

## Effective degrees of freedom in genetic algorithms

C. R. Stephens and H. Waelbroeck

*Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, Apartado Postal 70-543, México D.F. 04510, Mexico*

(Received 18 November 1996; revised manuscript received 14 May 1997)

An evolution equation for a population of strings evolving under the genetic operators, selection, mutation, and crossover, is derived. The corresponding equation describing the evolution of schemata is found by performing an exact coarse graining of this equation. In particular, exact expressions for schema reconstruction are derived that allow for a critical appraisal of the “building-block hypothesis” of genetic algorithms. A further coarse graining is made by considering the contribution of all length- $l$  schemata to the evolution of population observables such as fitness growth. As a test function for investigating the emergence of structure in the evolution, the increase per generation of the *in-schemata fitness* averaged over all schemata of length  $l$ ,  $\Delta_l$ , is introduced. In finding solutions to the evolution equations we concentrate more on the effects of crossover; in particular, we consider crossover in the context of Kauffman  $Nk$  models with  $k=0,2$ . For  $k=0$ , with a random initial population, in the first step of evolution the contribution from schema reconstruction is equal to that of schema destruction leading to a scale invariant situation where the contribution to fitness of schemata of size  $l$  is independent of  $l$ . This balance is broken in the next step of evolution, leading to a situation where schemata that are either much larger or much smaller than half the string size dominate those with  $l \approx N/2$ . The balance between block destruction and reconstruction is also broken in a  $k>0$  landscape. It is conjectured that the effective degrees of freedom for such landscapes are *landscape connective trees* that break down into effectively fit smaller blocks, and not the blocks themselves. Numerical simulations confirm this “connective tree hypothesis” by showing that correlations drop off with connective distance and not with intrachromosomal distance. [S1063-651X(98)10002-8]

PACS number(s): 87.10.+e, 02.50.-r, 05.50.+q

### I. INTRODUCTION

One of the most important steps in developing a qualitative or quantitative model of a system is to gain an understanding of the nature of its effective degrees of freedom (EDOF). An important feature that distinguishes them is that their mutual interactions are not very strong; that is to say, they must have a certain degree of integrity. In this sense, the aim of developing an effective model of a system is to arrive at a description of the system in terms of *relevant* (e.g., “macroscopic”) variables.

Identifying the correct EDOF in complex systems is generally speaking a very difficult task. To begin with, more often than not they are scale dependent, where what one means by “scale” depends on the particular problem under consideration. In the case of evolution theory and genetic algorithms (GAs), one expects to find different EDOF at different time scales. Generically if a system is complex at the relevant scale then it will admit a simple effective dynamics only in terms of complex degrees of freedom: one trades the complicated dynamics that results from the nonlinear interactions of the many “elementary” degrees of freedom for the simpler dynamics of more complicated EDOF. What one gains in the trade is effective predictability; what one loses is detail.

It is well worth recalling in this context the example of spin glass models of neural networks [1–3]. In this case the EDOF are the overlaps with a certain number of “patterns,” each of which is related to a local extremum of the energy landscape or Hamiltonian. Since a large number of uncorrelated patterns is involved in this effective representation, it

should be clear that the description of the EDOF themselves requires a large amount of information: One gets a measure of the complexity of the system by the information in its EDOF. Note that in this example the system’s dynamics is guided by large-scale attracting structures (the patterns), the EDOF (overlaps) being the instruments that measure how structure emerges as the system condenses from a disordered phase. Some other examples of structured complex systems include the brain, gene expression in eukaryotic cells [4], and of course evolution theory and GAs, among many others. We know that these systems are *structured* because their behavior is manifestly nonrandom; for instance, neural dynamics must be structured if the brain is to be of any use. Yet in most cases we have no idea what the nature of this structure is, much less how to identify EDOF.

In this paper we will begin to analyze the notion of EDOF in the context of GAs [5,6]. We emphasize, however, that GAs form only one area of interest where the results and conclusions of this paper are applicable, some others being statistical mechanics [7], biology [4], the Kauffman  $Nk$  model [8], and evolution theory [9].

Trying to ascertain what EDOF a GA is using in order to arrive at an optimal solution is in the strict sense a nonsensical question—roughly equivalent to asking “what are the EDOF of a block of material?” Of course, the answer depends on the type of material under consideration and its state. However, it is *not* nonsensical to think of what are the EDOF in a generic type of fitness landscape. The fitness landscapes we choose to consider as being representative of general classes of fitness landscapes are Kauffman’s  $Nk$  models with  $k=0$  and  $k=2$ .

As in the example of spin glasses, the dynamics of GAs can be viewed as a condensation process in a rugged landscape. So again one expects the EDOF to represent the emergence of certain structures, or “patterns,” which are related to local fitness optima. In GA theory one usually considers partly specified patterns, called “schemata,” and determines the fraction of all the individuals in the population that include a particular schema, this being a measure of order comparable to the “overlap” of spin glass models. Since one does not know *a priori* which schemata lead to a useful set of EDOF, some hypothesis must be made to this effect. The standard conjecture about the EDOF of GAs is the “building block hypothesis” [5,6], the essence of which is that a GA arrives at an optimal solution of a complex problem via the combination of *short*, fit schemata. In this paper we will present both analytic and numerical evidence that generically this is not the case. The argument for the block hypothesis is that large schemata are likely to be “broken” by the crossover operator. However, this argument neglects the possibility that a schema be reconstructed via parents that contain its constituent parts.

It is clear that the validity of the block hypothesis will depend on the nature of the fitness landscape. If there is a larger contribution to fitness from widely separated string bits, large schemata will be favored irrespective of the effect of crossover. On the contrary, if the landscape strongly favors smaller schemata, this would lend support to the block hypothesis. However, the intuition behind the block hypothesis is firmly based on the action of crossover not on the pathologies of particular landscapes. It is for this reason that we choose to consider the block hypothesis in the context of Kauffman  $Nk$  models. We will always assume that the fitness landscape is generic in the sense that there is no systematic bias in the fitness function that would favor one part of the string over another.

The format of the paper will be as follows. In Sec. II, as this paper is not intended for a dedicated GA audience, we will give a brief overview of various elements of GA theory. In Sec. III we will derive an evolution equation for the development of a population of strings under the genetic operators of selection, mutation, and simple crossover. We then “coarse grain” this equation to derive an effective evolution equation for the evolution of schemata of size  $l$  and order  $N_2$ . In Sec. IV we consider a further coarse graining, considering the effects of schemata of size  $l$  but of any order  $N_2 \leq l$ . We consider especially the increment in fitness per generation from such schemata showing that under general assumptions the coarse-grained variable is closely related to the spatial correlation function, so it provides information about the size distribution of the EDOF. In Sec. V we consider asymptotic solutions of the coarse-grained evolution equation near a random initial population for a simple “neutral” fitness landscape and also make some comments about what happens near the ordered population limit. In Sec. VI we consider a more nontrivial landscape with  $k=2$  where we introduce a conjecture about the EDOF of direct-encoded GAs on an  $Nk$  landscape, which we call the “connective tree hypothesis.” Finally in Sec. VII we summarize our conclusions.

## II. GENETIC ALGORITHMS AND THE BUILDING BLOCK HYPOTHESIS

GAs have become increasingly popular in the analysis of complex search and optimization problems and in machine learning, one of their chief attributes being their robustness (see [10] and references therein for a recent overview). One begins with a complex optimization problem that depends on many variables. The variables and the rules that govern them are subsequently coded in the form of a population of strings (“chromosomes”). The latter consist of a set of symbols (“alleles”), each symbol taking values defined over an alphabet.

The population evolves under the action of a set of genetic operators. Reproduction can be implemented in many different ways; all have the effect of increasing the relative numbers of “fit” strings, fitness being measured by a fitness function  $f: \mathcal{A}_s \rightarrow \mathbb{R}_+$ , where  $\mathcal{A}_s$  is the space of string states. The role of most other genetic operators is to encourage diversity in the population. We will restrict our attention to simple crossover and mutation. The former is a type of recombination and involves the splitting of two parents,  $C_i, C_j \in \mathcal{A}_s$ , at a particular crossover point  $k$ , and the subsequent juxtaposition and recombination of the left half of  $C_i$  with the right half of  $C_j$  and the right half of  $C_i$  with the left half of  $C_j$ , left and right being defined relative to the crossover point  $k$ . Crossover is one method of generating strings that were not originally in the population of a given generation, thus providing diversity. Mutation, on the other hand, offers a form of insurance, in that if a particular bit is lost it is irrecoverable using only reproduction and crossover. Mutation offers a way to recover lost bits that may subsequently be important in the construction of an optimum string. Using the language of statistical mechanics the evolution of the GA is a competition between the “ordering” tendency of reproduction and the “disordering” effects of crossover and mutation.

Theoretical analysis of how a GA seeks an optimum solution has focused on the notion of schemata. For strings of  $N$  bits a schema is a subset,  $N_2 \leq N$ , of bits defining a certain “word” constructed from the alphabet. In the  $N - N_2$  positions not defined by the schema one does not care about the value of the bit, and this is taken into account by use of the metasymbol, or “wild card”,  $*$ . The essential idea behind the notion of the schema is that the GA arrives at an optimum solution through combining fit schemata. As each string is an example of  $\sim 2^{N_2}$  schemata, it is clear that a very large number of them are being processed simultaneously by the GA, a phenomenon known as implicit parallelism [5]. Of course, not all these schemata survive crossover, which leads us to consider the size of a schema,  $l$ , which is defined as  $(l = j - i + 1)$ , where  $i$  and  $j$  are the first and last of the  $N_2$  defining elements of the schema, respectively. In terms of reproduction and mutation there is no preference for short versus long schemata, except as might be induced by the fitness function itself. However, if one considers the effects of crossover, purely in terms of the crossover point itself there is a higher probability of “breaking” a long schema than a short one. This apparent disfavor of large schemata led to what is known as the “building block” hypothesis, which claims that the joint effect of reproduction and cross-

over is to favor highly fit but *short schemata that propagate from generation to generation exponentially. It is these highly fit, short schemata that are then considered to be the EDOF in the system, the GA building a better solution through combining small subsolutions.*

### III. STRING EVOLUTION EQUATION

In this section we will derive an equation that describes the evolution of a GA induced by the effects of the three genetic operators: selection, crossover, and mutation. In particular, we will consider the change in the number  $n(\xi, t)$  of strings that contain a particular schema  $\xi$ , of order  $N_2$  and size  $l \geq N_2$ , as a function of time (generation) in a population of size  $n$ .

We will first derive evolution equations for the ‘‘microscopic’’ degrees of freedom themselves—the strings. Considering first selection in the absence of mutation or crossover, one has

$$P(C_i, t+1) = P'(C_i, t), \quad (1)$$

where  $P'(C_i, t) = [f(C_i, t)/\bar{f}(t)]P(C_i, t)$ ,  $f(C_i, t)$  is the fitness of string  $C_i$  at time  $t$ ,  $P(C_i, t) = n(C_i, t)/n$ , and  $\bar{f}(t) = \sum_i f(C_i, t)P(C_i, t)$  is the average string fitness. In Eq. (1) we are neglecting fluctuations in the numbers  $n(C_i, t)$ , an approximation that becomes exact in the infinite population limit. The effect of reproduction is to augment the number of fit strings. However, the trouble with using selection as the sole genetic operator is that the search space for optima is restricted to that of the initial population. In complex systems this number will often be small compared to the size of the total search space. This implies that finite-size effects are important.

Including the effects of mutation but not crossover gives rise to the quasispecies model [11], with the evolution equation

$$P(C_i, t+1) = \mathcal{P}(C_i)P'(C_i, t) + \sum_{C_j \neq C_i} \mathcal{P}(C_j \rightarrow C_i)P'(C_j, t), \quad (2)$$

where  $\mathcal{P}(C_i) = \prod_{k=1}^N [1 - p(k)]$  is the probability that string  $i$  remains unmutated,  $p(k)$  being the probability of mutation of bit  $k$ , which we assume to be a constant, though the equations are essentially unchanged if we also include a dependence on time.  $\mathcal{P}(C_j \rightarrow C_i)$  is the probability that string  $j$  is mutated into string  $i$ ,

$$\mathcal{P}(C_j \rightarrow C_i) = \prod_{h \in \{C_j - C_i\}} p(k) \prod_{h \in \{C_j - C_i\}^c} [1 - p(k)], \quad (3)$$

where  $\{C_j - C_i\}$  is the set of bits that differ between  $C_j$  and  $C_i$ , and  $\{C_j - C_i\}^c$ , the complement of this set, is the set of bits that are the same. In the limit where the mutation rate  $p$  is uniform,  $\mathcal{P}(C_i) = (1-p)^N$  and  $\mathcal{P}(C_j \rightarrow C_i) = p^{d^H(i,j)}(1-p)^{N-d^H(i,j)}$ , where  $d^H(i,j)$  is the Hamming distance between the strings  $C_i$  and  $C_j$ . The behavior of the solutions to Eq. (2) has been much discussed in the literature (see, for example, [7] and references therein), although mainly in the context of a flat fitness landscape. One of the principal features of interest is the existence of an ‘‘error threshold’’

separating an ‘‘ordered’’ (selection dominated) phase from a ‘‘disordered’’ (mutation dominated) phase that manifests itself as a second order phase transition at a certain critical mutation rate. In the case of uniform mutations, Eq. (2) can be mapped into an equilibrium statistical mechanics problem using transfer matrix techniques [12], where the role of inverse temperature is played by  $\beta = \frac{1}{2} \ln[p/(1-p)]$ .

We will now consider the effects of crossover without mutation. This is a much less studied case theoretically, but one that is very important from the point of view of EDOF, since unlike mutations it is sensitive to the linear disposition of bits along the string. Some work has been done, notably in the context of GAs by Prughel-Bennett and Shapiro [13], whose method models GA evolution via a small set of macroscopic quantities and assumes that the system entropy is maximized, and by Higgs [14] in an interesting generalization of the Eigen model to diploid models that include crossover. Neither of the two approaches focuses on schemata, the main interest of the present paper. With crossover the evolution equation can be written in the form

$$\begin{aligned} P(C_i, t+1) = & P'(C_i, t) - \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{k=1}^{N-1} \mathcal{C}_{C_i C_j}^{(1)}(k) \\ & \times P'(C_i, t)P'(C_j, t) \\ & + \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} \mathcal{C}_{C_j C_l}^{(2)}(k) \\ & \times P'(C_j, t)P'(C_l, t), \end{aligned} \quad (4)$$

where  $p_c$  is the probability of implementing crossover in the first place,

$$\mathcal{C}_{C_i C_j}^{(1)}(k) = \theta(d_L^H(i,j))\theta(d_R^H(i,j)) \quad (5)$$

and

$$\begin{aligned} \mathcal{C}_{C_j C_i}^{(2)}(k) = & \frac{1}{2} (\delta(d_L^H(i,j))\delta(d_R^H(i,l)) \\ & + \delta(d_R^H(i,j))\delta(d_L^H(i,l))), \end{aligned} \quad (6)$$

where  $d_R^H(i,j)$  is the Hamming distance between the right halves of the strings  $C_i$  and  $C_j$ , ‘‘right’’ being defined relative to the crossover point  $k$ . The other quantities are defined analogously.  $\mathcal{C}_{C_i C_j}^{(1)}(k)$  is the probability that given that  $C_i$  was one of the parents it is destroyed by the crossover process.  $\mathcal{C}_{C_i C_l}^{(2)}(k)$  is the probability that given that neither parent was  $C_i$  it is created by the crossover process, so this represents a gain term. It is naturally much easier to destroy an individual string by crossover than create it; hence  $\mathcal{C}_{C_j C_i}^{(2)}(k)$  is a very sparse matrix.  $\mathcal{C}_{C_i C_l}^{(2)}(k)$  represents a contact interaction term in Hamming space. Another important property of  $\mathcal{C}_{C_i C_j}^{(1)}(k)$  and  $\mathcal{C}_{C_j C_l}^{(2)}(k)$  is that they are completely population independent, depending only on string configurations and not string numbers.

Equation (4) is an extension of the ‘‘schema theorem,’’ or fundamental theorem of GAs, [5,6] that states that, for a schema  $\xi$ , of size  $l$ ,

$$P(\xi, t+1) \geq P'(\xi, t) \left[ 1 - p_c \left( \frac{l-1}{N-1} \right) \right], \tag{7}$$

to the case where the schema of interest is the entire string (an analogous equation was derived in [15]). The evolution equation we have derived takes into account exactly, given the approximation of a large population, the effects of destruction and reconstruction of strings.

Combining the effects of both crossover and mutation, where we assume that mutation is carried out after crossover, we have the evolution equation

$$P(C_i, t+1) = \mathcal{P}(C_i) P_c(C_i, t) + \sum_{C_j \neq C_i} \mathcal{P}(C_j \rightarrow C_i) P_c(C_j, t),$$

where

$$\begin{aligned} P_c(C_i, t) &= P'(C_i, t) - \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{k=1}^{N-1} \mathcal{C}_{C_i C_j}^{(1)}(k) \\ &\quad \times P'(C_i, t) P'(C_j, t) \\ &\quad + \frac{p_c}{N-1} \sum_{C_j \neq C_i} \sum_{C_l \neq C_i} \sum_{k=1}^{N-1} \mathcal{C}_{C_j C_l}^{(2)}(k) \\ &\quad \times P'(C_j, t) P'(C_l, t). \end{aligned} \tag{8}$$

We now turn our attention to the derivation of an evolution equation for schemata. Before doing this it is convenient to return to Eq. (4) to see that the notions of schema and coarse graining appear very naturally when considering the crossover of strings. Considering the destruction term: the matrix (5) restricts the sum to those  $C_j$  that differ from  $C_i$  in at least one bit both to the left and to the right of the crossover point. One can convert the sum over  $C_j$  into an unrestricted sum by subtracting off those  $C_j$  that have  $d_L^H(i, j) = 0$  and/or  $d_R^H(i, j) = 0$ . Similarly one may write the reconstruction term as

$$\begin{aligned} &\frac{p_c}{N-1} \sum_{k=1}^{N-1} \left( \sum_{C_j} \sum_{C_l} \mathcal{C}_{C_j C_l}^{(2)}(k) P'(C_j, t) P'(C_l, t) \right. \\ &\quad - 2 \sum_{C_j} \mathcal{C}_{C_i C_j}^{(2)}(k) P'(C_i, t) P'(C_j, t) \\ &\quad \left. - P'(C_i, t) P'(C_i, t) \right). \end{aligned} \tag{9}$$

The second and third terms cancel with corresponding expressions from the destruction term; hence Eq. (9) can be written as  $\sum_{C_j \supset C_i^L} \sum_{C_l \supset C_i^R} P'(C_j, t) P'(C_l, t)$ , where  $C_i^L$  is the part of  $C_i$  to the left of the crossover point and correspondingly for  $C_i^R$ . However, by definition  $\bar{f}(C_i^L, t) = (1/n_{C_i^L} \sum_{C_j \supset C_i^L} P'(C_j, t))$ , where  $n_{C_i^L}$  is the total number of strings in the population that contains  $C_i^L$ . As  $\bar{f}(C_i^L, t)$  is the average fitness of the substring  $C_i^L$ , one can think of this substring as a schema; likewise for  $C_i^R$ . In terms of these ‘‘schemata’’ the final form of the string equation is

$$\begin{aligned} P(C_i, t+1) &= P'(C_i, t) - \frac{p_c}{N-1} \sum_{k=1}^{N-1} [P'(C_i, t) \\ &\quad - P'(C_i^L, t) P'(C_i^R, t)], \end{aligned} \tag{10}$$

with  $P'(C_i^L, t) = \sum_{C_j \supset C_i^L} P'(C_j, t)$ , and similarly for  $P'(C_i^R, t)$ .

One thus sees that crossover naturally introduces the notion of coarse graining, even though we are working in terms of the microscopic degrees of freedom—the strings. The reconstruction probability depends on the relative fitness of strings that contain the constituent elements of  $C_i$ , but given that there can be many strings that contain  $C_i^L$  one must take an average over these strings. In this sense we are integrating out the ‘‘degrees of freedom’’ represented by the bits that are not contained in  $C_i^L$  or  $C_i^R$ . Equation (10) shows that the effects of reconstruction will outweigh destruction if the parts of a string are more selected than the whole.

Before deriving a schema evolution equation including crossover and mutation let us consider the effects of reproduction alone. The proportion of elements of the population,  $P(\xi, t)$ , that contains  $\xi$  satisfies the evolution equation

$$P(\xi, t+1) = P'(\xi, t), \tag{11}$$

where  $P'(\xi, t) = [\bar{f}(\xi, t)/f(t)] P(\xi, t)$ ,

$$\bar{f}(\xi, t) = \frac{\sum_{C_i \supset \xi}^{n(\xi, t)} f(C_i, t) n(C_i, t)}{n(\xi, t)}, \tag{12}$$

the sum is over all strings  $C_i$  that contain  $\xi$ , and  $\bar{f}(t) = \sum_{\xi} \bar{f}(\xi, t) P(\xi, t) / \sum_{\xi} P(\xi, t)$  is the average fitness per string or per schema of the population. Note that the sum over strings that contain  $\xi$  is a sum over the possible values of the bits that are not definite elements of  $\xi$ , i.e., the wild cards. In this sense, as above in Eq. (10), ‘‘degrees of freedom’’ have been integrated out of the problem and Eq. (11) represents an exact coarse graining of the original string evolution equation.

Considering mutation without crossover we ‘‘coarse grain’’ the microscopic equation (2) by summing over all  $C_i \supset \xi$ . One can write an effective evolution equation for schemata evolving under mutation

$$P(\xi, t+1) = \mathcal{P}(\xi) P'(\xi, t) + \sum_{\xi_i} \mathcal{P}(\xi_i \rightarrow \xi) P'(\xi_i, t), \tag{13}$$

where the effective coefficients  $\mathcal{P}(\xi)$  and  $\mathcal{P}(\xi_i \rightarrow \xi)$  are

$$\mathcal{P}(\xi) = \prod_{k=1}^{N_2} [1 - p(k)],$$

$$\mathcal{P}(\xi_i \rightarrow \xi) = \frac{\sum_{C_j \supset \xi_i} P'(C_j, t) \mathcal{P}(C_i \rightarrow C_j)}{P(\xi, t) \bar{f}(\xi, t)}. \tag{14}$$

In the latter the sum is over strings  $C_j$  that contain the schemata  $\xi_i$ , where  $\xi_i$ , differs in at least one bit from  $\xi$  on the  $N_2$  defining bits of the schema.

To derive an evolution equation for schemata, including in the effects of crossover, we return to Eq. (10) and sum over all strings  $C_i \supset \xi$ . One finds

$$\begin{aligned}
P(\xi, t+1) &= (1-p_c)P'(\xi, t) \\
&+ \frac{p_c}{N-1} \sum_{C_i \supset \xi} \sum_{k=1}^{N-1} P'(C_i^L, t)P'(C_i^R, t).
\end{aligned} \tag{15}$$

We now break the sum over crossover points into those that cut the schema itself and those that cut outside the schema. In the reconstruction term if the cut is outside the schema, to the right say, then the sum over  $C_i^R$  is 1. Similarly if the cut is to the left, the sum over  $C_i^L$  is 1. The remaining sums yield  $P'(\xi, t)$  and this term cancels with an analogous expression originating in the destruction term. For the reconstruction contribution from cuts in the schema we denote by  $\eta_L$  ( $\eta_R$ ) the bits to the left (right) of the crossover point that are *not* in the schema and note that  $\sum_{C_i \supset \xi} P'(C_i^L, t)P'(C_i^R, t) = \sum_{\eta_L} \sum_{\eta_R} P'(C_i^L, t)P'(C_i^R, t)$ . We will denote by  $\xi_L$  and  $\xi_R$  the parts of the schema to the left and right of the crossover point, respectively. Now,  $\sum_{\eta_L} P'(C_i^L, t) = P'(\xi_L, t)$ , where by definition  $P'(\xi_L, t) = [\bar{f}(\xi_L, t)/\bar{f}(t)]P(\xi_L, t)$ ,  $\bar{f}(\xi_L, t)$  being the average fitness of the schema  $\xi_L$ . Analogous expressions hold for  $\xi_R$ . With these results the final form of the schema evolution equation including crossover is

$$\begin{aligned}
P(\xi, t+1) &= P'(\xi, t) - \frac{p_c}{N-1} \sum_{k=1}^{l-1} [P'(\xi, t) \\
&- P'(\xi_L, t)P'(\xi_R, t)],
\end{aligned} \tag{16}$$

where the sum is only over crossover points that cut the schema.

The interpretation of this equation is very similar to that of Eq. (10). In the reconstruction term  $P'(\xi_L, t)P'(\xi_R, t)$  is the probability that one parent is selected that contains the left part of the schema and the other contains the right part. A schema will be augmented by the effects of crossover if, as in the string case, its constituent parts are selected more than the whole. Compared with Eq. (10) a further coarse graining has been carried out by summing over all the states of bits outside of  $\xi$ . Combining now the effects of selection, mutation, and crossover the schema evolution equation is

$$P(\xi, t+1) = \mathcal{P}(\xi)P_c(\xi, t) + \sum_{\xi_i} \mathcal{P}(\xi_i \rightarrow \xi)P_c(\xi_i, t), \tag{17}$$

where

$$\begin{aligned}
P_c(\xi, t) &= P'(\xi, t) - \frac{p_c}{N-1} \sum_{k=1}^{l-1} [P'(\xi, t) \\
&- P'(\xi_L, t)P'(\xi_R, t)].
\end{aligned} \tag{18}$$

This evolution equation is the fundamental equation governing the evolution of schemata and is written at a ‘‘semimicroscopic’’ level, in that it is written in terms of individual schemata. It represents an exact coarse graining of the corresponding string evolution equation after summing over all possible states of the nonschema degrees of freedom.

Another useful concept we will introduce here is that of ‘‘effective fitness,’’  $f_{\text{eff}}(\xi, t)$ , which we define via the relation

$$P(\xi, t+1) = \frac{f_{\text{eff}}(\xi, t)}{\bar{f}(t)} P(\xi, t). \tag{19}$$

Comparing with Eq. (17) one finds

$$\begin{aligned}
f_{\text{eff}}(\xi, t) &= \mathcal{P}(\xi) \left[ 1 - p_c \left( \frac{l-1}{N-1} \right) \right] \bar{f}(\xi, t) \\
&+ \sum_{\xi_i} \mathcal{P}(\xi_i \rightarrow \xi) \frac{P'(\xi_i, t)}{P(\xi, t)} \bar{f}(\xi_i, t) \\
&+ p_c \left( \frac{l-1}{N-1} \right) \frac{\mathcal{P}(\xi)}{P(\xi, t)} \bar{f}(t) P'(\xi_L, t) P'(\xi_R, t) \\
&- \frac{p_c}{N-1} \sum_{\xi_i} \mathcal{P}(\xi_i \rightarrow \xi) \bar{f}(t) \\
&\times \sum_{k=1}^{N-1} \left( \frac{P'(\xi_i, t) - P'(\xi_{i_L}, t)P'(\xi_{i_R}, t)}{P(\xi, t)} \right).
\end{aligned} \tag{20}$$

Thus we see that the effect of mutation and crossover is to ‘‘renormalize’’ the ‘‘bare’’ fitness  $\bar{f}(\xi, t)$ . The destructive effects of crossover and mutation give a multiplicative-type renormalization while the reconstruction terms give an additive-type renormalization. In the low ‘‘temperature’’ limit where mutation and crossover go to zero,  $f_{\text{eff}}(\xi, t) \rightarrow \bar{f}(\xi, t)$ . The above also leads to the idea of an effective selection coefficient  $s_{\text{eff}} = f_{\text{eff}}(\xi, t)/\bar{f}(t) - 1$ . If we think of  $s_{\text{eff}}$  as being approximately constant in the vicinity of time  $t_0$ , then  $s_{\text{eff}}(t_0)$  gives us the exponential rate of increase or decrease of growth of the schema  $\xi$  at time  $t_0$ .

#### IV. EFFECTIVE DEGREES OF FREEDOM AND COARSE GRADING

As mentioned, one of the most important steps in obtaining a qualitative and quantitative understanding of a system is deciding what are its relevant degrees of freedom. This is often difficult owing to the fact that they are ‘‘scale’’ dependent. In the case of evolution dynamics this implies that the effective dynamics depends on the time scale considered. Trying to understand such behavior quantitatively is very difficult as almost inevitably one will have to resort to an approximation technique, which invariably depends on focusing on the relevant EDOF, as in the methods of effective field theory. However, if they are time dependent then what starts as a good approximation focusing on a certain type at one time will usually break down as one approaches time scales where they are qualitatively quite different.

One feature that is very common, if one has found a reasonable set of EDOF, is that their mutual interactions are not very strong, so that they have a certain degree of integrity. Calling something an EDOF is not a very useful thing to do if it is not readily identifiable as such. For instance, in low energy QCD, gluons and quarks are not very useful concepts, as they are so strongly coupled via highly nonlinear interactions that they form baryons and mesons, bound states of the

former. The latter have a much higher degree of integrity than the former at such energies.

So how is the above related to the present discussion of GAs? GAs, as algorithmic representations of complex systems, have many degrees of freedom and therefore an exponentially large number of possible states. For instance, in the case where the state of a string of size  $N$  is defined as a binary word, for a population  $n$  the total number of possible states is  $\sim(2^N)^n$  in the case where strings are identifiable by a label other than the state of their bits, and  $\sim n2^N$  in the case where permutations of identical strings are not counted separately. Ideally, the search for optima proceeds in a smaller space, spanned by effective ‘‘coarse-grained’’ degrees of freedom. The traditional answer to the question, ‘‘What is the nature of these degrees of freedom?’’ is, as mentioned previously, given by the ‘‘building block hypothesis.’’

We can get some idea of the dynamical behavior of schemata due to crossover by restricting attention for the moment to a flat fitness landscape. In this case, for an uncorrelated population, crossover is completely neutral and we have a scale invariant situation. To solve the evolution equation in the case of a correlated population one needs to solve the corresponding equations for  $\xi_L$  and  $\xi_R$ ; these will involve reconstruction terms that contain  $\xi_{LL}$ ,  $\xi_{LR}$ ,  $\xi_{RL}$ , and  $\xi_{RR}$ . The first two are the components of  $\xi_L$  and the latter two of  $\xi_R$ . Naturally this process can be iterated relating fine grained degrees of freedom to more and more coarse grained degrees of freedom, where more and more bits ( $N-N_2$ ) have been summed over. Obviously when one arrives at one schema the process stops as one cannot split by crossover such schema. Thus crossover leads to a hierarchy of equations relating fine grained degrees of freedom to successively more and more coarse grained degrees of freedom.

Restricting attention to two schemata in the flat fitness landscape setting and considering the continuous time limit, one arrives at the following differential equation:

$$\frac{dP(ij,t)}{dt} = -p_c \frac{l-1}{N-1} [P(ij,t) - P(i,t)P(j,t)], \quad (21)$$

where  $i$  and  $j$  are the definite bits that define the 2 schema and also the two 1 schemata, respectively. As one cannot split a one schema  $P(i,t)$  and  $P(j,t)$  are conserved quantities; thus one finds

$$P(ij,t) = P(ij,0)e^{-p_c[(l-1)/(N-1)]t} + P(i,0)P(j,0) \times (1 - e^{-p_c[(l-1)/(N-1)]t}). \quad (22)$$

Thus one sees that  $P(ij,t)$  approaches an uncorrelated fixed point  $P^*(ij) = P(i,0)P(j,0)$  exponentially rapidly. The sole effect of the size of the schema is to govern the rate of approach to the fixed point, an exponentially small preference being given to smaller schemata.

The steady state solution for a schema  $\xi$  of order  $N_2$  is

$$P^*(\xi) = \prod_{i=1}^{N_2} P(\xi(i),0), \quad (23)$$

where  $P(\xi(i),0)$  is the probability of finding the one schema corresponding to the  $i$ th bit of  $\xi$  at  $t=0$ . One can verify that this steady state solution also is purely a result of the effects

of reconstruction. Without reconstruction there is no other fixed point other than zero. Thus reconstruction is the driving force of crossover and will always come to dominate. This is very much contrary to the standard block hypothesis point of view that treats schema destruction as the dominant effect. We can also make another interesting observation associated with the effective fitness  $f_{\text{eff}}(\xi,t)$  and crossover. Here the effect of crossover is to renormalize the fitness. The effective selection coefficient is

$$s_{\text{eff}} = -p_c \left( \frac{l-1}{N-1} \right) + p_c \left( \frac{l-1}{N-1} \right) \frac{P(i,0)P(j,0)}{P(ij,t)}. \quad (24)$$

Thus schema destruction gives a multiplicative renormalization that contributes negatively to the effective fitness advantage. However, schema reconstruction leads to an additive renormalization of the effective fitness that exceeds the contribution of the destruction term if  $i$  and  $j$  are negatively correlated.

In general the fitness landscape itself induces correlations between  $\xi_L$  and  $\xi_R$ . In this case there is a competition between the (anti-) correlating effect of the landscape and the mixing effect of crossover. Selection itself more often than not induces an anticorrelation between fit schema parts, rather than a positive correlation. Indeed, in the neutral case of a  $k=0$  landscape when  $\delta f_{\xi_L}, \delta f_{\xi_R} > 0$ , then  $1 + (2N_2/N)\delta f_{\xi} < [1 + (2N_L/N)\delta f_{\xi_L}][1 + (2N_R/N)\delta f_{\xi_R}]$ . So selection induces an anticorrelation; hence in an uncorrelated initial population,  $P'(\xi,t) < P'(\xi_L,t)P'(\xi_R,t)$ . This means that crossover plays an important role in allowing both parts of a successful schema to appear in the same individual.

We can analyze this effect in more detail, taking once again the case of 2 schemata. Defining the correlation  $\mathcal{C}(ij,t) \equiv [P(ij,t)/P(i,t)P(j,t)] - 1$ , then in terms of the selection coefficient,  $s_{\xi} = f(\xi,t)/\bar{f} - 1$ , one finds

$$\mathcal{C}(ij,t+1) = \left( 1 - p_c \frac{l-1}{N-1} \right) \left( \frac{(1+s_{ij})}{(1+s_i)(1+s_j)} \mathcal{C}(ij,t) - \frac{(s_i s_j + s_i + s_j)}{(1+s_i)(1+s_j)} \right). \quad (25)$$

Note that the effect of crossover is to diminish correlations induced by the fitness landscape; however, crossover cannot change the sign of the correlations. The larger the value of  $l$  in this simple case the more the correlations are damped. This is the effect that we saw previously in the context of a flat landscape. In the extreme case  $l=N$ ,  $p_c=1$  the effect of crossover is to eliminate all correlation between  $i$  and  $j$ . In the neutral ( $k=0$ ) case,  $s_{ij} = s_i + s_j$  and

$$\mathcal{C}(ij,t+1) = \left( 1 - p_c \frac{(l-1)}{N-1} \right) \left( \frac{(1+s_i+s_j)}{(1+s_i)(1+s_j)} \mathcal{C}(ij,t) - \frac{(s_i s_j + s_i + s_j)}{(1+s_i)(1+s_j)} \right). \quad (26)$$

Thus the effect of crossover is to weaken but not cancel completely the anticorrelations induced by  $k=0$  selection. In the remainder of this section we will consider this effect for general schemata.

In our search for the relevant EDOF and in analyzing the building block hypothesis we will now consider schemata of length  $l$ , irrespective of their order or their overall position in a string. This is a further coarse graining relative to the evolution equations considered earlier. The evolution equation (17) by itself is not very useful for analyzing schemata of size  $l$ , as any given string contains schemata of all sizes. However, consideration of just about any quantity in conjunction with Eq. (17) and a sum over schemata of a given length is meaningful. Our notation here is that for any function  $A(\xi, t)$ ,

$$\langle A(t) \rangle_l = \frac{\sum_{i=1}^{(N-l+1)} \sum_{N_2=2}^l \sum_{\{N_2\}} \sum_{\text{words}} P(\xi, t) A(\xi, t)}{(N-l+1)2^{l-2}}, \quad (27)$$

where  $l \geq 2$ . The first sum is over the possible beginning point  $i$  of the schema and the following two sums represent the different configurations of any number  $N_2 \leq l-2$  specified bits chosen among the  $l-2$  available sites. The number of available sites is  $l-2$  because we fix the ends bits.

Using Eq. (17) one may derive a recursion relation for the expectation value of  $A$ ; however, time dependence enters not only in the changing probability distribution  $P(\xi, t+1)$ , which can be substituted using Eq. (17), but also in  $A(\xi, t+1)$ . This occurs even though many observables of interest are time-independent functions of the string states as the summing over degrees of freedom associated with passing to a more coarse grained description induces an implicit time dependence in the coarse grained observables. For example,  $\bar{f}(\xi, t)$  is a population-dependent observable, even though  $f(C_i)$  is not. To simplify matters, to search for structure in the population we define a time-independent function on schemata: the average selective advantage that in-schema bits would enjoy if the schema were immersed in a random population,

$$\delta f_\xi = \left( \frac{N}{N_2} \right) \left( \frac{1}{2^{N-N_2}} \right) \sum_{\eta-\text{words}} (f(\xi, \eta) - \frac{1}{2}), \quad (28)$$

where  $\eta$  represents the out-of-schema bits and the average fitness in a random population has been normalized to  $1/2$ . Note that here and in the rest of the paper we are looking at the fitness deviation per schema bit as opposed to Sec. III, where the total fitness of a schema was being considered. This observable corresponds to the *effective fitness* of in-schema bits if the population is in fact random, or if the landscape assigns an independent fitness contribution to each bit in the chromosome ( $k=0$  in the terminology of the Kauffman  $Nk$ -model). In general, it is a useful test function with which one can probe for the emergence of structure during the first steps of evolution away from a random initial population. We will refer to this observable below as the *in-schema fitness*.

We will make use below of the following simplified averages: if  $A(\xi)$  is independent of the initial defining point of the schema, or if the landscape is “generic,” then we can sum over this point to find

$$\langle A(t) \rangle_l = \frac{1}{2^{l-2}} \sum_{N_2=2}^l \sum_{\{N_2\}} \sum_{\text{words}} P(\xi, t) A(\xi). \quad (29)$$

By “generic” we mean that within the class of landscapes we are considering, such as an  $Nk$  model for a particular value of  $k$ , there is no systematic bias in the fitness function for a particular part of the string, i.e., the sums over words, configurations, and  $N_2$  lead to an average that is effectively translation invariant (the system is effectively self-averaging). We will also use the notation  $\langle\langle A \rangle\rangle_1$  to represent the average over schemata and over crossover points, namely,  $\langle\langle A \rangle\rangle_l = 1/(l-1) \sum_{i=1}^{l-1} \langle A \rangle_l$ .

Considering the expectation value of the in-schema fitness, the equation that gives the improvement of  $\langle \delta f_\xi \rangle_l$  from generation  $t$  to generation  $t+1$  is

$$\Delta_l = \left\langle\left\langle \delta f_\xi \frac{\delta f_{\text{eff}}(\xi, t)}{f(t)} \right\rangle\right\rangle_l, \quad (30)$$

where  $\delta f_{\text{eff}}(\xi, t) = f_{\text{eff}}(\xi, t) - \bar{f}(t)$ . More explicitly, using the evolution equation for schemata, one finds

$$\Delta_l = \left\langle\left\langle \frac{\delta f_\xi \delta \bar{f}(\xi, t)}{\bar{f}(t)} \right\rangle\right\rangle_l - p_c \left( \frac{l-1}{N-1} \right) \left\langle\left\langle \frac{\delta f_\xi}{P(\xi, t)} [P'(\xi, t) - P'(\xi_L, t) P'(\xi_R, t)] \right\rangle\right\rangle_l, \quad (31)$$

where  $\delta \bar{f}(\xi, t) = \bar{f}(\xi, t) - \bar{f}(t)$ . The first term is independent of  $l$  in a random population if the fitness landscape itself is  $l$  independent.

As defined,  $\Delta_l$  measures the average improvement of the in-schema fitness over one step of evolution. How does this improvement come about? First of all, schemata with  $\delta \bar{f}(\xi, t) > 0$  will be more frequent in the parent population, thanks to the selection factor  $(1+s)$ , where  $s = \delta \bar{f}(\xi, t)/\bar{f}(t) = (2N_2/N) \delta f_\xi$ , where the latter equality is only true for a random initial population or  $k=0$  model. The next step is to consider the action of the crossover operator. On the other hand, selected parents with  $\xi$  may not pass it on to their offspring if crossover “breaks” the schema. However, there is a possibility that  $\xi$  will be reconstructed from parents that have parts of  $\xi$  but not all of it. The question is, which is larger for a particular value of  $l$ ? Before we turn to answering this question in particular cases, let us consider the relation between  $\Delta_l$  and the spatial correlation function.

If the population is uncorrelated; in other words, if  $P(\xi, t) = \prod_i P(\xi_i, t)$ , where  $\xi_i$  is the  $i$ th bit of  $\xi$ , then the expectation value of  $\delta f_\xi$  is independent of  $l$ , as  $N_2 \delta f_\xi = \sum_i \delta f_{\xi_i}$  and  $\langle \sum_i \delta f_{\xi_i} \rangle_l$  is just the uncorrelated sum of contributions from  $1$  schemata. The fact that the existence of correlations in  $P(\xi, t+1)$  implies an  $l$  dependence can be demonstrated explicitly. One writes

$$f_{\text{eff}}(\xi, t) \approx \frac{1}{N_2} \sum_{i=1}^{N_2} f_1(\xi_i) + \frac{1}{N_2(N_2-1)} \times \sum_{i=1}^{N_2} \sum_{j \neq i} \{f_2(\xi_i \xi_j) - \frac{1}{2}[f_1(\xi_i) + f_1(\xi_j)]\}, \quad (32)$$

where

$$f_1(\xi_i) = \frac{1}{2^{N_2-1}} \sum_{\{\xi_k, k \neq i\}} f_{\text{eff}}(\{\xi_k\}), \quad (33)$$

$$f_2(\xi_i \xi_j) = \frac{1}{2^{N_2-2}} \sum_{\{\xi_k, k \neq i, j\}} f_{\text{eff}}(\{\xi_k\}), \quad (34)$$

and we are considering only up to two-point correlations. Defining  $\delta s_\xi = \delta f_{\text{eff}}(\xi, t) - \delta f_\xi$ , which is a measure of the selective advantage over and above the in-schema fitness, one finds

$$\delta s_\xi \approx \frac{1}{N_2} \sum_{i=1}^{N_2} f_1(\xi_i) + \frac{1}{N_2(N_2-1)} \sum_{i=1}^{N_2} \sum_{j \neq i} \{f_2(\xi_i \xi_j) - \frac{1}{2}[f_1(\xi_i) + f_1(\xi_j)]\} - \bar{f}(t) - \delta f_\xi. \quad (35)$$

For a  $k=0$  landscape,

$$\delta s_\xi \approx \frac{1}{N_2(N_2-1)} \sum_{i=1}^{N_2} \sum_{j \neq i} \{f_2(\xi_i \xi_j) - \frac{1}{2}[f_1(\xi_i) + f_1(\xi_j)]\}. \quad (36)$$

So, in this case we see that the existence of a selective advantage is due to the existence of correlations in the effective fitness. Defining a selective coefficient  $s_l$  that represents the selective advantage for a schema to be of size  $l$ , one finds

$$\Delta_l = \langle\langle \delta f_\xi^2 \rangle\rangle_l (1 + s_l), \quad (37)$$

where

$$s_l = \frac{\langle\langle \delta f_\xi^2 \delta s_\xi \rangle\rangle_l}{\langle\langle \delta f_\xi^2 \rangle\rangle_l}. \quad (38)$$

In this expression for  $\Delta_l$ ,  $\langle\langle \delta f_\xi^2 \rangle\rangle_l$  is independent of  $l$  for a random initial population. Thus we see that any  $l$  dependence can be attributed to the existence of spatial correlations.

If the reconstruction term from crossover exceeds the destruction term for some  $l$ , one can conclude that the fitness improvement attributed to a particular bit in the string depends on its being part of selected schemata of this size. That the conditioning information on the existence of other specified bits should be useful, is a direct consequence of the correlations between the different bits in the string. We emphasize the relation between  $\Delta_l$  and the correlation function because correlations are intimately linked to the emergence of EDOF. In this sense, the function  $\Delta_l$  is related to the expected size distribution of the EDOF.

## V. ASYMPTOTIC SOLUTIONS

In this section we consider some asymptotic solutions of the evolution equation for  $\Delta_l$  derived in Sec. IV. In particular, we will consider two limiting cases: the evolution of schemata starting from a completely random initial state, and a random perturbation around a completely ordered state. As one of our principal considerations is in investigating the validity of the building block hypothesis, we will set the mutation rate to zero, as the effects of the latter do not depend on schema size. We will derive expressions for  $\Delta_l(t+1)$  and  $\Delta_l(t+2)$  starting out from an initial random population at time  $t$ .

For a random initial population at time  $t$ ,

$$\Delta_l(t+2) = \langle\langle \delta f_\xi(t+2) \rangle\rangle_l - \Delta_l(t+1). \quad (39)$$

Even though  $\delta f_\xi$  is time independent, we use the above notation to indicate that its expectation value is with respect to the probability distribution at time  $t+2$ . In the initial random population the effective schema fitness is the in-schema fitness  $\delta f_\xi$  and  $\delta \bar{f}(\xi, t) = (N_2/N) \delta f_\xi$ ,  $P(\xi, t) = 1/2^{N_2}$ . Thus one finds

$$\Delta_l(t+1) = \left(1 - p_c \left(\frac{l-1}{N-1}\right)\right) \langle\langle \alpha \rangle\rangle_l + p_c \left(\frac{l-1}{N-1}\right) \times \left\langle\left\langle \frac{N_L}{N_2} \beta_L + \frac{N_R}{N_2} \beta_R + \frac{4N_L N_R}{N^2} \delta f_\xi \delta f_{\xi_L} \delta f_{\xi_R} \right\rangle\right\rangle_l, \quad (40)$$

where we have introduced the notation for the quadratic terms

$$\alpha = \frac{1}{2^{N_2}} \sum_{\text{words}} \frac{2N_2}{N} \delta f_\xi^2 \quad \beta_L = \frac{1}{2^{N_2}} \sum_{\text{words}} \frac{2N_2}{N} \delta f_\xi \delta f_{\xi_L}, \quad (41)$$

with an analogous expression for  $\beta_R$ .

We will now derive explicit results in some concrete cases based on generic fitness landscapes. The Kauffman  $Nk$  models provide such a set of landscapes. Here we will specialize to the case  $k=0$ , which is neutral in the sense that it neither favors nor disfavors correlations between bits. In the  $k=0$  landscape,

$$\delta f_\xi = \frac{N_L}{N_2} \delta f_{\xi_L} + \frac{N_R}{N_2} \delta f_{\xi_R}. \quad (42)$$

We also have that  $\langle\langle \delta f_\xi \delta f_{\xi_L} \delta f_{\xi_R} \rangle\rangle_l = 0$ , which results in the complete cancellation of the destruction and reconstruction crossover terms, the final result being

$$\Delta_l(t+1) = \langle\langle \alpha \rangle\rangle_l. \quad (43)$$

The above expression is for an arbitrary  $k=0$  landscape. To find a more explicit solution we must consider a more explicit landscape. We will consider two: a binary landscape where the fitness of a bit may only take two values, 1 and 0; and a landscape where the fitness of a bit is selected uniformly at random from the interval  $[0,1]$ . Both landscapes conform with the requirement that the average fitness per bit



in a random population be  $1/2$ . Let  $x_i$  denote the deviation from the mean fitness of bit number  $i$ , i.e.,  $x_i = f_i - 1/2$ . The averaging over configurations then gives, for  $l \geq 3$ ,

$$\langle\langle \alpha \rangle\rangle_l = \frac{2}{N(N-l+1)l(l-1)(l-2)} \sum_{i=1}^N m_i x_i^2,$$

where

$$m_i = l(l-1)(l-2) \quad (l \leq i \leq N-l+1),$$

$$m_i = (l^2 - 3l) + i(l^2 - 5l + 8) + \frac{l-2i}{2^{l-2}} \quad (i < l),$$

and symmetrically for  $i > N-l$ . Explicitly,

$$\langle\langle \alpha \rangle\rangle_l = \frac{c}{6N}, \quad (44)$$

where  $c=3$  for the binary landscape and 1 in the random landscape. In the latter for large  $N$  we have assumed that the average over the  $N$  bits (weighted by  $n_i$ ) can be replaced by an average over the distribution of  $x_i$  used to generate the landscape. Thus one sees that crossover acts in a scale invariant way at the first time step of evolution from a random initial population: there is no preference whatsoever for small blocks at the expense of large blocks.

We will now consider what happens at time  $t+2$ . The extra ingredient we need relative to the above calculation is  $\langle\langle \delta f_{\xi}(t+2) \rangle\rangle_l$ . To calculate this we in turn need to calculate  $P'(\xi, t+1)$ ,  $P'(\xi_L, t+1)$ , and  $P'(\xi_R, t+1)$ , i.e., the selection probabilities at time  $t+1$ , calculation of which requires knowledge of  $\bar{f}(\xi, t+1)$ ,  $\bar{f}(\xi_L, t+1)$ ,  $\bar{f}(\xi_R, t+1)$ , and  $\bar{f}(t+1)$ . Specializing once again to a  $k=0$  landscape, one finds

$$P'(\xi, t+1) = \frac{1}{2^{N_2}(1+2\alpha_s)} \left\{ \left( 1 + \frac{2N_2}{N} \delta f_{\xi} \right)^2 + 2 \frac{(N-N_2)}{N} \alpha_{(N-N_2)} + \frac{p_c}{N-1} \right. \\ \times \left( 1 + \frac{2N_2}{N} \delta f_{\xi} \right) \sum_{k=1}^{l-1} \frac{4N_L N_R}{N^2} \delta f_{\xi_L} \delta f_{\xi_R} \\ + \frac{4p_c}{N(N-1)} \left[ \sum_{k=1}^{l-1} (N_R \delta f_{\xi_R} \beta_{(k-N_L)} + N_L \delta f_{\xi_L} \beta_{(N-k-N_R)}) + N_2 \delta f_{\xi} \right. \\ \left. \left. \times \left( \sum_{k < \xi} \beta_k + \sum_{k > \xi} \beta_{(N-k)} \right) \right] \right\}, \quad (45)$$

where

$$\alpha_{(N-N_2)} = \frac{1}{2^{N-N_2}} \sum_{\eta\text{-words}} \frac{2(N-N_2)}{N} \delta f_{\eta}^2, \quad (46)$$

$$\beta_{(k-N_L)} = \frac{1}{2^{k-N_L}} \left( \frac{2(k-N_L)}{N} \right)^2 \sum_{\eta_L\text{-words}} \delta f_{\eta_L}^2, \quad (47)$$

$$\beta_k = \frac{1}{2^k} \left( \frac{2k}{N} \right)^2 \sum_{L\text{-words}} \delta f_L^2. \quad (48)$$

In Eq. (48) the sum is over words associated with bits to the left of the crossover point given that all the schema  $\xi$  lies to the right of the crossover point. The expression for  $\beta_{(N-k)}$  is analogous, but with the sum over words being associated with bits to the right of the crossover point, given that the schema lies to the left. Equation (47) is associated with a sum over words for the bits to the left of the crossover point but excluding bits that are in the schema. Likewise the expression  $\beta_{(N-k-N_R)}$  contains a sum over words associated with bits that are out of the schema to the right of the crossover point. Finally,  $\alpha_s = 1 + (2N_2/N)\alpha + [2(N-N_2)/N]\alpha_{(N-N_2)}$ .

If one considers  $\xi_L$  and  $\xi_R$  as schemata on exactly the same footing as  $\xi$ , then the expressions for  $P'(\xi_L, t+1)$  and  $P'(\xi_R, t+1)$  are completely analogous to those above, except that one is now considering the bits of  $\xi_L$  and  $\xi_R$  that lie to the left and the right of the crossover point. Combining these expressions with Eqs. (45)–(48), after some lengthy but straightforward calculations one finds

$$\Delta_l(t+2) = \left( \frac{1-2\alpha_s}{1+2\alpha_s} \right) \langle\langle \alpha \rangle\rangle_l + \frac{p_c}{(1+2\alpha_s)(N-1)} \\ \times \left\langle \left\langle \frac{2N}{N_2} (\beta_R \beta_k + \beta_L \beta_{(N-k)}) \right\rangle \right\rangle_l \\ + \frac{2p_c}{(1+2\alpha_s)(N-1)} \\ \times \left\langle \alpha \sum_{k < \xi} \beta_k + \alpha \sum_{k > \xi} \beta_{(N-k)} \right\rangle_l. \quad (49)$$

The first term on the right-hand side of Eq. (49) is the result of the effect of selection at  $t+1$  on the population that was the result of selection at  $t$ . It is crossover independent. The last two terms are associated with the effects of selection on the population at time  $t+1$ , which has incorporated non-trivial contributions from crossover at time  $t$ . More precisely, the picture is the following:  $k=0$  selection on a random population induces anticorrelations in  $P'(\xi_L, \xi_R, t)$  when both  $\delta f_{\xi_L}$  and  $\delta f_{\xi_R}$  are positive due to the quadratic term  $\sim \delta f_{\xi_R} \delta f_{\xi_L}$ . Crossover reduces these anticorrelations, thereby enhancing the *whole* schema  $\xi = \xi_L + \xi_R$  relative to its parts. Selection at  $t+1$  reinforces this effect of crossover to enhance  $\xi$ , leading to the net positive contribution to  $\Delta_l(t=2)$ .

As above, we will consider the binary landscape and a landscape where the fitness of a bit is selected uniformly at random from the interval  $[0,1]$ . Similar calculations to the ones given above for  $\langle\langle \alpha \rangle\rangle_l$  lead to the final expression for  $\Delta_l(t+2)$ ,

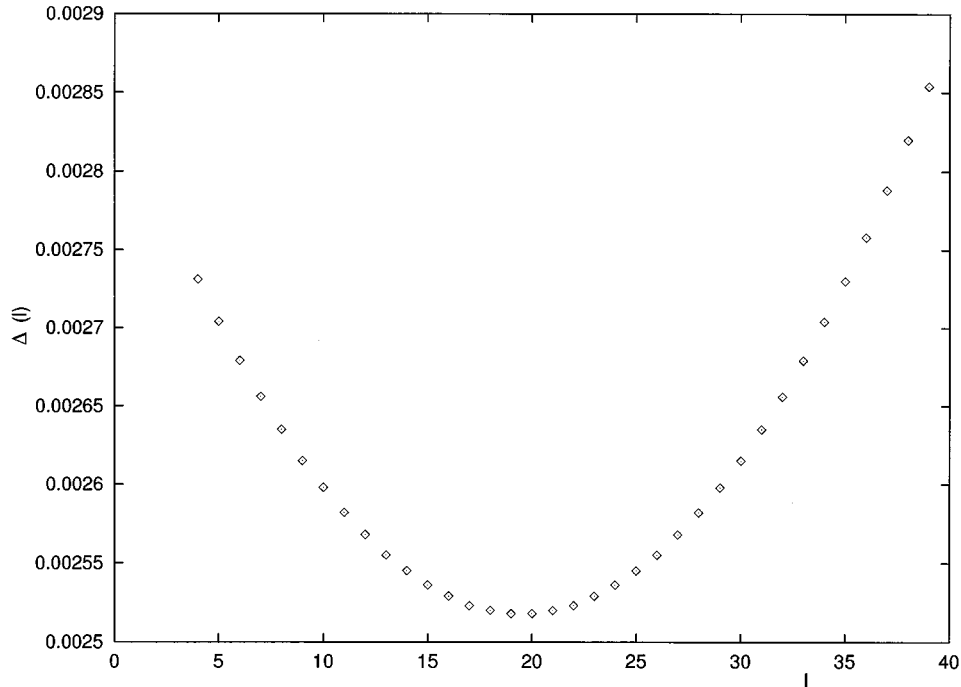


FIG. 1. Multiplicative renormalization of the effective fitness due to crossover,  $(1+s_l)$ , is represented as a function of the schema length  $l$ . The crossover term gives a positive contribution to fitness growth for all values of  $l$ , which is greater for schema sizes that are either much smaller or much larger than half the chromosome size.

$$\Delta_l(t+2) = \left( \frac{3N-c}{3N+c} \right) \frac{c}{6N} + p_c \left( \frac{2N^2 - Nl + l^2 + N + l - 8 - (8/2^l)}{144(3N+c)N^2(N-1)} \right). \quad (50)$$

From this expression one can readily see that the effects of crossover are always positive, i.e., the effects of schema reconstruction outweigh those of schema destruction. A graph of  $\Delta_l(t+2)$  versus  $l$  can be seen in Fig. 1.

We now turn our attention briefly to the limiting case of an almost organized population. In this limit, one can consider that the strings differ from the population consensus at most at one site; we will refer to the differing site as a “defect.” There are  $N$  possible defects, each with an effective negative fitness differential over the consensus string. The evolution equation implies in this case that the effect of crossover is strictly neutral: there is no net creation or destruction of defects by pure crossover without selection. So in this limit  $\Delta_l$  is once again strictly independent of  $l$ . The possibility of multiple defects in a single string raises the possibility of correlations in the distribution of defects along the string, which would induce mirroring correlations in the schemata, so  $\Delta_l$  may acquire a nontrivial  $l$  dependence as a second order effect in the mean density of defects, which is the perturbative expansion parameter near the ordered limit.

Taking once again the  $k=0$  landscape, the fitness penalty per bit for two defects is given by  $2\delta f_{ij} = \delta f_i + \delta f_j$ , so that

$$P'(ij,t) - P'(i,t)P'(j,t) \sim -\frac{\delta f_i \delta f_j}{\bar{f}^2 N^2}.$$

Since  $\delta f_i \delta f_j > 0$ ,  $P'(ij,t) < P'(i,t)P'(j,t)$ : selection induces an anticorrelation between the defects. Crossover enhances  $P(ij,t+1)$  to bring it closer to  $P(i,t+1)P(j,t+1)$ . Since the schema  $(ij)$  is more strongly damped than  $i$  or  $j$  separately, the selection at the next time step will destroy more defects than without crossover. So here again, as near the random limit, crossover has a beneficial effect due to the enhancement of the whole schema relative to its parts. Near the random limit this was beneficial because the whole schema was picked up by selection; here it is beneficial because the whole schema is more strongly damped by negative selection, so defects die out more rapidly.

## VI. EFFECTIVE DEGREES OF FREEDOM IN THE $N_2$ LANDSCAPE

The  $k=0$  landscape discussed in the preceding section has the virtue of being “neutral” from the point of view of the block hypothesis; however, it is not a realistic example of landscapes usually encountered in complex optimization problems; we will therefore turn our attention to the case  $k=2$  (for a description of the Kaufmann model, see [8]). There are two mechanisms by which connected landscapes can induce correlations. On the other hand, schemata that contain landscape-related bits have a sharper selective coefficient because there are fewer unspecified bits involved in their fitness contribution. On the other hand, the balance between the schema destruction and reconstruction terms from crossover is broken to first order.

Let us first restrict our attention to schemata of two definite bits ( $N_2=2$ ). There are three possible situations for a 2 schema. Either the two bits are not connected by the fitness landscape, one bit is the connected partner of the other, or the bits are connected both ways. This last situation is im-

probable for  $N \gg 1$ , so we will focus on the first two cases. If one has two unrelated bits in an otherwise random initial population, the effective fitness of each bit in this schema is equal to an average of four of the eight random numbers in the fitness table at that site, because one of the three bits is fixed and the other two are picked at random. If, on the other hand, one of the bits is connected to the other, then its fitness contribution is given by averaging over the two possible values that the other connected partner can take. This is an average of two out of the eight random numbers in the fitness table. The key point is that the average of two random numbers typically differs from 1/2 more than the average of four. Therefore, schemata that include landscape-related bits will have a stronger selective coefficient, in absolute value. This leads to a bias for the condensation of schemata that recognize the structure of the fitness landscape.

In order to make this argument more precise, we need to compute the expectation value of the best of  $m_1$  averages of  $m_2$  random numbers, where each random number is uniformly distributed in the unit interval. The probability distribution of the best of  $m_1$  averages of  $m_2$  random numbers is equal to the derivative of the probability that  $z$  is larger than all  $m_1$  averages. If we call  $P(x_1, \dots, x_{m_1})$  the distribution of the averages, the probability that  $z$  is greater than all of the averages is

$$p(z > \sup(m_i)) = \int_0^1 dx_1 \cdots \int_0^1 dx_{m_1} P(x_1, \dots, x_{m_1}) \times \prod_{i=1}^{m_1} \theta(z - x_i). \quad (51)$$

Since the  $m_1$  averages are statistically independent in this case, this expression reduces to

$$p(z) = \left( \int_0^1 dx P(x) \theta(z - x) \right)^{m_1}. \quad (52)$$

The expectation value of the best of the  $m_1$  averages is

$$\langle z_{\max} \rangle = \int z p'(z) dz. \quad (53)$$

For our purposes it is sufficient to consider the cases where  $m_1, m_2 \in \{2, 4\}$ . For  $m_2 = 2$  the distribution of the average of two uniformly distributed random numbers is given by  $P(x) = 4x$  for  $x < 1/2$  and the symmetry condition  $P(1/2 + x) = P(1/2 - x)$ . The expectation values for the best of  $m_1$ , such averages are, for  $m_1 = 2, m_2 = 2, \langle z_{\max} \rangle = 0.6167$ ; while for  $m_1 = 4, m_2 = 2, \langle z_{\max} \rangle = 0.7300$ . For  $m_2 = 4$  (averages of four uniformly distributed variables), one has

$$P(x) = \frac{128}{3} x^3 \quad \text{for } x < \frac{1}{4}, \quad (54)$$

$$P(x) = \frac{128}{3} x^3 - \frac{2}{3} \quad \text{for } \frac{1}{4} < x < \frac{1}{2}, \quad (55)$$

and the symmetry condition given above. One finds the following result: for  $m_1 = 2, m_2 = 4, \langle z_{\max} \rangle = 0.5673$ . Finally, the expectation value of the best of  $m_1$  uniformly distributed random variables is  $\langle z_{\max} \rangle = m_1 / (m_1 + 1)$ . Here we will need only the best of eight, which is equal to  $8/9 = 0.8889$ .

If the two bits are not related by a landscape connection the effective fitness of any one of these bits in a random population is given by the average of four random numbers from the fitness table, where the averaging is over the values of the two connected partners that determine the fitness contribution of this bit. Thus, the best schema can be expected to have a selective advantage  $s_1 = (\bar{f}_\xi / \bar{f}) - 1 = (4/N)(0.567 - 0.5)$ . Now, if there is a landscape connection between the two bits of the schema, the contribution of one of these bits to the string fitness is given by an average of two random numbers, since we only need to average over the other connected partner that is not in the schema. The best schema in this case will have a selective advantage  $s_2 = (4/N)[(0.567 + 0.73)/2] - 0.5$ .

In the case  $N = 40$  analyzed in the preceding section, the ratio of the growth rates of a 2 schema that recognizes a landscape connection to that of one that does not is  $r = [(1 + s_2)/(1 + s_1)] = 1.0081$ . This result should be compared to the effect of crossover, which we computed in the  $k = 0$  landscape at the second time step:  $1 + s_l$  was found to fluctuate between 1.0025 at  $l = N/2$  and 1.0029 at  $l = 2, l = N - 1$ . Clearly, the conclusion is that landscape correlations should be taken into account in a proper analysis of the condensation of ‘‘schemata.’’

In our discussion we neglected the possible existence of frustration and assumed that the fitness contribution of the two bits of the schema could be optimized independently without affecting the mean fitness contribution of the other bits in the string. A more careful analysis including frustration would be much more complicated; however, one expects that at least for small  $N_2$ , frustration should be marginal and that our conclusions should hold qualitatively. Of course, there are fewer 2 schemata that recognize a landscape connection than not, so the overall contribution of such schemata to the condensation of EDOF is diluted by a phase space factor  $2/N$ , relative to 2 schemata of landscape-related bits. Thus, one expects that the first stage of divergence from a random population will be dominated by schemata that do not ‘‘understand’’ the fitness landscape. The landscape-related schemata, which grow at a faster rate, will eventually overcome the contrary phase space factor and become more important in the condensation process.

Returning to the fundamental equation (16) for the growth of in-schema fitness, we can evaluate the effect of crossover in a  $k = 2$  landscape by calculating  $\Delta_t$  in the first step away from a random population:

$$\Delta_{N_2=2}^{(n)} = \left\langle \frac{4}{N} \delta f_\xi^2 \right\rangle^{(n)} - \frac{4p_c}{N(N-1)} \langle l-1 \rangle^{(n)} \times \left\langle \delta f_\xi \left( \delta f_\xi - \frac{N_L}{N_2} \delta f_{\xi_L} - \frac{N_R}{N_2} \delta f_{\xi_R} \right) \right\rangle^{(n)},$$

where we have used the identity, valid in a random population,  $\sum_{\text{words}} \delta f_{\xi_L} \delta f_{\xi_R} \delta f_\xi = 0$ , and the average  $\langle \rangle^{(n)}$  runs over the set of all schemata with  $N_2 = 2$  definite bits with  $n = 0, 1, 2$  landscape connections between schema bits. We are also assuming that there is not explicit  $l$  dependence in the fitness landscape itself.

The evaluation of  $\delta f_\xi$  depends on the number of in-schema connections. One must evaluate the contribution of

each of the two bits in the schema. If there are no in-schema connections, then the averaging over unspecified bits leads to a contribution to  $\delta f_\xi$  equal to the average of four of the eight random numbers in the fitness table. If one of the bits is connected to the other, then in evaluating its fitness contribution one has only one unspecified bit and the contribution to  $\delta f_\xi$  turns out to be the average of two of the eight random numbers. The values of  $\delta f_{\xi_L}$  and  $\delta f_{\xi_R}$  are always given by the average of four random numbers. Thus, if there are no in-schema connections,  $\delta f_\xi = (N_L/N_2)\delta f_{\xi_L} + (N_R/N_2)\delta f_{\xi_R}$  and the contribution of the crossover term vanishes as in the  $k=0$  case. If we denote by  $\sigma^2$  the variance of the random number distribution used to generate the tables of eight possible fitness contributions for each bit, the averaging over schemata with  $n=0$  landscape connections gives

$$\Delta_{N_2=2}^{(0)} = \left\langle \frac{4}{N} \delta f_\xi^2 \right\rangle^{(0)} = \frac{\sigma^2}{2N}.$$

On the other hand, if there is one in-schema connection, then  $\langle \delta f_\xi^2 \rangle^{(0)}$  is the variance of the average of two random numbers plus the variance of an average of four, while one of the terms  $\langle \delta f_\xi \delta f_{\xi_L} \rangle^{(0)}$  or  $\langle \delta f_\xi \delta f_{\xi_R} \rangle^{(0)}$  is equal to half the variance of two random numbers, the other being the variance of an average of four. Using  $\langle l-1 \rangle = (N+1)/3$ , one finds

$$\begin{aligned} \Delta_{N_2=2}^{(1)} &= \left( \frac{3}{4} - \frac{p_c(N+1)}{12(N-1)} \right) \frac{\sigma^2}{N} \\ &= \Delta_{N_2=2}^{(0)} + \Delta_{N_2=2}^{(0)} \left( \frac{1}{2} - \frac{p_c(N+1)}{6(N-1)} \right). \end{aligned}$$

Similarly, for  $n=2$  in-schema connections,

$$\Delta_{N_2=2}^{(2)} - \Delta_{N_2=2}^{(0)} = \Delta_{N_2=2}^{(0)} \left( 1 - \frac{p_c(N+1)}{3(N-1)} \right).$$

In these expressions, the  $p_c$ -independent correction is the result of the selective advantage of schemata that recognize landscape connections, which we discussed previously. These numbers appear somewhat magnified relative to  $r$ . This is only because here we are examining the in-schema fitness per bit, whereas  $r$  was associated with the growth rate of the entire string. The crossover contribution reduces this correlating effect of the landscape but only by a factor of  $2/3$  in the limit  $p_c \rightarrow 1$ ,  $N \rightarrow \infty$ . In conclusion, schemata that reflect the landscape connections contribute more (per bit) to the growth of fitness than schemata involving unrelated bits.

A similar conclusion can be expected to hold if one considers larger schemata with  $N_2 > 2$ . Extending the argument to general schemata, one is led to consider *fitness trees*: the fitness tree of a bit is the set that consists of the bit itself, its connected partners, the connected partners of these connected partners, and so on. We can define an order  $n$  truncated fitness tree by truncating this procedure after  $n$  steps. The dominant value of  $n$  depends on the degree of order in

the system, which is a function of the mutation rate. For a high mutation rate one expects the gene pool to be highly disordered and EDOF are mostly single bits ( $n=0$ ) or truncated fitness trees with small values of  $n$ . As the rate decreases, larger trees can condense and the dominant value of  $n$  increases. This leads us to propose the following conjecture on the nature of the EDOF, which we shall call the ‘‘fitness tree hypothesis.’’

*Hypothesis.* The EDOF of genetic algorithms with  $Nk$  fitness landscapes are the truncated order  $n$  fitness trees. The effective value of  $n$  increases as the condensation process allows for an increasingly structured gene pool.

In order to test this hypothesis we designed a numerical simulation with a population of 1000 individuals in an  $Nk$  landscape, with  $N=40$  and  $k=2$ . The crossover probability was taken to be equal to 1. The spatial correlation function measures the correlation of bits at distance  $d$  along the string and tests the block hypothesis directly. A second correlation function measures this correlation as a function of the connective distance between bits, defined as the smallest number of landscape connections from one bit to the other. The results are shown in [Figs. 2(a)–2(c)]. At generation 15 [Fig. 2(a)] the spatial correlation function reflects the preference for small schemata, as suggested by the block hypothesis. After 100 generations [Fig. 2(b)] the spatial correlation function has become weak and roughly independent of the distance; on the other hand, the correlation of landscape-related bits becomes significant at connective distance 1. By generation number 150 one finds statistically significant correlations up to connective distance 4, which are progressively reinforced. In Fig. 2(c) we show the correlation functions at generation 200. Since the mutation rate is equal to zero in these simulations, population diversity eventually decreases and becomes insufficient to derive statistically relevant correlation coefficients. At generation 350 the strings are totally condensed up to connective distance 2 (the first two correlation coefficients are equal to 1); the gene pool is completely organized at the 500th generation.

Throughout this article, with the exception of the numerical experiments, finite-size effects were neglected. If one considers their contribution, the failure of the block hypothesis only becomes more apparent. Here we will mention only briefly two arguments to this effect. In a finite population the difficulty of *finding* a good schema must be considered, since not all schemata are present in the initial population. Since the number of schemata with fixed  $N_2$  grows with  $l$  as  $l^{-2}C_{N_2-2}$ , one expects it to be easier to discover good large schemata than small ones. Another important finite-size effect is the effective nonlinearity of selection emphasized in the neutral theory of molecular evolution [9]: Schemata with only weak selective coefficients are not necessarily selected, as the neutral drift due to fluctuations in the selection of parents dominates over selection unless  $|s_{\text{eff}}| > 1/P$ ,  $P$  being the effective breeding population. This leads to an effective nonlinearity of selection due to the existence of a threshold in favor of schemata with a selective coefficient above this value. Since the selective coefficient of a schema grows in proportion to  $N_2$ , this effect favors schemata with large  $N_2$ . Combining this result with the previous comment on the probability of finding good schemata being proportional to  $l^{-2}C_{N_2-2}$ , we find that schemata with small values of  $l$  are strongly disfavored by the finite-size effects.

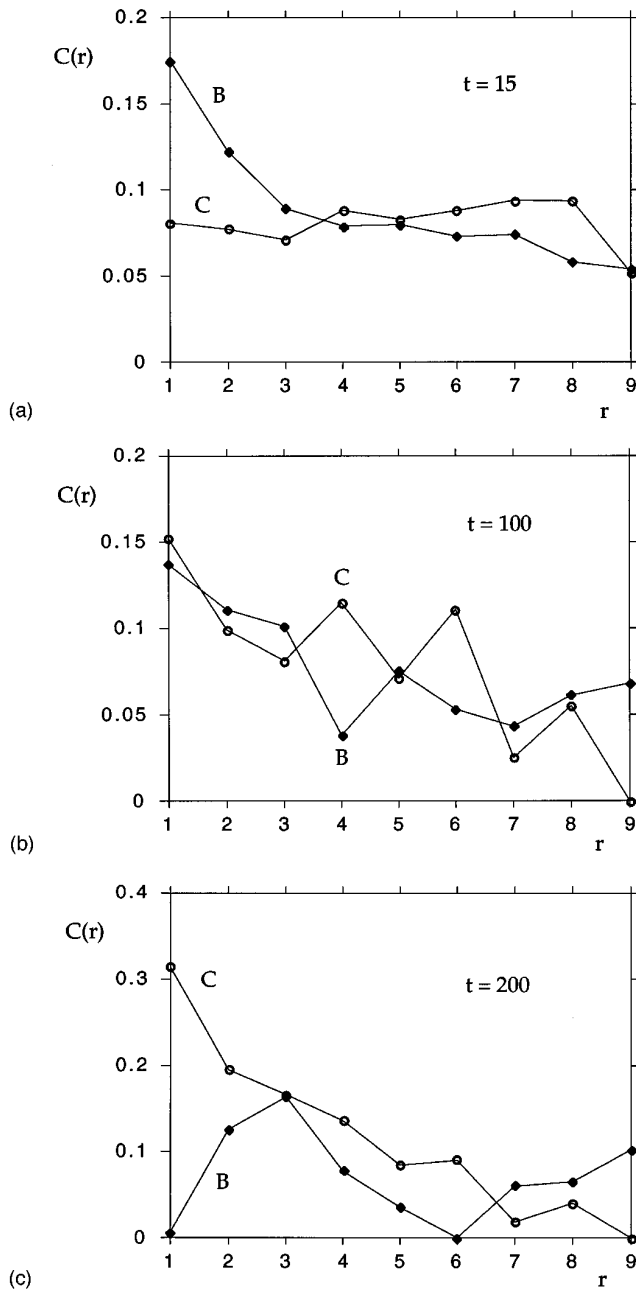


FIG. 2. Average absolute correlations between bits in the chromosome are given in terms of ( $B$ ) the linear distance that separates the bits on the chromosome, and ( $C$ ) the connective distance defined as the smallest number of landscape connections to go from one bit to the other. Very early on one notes a slight preference for correlations between bits that are near each other on the chromosome, i.e., with  $l \ll N$  (a). By  $t=100$  the correlations between landscape-related bits become important (b), and they come to dominate at  $t=200$  (c). At this point the population is highly organized and correlations on the basis of linear chromosome distance are no longer significant.

## VII. CONCLUSIONS

The bulk of this paper has been devoted to deriving equations that describe the evolution of string populations in GAs, and in particular how EDOF may emerge during this evolution. We started with an equation that governed the evolution of the strings themselves under the joint action of

selection, mutation, and crossover. We found that this equation could be elegantly expressed in terms of the evolution of a string  $C_i$  and its constituent parts. This naturally introduced the notion of a coarse graining relative to a description in terms of the strings themselves, the coarse graining being associated with sums over strings that contained a part of  $C_i$ . Subsequently we derived an analogous equation for the evolution of schemata, this time in terms of a schema and its constituent parts. Schema evolution is coarse grained relative to string evolution because of the summing over the  $N - N_2$  nonschema bits. The evolution of a schema of  $O(N_2)$  is described in terms of its constituent parts, which are schemata of order less than  $N_2$ . Thus the action of crossover invokes a natural hierarchy of coarse grainings. Such a hierarchy is reminiscent of a renormalization group transformation, where there is a coarse graining over a subset of degrees of freedom, such as in the one-dimensional Ising model, where one may sum over every other spin in the partition function, for instance. In the GA case this coarse graining stops naturally when one arrives at the evolution of 1 schemata.

In one sense it is remarkable that one may solve analytically a GA, albeit for a simple fitness landscape and over a short time interval; however, what is lacking is a reasonable approximation scheme with which one may attack the evolution equations. Just as solving an exact renormalization group equation is almost impossible, so with GAs finding exact solutions is probably hopeless. However, implementing renormalization group transformations approximately has been remarkably successful in explaining many physical phenomena. We hope that finding analogous techniques in the study of GAs might lead to similar success.

Starting from the evolution equation for schemata, a further coarse graining was performed to arrive at an expression for the average contribution of all schemata of size  $l$  to the improvement of fitness. Applying this equation to the particular case of a  $k=0$  landscape, where each bit contributes independently to fitness, we showed that the net effect of crossover on fitness growth is slightly positive for all  $l$ : the effect of schema reconstruction always exceeds that of destruction. Schemata that are either much smaller or much larger than half the string size are most enhanced. A different situation arises if one considers a  $k>0$  landscape. In this case the sum of the effective selective advantages of the parts of a schema is not necessarily equal to the effective selective advantage of the entire schema. Only when the parts of a selected schema are less selected than the whole (the deceptive case), does crossover lead to a net destructive force as schemata are broken down into pieces that are then lost due to their low selectivity. The schemata that are selected over a long time scale are those that break down into useful parts, independently of their size.

Finite-size effects break the apparent symmetry of the geometrical effect of crossover about  $l=N/2$ : The existence of a selection threshold favors highly fit schemata with a large number of specified bits  $N_2$ , and these can be found with a reasonable probability only if their length  $l$  is large. Combining this argument with the  $l$  dependence of in-schema fitness growth  $\Delta_l$ , one concludes that the effective degrees of freedom will be schemata with large  $N_2$  and  $l > N/2$ .

This conclusion has important and surprising conse-

quences for the designer of GAs. It is often thought that GA designers should strive to find a coding such that bits that “cooperate” are placed near each other on the chromosome, so as to resist the destructive effect of crossover. This is generally speaking a very difficult task, since the structure of the optimization problem usually does not match the linear topology of the strings. Our results show that this task is pointless: if anything, one should try to place cooperating bits as far from each other as possible. Of course this is the most probable outcome if no attention is placed on the linear disposition of the bits, so this is not a problem one should worry about.

We should stress that the above comment by no means implies that the choice of coding is irrelevant. The choice of a genetic interpreter is crucial to generate a high density of states near desired fitness extrema and perhaps also to guide the emergence of an algorithmic language [16] that facilitates the search for new highly fit schemata. These issues, however, lie beyond the scope of the present paper.

With the results of this paper in mind, it is interesting to recall the analogy between GAs and spin glass dynamics discussed in the Introduction. In both cases one is describing

a condensation process in a rugged landscape, guided by the emergence of overlaps with certain structures or “patterns.” One of the chief reasons why in GAs the overlaps with schemata are considered rather than with entire strings ( $N_2 = N$  schemata) is that genetic populations are generally too disordered for such a rigid structure as a completely specified string to be of much relevance. Of course the same can be said of spin glasses far from equilibrium. This suggests that the notion of “schema” may find some usage in the study of the condensation of spin glasses from an initial disordered phase. One can carry the analogy between GAs and spin glasses one step further and suggest that, in the case of sparsely connected neural networks, the truncated connective trees may form a privileged class of schemata for the purpose of developing an effective theory of neural dynamics.

#### ACKNOWLEDGMENTS

We would like to thank Peter Stadler for discussions during his stay at the Institute of Nuclear Sciences. This work was partially supported through Conacyt Grant No. 211085-5-0118PE.

- 
- [1] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
  - [2] D. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).
  - [3] F. Zertuche, R. López-Peña, and H. Waelbroeck, J. Phys. A **27**, 5879 (1994).
  - [4] B. Lewin, *Genes V* (Oxford University Press, Oxford, 1995).
  - [5] J. H. Holland, *Adaptation in Natural and Artificial Systems* (MIT Press, Cambridge, MA, 1975).
  - [6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley, Reading, MA, 1989).
  - [7] L. Peliti, e-print cond-mat/9505003.
  - [8] S. A. Kauffman, *The Origins of Order* (Oxford University Press, Oxford, 1993).
  - [9] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, 1983).
  - [10] M. Mitchell, *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, MA, 1996).
  - [11] M. Eigen, Naturwissenschaften **58**, 465 (1971); M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
  - [12] I. Leuthausser, J. Chem. Phys. **84**, 1884 (1986); J. Stat. Phys. **48**, 343 (1987).
  - [13] A. Prugel-Bennett and J. L. Shapiro, Phys. Rev. Lett. **72**, 1305 (1994).
  - [14] P. G. Higgs, Gen. Res. Cambridge **63**, 63 (1993).
  - [15] C. L. Bridges and D. E. Goldberg, in *Genetic Algorithms and their Applications*, edited by John J. Grefenstette (Lawrence Erlbaum, Hillsdale, NJ, 1987).
  - [16] O. Angeles, H. Waelbroeck, and C. R. Stephens, BioSystems (to be published).