# Statistical Dynamics of the Royal Road Genetic Algorithm

**Erik van Nimwegen,**[†]
**James P. Crutchfield,**[†‡] **and**
**Melanie Mitchell**[†]
[†]**Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501**
[‡]**Physics Department, University of California, Berkeley, CA 94720-7300**

To appear in **Theoretical Computer Science** (1998).

**Abstract.**

Metastability is a common phenomenon. Many evolutionary processes, both natural and artificial, alternate between periods of stasis and brief periods of rapid change in their behavior. In this paper an analytical model for the dynamics of a mutation-only genetic algorithm (GA) is introduced that identifies a new and general mechanism causing metastability in evolutionary dynamics. The GA's population dynamics is described in terms of flows in the space of fitness distributions. The trajectories through fitness distribution space are derived in closed form in the limit of infinite populations. We then show how finite populations induce metastability, even in regions where fitness does not exhibit a local optimum. In particular, the model predicts the occurrence of "fitness epochs"—periods of stasis in population fitness distributions—at finite population size and identifies the locations of these fitness epochs with the flow's hyperbolic fixed points. This enables exact predictions of the metastable fitness distributions during the fitness epochs, as well as giving insight into the nature of the periods of stasis and the innovations between them. All these results are obtained as closed-form expressions in terms of the GA's parameters.

An analysis of the Jacobian matrices in the neighborhood of an epoch's metastable fitness distribution allows for the calculation of its stable and unstable manifold dimensions and so reveals the state space's topological structure. More general quantitative features of the dynamics—fitness fluctuation amplitudes, epoch stability, and speed of the innovations—are also determined from the Jacobian eigenvalues. The analysis shows how quantitative predictions for a range of dynamical behaviors, that are specific to the finite population dynamics, can be derived from the solution of the infinite population dynamics. The theoretical predictions are shown to agree very well with statistics from GA simulations. We also discuss the connections of our results with those from population genetics and molecular evolution theory.

## Contents

# 1. Epochal Evolution

Metastability is a commonly observed phenomenon in many population-based dynamical systems. In such systems—including evolutionary search algorithms, models of biological evolution, and ecological and sociological systems—the state of a population is often described as the distribution of certain features of interest over the population. A commonly observed qualitative behavior is that the distribution of these features alternates between periods of stasis and sudden change. For extended periods of time the system seems to stabilize on some feature distribution, which is then disrupted by a brief burst of change. We call this type of behavior "epochal evolution", where the term "epoch" denotes an extended period of apparent stability. We use the term "innovation" to refer to the sudden change between epochs. Such behavior, often referred to as "punctuated equilibria", has been reported in many applications of evolutionary search as well as in models of biological and molecular evolution (e.g.,[3, 15, 23, 25, 31, 33, 41]). Epochal behavior has also been observed in natural phenomena such as the outbreak of epidemics [4], the progression of diseases such as cancer [1] and AIDS [5] in individuals, rapid large-scale ecological changes, and the sudden rise and fall of cultures. In natural systems and in many models, epochal behavior is undoubtedly the result of a complicated and poorly understood web of mechanisms.

In this paper we identify the mechanism that underlies the occurrence of epochal behavior in a simplified mutation-only genetic algorithm. In the mechanism's most general form, metastability is induced in an area of state space where the local "flow" of the dynamics is small compared to a scale set by the population's finite size. The dynamics becomes too weak to drive changes in the finite population. More specifically, we will see that the metastability can be associated with an "entropic barrier" that the finite population must pass in moving through the "slow" part of state space. Metastability due to such entropic barriers is contrasted here with the more traditional explanation of metastability as being induced by "fitness barriers". In the latter the population stabilizes around a local fitness optimum in sequence space and must pass through a "valley" of lower fitness to find a higher-fitness optimum. We believe that the generality and simplicity of the mechanism for metastability introduced here makes it likely to play a role in the occurrence of epochal dynamics in the more general and complicated cases alluded to above.

## 1.1 Search and Evolution

Genetic algorithms (GAs) are a class of stochastic search techniques, loosely based on ideas from biological evolution, that have been used successfully for a great variety of different problems (e.g., [2, 10, 16, 20]). However, the mechanisms that control the dynamics of a GA on a given problem are not well understood. GAs are nonlinear population-based dynamical systems. The complicated dynamics exhibited by such systems has been appreciated in the field of mathematical population genetics for decades. On the one hand, these complications make an empirical approach to the question of when and how to use evolutionary search problematic. On the other hand, the lack of a unified theory capable of quantitative predictions in specific situations has rendered the literature largely anecdotal and of limited generality. The work presented in this paper tries to unify and extend theoretical work that has been done in the areas of GA theory, the theory of molecular evolution, and mathematical population genetics. The goal is to obtain a more general and unified understanding of the mechanisms that control the dynamics of GAs and other population-based dynamical systems.

Vose and his colleagues have previously studied GA dynamics by describing the state of a genetic algorithm at a certain time as a vector in a high-dimensional Euclidean space. Each dimension of this space either represents a certain string [45] or the state of the population as a whole [35]. The dynamics is then described by a nonlinear matrix operator acting on this vector to produce the state at the next time step. Although this "microscopic" approach is formally very clear and precise, in practice the huge sizes of these matrices make it impossible to obtain specific quantitative results. In this paper, a genetic matrix operator is constructed that is similar in spirit to the genetic operators discussed in [35] and [45] but that acts on vectors representing fitness distributions only, *averaging out* all other structure of the microscopic state of the population. The operator, therefore, has a much lower dimensionality. This will make quantitative analyses of this operator possible, allowing specific quantitative predictions to be made about the GA's observed behavior.

A more macroscopic theoretical approach was developed by Prügel-Bennett, Shapiro, and Rattray. Their approach uses ideas from statistical mechanics to analyze the dynamics of GAs [38, 39, 40]. Their formalism also focuses on the evolution of fitness distributions, but generally only the *average* evolution of the first few cumulants of the fitness distribution is calculated. This averaging of the dynamics over a large number of runs makes it impossible to describe the epochal structure of the dynamics in which we are interested. The statistical mechanics approach does, however, provide some insights into the roles that the different genetic operators play in the dynamics and shares with our approach the idea of averaging out most microscopic degrees of freedom to obtain a macroscopic description of the dynamics.

Another theoretical framework of relevance is Eigen's theory of molecular evolution [11, 14]. In the simplest form of this theory, one considers a large population of self-replicating molecules in a reaction vessel. Since the total concentration of molecules is kept constant, there is an effective selection for molecules that replicate fast and efficiently. It is assumed that the molecules make errors during replication, thus introducing mutations. The differential equations that describe the change in concentrations of the different molecular types are analogous to the genetic operator equations that we will develop in this paper. Although, in contrast to our mesoscopic approach, they are defined only on the microscopic states of concentrations of individual genotypes. We will explain how some theoretical concepts from molecular evolution theory, such as the quasispecies and the error threshold, generalize to analogous concepts in our description of the GA dynamics.

Finally, the theory of mathematical population genetics has a long history of analyzing the behavior of evolving populations. Many important results were obtained in the 1930s by the trio of Fisher, Wright, and Haldane. In the 1960s Kimura developed a new way of analyzing evolving populations using diffusion equations [27] that were originally developed in the context of statistical physics [19, 30, 37]. We will make use of this type of analysis several times and will show how methods developed in the context of mathematical population genetics bear on the dynamics of GAs as well.

## 1.2 Organization of the Analysis

Our analysis of epochal evolution in a mutation-only genetic algorithm first appeared in [44]. The present work goes into considerably more depth. Section 2 introduces the simplified GA used. In section 3 we present an overview of the wide range of different dynamical behaviors that our simple GA exhibits. We discuss the qualitative features of these different dynamical behaviors and pose ourselves a set of questions that we would like to answer using our theoretical model.

The bulk of the remainder is devoted to the development and analysis of this theoretical model. We have termed our type of analysis "statistical dynamics", since it combines a dynamical systems approach, on the one hand, with a statistical physics and stochastic process approach, on the other. The infinite population behavior is treated as the dynamics of a deterministic nonlinear dynamical system. In constructing this dynamical system we have to choose suitable "mesoscopic" state variables (in this case, fitness distributions) that capture the complicated microscopic state of the system in terms of a much lower dimension state space. Moreover, we require that the description of the system and its behavior in terms of these variables should be closed in the limit of infinite populations. That is, for infinite populations we assume that the dynamics of fitness distributions is fully specified by the fitness distributions themselves and does not depend on the exact underlying ("microscopic") distribution of genotypes. This condensation of the microscopic states using a few "order parameters", that describe the dynamics in the limit of infinite system size, is a well-known procedure from statistical physics. With this setting established, we augment the solution of the nonlinear dynamical system with a statistical treatment of the finite population behavior. In doing so, we make use of simple stochastic differential equations, such as the Fokker-Planck equation. These three features—describing the system in terms of a small set of statistical order parameters, deriving and solving the deterministic nonlinear dynamical systems equations in the infinite population limit, and then augmenting this solution with simple stochastic differential equations to capture the finite-population dynamics—is the essence of our statistical dynamics approach.

In section 4 we introduce the state space of fitness distributions in terms of which the GA's dynamics is defined, as well as motivate the use of this particular state space. Section 5 develops an exact solution of the dynamics in this state space in the limit of infinite populations. Specifically, we solve in closed form for the trajectory through fitness distribution space that is followed by an infinite population and analytically characterize the asymptotic behavior of the dynamics. Section 6 is concerned with the finite-population dynamics and presents the main results. This section builds on the results from section 5 to quantitatively analyze a wide range of dynamical features that derive from the finite-population dynamics. In particular, we identify the mechanism that leads to the fitness epochs, solve for their locations in fitness distribution space, and show that the fitness levels of the epochs are left unaltered under the introduction of genetic crossover. We then calculate the stable and unstable manifold dimensions of the metastable epochs, the fitness fluctuation amplitudes during the epochs, the speed of innovations between epochs, and the stability of the epochs. All these results are obtained analytically as functions of the model parameters and are shown to agree with statistics estimated from our GA simulations. Major players in the derivation of the results of section 6 are the eigenvalues and eigenvectors of the Jacobian of the generation operator that describes the dynamics in the limit of infinite populations. Section 7 discusses the average durations of the epochs and describes how the model breaks down in predicting these average durations. Section 8 discusses the results of our paper and looks ahead to future work.

## 2. A Simple Genetic Algorithm on the Royal Road Fitness Function

### 2.1 The Fitness Function

The Royal Road fitness functions assign a fitness $f(s)$ to a string $s$ as the sum of fitness contributions $f_i$ from $N$ different nonoverlapping bit sets ("blocks") $s_i$ of $s$. We will consider bit strings $s$ of length $L = NK$, each of which can be thought to consist of $N$ blocks of length $K$.

$$\underbrace{\overbrace{101\cdots011}^{K}\,101100010\cdots10100111001}_{NK}$$

For each block of length $K$ there is a particular desired bit configuration (schema). In the above illustration we took the blocks to be sets of $K$ contiguous bits, but it's easy to see that the dynamics of a mutation-only GA is invariant under random permutations of the bits in the string representation. Formally, we have

$$f(s) = \sum_{i=1}^{N} f_i \delta_{s_i,x_i}, \tag{2.1}$$

where the $x_i$ are desired configurations for each block $s_i$ of $s$ and $\delta_{s_i,x_i} = 1$ if and only if $x_i = s_i$, otherwise $\delta_{s_i,x_i} = 0$. A block $s_i$ that exhibits the desired configuration $x_i$ will be called an "aligned" block and blocks that do not exhibit the desired configuration will be called "unaligned". Without loss of generality, this desired configuration can be taken to be the configuration of $K$ 1s: $x_i = 1^K$. The $f_i$ are the fitness contributions from each aligned block $i$. For simplicity we shall take all $f_i$ to be equal: $f_i = 1$. The fitness of $s$ can then be simply defined as the number of aligned blocks in $s$. Thus $0 \leq f(s) \leq N$. The number of blocks $N$ and the number of bits per block $K$ are parameters of the fitness function.

Royal Road fitness functions were initially designed to address questions about the processing and recombination of schemata in genetic algorithms. They were thought to lay out a "royal road" for genetic algorithm search [34] and so could test a GA's ability to preferentially propagate useful genetic "building blocks". The Royal Road functions defined in [34] are more general than the ones we are considering in this paper. For instance, the fitness $f(s)$ of a string $s$ does not in general need to be a simple *sum* of the fitnesses $f_i$ of the aligned blocks. Here we will not be concerned with the issues of schemata processing and the building block hypothesis. We use the simple Royal Road fitness functions defined above, because they are simple enough to be analyzed and because the GA's behavior on these fitness functions exhibits a range of qualitatively distinct epochal dynamics.

### 2.2 The Genetic Algorithm

In our study, we use the following mutation-only genetic algorithm:

1. Generate a population of $M$ strings of length $L = NK$, chosen with uniform probability from the space of all $L$-bit strings.
2. Evaluate the fitness $f(s)$ of each string $s$ in the population.

3. Create a new population of $M$ strings by choosing strings from the current population with replacement and with probability proportional to fitness. This is sometimes called fitness-proportionate selection.
4. Mutate (i.e., change) each site value in all strings with a fixed probability $q$.
5. Go to step 2.

As noted before, this algorithm does not include crossover, a genetic operator purported to aid in the evolutionary propagation of important genetic building blocks. Crossover is left out of this first work for two reasons. First, it considerably simplifies the analysis of the GA. Second, the main qualitative features of the dynamics—the occurrence of fitness epochs and innovations between them—are not changed by leaving out crossover. We will address in more detail the effects of crossover on this GA in section 6.5 and show that the fitness levels of the epochs are not changed by including crossover.

This defines our Royal Road GA. It has four control parameters: the number $N$ of blocks, the number $K$ of bits in a block, the mutation rate $q$, and the population size $M$.

## 3. Observed Behavior of the Royal Road GA

In general, the behavior of an evolving population is governed by a complicated interplay of a few major genetic and populational forces—selection, mutation, crossover (if present), and genetic drift.

Selection tends to converge the population onto the current best strings that are found in the population. In this sense selection is installing information in the population: strings in the population become much more similar to one another than a collection of random strings and the entropy of the distribution of strings is therefore decreased. To some extent the bit values that are shared by all members of the population are a reflection of their functionality. For example, under the Royal Road fitness function, once a string with a desired block is discovered, that string is preferentially reproduced and the population converges quickly to strings containing that block. Such convergence reflects the functionality conferred on strings by this block.

Mutation and crossover, for that matter, are largely forces that drive genetic mixing. Information that has been installed in a string is destroyed by mutation randomly flipping bits. At the same time mutation is a force that can provide something *new*: it can create bit strings that were never before present in the population and that might have improved fitness.

A third important force is genetic drift, which is due to the sampling fluctuations induced by the finite size of the population. Genetic drift, which is recognized as a major player in the theory of population genetics, seems to be somewhat neglected in the theory of GAs. For small populations, it turns out that information can be stored in the strings *by accident*. Suppose that there is no mutation or selection. The initial population is chosen at random, so it is likely to contain $M$ different genotypes. At each generation a new population is created by randomly sampling $M$ strings from the old population. At each time step it is likely that certain genotypes will be lost, since they are not sampled, and that other genotypes will multiply. After a number of generations on the order of the population size $M$, it is likely that there will be only one genotype left in the population, the particular genotype being purely accidental. This process is known as "random genetic drift" and it is one way in which arbitrary information is stored in the string population. Genetic drift plays a major role in the dynamics of GAs with selection and mutation: small populations tend to spontaneously converge, regardless of selection. Note that populations also converge in the presence of crossover. Thus, the only genetic operator capable of prohibiting complete convergence of the population is mutation.

We now present a set of examples of the empirical behavior of our Royal Road GA in order to demonstrate how varied it can be for different parameter settings. This should make it clear that, although we can qualitatively identify the main evolutionary forces, the actual interplay of these forces can be complicated and subtle.

Caption for figure 1: Average fitness (solid lines) and best fitness (diamonds, often appearing together as thick solid lines) in the population over time for nine runs of the Royal Road GA with different parameter settings. Our canonical example of epochal behavior can be found in run 1(d). The parameter settings for this run 1(d) are: $N = 10$ blocks of length $K = 6$, a mutation rate of $q = 0.001$, and a population size of $M = 500$. All other runs were done with parameter settings that can be obtained by consecutive single-parameter changes from the parameters of run 1(d). Note that all runs were done with a fixed string length $NK = 60$. Moving out from run 1(d) by following the large arrows from run to run, a single parameter is changed. The changed parameter and its new setting (increased or decreased according to the thin up or down arrow, resp.) are indicated next to the large arrows. For example, following the large arrows up and down from run 1(d) we arrive at runs 1(a) and 1(g), respectively. As indicated, only the population sizes of run 1(a) and 1(g) differ from the parameters in run 1(d). For run 1(a) a large population size of $M = 5000$ was used and for run 1(g), a small population size of $M = 50$ was used. From the three runs 1(a), 1(d), and 1(g) three arrows point to the three runs 1(b), 1(e), and 1(h) in the middle column. All vary in only one parameter from the settings in the corresponding runs in the left-hand column. Run 1(b) differs from run 1(a) by a decrease of the block size to $K = 3$ and an increase of the number of blocks to $N = 20$. Run 1(e) differs from run 1(d) by an increase of the mutation rate to $q = 0.0075$. And, run 1(h) differs from run 1(g) by an increase of the mutation rate to $q = 0.005$. The three runs in the middle column show how a change in a single parameter with respect to the runs in the left-hand column can make epochal dynamics disappear. The runs in the right-hand column show that a further change in a single parameter can make epochal dynamics reappear. Run 1(c) differs from run 1(b) by a decrease in population size to $M = 300$. Run 1(f) differs from run 1(e) by an increased population size of $M = 7500$. And finally, run 1(i) differs from run 1(h) by an increased block size of $K = 15$. These nine runs demonstrate the variety of dynamical behaviors exhibited by our simple Royal Road GA.

Figures 1(a) through 1(i) show the results of nine runs of the Royal Road GA with different parameter settings. In each, the average fitness in the population over time is shown (solid lines) together with the best fitness in the population at each time step (diamonds). In all runs the total length of the string $L = NK$ was kept constant at $L = 60$ to reduce the number of free parameters. Figure 1(d) is the central parameter setting of our analysis. Following the large arrows, each successive run differs from the predecessor by a single parameter change.

Run 1(d), selected as the canonical example of epochal evolution, was performed with

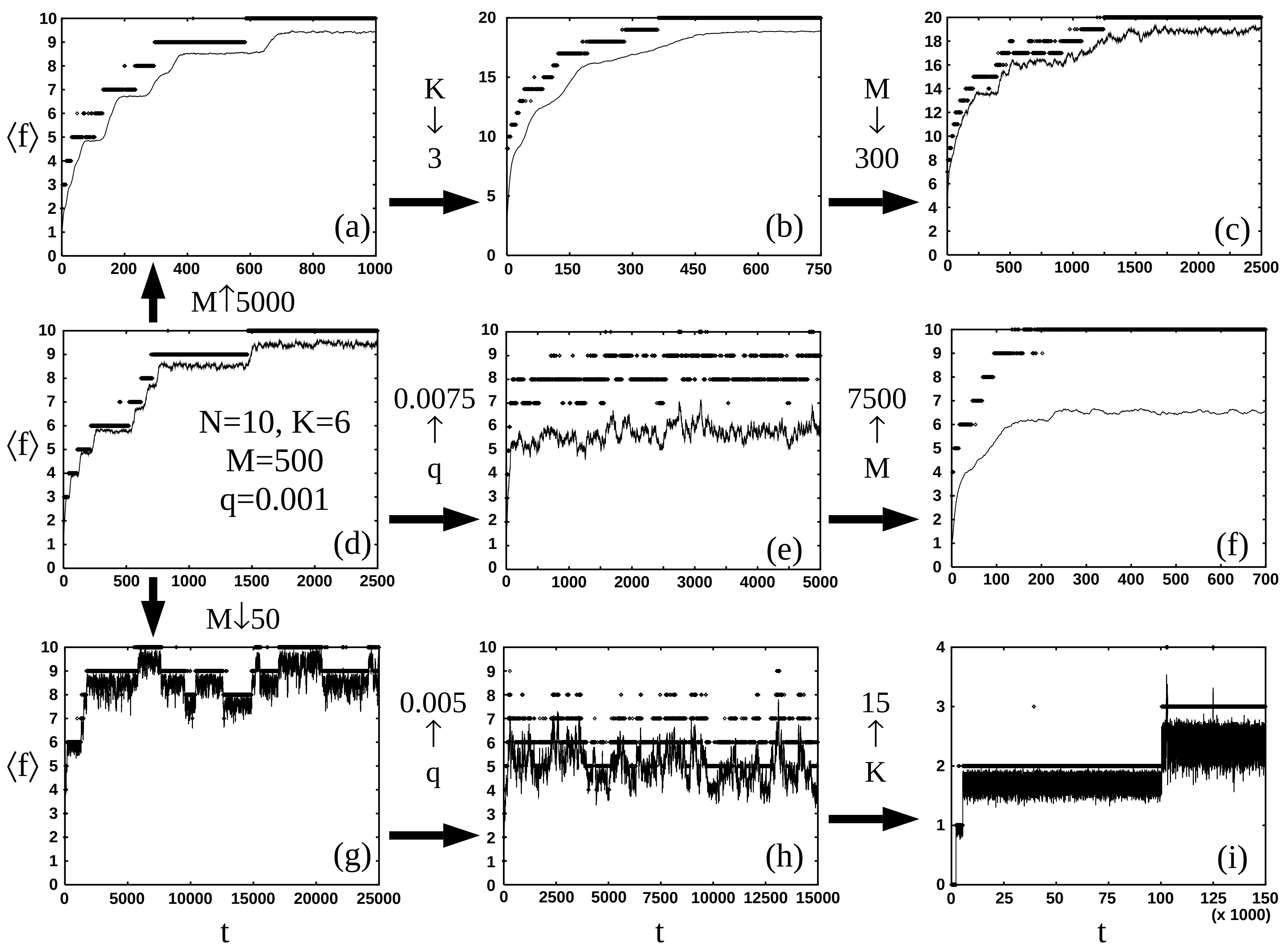

Figure 1: For caption see text.

$N = 10$ blocks of length $K = 6$ bits, a mutation rate of $q = 0.001$, and a population size of $M = 500$. This run clearly shows epochal dynamics in the evolutionary search. This behavior is encountered over a wide range of parameter settings around those of run 1(d). Note that experimentally, the length of the epochs varies greatly from run to run, but the fitness levels at which they occur are the same in each run. The epoch levels therefore constitute a *reproducible* macroscopic feature of the dynamics. Runs 1(a) and 1(g) in the left-hand column show the effect on the search behavior of changing the population size $M$. Run 1(a) was performed with a relatively large population size of $M = 5000$. The fitness epochs are also clear in this plot. The behavior is less noisy, the search for higher fitness quicker on average, and the innovations between the fitness epochs are less rapid. For the small population size of $M = 50$ in figure 1(g), however, the behavior becomes much noisier and the average fitness in the population shows an interesting intermittent behavior at high fitness levels. The average fitness jumps up and down irregularly between the different fitness epochs. We would like to understand what causes the epochal dynamics in Figures 1(a), 1(d), and 1(g). In particular, we would like to understand how the different population sizes lead to these different classes of search behavior.

The runs shown in the middle column—Figures 1(b), 1(e), and 1(h)—illustrate how change in a single parameter can cause fitness epochs to disappear. In run 1(b), for example, the block length was lowered from $K = 6$ to $K = 3$, otherwise the parameter settings are those of 1(a). Note that since we are keeping string length constant ($L = 60$) throughout the figure mosaic, in run 1(b) there are $N = 20$ blocks. In this plot we see the average fitness increasing strictly monotonically over time. Although well-defined fitness epochs cannot be distinguished anymore, the rate of fitness increase does vary in different regions of the run. Note that the fluctuations in this rate of fitness increase are *not* the same between runs. That is, although the rate of fitness increase is decreasing over time on average, the precise locations of the inflection points in the curve of average fitness against time are not the same between runs.

Run 1(e) has a higher mutation rate ($q = 0.0075$) than that of the canonical run 1(d). Here too the fitness epochs have disappeared from the behavior, but in a markedly different way from that in going from run 1(a) to run 1(b). Overall, fitness jumps up quickly at first after which it moves up and down irregularly in a band of fitnesses between 4.5 and 7. Note also that the best fitness in the population alternates between different values in a fitness band between 7 and 10.

Run 1(h) has a higher mutation rate ($q = 0.005$) than that in run 1(g) ($q = 0.001$). The behavior of run 1(h) is similar to that in run 1(e) although the fitness fluctuations occur over a wider range of amplitudes and seem to have larger variation in the associated time scales. In contrast, in run 1(e) the fitness fluctuates up and down on a time scale that is roughly constant. In run 1(h) fluctuations occur both on short time scales as well as on long time scales. Moreover, a hierarchy of fluctuation amplitude bands can be distinguished. As in the transition from 1(d) to 1(e), we would like to understand why the fitness epochs disappear in going from 1(g) to 1(h) by increasing mutation and what mechanisms cause the different types of fluctuation behavior in 1(e) and 1(h).

Finally, the runs in the rightmost column—Figures 1(c), 1(f), and 1(i)—illustrate how change in another (different) single parameter can make epochal evolution dynamics reappear. In run 1(c) $M$ was lowered to 300, compared with run 1(b) in which $M = 5000$. Although the behavior is quite noisy and fitness increases more gradually on average over a wide range of time steps in the run, at least two fitness epochs can be distinguished in this run. One occurs around an average fitness of 13.5 and one around an average fitness of 16. In general, different fitness epochs occur for different runs with these parameter settings, but if a certain epoch occurs, it always occurs at the same fitness level.

In run 1(f) $M$ was raised from 500 to 7500. Here the noise seen in run 1(e) has largely disappeared. Although the fitness increases smoothly almost everywhere in the run, an epoch can be seen around a fitness value of 6. Notice that in contrast to run 1(c), the epochs reappeared here by *increasing* the population size instead of decreasing it. This illustrates the strong interdependence of the parameter settings. That is, the effect of increasing or decreasing a certain parameter is highly dependent on the values of all other parameters. Notice alsothe large gap between the best fitness and average fitness in this run.

Lastly, in run 1(i) the block size was increased to $K = 15$. Here the fitness epochs that disappeared in run 1(h) can be clearly seen again. The behavior is still very noisy within the epochs, but average fitness increases sharply between them. From the run it is not clear if the population will eventually find the highest possible fitness of 4. Note that the best fitness did reach this highest level around generation 100,000, but did not successfully establish itself in the population, almost immediately dropping back to the lower fitness epoch. Again, our task is to explain why epochal behavior reappeared by changing each parameter in these runs.

The goal of this paper is to understand the dynamical features observed in these nine runs. Why does average fitness follow a pattern of stepwise increase for a large range of parameters? At what fitness levels are epochs going to occur and how many can occur for a given parameter setting? What is the average length of the fitness epochs? What determines the size of the fitness fluctuations within these epochs and what determines the speed of the innovations between the fitness epochs? Why do epochs disappear from the dynamics if some parameters are changed? What causes these different kinds of dynamical behaviors, such as the intermittency of run 1(g), the bounded fluctuations seen in runs 1(e) and 1(h), and the relatively smooth dynamics of run 1(b) and run 1(f)? In this paper we will present quantitative answers to almost all of these questions.

Explaining the GA's behaviors at different parameter settings is a first step towards understanding *how* one should optimally design an evolutionary search algorithm. Unfortunately, making general statements about how to set parameters is problematic at best. Since a GA's algorithmic formulation is naturally specified by independent operations on the population—such as selection, mutation, and crossover—it is tempting to base an explanation of GA dynamics on an understanding of the behaviors under selection, under mutation, and under crossover individually. As illustrated in our discussion of figure 1, changing a single parameter such as population size or mutation rate can have very different quantitative and even qualitative effects on GA behavior. Moreover, the resulting behavior depends on the settings of *all* other GA parameters. Thus, although one intuitively thinks of the GA as being composed of selection, mutation, and crossover, its actual behavior is generally not decomposable into these operators' separate effects. The direct consequence is that an integrative approach to the dynamics—one that treats the component operators simultaneously and on a equal footing and that focuses on the emergent effects of their interaction—is necessary to obtain a basic understanding of GA behavior. The following analysis demonstrates how a subtle interplay between evolutionary operators is responsible for a wide range of GA epochal search behaviors. Furthermore, our statistical dynamics approach enables us to describe the behaviors in terms of the emergence of "effective" dynamical forces, which generally involve selection and mutation as well as finite population effects. In our view the identification and analysis of these forces is prerequisite to developing, for example, prescriptions for optimizing evolutionary search parameters.

## 4. The Royal Road GA's State Space

Just as the state of a gas or a fluid in a box is fully described only when the locations and velocities of all the molecules are given, so the state of a population in an evolutionary search algorithm is fully specified only when all of its bit strings are listed. Any knowledge we have of the algorithm's behavior must be rooted in an understanding of the behavior of the algorithm at the lowest level of distributions of actual bit strings. The same holds for the behavior of gases and fluids in nature: they can be understood only in terms of the kinetics of their constituent molecules. It is, however, practically impossible to describe the locations of all the molecules and their velocities, just as it is impractical to describe the state of all individuals in the GA population.

This impracticality is much more of a problem from the experimental point of view than from the theoretical point of view. In fact, it is relatively easy to formally write down the equations of motion on the level of whole populations of bit strings for a simple GA, just as it is easy to formally write down the Newtonian equations of motion for all molecules in a gas. The problem is that no experimentalist will produce lists of all strings in the time series of populations to analyze, or even just describe, the GA's behavior. Instead, the experimentalist produces much more coarse-grained data from simulations—recording the statistics of most interest, such as the average fitness and best fitness in the populations or the number of generations necessary to find a string of a certain fitness.

Thus, the problem facing the theoretician is to develop a theory capable of making quantitative predictions about the GA's behavior on the level of observables that an experimentalist measures. In statistical physics this problem is solved by finding suitable "macroscopic" variables, such as temperature, pressure, and volume, that fully and self-consistently describe the thermodynamic state of the system. Specifically, in the limit of infinite system size—the thermodynamic limit—the description of the system in terms of these macroscopic variables becomes closed. That is, in this limit, one does not have to know the detailed microscopic state of the system to obtain the dynamics of the macroscopic variables.[1]

The strategy of our GA analysis is precisely this. Rather than focus on distributions of individuals in the population, we choose distributions of fitnesses in the population as our macroscopic variables. We claim that in the limit of infinite population size—the analogue of the thermodynamic limit—the evolutionary system can be deterministically and self-consistently modeled in terms of these variables. As is generally true in statistical physics, this claim is not based on rigorous mathematical proof. Rather we assume that for very large populations, the dynamics of the population's fitness distribution can be described solely in terms of the fitness distribution. This assumption is supported by the results of our simulations with large populations and by mathematical arguments that describe how the large population dynamics converges to the behavior of infinite populations. Again, the situation is analogous to that in statistical physics. Although it may not be widely appreciated, at present there is no general mathematical proof that the behavior of observables such as pressure, temperature, and volume of a gas can be described in terms of these macroscopic variables only. This fact derives from *experimental* observation.

Mathematically, the dynamics of the macroscopic variables is obtained by making the maximum entropy assumption. This says that all microscopic states that are consistent with a certain macroscopic state are equally likely. In our case, we assume that if a population has a

---

[1] This is strictly only true for systems in thermodynamic equilibrium.

certain fitness distribution $\Pr(f)$, then the microscopic state of the population—the distribution of genotypes—is equally likely to be any of the microscopic population states consistent with $\Pr(f)$. Starting from the "microscopic" specification of how the GA acts on string populations and using the maximum entropy assumption, a generation operator $\mathbf{G}$ can be constructed that, in the limit of infinite populations, deterministically describes the action of the GA on fitness distributions. In section 5.1 we will give theoretical arguments in support of the maximum entropy assumption, but ultimately its validity must be tested by comparing our theory's predictions against statistics gathered from the simulation experiments.

To implement this approach we describe the state of the population at a given generation by its distribution $\vec{P}(t)$ of string fitnesses, rather than by the distribution of string probabilities $\Pr(s)$ directly. This projects from the $2^L$-dimensional space of strings to an $N$-dimensional space. $\vec{P}(t)$'s components, $P_f(t), 0 \leq f \leq N$, represent the fraction of individuals in the population with fitness $f$ at time $t$. Thus, the state of the population is described by a $N+1$-dimensional vector $\vec{P}$ whose components sum up to 1; that is,

$$\sum_{f=0}^{N} P_f = 1. \tag{4.1}$$

Therefore, the state space has $N$ independent components (dimensions). The average fitness in the population is given by

$$\langle f \rangle = \sum_{f=0}^{N} f P_f. \tag{4.2}$$

We denote the set of all possible states $\vec{P}$ for a finite population of size $M$ as $\Lambda_M$. This state space is given by

$$\Lambda_M = \{\vec{P} \, : \, P_f = \frac{n_f}{M}, \; n_f \in \mathbb{N}, \; \sum_{f=0}^{N} n_f = M\}, \tag{4.3}$$

where $n_f$ is the number of individuals with fitness $f$ in the population. The number of points in the state space $\Lambda_M$ for finite populations is

$$|\Lambda_M| = \binom{M+N}{N}, \tag{4.4}$$

which is on the order of $M^N$ states for $N \ll M$. The above set $\Lambda_M$ forms a rectangular lattice within the simplex in $N+1$ dimensions with lattice spacing $1/M$. In the limit of infinite populations the lattice points become dense in the simplex and we can take the state space to be the simplex itself:

$$\Lambda_\infty = \{\vec{P} \in \mathbb{R}^{N+1} \, : \, \sum_{f=0}^{N} P_f = 1, \; P_f \geq 0\}. \tag{4.5}$$

Another important assumption in considering a state space of fitness distributions is that the dynamics of the fitness distribution is not sensitive to the underlying microscopic state of the population. Note that this is an extra restriction in addition to the maximum entropy assumption. Even if it were true that for a certain fitness distribution $\vec{P}$ the GA state is equally likely to be in any microscopic population $\Pr(s)$ consistent with $\vec{P}$, then there could still be an inconsistency if the fitness distribution dynamics was sensitive to some more detailed feature of $\Pr(s)$. In the statistical physics literature this problem is called: lack of "self-averaging". Thus, our approach assumes that the dynamics on the level of fitness distributions is "self-averaging". In other words, for large populations we assume that two microscopic population states with the same fitness distribution tend to give rise to the same fitness distribution dynamics. As before, this assumption is supported by statistics gathered from our simulation experiments. We expect though that self-averaging is not generally valid for arbitrary GA dynamics. That it works so well is, therefore, a specific feature of our mutation-only GA and the Royal Road fitness function. For example, if crossover were employed by our GA, the dynamics of strings with fitness $n$ would be sensitive to *where* the $n$ aligned blocks were located in the strings. Fitness distributions then would be an overly coarse-grained set of macroscopic variables. We would have to augment them to include other variables that appropriately accounted for statistical properties of the aligned-block locations.

## 5. Infinite Population Dynamics on Fitness Distribution Space

We will first solve for the dynamics of the fitness distribution in the infinite-population limit ($M \to \infty$) by constructing a generation operator **G** that incorporates the effects of selection and mutation on the fitness distribution. The state $\vec{P}(t+1)$ of the population at time $t+1$ is obtained by acting with the operator **G** on the state $\vec{P}(t)$ of the population at time $t$:

$$\vec{P}(t+1) = \mathbf{G}[\vec{P}(t)], \tag{5.1}$$

where the operator **G** formally can be written as the product,

$$\mathbf{G} = \mathbf{M} \cdot \mathbf{S}, \tag{5.2}$$

of the selection operator **S** and the mutation operator **M**.

This construction is similar in spirit to the generation operators in [45], although since we focus on fitness distributions rather than entire populations, we will be able to analyze the effect of this operator explicitly and quantitatively. The focus on the dynamics of fitness distributions is similar to the analysis in [38], where the dynamics of the first few cumulants of the fitness distribution is analyzed. Here we track the dynamics of the entire fitness distribution.

Equation 5.1 is analogous to a discrete-time version of the differential equations of the Eigen model of molecular evolution [11, 13]. In that model one considers the evolutionary dynamics of the distribution of different self-replicating molecular "genotypes". Therefore, the Eigen model equations are again defined on the microscopic level, not on a coarse-grained level such as that of fitness distributions.

We will now construct the operator **G** explicitly for the case of our simple GA. We first consider the alignment dynamics of unaligned blocks and then build the mutation and selection components of **G**. Once we have this composite expression for **G**, we turn to explicitly solving for the dynamics.

### 5.1 Block Alignment Dynamics

We will now present some heuristic theoretical arguments for our maximum entropy assumption: given a fitness distribution, all microscopic populations with that fitness distribution are equally likely to occur. Since the dynamics of the mutation-only GA is obviously independent of where the aligned blocks in strings occur, our maximum entropy assumption boils down to assuming that the bit values in unaligned blocks are essentially random and statistically independent of each other.

Assuming that the unaligned blocks are random we can calculate the probability $A$ that mutation will transform an unaligned block into an aligned block. A random unaligned block is equally likely to be in any of the $2^K - 1$ unaligned block states. If the block has $d$ zeros, the probability that it will turn into an aligned block under mutation is just $q^d(1-q)^{K-d}$. There are $\binom{K}{d}$ different block states which have $d$ zeros. We thus have:

$$A = \frac{1}{2^K - 1} \sum_{d=1}^{K} \binom{K}{d} q^d (1-q)^{K-d} = \frac{1-(1-q)^K}{2^K - 1}. \tag{5.3}$$

The above expression for $A$ is a direct consequence of the maximum entropy assumption that all microscopic populations with the same fitness distribution are equally likely. This is, of course, no guarantee that the GA will actually behave according to this assumption. The reason we expect unaligned blocks to be random on average is because (i) selection in our GA only counts the number of aligned blocks and therefore acts on all unaligned blocks equally and (ii) mutation randomly mixes bit values in blocks. But let's be more precise about our assumption.

Consider a particular unaligned block $b$ in a given string $s$. We consider two cases. First, string $s$ either has an ancestor in which block $b$ was aligned and then was destroyed or, second, none of its ancestors had $b$ aligned. For low mutation rates, block $b$ is likely to have more bits set in the first case than in the second, since in the first case it is the remnant of a previously aligned block and in the second it is the descendant of well-mixed random unaligned blocks.

Now consider unaligned blocks in strings that belong to the highest fitness class currently in the population at time $t$ during a run. Let $\mathcal{P}(t)$ denote the current population and let $\mathcal{P}_h(t)$ denote the set of strings having the highest fitness in $\mathcal{P}(t)$:

$$\mathcal{P}_h(t) = \{s' \in \mathcal{P}(t) : f(s') \geq f(s) \; \forall s \in \mathcal{P}(t)\}$$

Unaligned blocks in such strings are very likely to have descended from unaligned blocks in ancestral strings. This is because it is unlikely that strings with fitness higher than those of $\mathcal{P}_h(t)$ were ever present in the population in appreciable numbers. Therefore, at any time $t$ in the run we can assume that these unaligned blocks have been subject to mutation for $t$ time steps without ever having been aligned. Assuming that we can take different unaligned blocks in different strings in $\mathcal{P}_h(t)$ to be independent, we can take the state of these blocks to be essentially random. Different unaligned blocks within the *same* string in $\mathcal{P}_h(t)$ can in general be taken to be independent to a high approximation. The assumption that the unaligned blocks in different strings in $\mathcal{P}_h(t)$ are independent is only appropriate for very large populations, as we will see later on. For infinite populations this assumption of independence is exact for blocks in strings that are in $\mathcal{P}_h(t)$. We have therefore argued that unaligned blocks in $\mathcal{P}_h(t)$ can be taken as random, at least for very large populations.

We now turn to the case of unaligned blocks in strings not in $\mathcal{P}_h(t)$. Let $s$ be in $\mathcal{P}(t)$ but not in $\mathcal{P}_h(t)$. The unaligned blocks in $s$ can be divided into two types: blocks that were never aligned in ancestors of $s$ and blocks that were aligned in some higher-fitness ancestor of $s$ but that were destroyed by mutation. The first type of block can also be taken as random and has the same probability $A$ given by equation 5.3 of becoming aligned by mutation. The second type of block is likely to have more bits set to 1 than a random block, so it has a higher probability of being aligned by mutation. In general we will not try to solve for the relative proportions of the two types of blocks in strings $s$. We can give, however, lower and upper bounds on the probability $A_s$ that an unaligned block in $s$ will become aligned through mutation. The lower bound is obtained by assuming that *all* unaligned blocks are of the first type and thus have never been aligned before, which yields:

$$A_s \geq A. \tag{5.4}$$

An upper bound is obtained by assuming that all blocks have $K - 1$ of the $K$ bits set, giving

$$A_s \leq q(1-q)^{K-1}. \tag{5.5}$$

We will see later on that many of our results are largely insensitive to the value of $A_s$ within these bounds. For convenience we will use $A_s = A$ and employ the same probability for all unaligned blocks in all strings.

Again we stress that the above arguments do not prove that unaligned blocks behave as random blocks for our GA. In fact, we will later see some examples were this approximation breaks down for the dynamics of finite populations. However, for very large populations, our experiments show that this maximum entropy assumption gives excellent predictions of the dynamics of the actual GA. In the following sections we will solve for the infinite population dynamics of the GA under this assumption.

## 5.2 The Mutation Operator M

In the last section we derived an expression for the probability $A$ to align a block through mutation in one time step under the random block approximation. The probability $D$ that mutation will destroy an aligned block is simply given by

$$D = 1 - (1-q)^K. \tag{5.6}$$

Using $A$ and $D$, we now consider the probability $\mathbf{M}_{ij}$ that mutation turns a string with fitness $j$ (i.e. $j$ aligned blocks) into a string with fitness $i$. In other words, $\mathbf{M}_{ij}$ is the probability that by mutating every bit with probability $q$ in a string of fitness $j$ this string will turn into a string with fitness $i$. We can write $\mathbf{M}_{ij}$ as the sum over all the probabilities that $k$ unaligned blocks in the string will be aligned and $l$ aligned blocks will be destroyed such that $j + k - l = i$. Thus, we have

$$\mathbf{M}_{ij} = \sum_{k=0}^{N-j} \sum_{l=0}^{j} \delta_{j+k-l,i} \binom{N-j}{k} \binom{j}{l} A^k (1-A)^{N-j-k} D^l (1-D)^{j-l}. \tag{5.7}$$

In the limit of infinite populations, the operator $\mathbf{M}_{ij}$ acting on a fitness distribution $\vec{P}$ will give the fitness distribution $\vec{P}^m$ after mutation:

$$P_i^m = \sum_{j=0}^{N} \mathbf{M}_{ij} P_j, \tag{5.8}$$

where $P_i$ and $P_i^m$ are the proportions of strings with fitness $i$ before and after mutation, respectively. The mutation operator $\mathbf{M}$ is an ordinary linear matrix operator with the property that

$$\sum_{i=0}^{N} \mathbf{M}_{ij} = 1. \tag{5.9}$$

This is, of course, just another way of saying that summing the probabilities of *all* possible outcomes of mutation gives unity. That is, $\mathbf{M}$ is a stochastic matrix.

### 5.3 The Selection Operator S

Our simplified GA uses fitness proportionate selection. This means that the proportion $P_i^s$ of strings with fitness $i$ after selection is proportional to both $i$ and the fraction $P_i$ of strings with fitness $i$ before selection:

$$P_i^s = c\, i\, P_i, \tag{5.10}$$

where $c$ is a constant.[1] The constant can easily be obtained by demanding that the vector $\vec{P}^s = (P_0^s, \ldots, P_N^s)$ is normalized:

$$\sum_{i=0}^{N} P_i^s = 1. \tag{5.11}$$

Therefore, we have

$$c = \left[ \sum_{i=0}^{N} i P_i \right]^{-1} = \frac{1}{\langle f \rangle}, \tag{5.12}$$

where $\langle f \rangle$ is the average fitness of the population. We can write the entries $\mathbf{S}_{ij}$ of the selection operator as the diagonal matrix

$$\mathbf{S}_{ij} = \frac{i \delta_{ij}}{\langle f \rangle}. \tag{5.13}$$

Notice that, in contrast to the mutation operator, the selection operator is nonlinear because it depends on the average fitness of the distribution it acts on.

### 5.4 The Generation Operator G

We can now construct the generation operator $\mathbf{G}$ that maps the fitness distribution of a population into the fitness distribution of the population at the next generation. $\mathbf{G}$ is the product of the selection and mutation operators:

$$\mathbf{G}_{ij} = \sum_{k=0}^{N} \mathbf{M}_{ik} \mathbf{S}_{kj}. \tag{5.14}$$

To analyze the dynamics of this operator we first construct a linearized version $\tilde{\mathbf{G}}$. We note that all entries in the generation operator $\mathbf{G}$ are independent of the vector $\vec{P}$ it acts on, apart from the normalization factor $1/\langle f \rangle$ in $\mathbf{S}$. We can take this factor outside the matrix and write

$$\mathbf{S} = \frac{1}{\langle f \rangle} \tilde{\mathbf{S}}, \tag{5.15}$$

---

[1] In the case where all strings have zero fitness, the GA is not well defined. In all practical situations considered here the fitness distribution will have nonzero proportions of nonzero fitness strings at all times.

giving

$$\tilde{\mathbf{G}} = \mathbf{M} \cdot \tilde{\mathbf{S}}. \tag{5.16}$$

The operator $\tilde{\mathbf{G}}$ is an ordinary $(N+1)$ by $(N+1)$ matrix with nonnegative entries.

The fitness distribution in the population at time $t$ is given by the $t^{th}$ iterate of $\mathbf{G}$ acting on the initial fitness distribution $\vec{P}(0)$. That is,

$$\vec{P}(t) = \mathbf{G}^t[\vec{P}(0)]. \tag{5.17}$$

Since at each iteration the operator $\mathbf{G}$ is just the matrix $\tilde{\mathbf{G}}$ times a constant that depends on the fitness distribution it acts on, $\mathbf{G}^t$ is proportional to the linear operator $\tilde{\mathbf{G}}^t$ times a constant that depends only on $\vec{P}(0)$ and $t$. Therefore, we can express the fitness distribution $\vec{P}(t)$ at time $t$ as a constant times $\tilde{\mathbf{G}}^t$ acting on $\vec{P}(0)$. Thus, we have

$$\vec{P}_i(t) = \sum_j C[t, \vec{P}(0)] \tilde{\mathbf{G}}^t_{ij} P_j(0). \tag{5.18}$$

We can determine the constant $C[t, \vec{P}(0)]$ easily by requiring that $\vec{P}(t)$ is normalized. The result is that

$$C[t, \vec{P}(0)] = \left[ \sum_{i,j} \tilde{\mathbf{G}}^t_{ij} P_j(0) \right]^{-1}. \tag{5.19}$$

Since the initial population consists of random strings of length $L = NK$, the initial fitness distribution $\vec{P}(0)$ can be obtained by considering that at $t = 0$ each block in each string has a probability $2^{-K}$ to be aligned. We then find that

$$P_i(0) = \binom{N}{i} 2^{-Ki} \left(1 - 2^{-K}\right)^{N-i}. \tag{5.20}$$

We can solve explicitly for $\vec{P}(t)$ by diagonalizing the matrix $\tilde{\mathbf{G}}$. In general, $\tilde{\mathbf{G}}$ will have $N+1$ distinct eigenvectors and eigenvalues.[2] We will denote these eigenvectors $\vec{V}^i$ and their corresponding eigenvalues $g_i$. These obey

$$\tilde{\mathbf{G}} \cdot \vec{V}^i = g_i \vec{V}^i, \tag{5.21}$$

for each value of $i$ from 0 to $N$. We further normalize these eigenvectors in probability, so that

$$\sum_{i=0}^{N} V^j_i = 1. \tag{5.22}$$

Defining the matrix $\mathbf{R}$ to contain $\tilde{\mathbf{G}}$'s $N+1$ normalized eigenvectors $\vec{V}^i$ as its columns, the matrix $\mathbf{R}$ and its inverse diagonalize the generation operator, in the sense that

$$\mathbf{G}'_{ij} = (\mathbf{R}^{-1} \cdot \tilde{\mathbf{G}} \cdot \mathbf{R})_{ij} = g_i \delta_{ij}. \tag{5.23}$$

---

[2] In the very rare cases were the characteristic polynomial of $\tilde{\mathbf{G}}$ has multiple roots, these roots can be separated by an infinitesimal change in the mutation rate $q$.

Since the eigenvectors are normalized, the matrix $\mathbf{R}$ has the additional property that its columns add up to 1:

$$\sum_{i=0}^{N} \mathbf{R}_{ij} = \sum_{i=0}^{N} V_i^j = 1. \tag{5.24}$$

The generation operator $\tilde{\mathbf{G}}^t$ can now be written in terms of its eigenvalues $g_i$, the matrix $\mathbf{R}$, and its inverse, as follows

$$\tilde{\mathbf{G}}_{ij}^t = \sum_{k=0}^{N} \mathbf{R}_{ik} g_k^t \mathbf{R}_{kj}^{-1}. \tag{5.25}$$

This allows us to solve for $\vec{P}(t)$ in equation 5.18, obtaining

$$\vec{P}_i(t) = C[t, \vec{P}(0)] \sum_{k,j} \mathbf{R}_{ik} g_k^t \mathbf{R}_{kj}^{-1} P_j(0). \tag{5.26}$$

Equation 5.26 gives an exact expression for the fitness distribution in the population as a function of time in the limit of infinite population size. We can make equation 5.26 more transparent by moving to the basis of $\tilde{\mathbf{G}}$'s eigenvectors. First, we write the fitness distributions in this basis as

$$\alpha_i(t) = \sum_{j=0}^{N} \mathbf{R}_{ij}^{-1} P_j(t). \tag{5.27}$$

Note that the $\alpha_i(t)$ are normalized. We further simplify the expression for $C[t, \vec{P}(0)]$ using equations 5.25 and 5.27 in equation 5.19:

$$C[t, \vec{P}(0)] = \left[ \sum_{i,j,k=0}^{N} \mathbf{R}_{ik} g_k^t \mathbf{R}_{kj}^{-1} P_j(0) \right]^{-1} = \left[ \sum_{k=0}^{N} g_k^t \alpha_k(0) \right]^{-1}. \tag{5.28}$$

Transforming the equations of motion (equation 5.26) to the basis of the eigenvectors $\vec{V}^i$ we find that the fitness distribution is given by

$$\alpha_i(t) = \frac{g_i^t \alpha_i(0)}{\sum_{j=0}^{N} g_j^t \alpha_j(0)}. \tag{5.29}$$

From this we get a very simple expression for the average fitness $\langle f(t) \rangle$ as a function of time. Again using the fact that the rows of $\mathbf{R}$ sum up to 1, a little algebra gives:[3]

$$\langle f \rangle = \sum_i i P_i = \sum_i g_i \alpha_i. \tag{5.30}$$

From the two preceding expressions, we are left with a simple, direct expression for the behavior of the average fitness as a function of time for infinite populations:

$$\langle f(t) \rangle = \frac{\sum_i g_i^{t+1} \alpha_i(0)}{\sum_j g_j^t \alpha_j(0)}. \tag{5.31}$$

---

[3] See the next section for a detailed derivation of this property.

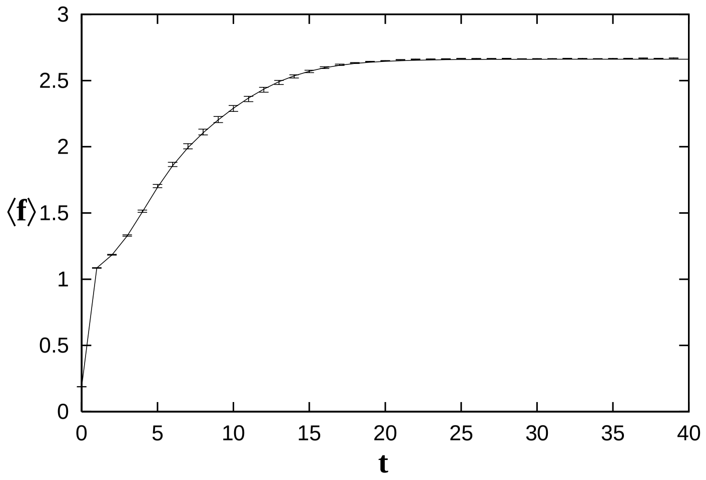

Figure 2: Average fitness $\langle f(t) \rangle$ over time $t$ averaged over 20 GA runs along with the theoretical prediction (solid line) for infinite population obtained from equation 5.31. The error-bars show $\pm 2$ standard deviations from the average fitness over the 20 runs. The parameters here are $N = 3$ and $K = 4$ with $q = 0.01$ and a population size of $10^4$.

Figure 2 shows the predicted infinite population dynamics for a particular parameter setting together with the empirical results averaged over 20 runs with a large population. The figure shows that for large populations, typically for $M > 2^{NK}$, the actual average fitness behavior follows the infinite population dynamics, obtained using our maximum entropy assumption, quite closely. This supports our assumption that for large populations, the evolution of the fitness distribution can be described in terms of the fitness distribution only. (Appendix B proves that under the maximum entropy assumption the dynamics of large populations converges towards the deterministic dynamics as given by $\mathbf{G}$.) The small error bars in the figure also show that for these large populations, there are very small fluctuations in the evolution of the average fitness between runs. This fact supports our assumption that the dynamics on the level of fitness distributions is self-averaging in the limit of large populations. Notice, though, that for populations of this size the average fitness increases smoothly as a function of time and there are no fitness epochs. Apparently, fitness epochs occur only for population sizes that are not too large.

Recapitulating, by obtaining the probabilities $A$ and $D$ to align and destroy blocks we constructed a matrix operator that describes the dynamics of the genetic algorithm in the limit of infinite populations. By linearizing $\mathbf{G}$ to form $\tilde{\mathbf{G}}$ and then computing $\tilde{\mathbf{G}}$'s eigenvalues and eigenvectors, we were able to solve exactly for (i) the dynamics of the fitness distribution (equation 5.29) and (ii) the average fitness $\langle f(t) \rangle$ in the population (equation 5.31).

### 5.5 Properties of the Generation Operator G

The fixed points of the operator $\mathbf{G}$ are those vectors $\vec{V}$ that are mapped to themselves under its action. Whenever the fitness distribution $\vec{P}$ describing the infinite population equals one of those vectors $\vec{V}$, it will remain at $\vec{V}$ from then on. Consider the fixed point equation for the dynamics, $\vec{P} = \mathbf{G}(\vec{P})$. We have

$$\mathbf{G}(\vec{P}) = \frac{\tilde{\mathbf{G}}}{\langle f \rangle} \cdot \vec{P} \; ; \qquad (5.32)$$

so for fixed points $\vec{P}$ of $\mathbf{G}$,

$$\tilde{\mathbf{G}} \cdot \vec{P} = \langle f \rangle \vec{P}. \tag{5.33}$$

In words, the fixed points of $\mathbf{G}$ are given by normalized eigenvectors of $\tilde{\mathbf{G}}$, with the extra restriction that $\tilde{\mathbf{G}}$'s eigenvalues are equal to the fitness average of the eigenvector.

Using the stochasticity property of the mutation operator $\mathbf{M}$ (equation 5.9), we find that *all* eigenvectors $\vec{V}^i$ of $\tilde{\mathbf{G}}$ fulfill the above restriction (equation 5.33). First, using equation 5.21, note that

$$\sum_{i,j} \tilde{\mathbf{G}}_{ij} V_j^k = \sum_i g_k V_i^k = g_k. \tag{5.34}$$

This simply states the fact that $\vec{V}^k$ is a normalized eigenvector with eigenvalue $g_k$. Furthermore, substituting the definition of $\tilde{\mathbf{G}}$ in terms of $\mathbf{M}$ and $\mathbf{S}$, we have

$$\sum_{i,j} \tilde{\mathbf{G}}_{ij} V_j^k = \sum_{i,j} \mathbf{M}_{ij} j V_j^k = \sum_j j V_j^k = \langle f \rangle. \tag{5.35}$$

Thus, we see that

$$g_k = \langle f \rangle \tag{5.36}$$

for all eigenvectors $\vec{V}^k$, which gives a simple interpretation of their eigenvalues as average fitnesses.

This implies, in turn, that all eigenvectors of $\tilde{\mathbf{G}}$ are fixed points of the generation operator $\mathbf{G}$. This might lead one to believe that the infinite population GA dynamics could end up in $N$ different stable states in fitness distribution space.[4] This is not true. The normalized eigenvectors of $\tilde{\mathbf{G}}$ have to be positive definite to be interpreted as fitness distributions and, in general, they are not. In Appendix A we prove that the matrix $\tilde{\mathbf{G}}$ can have only a single positive definite eigenvector. Therefore, there is a unique fixed-point fitness distribution towards which the Royal Road GA will always converge asymptotically. All other fixed points lie outside the simplex $\Lambda_\infty$ and are therefore unreachable. The asymptotic fixed point corresponds to $\tilde{\mathbf{G}}$'s principal eigenvector—the eigenvector of $\tilde{\mathbf{G}}$ with the largest eigenvalue. This can also be understood by considering equation 5.31, which shows that the largest eigenvalue $g_N$—having chosen the ordering $g_0 < g_1 < \cdots < g_N$—will dominate the average fitness in the limit of $t \to \infty$. That is,

$$\lim_{t\to\infty} \langle f(t) \rangle = g_N. \tag{5.37}$$

In addition, we see that the fitness distribution $P(t)$ asymptotically approaches the principal eigenvector:

$$\lim_{t\to\infty} \vec{P}(t) = \vec{V}^N. \tag{5.38}$$

Globally, the map $\mathbf{G}$ has one fixed point lying inside the state space $\Lambda_\infty$ and $N-1$ fixed points lying outside.

Notably, the asymptotic distribution $\vec{V}^N$ over the different fitness classes is analogous to what is called a quasispecies in molecular evolution theory [11]. A species—a particular virus for instance—cannot in general be identified with a certain unique genotype. In fact, the genotypes of members of a certain species can often be thought of as a cloud of points in sequence space that is centered around a genotype of highest fitness. The size and shape of this cloud depend on

---

[4] Remember that we excluded the possibility of all strings having 0 fitness.

the structure of the fitness variations around this peak and on the interplay between mutation and selection. Mutation causes points to drift away from the peak, while selection tends to replicate only individual genotypes whose fitness remains close to that of the peak.[5] This cloud is called a "quasispecies" and is equivalent to our asymptotically stable eigenvector $\vec{V}^N$. In our case, $\vec{V}^N$ represents a cloud of strings with different fitnesses centered around the string with all blocks aligned. Here, as in the molecular evolution setting, the shape and size of the cloud are obtained by solving for the principal eigenvector $\vec{V}^N$ of the generation operator. Note though that since $\vec{V}^N$ is a fitness distribution, the lower components of $\vec{V}^N$ do not correspond to particular genotypes as in the molecular quasispecies case, but rather to sets of genotypes with equal fitness. Our quasispecies distribution is therefore a "phenotypic" or fitness quasispecies cloud.

### 5.6 GA Dynamics as a Flow in Fitness Distribution Space

With equation 5.26 we have an exact expression for the evolution of the fitness distribution $\vec{P}(t)$. As mentioned above, this dynamics shows a smooth and strictly monotonic increase of average fitness starting from a random initial population; figure 2. It does not exhibit the fitness epochs illustrated in figure 1. To begin to explain the occurrence of fitness epochs for small populations we must first adopt a different and more geometric view of the infinite population dynamics. We will visualize the infinite population dynamics as a flow in the space of fitness distributions $\Lambda_\infty$. As a simple example, figure 3 illustrates the dynamics for the parameter settings $N = 3$, $K = 4$, and $q = 0.01$. The state space $\Lambda_\infty$ can be projected into 3-dimensional Euclidean space for this case. Figure 3 has $P_0$ on the $x$-axis, $P_1$ on the $y$-axis, and $P_2$ on the $z$-axis. Of course, we have $P_3 = 1 - P_0 - P_1 - P_2$.

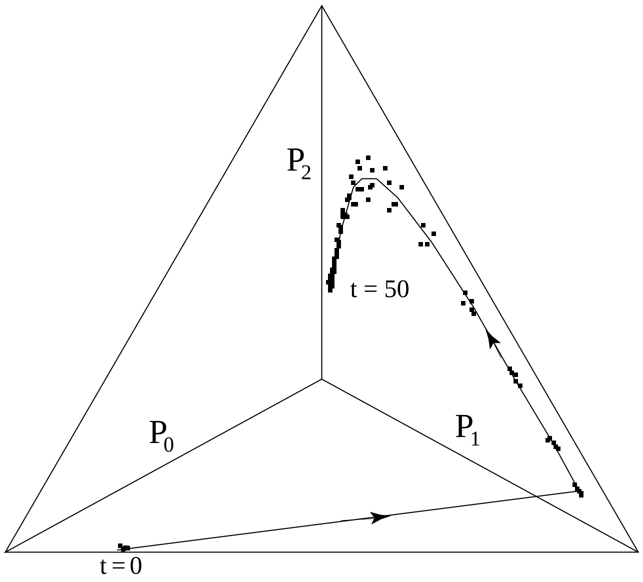

Figure 3: Infinite population dynamics as a trajectory (solid line) in $\Lambda_\infty$ for $N = 3$, $K = 4$, and $q = 0.01$, together with the results (squares) of five runs with these parameters and $M = 10^4$. Cf. figure 2, which plots the average fitness dynamics.

---

[5] It is often tacitly assumed that genotypes with fitness close to that of the peak remain genotypically close to the peak as well. See [25] for a discussion of the implications of this *not* being the case.

Figure 3 shows the theoretical infinite population trajectory of the dynamics in $\Lambda_\infty$ for these parameter settings, together with the empirical dynamics for five runs with a large population ($M = 10^4$). We see that the empirical dynamics follows the infinite population trajectory quite closely. (Compare the average fitness dynamics plotted in figure 2.) This again shows that the evolution of the fitness distribution is well approximated by our maximum entropy assumption for large populations.

We can get an idea of the force driving the fitness distribution along its trajectory through the simplex by considering the difference $d\vec{P} = \mathbf{G}(\vec{P}) - \vec{P}$ of the fitness distribution $\vec{P}$ and the fitness distribution $\mathbf{G}(\vec{P})$ at the next time step. The vector $d\vec{P}$ gives the local direction and magnitude of the change of the fitness distribution at $\vec{P}$ over one time step. We will refer to this change $d\vec{P}$ as the "flow" induced by the generation operator at the point $\vec{P}$ in the state space.[6] Figure 4 shows $d\vec{P}$ in the simplex for the same parameter settings as in figure 3. We now see how the trajectory as shown in figure 3 can be understood as the fitness distribution following the flow $d\vec{P}$ through fitness distribution space. The geometric view of the GA dynamics is that the flow $d\vec{P}$ in the simplex is followed by the fitness distribution in the limit of infinite populations. This will help us understand the occurrence of fitness epochs for finite populations. We will see that in the finite population case the fitness distribution *attempts* to follow this flow, but for reasons to be discussed in the next section, cannot always (and not everywhere) exactly follow it. This, it turns out, is precisely the origin of the fitness epochs, seen only for finite populations.

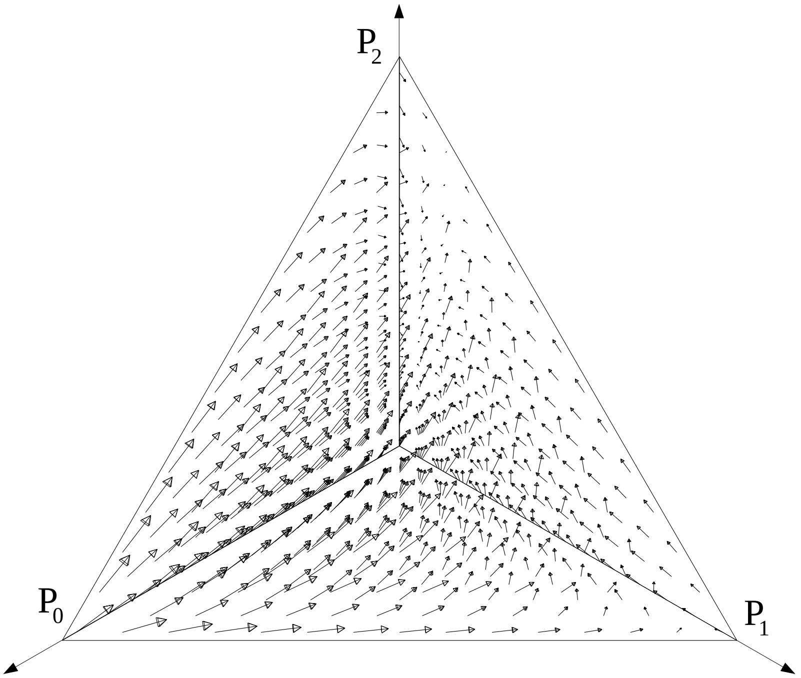

Figure 4: Fitness distribution flow $d\vec{P}$ in the simplex for $N = 3$, $K = 4$, and $q = 0.01$.

---

[6] We are using "flow" in a somewhat loose, but convenient, sense. In dynamical systems theory "flow" is normally reserved for the collection of all trajectories through state space over some time interval. Our usage of the term "flow" corresponds to the vector field that points along the trajectory for each point in state space. It can therefore be seen as the vector field *generating* the flow at each point of state space.

## 6. Finite Population Dynamics

As we saw in section 4, the fitness distribution state space for a finite population is given by a rectangular lattice $\Lambda_M$ over the $N+1$ dimensional simplex $\Lambda_\infty$ with lattice spacing $1/M$. As in the infinite population case, applying the generation operator $\mathbf{G}$ to a fitness distribution $\vec{P} \in \Lambda_M$ gives the *expected* fitness distribution $\langle\vec{P}\rangle$ at the next time step. Viewed in a slightly different way, $\mathbf{G}_i(\vec{P}) = \langle P_i \rangle$ is the probability that if one string is selected from $\vec{P}$ and mutated, it will turn out to have fitness $i$. Since the new population is created from $M$ of these selections and mutations, it is clear that the new population is a sample of size $M$ of the distribution $\langle\vec{P}\rangle$. In the infinite population case the *expected* distribution $\langle\vec{P}\rangle$ is always attained; there are no fluctuations in the $\langle P_i \rangle$ in this limit. For a finite population of size $M$, the finite sampling from $\langle\vec{P}\rangle$ leads to stochasticity in the dynamics. At each time step the population that is actually attained will fluctuate around the expected distribution $\langle\vec{P}\rangle$. We therefore see that the flow in state space for a finite population is given by the same flow operator $\mathbf{G}$ but now over a discrete space and with noise added due to the population's finiteness. It is important to note that the finite population sampling noise added to the operation of $\mathbf{G}$ takes into account the combined effect of the fluctuations in both mutation and selection. Normally, one would consider the sampling fluctuations in mutation and selection to be two distinct sources of fluctuations. In combining the selection and mutation effects into the operator $\mathbf{G}$, all fluctuations are due to the finite population sampling applied to the operation of $\mathbf{G}$.

In general, the probability $\Pr[\vec{P}^n \rightarrow \vec{P}^m]$ that a fitness distribution $\vec{P}^n = (n_0, n_1, \ldots, n_N)/M$ will go to a fitness distribution $\vec{P}^m = (m_0, m_1, \ldots, m_N)/M$ under mutation and selection is given by a multinomial sampling distribution with mean $\mathbf{G}(\vec{P}^n)$:

$$\Pr[\vec{P}^n \rightarrow \vec{P}^m] = M! \prod_{i=0}^{N} \frac{[\mathbf{G}_i(\vec{P}^n)]^{m_i}}{m_i!}, \tag{6.1}$$

where $\mathbf{G}_i(\vec{P}^n)$ is the expected proportion of individuals with fitness $i$ at the next generation. The multinomial distribution is nothing more than the distribution of a random sample of size $M$ of the expected distribution $\mathbf{G}(\vec{P})$. In Appendix B this distribution is used to prove that for large populations the finite population dynamics approaches the infinite population dynamics arbitrarily closely.

### 6.1 Fitness Epochs

We will now turn to the fitness epochs that occur for finite population size. As we have seen, $\mathbf{G}(\vec{P})$ gives the expected distribution at the next time step if the current distribution is $\vec{P}$. The expected flow $\langle d\vec{P}\rangle$ in fitness distribution space is given by $\langle d\vec{P}\rangle = \mathbf{G}(\vec{P}) - \vec{P}$. Now let us consider what happens if the absolute value $|\langle dP_i \rangle|$ of a certain component $\langle dP_i \rangle$ of the expected flow is much smaller than the lattice spacing; that is, if $|\langle dP_i \rangle| \ll 1/M$. The fitness distribution $\vec{P}$ for a finite population can only be on the lattice points of $\Lambda_M$, since the actual difference $dP_i$ can only be $\ldots -1/M\ 0, 1/M, 2/M \ldots$. Thus, we expect the actual difference $dP_i$ to be 0 most of the time; this means that the actual component $P_i$ will likely not change for some time. For example, in the previous case with $N = 3$, $K = 4$, and $q = 0.01$, but with finite $M = 100$, we find that if $P_3 = 0$, the dynamics will push $P_0$, $P_1$, and $P_2$ into a region where

$$|\langle dP_3 \rangle| \approx 6.5 * 10^{-4} \ll \frac{1}{M} = 0.01. \tag{6.2}$$

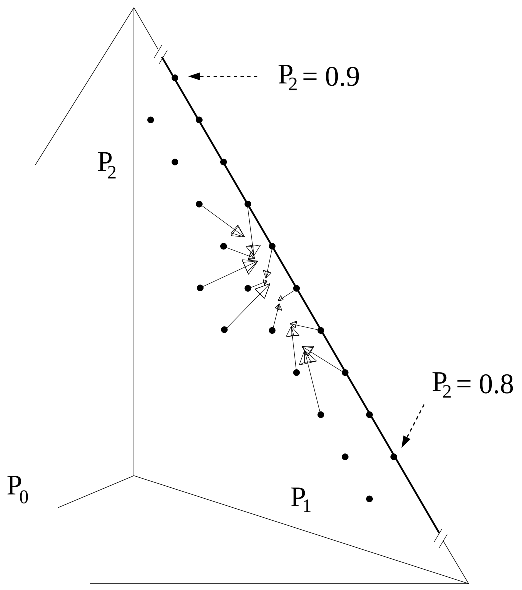

Figure 5: Face view of the flow $\langle d\vec{P}\rangle$ within the surface $P_3 = 0$. The flow of fitness distributions centers around an area where $P_2 \approx 0.85$, $P_1 \approx 0.14$ and $P_0 \approx 0.01$. This region is shown substantially enlarged. Parameters are set to $N = 3$, $K = 4$, $q = 0.01$, and $M = 100$. The dots indicate the allowed finite populations on a portion of $\Lambda_M$.

Since $\langle dP_3\rangle$ is so small[7] compared to $1/M$, we expect the distribution to stabilize itself in that region for some time. Figures 5 and 6 illustrate this behavior. Figure 5 shows the flow at points in the plane $P_3 = 0$ as arrows whose length is the value of $|\langle d\vec{P}\rangle|$. We see how the flow stabilizes the dynamics in an area on the surface where $\vec{P} = (P_0, P_1, P_2, P_3) \approx (0.01, 0.14, 0.85, 0.0)$. Figure 6 shows a side view, perpendicular to the $P_3 = 0$ plane. The arrows show the expected change in the component $dP_3$ off the $P_3 = 0$ surface. We see that the flow off the surface is so small that the dynamics is likely to remain on that surface for some time. This region on the $P_3 = 0$ surface is a likely place for a fitness epoch to occur.

We would like to have a way to determine *where* metastable regions like the one shown in figure 6 are to be found in the state space. In section 5.5 we saw that all eigenvectors of the operator $\tilde{\mathbf{G}}$ correspond to fixed points of $\mathbf{G}$. It is therefore natural to assume that the metastable regions of the finite population dynamics are to be found in the vicinities of the eigenvectors $\vec{V}^i$. We saw, in addition, that the average fitness in the population in the neighborhood of the eigenvectors $\vec{V}^i$ is equal to the eigenvalues $g_i$ of the operator $\tilde{\mathbf{G}}$. We therefore expect the population to have mean fitness close to $g_i$ during fitness epoch $i$.

### 6.2 Predicted Epoch Levels

This section assesses under what circumstances we can predict the mean fitness of the population during an epoch and lists additional aspects of the empirical GA behavior that will be explained in subsequent sections.

Figures 7(a) through 7(i) show the same runs as those in figures 1(a) through 1(i) together with the theoretical predictions of the fitness levels (horizontal lines) at which we expect the

---

[7] The expression for $\langle dP_3\rangle$ can be obtained by plugging the epoch center $\vec{P} = (0.01, 0.14, 0.85, 0.0)$ into $\langle d\vec{P}\rangle = \mathbf{G}(\vec{P}) - \vec{P}$. Epoch centers are defined in section 6.3.

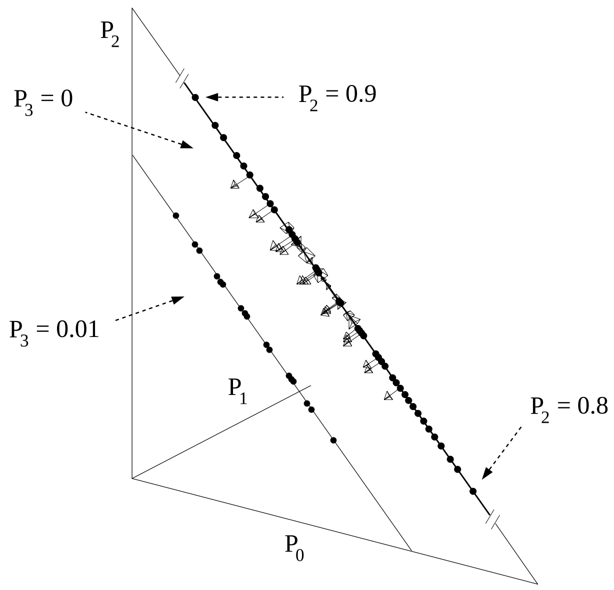

Figure 6: Side view of the simplex perpendicular to the surfaces $P_3 = 0$ and $P_3 = 0.01$, which are shown as the upper and lower diagonal solid lines, respectively. The dots indicate the allowed finite populations in $\Lambda_M$. The expected change $\langle dP_3 \rangle$ (arrows) is so small that the dynamics is likely to stay on the surface $P_3 = 0$ for some time. Note that the arrows emanating from the $P_3 = 0$ surface have been magnified five times to make them visible. Parameters are $N = 3$, $K = 4$, $q = 0.01$, and $M = 100$.

epochs to occur. The levels were calculated numerically by determining the eigenvalues $g_i$ of $\tilde{\mathbf{G}}$ for the different parameter settings of each run. Recall that the fitness levels are not a function of the population size $M$; they are determined by the infinite population dynamics.

Caption for figure 7: Average fitness over time for nine runs of the Royal Road GA with different parameter settings together with the theoretical predictions (horizontal lines) for the epoch levels. See figures 1(a) through 1(i) for the corresponding parameter settings for each run. Although there are $N$ possible fitness levels in each case, for the sake of clarity only subsets are shown. Figures 7(a) and 7(d) show epoch levels $g_3$ through $g_{10}$ and figure 7(g) shows levels $g_4$ through $g_{10}$. Note that the theoretical levels are the same for these three runs since the epoch levels don't depend on population size. Figure 7(b) shows predicted levels $g_9$ through $g_{20}$. Only the asymptotic epoch $g_{20}$ can be seen to occur in this run. The same epoch levels $g_9$ through $g_{20}$ are shown in figure 7(c). In run 7(c) epochs can be seen to occur at $g_{14}$, $g_{17}$, and $g_{20}$. Figure 7(e) shows epoch levels $g_5$ through $g_{10}$. No epochs can be seen to occur for these parameter settings. Figure 7(f) shows the last two of the epoch levels $g_9$ and $g_{10}$ as dashed lines. The solid lines in 7(f) show the upper bounds on these epoch levels (see text for discussion). Epoch levels $g_3$ through $g_{10}$ are shown in figure 7(h) in which no clear epochs can be distinguished. Although, one can see time intervals for which the average fitness fluctuates around $g_5$. Finally, figure 7(i) shows all epoch levels $g_1$ through $g_4$. In all of these figures, the epoch levels $g_i$ were obtained by numerically solving for the eigenvalues of $\tilde{\mathbf{G}}$ for each of the parameter settings. In section 6.4 we calculate simple analytical expressions for these epoch levels.

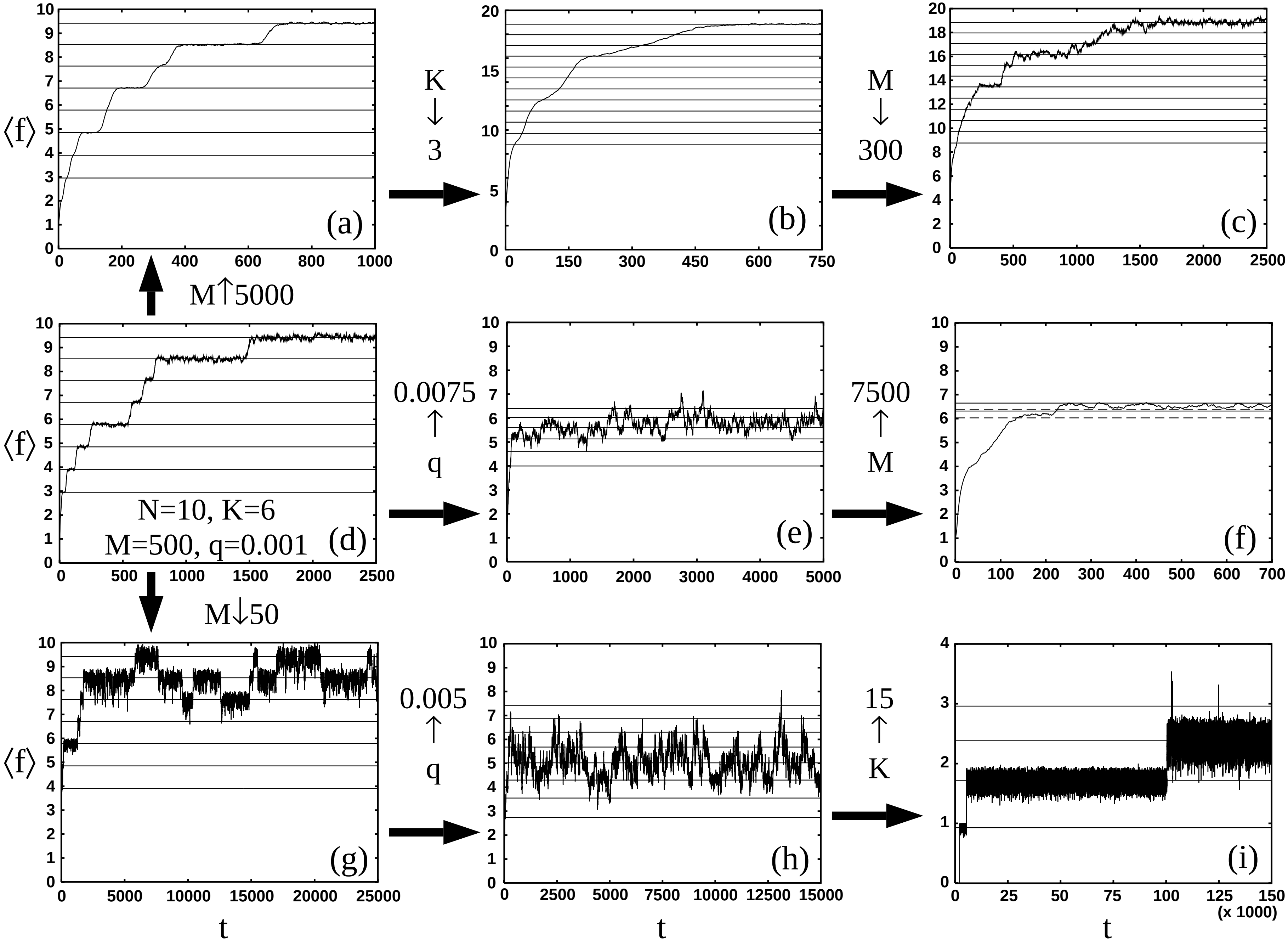

Figure 7: For caption see text.

In runs 7(a), 7(d), and 7(g) (first column of figure 7), the theory correctly predicts the fitness levels at which epochs occur. The variation of epoch durations across fitness levels and across the runs, sizes of fitness fluctuations in the epochs, and the intermittent behavior of run 7(g) need to be explained.

For the runs in the middle column (figures 7(b), 7(e), and 7(h)) we have plotted the theoretical fitness level predictions in a band of fitness values around which most of the behavior takes place. The theoretical predictions of the fitness levels correspond in only a limited way to the GA's behavior for these parameter settings. In run 7(b) the predicted highest level matches the empirical asymptotic fitness level. This is, of course, in accord with the analysis. For large populations the behavior approaches the infinite population dynamics we described in section 5; epochs do not occur. But we must explain why the epochs appear in run 7(a) but disappear in run 7(b) as a result of decreasing the block size $K$ to 3.

For run 7(e) we have plotted only the predicted fitness levels $g_5$ through $g_{10}$. The low value of $g_{10}$ explains why the average fitness stays low throughout the run. We must explain why the average fitness does not stabilize onto the highest fitness level $g_{10}$, but instead fluctuates in a band between $g_6$ and $g_{10}$.

In run 7(h) we have plotted the theoretical fitness levels of epochs 3 to 10. Again, the low value of $g_{10}$ explains the apparent ceiling on the average fitness. We must explain why the average fitness does not stabilize onto distribution $V^{10}$ at fitness $g_{10}$ and why the fluctuation amplitudes are larger than those in 7(e).

The theoretical epoch fitnesses plotted in run 7(c) correctly predict the levels of the epochs that can be distinguished in this case as well as the asymptotic fitness level.

In run 7(f), however, it turns out that we see for the first time the consequences of the breakdown of our approximation of the probability $A$ to create a block. The dashed lines show the predicted last two epoch levels using the analysis of the foregoing sections. They clearly underestimate the levels at which the epochs occur. Recall that in calculating the probability $A$ to align a block we assumed that all unaligned blocks have never been aligned before. This approximation breaks down for the parameters used in this case. With this run's high mutation rate, $q = 0.0075$, many blocks in $\mathcal{P}_h$ are destroyed through mutation. Therefore, the blocks in the fitness class just below $\mathcal{P}_h$ have a relatively large proportion of their bits set. This means that the probability to align a block is much higher for these strings than we assume it to be. Our prediction should thus be seen as a lower bound. In section 5.1 we obtained an upper bound of $A = q(1-q)^{K-1}$ by assuming all blocks have $K-1$ of the $K$ bits set. We plotted the upper bounds for the $9^{th}$ and $10^{th}$ epochs as solid lines in run 7(f). These upper and lower bounds still give quite accurate predictions for the observed fitness levels.

Finally, in run 7(i) the theoretical values correctly predict the epoch levels. The large fitness fluctuations within the epochs of this run remain to be explained.

To understand some of the remaining phenomena, we next analyze in more detail the structure of the epochal dynamics in the simplex.

### 6.3 Epochal Dynamics as the Evolutionary Unfolding of the State Space

We saw earlier that of all $\tilde{\mathbf{G}}$'s eigenvectors $\vec{V}^i$, only the principal eigenvector $\vec{V}^N$ is positive definite and thus interpretable as a fitness distribution. This means that the nonprincipal eigenvectors are not points in the simplex $\Lambda_\infty$. The fitness epochs could only occur at fitness distributions that are *close* but not at the eigenvectors $\vec{V}^i$. Let the term "epoch center" refer to the average position in $\Lambda_M$ of the fitness distribution during an epoch. We now present a way to exactly obtain the epoch centers for each epoch.

Intuitively, the $N$ epochs come about because there are $N$ different possible blocks to align. In the first epoch, some individuals have one aligned block (fitness 1) and some have no aligned blocks (fitness 0) but none has more than one block aligned. In the second epoch, some individuals have two aligned blocks, some have one, and some have none. Whenever a new block is aligned and spreads through the population, the current fitness epoch becomes unstable and the population moves up to the next epoch. In terms of the dynamics in the simplex, the population starts on a low-dimensional subset of the simplex and finds its way to higher dimensions step by step. Initially (for small populations), there are typically only individuals with fitness 0 or 1. The fitness distribution then stabilizes somewhere on the line $P_0 + P_1 = 1, P_i \geq 0$. After a new, second aligned block is discovered the population moves off that line and onto the plane $P_0 + P_1 + P_2 = 1, P_i \geq 0$ and stabilizes there. When the third block is found the population moves into the three-dimensional space $P_0 + P_1 + P_2 + P_3 = 1, P_i \geq 0$. The evolution proceeds in this incremental fashion until all blocks are discovered or, as we will discuss later on, until the population cannot stabilize within successively higher-dimensional subspaces.

This rough picture illustrates how the subsimplices unfold dimension by dimension through the evolutionary search. When the population is in the $n^{th}$ epoch it will move into the $(n+1)$-dimensional subsimplex with

$$\sum_{i=0}^{n} P_i = 1, P_i \geq 0. \tag{6.3}$$

This points to a way of calculating the actual epoch centers. While a population is in the $n^{th}$ epoch, the selection and mutation operators, by definition, act only on strings $s$ with $f(s) \leq n$ and produce new strings $s'$ with $f(s') \leq n$. We can find the dynamics of the GA on this subspace by projecting $\mathbf{G}$ onto this subspace to form a restricted operator $\mathbf{G}^n$:

$$\mathbf{G}^n_{ij} = \begin{cases} \mathbf{G}_{ij} & \text{if } i, j \leq n \\ 0 & \text{otherwise.} \end{cases} \tag{6.4}$$

By acting on the fitness distributions with this restricted operator we can find the flow of the population in the $n$-dimensional subspace of the full simplex. The center $\vec{P}^n$ of the $n^{th}$ fitness epoch is given by the principal eigenvector $\vec{P}^n$ of the restricted linearized operator $\tilde{\mathbf{G}}^n$. That is,

$$\tilde{\mathbf{G}}^n \cdot \vec{P}^n = e_n \vec{P}^n, \tag{6.5}$$

where $e_n$ is $\tilde{\mathbf{G}}^n$'s principal eigenvalue. In short, by restricting $\mathbf{G}$ to the subsimplex of dimension $n$ we can obtain the exact center of the $n^{th}$ fitness epoch.

## 6.4 Eigensystem of the Restricted Generation Operators

We will now derive expressions for the different epoch centers $\vec{P}^n$ and the associated average fitness $f_n$ in each epoch for small mutation rates $q$. For notational simplicity we will refer to the restricted operators $\tilde{\mathbf{G}}^n$ as $\tilde{\mathbf{G}}$ when calculating $e_n$ and $\vec{P}^n$ in the following.

We first expand $\tilde{\mathbf{G}}$ to first order in $q$. Starting from the definition of $\mathbf{G}$ in terms of $\mathbf{M}$ and $\mathbf{S}$, we obtain

$$\tilde{\mathbf{G}}_{ij} = j\left[\delta_{ij}\left(1 - q(A_1(N-j) + Kj)\right) + \delta_{(i-1)j}A_1(N-j)q + \delta_{(i+1)j}Kjq\right] + \mathcal{O}(q^2). \tag{6.6}$$

where $A_1 q$ is the probability $A$ to align a block to first-order in $q$ (from equation 5.3):

$$A = A_1 q + \mathcal{O}(q^2) = \frac{K}{2^K - 1}q + \mathcal{O}(q^2) \tag{6.7}$$

Formally, we can split equation 6.6 into a term that is zeroth order in $q$ and a term that is first order in $q$:

$$\tilde{\mathbf{G}} = \tilde{\mathbf{H}}^{\mathbf{0}} + q\tilde{\mathbf{H}}^{\mathbf{1}} + \mathcal{O}(q^2), \tag{6.8}$$

where $\tilde{\mathbf{H}}^{\mathbf{0}}_{ij} = j\delta_{ij}$. In the limit $q \to 0$, $e_n = n$ and $P^n_i = \delta_{in}$; that is, all strings have fitness $n$ during the $n^{th}$ epoch in this limit. For nonzero $q$ we can expand $\vec{P}^n$ and $e_n$ to first order in $q$:

$$P^n_i \equiv \delta_{in} + q\Delta^1_i \tag{6.9}$$

and

$$e_n = n + qe^1_n. \tag{6.10}$$

Using the eigenvalue equation

$$\tilde{\mathbf{G}} \cdot \vec{P}^n = e_n \vec{P}^n, \tag{6.11}$$

and equating coefficients of equal powers in $q$, we obtain the $n^{th}$ and $(n-1)^{st}$ components of $\vec{P}^n$:

$$P^n_n = (1 - n^2Kq) \text{ and } P^n_{n-1} = n^2Kq. \tag{6.12}$$

All other components are 0 to first order in $q$. For $e_n$ we find

$$e_n = n - \left[n^2K + n(N-n)A_1\right]q. \tag{6.13}$$

For the average fitness

$$f_n = \sum_{i=0}^{n} iP^n_i, \tag{6.14}$$

we obtain

$$f_n = n - n^2Kq. \tag{6.15}$$

Notably, the epoch fitness levels can also be approximated in a more straightforward way that gives an insightful result, if we assume that the number of fitness $n$ strings generated by block-aligning mutations of lower fitness strings is negligible. This assumes that the proportion $P^n_n$ of strings in the highest fitness class is kept constant by a balance between block destroying mutations and selection. We then simply find that under selection and mutation:

$$P^n_n \to \frac{n}{f_n}P^n_n \to P^n_n \frac{n}{f_n}(1-q)^{nK}, \tag{6.16}$$

where the last factor $(1-q)^{nK}$ arises because only the strings that do not mutate the $nK$ bits in their aligned blocks remain in fitness class $n$. From the above equation it follows that:

$$f_n = n(1-q)^{nK}. \tag{6.17}$$

This equation nicely shows that the average fitness in epoch $n$ is proportional to the probability of all $(nK)$ defining bits replicating without mutations. Note that equation 6.15 follows immediately to first order in $q$.

While the population resides in a fitness epoch there is a (metastable) equilibrium distribution of the population over the different fitness classes. We see that for very small $q$ almost all individuals in the population are in the current highest fitness class, which is $n$ for the $n^{th}$ epoch. The higher the mutation rate, the more individuals will have lower fitness and accordingly the average fitness of the population decreases as $q$ increases. Also, for higher epochs (larger $n$) the proportion of individuals in the current highest fitness class decreases with the square of $n$. This is also the case for the asymptotic fitness distribution $\vec{P}^N = \vec{V}^N$. The main difference between the metastable fitness distributions and the asymptotic distribution $\vec{V}^N$ is that the epoch distributions can become unstable when a new block is found and spreads through the population. Until this happens there is a metastable equilibrium between the effects of mutation and selection on the strings.

We noted earlier that the asymptotic distribution $\vec{V}^N$ is the phenotypic analogue of a quasispecies of replicating molecules or genomes. In light of the results above it is natural to extend this notion of quasispecies to include *metastable* (phenotypic) quasispecies that occur at earlier fitness epochs. Since in general the state space of molecular or genetic sequences is vastly larger than any realistic population size, we also expect finite-population-induced fitness epochs to occur in molecular evolution. During the evolution of a population there are large proportions of the sequence space that the population has never visited which might contain higher fitness strings than the current fittest genotypes. As already pointed out by Eigen and his colleagues [12, 32], the population will therefore reside in a metastable quasispecies distribution (or "metaspecies") until a mutation discovers one of the higher fitness strings. In general, though, the mechanism commonly proposed for metastability, e.g., [32], is that of a metastable quasispecies that is *localized* on a fitness peak in sequence space and that remains stable until one of the mutants at the edge of the quasispecies cloud finds a higher fitness peak. In this view, metastability is caused by local optima in the fitness landscape. The population will remain stable until one of the mutants crosses the fitness valley to a higher peak. In other words, the population has to cross a "fitness barrier" towards a higher peak.

The fitness-barrier mechanism is in sharp contrast to the mechanism presented here. There are no local optima in Royal Road fitness landscape. Rather, there are large connected subspaces of strings with equal fitness. Let $S_n$ denote the subspace of strings with fitness $n$. The metastable distributions $\vec{P}^n$ are *not* strictly localized in sequence space but diffuse around randomly in the subspace $S_n$—a "subbasin" of the transient evolutionary dynamics—until one of the individuals discovers a rare "portal" to the subspace $S_{n+1}$ by aligning a new block.[8] Therefore, the metastability in this view is the result of an "entropic barrier" that has to be crossed by the population. As a consequence, the evolutionary dynamics naturally splits into a neutral and an adaptive regime. The population spends most of its time randomly diffusing in the subbasin, or "neutral network", $S_n$ of strings with equal fitness until one mutant finds one of the rare portal connections to $S_{n+1}$, after which the population adapts by moving into this new subbasin. In [25] this entropic view of evolution was also advocated in the context of molecular evolution.

---

[8] The nested-simplex state-space architecture also describes the transient behavior of diffusive annihilating particles, from which we have borrowed the subbasin and portal terminology [6]. In the evolutionary setting the simplices unfold; in the particle case they collapse, due to annihilation.

### 6.5 Crossover

So far we have restricted our attention to a mutation-only GA while, in fact, most applications of GAs employ some form of crossover—an operator that combines portions of parental chromosomes to produce offspring. It might come as somewhat of a surprise, then, that the epoch levels for the GA including crossover are empirically found to be the same as the epoch levels for the GA without crossover. Figure 8 shows the results of a GA run with parameters $N = 10$, $K = 6$, $M = 100$, and $q = 0.001$ and single-point crossover probability 1.0, together with the theoretical predictions of the epoch levels $f_8$, $f_9$, and $f_{10}$. Pairs of parents are selected in proportion to fitness from the old population and are always crossed over at a single point to produce two strings for the new generation. From figure 8 it is clear that the same theoretical predictions for the epoch levels of the GA without crossover also correctly predict the epoch levels for the GA with crossover. This can be understood in the following way.

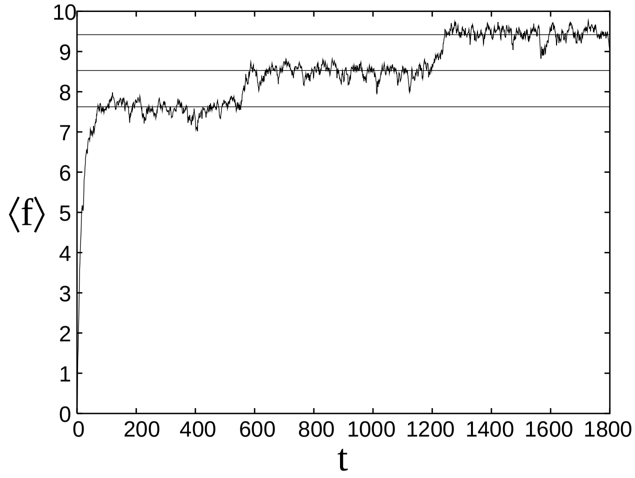

Figure 8: A run of the GA with single-point crossover, displaying the average fitness (solid curve) as a function of generation, together with the theoretically predicted epoch fitness levels (solid lines) for the $8^{th}$, $9^{th}$, and $10^{th}$ epochs. The parameters for this run are $N = 10$, $K = 6$, $q = 0.001$, and $M = 100$. Cf. figures 1(d) and 7(d).

When the population resides in the $n^{th}$ epoch all individuals in the population have fitness $n$ or less. An innovation occurs when one or more individuals with fitness higher than $n$ are discovered. It is unlikely that more than one such individual is discovered in the same generation. Therefore, it is almost always the case that one individual is the founder of a new epoch. When the population has moved up to epoch $n + 1$ all the strings with fitness $n + 1$ in the population are descendants of that founder string and thus share its aligned blocks (as well as other bits). This means that the $n + 1$ aligned blocks in each string with fitness $n + 1$ will be in corresponding positions in each string. If the population is dominated during the epoch by strings of fitness $n + 1$, crossover on pairs of selected strings will generally not break up the aligned blocks. Blocks that are destroyed are mostly destroyed through mutations, *not* through crossover. Because of this, the balance between block destruction and selection is hardly affected and the epochs occur at the same fitness levels as in the case without crossover.

However, at the very beginning of a run, the population will not be converged and aligned blocks will be in different positions along the strings. Therefore, in contrast to the mutation-only GA, crossover *can* combine different blocks from different strings at the start of the run to create higher fitness strings. Hence, the number of initial epochs expressed in runs with crossover is smaller than the number of epochs in the runs without crossover. Once crossover has put together the best string that can be produced from the aligned blocks in the initial population, the population will converge onto copies of that string. From that point on crossover will act as a "localized" mutation operator. It will introduce mixing only in the bits of the unaligned blocks, since all the bits in the aligned blocks are shared by the strings in the population.

This illustrates again that phrases such as "crossover combines building blocks" or "crossover introduces nonlocal mixing", are meaningless without precise specification of *all* other components of the dynamics. For crossover to be able to perform such actions as referred to above, it is necessary that *different* useful building blocks be present in the population at the same time and that generally the population is genetically diverse. We have argued that very early in the run, due to the randomness of the initial population, crossover can indeed combine building blocks. However, as soon as the epochal behavior sets in, all aligned blocks occur in corresponding positions in the strings and it is highly unlikely that different building blocks will occur in the population at the simultaneously ever again.

A similar argument holds for the nonlocal mixing that crossover is purported to introduce. For the random initial population this will certainly be the case, but as soon as the epochal dynamics sets in crossover can only introduce nonlocal mixing in the unaligned blocks. This nonlocal mixing is furthermore restricted by the diversity that *mutation* has introduced in bits of the unaligned blocks. In summary, for systems that show mainly epochal evolution, it seems unlikely that crossover's "combining building blocks" or "introducing nonlocal mixing" are important contributions to the search dynamics.

### 6.6 Stable and Unstable Manifolds

It is clear from figure 1 that not all of the $N$ possible fitness epochs are visited for a given GA run; later epochs tend to be visited in more runs than earlier epochs. Later epochs also tend to have longer durations than earlier epochs in a given run. In addition, for higher mutation rates, epochs appear less distinct and innovations less steep. Since, for small mutation rate $q$, the metastable fitness distributions in which a finite population can get trapped are close to the fixed points of $\mathbf{G}$, we can obtain a qualitative understanding of these features by analyzing the local stability of these fixed points.

We will analyze the topological structure in $\Lambda_\infty$ that determines the global stability of these metastable states by looking at the local stability of the fixed points themselves. The local stability around the fixed points is determined by the Jacobian matrix $\mathbf{DG}$ at each fixed point, since it gives the first-order approximation of the dynamics in the vicinity of the fixed points. That is, around a fixed point $\vec{V}$ we have:

$$\mathbf{G}(\vec{V} + \vec{\epsilon}) = \vec{V} + \mathbf{DG} \cdot \vec{\epsilon}, \tag{6.18}$$

where $\vec{\epsilon}$ is a small deviation vector. Consider the Jacobian matrix at the fixed point $\vec{V}^n$ of the $n^{th}$ fitness epoch. Using the basic definitions, $\tilde{\mathbf{G}} = \mathbf{M} \cdot \tilde{\mathbf{S}}$, and the fact that $\vec{V}^n$ is a fixed point $\tilde{\mathbf{G}} \cdot \vec{V}^n = g_n \vec{V}^n$, we find that

$$\mathbf{DG}_{ij}(\vec{V}^n) = \left[ \frac{\partial \mathbf{G}_i(\vec{P})}{\partial P_j} \right]_{\vec{P}=\vec{V}^n} = \frac{\tilde{\mathbf{G}}_{ij} - j V_i^n}{g_n}. \tag{6.19}$$

The local stable and unstable manifolds of this fixed point can be obtained by solving for the eigenvectors of this matrix. Eigenvectors with eigenvalues $\lambda_i > 1$ give the directions in which this fixed point is unstable and eigenvectors with eigenvalues $\lambda_i < 1$ give the directions in which the fixed point is stable. This follows from equation 6.18, since the deviations grow and shrink for eigenvalues greater or less than 1, respectively. By inspection of equation 6.19, it is easy to see that the eigenvectors $\vec{U}^i$ of the Jacobian matrix $\mathbf{DG}(\vec{V}^n)$ are given by the differences between the eigenvectors $\vec{V}^i$ and the eigenvector $\vec{V}^n$:

$$\vec{U}^i = \vec{V}^i - \vec{V}^n. \tag{6.20}$$

Substituting these vectors into the eigenvalue equation, we find that

$$\mathbf{DG}(\vec{V}^n) \cdot \vec{U}^i = \frac{g_i}{g_n}\, \vec{U}^i. \tag{6.21}$$

Thus, the eigenvalues $\lambda_i^n$ of this fixed point are given by[9]

$$\lambda_i^n = \frac{g_i}{g_n}. \tag{6.22}$$

This means that, since the $g_i$ are ordered in increasing magnitude, the fixed point $\vec{V}^i$ has $i$ stable directions and $N - i$ unstable directions. This is intuitively clear from the fact that in the $i^{th}$ epoch the subsimplex that is already *discovered* by the population is $i$-dimensional (corresponding to the $i$ aligned blocks) and the undiscovered part of the simplex is $(N - i)$-dimensional (corresponding to the $N-i$ yet-to-be-aligned blocks). Furthermore, we see that the strength of the Jacobian eigenvalues $\lambda_i^n$ is determined by the ratios of the different generation operator eigenvalues $g_i$ with respect to the eigenvalue $g_n$ of the epoch under consideration. As we have seen, to a high approximation the generation operator eigenvalues are equal to the average fitnesses in the different epochs. So we see that the eigenvalues of the Jacobian matrix are simply determined by the ratios of average fitnesses in the different epochs.

Thus, for small mutation rate $q$ we can use equation 6.17 to obtain analytical expressions for the eigenvalues of $\mathbf{DG}$ in the $n^{th}$ fitness epoch. These are

$$\lambda_i^n = \frac{i}{n}(1-q)^{(i-n)K}\ , 0 \leq i \leq N\ . \tag{6.23}$$

The relative sizes of these eigenvalues as a function of $i$, $n$, $K$, and $q$ control the dynamical features of the different fitness epochs.

Now we can qualitatively explain the first of our observations, namely, that the higher fitness epochs are longer. The population will remain in an epoch until it finds, and spreads into, one of the undiscovered dimensions of the simplex. Since the later epochs have more stable dimensions and fewer unstable ones, the population is less likely to find one of the unstable dimensions and the epochs will be longer. This is related to the simple fact that it is easier for mutations to align one or more new blocks when there are more blocks still unaligned.

We have also observed that later epochs typically appear in more runs than earlier epochs. The first and obvious reason for this is that the initial population is likely to contain strings

---

[9] $\vec{U}^n$ is the null-vector which, of course, has eigenvalue 0. It does not, however, give an independent direction in the simplex.

with different fitness values and that the first epoch $n$ that will be visited corresponds to the fitness $n$ of the highest fitness strings in the initial population. Let $\Pr(f < n)$ denote the probability that *none* of the $M$ individuals of the initial population $\vec{P}(0)$ has a fitness of $n$ or higher:

$$\Pr(f < n) = \Pr(P_i = 0,\ i \geq n). \tag{6.24}$$

We then have

$$\Pr(f < n) = \left[\sum_{i=0}^{n-1} \binom{N}{i} 2^{-Ki} \left(1 - 2^{-K}\right)^{N-i}\right]^M . \tag{6.25}$$

It turns out that as $n$ increases this probability jumps up sharply from almost 0 to almost 1 at some value of $n$ (provided that $K$ is not too small). If $n_f$ is the first value for which $\Pr(f < n) \approx 1$ then it is likely that the first epoch to appear will be the $(n_f - 1)^{th}$ epoch, since strings with fitness $n_f - 1$ are likely to occur in the initial population and strings with fitness $n_f$ are very unlikely to occur in the initial population. Furthermore, when one or more new blocks are discovered in epoch $n$, there are $N - n$ possible unstable dimensions into which the population can move out of the subsimplex. Each of these dimensions corresponds to a portal through which the next epoch can be visited. From this it follows that higher epochs with a larger number of attracting dimensions are more likely to be visited.

Next we consider the steepness of the innovations. The steepness of an innovation out of epoch $n$ is related to the size of the eigenvalues $\lambda_i^n$ for the unstable dimensions $i > n$. The larger these eigenvalues, the more quickly the population will move away from the metastable region once an unstable dimension is explored. From equation 6.23 we see that for larger $q$ the eigenvalues $\lambda_i^n$, given by the fitness ratios $f_i/f_n$, approach 1. This means that the innovations become less steep for larger values of $q$, which is indeed what we have observed.

From the fact that the eigenvalues $\lambda_i^n$ approach 1 for larger $q$, it also follows that the fitness fluctuations in an epoch increase with $q$. This is because the smaller the stable-direction eigenvalues $\{\lambda_i^n : i < n\}$, the more strongly the population is restored to the epoch center after a fluctuation. So if these eigenvalues increase toward 1, the fitness fluctuations increase as well. From the fact that the stable-direction eigenvalues $\{\lambda_i^n : i < n\}$ are larger and so closer to 1 for higher $n$, we also see that later epochs have larger fitness fluctuations than early ones (e.g., see figure 1(d)). We will discuss the size of the fitness fluctuations in more detail in section 6.8 below.

Finally, we see that the unstable-direction eigenvalues $\{\lambda_i^n : i > n\}$ for large $n$ are smaller than those for small $n$. From this, it follows that later innovations are, in general, less steep than early ones (e.g., see figures 1(a) and 1(d)). Following up this line of qualitative analysis, in the next section we estimate the innovation times quantitatively from the values of $\lambda_i^n$.

Recapitulating, we have argued that the main qualitative features of epochal evolution can be understood in terms of the eigenvalues $\lambda_i^n$ of the Jacobian matrix in the vicinity of the $n^{th}$ epoch. In the following sections we will use these eigenvalues to make more quantitative predictions about the epochal behavior.

### 6.7 Innovation Durations

We now estimate the expected time for an innovation from the $n^{th}$ to the $(n+1)^{st}$ epoch to take place. During the $n^{th}$ epoch the proportion of individuals with fitness $n+1$ is 0. When some individual finds a new aligned block it will either start spreading through the population or it will be lost through a sampling fluctuation. (See figure 1 for examples of this loss.) We are interested in the time it takes a string with the new aligned block to fully spread through the population *once* the string appears and continues to spread through the population. (For convenience in this section we take the time of this first appearance to be $t = 0$.) Thus, we will assume that the effects of the finite population sampling noise can be neglected during the innovation.

Thus, we refer to the time at which the epoch $n+1$ begins as $t = 0$. At this time, the number of individuals with fitness $n+1$ is 1; that is, $P_{n+1}(0) = 1/M$. From that time on, $P_{n+1}$ increases until it reaches the equilibrium value $P^{n+1}_{n+1}$ of the $(n+1)^{st}$ epoch. In going from the $n^{th}$ to the $(n+1)^{st}$ epoch, the population moves in the direction of the Jacobian eigenvector $\vec{U}^{n+1} = \vec{P}^{n+1} - \vec{P}^{n}$. Thus, during the innovation from the $n^{th}$ to the $(n+1)^{st}$ epoch, we can write the fitness distribution $\vec{P}(t)$ of the population as a linear combination of the $n^{th}$ epoch eigenvector $\vec{P}^{n}$ and the $(n+1)^{st}$ epoch eigenvector $\vec{P}^{n+1}$:

$$\vec{P}(t) = (1-\alpha(t))\vec{P}^{n} + \alpha(t)\vec{P}^{n+1}, \tag{6.26}$$

with the initial condition $\alpha(0) = 1/M$. One should interpret $\alpha$ as giving the decomposition of the fitness distribution during the innovation in terms of the lower and higher epoch distributions. During the innovation the balance shifts from the lower epoch $\alpha \approx 0$ to the higher epoch $\alpha \approx 1$.

Applying the generation operator to the fitness distribution in equation 6.26 and using the fact that the epochs correspond to eigenvectors of the generation operator, we find for the evolution of $\alpha(t)$ that

$$\alpha(t+1) = \frac{f_{n+1}}{f_{n+1}\alpha(t) + (1-\alpha(t))f_n}\alpha(t). \tag{6.27}$$

Assuming that $\alpha(t)$ changes slowly and smoothly as a function of time, we can deduce a differential equation from the above equation:

$$\frac{d\alpha}{dt} \approx \alpha(t+1) - \alpha(t) = \gamma_n \alpha \frac{1-\alpha}{\gamma_n \alpha + 1}, \tag{6.28}$$

where

$$\gamma_n = \frac{f_{n+1} - f_n}{f_n} \tag{6.29}$$

is the relative increase in fitness from the $n^{th}$ to the $(n+1)^{st}$ epoch. It is possible to solve equation 6.28 analytically to obtain $t$ as a function of $\alpha$:

$$t = \frac{1}{\gamma_n}\left[\log(M\alpha) - (1+\gamma_n)\log\frac{M(1-\alpha)}{M-1}\right], \tag{6.30}$$

where we have used the boundary condition $\alpha(0) = 1/M$. In general it is impossible to invert equation 6.30 analytically to obtain $\alpha$ as a function of $t$. Nonetheless, we can use equation 6.30 to obtain an estimate of the duration $t_n$ of the innovation from the $n^{th}$ to the $(n+1)^{st}$ epoch. We consider the innovation to have ended once $\alpha$ has reached the level $1 - 1/M$. Since we treat $\alpha$ as a continuous variable, it will approach 1 only asymptotically. For the finite population case we truncate this continuous dynamics at $\alpha = 1 - 1/M$ and solve for $t_n$, finding that

$$t_n = \frac{2+\gamma_n}{\gamma_n}\log[M-1]. \tag{6.31}$$

This equation has a simple interpretation. $\gamma_n$ gives the relative fitness increase from the $n^{th}$ to the $(n+1)^{st}$ epoch. The innovation duration is roughly inversely proportional to this relative fitness increase. The innovation duration, in addition, is proportional to the logarithm of the population size. The logarithm of the population size apparently controls the "inertia" of the distribution with respect to innovations. Certainly, a single more-fit string will take longer to dominate a larger population. And since replication is exponential in time, one obtains a logarithmic dependence on $M$.

Expanding $\gamma_n$ to first order in the mutation rate $q$ and setting $\log[M-1] \approx \log[M]$ we find that

$$t_n = [1 + 2n + 2Kn(n+1)q] \log M + \mathcal{O}(q^2). \tag{6.32}$$

We see that the innovation duration is proportional to the epoch level and the logarithm of the population size. Furthermore, it is clear that increasing the mutation rate makes the innovations less steep.

We can also approximate $\alpha(t)$ from equation 6.28 by neglecting the small term $\gamma_n \alpha$ in the denominator.[10] Equation 6.28 then turns into the well known logistic growth equation, which can be solved analytically:

$$\alpha(t) = \frac{\exp(\gamma_n t)}{\exp(\gamma_n t) + M - 1}. \tag{6.33}$$

Again, the speed of the innovation is controlled by the relative fitness increase $\gamma_n$ from the $n^{th}$ to the $(n+1)^{st}$ epoch. The average fitness $\langle f(t) \rangle$ during the innovation is proportional to $\alpha(t)$:

$$\langle f(t) \rangle = f_n + (f_{n+1} - f_n)\alpha(t). \tag{6.34}$$

Figures 9(a) and 9(b) plot the theoretical predictions for the average fitness (thick lines) during the innovations between the third and fourth and eighth and ninth epochs, respectively, for the parameter settings $N = 10$, $K = 6$, $q = 0.001$, and $M = 500$. The thin lines in figures 9(a) and 9(b) give some examples of the empirically observed innovations from runs of the GA. Figure 9(a) shows the innovation from just one run with these parameter settings. It is clear that the logistic growth approximation of the innovation gives an accurate prediction of the shape and length of this innovation. Figure 9(b) shows the innovation between the eight and ninth epoch from three different runs. At higher epochs there are large fitness fluctuations[11] and the exact shape of the innovations differs from run to run. Still the theory accurately predicts the *average* shape of the innovation as can be seen from figure 9(b). The predicted innovation durations in the above cases are $t_3 \approx 32$ and $t_8 \approx 98$, respectively.

### 6.8 Fitness Fluctuations

The finite population dynamics in and around the epochs is, of course, essentially stochastic and cannot be modeled by deterministic dynamical equations. From here on, we therefore have to turn to stochastic equations describing the evolution of a probability distribution $\Pr[\vec{\epsilon}, t]$ which gives the probability that the population is a distance $\vec{\epsilon}$ away from some epoch center $\vec{P}^n$ at time $t$. We will model the evolution of $\Pr[\vec{\epsilon}, t]$ by means of a diffusion (or Fokker-Planck) equation, where the *average* change per generation of $\vec{\epsilon}$ is determined by the Jacobian

---

[10] Since $\gamma_n \approx 1/n$ and $\alpha < 1$ this approximation is justified, especially for later epochs and for the small $\alpha$ at the start of the innovation.

[11] This will be discussed in the next section.

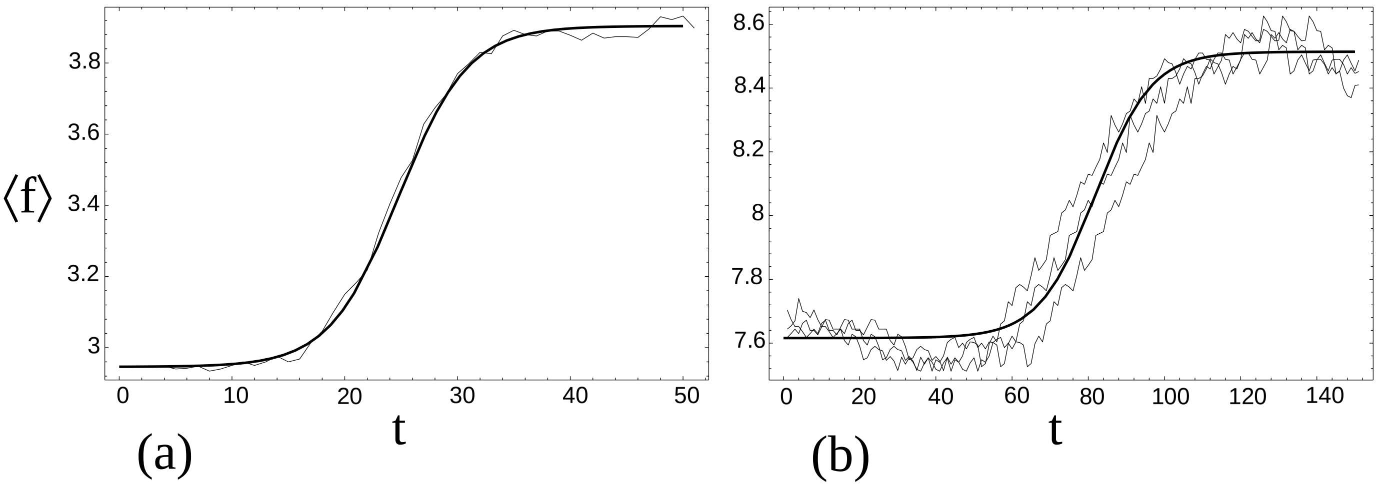

Figure 9: Innovation between the third and fourth (a) and eight and ninth (b) epochs for GA runs with parameters $N = 10$, $K = 6$, $q = 0.001$, and $M = 500$. (See figure 1(d).) The thick lines give the theoretical predictions of the innovation curves. The thinner lines show examples of these innovations from runs of the GA.

eigenvalues $\lambda_i^n$ of epoch $n$ and the fluctuations due to the finite population sampling occur as a Gaussian-noise diffusion. The idea of using continuous-variable diffusion models, like the Fokker-Planck equation, to solve for the stochastic dynamics of gene frequencies in a population was developed by the population geneticist Kimura, see [27]. This type of stochastic model assumes that gene frequencies, such as $P_i$ or $\epsilon_i$, can be approximated by real variables and that the gene frequencies change slowly over a single time step. (See [22] for a recent overview of the validity of the diffusion models in this context.)

Using the average fitness levels in each epoch, we can now estimate the size of the fluctuations about those levels. During a fitness epoch the population jumps around on a set of different lattice points in $\Lambda_M$ surrounding the epoch's center $\vec{P}^n$. From the Jacobian matrix $\mathbf{DG}(\vec{P}^n)$ we can derive the form of the flow in the vicinity of the fixed point. The difference vectors $\vec{U}^i = \vec{P}^i - \vec{P}^n$ between the different epoch locations and the location of the epoch of interest form a natural basis for modeling the fluctuations around the epoch center. We will therefore expand the fluctuation dynamics in terms of the Jacobian eigenbasis.

First, assume that the population fitness distribution $\vec{P}$ resides near the epoch center $\vec{P}^n$:

$$\vec{P} = \vec{P}^n + \sum_{i=0}^{n-1} \epsilon_i \vec{U}^i, \tag{6.35}$$

where $\vec{\epsilon} = (\epsilon_0, \ldots, \epsilon_{n-1})$ is a small deviation vector. In one time step, the vector $\vec{P}$ is expected to go to $\langle \vec{P}' \rangle$, which is given by

$$\langle \vec{P}' \rangle = \mathbf{G}(\vec{P}) = \vec{P}^n + \sum_{i=0}^{n-1} \frac{f_i}{f_n} \epsilon_i \vec{U}^i. \tag{6.36}$$

That is, the deviation $\vec{\epsilon}$ is expected to scale down by a factor of $f_i/f_n$ in each direction $\vec{U}^i$. From the above equation we can now calculate the expected change $\langle d\epsilon_i \rangle$ of the deviation in direction $\vec{U}^i$:

$$\langle d\epsilon_i \rangle = \left( \frac{f_i}{f_n} - 1 \right) \epsilon_i. \tag{6.37}$$

To simplify notation, we define the relative fitness decrease of epoch $i$ with respect to epoch $n$:

$$\mu_i = \frac{f_n - f_i}{f_n}. \tag{6.38}$$

Then equation 6.37 takes on a simple form:

$$\langle d\epsilon_i \rangle = -\mu_i \epsilon_i. \tag{6.39}$$

Thus, fluctuations in direction $\vec{U}^i$ die off exponentially *on average* with rate $\mu_i$. Note that this rate is smallest in the direction of the higher-fitness epochs.

Equation 6.39 gives the *expected* change of $\epsilon_i$ around the epoch center $\vec{P}^n$. Of course, there will be fluctuations in this change due to finite-size sampling. In equation 6.1 we saw that the probability for going from state $\vec{P}$ to state $\vec{P}'$ is a multinomial with mean $\mathbf{G}(\vec{P})$. From this we can derive expressions for the expected second moments of the change in the fitness distribution:

$$\langle dP_i dP_j \rangle = \frac{P_i(\delta_{ij} - P_j)}{M}. \tag{6.40}$$

From this we can derive the second moments of the change in the fluctuation vector $\langle d\epsilon_i d\epsilon_j \rangle$. To this end we transform the components $P_i$ to the basis of the epoch centers $\vec{P}^i$. We first define the similarity transformation matrix $\mathbf{R}_{ij} = P_i^j$. Using its inverse, we can calculate the change of fluctuations in direction $\vec{U}^i$:

$$\epsilon_i = \sum_{i=0}^{n-1} \mathbf{R}_{ij}^{-1} P_j. \tag{6.41}$$

Note that the sum goes from 0 to $n-1$ and not up to $n$, since the component $\epsilon_n$ is not independent of the others. (It is determined from the requirement that the fitness distribution vector be normalized.) Using equations 6.40 and 6.41 the second moments of the change in the fluctuations are given by

$$\langle d\epsilon_i d\epsilon_j \rangle = \sum_{k,m=0}^{n-1} \mathbf{R}_{ik}^{-1} \mathbf{R}_{jm}^{-1} \frac{P_k(\delta_{km} - P_m)}{M}. \tag{6.42}$$

Assuming that the fluctuations $\epsilon_i$ are small compared to the epoch center components $\vec{P}_i^n$, we can approximate $\vec{P}_i$ by $\vec{P}_i^n$ in the above formula and find

$$\langle d\epsilon_i d\epsilon_j \rangle = \sum_{k,m=0}^{n-1} \mathbf{R}_{ik}^{-1} \mathbf{R}_{jm}^{-1} \frac{P_k^n(\delta_{km} - P_m^n)}{M} \equiv \frac{\mathbf{B}_{ij}}{M}, \tag{6.43}$$

where the components $\mathbf{B}_{ij}$ depend only on the location of the current epoch center $\vec{P}^n$.

So that we can use the Jacobian matrix approximation to the dynamics, we assume that the population size is large enough to keep the fluctuations localized in the area around the epoch center. Then we can use equations 6.39 and 6.43 to approximate the stable limit distribution $\Pr[\vec{\epsilon}\,]$ of the fluctuations that occur while the population resides in epoch $n$. The distribution $\Pr[\vec{\epsilon}, t]$ gives the probability of finding the population at a deviation $\vec{\epsilon}$ from the epoch center at any particular time $t$ during the epoch. This distribution can be obtained by solving the multivariate Fokker-Planck equation associated with the drift term of equation 6.39 and the diffusion term given by equation 6.43. This is

$$\frac{\partial \Pr[\vec{\epsilon}, t]}{\partial t} = -\sum_i \frac{\partial \langle d\epsilon_i \rangle \Pr[\vec{\epsilon}, t]}{\partial \epsilon_i} + \frac{1}{2} \sum_{i,j} \frac{\partial^2 \langle d\epsilon_i d\epsilon_j \rangle \Pr[\vec{\epsilon}, t]}{\partial \epsilon_i \partial \epsilon_j}. \tag{6.44}$$

As $t \to \infty$, the asymptotic (stationary) solution of the above equation is a multi-dimensional Gaussian peak around $\vec{\epsilon} = 0$ for the case of constant $\langle d\epsilon_i d\epsilon_j \rangle$ [43]. This is given by

$$\Pr[\vec{\epsilon}\,] = \frac{1}{\sqrt{(2\pi)^n Det[\mathbf{C}]}} \exp\left[ -\frac{1}{2} \sum_{i,j=0}^{n-1} \epsilon_i \mathbf{C}^{-1}_{ij} \epsilon_j \right], \tag{6.45}$$

where the matrix $\mathbf{C}$ determines the second moments of the distribution. It is given by

$$\mathbf{C}_{ij} = \langle \epsilon_i \epsilon_j \rangle = \frac{\mathbf{B}_{ij}}{M(\mu_i + \mu_j)}. \tag{6.46}$$

Of course, the means are all 0:

$$\langle \epsilon_i \rangle = 0. \tag{6.47}$$

Using equation 6.45 we can solve for the expected fitness fluctuations during an epoch. With some algebra we find that the fitness variance $\mathrm{Var}[f]$ in epoch $n$ is given to first order in $q$ by

$$\mathrm{Var}[f] = \frac{K n^3 q}{2M}. \tag{6.48}$$

This shows rather transparently that the fluctuations scale inversely with the population size $M$ and proportionally to the cube of the epoch number $n$.

Figures 10, 11, and 12 show the above analytical prediction of the fitness fluctuations during various epochs. The grey bands show the average fitness in the epochs plus and minus two standard deviations, given by $\sigma_n = \sqrt{\mathrm{Var}[f_n]}$. Figure 10 shows the fitness fluctuations for our canonical parameter settings $N = 10$, $K = 6$, $q = 0.001$, and $M = 500$. Figure 11 shows the fitness fluctuations for the same parameter settings, only with an increased mutation rate of $q = 0.002$. By comparing these two figures, it is clear that increased mutation rate increases the size of the fitness fluctuations during the epochs. As predicted by the analysis, the fitness fluctuations increase by a factor of roughly $\sqrt{2}$. Finally, figure 12 has the same parameter settings as figure 10, except for a lower population size of $M = 50$. As predicted, the fitness fluctuations increase by a factor of approximately $\sqrt{10}$.

### 6.9 Destabilizing Fluctuations and the Error Threshold

We saw that for high-fitness epochs the proportion of individuals in the highest fitness class decreases with the square of the epoch number $n$:

$$P^n_n = (1 - n^2 K q) + \mathcal{O}(q^2). \tag{6.49}$$

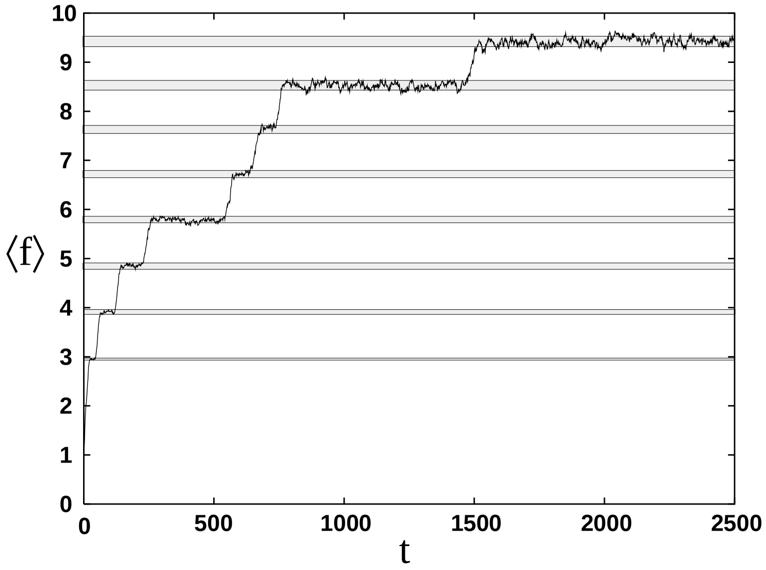

Figure 10: Predicted size of the epoch fitness fluctuations to two standard deviations, given by the grey bands. The parameters for this run are $N = 10$, $K = 6$, $q = 0.001$, and $M = 500$. This is the same run as was plotted in figure 1(d).

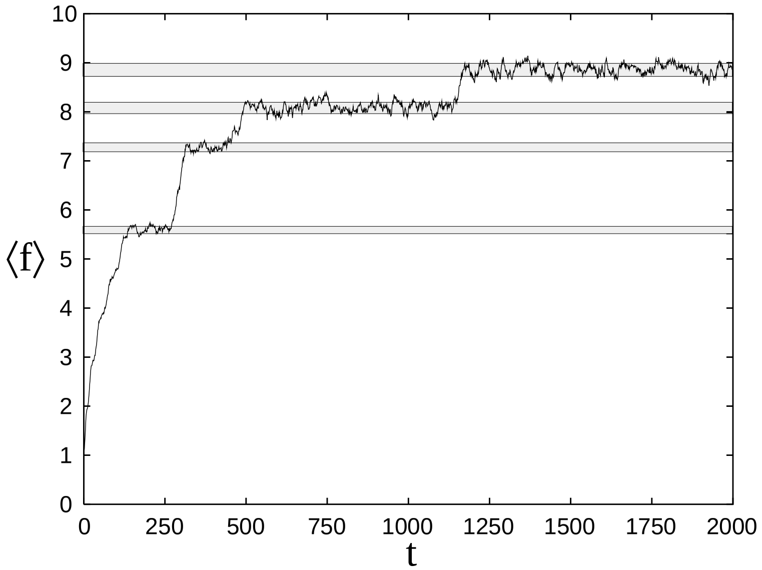

Figure 11: Predicted size of the fitness fluctuations during the epochs, plotted up to two standard deviations using the grey bands. The parameters for this run are $N = 10$, $K = 6$, $q = 0.002$, and $M = 500$. Cf. figure 10.

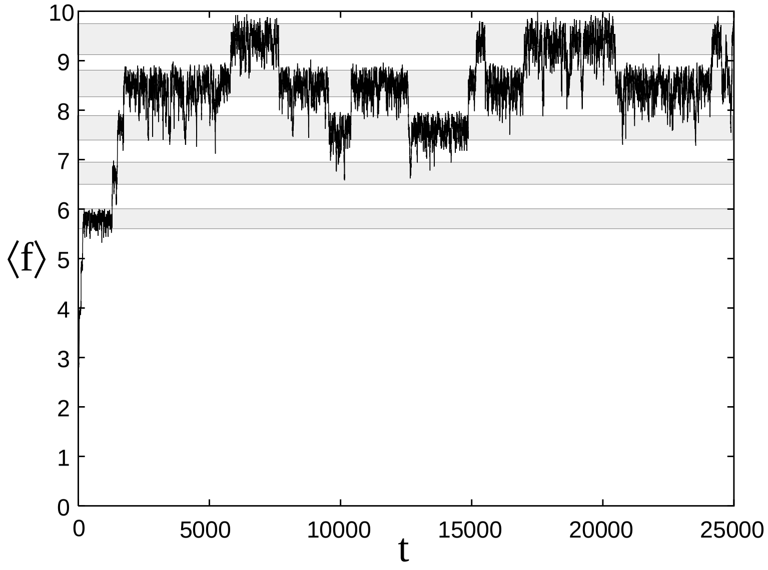

Figure 12: Predicted size of the fitness fluctuations during the epochs, plotted up to two standard deviations using the grey bands. The parameters for this run are $N = 10$, $K = 6$, $q = 0.001$, and $M = 50$. This is the same run as was plotted in figure 1(g).)

For high-fitness epochs, in the large $N$ case when there are many possible epochs, the proportion $P_n^n$ of strings in the highest fitness class can eventually become so small that there is an appreciable chance that all individuals in the highest fitness class will be lost through a sampling fluctuation. When this happens, the fitness distribution will fall back to the distribution $\vec{P}^{n-1}$ of epoch $n-1$ just below. If, after some time, the $n^{th}$ block is rediscovered and spreads through the population, the distribution will move up again to the $n^{th}$ epoch. This is exactly the process that causes the intermittent epochal behavior seen in figure 1(g).

We can obtain a rough magnitude estimate for the epoch at which one begins to observe this intermittency. Assume that the population resides at the epoch center. In generating the population for the next time step, each string has a probability $P_n^n$ to be of fitness class $n$. Therefore, using equation 6.12, the probability $\Pr(f \neq n)$ that none of the strings will be in class $n$ is given by

$$\Pr(f \neq n) = (1 - P_n^n)^M \approx (n^2 K q)^M. \qquad (6.50)$$

Demanding that this probability is of the order of (say) 1%, we can calculate for which epoch level $n$ this condition is satisfied as a function of $K$, $q$, and $M$. For the case of $N = 10$, $K = 6$, $q = 0.001$, and $M = 50$, we find $n \approx 12$. From figure 12, we see that by epochs 9 and 10 the intermittency has set in. The crude estimate given above apparently underestimates the fluctuations in the proportion of individuals in the highest fitness class $P_n$. Using an analysis similar to that in the last section, we will now calculate more precisely the average time —the "destabilization" time— that the population spends in epoch $n$ until all individuals with fitness $n$ are lost through a sampling fluctuation.

To calculate the average destabilization time, we consider the dynamics of the fluctuations $\epsilon_n$ in the component $P_n$ of the fitness distribution near the epoch center. That is, let

$$P_n = P_n^n + \epsilon_n. \qquad (6.51)$$

The deviation $\epsilon_n$ will fluctuate up and down while the population resides in the epoch. It can become at most $\epsilon_n = 1 - P^n_n$, corresponding to all strings fitness $n$, and when it becomes $-P^n_n$ all individuals with fitness $n$ are lost and the epoch becomes unstable. We therefore want to calculate the average time it takes until $\epsilon_n = -P^n_n$ for the first time.

The deviation $\epsilon_n$ just compensates the sum of the deviations $\epsilon_i$ in the eigenvector directions $\vec{U}^i$ of the Jacobian[12]. That is,

$$\epsilon_n = -P^n_n \sum_{i=0}^{n-1} \epsilon_i. \tag{6.52}$$

After one time step the component $P_n$ will on average move to

$$\langle P'_n \rangle = P^n_n - P^n_n \sum_{i=0}^{n-1} \frac{f_i}{f_n} \epsilon_i \equiv P^n_n + \lambda \epsilon_n, \tag{6.53}$$

where we have defined $\lambda$ as the factor by which the fluctuation $\epsilon_n$ is scaled down. Of course, in general $\lambda$ depends on the particular distribution of the fluctuations $\epsilon_i$ over the different directions. In the previous section we saw that the fluctuations $\epsilon_i$ during an epoch are all approximately normally distributed around the epoch center. As an approximation, we will assume that the fluctuation in direction $i$ is proportional to the variance $\langle \epsilon_i^2 \rangle$. We then obtain

$$\lambda = \frac{\sum_{i=0}^{n-1} f_i \langle \epsilon_i^2 \rangle}{f_n \sum_{j=0}^{n-1} \langle \epsilon_j^2 \rangle}. \tag{6.54}$$

With the above definition of the average scale factor, we can obtain the expected change in the fluctuation:

$$\langle d\epsilon_n \rangle = -(1-\lambda)\epsilon_n \equiv -\mu \epsilon_n, \tag{6.55}$$

where the second equality defines the coefficient $\mu$. We can again approximate the expected change in the square of the deviation by the size of the sampling fluctuations at the epoch center:

$$\langle (d\epsilon_n)^2 \rangle = \langle (dP_n)^2 \rangle \approx \frac{P^n_n (1 - P^n_n)}{M}. \tag{6.56}$$

Since the above (diffusion) term is again a constant, the problem reduces to that of the first passage time of a homogeneously diffusing particle in a potential field (see [21]). The solution for the average time $T(0)$ for the fluctuation to reach $\epsilon_n = -P^n_n$ for the first time, given that the process starts with a fluctuation $\epsilon = 0$, is given by

$$T(0) = \frac{M P^n_n}{1 - P^n_n} + \frac{\pi}{2\mu} \mathrm{erfi}\left[ \sqrt{\frac{M\mu P^n_n}{1 - P^n_n}} \right] \mathrm{erf}\left[ \sqrt{\frac{M\mu (1 - P^n_n)}{P^n_n}} \right], \tag{6.57}$$

where $\mathrm{erf}(x)$ is the error function and $\mathrm{erfi}(x) = \mathrm{erf}(ix)/i$ is the imaginary error function. Similar waiting time distributions for evolutionary processes were derived by Kimura [29].

Table 1 shows the average times $T_n(0)$, starting from the epoch center $\epsilon = 0$, for some of the epochs of figure 1(g) with $N = 10$, $K = 6$, $q = 0.001$, and $M = 50$. Comparing these average destabilization times with figure 1(g) we see that they give reasonable predictions of the average epoch stability times. The above numbers should be seen as an "order of magnitude" estimate. They are rather sensitive to the exact value of $P^n_n$ and since our calculation of $P^n_n$ essentially involves approximations, so do the above destabilization times. They do, however,

| $\mathbf{T_6(0)}$ | $\mathbf{T_7(0)}$ | $\mathbf{T_8(0)}$ | $\mathbf{T_9(0)}$ | $\mathbf{T_{10}(0)}$ |
|---|---|---|---|---|
| $3 \times 10^{13}$ | $4.8 \times 10^8$ | $9.9 \times 10^5$ | $2.5 \times 10^3$ | 2600 |

Table 1: Average destabilization times $T_n(0)$ for some epochs ($n = 6 - 10$) in the Royal Road GA with $N = 10$, $K = 6$, $q = 0.001$, and $M = 50$. These were the parameters used for the run plotted in figure 1(g).

| $\mathbf{T_3(0)}$ | $\mathbf{T_4(0)}$ | $\mathbf{T_5(0)}$ | $\mathbf{T_6(0)}$ |
|---|---|---|---|
| $1.7 \times 10^{22}$ | $5.5 \times 10^8$ | $21 \times 10^3$ | 375 |
| $\mathbf{T_7(0)}$ | $\mathbf{T_8(0)}$ | $\mathbf{T_9(0)}$ | $\mathbf{T_{10}(0)}$ |
| 65 | 28 | 15 | 9 |

Table 2: Average destabilization times $T_n(0)$ for some epochs ($n = 3 - 10$) in the Royal Road GA with $N = 10$, $K = 6$, $q = 0.005$, and $M = 50$. These were the parameters used for the run plotted in figure 1(h).

nicely explain the occurrence of the intermittent behavior seen around epochs 9 and 10 in this run.

Table 2 shows the average destabilization times for some of the epochs of figure 1(h), which has an increased mutation rate of $q = 0.005$. These predictions demonstrate why the fitness of the population gets trapped in a band that is set by the average fitnesses of the fifth and eighth epochs. Epoch 8 destabilizes so quickly that the population has almost no chance to find a ninth aligned block. Even if it could, it would not have time to stabilize on that epoch, since the innovations duration itself is longer than the destabilization time. The destabilization times are so short compared to innovation durations in this case that epochs are very hard to distinguish, if they can be distinguished at all. For these parameters the dynamics is almost completely governed by the fluctuations of the finite population sampling. Selection and mutation cannot stabilize the population against these sampling fluctuations, though they do set the bounds on this range. We find $f_5 - 2\sigma_5 \approx 4$ and $f_8 + 2\sigma_8 \approx 7$ which exactly matches the band within which the fitness fluctuates in figure 1(h). For fitnesses below 4 selection pushes the population up against the sampling fluctuations. For fitnesses above 7 the mutations push the population down against the sampling fluctuations. The same mechanism is at work in figure 1(e). Here, too, the size of the fluctuation band is explained by the above analysis. For $N = 10$, $K = 6$, $q = 0.0075$, and $M = 500$ (run 1(e)) we find that epoch 7 is still stable for $48 \times 10^3$ time steps, epoch 8 for about 226 time steps, epoch 9 for 35 time steps, and finally the tenth epoch is stable only for about 13 time steps on average. We find for the fluctuation band $f_7 - 2\sigma_7 \approx 4.7$ and $f_9 + 2\sigma_9 \approx 7.1$ which again explains the data from figure 1(e).

Destabilization is very closely related to the so-called "error threshold" of self-replicating molecules in the theory of molecular evolution [42]. It was found that when the size of the genome and the selection pressure are kept constant and the mutation rate is increased, there is a sharp transition from a regime where the most fit genotype is always in the population to a regime where it will almost always be lost. This transition point is referred to as the "error threshold". In a complementary way, for a fixed mutation rate, there will be an equivalent "size threshold" on increasing genome length since mutations *anywhere* in the genome will reduce its fitness. Above a certain critical size the mutations will out-compete selection and the fittest

---

[12] Note that $U_n^i = -P_n^n$ since $P_n^i = 0$ for all $i < n$.

genotype will be lost.

This is analogous to what happens in the Royal Road GA. Under constant mutation rate there is a certain upper limit on the number of aligned blocks that selection can keep in the population. In the above cases this threshold occurred around 9 and around 5 blocks for runs 1(g) and 1(h), respectively. In the region of the critical number of blocks this leads to intermittent behavior in the average fitness. The population "hops" between different epochs when either the highest fitness string is lost through a fluctuation or a new block is (re)discovered and spreads through the population. When the intermittency time becomes shorter than the innovation durations, epochal behavior disappears, and the population seems to fluctuate randomly in a wide band that encompasses several epoch levels.

We see that there is a functional genome-length $L_n = nK$ versus mutation rate threshold in this GA. Our analysis also establishes a population size $M$ versus mutation rate $q$ error threshold. This suggests that there is a critical error-threshold surface in the three dimensional phase space spanned by the parameters $M$, $q$, and $L_n$.

## 7. Epoch Durations

We will now turn to the most important feature of Royal Road GA behavior that remains to be addressed—namely, the average length of the fitness epochs. Until now, almost all behavioral features could be understood in terms of the epoch fitness levels $f_n$ and the epoch centers $\vec{P}^n$. We will investigate to what extent the same analysis can be used to predict the average epoch durations.

The ending of an epoch has two phases. First, a string has to be created of higher fitness than currently exists in the population. Thus, if the population resides in epoch $n$, a string of at least fitness $n+1$ has to be created by mutation. Second, this string has to be able to spread through the population if the population is to leave epoch $n$. Especially for higher-fitness epochs, where the relative fitness increase of new strings with respect to the old strings becomes small (i.e., proportional to $1/n$), it is likely that the best string will be lost through a sampling fluctuation before it gets a chance to spread through the population as was first shown by Fisher [18].

### 7.1 Creation of a Higher Fitness String

We will first calculate the probability that a string with fitness $n+1$ or higher is created while the population resides in the $n^{th}$ epoch. During the $n^{th}$ epoch the fitness distribution is given by $\vec{P}^n$ on average and the population resides in the $n$-dimensional subsimplex. The probability for the population to remain in the subsimplex over one generation is the chance that all individuals of the new generation have fitness smaller or equal to $n$. When an individual is selected and mutated for the new generation, the probability $\Pr[in]$ that it will have a fitness $i \leq n$ and thus remain within the subsimplex is given by:

$$\Pr[in] = \sum_{i=0}^{n} \mathbf{G}_i(\vec{P}^n) = \sum_{i=0}^{n} \mathbf{G}_i^n(\vec{P}^n) = \sum_i \frac{e_n}{f_n} \vec{P}_i^n = \frac{e_n}{f_n} \ . \tag{7.1}$$

The first equality follows from the fact that, given the epoch distribution $\vec{P}^n$, the probability to create a string of fitness $i$ is determined by the $i$th component of the generation operator acting on $\vec{P}^n$. The second equality notes that by restricting ourselves to $i \leq n$, the component $i$ of the generation operator acting on $\vec{P}^n$ is equal to the restricted operator component $\mathbf{G}_i^n$. The third equality uses the fact that $\vec{P}^n$ is an eigenvector of the linearized restricted operator $\tilde{\mathbf{G}}^n$ with eigenvalue $e_n$. The final equality uses the fact that $\vec{P}^n$ is normalized to one.

The probability that *all* $M$ individuals remain in the subsimplex is given by $\Pr[in]^M$. The probability $\Pr[out]$ that one or more individuals have jumped out of the subsimplex, by creating a string with fitness greater than $n$, is

$$\Pr[out] = 1 - \left[\frac{e_n}{f_n}\right]^M. \tag{7.2}$$

We will assume that the population resides at the epoch center and therefore has the same probability $\Pr[out]$ at each time step to jump out of the epoch. The expected number of time steps $\tau_n$ until the population jumps out of the epoch $n$ subsimplex is then given by

$$\tau_n = \frac{1}{Pr[out]} = \frac{f_n^M}{f_n^M - e_n^M}. \tag{7.3}$$

For small $q$ or large $K$, the probability $A$ to align a block is small and we can expand to leading order in $A$ to find:

$$\tau_n = \frac{1}{M(N-n)A}. \tag{7.4}$$

Thus, the expected number of generations $\tau_n$ to jump out of epoch $n$ is inversely proportional to the probability $A$ to align a block, the population size $M$, and the number of unaligned blocks $N-n$. We now investigate the probability that the new higher-fitness string will spread through the population.

### 7.2 Takeover of the Population by a Higher Fitness String

When a string of fitness $(n+1)$ is created, the initial proportion $P_{n+1}$ of such strings is $1/M$.

Using the results from section 6.7 we see that the expected change $\langle dP_{n+1}\rangle$ per time step is given by

$$\langle dP_{n+1}\rangle = \frac{f_{n+1} - f_n}{f_n} P_{n+1} = \gamma_n P_{n+1}, \tag{7.5}$$

for small $P_{n+1}$. The second moment of the change $dP_{n+1}$ is given by the sampling fluctuations:

$$\langle (dP_{n+1})^2\rangle = \frac{P_{n+1}(1-P_{n+1})}{M}. \tag{7.6}$$

Assuming that the change per time step is dominated by these first two moments, we can solve for the probability $\pi(p)$ that the new higher-fitness string will spread through the population and eventually reach proportion $P_{n+1}^{n+1}$, given that it initially has proportion $p$. To solve for $\pi(p)$ we use the backward Fokker-Planck equation:

$$\frac{\partial \pi(p,t)}{\partial t} = \langle dP_{n+1}\rangle \frac{\partial \pi(p,t)}{\partial p} + \frac{\langle (dP_{n+1})^2\rangle}{2} \frac{\partial^2 \pi(p,t)}{\partial p^2}, \tag{7.7}$$

where $\pi(p,t)$ is the probability that the higher fitness string will have reached $P^{n+1}_{n+1}$ by time $t$. The probability $\pi(p)$ that the mutant will spread, is given by the limit of $\pi(p,t)$ as $t$ goes to infinity. This calculation was first done by Kimura in 1962 [26] in the context of the drift of the frequency of a certain genotype in a population. The solution is

$$\pi(p) = \frac{\int_0^p G(x)dx}{\int_0^{P^{n+1}_{n+1}} G(x)dx}, \tag{7.8}$$

where, for our case, the function $G(x)$ is given by[13]

$$G(x) = (1-x)^{2M\gamma_n}. \tag{7.9}$$

Performing the integral, we obtain

$$\pi_n \equiv \pi(\frac{1}{M}) = \frac{1-\left(1-\frac{1}{M}\right)^{2M\gamma_n+1}}{1-\left(1-P^{n+1}_{n+1}\right)^{2M\gamma_n+1}} \approx 1 - e^{-2\gamma_n}, \tag{7.10}$$

where we have set the initial proportion $p = 1/M$. The approximation on the right-hand side holds only for large population sizes. Equation 7.10 tells us that the population has to find a better string $1/\pi_n$ times on average before it finally moves from the $n^{th}$ to the $(n+1)^{st}$ epoch. Therefore, the total average time $T_n$ the population spends in epoch $n$ is

$$T_n = \frac{\tau_n}{\pi_n} = \frac{f_n^M}{(f_n^M - e_n^M)(1-\exp(-2\gamma_n))}. \tag{7.11}$$

For small $q$ or large $K$ this becomes

$$T_n = \frac{1}{M(N-n)A\left[1-\exp(-2/n)\right]}. \tag{7.12}$$

As we found in section 6.8, the fitness fluctuates around $f_n$ during the $n^{th}$ epoch in an approximately Gaussian way with the standard deviation given by

$$\sigma_n = \sqrt{\frac{Kn^3q}{2M}}. \tag{7.13}$$

We thus find the average number of time steps $T_n(f)$ that the population has fitness $f$ during epoch $n$ is given by

$$T_n(f) = \frac{T_n}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{1}{2}\left(\frac{f-f_n}{\sigma_n}\right)^2\right]. \tag{7.14}$$

We performed experiments to test these theoretical predictions by accumulating histograms of the average number of time steps $\langle T\rangle$ that the population has fitness $f$ during a run, averaged over a large number of runs. Figure 13 shows the results of such an experiment for the parameter setting $N = 10$, $K = 6$, $M = 500$, and $q = 0.001$ of run 1(d) together with the theoretical predictions for the peaks at these parameter settings. The inset plot shows a magnification of

---

[13] This function is an essential quantity in the diffusion equation method. It's obtained by taking the exponential of the integral of the ratio $\langle dx\rangle/\langle dx^2\rangle$.

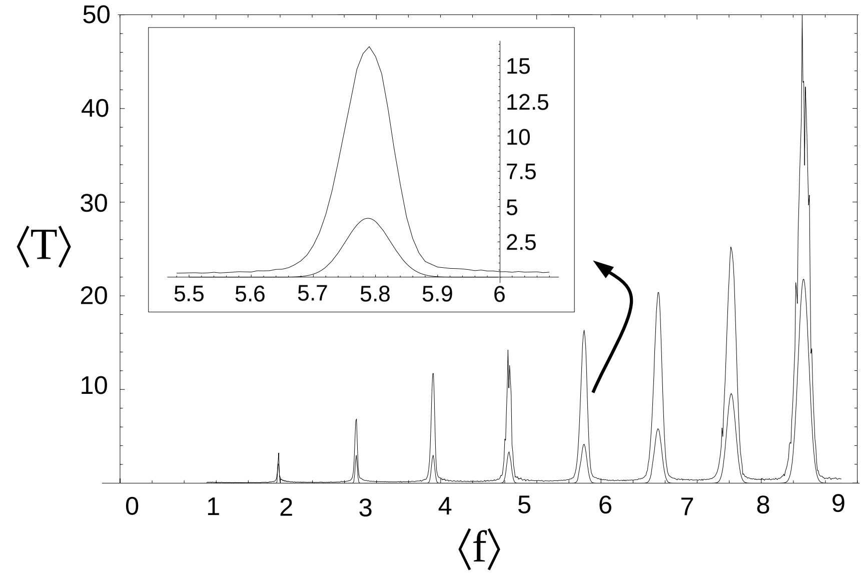

Figure 13: Empirical (upper curve) and theoretical (lower curve) fitness histograms. The horizontal axis shows the average fitness and the vertical axis shows the average number of time steps the population has that average fitness during a GA run, averaged over 500 runs. The parameters for this run are $N = 10$, $K = 6$, $q = 0.001$, and $M = 500$. The inset plot shows a magnification of the peak at the $6^{th}$ epoch.

the peak at the $6^{th}$ epoch. The upper curves are from the experiment and the lower curves plot the theoretical predictions (equation 7.14). As can be clearly seen from the figure the theory substantially underestimates the average lengths of the epochs found at this parameter setting. As shown before, the widths and locations of the peaks are correctly predicted by the theory. The empirically observed averages are offset vertically from the theoretical predictions outside of the peak region since the theoretical curve does not take into account the time the population spends in the innovations. This does not account, however, for the fact that the predicted peaks are about a factor of 6 too small.

Figure 14 shows the results for the parameter setting $N = 20$, $K = 3$, $M = 300$, and $q = 0.001$ of run 1(c) together with the theoretical predictions of the epoch duration peaks for these parameter settings. The inset plot shows a magnification of the peaks around the $16^{th}$, $17^{th}$, and $18^{th}$ epochs. Again it is clear that the theory underestimates the epoch durations; here it does so by a factor of approximately 3. Nonetheless, it is instructive to compare the above histogram with run 1(c). From that figure it is very hard to say how many epochs occur and what their durations and exact locations are. The fitness histogram of figure 14 clearly shows that the epochs are *still there* in the GA dynamics. They are reflected in the peaks in the plots. With this observation in mind, we propose to *define* the existence of an epoch in the dynamics by the appearance of a peak in the average-time histogram. In the empirical curve in figure 14 at least 10 peaks can be counted. It would have been very hard to get such a clear view of the epochal behavior from plots such as those in figure 1(c). Averaging the average fitness at each time step from a large number of runs like that in figure 1(c) only makes detecting epochs more difficult. Since the onset of each epoch shifts in time from run to run, averaging the fitness values at each time step over many runs completely washes out any trace of epochal dynamics.

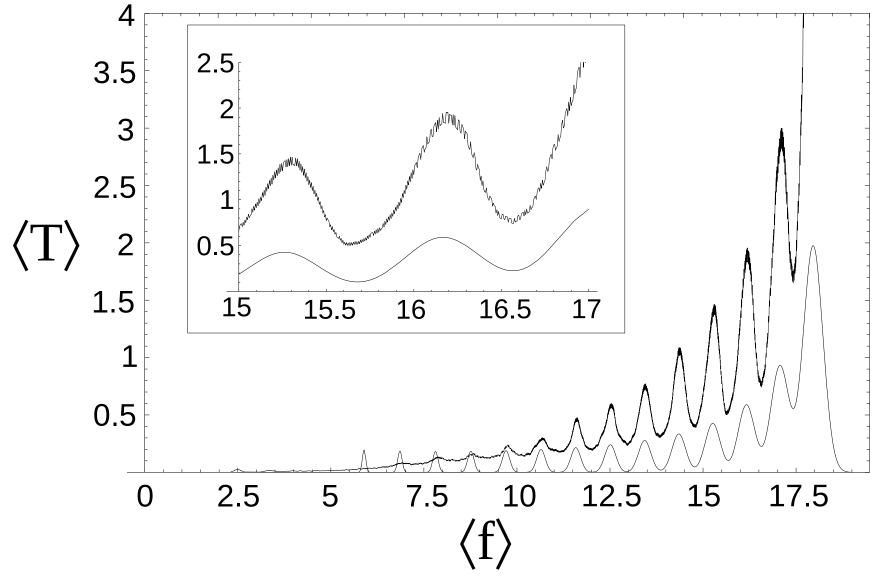

Figure 14: Empirical (upper curve) and theoretical (lower curve) fitness histograms. The horizontal axis shows the average fitness and the vertical axis shows the average number of time steps the population has that average fitness during the run, averaged over 500 runs. The parameters for this run are $N = 20$, $K = 3$, $q = 0.001$, and $M = 300$. The inset plot shows a magnification of the peaks at the $16^{th}$, $17^{th}$, and $18^{th}$ epochs.

We also see from figure 14 that the distributions associated with each epoch start overlapping at these parameters. This is, of course, the reason why the epochs are hardly discernible in figure 1(c). In obtaining the theoretical curve for figure 14 we summed the contributions of the peaks corresponding to the $6^{th}$ through $19^{th}$ epochs. It is clear from the figure that the peaks of the $6^{th}$, $7^{th}$ and $8^{th}$ epochs do not actually occur in the behavior. The reason for this is that in calculating the peaks for the epochs we assumed that the population starts in the epoch center. In the actual behavior seen in the run the population never reaches these epoch centers.

Why do the theoretical predictions of the epoch durations markedly underestimate the actual epoch lengths in the behavior? As noted in section 6.5 a new epoch is almost always founded by a single individual. This means that at the start of the epoch all strings of the current highest fitness class are essentially the same, the population is highly converged. It is obvious in this case that assuming all unaligned blocks are statistically independent breaks down, because all unaligned blocks in the strings of the current highest fitness class are almost identical at the start of the epoch. The length of the epoch is very sensitive to the number of bits set to 1 in the unaligned blocks of the epoch's founding string. The fact that all unaligned blocks in the population are the *same* at the start of an epoch will cause the epoch to be longer on average than it would be if all unaligned blocks were independent. A more subtle but unfortunately poorly understood effect, as pointed out in [24], is the fact that finite population sampling in general prohibits the unaligned blocks in the population from becoming completely independent. That is, sampling causes the strings in the population to remain correlated for all time thereby effectively reducing the proportions of unaligned blocks that have a high fraction of bits set to 1. These factors lead to the theory's underestimation of the epoch durations. We are currently studying the exact dynamics of a finite population of strings searching for a new block while it resides in an epoch starting from a completely converged population at the start of the epoch. The results, which will be presented elsewhere, lead to greatly improved epoch duration predictions.

## 8. Discussion

We have seen how most of the behavioral features of the Royal Road genetic algorithm—the appearance and disappearance of epochs, the structure of the innovations, the variation in fitness levels and their fluctuations, and the epoch durations—can be understood in terms of the properties of the infinite-population generation operator **G**. The analysis showed how the basic balance of evolutionary “forces”—ordering due to selection, increased diversity and aligned-block creation and destruction due to mutation, discreteness of the state space due to the finite population, and stochasticity from finite population sampling—competed and cooperated to produce a wide range of phenomena. Some of the trade-offs between these pressures, as controlled by the GA parameters, were nonmonotonic and occasionally counterintuitive. As a concise summary of these trade-offs, table 3 presents an overview of the major analytical results we obtained for the different dynamical quantities, for small mutation rates $q$.

| | |
|---|---|
| $f_n = n(1-q)^{nK}$ (6.17) | Epoch fitness |
| $\sigma_n^2 = Kn^3q/2M$ (6.48) | Epoch fitness fluctuations |
| $P_n^n = 1 - n^2Kq$ and $P_{n-1}^n = n^2Kq$ (6.12) | Epoch population |
| $\lambda_i^n = f_i/f_n = i(1-q)^{(i-n)K}/n$ (6.23) | Epoch stability |
| $T_n = [M(N-n)A(1-\exp(-2/n))]^{-1}$ (7.12) | Epoch duration |
| $t_n = [1 + 2n + 2Kn(n+1)q] \log M$ (6.32) | Innovation duration |

Table 3: Low mutation behavior of the Royal Road genetic algorithm: an overview of the analytical results for small mutation rates $q$. $n$ denotes the epoch number, $K$ is the number of bits in a block, $N$ the total number of blocks, $q$ is the mutation rate, and $M$ is the population size.

### 8.1 Low Mutation Rate Results

The first line shows the average fitness $f_n$ of the $n^{th}$ epoch. The fitness is decreased from $n$ by a factor that drops geometrically as a function of the number of defining (aligned-block) bits $(nK)$ of the epoch.

The second line gives the variance $\sigma_n^2$ of an epoch's average fitness fluctuation. The fluctuation amplitudes $(\sigma_n)$ are proportional to the epoch level $n$ to the power $3/2$ and inversely proportional to the square root of the population size $M$. They are also proportional to the square root of the block size $K$.

The third line shows that the proportion $P_n^n$ of individuals in the highest fitness class drops proportional to the block size $K$ and with the square of the epoch number $n$. Likewise it shows that the proportion $P^n_{n-1}$ of strings in the $(n-1)^{th}$ fitness class also grows with the same coefficient.

The fourth line gives the eigenvalues $\lambda^n_i$ of the Jacobian matrix around the $n^{th}$ epoch center. These eigenvalues determine the bulk of the epoch's stability. In particular, they control the epoch's innovation and fluctuation dynamics. They can be simply expressed in terms of the relative sizes of the epoch fitness levels $f_i$.

The fifth line shows the theoretical predictions of the average epoch duration $T_n$ of epoch $n$ to leading order in $q$. We have seen that these predictions underestimate the average epoch duration. The expression is included for completeness. The theoretical epoch duration is inversely proportional to the probability to create a block $A$, the population size $M$, and a factor that depends on the probability, $1-\exp(-2/n)$, that a fitter string will spread in the population. Note, that since $1/A$ is proportional to $2^K$, the epoch duration increases exponentially with the block length $K$. This explains why epochal behavior is mainly seen for large blocks.

Finally, the last line shows the average time $t_n$ taken for the innovation from the $n^{th}$ to the $(n+1)^{st}$ epoch. The result shows that the innovation time is proportional to $2n$ and to the logarithm of the population size $M$. It also shows that increasing mutation decreases the steepness of the innovations roughly in proportion to the square of the epoch number $n$, the size of the blocks $K$, and the logarithm of the population size $M$.

In focusing our attention on fitness distributions we assumed that the exact inner structure of the unaligned blocks is unimportant and can be taken as random. As we demonstrated, this maximum entropy assumption breaks down for the calculation of the average epoch durations. In general, analyzing GA behavior solely in terms of fitness distributions will work if strings within the same fitness class act similarly under the GA dynamics. In cases where this simplification does not work one must include additional order parameters to the describe the projected state of the microscopic system. In the Royal Road GA a number of alternatives come to mind. These include using a distribution of the number of 1-bits contained in the unaligned blocks or an order parameter that describes the convergence of the bits in the unaligned blocks. If the number of fitness classes or the number of "order parameters" in general becomes too large, the analysis, though still appropriate in principle, will break down from a practical point of view. The generation operator **G** could simply acquire too many components for it to be theoretically or even numerically analyzed. However, under some circumstances, such as when fitness is determined in some noisy way, different fitness classes may be grouped together to obtain a low-dimensional **G**. The resulting coarse-grained fitness "landscape" itself may be rather complicated, but as long as it is known, our analysis can also be performed even in these cases. More interestingly, our analysis suggests that it might be enough to determine certain statistics of the fitness landscape to be able to predict the population dynamics. For instance, the fitness levels of the epochs depend only on the number of "defining bits" of the epoch's fitness class (which is roughly the logarithm of the number of strings in the fitness class).

### 8.2 Metastability, Unfolding, and Landscapes

The main result of the preceding analysis is our explanation of epochal evolution as an interplay between the infinite-population flow given by **G** and the coarse-graining of the state space due to finite population size. Note that this mechanism for metastability is quite general and applies outside of evolutionary search behavior. A large number of dynamical systems in nature, as well as evolutionary computation in general, are stochastic dynamical systems in which a large set of identical subsystems evolve through a state space, in parallel, and under the influence of one another. Macroscopic states for these systems are often defined in terms of the first moments of the distributions over the state variables of the components. A commonly observed qualitative behavior is that the mean of some state variable alternates between periods of stasis and sudden change. We would like to suggest that this sort of punctuated equilibrium behavior [23] can be explained in terms of the simple mechanism presented here. In the limit of an infinite number of subsystems the global dynamics is often much more tractable. Solving for the "flow" through the appropriate state space in the limit of an infinite number of subsystems can be used to identify state space regions where the flow is weakest. Then, in the case of a finite number of subsystems, we expect the dynamics to get trapped in the weak-flow regions. The behavioral features of the finite-size dynamics can be almost completely understood in terms of the eigenvalues and eigenvectors of the infinite-population flow operator.

The behavior of evolutionary search algorithms is often informally described as moving along a "fitness landscape" directly defined by a fitness function. It is clear from both our experiments and our analysis that this geographic metaphor, originally due to Wright [46], can be misleading. The fitness function is only a partial determinant of the dynamics. Even with a fixed fitness function, population size, mutation rate, and other parameters of the search process can radically alter the population dynamical behavior of the system, revealing or hiding much of the structure in the fitness landscape. Moreover, significant features of the population dynamics, such as the metastability and destabilization of epochs, are endogenous in the sense that they cannot be understood from a naive analysis of the landscape alone.

Let us recall that for typical evolutionary problems the genetic sequence spaces explored by populations are vast. In our case of bit strings with modest length $L = 60$, the size of the sequence space is already $2^L \approx 10^{18}$. It is clear that at any point in time the population can occupy only a minute proportion of the sequence space. It is therefore logical to assume that the population in principle could become trapped in certain regions of this state space. It has become fashionable to assume that the fitness functions over the sequence spaces that occur for typical evolutionary search problems can be modeled as "rugged landscapes" [36]. These rugged landscapes have wildly fluctuating fitness values even on very small genome-variation scales and possess a large number of local fitness optima. It has been assumed that after some time, the population is likely to be found *localized* around some suboptimal fitness peak in the sequence space. This notion of locality is also common in the analysis of quasispecies in the molecular evolution theory [12]. There, the population consists of a cloud of mutants around a local fitness optimum called the "wild type". A balance exists between the forces of mutation and crossover that makes the population diffuse away from the peak and the force of selection that tends to restore the population towards the peak. Within this view it becomes natural to associate metastable evolutionary behavior with hopping between these local fitness optima. The population stabilizes at the local peak until one of the mutants at the outer edge of the quasispecies cloud discovers a new and higher fitness peak. The population is metastable until a mutant crosses a "valley" of lower fitness towards a higher fitness peak. The local fitness optimum therefore presents a fitness barrier that has to be crossed by an individual of the population. The height of this barrier is determined by some measure of the "steepness" of the local fitness peak.

This view of metastability is in sharp contrast with the mechanism presented in this paper. Kimura was the first to advocate that many of the point mutations that occur on the molecular level are neutral with respect to fitness [28]. There is often a large degeneracy between the genotype and the fitness or functionality of its phenotype—many genotypes lead to the same or similar fitness. This means that although the fitness landscape might be rugged, there are always neutral ridges along which the genotype can move without affecting fitness. In some cases local optima might disappear completely from the fitness landscape, as in the Royal Road fitness function. As we discussed in section 6.4, if we let $S_n$ denote that subspace of all sequences containing strings of fitness $n$, then there are always points in $S_n$ that are only a single mutation away from $S_{n+1}$. In any particular time interval, the population is not likely to be localized in sequence space, but instead diffuses randomly within subspace $S_n$ until one of its mutants moves into the higher-fitness subspace $S_{n+1}$. Metastability occurs here on the phenotypic level of fitness or functionality. As we have seen, the fitness distribution stabilizes on an epoch center $\vec{P}^n$, while the best individuals in the population randomly diffuse through the subspace $S_n$. The fitness distribution remains stable until one of the mutants moves into $S_{n+1}$. The time that this takes depends on the relative number of points in $S_n$ that connect to the subspace $S_{n+1}$. In general, there is only a very small proportion of such "portal" genotypes in $S_n$. (In our case the proportion is on the order of $2^{-K}$.) The consequence is that the metastability here is due to an entropic barrier—in which large fitness-neutral volumes must be traversed. The entropy of the population in $S_n$ has to increase until almost all points in $S_n$ have been visited by the population and a connection to $S_{n+1}$ is discovered. Thus, increased phenotypic sophistication is reached by passing through conditions of increased genotypic disorder. This mechanism is quite different from the metastable behavior in sequence space due to local fitness barriers. We believe that the kind of entropy-induced metastable behavior described above is very common in evolutionary search. It will occur as long as there is a large degeneracy between the genotype and its actual functionality.

## 8.3 Future Work

We are currently studying in more detail the process of aligning blocks from a converged population. We hope that this will eventually enable us to predict the distribution of epoch durations in these entropic metastable states. In addition, we plan to include in the analysis other aspects of evolutionary systems such as crossover, geographically distributed populations with migration, and noise in the fitness function. We are currently studying more general fitness functions to see how the metastable behavior generalizes to those that have both entropic barriers and local optima and, moreover, to determine which macroscopic variables defined on fitness landscapes are most informative in predicting actual population dynamical behavior.

We are also investigating epochal evolution from an information and computation theoretic point of view [7]. Selection can be seen as installing structural information from the fitness function into the genotypes. Mutation and crossover, in contrast, can be seen as randomizing the bits in the genetic representation and thereby destroying the information that selection has stored in them, while providing a necessary source of genetic novelty. The genetic representation itself—the string length and the strings' blockiness in our simple GA—also imposes restrictions on how much of the fitness function's structure can be stored in the gene pool.

It is already evident from the preceding investigations that epochal dynamics leads to interesting informational behaviors of the population. At an innovation, for instance, the whole population becomes converged. This means that essentially *all* bits of the founder string are stored in the population, not just the "functional" bits in the aligned block, but also the arbitrary bits this founder string happens to have in the unaligned blocks. (This kind of phenomenon has also been called "hitchhiking" in the GA literature [34].) Thus, at an innovation, more raw information is stored in the population than is necessary for improved fitness. During the epoch the unaligned blocks start to diversify again and the founder string information originally stored in these unaligned blocks is destroyed. This thermalization process increases the population entropy, but is actually a prerequisite for the search being able to find increasingly better genotypes. We shall present this complementary thermodynamic analysis elsewhere.

Eventually, we would like to extend the analysis to evolutionary processes in which there is a nontrivial mapping between the genotype and the phenotype, such as found in our other work on a genetic algorithm that evolves cellular automata [8, 9], work that originally motivated the preceding investigations. We presume that entropic metastability will also be observed in these more complex adaptive systems as long as there is enough redundancy between a genotype and its phenotype.

## 9. Acknowledgments

The authors thank Lee Altenberg, Marc Feldman, Martijn Huynen, Mark Newman, Richard Palmer, Christian Reidys, and Andreas Wagner for helpful discussions. This work was supported at the Santa Fe Institute by NSF IRI-9320200, DOE DE-FG03-94ER25231, and ONR N00014-95-1-0975, and at UC Berkeley under ONR N00014-95-1-0524.

## A. Uniqueness of the Asymptotic Fitness Distribution

We will show that the matrix $\tilde{\mathbf{G}}$ has only one positive definite eigenvector as was claimed in section 5.5. Since individuals with fitness 0 have probability 0 of being selected we have $\mathbf{S}_{i0} = 0$ and $\tilde{\mathbf{G}}_{i0} = 0$. Because there is positive probability for any nonzero-fitness string to be selected and since, with positive probability, mutation can take a string with fitness $j$ into a string with fitness $i$ for any $j$ and $i$, we have

$$\tilde{\mathbf{G}}_{ij} > 0, \; j > 0. \tag{A.1}$$

Define a new matrix $\mathbf{H}$ that is the restriction of $\tilde{\mathbf{G}}$ to the positive fitness subspace. This is given by

$$\mathbf{H}_{ij} = \tilde{\mathbf{G}}_{ij}, \; i, j > 0. \tag{A.2}$$

In addition, define the vector $\vec{Q}$ to be the $N$-dimensional nonzero fitness projection of the $(N+1)$-dimensional fitness distribution $\vec{P}$ to be

$$Q_i = P_i \; , \; i > 0. \tag{A.3}$$

We can now turn the eigenvector equation

$$\tilde{\mathbf{G}} \cdot \vec{P} = f \vec{P} \tag{A.4}$$

into an eigenvector equation for $\mathbf{H}$,

$$\mathbf{H} \cdot \vec{Q} = f \vec{Q}, \tag{A.5}$$

and into an equation for the zeroth component $P_0$,

$$P_0 = f^{-1} \sum_{j=1}^{N} \tilde{\mathbf{G}}_{0j} P_j. \tag{A.6}$$

Since **H** is a positive definite matrix, Perron's theorem applies: **H** has a unique positive definite eigenvector, the eigenvalue of which is larger than all other eigenvalues. This means that **H** has a unique positive eigenvector $\vec{Q}^{max}$ with maximal eigenvalue $f_{max}$. Since all components $\tilde{\mathbf{G}}_{0j}$ are positive we have from equation A.6 that $P_0$ is positive. Specifically, we have

$$P_0 = \frac{\sum_{j=1}^{N} \tilde{\mathbf{G}}_{0j} Q_j^{max}}{f_{max}} \tag{A.7}$$

and so $\vec{P} = (P_0, Q_1, \ldots, Q_N)$ is the unique eigenvector of $\tilde{\mathbf{G}}$ with maximal eigenvalue. Since all other eigenvectors of **H** have at least one negative component, the above eigenvector $\vec{P}$ is the unique positive definite eigenvector of **G** and therefore $\vec{P}$ is the only eigenvector of **G** that can be interpreted as a fitness distribution.

## B. Finite Population Dynamics Convergence in the Infinite Population Limit

We will show that as the population size increases, the finite population dynamics approaches arbitrarily closely the infinite population dynamics for any finite number of time steps $T$. The proof presented here is a more elaborate version of a proof that was outlined in [35]. (Useful mathematical background can be found in [17].) Note that we will prove that the finite population dynamics converges towards the infinite population dynamics as given by **G**, *provided* that **G** accurately describes the dynamics of fitness distributions for an infinite population.

In the infinite population limit the dynamics is deterministic. Let $\vec{I}(t)$ denote the fitness distribution at time $t$ in the infinite-population limit. For the finite population the dynamics in fitness distribution space is stochastic. Let $\vec{P}(t)$ denote the finite population fitness distribution at time $t$. We define the distance between the $i^{th}$ component of the finite population fitness distribution and the infinite population fitness distribution at time $t$ to be

$$\delta_i(t) = |I_i(t) - P_i(t)|. \tag{B.1}$$

At time $t = 0$ the fitness distribution for the finite population is taken to be the infinite population distribution $\vec{I}(0)$. Using equation 6.1 for the transition probabilities we can calculate the probability that $\delta_i(1) > \epsilon$ for some arbitrary component, $0 \leq i \leq N$. The vector $\vec{P}(1)$ is given by a multinomial sample of size $M$ of the expected distribution $\vec{I}(1) = \mathbf{G}(\vec{I}(0))$. Using Chebysev's inequality we find that

$$\Pr[\delta_i(1) > \epsilon] \leq \frac{I_i(1)(1 - I_i(1))}{M\epsilon^2} \leq \frac{1}{4M\epsilon^2}, \tag{B.2}$$

using the inequality $x(1-x) \leq 1/4$.

With the above inequality on the transition probabilities we can prove that for any $\gamma < 1$, any $\beta > 0$, and any finite number of time steps $T$, there is a population size $M$ such that for populations larger than $M$ with probability at least $\gamma$, any component $\vec{P}_i(t)$ of the fitness distributions $\vec{P}(t)$ stays within $\beta$ of the infinite population trajectory. Specifically, we want to prove that for sufficiently large populations one has

$$\Pr\left[\delta_i(1) \leq \epsilon(1) \text{ and } \delta_i(2) \leq \epsilon(2) \text{ and} \ldots \text{and } \delta_i(T) \leq \epsilon(T)\right] > \gamma, \tag{B.3}$$

where the $\epsilon(t)$ are uniformly smaller than the chosen bound $\beta$ for all $t$. Since the process has the Markovian property that the next fitness distribution depends only on the current one, we can factor the left-hand side of B.3 into conditional probabilities:

$$\Pr\left[\delta_i(1) \leq \epsilon(1)\right] \prod_{t=1}^{T-1} \Pr\left[\delta_i(t+1) \leq \epsilon(t+1) | \delta_i(t) \leq \epsilon(t)\right]. \tag{B.4}$$

Thus, we need to bound each of these conditional probabilities. Given a population with fitness distribution $\vec{P} = \vec{I}(t) + \vec{\delta}(t)$, the expected distribution $\vec{P}'$ at the next time step is $\mathbf{G}(\vec{P})$. To first order in $\vec{\delta}(t)$ this is given by

$$\langle \vec{P}(t+1) \rangle = \mathbf{G}(\vec{I}(t) + \vec{\delta}(t)) = \vec{I}(t+1) + \mathbf{DG} \cdot \vec{\delta}(t) + \mathcal{O}(\delta^2), \tag{B.5}$$

where the Jacobian matrix $\mathbf{DG}$ is evaluated with $\vec{I}(t)$:

$$\mathbf{DG}_{ij} = \frac{\tilde{\mathbf{G}}_{ij} - jI_i(t+1)}{f(t)}, \tag{B.6}$$

where we denote the average fitness of the infinite population distribution $\vec{I}(t)$ as $f(t)$. We now need to place an upper bound $B$ on the absolute values of the eigenvalues of this matrix such that we can write

$$\left| \sum_j^N \mathbf{DG}_{ij} \delta_j(t) \right| \leq B\delta_i(t). \tag{B.7}$$

The bound $B$ is nothing other than the norm of the matrix $\mathbf{DG}$, which we can obtain by explicitly substituting a vector $\vec{\delta}$ into equation B.6,

$$\mathbf{DG} \cdot \vec{\delta} = \frac{\tilde{\mathbf{G}} \cdot \vec{\delta} - \vec{I}(t+1)\langle \delta \rangle}{f(t)}, \tag{B.8}$$

where $\langle \delta \rangle = \sum_j^N j\delta_j$. Since $f_i \geq 0$, $I_i(t+1) > 0$, and $\delta_i > 0$ for all $i$, the two terms in the above expression are of opposite sign and therefore the norm of the above expression is bounded by the norm of the larger of the two terms. That is, we have

$$\left| \sum_j^N \mathbf{DG}_{ij} \delta_j(t) \right| \leq \max\{\frac{|(\tilde{\mathbf{G}} \cdot \vec{\delta})_i|}{f(t)}, \frac{I_i(t+1)\langle \delta \rangle}{f(t)}\} \leq \frac{N}{f(t)} \delta_i. \tag{B.9}$$

Here we have used the inequalities $\langle \delta \rangle \leq N$, $I_i(t+1) \leq 1$, and $f_N \leq N$, with $f_N$ being the largest eigenvalue of $\tilde{\mathbf{G}}$. We can therefore write

$$|\langle P_i(t+1) \rangle - I_i(t+1)| \leq \frac{N}{f(t)} \delta_i(t). \tag{B.10}$$

Using Chebysev's inequality again on the multinomial transition probabilities we obtain

$$\Pr\left[|P_i(t+1) - \langle P_i(t+1) \rangle| > \epsilon\right] \leq \frac{1}{4M\epsilon^2}. \tag{B.11}$$

Furthermore, we have that

$$|P_i(t+1) - \langle P_i(t+1)\rangle| \leq \epsilon \Rightarrow \delta_i(t+1) \leq \epsilon + \frac{N}{f(t)}\delta_i(t)\ . \tag{B.12}$$

Now, if we define

$$\epsilon(t+1) = \epsilon + \frac{N}{f(t)}\epsilon(t), \tag{B.13}$$

then we find that

$$\Pr\left[\delta_i(t+1) \leq \epsilon(t+1) | \delta_i(t) \leq \epsilon(t)\right] > 1 - \frac{1}{4M\epsilon^2}. \tag{B.14}$$

Looking over a series of time steps, if we define

$$\epsilon(t) = \sum_{n=0}^{t-1} \epsilon \left[\frac{N}{f(t)}\right]^n \tag{B.15}$$

and take $\epsilon(1) \equiv \epsilon$, then we find for the joint probability of equation B.4 that

$$\Pr\left[\delta_i(1) \leq \epsilon(1) \text{ and } \delta_i(2) \leq \epsilon(2) \text{ and} \ldots \text{and } \delta_i(T) \leq \epsilon(T)\right] > \left[1 - \frac{1}{4M\epsilon^2}\right]^T . \tag{B.16}$$

Requiring this probability to be greater than $\gamma$ we obtain

$$M > \frac{1}{4\epsilon^2(1-\gamma^{\frac{1}{T}})}. \tag{B.17}$$

We also require that $\epsilon(t) < \beta$ for all $t \leq T$. From the definition of $\epsilon(t)$, we see that this requirement implies

$$\epsilon(T) < \beta \Rightarrow \epsilon < \beta \left[\sum_{n=0}^{T-1} \left(\frac{N}{f(t)}\right)^n\right]^{-1} \equiv \frac{\beta}{c[T]}, \tag{B.18}$$

where

$$c[T] = \sum_{n=0}^{T-1} \left(\frac{N}{f(t)}\right)^n . \tag{B.19}$$

Concluding, we see that if we choose $M$ such that

$$M > \frac{c[T]^2}{4\beta^2(1-\gamma^{\frac{1}{T}})}, \tag{B.20}$$

then, for all $t \leq T$, the $\delta_i(t)$ are smaller than $\beta$ with probability greater than $\gamma$. Since the right-hand side in B.20 is finite for any finite $T$, $\gamma < 1$, and $\beta > 0$ we can always find a population size $M$ such that the above inequality is satisfied. This concludes the proof. Note that if we take the limit $T \rightarrow \infty$ there is no finite population $M$ for which the deviations $\delta_i(t)$ remain arbitrarily small for arbitrary time.

## References

[1] R. Baserga, editor. *Cell Proliferation, Cancer, and Cancer Therapy*, volume 397 of *Ann. New York Acad. Sci.*, 1982.

[2] R. K. Belew and L. B. Booker, editors. *Proceedings of the Fourth International Conference on Genetic Algorithms.* Morgan Kaufmann, San Mateo, CA, 1991.

[3] A. Bergman and M. W. Feldman. Question marks about the period of punctuation. Technical report, Santa Fe Institute Working paper 96-02-006, 1996.

[4] T.D. Brock. *Biology of microorganisms*, chapter 23, pages 942–943. Prentice Hall, 1997.

[5] T.D. Brock. *Biology of microorganisms*, chapter 23, pages 956–957. Prentice Hall, 1997.

[6] J. P. Crutchfield. Subbasins, portals, and mazes: Transients in high dimensions. *J. Nucl. Phys. B*, 5A:287, 1988.

[7] J. P. Crutchfield. The calculi of emergence: Computation, dynamics, and induction. *Physica D*, 75:11 – 54, 1994.

[8] J. P. Crutchfield and M. Mitchell. The evolution of emergent computation. *Proc. Natl. Acad. Sci. U.S.A.*, 92:100742–10746, 1995.

[9] R. Das, J. P. Crutchfield, M. Mitchell, and J. E. Hanson. Evolving globally synchronized cellular automata. In L. J. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 336–343, San Francisco, CA, 1995. Morgan Kaufmann.

[10] L. D. Davis, editor. *The Handbook of Genetic Algorithms.* Van Nostrand Reinhold, 1991.

[11] M. Eigen. *Naturwissenschaften*, 58:465–522, 1971.

[12] M. Eigen, J. McCaskill, and P. Schuster. The molecular quasispecies. *Adv. Chem. Phys.*, 75:149–263, 1989.

[13] M. Eigen and P. Schuster. *Naturwissenschaften*, 64:541–565, 1977.

[14] M. Eigen and P. Schuster. *The Hypercycle.* Springer, Berlin, 1979.

[15] S.F. Elena, V.S. Cooper, and R.E. Lenski. Punctuated evolution caused by selection of rare beneficial mutations. *Science*, 272:1802–1804, 1996.

[16] L. Eshelman, editor. *Proceedings of the Sixth International Conference on Genetic Algorithms.* Morgan Kaufmann, San Mateo, CA, 1995.

[17] W. Feller. *An Introduction to Probability Theory and its Applications.* Wiley, New York, third, revised edition, 1970.

[18] R. A. Fisher. *The Genetical Theory of Natural Selection.* Oxford, The Clarendon Press, 1930.

[19] A. D. Fokker. *Ann. Phys. (Leipzig)*, 43(310), 1915.

[20] S. Forrest, editor. *Proceedings of the Fifth International Conference on Genetic Algorithms.* Morgan Kaufmann, San Mateo, CA, 1993.

[21] C. W. Gardiner. *Handbook of Stochastic Methods.* Springer-Verlag, 1985.

[22] J. H. Gillespie. Chapter 4. In M. W. Feldman, editor, *Mathematical Evolutionary Theory.* Princeton University Press, 1989.

[23] S.J. Gould and N. Eldredge. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*, 3:115–251, 1977.

[24] R. Haygood. The structure of royal road fitness epochs. *Evolutionary Computation, submitted*, ftp://ftp.itd.ucdavis.edu/pub/people/rch/StrucRoyRdFitEp.ps.gz, 1997.

[25] M. Huynen, P. F. Stadler, and W. Fontana. Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci.*, 93:397–401, 1996.

[26] M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47:713–719, 1962.

[27] M. Kimura. Diffusion models in population genetics. *J. Appl. Prob.*, 1:177–232, 1964.

[28] M. Kimura. *The neutral theory of molecular evolution.* Cambridge University Press, 1983.

[29] M. Kimura and T. Ohta. The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61:763–771, 1969.

[30] A. N. Kolmogorov. *Math. Annal.*, 104:415–418, 1931.

[31] K. Lindgren. Evolutionary phenomena in simple dynamics. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 295–312, Reading, MA, 1992. Addison-Wesley.

[32] J. S. McCaskill. A stochastic theory of macromolecular evolution. *Biological Cybernetics*, 50:63–73, 1984.

[33] M. Mitchell, J. P. Crutchfield, and P. T. Hraber. Evolving cellular automata to perform computations: Mechanisms and impediments. *Physica D*, 75:361 – 391, 1994.

[34] M. Mitchell, J. H. Holland, and S. Forrest. When will a genetic algorithm outperform hill climbing? In J. D. Cowan, G. Tesauro, and J. Alspector (editors), *Advances in Neural Information Processing Systems 6*. San Mateo, CA: Morgan Kaufmann.

[35] A. E. Nix and M. D. Vose. Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 5, 1991.

[36] A. S. Perelson and S. A. Kauffman, editors. *Molecular evolution on rugged landscapes: proteins, RNA and the immune system*, volume IX of *Santa Fe Institute Studies in the Science of Complexity*, 1989.

[37] M. Planck. *Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl.*, 325, 1917.

[38] A. Prugël-Bennett and J. L. Shapiro. Analysis of genetic algorithms using statistical mechanics. *Physical Review Letters*, 72(9):1305–1309, 1994.

[39] A. Prügel-Bennett and J. L. Shapiro. The dynamics of a genetic algorithm in simple random Ising systems. *Physica D*, 104 (1):75–114, 1997.

[40] M. Rattray and J. L. Shapiro. The dynamics of a genetic algorithm for a simple learning problem. *J. of Phys. A*, 29(23):7451–7473, 1996.

[41] T. S. Ray. An approach to the synthesis of life. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 371–408, Reading, MA, 1992. Addison-Wesley.

[42] J. Swetina and P. Schuster. Self replicating with error, a model for polynucleotide replication. *Biophys. Chem.*, 16:329–340, 1982.

[43] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry.* North-Holland, 1992.

[44] E. van Nimwegen, J. P. Crutchfield, and M. Mitchell. Finite populations induce metastability in evolutionary search. *Phys. Lett. A*, 229:144–150, 1997.

[45] M. D. Vose and G. E. Liepins. Punctuated equilibria in genetic search. *Complex Systems*, 5, 1991.

[46] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proc. of the Sixth International Congress of Genetics*, volume 1, pages 356–366, 1932.