da}{\int_{-\infty}^{r} e^{-(k+1)(a-\mu_0)^2/2\sigma_0^2}da}\right]dr. \tag{4.5b}$$

The recursions are terminated when the mean energy of a locally optimal configuration $\mu_0-\sigma_0[2\ln(k+1)/(k+1)]^{1/2}$ is reached. For sufficiently large N, the central limit theorem implies that the energy of the configuration, i.e., the average over all of the site energies, is asymptotically Gaussian with mean μ_n and variance σ_n^2/N. Finite-N corrections to the Gaussian distribution will be discussed below.

As a check on the above analysis, we compared μ_n and σ_n^2 as computed from (4.4) and as averaged over 100 simulations of downhill walks for $N=96$ and $k=16$ and 48, adjacent neighborhoods. The integrals were done numerically, using the NAG Library routines D01AMF and D01AJF (Mark 13) for the inner and outer integrals, respectively. All of the simulated walks started from the same initial configuration, which had energy 0.500 885 (as compared with $\mu_0=0.5$, exactly). As in Sec. III, both the calculation and the simulation assumed the underlying distribution of site energies to be the uniform distribution on the unit interval.

We note good agreement for $k=48$ (Fig. 2), but only qualitative agreement for $k=16$ (Fig. 3). This is not too surprising, because we expect the Gaussian approximation used in (4.5) to compute the conditional mean and variance in (4.4) to break down for small k, especially at the "tails" of the distribution, which are especially important in computing the conditional moments in (4.4). For sufficiently small k, it is probably not too much work to dispense with the Gaussian approximation and calculate these conditional moments directly, although we do not attempt this here. For intermediate k values, improved estimates of the conditional mean and variance can be computed via a more refined approximation involving corrections to the Gaussian distribution. For example [23], the density p_{k+1} of the average of $k+1$ independent identically distributed random variables with mean μ, variance σ^2, and third moment τ is

$$p_{k+1}\left[\sqrt{k+1}\,\frac{x-\mu}{\sigma}\right] = \mathcal{N}(x)\left[1+\frac{\tau}{6\sigma^3\sqrt{k+1}}H_3(x)\right]+o(1/\sqrt{k+1}),$$

uniformly in x. Here, $\mathcal{N}(x)$ is the Gaussian density with mean zero and unit variance, and $H_3(x)$ is the Hermite polynomial

$$H_3(x)=x^3-3x .$$

The unknown third moment τ_n of the site energy distribution after the nth step can be computed via the recursion

$$\tau_{n+1}=\rho\tau_n+(1-\rho)(k+1)^2 \times \mathcal{E}\left[(\bar{r}^{(n)}-\mu_{\bar{r}^{(n)}})^3 \,\middle|\, \sum_{j\in\nu_i} r_j^{(n)} < \sum_{j\in\nu_i} h_j^{(n)}\right] .$$

A similar procedure can be used to find the fourth and higher moments of the distribution.

For large N, we can approximate the recurrence relations (4.4) by differential equations. To do this, we introduce functions m and s of the continuous argument t and require that $m(n/N)=\mu_n$ and $s(n/N)=\sigma_n^2$. Using the approximations $\mu_{n+1}-\mu_n\approx(1/N)\dot{m}(t)$ and $\sigma_{n+1}^2-\sigma_n^2\approx(1/N)\dot{s}(t)$, we then have the differential equations

$$\dot{m}(t)=-(k+1)m+(k+1)\mu_{\bar{r}(t)} \tag{4.6a}$$

and

$$\dot{s}(t)=-(k+1)s +(k+1)^2\mathrm{Var}\left[\bar{r}(t)\,\middle|\, \sum_{j\in\nu_i} r_j(t) < \sum_{j\in\nu_i} h_j(t)\right] , \tag{4.6b}$$

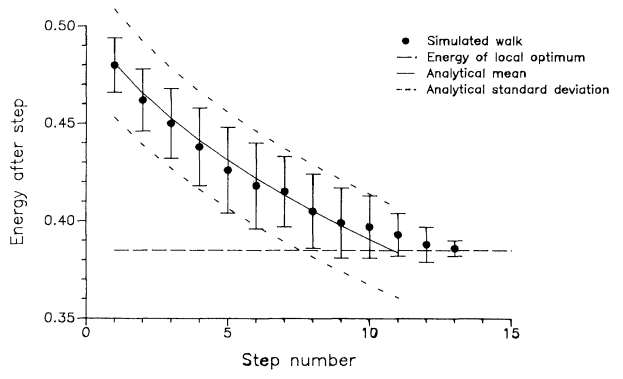

FIG. 2. Comparison of analytical estimates of the mean and standard deviation of the energies encountered along a random adaptive walk as a function of step number for $N=96$, $k=48$. The dashed horizontal line at the bottom represents the analytical estimate of the energy at a local optimum (≈ 0.38), the solid curve represents the analytically derived mean energies at each step, the dashed curves surrounding the solid curve represent the analytically derived standard deviation estimates. Each solid circle and the $\mathbb{I}$ surrounding it represents the mean and standard deviation of the 100 random adaptive walks described in the text. The standard deviations of the last several steps of the simulated walks are artificially low because some walks found local optima in previous steps.

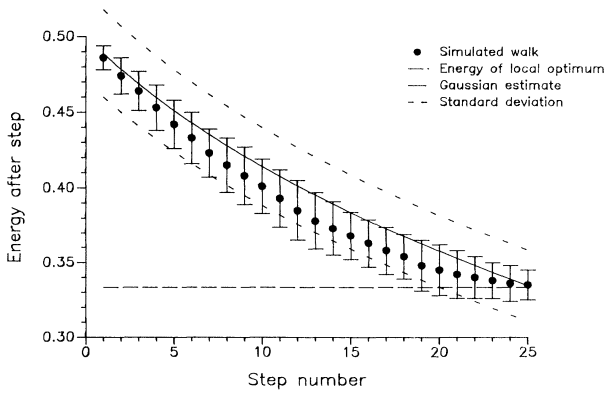

FIG. 3. Same as Fig. 3 for $N=96$, $k=16$. Here, the mean energy of a local optimum is about 0.34.

where the conditional moments in the above equations are now functions of the continuous variable t, and the quantity $\mu_{\bar{r}(t)}$ is the continuous time analog of $\mu_{\bar{r}(n)}$ in (4.5a).

An approximate solution to (4.6a) is straightforward, given the observations that much of the walk is spent near the local optimum and that much of the contribution to the outer integral in

$$\mu_{\bar{r}(t)}=\mu_0-\frac{\sigma_0}{\sqrt{2\pi(k+1)}}\int_{-\infty}^{\infty}d\xi\, e^{-\xi^2/2}\frac{e^{-[\xi\sqrt{s}+(m-\mu_0)\sqrt{k+1}]^2/2\sigma_0^2}}{\int_{-\infty}^{[\xi\sqrt{s}+(m-\mu_0)\sqrt{k+1}]/\sigma_0}e^{-a^2/2}da} \tag{4.7}$$

comes from values of $\xi\approx -1$. [Note that we must have $m(0)=\mu_0$, the mean of the underlying site energy distribution, and $s(0)=\sigma_0^2$, its variance.] Thus, for even moderate values of k, the quantity $\eta=[\xi\sqrt{s}+(m-\mu_0)\sqrt{k+1}]/\sigma_0\lesssim -3$, justifying the approximation

$$\int_{-\infty}^{\eta}e^{-a^2/2}da\approx -e^{-\eta^2/2}\left[\frac{1}{\eta}-\frac{1}{\eta^3}\right], \tag{4.8}$$

which is accurate to 3% when $\eta=-3$ and is asymptotically exact for large negative values of η. We then have

$$\begin{aligned}\mu_{\bar{r}(t)}&\approx\mu_0+\frac{1}{\sqrt{2\pi(k+1)}}\int_{-\infty}^{\infty}d\xi\, e^{-\xi^2/2}\frac{\xi\sqrt{s}+(m-\mu_0)\sqrt{k+1}}{1-\dfrac{\sigma_0^2}{[\xi\sqrt{s}+(m-\mu_0)\sqrt{k+1}]^2}}\\&\approx\mu_0+(m-\mu_0)+\frac{\sigma_0^2}{(k+1)(m-\mu_0)\sqrt{2\pi}}\int_{-\infty}^{\infty}d\xi\frac{e^{-\xi^2/2}}{1+\dfrac{\xi\sqrt{s}}{(m-\mu_0)\sqrt{k+1}}}\\&\approx\mu_0+(m-\mu_0)+\frac{\sigma_0^2}{(m-\mu_0)(k+1)}\,.\end{aligned}$$

The second line follows from the first by expanding the denominator as a geometric series, truncating after the second term, multiplying everything out, and observing that some of the terms in the result are odd functions of ξ, which make no net contribution to the integral. The third line follows from the second by simply ignoring the fraction in the denominator of the integrand in the previous line, and thus some higher-order correction terms. Strictly speaking, the integrands of all of the integrals above blow up for sufficiently positive ξ. However, this is an artifact of approximation (4.8) that could have been avoided by retaining higher-order terms that are discarded later. The simplified version of (4.6a) that results from the above calculation has solution $[m(t)-\mu_0]^2\approx 2\sigma_0^2 t$.

The walk stops at t^* such that

$$m(t^*)\approx\mu_0-\sigma_0\left[\frac{2\ln(k+1)}{k+1}\right]^{1/2},$$

the approximate mean value of the local optima computed in Sec. III. The length of the walk is thus roughly $\ln(k+1)/(k+1)$ in the scaled units, or $N\ln(k+1)/(k+1)$ in the original unscaled number of discrete steps, regardless of the mean μ_0 or variance σ_0^2 of the underlying site energy distribution. This estimate is compared in Fig. 4 with the simulation data for $N=96$ and various k values.

V. SUMMARY AND CONCLUSIONS

The primary focus of this article has been the study of energy landscapes that represent an intermediate case between the quadratically coupled landscapes of spin-glass physics, and the random energy model. The N-k model is, of course, a k-ary spin glass: It is always possible to write *any* function $f(\underline{e})$ that assigns a real number to a $k+1$ bit string $\underline{e}=(e_1,e_2,\ldots,e_{k+1})$ in the form

$$f(\underline{e})=\alpha_0+\sum_{p_1}\alpha_{p_1}e_{p_1}+\sum_{p_1,p_2}\alpha_{p_1p_2}e_{p_1}e_{p_2}+\cdots+\sum_{p_1,\ldots,p_{k+1}}\alpha_{p_1\cdots p_{k+1}}e_{p_1}\cdots e_{p_{k+1}},$$

where the α's are real numbers. In particular, the site energy function h_i can be written in this form. Because the N-k Hamiltonian is just the average of such expressions, this Hamiltonian contains some of the $\binom{N}{k+1}$ products $e_{p_1}\cdots e_{p_{k+1}}$ plus some of the products of every lower order, and is therefore a generalization of the familiar quadratically coupled spin-glass Hamiltonian. The need for such a generalization is suggested by an unsuccessful attempt [7] to fit a "landscape" of sequence-dependent RNA kinetic parameters to the quadratic scheme. Given the assumptions described in the introduction, N-k landscapes provide this generalization, in the sense that

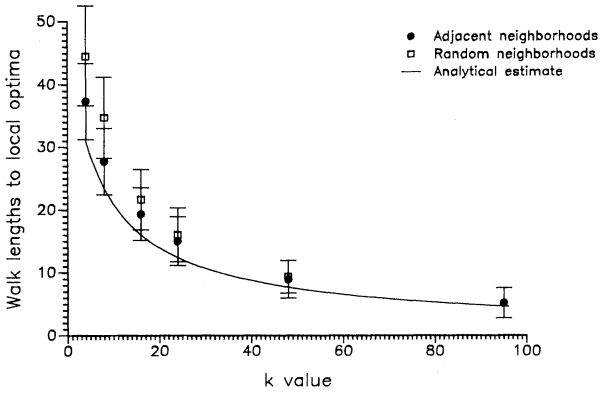

FIG. 4. Comparison of the length of the random adaptive walks used in Fig. 2 with the analytical estimate derived in the text.

all AR(1) landscapes on the Boolean hypercube must have approximately the same statistical properties as N-k landscapes. N-k landscapes with $k \approx 40$ have been used to capture many of the important statistical properties of the maturation of the immune response in mice [2]. Preliminary results obtained in fitting the free energy landscapes for RNA folding [24] are also encouraging.

As mentioned in the Introduction, the high-correlation ($k=0$) limit of the N-k model leads to a landscape with a single local optimum that is also the global optimum. If $k>0$ but $k=O(1)$ and $N \gg 1$, widely separated points on the landscape have significant correlations. For both the random and adjacent neighor models, the form of the autocorrelation functions $R(d)$ for $d \ll N$ implies that

$$\mathrm{Var}[H-H'] \approx \frac{(k+1)d}{N}\sigma^2$$

for energies H and H', separated by Hamming distance $d \ll N$. Sorkin [21] has noted the similarity between this scaling law and that of Brownian motion, for which

$$\mathrm{Var}[X(t_2)-X(t_1)] \propto |t_2-t_1| ,$$

and he argues that many of the same fractal properties of Brownian motion should be found in landscapes where the above scaling law applies. In particular, we expect that energy "valleys" should appear on all scales, including the largest possible scale, the diameter of the landscape. The bowl shape of highly correlated N-k landscapes has been confirmed by computer simulations showing that the deepest local minima lie nearest the global minimum [27]. So many of the properties of Gaussian energy landscapes seem to be related to the correlation coefficient ρ that we speculate that all such landscapes lie in the same "universality class." In particular, the fact that the topography of N-k landscapes seems to be independent of how the neighborhoods are chosen lends strong support to the argument that the Parisi ansatz for the solution to the Sherrington-Kirkpatrick spin-glass model—for which $\rho=1-O(N^{-1})$—is also correct for "real" (i.e., Edwards-Anderson) spin glasses.

As we saw in Secs. III and IV, the long-range (ultrametric?) correlations in highly correlated (small-k) landscapes disappear as k increases and the correlation decreases. Instead, the landscape breaks up into uncorrelated patches whose radii are on the order of the correlation length. We saw, in particular, that both gradient and random adaptive walks from random starts extend for somewhat more than one correlation length. Increasing k also increases the energy of local minima, in contrast to the energy of the global minimum, which stays roughly constant [15].

In the high-k ($k=N-1$) limit, we recover the fully uncorrelated, random energy model. Similar questions to ours have been asked about the random energy model by other authors [19], and it is of interest to compare their results with ours. For example, the number of local optima in the uncorrelated case was $2^N/(N+1)$ but $(2\lambda)^N$, where λ is the quantity in the second set of square brackets in (3.5), for the correlated case. In both cases, there is a boundary layer of local optima, but the location of the boundary layer depends on the distribution of energies in the landscape, which differs in the two models. The final and most striking difference between the two models is the difference in the magnitude of downhill walk lengths: $O(\ln N)$ in the uncorrelated case and $O(N|\ln\lambda|)$ in the correlated case.

The underlying premise of the N-k model—that the energy of a given configuration is a sum of independent random variables, some of which are replaced upon moving to a neighboring configuration—is extremely general. For example, the traveling salesman problem is essentially an N-k model on the permutation group of the N cities to be visited: interchanging the order in which two cities are visited replaces four intercity distances by other effectively random distances. As a result, the number of local optima is close to what one would expect from the N-k estimates; that is, approximately the number of configurations divided by the number of configurations within a correlation length of each other [20]. The general situation is that any collection of jointly Gaussian random variables $\{X_1, X_2, \ldots, X_M\}$ has a "spectral resolution," that is, a representation of the form [28]

$$X_i = \sum_{j=1}^{M} T_{ij} Z_j ,$$

where the T_{ij} are constants and $\{Z_1, Z_2, \ldots, Z_M\}$ is a sequence of M independent random variables whose variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2$ are the eigenvalues of the covariance matrix of the X's. For a wide class of landscapes, it can be shown [29,30] that relatively few of the Z's with significant variance change when a single step is made in a highly correlated landscape, but many more of them change when the landscape is relatively uncorrelated. The number of such "significant variables" is crudely measured by the parameter k in the N-k model.

When there can be $s>2$ symbols per site, the analysis changes only in minor ways. In fact, nothing changes in (2.3), provided we reinterpret $\underline{0}_i$ as a configuration with any of the $s-1$ symbols other than 0 at site i and 0's elsewhere. Because there are now $s-1$ such inequalities for each i, rather than just one,

$$Z(\beta,s)=\int_{-\infty}^{\infty} dP(h_1)\int_{-\infty}^{\infty} dP(h_2)\cdots\int_{-\infty}^{\infty} dP(h_N)e^{-\beta(\underline{1}^T\underline{h})/N}\left\{\prod_{i=1}^{N}\left[1-P_{k+1}\left[\sum_{j\in\nu_i} h_j\right]\right]\right\}^{s-1}. \tag{5.1}$$

This expression can also be evaluated by the techniques described above. We thus conclude that there are $O((s\lambda)^N)$ local optima and that the typical distance between optima is $N|\log_s\lambda|$, where λ now depends on both k and s. If the site energy distribution has finite variance, a trivial extension of the methods of Sec. III yields the resulting bounds for h_m:

$$\mu-\sigma\left[\frac{(2-\delta)\ln[(s-1)(k+1)]}{k+1}\right]^{1/2}<h_m<\mu-\sigma\left[\frac{2\ln[(s-1)(k+1)]}{k+1}\right]^{1/2}, \tag{5.2}$$

and the estimate

$$N\left[\frac{s-1}{s}\right]|\log_s\lambda|\approx N\frac{(s-1)\log_s[(s-1)(k+1)]}{s(k+1)}$$

for the typical distance from a randomly chosen configuration to a local optimum. [Typically, a randomly chosen symbol will fail to have a specified value with probability $(s-1)/s$.] Figure 5 shows a comparison between (5.2) and the mean energy of a local optimum for 100 simulated random adaptive walks in which the underlying site energy distribution is, once more, the uniform distribution on the unit interval. The only change in the derivation of the walk length estimate is that (5.2) must be used instead of (3.4) in estimating $m(t^*)$, resulting in the s-dependent estimate of $\log_s[(s-1)]/(k+1)$ for the walk length. A comparison of this estimate with the simulation data is shown in Fig. 6.

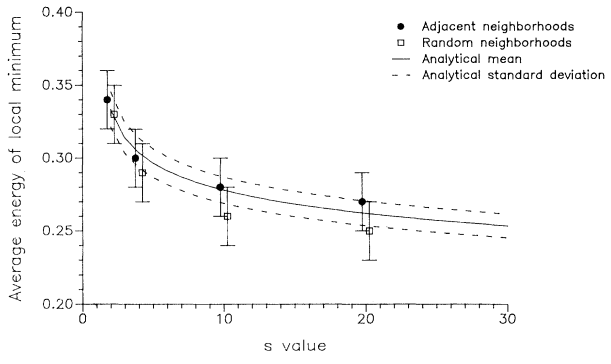

FIG. 5. Dependence of the energy of local optima for both the random and adjacent neighbor models with s, the number of possible symbols per site for $N=48$, $k=16$, as estimated from 100 simulations of random adaptive walks from random starts, and analytical estimates. Simulation data is for $s=2$, 4, 10, and 20, but the s values of the plotted points are shifted slightly to distinguish the two different kinds of data. The $\mathbb{I}$ surrounding each data point shows the standard deviation of 100 simulation runs, and the solid and dotted lines show the analytical mean and standard deviation.

A limitation in our method is the requirement that the site energy distributions satisfy the hypotheses of the central limit theorem. One distribution that fails to satisfy these hypotheses is the Cauchy distribution [23], whose density is given by

$$\text{Prob}\{x\le X\le+dx\}=\frac{1}{\pi(1+x^2)}dx$$

for $-\infty<x<\infty$ and therefore fails to have a finite mean. It is easy to prove the remarkable fact that the average of n such random variables has the same distribution as one of them, and thus the magnitude of a sum of Cauchy random variables will almost surely be determined by only a single one of its terms. This characteristic of such sums is reminiscent of a typical biological enzyme, whose effectiveness is largely due to the correct placement of a few specific amino acids in a short "active region." In contrast, the statistical properties of sums of random variables that satisfy the central limit theorem are determined essentially equally by all terms in the sum, corresponding to a scenario in which all amino acids in the enzyme make equal contributions to its activity. Thus, the Cauchy distribution may be a more biologically appropriate distribution than Kauffman's original choice of the uniform distribution for the h_i's, but it can be shown [31] that the resulting landscapes still have the same qualitative asymptotic N dependence: the probability that a randomly chosen point is a local optimum is λ^N for λ independent of N, the expected energy of a local optimum is independent of N, and the variance of the distribution of locally optimal energies is $O(1/N)$.

The quantity we defined as $Z(\beta)$ is actually the true partition function when the sum is restricted to local op-

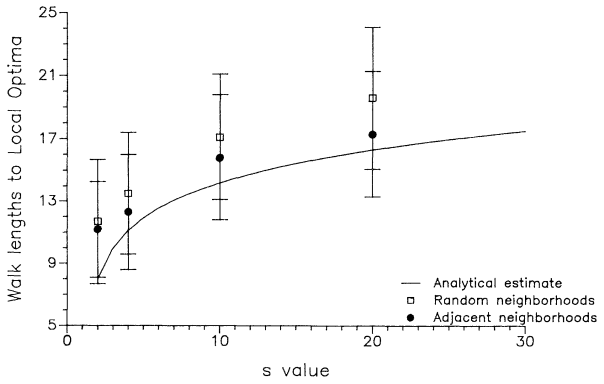

FIG. 6. Comparison of the length of the random adaptive walks used in Fig. 5 with the analytical estimate derived in the text.

tima and the results are averaged over the ensemble of possible energy tables. (Let us refer to this quantity by $\langle Z \rangle$ in this paragraph to emphasize the average that is being taken.) In view of the comments in the Introduction, the restriction to local optima may be physically sensible. However, our results cannot be directly applied to thermodynamic calculations on N-k landscapes because it is only the average $\langle \ln Z \rangle$, rather than $\ln\langle Z \rangle$, which is extensive, and is therefore a sensible expression for the average free energy in the thermodynamic limit. Our method for identifying local optima, combined with the famous "replica trick," might yield a result for $\langle \ln Z \rangle$. More directly obtainable from our formalism and transfer matrix arguments [31] is the fact that Z (this time the unaveraged version) is still the trace of a (random) product of transfer matrices whose eigenvalues depend analytically on β for finite k. For $k=N-1\rightarrow\infty$, though, the N-k model becomes the random energy model, which is known to have a phase transition. This shows once more that the important thing about a spin-glass model is not whether its interactions are long range, but whether each spin interacts with a significant fraction of the other spins.

ACKNOWLEDGMENTS

This work began as a collaboration with Marco Isopi and Ronnie Mainieri at the First Summer School on Complex Systems, held under the auspices of the Santa Fe Institute in Sante Fe, New Mexico. The author would like to thank Stuart Kauffman for many hours of useful discussions. Additional discussions were held with Manfred Eigen, Jerry Percus, Richard Palmer, Peter Stadler, and Dan Stein. Funding from the Department of the Navy under ONR Grant No. N00014-85-K-0258 and from the Max Planck Society in support of various stages of this work is gratefully acknowledged.

[1] S. Kauffman and S. Levin, J. Theor. Biol. **128**, 11 (1987).
[2] S. Kauffman and E. Weinberger, J. Theor. Biol. **141**, 211 (1989).
[3] S. Kauffman, E. Weinberger, and A. Perelson, in *Theoretical Immunology, Santa Fe Institute Studies in the Sciences of Complexity,* edited by A. S. Perelson (Addison-Wesley, Reading, MA, 1988), Pt. 1.
[4] J. McCaskill, J. Chem. Phys. **80**, 5194 (1984).
[5] M. Eigen, Ber. Bunsenges. Phys. Chem. **89**, 658 (1985).
[6] M. Eigen, Chem. Scr. **26B**, 13 (1986).
[7] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[8] J. Holland, *Adaptation in Natural and Aritificial Systems* (University of Michigan Press, Ann Arbor, MI, 1975).
[9] R. Brady, Nature **317**, 804 (1985).
[10] R. G. Palmer, in *Complex Systems,* Proceedings of the 1988 Summer School on Complex Systems in Santa Fe, New Mexico, edited by D. L. Stein (Addison-Wesley, Reading, MA, 1989).
[11] D. L. Stein, *Complex Systems* (Ref. [10]).
[12] K. Binder and A. Young, Rev. Mod. Phys. **54**, 801 (1986).
[13] A. J. Bray and M. A. Moore, J. Phys. C **13**, L469 (1980).
[14] J. Gillespie, Evolution **38**, 1116 (1984).
[15] E. Weinberger, Biol. Cybern. **63**, 325 (1990).
[16] W. Ewens, *Mathematical Population Genetics* (Springer, Berlin, 1979).
[17] B. Derrida, Phys. Rev. B **24**, 2613 (1981).
[18] E. Weinberger, J. Theor. Biol. **134**, 125 (1988).
[19] C. Macken and A. Perelson, Proc. Natl. Acad. Sci. USA **86**, 6191 (1989).
[20] P. Stadler and W. Schnabl, Phys. Lett. A (to be published).
[21] G. Sorkin, IBM Research Report No. RC13674 (No. 61253) 1988 (unpublished).
[22] L. Breiman, *Probability* (Addison-Wesley, Reading, MA, 1968).
[23] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1965).
[24] W. Fontana, T. Griesmacher, P. Schuster, P. Stadler, P. Tarazona, and E. Weinberger (unpublished).
[25] E. Weinberger, Acta Appl. Math. (to be published).
[26] W. Fontana, W. Schnabl, and P. Schuster, Phys. Rev. A **40**, 3301 (1989).
[27] S. Kauffman, *Complex Systems* (Ref. [10]).
[28] M. Priestley, *Spectral Analysis and Time Series* (Academic, London, 1981).
[29] E. Weinberger, *Measuring Correlations in Energy Landscapes and Why it Matters,* in *Information Dynamics,* Vol. 256 of NATO *Advanced Study Institute Series B: Physics* edited by H. Atmanspacher and H. Scheingraber (Plenum, New York, 1991).
[30] E. Weinberger, Biol. Cybern (to be published).
[31] M. Isopi, R. Mainieri, and E. Weinberger (unpublished).