The Statistical Dynamics of Epochal Evolution

De Statistische Dynamica van Gefaseerde Evolutie
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. Dr. H.O. Voorma,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op
maandag 27 september 1999 des middags te 4.15 uur

door

Erik Jan van Nimwegen

geboren op 5 november 1970
te Amsterdam.

Promotores:

Prof. P. Hogeweg
Faculteit Biologie
Universiteit Utrecht.

Prof. J. P. Crutchfield
Santa Fe Institute

Promotie Commissie:

Prof. B. Derrida
Laboratoire de Physique Statistique
Département de Physique
Ecole Normale Supérieure, Paris.

Prof. P. Schuster
Institut für Theoretische Chemie
Universität Wien.

Prof. J. Shapiro
Computer Science Department
University of Manchester.

The studies described in this thesis were performed at the Santa Fe Institute in Santa Fe, New Mexico, U.S.A. and at the department of theoretical biology and bioinformatics at Utrecht University. The investigations were financially supported by the Santa Fe Institute and by the Priority Program Nonlinear Systems of the Netherlands Organization for Scientific Research (NWO).

ISBN 90-393-2162-0

Keine Annahme scheint mir natürlicher, als daß dem Assoziieren oder Denken kein Prozeß in Gehirn zugeordnet ist; so zwar, daß es also unmöglich wäre, aus Gehirnprozessen Denkprozesse abzulesen. Ich meine das so: Wenn ich rede oder schreibe, so geht, nehme ich an, ein meinem gesprochenen oder geschriebenen Gedanken zugeordnetes System von Impulsen von meinem Gehirn aus. Aber warum sollte das *System* sich weiter in zentraler Richtung fortsetzen? Warom soll nicht sozusagen diese Ordnung aus dem Chaos entspringen? Der Fall wäre ähnlich dem—daß sich gewisse Pflanzenarten durch Samen vermehrten so daß ein Same immer dieselbe Pflanzenart erzeugt, von der er erzeugt wurde,— daß aber *nichts* in dem Samen der Pflanze, die aus ihm wird, entspricht; so daß es unmöglich ist, aus den Eigenschaften oder der Struktur des Samens auf die der Pflanze, die aus imh wird, zu schließen,—daß man dies nur aus seiner *Geschichte* tun kann. So könnte also aus etwas ganz Amorphem ein Organismus sozusagen ursachelos werden; und es ist kein Grund, warum sich dies nicht mit unserem Gedanken, also mit unserem Reden oder Schreiben etc. wirklich so verhalten sollte.

No supposition seems to me more natural than that there is no process in the brain correlated with associating or with thinking; so that it would be impossible to read off thought-processes from brain-processes. I mean this: if I talk or write there is, I assume, a system of impulses going out from my brain and correlated with my spoken or written thoughts. But why should the *system* continue further in the direction of the center? Why should this order not proceed, so to speak, out of chaos? The case would be like the following—certain kinds of plants multiply by seed, so that a seed always produces a plant of the same kind as that from which it was produced—but *nothing* in the seed corresponds to the plant which comes from it; so that it is impossible to infer the properties or structure of the plant from those of the seed that comes out of it—this can only be done from the *history* of the seed. So an organism might come into being even out of something quite amorphous, as it were causelessly; and there is no reason why this should not really hold for our thoughts, and hence for our talking, and writing.

Ludwig Wittgenstein, Zettel [145]

Contents

1

Introduction

1.1 Formalizing the Darwinian Paradigm

The Darwinian theory of evolution [26] of reproducing entities with heritable variation that effects their reproductive success is enormously flexible. In essence it is a paradigm for an explanative theory that can be applied to *any* system which contains objects

- that reproduce with variation.
- that have heritable characters.
- whose reproductive success is affected by their characters.

This flexibility and general applicability is the paradigm's strength, on the one hand, while on the other hand it leaves unspecified *how* exactly the theory of evolution should be mathematically formalized.

This is in contrast to the typical situation found in physics where theories are generally formulated in mathematical terms from the start. The analysis of the system under study is generally conceptually clear. The dynamics is rigorously defined at some microscopic level and one imagines that the (higher level) behavior of the system is in principle derivable from these microscopic equations of motion when combined with particular boundary conditions. One in general focuses on the task of deriving the behavior of interest from these underlying microscopic laws. This task is in itself highly nontrivial and may not always be accomplished—such as in chaotic dynamical systems, where closed-form solutions of the equations of motion are often not available. In other words, one explores, to the best of one's abilities, the consequences of rigorously defined microscopic equations of motion.

From this particular viewpoint, the Darwinian theory of evolution doesn't seem so much a *theory* of evolution but instead more of a *paradigm* for a theory of the evolution of life on earth. That is, with respect to the theory of evolution, the question is still of a conceptual nature: What is the "right" way to formalize the theory in a mathematical framework?

For example, which objects in the evolving system should one take as the fundamental reproducing entities when implementing the Darwinian paradigm? Should one

take species, individuals, genomes, genes, or other less obvious constructions, such as collections of interacting individuals? Indeed, precisely this question has been the source of heated debates among evolution theorists; [29] for instance argues that genes are *the* units at which natural selection acts. The main problem, in the opinion of the present author, is that there is no such *right* or *fundamental* level: the Darwinian paradigm is capable of describing evolutionary dynamics on many different levels without any of these providing a fundamental level of description by itself.

Surely, different genes are competing with each other within the genomes of organisms and in that sense these genes undergo Darwinian evolution. However, this evolutionary dynamics cannot be rigorously defined without taking into account interactions with the evolution on higher levels—such as between genes within a genome and on the level of individuals. Similarly, single-celled organisms appear to be undergoing Darwinian evolution, competing with each other for molecular resources. Cells, however, interact, exchange genetic material and may form intricate spatial patterns. It is thus clear that one cannot define evolutionary equations of motions that only have "single-cell types" as variables. One may also attempt to formalize Darwinian evolution on the level of different *species* of single-celled organisms that compete with each other. At an even higher level, collections of interacting cells form multicellular organisms that can be considered to undergo Darwinian evolution as well. Finally, one may even formalize evolution on the level of different *groups* of organisms. See for instance [144] and references therein for an overview of the *group selection* versus *individual selection* debate.

One typically finds that the theorist attempts to circumvent these issues by introducing the mathematical formalization directly at the level that happens to be of immediate interest to him. The equations of motion are then constructed at this level and the behavior at the level of interest can be directly studied. However, as the examples above point out, the dynamics might not have *integrity* at this chosen level, since it depends on interactions with the dynamics on different levels. That is, evolution takes place at different levels at the same time, and the evolutionary *equations of motion* on different levels of description are often interdependent [73].

To give an illustrative example of such cross-level interdependence, which can be found in [13]: In a spatial environment, hypercycles [35] of replicating molecules self-organize into rotating spiral waves of different molecular species. Because of this, selection not only takes place on the level of the individual molecules, but also on the level of the rotating spirals. The competition between these spirals feeds back on the evolutionary dynamics on the level of the individual replicating molecules. That is, due to the selection on the level of spirals, the selection on the level of the replicating molecules is changed. In a completely mixed environment, molecules with a high decay rate are out competed by molecules with a low decay rate. In a spatial environment, however, these fast decaying molecules give rise to spirals that rotate faster than those made up of molecules with a low decay rate. Since on the level of spirals the faster spirals out compete slow spirals, the selection on the level of molecules is reversed, i.e. the molecules with a high decay rate now out compete those with a low decay rate.

From the viewpoint of mathematical formalization such situations present fundamental problems. The dynamics on the level of the individuals cannot be understood without including the dynamics the level of the higher-order structures, i.e. the spir-

als. At the same time, the dynamics on the level of these higher-order structures cannot be separated from the dynamics on the level of the individuals making it impossible to define "closed" equations of motion on either of these levels.

This problem seems to occur frequently in evolutionary theory. It would in principle be possible to define the evolutionary model at some fundamental level—i.e. involving such things as distributions of genotypes, where in space the individuals are located, what the states of the local environments are, which individuals interact with which, what the internal states of the different individuals are, etc. However, one is tempted to immediately move to the level of description of interest and invent ad hoc evolutionary equations of motion at this level. For instance, in mathematical population genetics, one is often interested in the evolution of the frequencies of certain genes in the population. The complicated influences of the rest of the genomes and the dynamics on the level of the individuals are often ignored and the evolutionary dynamics is simply defined on the level of the gene frequencies only. On the other end of the spectrum, models are constructed in which the dynamics on the level of genes is ignored and evolution is described at the level of frequencies of certain *adaptive phenotypic traits* in the population, see for instance [72, 129].

As said, I do not believe there is any one level that is *the* right level for describing evolutionary dynamics. The problem in evolution theory is not to formulate *new* dynamical laws for how natural systems behave at a *fundamental* level. Such laws already exist: they are the laws of physics. The problem is to investigate the consequences that emerge out of these physical laws in the context of biological systems. That is, we need to discover what higher level behavior is induced in biological systems by the underlying laws. The Darwinian evolution theory provides a theoretical paradigm that seems to capture the essence of most of this higher level behavior.

Fortunately, it is not necessary to formalize the Darwinian evolutionary dynamics at *one particular* level exclusively. Defining the equations of motion at a certain microscopic level doesn't necessarily mean that one can only study its behavior *at that level*. It is possible to start at a certain microscopic level of description and to *infer* the dynamics at a higher level of interest. Precisely this reconstruction of the behavior at the level of interest from the underlying microscopic equations of motion is the goal of *statistical mechanics* in physics. The main mathematical methodology for systems in thermodynamic equilibrium was developed by Gibbs [57] and Boltzmann [14]. In the 1950s Jaynes generalized this methodology and put it on a stronger conceptual basis, which includes exposing its close connections with information theory [84]. This general methodology allows one to start at the fundamental microscopic equations of motion of a system, and derive macroscopic equations of motion for the dynamics at some other level of interest. In the opinion of this author, this methodology is important for evolutionary theory since it may bridge the conceptual gap between the precise microscopic equations of motion of an evolutionary system that do not address the questions that interest us, and the ad hoc macroscopic evolutionary models that are not properly founded in a realistic microscopic dynamics. Although this methodology is conceptually straightforward, its technical implementations for different situations are far from fully developed.

In this thesis, the benefits of this methodology are demonstrated by application to some very simple evolutionary systems that are of interest in their own right. In these

models, the microscopic evolutionary dynamics is precisely defined on the level of distributions of genotypes in the population. At the same time, we are only interested in the dynamical behaviors of the population at some much coarser level—such as the average and best fitness in the population. Although the exact microscopic dynamics is almost impossible to solve analytically in detail, we show how the approximate macroscopic dynamics may still be constructed from the underlying microscopic equations of motion by using the statistical mechanics methodology. Since we focus on dynamics as opposed to the stationary states typically studied in statistical mechanics, we denote this approach as *statistical dynamics*[1] The methodology in itself is general enough that we expect that it can be applied in more or less modified form to many different evolutionary systems including much more complicated and realistic ones.

1.2 Simple Evolutionary Systems

Apart from providing illustrative examples of how one can, starting from the microscopic equations of motion, derive the evolutionary dynamics on a macroscopic level of interest, the models in this thesis also have some more direct relevance for understanding the evolutionary dynamics of simple evolutionary systems. The particular evolutionary systems analyzed in this thesis, and some of the assumptions that we make in defining them, will now be briefly introduced.

All evolutionary systems that are studied in this thesis consist of populations evolving in a constant and homogeneous external environment. That is, for each individual in the population its reproductive success is only a function of its "type". This assumption excludes a vast array of possible complications that typically occur in more realistic situations. To name a few of these: in many situations the reproductive success of an individual is not only a function of its own type but also of its interactions with individuals of other types. In these situations one often says that the selective environment is not externally imposed but is a function of the internal state of the system. One also talks of *frequency dependent selection*—see for instance [17, 39]—when the reproductive success of one type of individual depends on the frequencies with which other types occur in the population. In case where the different types evolve and determine each other's reproductive success, the word *co-evolution* is also used to describe these situations—see for instance [124]. Even if selection is entirely externally determined one may still have that the reproductive success of an individual depends on where in time and space the individual finds itself, since the selective environment may vary in space and time. On top of these complications, individuals cannot generally be considered to be characterized by a particular "type". That is, individuals can occur in many different states themselves and their reproductive success may depend on when and where the individual finds itself in which states. In other words, although an individual can be mainly identified with its *genotype*—the particular DNA sequence that constitutes its genome—it goes through a complicated developmental process of interactions with its environment that eventually leads to a *phenotype*—-the actual shape, form, and behavior of the organism. These phenotypes generally show complicated behaviors and interactions with

[1] This terminology was first coined in an evolutionary context by Crutchfield. See also [131, 132] for a development of approximately the same analytical approach under the same name.

their environment, which in turn determine the reproductive success of the genotype.

As noted, none of these complications are studied in this thesis. We will look at populations of individuals whose fitness is a constant function of their genotypes only and that reproduce with certain definite genetic operators acting on their genotypes. Such simple systems occur in practice only in very specific situations. For instance, one may think of recent experiments where bacteria of a single species are evolved in a constant environment for many thousands of generations [97]. *In vitro* experiments with replicating biomolecules are also examples of the kinds of systems that may be modeled by the systems studied in this thesis. Particular examples are the *in vitro* experiments with replicating RNA molecules [10, 102] that are closely related to the Eigen model of molecular evolution [33]. Even simpler cases of relevance are examples of *in vitro* evolution where biomolecules are evolved to perform a certain biochemical task, such as binding a particular ligand [135, 147]. Similar *in vitro* evolution experiments have also been simulated in detail on computers [46, 47, 76, 81]. These artificial evolution experiments are closely related to the systems studied in this thesis. In all these cases the dynamics is fairly well understood on an microscopic level: simple "individuals", i.e. molecules, that can be identified with their genotype, reproduce and undergo simple genetic operators—such as mutation and recombination—in a constant selective environment.

Finally, the models studied here are examples of what are generally called *evolutionary algorithms*. Evolutionary algorithms are stochastic search methods that are based on the Darwinian evolution paradigm. Populations of candidate solutions, represented as "genotypes" in a computer, evolve under selection for solving the problem under consideration and undergo (random) genetic operations that introduce variation; see for instance [4, 18, 94, 103]. Although most of the interest in evolutionary algorithms is focused on the performance of these algorithms as search methods, they can also be regarded as simple prototype models for evolutionary dynamics. This is the way in which such evolutionary algorithms are studied in most parts of this thesis. Understanding the dynamics of such evolutionary algorithms will of course also lead to a better understanding of their optimization and search properties. This type of analysis is also presented, see chapters 5 and 6 in this thesis.

1.3 Epochal Evolutionary Dynamics

Now that some of the conceptual motivations behind the analysis in this thesis, as well as the natural and artificial systems that are most directly related to the models studied here, have been introduced, I finally introduce briefly the type of dynamical behaviors that this work focuses on.

As emphasized above, we are not interested in the precise dynamics on the microscopic level of distributions of genotypes at which the evolutionary dynamics is defined. Especially for the kinds of natural and artificial systems that are related to the evolutionary models that we study, it is natural to be interested only in the dynamics on the coarser level of average or best fitness in the population. For instance, when running an evolutionary algorithm, one is interested in the quality of solutions that are evolved and related statistics—such as the average time to evolve a solution of a certain quality.

In most of this thesis, the analysis thus attempts to predict the dynamics on the level of fitness in the population. In principle, the same type of analysis could be used to study the dynamics on the level of phenotypes instead of fitness or, even, on the level of certain phenotypic traits.

This thesis further focuses on a dynamical behavior in evolving populations that we have called *epochal evolution*. In epochal evolution, long periods of stasis in the population average of a macroscopic variable are punctuated by sudden "innovations" of rapid change. That is, the evolutionary dynamics shows metastable behavior on the level of the macroscopic variable of interest. In cases where fitness is the macroscopic variable of interest, innovations generally lead to a sudden increase of the average fitness in the population. On the level of phenotypes in the population, innovations can often be identified with the emergence of new phenotypic functionality.

Such epochal evolution has been observed in a wide class of evolutionary dynamical systems. First there is the common occurrence of *punctuated equilibria* in the fossil record [64]. On macro-evolutionary time scales the fossil data seems to indicate that phenotypic evolution occurs in short bursts that are separated by long periods of phenotypic stasis. Such behavior has recently also been observed on much shorter time scales in the evolution of bacterial colonies [37]. There, the epochal behavior occurred both on the level of fitness (reproduction rate) and a phenotypic trait (cell size). The behavior has also been observed in the evolutionary dynamics of replicating computer programs [1], where innovations correspond to the sudden increase of the reproduction rate of these computer programs. In computer simulations of *in vitro* evolution of biomolecules, such epochal evolution seems to be a common occurrence as well. For instance in cases where RNA sequences are evolving towards a certain secondary structure [47, 79]. And finally, it is also commonly observed in the behavior of evolutionary search algorithms, for instance in the evolution of cellular automata toward performing a certain computational task [22]. In those evolution experiments, the innovations correspond to sudden changes in the dominant types of computational strategies that occur in the population.

In general, the epochal behavior in natural and artificial evolutionary systems can be the result of a variety of different mechanisms. In this thesis, we focus on some mechanisms that we believe may be responsible for much of the observed metastability in evolutionary systems. These mechanisms for epochal evolution are now briefly introduced.

On an abstract level, the different mechanisms take on the same mathematical form, which results from an interplay between the finite sizes of evolving populations, and the expected *flow* of these populations through the state space of macroscopic variables. That is, from the original microscopic dynamics on the level of the genotypes one constructs an expected dynamics on the level of the macroscopic variables of interest. The state space on this macroscopic level then consists of all possible population distributions over these macroscopic variables—such as, fitness or phenotypic traits. For a finite population such a state space is a discrete lattice; the percentage in the population of a certain macrovariable must be a multiple of the inverse population size. Metastability occurs when the expected flow through this state space is *small* in comparison to the state space lattice spacing, which is equal to the inverse population size.

On a more concrete level, such small flows occur when the expected number of individuals of a certain macroscopic type that is produced per unit time is much smaller than

one. This is quite likely to happen for macroscopic types that are rare in genotype space, since genotype spaces are generally much larger than population sizes. That is, even if the population spreads uniformly through genotype space, it can only visit a minute fraction of this space, so that rare macroscopic types will not be generated for long periods of time. Still, there might be different reasons for the low rate of production of these rare macroscopic types. Most prosaically put, their production may be inhibited because evolutionary forces are keeping the population from approaching them or, alternatively, they may just be hard to find since the evolving population has no way of directing itself *towards* them. Let us briefly review these two distinct mechanisms in the context of populations evolving under selection and mutation in a constant environment.

The first of these two mechanisms is the most commonly evoked explanation for metastability in evolving populations and involves the notion of a *fitness barrier*. Assume that the most-fit genotype occurring in the population is fitter than any of its single mutant neighbors. Selection then concentrates the population around this *local optimum* in genotype space. In this way, the population is kept from even higher-fitness genotypes by a "valley" of low-fitness genotypes around the local optimum at which it is currently located. Note that in this fitness barrier mechanism, the population is metastable on the microscopic level of the genotypes themselves. The evolutionary dynamics of crossing such fitness barriers is the topic of chapter 8. In most of this thesis, however, it is the alternative mechanism—that in which higher fitness types are simply hard to find—that is studied.

The central concept involved in this alternative mechanism is the occurrence of *neutrality* in the genotype-to-phenotype and genotype-to-fitness mapping over the microscopic genotype space. The concept will be discussed in more detail later. For now it suffices to point out how it is involved in bringing about the epochal evolutionary dynamics. The main idea is that there are large degeneracies in the genotype-to-fitness mappings. That is, there are many more genotypes than there are distinct fitness types. Neutrality refers to the fact that (almost) any genotype in genotype space has one or more mutant neighbors that are *neutral* to it with respect to fitness. Instead of the fittest individuals in the population occurring at a local optimum in genotype space, trapping the population genotypically around this local optimum, neutrality causes the population to diffuse through neutral *subbasins* of equal fitness genotypes. Since genotypes with even higher fitness might be rare, the population may have to search through this subbasin for a long time until a *portal* is discovered to a genotype of higher fitness. Nothing keeps the population from visiting these higher fitness genotypes, they are just hard to find. We refer to this situation as an *entropy barrier*. Note also that in this case the metastability does not occur on the microscopic level of genotypes, but only occurs on a coarser level of phenotypes or fitness.

This mechanism for metastability received increasing attention recently [6, 47, 77, 79, 108]. A growing consensus has developed that neutrality is a common occurrence in genotype-to-phenotype and genotype-to-fitness mappings that typically occur in biological systems. The idea goes back to Kimura's contention that most evolutionary change at the genotype level is selectively neutral [91]. More recently, "biological" genotype-to-phenotype mappings have been studied in much more detail. Large scale neutrality was observed both in the sequence to secondary-structure mapping of RNA [48, 65, 126] and in sequence to structure mappings of proteins [3, 9]. Neutrality seems to commonly oc-

cur in combinatorial optimization problems [22] as well, especially when there is some redundancy in the coding of candidate solutions.

Since neutrality seems to be so common in the fitness functions of evolutionary systems, we believe that the epochal evolutionary dynamics caused by the entropy barriers that it induces might be a common occurrence in experimental and natural systems, and play a role in the metastable behavior of many of the different evolutionary systems mentioned above. Much of the analysis in this thesis is concerned with studying in detail the dynamics of epochal evolution in selective environments that possess neutrality.

1.4 Outline of the Thesis

Chapter 2 gives a general introduction to the conceptual foundations of deriving macroscopic laws from underlying microscopic laws. Such "macroscopic" laws are very common in physics; any *useful* physical law that has direct applications is almost always a macroscopic (approximate) law. Historically, statistical mechanics might be considered the first theory that attempts to formalize deriving macroscopic laws from the underlying microscopic ones. It has been very successfully applied to systems in thermodynamic equilibrium. Unfortunately, the conceptual foundations of the statistical mechanics approach have been, and to a certain extent still are unnecessarily muddled. Since the statistical mechanics framework is going to be used in this thesis on systems relatively far removed from the kind of systems normally studied in statistical physics, it is important to get these conceptual issues laid out very transparently. This is what is being attempted in chapter 2. The exposition is based on the idea that statistical mechanics is a special example of a general method of statistical inference that goes under the name of the *maximum entropy method*. As will also be discussed, this methodology is intimately related to information theory. The chapter ends with some comments on the (ir)relevance of the second law of thermodynamics for evolutionary systems.

In chapter 3 we first describe in more detail the kind of evolutionary systems that will be studied in this thesis. In particular, we describe the general type of microscopic equations of motion that is obeyed by these systems. After that, we discuss what kind of macroscopic variables seem reasonable candidates in the context of evolutionary systems, and how the maximum entropy methodology is applied in these situations. In particular, we derive the general form of the macroscopic equations of motion that are used throughout the applications in this thesis. We then describe how metastability is induced by finite populations in this framework. Finally, this chapter describes the general view of the dynamics through macroscopic state space in epochal evolution that has arisen from the analysis in this thesis. Specifically, the sudden innovations in epochal dynamics correspond to the *unfolding* of new macroscopic, i.e. fitness or phenotypic, dimensions. This unfolding of a new dimension in macroscopic state space can, to a certain extent, also be identified with *phase transitions* through a dynamic symmetry breaking on the microscopic level of the dynamics, i.e. on the level of genotypes. In this way, the epochal evolution unfolds its macroscopic state space over time through evolutionary innovations. This view on the evolutionary dynamics seems especially nice since it can accommodate more naturally the conception of evolution as a dynamic open ended process that doesn't take place in an a priori defined state space, but builds its

effective state space as it goes along. As an aside, it is also consonant with the literal meaning of the word evolution as unfolding.

Chapter 4 presents in detail our statistical dynamics approach to a model of a simple evolutionary system. In particular, we study the dynamics on the level of fitness of a "genetic algorithm" on a set of fitness functions known as *Royal Road* fitness functions. These fitness functions divide the genotype space into a relatively small number of nested neutral subbasins, and the epochal dynamics is the result of the entropy barriers that are generated in this way. We analyze where in the space of fitness distributions the epochs occur and study the stochastic finite population dynamics in these areas using a linearization of the dynamics together with a diffusion equation approach. This allows us to calculate many quantitative features of the epochal dynamics on these sets of fitness functions such as fitness fluctuations during epochs, the shape of innovation curves between epochs, epoch stability and destabilization times. We also study and give approximations to the average durations of the metastable epochs. The quantitative theory is then used to understand a wide range of qualitatively different dynamical behaviors that evolving populations exhibit on this class of fitness functions under changes of the evolutionary parameters.

This analysis is then used and extended in chapters 5 and 6 to derive optimal evolutionary parameter settings for epochal evolutionary search. That is, we derive how to set mutation rates and population sizes such that evolving populations will on average reach the global optimum as quickly as possible. In these chapters we thus study the evolutionary dynamics particularly in the context of evolutionary search algorithms. To model the epochal evolution, we use a class of fitness functions that we have called *Royal Staircase* fitness functions, which are closely related to the Royal Road functions studied in chapter 4. These functions too consist of nested neutral subbasins of increasing fitness. Apart from providing quantitative predictions, the understanding of the general mechanisms involved in parameter optimization should have relevance to more general evolutionary-search settings. In chapter 5 we also study in more detail the role of crossover for evolutionary search in systems that show epochal evolution. Comparison of the *scaling* of the total number of fitness function evaluations needed to reach the global optimum with those for hill climbing algorithms are also made. In chapter 5 we focus mainly on the optimization of the mutation rate, while in chapter 6 we focus on the population size effects which turn out to be mainly related to destabilization of epochs due to finite population size. At the end of chapter 6 we discuss in more generality how the evolutionary dynamics effectively induces a *coarse graining* of the fitness function on which the population evolves. We relate this to the question of what kinds of fitness functions may be efficiently searched by an evolutionary algorithm.

In chapter 7 we move away from studying the dynamics of populations *between* different neutral subbasins and study the dynamics of an evolving population *within* a single neutral subbasin. Such neutral subbasins are generally referred to as *neutral networks* within the theory of molecular evolution. We describe a general neutral subbasin/network architecture by a graph embedded in genotype space and study how an evolving population distributes itself over such a neutral network. The analysis shows that this distribution is *independent* of evolutionary parameters such as mutation rate and population size and is a function only of the topology of the neutral network. Additionally, the theory shows that the population will evolve a *mutational robustness* (the

insensitivity of the phenotype to mutations) that is higher than if the population were to spread randomly over the neutral network. The independence of this mutational robustness to the evolutionary search parameters then allows us to classify neutral network topologies according to the mutational robustness that they generate. Finally, we discuss the possibility of *in vitro* or *in vivo* experiments that would allow one to infer the topology of neutral networks of biomolecules from simple population statistics.

In most of this thesis, we have studied metastability in the context of entropy barriers in genotype space. The general occurrence of fitness barriers and local optima are however the more traditionally invoked explanations for the occurrence of metastability in the population dynamics. In chapter 8 we turn to studying this mechanism for metastability so as to be able to compare it with the metastability caused by entropy barriers. In the first parts of the paper we define simple fitness functions that consist of a single local optimum, a fitness "valley" and a portal genotype. We then study in detail the dynamics of crossing this single fitness barrier between the local optimum and the portal using a branching process approach. We derive accurate analytical expressions for the barrier crossing time as a function of the evolutionary parameters and the parameters defining the fitness function, i.e. barrier height and width. We then also derive the rough scaling behavior of these barrier crossing times. Notably, we show that these barrier crossing times scale very differently from the barrier crossing times typically found in physics (such as the waiting time for a physical system to cross an energy barrier). We discuss how these different scalings arise and clarify the danger in the "landscape" metaphors so often used in evolution theory. We then study in more detail the dynamics of entropy barrier crossing and in particular show that the waiting times for entropy barrier crossing exhibit anomalous scaling in both population size and mutation rate.

In the second half of the paper we combine our previous statistical dynamics analysis with the fitness barrier crossing analysis to study the general metastable population dynamics in fitness functions that contain both fitness and entropy barriers. These fitness functions are called *Royal Staircase with Ditches* and provide a further extension of the class of fitness functions that we study. This wide class of fitness functions allows us to compare the fitness and entropy barrier crossing times directly within the same model setting. The analysis shows clearly how much faster entropy barrier crossing occurs than fitness barrier crossing. The analysis thus quantifies the intuitive idea that the possibility for further evolution of a population is largely determined by where neutral evolution may take it.

Finally, a short comment on the methodology followed in most of this thesis. Generally, we first define a set of microscopic evolutionary equations of motion whose macroscopic behavior on the level of, for instance, fitness we want to understand. After that, we present our theoretical approach that attempts to quantitatively describe and predict this macroscopic behavior. Finally, we compare our predictions with results from simulations of the precise microscopic equations of motion. In particular, apart from showing parameter values for which our predictions are in accordance with simulation data, we also study where in parameter space our predictions break down, and why they do so. We feel that this last step is essential for giving our results the power of generalization. By analyzing the regime of validity of the theory, we are immediately pointed to the *restrictions* on the theoretical results. Additionally, these restrictions may point the way to further improvements of the theory. Without this, it seems that theoretical developments

may start to suffer from inbreeding; it may become a theoretical game that relates to nothing whatsoever but itself.

2

Statistical Mechanics and Thermodynamics

The fundamental physical laws describe the dynamics of physical systems only at a very microscopic level. In order to make predictions about the phenomena we see around us, we have to make *inferences* about the compounded effect of the physical laws acting on all the microscopic variables in parallel. In general, analytical derivations alone are incapable of deducing the macroscopic phenomena from the underlying laws. Statistical mechanics can be considered the first theory that successfully inferred the stationary macroscopic behavior of such systems as fluids, gases, and solids. It explained thermodynamic concepts such as *heat*, *temperature*, and *entropy* that cannot be simply understood, let alone deduced, from the underlying microscopic laws. It has by now become clear that the essentials of the statistical mechanical approach lie in the method of *statistical inference* that is used to move from the microscopic level of the physical laws to the macroscopic level of the thermodynamic states. This methodology has far more general applicability than equilibrium thermodynamics. Since this methodology will be used in an evolutionary context later on, we will introduce the essentials of this approach in this chapter.

In the course of the last three centuries, the fundamental laws of physics have been formulated at an ever more microscopic level. While Newton's theory of gravitation described such day to day phenomena as the falling of apples and the orbit of the moon, one can hardly imagine concepts further away from the experiences of our daily life than the currency of contemporary physics: electrons, quarks, gauge-fields, and space-time curvatures. Ironically, even those electrons and quarks are, in a certain sense, metaphors for the much more abstract mathematical concepts that feature in the fundamental laws.

What then is their bearing on the experiences of our daily life? On one side, a reductionist may stress that knowledge of the fundamental laws combined with explicit and detailed knowledge of the boundary conditions would *in principle* provide enough information to derive all the macroscopic phenomena from the underlying laws. On the other side, an "emergentist" might argue that it is not the fundamental laws themselves, but "emergent" higher level concepts, which are central to our understanding of the world around us. Both are right, and both are in danger of neglecting the essential facts that might marry their seemingly different viewpoints.

Two obstacles lie between a strong reductionist viewpoint and any useful application. First, the fact that the explicit and detailed knowledge of the boundary conditions, which are necessary to analytically derive macroscopic phenomena from the microscopic laws, are consistently absent in practical situations. Only in very exceptional situations, far

removed from day to day reality, and specifically set up in physics laboratories, are the boundary conditions sufficiently controlled for the fundamental laws to be able to make predictions of behavior that are testable.

Second, even if all boundary conditions are known specifically, it still may be computationally infeasible to derive predictions of macroscopic behavior from the fundamental laws. That is, one might in principle write down the equations whose solutions would give the desired predictions, but solving these equations might be impossible either analytically or numerically. Both these obstacles occur as a rule rather than as an exception. For instance, the common occurrence of deterministic chaos in nonlinear differential equations shows that uncertainties in the boundary conditions, no matter how small, may be amplified so rapidly that detailed predictions from knowledge of the underlying microscopic dynamics are generally impossible.

Considering these facts, the "emergentist" is right that the main problem is to find higher level representations of macroscopic systems that are able of making useful predictions. First, the variables in these higher level representations should be open to direct measurement, unlike, for instance, the joint quantum state of all atoms and electrons in a system. And second, one should be able to make predictions of macroscopic behavior in terms of these variables.

The search for an informative representation is however not independent of the fundamental laws. It would be foolish to ignore that the macroscopic phenomena that we see around us are rooted in the fundamental laws of physics. This is, of course, the sense in which the reductionist is right. Any description on an "emergent" higher level should in the end be understood in terms of the underlying laws. This however does not imply that this higher level *description* is somehow contained in the fundamental laws. The situation is similar to after just being taught the rules of the game of Go, trying to understand the play of two Japanese professionals. From the reductionist point of view, even the most subtle play in the game of Go doesn't require any other initial information than its rules. On the other hand, the process of taking these rules and *unfolding* them, so to speak, towards an understanding of high level behavior (i.e. professional game play), is so vastly complicated and rich, that it has been studied by Go scholars for centuries, and has filled hundreds and hundreds of books. The contents of these books, for instance, are not contained in the rules of Go themselves, they *emerge* through human investigations of the consequences of the rules in *playing* Go.

The topic of this chapter is then to briefly discuss the methodologies that have been developed to relate underlying microscopic laws to macroscopic behavior.

2.1 Macroscopic Laws

Many of the common "laws" of physics taught in high school are very far from the fundamental microscopic laws but are in fact macroscopic "emergent" laws. For the purpose of illustrating the ubiquity of macroscopic laws, I will discuss a couple of such macroscopic laws, how they represent a system's state, and how they relate to the underlying microscopic laws.

2.1.1 Elasticity: Hooke's Law

Imagine a beam of some material that has a length L when no forces are exerted on it. We then start pulling or pushing on this beam, with a force F. Hooke's law states that the change in length of this beam ΔL is proportional to the force F that is exerted:

$$F = k\,\Delta L. \tag{2.1}$$

The constant k is called the *force constant* and is material dependent. Instead of representing the state in terms of the locations of all molecules in the material, let alone in terms of some joint quantum state, the above formula relates its macroscopic length to a macroscopic "force" that is being exerted. This macroscopic law can be understood in terms of the forces between the molecules in the material that resist being pushed together or pulled apart. This resistance in turn should eventually be explained in terms of the quantum electrodynamics of the electrons in the solid. Deriving Hooke's law, and its limitations, from quantum electrodynamics is hardly possible however. The interactions between the molecules in the material are highly complex in general and Hooke's law is only an approximate and largely phenomenological law. For small forces and deformations it does work to some extent, and it nicely explains the common occurrence of oscillations in mechanical systems.

2.1.2 Electric current: Ohm's Law

Ohm's law states that the local current density in a material is proportional to the local electric field. In a more familiar form, it states that the current I through a wire in a circuit is proportional to the Voltage drop V over the wire:

$$V = I\,R, \tag{2.2}$$

where R is the *resistance* of the wire. Again, Ohm's law doesn't formulate the behavior of the system in terms of its microscopic constituents such as the electrons moving through the wire, or the atoms that make up the wire. Similarly, the resistance R is a macroscopic quantity that cannot be easily derived from microscopic laws—i.e. quantum electro dynamics in this case. Note also that it isn't obvious at all from the microscopic laws that (for most materials) there is a material dependent *constant* that relates the electron flux through the material to the potential difference over the material. Of course, Ohm's law is only an approximate law. It does not work for many materials, such as semi-conductors, and it is in blatant contradiction with phenomena such as super-conductivity. It is also an oversimplification to call resistance a constant. Resistance is known to be dependent on external conditions such as temperature. But in many situations, Ohm's law works fine, and electrical circuits can be constructed using it. In this sense, Ohm's law is much more useful than the fundamental laws when one is concerned with simple applications.

2.1.3 Diffusion: Fick's Law

Fick's law relates the time derivative of the concentration of particles at a certain location to the Laplacian of the concentration at that location:

$$\frac{\partial C}{\partial t} = D\,\nabla^2 C, \tag{2.3}$$

where D is the *diffusion coefficient*. The above macroscopic law describes how concentrations of particles smear out over a volume over time. Relating this macroscopic law to the underlying microscopic motions of the particles is somewhat less straightforward than in the case of Ohm's law. There it was still possible *in principle* to derive Ohm's law from quantum electro dynamics. In contrast, the original derivation of the above law, given by Einstein, might be considered one of the first examples of *statistical mechanics*. It involves assuming that collisions between a reference particle and background particles randomize the movements of this reference particle. One then calculates the probability distributions of where a reference particle might end up after a short period of time. Notice that this derivation involves the introduction of *random* influences on the reference particle. Although the laws of Newtonian mechanics in principle give a deterministic description of the movement and collisions of all particles, the macroscopic law can only be derived by introducing some averaging procedure that effectively treats the movements of a reference particle as random. This is an essential ingredient of the statistical mechanical approach: details of the microscopic state are modeled as *random* in order to obtain the macroscopic behavior. This is something different than what could be called a derivation of the macroscopic behavior from the underlying microscopic laws. It additionally involves somehow taking into account the locations and velocities of all particles without any specific knowledge of them. A general procedure for doing this will be described below.

2.1.4 Ideal Gas: Equation of State

Lastly, the ideal gas law

$$PV = nkT, \tag{2.4}$$

relates the macroscopic variables pressure P, volume V, temperature T, and particle number n to each other. The constant k is Boltzmann's constant. In this equation, all variables that appear are clearly macroscopic variables. The particle number n and the pressure P can both be understood in a straightforward way from the microscopic variables, pressure just being the sum of the forces applied to the walls by all the microscopic particles per unit of area. The temperature T, however, has a much more complicated interpretation—for instance, it is often interpreted as the average energy per degree of freedom in the system. Again, deriving this law from the microscopic intuition of a large number of particles bouncing around in a box involves assuming some *randomness*. Doubling the size of the box makes the particles hit the walls half as often, making the pressure half as large. This for instance assumes that particles don't cluster together in some complicated distribution but spread randomly through the box. For particles of an ideal gas, the only degrees of freedom are their velocities. Since the particle's energy is proportional to the *square* of its speed, doubling the temperature roughly amounts to

multiplying the particle speeds by $\sqrt{2}$[1]. Thus, particles hit the walls $\sqrt{2}$ times as often, but also $\sqrt{2}$ times as hard, leading to a doubling of the pressure.

2.2 Equilibrium Statistical Mechanics

We have seen above that many of the (older) laws of physics are in fact macroscopic laws that relate macroscopic quantities to each other. For most practical cases, these macroscopic (approximate) laws are of much more use than stating the detailed microscopic laws. If one builds a circuit one is not generally interested in the quantum wave functions of all the electrons that provide the current. Moreover, it is experimentally entirely infeasible to measure the state of the system in such detail. Similar considerations of course apply to gas particles in a box. The essential problem then becomes how to deal with a lack of *information*. That is, how is one to infer predictions of macroscopic quantities given that one lacks knowledge of the detailed microscopic state of a system but only has knowledge of the values of some macroscopic quantities? Stated in this way, it becomes clear that the problem is not that of constructing some new *physical theory* but rather simply a problem of *inference* from incomplete information. What is needed is a combination of the information available from the microscopic laws and the information about the state of its system in terms of some macroscopic quantities to predict other macroscopic quantities of interest. It turns out that there is a definite and very straightforward procedure that describes how this statistical inference is to be performed. This procedure goes under the name of the "maximum entropy method". Below we will introduce this general method and show how equilibrium statistical mechanics provides an example of this method.

2.2.1 The Maximum Entropy Method

Consider a system with N possible microscopic states $i \in \{1, 2, \ldots, N\}$. We want to make predictions about measurable macroscopic quantities of this system, given that we only have some very limited knowledge of the system's state. For instance, we only know that some quantity f—such as the total current through a wire, or the pressure of a gas—is fixed at a certain value. Alternatively, we even might not know that f is fixed, but only know its average value $\langle f \rangle$. We further know that for each possible microstate i the quantity f takes on some value f_i. For instance, we know the pressure of the gas given the exact locations and velocities of all particles contained in it. Other than that, we have no information whatsoever about which of the microscopic states i the system is more or less likely to be in. We now want to predict the value $\langle g \rangle$ of some other observable quantity g. To this end we will have to construct a distribution $\{p_i\}$ of probabilities to find the system in a microscopic state i. Since N tends to be very large for physical systems, we have to determine a huge number of probabilities $\{p_i\}$ and this may seem an impossible task given that we only have *one* constraint (the value $\langle f \rangle$) to go by. This would indeed be a problem if we were to analytically *deduce* the probabilities $\{p_i\}$, but what we want to do amounts to statistically inferring the most

[1] Note that this reasoning implicitly assumes some features of the distribution of particle speeds.

reasonable assignment of probabilities $\{p_i\}$ given our knowledge $\langle f \rangle$. The maximum entropy method solves this problem by prescribing that one **choose that assignment of probabilities $\{p_i\}$, which is consistent with our knowledge about f, and has the highest "uncertainty" associated with it**.

The uncertainty of a distribution $\{p_i\}$ is quantified by its *entropy* H which is defined as

$$H = -k \sum_{i=1}^{N} p_i \log(p_i), \tag{2.5}$$

where k is a proportionality constant. The entropy H is the central quantity in information theory, as originally developed by Shannon [127, 128]. It quantifies our degree of ignorance, or lack of knowledge or uncertainty, about the specific microscopic state i of the system given the probability distribution $\{p_i\}$. Viewed slightly differently, it quantifies the amount of information one would obtain if one were to be told what microscopic state i the system is actually in. One can think of it as proportional to the average number of independent yes/no questions that need to be answered before one knows the exact microscopic state of the system. If one, for instance, knows the exact microscopic state i of the system, the entropy is zero. If one knows nothing whatsoever about the system's state, the entropy equals $k \log(N)$. In appendix A entropy and some related information theoretic concepts are introduced. There it will also be shown that the functional form (2.5) is unique given a small set of axioms.

If one accepts this interpretation of entropy, the prescription to maximize it becomes nothing other than common sense. If one were to choose a distribution $\{p_i\}$ that does *not* maximize the entropy, one would choose a distribution that has less uncertainty associated with it than the maximum entropy distribution. Such a choice amounts to making additional implicit assumptions about the state of the system which are not warranted based on the information that we have. The maximum entropy prescription is nothing else than demanding that we honestly confess our ignorance to the fullest of its extent: Know thy ignorance!

We can now apply this method to the problem of predicting the expected value $\langle g \rangle$ of a quantity g. If we know that the system has some fixed value f_c for the variable f, we will find a maximum entropy distribution $\{p_i\}$ which is zero for all i that have $f_i \neq f_c$, and *uniform* over all i that have $f_i = f$. As was of course to be expected, all microstates i that have a value f_i which does not equal f_c are excluded. That one should take a uniform distribution over the other states is of course intuitive but not entirely obvious. Equation (2.5) has the feature that it is maximized when all (nonzero) p_i are taken equal. Our prediction $\langle g \rangle$ for the quantity g becomes in this case:

$$\langle g \rangle = \sum_{\{i | f_i = f_c\}} \frac{g_i}{N(f_c)}, \tag{2.6}$$

where $N(f_c)$ is the number of microstates i that have $f_i = f_c$, and g_i is the value of the quantity g for microstate i.

In the case where we only know the average value $\langle f \rangle$ of the quantity f, the maximum entropy distribution $\{p_i\}$ can be found using the method of Lagrange multipliers

which is explained in appendix B. The answer is

$$p_i = \frac{1}{Z(\lambda)} e^{-\lambda f_i}, \tag{2.7}$$

where the normalization constant $Z(\lambda)$ is given by

$$Z(\lambda) = \sum_i e^{-\lambda f_i}. \tag{2.8}$$

The parameter λ is determined by demanding that

$$\sum_i f_i p_i = \langle f \rangle. \tag{2.9}$$

Our prediction for $\langle g \rangle$ now becomes

$$\langle g \rangle = \frac{1}{Z(\lambda)} \sum_i g_i e^{-\lambda f_i}. \tag{2.10}$$

In a certain sense, this is all there is to statistical mechanics. For example, an isolated physical system cannot exchange matter or energy with its surroundings. Its energy E is therefore constant. In this case, given that the energy of the system is E, the maximum entropy method prescribes that *all* microstates i with this energy E are equally likely to occur. The maximum entropy distribution is a uniform distribution over all states i that have energy E. In statistical mechanics, this is known as the *micro-canonical* distribution.

In the case where a system is closed (no exchange of matter but possible exchange of energy) one generally only knows the average energy $\langle E \rangle$ of the system. In this case, the maximum entropy distribution is given by equation (2.7), with f_i in the role of the energy of state i. The parameter λ is inversely proportional to *temperature* in this case. This distribution is commonly referred to as the Boltzmann distribution. In cases where the system can exchange matter as well as energy, one would generally condition the distribution on the average energy $\langle E \rangle$ and the average number of particles $\langle n \rangle$. In this case, the distribution looks like

$$p_i = \frac{1}{Z(\beta, \mu)} e^{-\beta(E_i + \mu n)}, \tag{2.11}$$

where the standard notation $\beta = 1/(kT)$ for inverse temperature and μ for *chemical potential* is being used. Essentially all distributions over microstates that are commonly used in statistical mechanics can be derived in this almost trivial way. One specifies the macroscopic variables that are used to describe a system and derives the maximum entropy distribution given their average values. Other quantities (such as specific heats etc.) can then be derived by using the maximum entropy distribution.

Just to mention some more examples of maximum entropy distributions: If a quantity x can take on values in $\mathbb{R}$, and we only know its mean μ and variance σ^2, its maximum entropy distribution is a *Gaussian* distribution: $p(x) \propto \exp((x-\mu)^2/\sigma^2)$. More

generally, if we only know the average values of functions $f_i(x)$ of the variable x, the maximum entropy distribution is given by $p(x) \propto \exp(\sum_i c_i f_i(x))$, where the c_i are chosen such as to obey the contraints $\langle f_i(x) \rangle = f_i$. For example, if we only know the average *scale* $\langle \log(x) \rangle = s$ of a positive variable x, the maximum entropy distribution is a power-law distribution: $p(x) \propto x^\alpha$, where the exponent α is chosen such that $\langle \log(x) \rangle = s$. Appendix B describes a general method for deriving such maximum entropy distributions.

Why would predictions based on maximum entropy distributions be successful? Note that the maximum entropy methodology does not guarantee success. It might fail for several reasons. First of all, we predict the average $\langle g \rangle$ of a quantity g from the maximum entropy distribution, but if the variance $\langle g^2 \rangle - \langle g \rangle^2$ (calculated from the maximum entropy distribution as well) is large compared to the mean, we cannot expect our prediction for $\langle g \rangle$ to be accurate for any particular case. Thus, accurate predictions for $\langle g \rangle$ can only be expected when the variance is much smaller than the mean. For the systems that are typically studied in statistical mechanics the size of the variance relative to the mean scales (away from phase transition points) with the inverse of the square root of the number of particles. Since the number of particles in these systems is very large (order 10^{23}), the predictions are typically very accurate.

However, even if the relative variance is small, it may still turn out that the predictions give wrong answers. In this case, we are forced to conclude that our counting of the microstates is wrong. For instance, if an experimenter derives maximum entropy distributions for a box of gas based on volume and particle number alone, his predictions are not likely to agree with experiments. Moreover, the outcomes on hot summer days will turn out to be consistently different from the outcomes on cold winter days. His predictions failed because he has taken into account microstates with very different mean energies (temperatures) while the mean energy of his system is a well defined macroscopic quantity that varies with temperature. If he had taken the mean energy or temperature into account, he would have obtained correct predictions.

One might think of this as a weakness of the theory but in fact it is one of its strengths. The failure of the predictions, even when the relative variance was small, points us to additional macroscopic variables that have to be taken into account. In fact, new laws of physics have been discovered in the past as the result of such discrepancies. It was found, for instance, that the number of different microstates with N identical particles each having the same energy is not $N!$ but 1 for certain types of particles. This counterintuitive counting of states was eventually explained by the theory of quantum mechanics. Even the birth of quantum mechanics can be traced back to the realization that in order to obtain correct predictions, one has to assume that certain variables cannot take on continuous values but are "quantized".

In summary, we use our microscopic knowledge of the system under study to construct the set of all possible microstates and to express macroscopic variables as functions of the microstates. That is, we assume that we know all microstates that are possible, and that given a specific microstate we can calculate any macroscopic variable associated with that microstate. We then choose a set of macroscopic variables and derive the maximum entropy distribution conditioned on the values of these macroscopic variables. After that, we can make predictions of other variables by averaging over the maximum entropy distribution.

2.2.2 Is that all?

At this point, it is helpful to note that this view of statistical mechanics is by no means standard in the scientific community in general and the physics community in particular. In the view presented here, the maximum entropy distributions are obtained as a result of *statistical inference* based on partial information (for instance, the mean energy). This informational view of statistical mechanics contends that, given only knowledge of the mean energy of the system, we cannot do anything other than choose the maximum entropy distribution for making predictions about other macroscopic quantities. The maximum entropy distribution corresponds to the "most ignorant" distribution given the partial information. Any other distribution over the microstates would introduce unwarranted biases.

Most physicists would however probably subscribe to the view that the success of statistical mechanics merely shows that *apparently* the maximum entropy distribution gives an accurate description of the actual distribution over microstates that occurs in physical systems. From this point of view, one obviously has to justify the use of the maximum entropy distribution by independent arguments. From the statistical inference point of view, no such justification is necessary; the maximum entropy distribution is simply the least biased way to deal with a lack of information. But from the more commonly held point of view, one would have to show that the maximum entropy distribution is the actual distribution with which systems visit their microstates.

Therefore, the proponents of this "objective" view are engaged in a program which tries to show that the sort of physical systems that are studied in statistical mechanics generally evolve towards the maximum entropy distribution. One for instance tries to show that particles bouncing around in a box form an *ergodic* system. That is, for long times, the time average of the microstates that such systems visits is equal to the maximum entropy distribution independent of the initial distribution over microstates. This *ergodic program* has not been entirely successful due to some fundamental difficulties. For instance, since the laws of physics are reversible, one can uniquely determine the microstate of the system arbitrary amounts of time in the past or future given the microstate of the system at any particular time. This implies that the entropy of a distribution over microstates can only be *constant* over time. One of the ways of dealing with this problem is to calculate the entropy with respect to some *coarse graining* of the space of microstates. Although one assumes that the momenta and positions of particles can still take on a continuum of values, one only calculates the entropy with respect to the frequencies of particles in discrete *bins* of momentum and position. However, even under these ad hoc simplifications, only a small class of simplified systems can be shown to evolve towards the maximum entropy distribution with respect to these bin-frequencies.

In contrast to the objective view, the informational view provides a conceptually simple justification for use of maximum entropy distributions. It however does not specify how one should choose the set of macroscopic variables with respect to which the maximum entropy distribution is calculated; it can be applied to any set of macroscopic variables. The question: "Which macroscopic variables should one choose?" does not have a simple general answer. There are, however, some constraints and qualitative considerations to go by. First, one generally chooses macroscopic variables that can be easily measured or controlled. Additionally, the macroscopic variables have to be suf-

ficient in the sense that definite predictions of other variables can be obtained from the maximum entropy distributions. For instance, if we don't condition the maximum entropy distribution on macroscopic variables at all, we obtain a uniform distribution over all microstates. In that case, our predictions might have a variance associated with them that is not small at all relative to the mean and we cannot expect our predictions to be accurate.

In equilibrium statistical mechanics one is interested in the properties of the system that are constant over time. A macroscopic variable that fluctuates wildly over time is therefore not very suitable; measuring it at one time doesn't provide a good prediction for its value at some other time. The microscopic laws tell us that energy is a conserved quantity. This implies that it cannot change over time when the system is isolated and therefore provides a useful macroscopic quantity. Momentum and angular momentum are other well known conserved quantities. Why are these not included in the common description of systems that are typically studied in statistical mechanics? The reason is simply that they are zero for systems that are not moving or rotating macroscopically, and this is typically true for the systems studied. However, some systems studied in astrophysics do in fact move and rotate macroscopically. For these systems the momenta and angular momentum do appear as macroscopic variables in the statistical mechanical description.

Choosing an appropriate set of macroscopic variables is further complicated by the fact that a physical system does not generally have *one* set of macroscopic variables that describes its behavior adequately for all parameter settings. That is, under a change of external parameters, the same physical system may occur in different "phases" that require different sets of macroscopic variables to describe it. This phenomenon will be discussed in the next section.

2.3 Phase Transitions

It is of course well known that systems in nature can exhibit different "phases" and that by changing some macroscopic variables, a system can be made to undergo phase transitions between these different phases. The boiling and freezing of water are probably the most commonly observed examples of such phase transitions. There are many fascinating phenomena involved with phase transitions but in this section we will only focus on those features of phase transitions that bear on the evolutionary phenomena studied later on.

In the previous sections it was pointed out that under certain conditions the fluctuations in a macroscopic quantity as calculated from the maximum entropy distribution might become large relative to its mean, calculated from the same maximum entropy distribution. In general, a phase transition can be associated with such large fluctuations in one or more macroscopic quantities. Under a change in some macroscopic control variable (such as temperature) the fluctuations in some other macroscopic variables suddenly become very large. The control parameters for which the fluctuations "blow up" define the points where the phase transition occurs. As was also pointed out in the previous sections, the fact that fluctuations in some macroscopic quantity are large indicates that we need additional macroscopic variables to describe the state of the system. In

statistical mechanics, these new macroscopic quantities that emerge through a phase transition are generally referred to as *order parameters*. As will be shown below, they are often associated with the spontaneous *breaking of a symmetry* that was present in the system before the phase transition.

The simplest and most illustrious example of the breaking of symmetry and the appearance of a new macroscopic variable through a phase transition is probably the transition from a paramagnetic to a ferromagnetic phase in ferromagnetic materials. A well studied prototype model for such ferromagnetic systems is the Ising ferromagnet model. It consists of a one, two, or three dimensional square lattice with a microscopic magnetic spin located at each site. Each magnetic spin can occur in two states only: up or down. The microscopic state of the system is given by specifying which spins are up and which spins are down. Spins only interact with their neighbors on the lattice. Each pair of neighboring spins has a contribution to the energy of the system that depends on the relative orientation of the spins. A pair of aligned spins contributes less than a pair of anti-aligned spins. The energy E_s of a microscopic state s is thus only dependent on the number of aligned and anti-aligned pairs of neighboring spins. The equilibrium (maximum entropy) distribution at a temperature $T = 1/(k\beta)$ is given by the Boltzmann distribution $p_s = \exp(-\beta E_s)/Z$. This equation in principle contains all necessary information regarding the equilibrium properties of the Ising ferromagnet.

With respect to the symmetries of this system, one obviously notes that the system possesses a spin flip symmetry. That is, the energy E_s of a state s stays the same when all spins are reversed $s \rightarrow \bar{s}$. Since the probability p_s of a microstate s depends only on its energy E_s, it follows that a state s has the same probability of occurrence as its spin reverse $\bar{s}$, i.e. $p_s = p_{\bar{s}}$. This symmetry guarantees that the expectation value of the macroscopic magnetization M, which can be roughly defined as the difference between the number N_+ of spin "up" pairs and the number of spin "down" pairs N_- is zero:

$$\langle M \rangle \equiv \langle \frac{N_+ - N_-}{N_+ + N_-} \rangle = 0. \tag{2.12}$$

For every state s with a magnetization M_s, there is an equally probable opposite state $\bar{s}$ with the opposite magnetization $M_{\bar{s}} = -M_s$.

One might expect that equation (2.12) implies that the Ising ferromagnet will never exhibit any macroscopic magnetization. For high temperatures (small β), this is indeed the case. However, lowering the temperature to some critical temperature T_c, the fluctuations $\langle M^2 \rangle$ can suddenly become very large[2]. When this happens, we can no longer expect our theoretical predictions ($\langle M \rangle = 0$) to be accurate in the sense that any particular example of an Ising ferromagnet should have no macroscopic magnetization. What happens is that for high temperatures, the Boltzmann distribution is concentrated around microstates s that have no macroscopic magnetization. There will be regions where most spin pairs are up, but there will be roughly equally many regions where spin pairs are mostly down making the total magnetization roughly zero. For low temperatures however, the Boltzmann distribution has two peaks; one at a magnetization $M = +1$ and one at a magnetization $M = -1$. The mean magnetization $\langle M \rangle$ is still zero, but its variance is large, i.e. $\langle M^2 \rangle = 1$.

[2]For a 1-d Ising ferromagnet this strictly only occurs at $T_c = 0$, which makes the 1-d case somewhat degenerate.

For real ferromagnets one observes that as the temperature falls below T_c, the system spontaneously acquires either a positive or a negative magnetization. What happens is that the regions in which most spin are aligned become as large as the magnet itself. That is, either the spin up regions take over the entire magnet, or the spin down regions will take over the entire magnet. The system thus undergoes a phase transition from a state with magnetization zero, to a state with either positive or negative magnetization. At any point in time, and for any particular system, one will find the system *either* in the positive magnetization state or the negative magnetization state. The macroscopic prediction $M = 0$ no longer describes the physical situation accurately as was to be expected from the large variance $\langle M^2 \rangle$. However, if we introduce one additional macroscopic parameter we can again accurately describe the system's macroscopic behavior. We simply need to add to the description a *directional* variable that tells whether the system has positive or negative magnetization. For systems with for instance positive magnetization, all states with negative magnetization are excluded and the Boltzmann distribution will thus again be concentrated in a single peak.

Obviously, this phase transition has a spontaneous symmetry breaking associated with it. The symmetry between positive and negative magnetization states gets broken. Below the critical temperature, tiny fluctuations in magnetization away from zero will be amplified into a large macroscopic magnetization that gets fixed in the ferromagnet. Once the system has developed a large magnetization, it is very unlikely to spontaneously revert to the opposite magnetization. At that point, we need a new macroscopic variable (magnetization direction) to accurately describe the macroscopic behavior of the system. This is typical of phase transitions (especially "continuous" phase transitions): Fluctuations in a macroscopic quantity become macroscopic themselves, a symmetry of the system is broken, and a new macroscopic variable, often called an order parameter, appears. Of course, the situation is reversed when the phase transition is approached from the opposite direction, i.e. increasing the temperature in the case of the ferromagnet. In more complicated cases, more than one new macroscopic parameter may appear at the same time. That is, more complicated symmetries might be broken. There are even cases in which infinitely many new macroscopic variables may appear at once.

Finally, it should be noted that the symmetry breaking and phase transition are induced by a change in a control parameter of the system, such as temperature. In the evolutionary setting to be described later on, the symmetry breaking occurs *dynamically* due to the internal dynamics of the system. In that context, the occurrence of a new macroscopic variable corresponds to the appearance of a new (functional) phenotypic trait in the evolving population.

2.4 Non-equilibrium Statistical Mechanics

Until now, the discussion of statistical mechanics has focused on the behavior of systems in *thermodynamic equilibrium*. Thermodynamic equilibrium is characterized by a stationary state of the system in which no macroscopic flows between the system and its environment are present. The methodology of statistical mechanics, as explained in the previous sections, has proven to be able to accurately describe the macroscopic behavior

of systems in equilibrium. However, as soon as one moves away from the equilibrium situation, things start to become rather complicated if one takes the common view that the maximum entropy distribution somehow corresponds to the actual "objective" distribution over microstates.

One will be faced with the problem of having to somehow *derive* new distributions over microstates for systems out of equilibrium. For equilibrium, it is known that the maximum entropy distributions provide accurate predictions, but for non-equilibrium systems it is not known which distributions to use. Apart from first-order perturbations around an equilibrium state [30], almost no general methods have been developed to derive non-equilibrium microstate distributions. Further, the ergodic hypothesis is entirely unsuited for non-equilibrium situations, since one is generally interested in *dynamics*. For equilibrium situations, the ergodic hypothesis justifies using the maximum entropy distribution by assuming that averages calculated with the maximum entropy distribution give the same results as *time averages* over the stationary macrostate. However, this justification precludes calculation of the *dynamics* of macroscopic variables. In short, the "objective" view that the maximum entropy distribution corresponds to the actual physical distribution over microstates, renders non-equilibrium problems almost inapproachable.

In contrast, if one takes the informational view advocated in section 2.2.1, non-equilibrium situations cause considerably fewer problems. For example, there are no conceptual difficulties. The method remains essentially unchanged: One assumes complete knowledge of the microscopic state space and microscopic dynamics of the system under study, and only partial information on the specific state that the system is in. In the dynamic, non-equilibrium situation, this information may be in the form of the averages of some macroscopic quantities at different times, or knowledge of certain macroscopic flows or gradients in the system. One then again constructs the maximum entropy distribution $p_s(t)$, which now depends on time as well. Other macroscopic quantities of interest can then again be predicted using this maximum entropy distribution. This general approach to non-equilibrium statistical mechanics was introduced in the 1970s by Jaynes [84], similar treatments can be found in [5, 55, 131, 132].

To make the ideas more precise, consider the case in which the average values $f(t_i)$ of some macroscopic variable f are given for a set of times $\{t_i\}$ with $i = \{1, 2, \ldots, n\}$. We set the origin at the first time $t_1 = 0$. For each of these times, the average $\langle f \rangle$ is thus given by some number $f(t_i)$. Additionally, for any microstate s, we can uniquely predict its microscopic time behavior

$$s(t) = G\left[s(0), t\right], \tag{2.13}$$

where the operator G represents the microscopic equations of motion of the system. From this information we derive a maximum entropy distribution $p_s(0)$ over microstates at time $t = t_1 = 0$. By using equation (2.13), we can then obtain the distribution $p_s(t)$ for arbitrary times. Note that this distribution $p_s(t)$ is generally *not* a maximum entropy distribution with respect to the macroscopic variable $f(t)$ at t. That is, if we had constructed the maximum entropy distribution based on the information that $\langle f \rangle$ equals $f(t)$ at time t, we would have generally obtained a different maximum entropy distribution $p'_s(t)$.

Constructing the maximum entropy distribution $p_s(0)$ is done in complete analogy to the methods described in section 2.2.1. Using the equations of motion (2.13), the constraints take on the form

$$\sum_s p_s(0) f_{G[s,t_i]} = f(t_i), \tag{2.14}$$

where f_s denotes the value of the quantity f for microstate s, such that $f_{G[s,t_i]}$ denotes the value of f at time t_i given that the system was in state s at time $t = 0$. Thus, given the initial distribution $p_s(0)$ at time $t = 0$, we demand that the average values $\langle f \rangle$ at the different times t_i are given by $f(t_i)$. But in this form, the problem is exactly the same as the ones described in the appendix B. From a mathematical point of view, averages of the same quantity at different times are equivalent to a set of constraints given by averages of different quantities at one particular time. The averages $\langle f \rangle$ at different times can be inferred from $p_s(0)$ using the equations of motion (2.13). Therefore, the maximum entropy distribution is simply given by

$$p_s(0) = \frac{1}{Z} exp\left(-\sum_i \lambda_i f_{G[s,t_i]} \right), \tag{2.15}$$

where Z is again a normalization constant and the Lagrangian multipliers λ_i are chosen such that the constraints (2.14) are all satisfied.

Intuitively, the maximum entropy method distributes probability as uniformly as possible over all states that are consistent with the macroscopic information that is provided. In this time-dependent case, this means that the distribution $p_s(0)$ gives high (uniform) probability to all states s that exhibit values of the macroscopic quantity f which are consistent with the values $f(t_i)$ at each of the times t_i for which the macroscopic quantity f was specified.

When the maximum entropy distribution $p_s(0)$ has been determined, it can be used to predict the dynamics of other macroscopic quantities. For instance, for any time t, the predicted value $\langle g(t) \rangle$ for the average of some macroscopic variable g is given by

$$\langle g(t) \rangle = \sum_s p_s(0) g_{G[s,t]}. \tag{2.16}$$

Using such equations, we may derive macroscopic equations of motion for macroscopic variables such as $g(t)$ from the maximum entropy distribution.

Notice that in this formalism, predictions for future time are not necessarily only dependent on the values of the macroscopic quantities at the current time, but may depend on the values of the macroscopic variables in the past as well. This contrasts with the microscopic dynamics as generated by the operator G. There, the future microscopic state of a system is always only dependent on the current microscopic state[3]. It may seem counterintuitive that the dynamics on a macroscopic level exhibits such *memory effects* while the underlying microscopic laws do not. The maximum entropy method is however not a law of physics, it is a law of statistical inference. Significantly, this may easily lead to memory effects. In particular, it will typically be impossible to derive

[3]This holds in particular for all the fundamental laws of physics.

differential equations for the macroscopic quantities such as $g(t)$. If one wants to obtain memoryless equations of motion for the macroscopic variables, one has to make further approximations. This will be done in the next section.

2.4.1 Memoryless Approximation

In many cases, the formal microscopic equations of motion (2.13) can be rather complicated, since G is often nonlinear. Moreover, deriving macroscopic equations of motion using equations such as (2.16) can be rather tedious. In those cases, we might want to make some additional assumptions to simplify the analysis.

One possible approach is to assume that the dynamics for the macroscopic variables of interest is slow compared to the dynamics on the microscopic level. Under that assumption, one can construct the dynamics on the level of the macroscopic variables by assuming that the microscopic maximum entropy distribution remains valid at all times. That is, given that we know the value $f(t)$ of the macroscopic variable f at time t, we have a maximum entropy distribution $p_s \propto \exp(-\lambda f_s)$ at time t, with mean $\langle f \rangle = f(t)$. From this, we derive the value f at the next instant $t+dt$, by integrating the equations of motion (2.13) for a very small time interval dt. From this we find a small change df in the macroscopic variable f, i.e. $f(t) \to f(t) + df$ in the time dt. This is essentially the same procedure as outlined before. However, at this point we reassume a new maximum entropy distribution $p_s \propto \exp(-\lambda' f_s)$ with mean $\langle f \rangle = f(t) + df$. We then repeat the procedure and, in this way, iteratively derive the macroscopic dynamics of the variable f. One can think of this approximation as neglecting any memory of the state at time t in the distribution at time $t+dt$ apart from the value $f(t)+df$ of the macroscopic variable f. This absence of memory makes this approximation appealing. In particular, we are able to derive phenomenological equations of motion for the macroscopic variable $f(t)$. In this thesis, we will take this memoryless approximation in predicting macroscopic variables for an evolving population.

It is important to note once more that it is not necessary that the maximum entropy distributions somehow accurately describe the "real" distributions of the non-equilibrium dynamical systems. In general, we are not specifically interested in the actual microscopic distributions but are only interested in predicting some macroscopic observables. The maximum entropy distribution might be inaccurate in the sense of not predicting the actual microscopic distribution but might still accurately predict the macroscopic observables.

The same argument holds for assuming no memory and reassuming maximum entropy after each time increment. This assumption is typically provably wrong from a microscopic point of view. But this does not necessarily mean that it produces inaccurate predictions for macroscopic observables. Roughly speaking, the maximum entropy distribution at time $t+dt$ might differ from the distribution obtained by direct integration of the microscopic equations of motion from the maximum entropy distribution at time t by assigning too much probability to certain states and too little to others. If, with respect to predicting a macroscopic quantity g, the states with "high g" and the states with "low g" are spread roughly equally among the states that obtain too much and too little probability, the predictions for the quantity g will not be affected. In order for the predictions to fail, the states with "high g" should consistently fall in the set of states that

obtains too much probability or fall consistently in the set of states that obtains too little probability. The fact that this does not often occur in practice accounts for the robustness of the maximum entropy approach. Notably: in cases where the approach does fail, it often points the way to interesting relations between certain macroscopic variables.

2.5 The Second Law

One of the most frequently discussed laws of physics is the second law of thermodynamics. It comes in many different varieties and is often presented as an extremely fundamental law that describes a natural tendency common to all processes in nature. It is not uncommon to find a description of the second law along the lines of: Whatever happens in the universe, its entropy can only go up. Such a colloquial description can even be found in the famous *Lectures on Physics* by Richard Feynman [43]. These generalized interpretations of the second law have lead to conceptual misunderstandings, especially regarding its relevance for biology and evolution. One can think of another authority, Erwin Schrödinger, and the concept of negative entropy, in this regard [125]. In combination with the conceptualization of entropy as representing disorder, these ideas have even lead some to claim that evolution as a whole is at odds with the second law of thermodynamics, since order seems to spontaneously arise through evolution—something which in the popular view is prohibited by the second law.

Since evolution is the topic of this thesis, and since some methodology borrowed from statistical mechanics is used in this thesis as well, it seems appropriate to briefly discuss these misconceptions. The second law is, in fact, far more restricted and of limited applicability than is often suggested.

In the original version of the second law, which will be discussed below, it is stated that under a certain class of operations on a system, the thermodynamic entropy of the system at the end of the operations cannot be smaller than its initial entropy. The first thing to note here is that thermodynamic entropy as such is only defined for systems in *thermodynamic equilibrium*. This immediately implies that the second law applies only to processes for which the system is in equilibrium both at the beginning and end. For systems that are out of equilibrium, and continue to remain out of equilibrium, the second law simply does not apply. Living systems in particular are very far from equilibrium—since they, for instance, are sustained by energy fluxes—and the second law therefore does not speak to them. The same holds for the biosphere as a whole. The periodic heating due to sunlight and cooling at night guarantees that the biosphere can never reach a state of equilibrium as long as the sun is burning.

We could stop our discussion of the second law here, but it is instructive to briefly review how it arises and what its general implications for non-equilibrium dynamics and evolution might be. In the following, we will again present the informational view on these matters in which the entropy of a system measures the amount of uncertainty about the precise microstate of a system, given knowledge of some of its macroscopic variables. This contrasts with the "objective" view that the entropy of a system corresponds to a *physical property* of that system.

2.5.1 Definitions

Let's first make a precise statement of the second law. We will use the traditional and commonly taught statement due to Clausius. Before the law can be stated, we will need a definition of entropy. Historically, the discovery of entropy as a state variable can be traced back directly to Carnot's principle that the most efficient heat engine must be a reversible engine, as lucidly explained in [85]. If a reversible engine wasn't the most efficient engine, one could use a combination of a more efficient engine and a reversible engine which is running backwards to create a perpetuum mobile. Kelvin then showed how the efficiency of such reversible engines is a function only of the temperatures of the hot and cold reservoirs between which it operates. One can generalize this, and Kelvin did so, to the case in which the system exchanges amounts dQ_i of heat with different heat baths i at different temperatures T_i. The fact that the efficiency of an arbitrarily complicated heat engine, which exchanges heat with a variety of heat baths is still bounded by the efficiency of a reversible one, then yields the inequality

$$\sum_i \frac{dQ_i}{T_i} \leq 0, \tag{2.17}$$

where the ratio dQ_i/T_i can be thought to measure the efficiency of the energy exchange with reservoir i. The above form invites one to take the limit to an infinite number of heat baths with which the engine, or system in general, exchanges heat. In this way, the notion of a thermodynamic *path* is introduced. As the system exchanges heat with the different reservoirs, its macroscopic state variables, such as volume and temperature, trace out a path in the space of macroscopic variables. One then has from equation (2.17) that the heat exchange dQ divided by temperature T, integrated over a closed path, obeys the inequality

$$\oint \frac{dQ}{T} \leq 0, \tag{2.18}$$

where the equality holds only for a reversible path. Reversible refers to reversibility of the macroscopic state, i.e. slowly moving a piston induces a reversible change in the volume V. Reversibility is thus fundamentally defined in terms of the possibility to *control* certain macroscopic variables in the laboratory. Notice also that the temperature in equation (2.18) refers to the temperatures of the heat baths with which the system interacts. One can view this collection of heat baths as an *environment* with which the system interacts.

Once one realizes that the integral along a closed reversible path is zero, one realizes immediately that the integral from some point A to some point B is independent of whatever reversible path is taken:

$$\int_A^B \frac{dQ}{T} = H_B - H_A. \tag{2.19}$$

That is, one can uniquely associate a number H_P with each point P in the space of macroscopic variables. This number, is of course, the thermodynamic *entropy* of the system. Say that we move from A to B along a reversible path, and then back from B to A along an *arbitrary* path, either reversible or not. According to equation (2.18), the

sum of the integrals along both paths must be smaller than or equal to zero, and since the reversible path from A to B integrates to $H_B - H_A$, we have

$$\int_B^A \frac{dQ}{T} \leq H_A - H_B, \tag{2.20}$$

for the arbitrary path leading back from B to A. By the definition of the entropy, $H_A - H_B$ is the change dH_{sys} in the system's entropy due to its (possibly irreversible) movement from B to A. Note further that the left hand side of equation (2.20) is precisely the negative change dH_{env} in the entropy of the environment. Moving the left hand side to the right we have:

$$dH_{\text{sys}} + dH_{\text{env}} \geq 0. \tag{2.21}$$

This is Clausius' form of the second law. It states that when a system and its environment both start in equilibrium, undergo some changes, and eventually both end up in some new equilibrium state, that the total entropy of these equilibria is greater than or equal to zero. It is only zero when the changes in system and environment can be reversed.

In this form, the second law hardly has any implications for systems that occur consistently out of equilibrium. It demands that we compare the entropies of the initial equilibrium states and final equilibrium states. Furthermore, the interactions that are considered between the system and its environment are fairly limited. Only heat is exchanged. Note also that these entropies are defined thermodynamically, namely as an integral over a function of temperature and heat exchange. However, through the work of Gibbs, Shannon, and Jaynes we now know that this entropy is equal (up to a constant) to a measure of our ignorance regarding a system's microstate, given knowledge of the macroscopic variables that define its (equilibrium) state. In the next section, we will discuss the second law from the information theoretic viewpoint, see how it arises, and how it bears on more general situations than the ones originally considered by Clausius.

2.5.2 Information Theoretic Viewpoint

As mentioned before, the thermodynamic definition of entropy in the previous section appears very different from the information theoretic definition given in equation (2.5). In principle they *are* different, simply because they have different definitions. It was Gibbs' contribution to postulate that the form (2.5) and the definition (2.19) are in fact the same for systems in thermodynamic equilibrium. Gibbs postulated that the equilibrium distributions and their entropies can be obtained by maximizing the form (2.5) within the constraints set by the macroscopic variables that define the macroscopic equilibrium state. Although Gibbs' methods work and have been (almost exclusively) used in statistical mechanics, it has been unclear until quite recently *why* they should work. However, after the work of Shannon [127, 128] had shown that the form (2.5) is a measure of uncertainty, we have now come to understand through the work of Jaynes [84, 83] that the method introduced by Gibbs is not a law of physics, it is a law of statistical inference. So how would the second law arise from this point of view?

As might perhaps have been expected, the second law turns into almost a triviality if one adopts the informational view of entropy as uncertainty. We consider the same

setup as in the previous section: A system and an environment are both found in some thermodynamic equilibrium states. We then let the system and environment interact, typically exchanging some heat, after which both are found in some other equilibrium states. Let's assume that initially, the system is in a macroscopic state which has an entropy H^i_{sys}. That is, using Gibbs' method of choosing the maximal entropy distribution over microstates consistent with the system's macrostate gives an initial entropy H^i_{sys}. Likewise, the initial entropy of the environment, maximized conditioned on its macrostate is given by H^i_{env}. Before interaction, the joint entropy H of the system and its environment is simply given by the sum of the two entropies

$$H = H^i_{\text{sys}} + H^i_{\text{env}}. \tag{2.22}$$

The essential ingredient at this point is the realization that this *joint entropy* H cannot change due to the interaction between system and environment. The reason for this is that the laws of physics are *reversible*. That is, on the microscopic level, the laws are invariant under time reversal; any microscopic dynamics that is possible forward in time is equally possible backward in time. A consequence of this reversibility is known as *Liouville's theorem*, which states that the *phase space volume* cannot change under reversible dynamics. If we start with a volume in phase space and evolve it forward in time, the volume of this volume cannot change. It can bend and twist in all kinds of ways, but its total volume will remain unaltered. If our maximum entropy distribution is a uniform distribution over some volume in phase space, then the entropy H is proportional to the logarithm of this volume. We immediately see that Liouville's theorem implies that H is constant. For non uniform distributions in phase space it can also be shown that H is still a constant under the microscopic equations of motion. Summarizing, we initially had an uncertainty H^i_{sys} as to the precise microstate of the system and an uncertainty H^i_{env} as to the microstate of the environment. We thus had an uncertainty $H = H^i_{\text{sys}} + H^i_{\text{env}}$ as to the joint state of system and environment. From Liouville's theorem, this uncertainty H cannot change through time.

At this point it seems that we have just disproved what we wanted to establish in the first place, since we wanted to show that entropy increases but we have just shown it to be constant. However, the entropy which is constant is the *joint entropy* H of both system and environment while the second law makes a statement about the sum of the final entropies H^f_{sys} and H^f_{env} of system and environment respectively. Since the system and environment interacted, they became entangled in such a way that their respective microstates became correlated. That is, there is now some information about the environment's microstate in the system and vice versa. In information theory, this phenomenon is expressed most transparently by the following general equality

$$H = H_{\text{sys}} + H_{\text{env}} - I_{\text{sys,env}}, \tag{2.23}$$

where $I_{\text{sys,env}}$ is the *mutual information* between system and environment, see appendix A.

The mutual information measures how much the uncertainty about the environment's microstate is reduced by knowing the microstate of the system. Mutual information is, of course, always a positive quantity. That is, if it is zero initially, it can only go up. This is precisely what happens in our case. Before the system and environment interacted their

mutual information was zero; knowing the microstate of the environment didn't reduce our uncertainty about the system's microstate. But as soon as system and environment start interacting, their mutual information can only go up. Let's assume that after the environment and system have come to equilibrium again, they will have evolved an amount I of mutual information. We then have for the final entropies:

$$H = H^f_{\text{sys}} + H^f_{\text{env}} - I. \tag{2.24}$$

Combining this with equation (2.22), we obtain a version of the second law:

$$dH_{\text{sys}} + dH_{\text{env}} = I \geq 0. \tag{2.25}$$

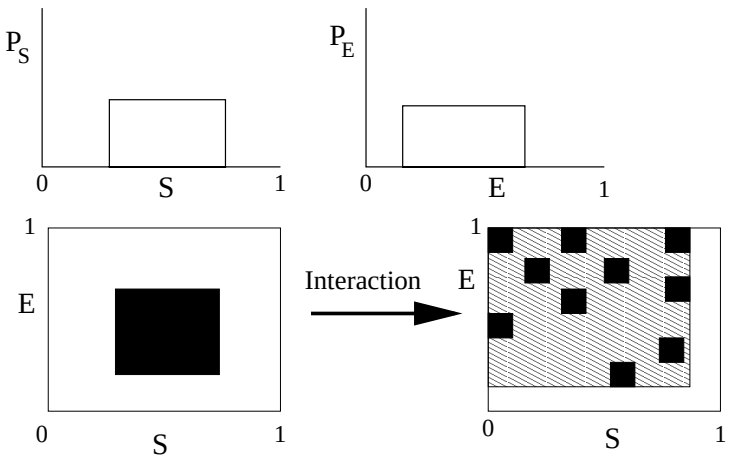

Figure 2.1: Intuitive illustration of the second law. A system and environment both start out in some macroscopic state that restrict their possible microstates—denoted by S and E respectively—to some regions of their phase spaces. The maximum entropy distributions P_S and P_E over these regions are uniform and shown in the upper two plots. The joint distribution forms a rectangle in the joint phase space shown as the black rectangle in the lower left plot. After interaction between system and environment, the *same volume* of microstates has spread through phase space as shown by the small black squares. The new entropies of system and environment, i.e. the maximal entropy consistent with the new macrostates, are effectively given by the shaded envelope. The sum of entropies of system and environment increases because they develop positive mutual information.

This result is illustrated in figure 2.1. For simplicity, we have assumed that the maximum entropy distributions are all uniform distributions. The possible microstates of the system are denoted by S, and the possible microstates of the environment by E. The initial macroscopic states of the system and environment both define some regions in their microscopic state spaces that are consistent with the macroscopic state. The maximum entropy distributions P_S and P_E are then uniform over those regions, as shown in the upper plots of figure 2.1. The initial joint state of system and environment

covers a rectangular region in their joint state space shown as the black rectangle in the lower left plot.

After system and environment have interacted, we obtain the situation in the lower right plot. The original rectangular region has broken up and spread over a large part of the joint phase space. Notice that the total volume of the black boxes in the lower right plot is the *same* as the volume of the rectangle in the lower left plot. The joint entropy has not changed. However, if we project the joint distribution on either the horizontal or vertical axis, we see that the separate entropies $H^f_{\rm sys}$ and $H^f_{\rm env}$ of system and environment have gone up. If we determine their macroscopic states, there is a much larger volume of phase space that is consistent with these new macroscopic states. This larger volume is indicated by the shaded region in the lower right plot. Although we *know* from the reversibility of the laws that the combination of system and environment can only occur in a subset of the shaded region, our inability to assess or control the precise correlations between system and environment effectively forces us to consider the whole shaded region as potential microstates for system and environment. In this way, the sum of their entropies has increased.

2.5.3 Implications

Finally, we draw out several consequences of the mutual information version of the second law just derived. First of all, this version is *not* restricted to equilibrium states only. For any combination of a system and environment that are initially without mutual information, their interaction can only lead to an increase in their mutual information. When this happens, the sum of their separate entropies must go up. This holds in general for reversible dynamics. Thus if we let two systems interact, and we notice from the evolution of their macrostates that the sum of their separate entropies (as determined from maximizing entropy with respect to their macroscopic states) has gone up, we know that they must have evolved some mutual information. That is, there are interdependencies to be discovered between them. Or, in other words, there are variables that are correlated between them which, as yet, we cannot measure or control. From this point of view, irreversible phenomena provide a potential for discovering additional "variables".

In connection with this, it is interesting to point out that considerations of engine efficiencies, which lie at the historical roots of the second law, can be simply described from an informational point of view as well. If one wants to use heat baths to drive an engine, one is essentially obtaining energy from a system whose microscopic state has a great deal of uncertainty. That is, we only know the temperature of the heat bath, which can be considered proportional to the average energy per degree of freedom of the heat bath system. With such enormous uncertainties, it is not surprising that we cannot control exactly *where* the energy is going to be delivered to our engine. That is, some of the energy will end up in degrees of freedom of our engine that do not play a role in its functionality—such as, the vibrations of its walls.

If, on the other hand, we use as an energy source some system other than a heat bath, whose state we know much better, we might be able to set up the interaction between the engine and this system in a much more controlled way, such that more of the energy will be delivered to those degrees of freedom that carry its functionality—such as, the

movement of a piston. In a certain sense, this is why microwave ovens manage to heat up your drink much more quickly than the fire on the stove does (since the microwaves make "use" of knowledge such as the resonance frequency of water). Another example is the surprisingly high efficiency of muscle fiber, which was beautifully analyzed along information theoretic lines in [86].

With respect to biological evolution, it is interesting to speculate that through evolution organisms acquire more and more "control" over the energy sources around them. That is, evolutionary innovations might be viewed in terms of organisms discovering new ways of channeling the energy stored in certain degrees of freedom of the environment. These channels may then deliver the enery to places where the organism "wants it", whereas the energy may previously have only excited degrees of freedom which were of no "use" to the organism. Maybe somewhat in the way that sunlight shining on a roof of solar cells provides more useful energy to a person than when it shines on their head and makes them sweat.

3

Macroscopic Evolutionary Dynamics

In this chapter we discuss how the maximum entropy methodology of chapter 2 can be applied to simple evolutionary processes. First we describe the formulation of the microscopic equations of motions for such simple evolutionary systems. After that, we discuss which kinds of variables might provide reasonable choices for macroscopic variables in the context of such evolutionary processes. We explicitly construct the macroscopic state spaces and dynamics on the level of the macroscopic states for several simple examples. Additionally, we discuss the qualitative behavior of finite-population dynamics which plays a central role in the rest of the thesis. Finally we discuss how, in the evolutionary context, dynamic symmetry breaking can lead to the appearance or disappearance of dimensions in macroscopic state spaces and discuss its relation to phase transitions.

3.1 Microscopic Description

The maximum entropy methods explained in chapter 2 can only be brought to bear once the microscopic dynamics of the system under study is known and formulated explicitly. For physical systems, such as a fluid or a gas, one might assume that the microscopic equations of motion are formulated in terms of elementary-particle interactions. Although these microscopic laws are known to a certain extent, there is probably no statistical physicist who would take quantum electrodynamics as a starting point for their investigations of the behavior of a gas. Instead, what one typically does is to distill out of the "real" microscopic equations of motion a mathematical model which contains all the ingredients that are thought to be of relevance. For instance, one can model a gas as a collection of spherical particles that attract each other at moderate distances and strongly repel each other at close distances. Although molecular interactions in a real gas are much more complicated than implied by such a model, the model is thought to contain all ingredients that are important for understanding the behavior of real gases.

Although the elementary-particle dynamics of a real gas might be too complicated to take as a starting point for a statistical mechanics approach, in comparison with the microscopic dynamics of "real" evolving populations it is astonishingly simple. There is simply no chance of somehow starting an analytical approach with the actual microscopic dynamics of an evolving population. The microscopic mathematical models for evolving populations that we construct below are not more than caricatures that capture and explain some typical qualitative behaviors of real evolutionary systems.

In this thesis, we restrict ourselves to cases where the microscopic state of the population is given by a list $\mathcal{S}$ of the population's current genotypes, together with their relative frequencies. Thus, first of all, we assume that every individual in the population experiences the same environment. This is in contrast to spatial models, for instance, where the microscopic state of a population includes the spatial distribution of concentrations of the different genotypes at different spatial locations. Additionally, we assume that the environment is constant. In environments that vary in time and space, the precise state of the environment and its evolution, are also be part of the microscopic description of an evolutionary system. Finally, we assume that the "state" of each individual in the population is simply a function of its genotype. That is, we neglect the possibility of complicated internal dynamics of the individuals affecting their reproductive success—such as in models including developmental processes. In particular, the fitness of an individual in the constant environment is a direct function of its genotype.

Since the members of the population are indistinguishable in our models, only the frequencies of the different genotypes determine the microscopic state $\mathcal{S}$. Once such a list $\mathcal{S}$ is given, the state of the evolutionary system is completely specified, and the dynamics is a function of $\mathcal{S}$ only. This kind of microscopic description is typical of that found in mathematical population genetics [39, 69]. Additionally, we view the genotypes that can occur as embedded explicitly in a genotype *space*. The idea that the genotypes are embedded in a space of possible genotypes, often the hypercube of all length-L symbol sequences over a finite alphabet $\mathcal{A}$, is somewhat less common in mathematical population genetics and was first stressed by Eigen [32].

It follows from these restrictions on the formal microscopic definition of an evolutionary system, that the evolutionary dynamics can be generally described as a *Markov chain* with conditional transition probabilities $\Pr(\mathcal{S}'|\mathcal{S})$ that the population at the next time step will be the "microscopic" collection $\mathcal{S}'$ given that its current microscopic state is $\mathcal{S}$. See Refs. [39] and [109] for this microscopic formulation in the context of mathematical population genetics and the theory of genetic algorithms, respectively. In formulating the microscopic dynamics as a Markov chain, a few assumptions are made implicitly:

1. The dynamics takes place in discrete time steps rather than in continuous time.
2. The microscopic state spaces are *discrete* and *finite*.
3. The Markovian assumption states that the state at the next time step is only dependent on the current state of the system. In particular, it does not depend on an infinite sequence of states of the system at previous time steps. When the next state only depends on a finite set of states in the past, these can all be grouped together in a single effective state, and then the dynamics on these new states will be Markovian again.

The first assumption is not so much a restriction but more a matter of having to choose between discrete and continuous time. In most cases, the dynamics of Markov models in continuous time can be easily mapped to the dynamics of the analogous discrete-time models. In any case, the distinction between discrete and continuous time is typically not considered to be a determining factor for the qualitative dynamical behaviors of the model. Specifically, we are careful in the following to construct our models

such that the discretization of time is not a determining factor in the behavior of the model. For instance, we take the time steps small with respect to the time scale on which the phenomena of interest occur. In particular, we consider populations evolving in discrete generations. The dynamical behaviors that we observe in our models are exactly the same as the dynamical behaviors of analogous continuous-time models.

A similar argument holds with respect to the second assumption. For a constant, homogeneous environment, the microscopic state of the population is only dependent on the frequencies of occurrence of the different genotypes. Since genotype spaces are discrete and finite, so are the population state spaces. Moreover, as will become clear, the analysis presented here can be easily extended to cases where genotype spaces are not of constant size, but are allowed to grow over time—i.e. genomes may grow in length.

The third assumption entails that the dynamics of the system has only finite memory. This does not mean that the system cannot show behavior which has memory over arbitrarily long times. It is possible that an event occurring at a certain point in time has repercussions for the entire future of the system. It is only assumed that the probabilities for the different possible futures of the system are determined only by the current state of the system. That is, the microscopic equations of motion do not exhibit memory effects, but the dynamical behaviors can.

Under these assumptions, the microscopic dynamics of an evolving population can thus be described by a Markov chain. Of course, the general theory of Markov chains can be brought to bear on these systems: this generally involves manipulating the transition matrix $\Pr(\mathcal{S}'|\mathcal{S})$ of microscopic transition probabilities. For any reasonable genetic representation, however, there is an enormous number of these microscopic states $\mathcal{S}$ and so too of their transition probabilities. For instance, for binary sequences of length L and a population of size M, the number of microscopic states is on the order of $\mathcal{O}(2^{LM})$, which is huge for reasonable sequence lengths and population sizes. This large number of microscopic states makes it almost impossible to concretely analyze the dynamics at this microscopic level. At most, one can use abstract Markov chain theory to obtain results on various kinds of asymptotic properties: such as, in the limit of infinite time, the dynamics reaches a unique fixed-point distribution over the microscopic states. Such results tend to be useless unless one is able to predict a priori how long "asymptotic" is and what this unique distribution looks like.

More practically, a full description of the dynamics on the level of microscopic states $\mathcal{S}$ is neither useful nor typically of interest. One is much more likely to be concerned with relatively coarse statistics of the dynamics, such as the evolution of the best and average fitness in the population or the expected waiting times for evolution to produce a genotype of a certain fitness or with certain phenotypic characteristics. It is hard to imagine that one would be interested in a precise description of the evolution of the probability distribution $\Pr(\mathcal{S})$ over all possible lists of genotypes $\mathcal{S}$, unless this description could be directly used to predict coarser statistics of interest. However, the huge size of the space of microscopic states $\mathcal{S}$ is precisely what prohibits such a direct derivation of the distribution $\Pr(\mathcal{S})$. The result is that quantitative mathematical analysis faces the task of finding a macroscopic description of the microscopic evolutionary dynamics that is simple enough to be tractable numerically or analytically and that, moreover, facilitates predicting the quantities of interest. Additionally, one would hope that such

a macroscopic description of the dynamics would provide insight into the qualitative mechanisms by which certain observed dynamical behaviors at the macroscopic level.

3.2 Evolutionary Macrostates

Thus, we are interested in describing the dynamics of an evolving population on a relatively coarse and macroscopic level. The first step in constructing such a description is to choose a set of macroscopic variables. As already pointed out in chapter 2 there is, as of yet, no general algorithm by which to choose a "suitable" set of macroscopic variables.[1] The macroscopic description should be capable of predicting the statistics of interest, but it should also be simple enough to allow for analyzing the dynamics in terms of these variables.

As a first step, one might attempt to remove all degrees of freedom in the dynamics that do not play a role in determining the statistics of interest. For example. there might be *symmetries* in the microscopic dynamics that can be *factored out*. A symmetry consists of a set of transformations that leave the microscopic dynamics invariant. More formally, let a transformation map each microscopic state $\mathcal{S}$ to a microscopic state $\mathcal{T}(\mathcal{S})$. If we have for all $\mathcal{S}$ and $\mathcal{S}'$ that $\Pr(\mathcal{S}'|\mathcal{S}) = \Pr(\mathcal{T}(\mathcal{S}')|\mathcal{T}(\mathcal{S}))$, then this transformation is a symmetry of the microscopic dynamics. It should be easy to see that we can *group* together all microstates that are related via such symmetry transformations and describe the dynamics on the level of these grouped states. For instance, if the genotypes consist of sequences over a binary alphabet and the dynamics is symmetric under flipping of the nth bit in all sequences of the population, then we can group together the pairs of states $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ into effective states and describe the dynamics on the level of these effective states. If there are groups of transformations $\mathcal{T}_1$, $\mathcal{T}_2$ and so on, the grouping may lead to a large reduction in the number of states in the microscopic phase space. By finding all such symmetries, the microscopic dynamics may be reduced to a minimal number of degrees of freedom. This could be potentially very helpful in analysis.

There are two problems with this formal approach however. First, the symmetries of the microscopic dynamics do not necessarily respect the statistics of interest. For instance, we might be interested in the evolution of the average fitness in the population, but the average fitness of state $\mathcal{S}$ is not necessarily the same as the average fitness of the state $\mathcal{T}(\mathcal{S})$. Second, microscopic symmetries are typically not very abundant except in the simplest cases.

Although the microscopic dynamics may not harbor many microscopic symmetries, it may still contain many symmetries with respect to some macroscopic statistic. If we are only interested in, say, the dynamics of the average fitness in the population, we may find a set of transformations that do not form a microscopic symmetry, but that keep the dynamics of the average fitness invariant. Formally, we consider the probabilities $\Pr(f(t)|\mathcal{S})$ that, given a current population state $\mathcal{S}$, the average fitness $\langle f \rangle$ follows the function $f(t)$ for all future times t. That is, given the microscopic state $\mathcal{S}$, $\Pr(f(t)|\mathcal{S})$ denotes the probability that the average fitness will be $f(1)$ at the next time step, $f(2)$

[1] One might guess that the lack of such an algorithm is simply caused by the absence of a clear definition of "suitable" in most circumstances.

at the time step after that, and so on. If two states $\mathcal{S}$ and $\mathcal{S}'$, have the same probabilities $\Pr(f(t)|\mathcal{S}) = \Pr(f(t)|\mathcal{S}')$ for *all* possible average fitness futures $f(t)$, then the microscopic states $\mathcal{S}$ and $\mathcal{S}'$ are *equivalent* with respect to the dynamics of the average fitness. This idea of equivalencing microscopic states with respect to the probabilities of possible futures is one of the key concepts of the computational mechanics approach [21, 23, 25, 41] to natural complexity. All states that are equivalent with respect to the probability of possible fitness futures are grouped into one equivalence class C which is called a *causal state*. One may use the microscopic equations of motion $\Pr(\mathcal{S}'|\mathcal{S})$ to construct the dynamics $\Pr(C'|C)$ on the level of these causal states. The important feature to note is that it is not necessary that the collections of microscopic states that form the causal states be related to symmetries of the microscopic dynamics. The causal states encode symmetries on the level of the macroscopic statistic of interest, such as average fitness. This makes this approach much more powerful in reducing the dimensionality of the state spaces than looking for actual symmetries of the microscopic dynamics.

Still, one may typically find, as one can easily imagine, that an enormous number of (causal) states remain, even with respect to the dynamics of average fitness only. Bolder choices for the macroscopic variables are then necessary to reduce the description of the dynamics to proportions that facilitate analysis. At this point, however, we have exhausted all mathematically rigorous possibilities of reducing the complexity of the macroscopic evolutionary dynamics. If we *do* want to predict the dynamics of the average fitness, the causal states form the smallest set of states that is capable of describing the dynamics of the average fitness completely [23]. The only way of further reducing the size of the state space is by giving up some of the accuracy in predicting the dynamics of the average fitness for all possible microscopic states.

There is much to be gained by this, however. For the cases studied here, it proved possible to reduce the description to a small number of macroscopic variables and, still, capture most of the dynamical behaviors on the level of the fitness that occur in the population. The main reason for this effectiveness is that in order to get reasonably good predictions, one does not need to find a set of variables that describes the average fitness dynamics *exactly*, one only has to find a set of macroscopic variables that describes the average fitness dynamics in typical situations with a reasonable accuracy. Once a macroscopic description is chosen, there might be microscopic states $\mathcal{S}$ for which this macroscopic description breaks down, but if such microscopic states are very unlikely to occur in practice, they will minimally influence the accuracy of the predictions.

What kinds of macroscopic variables would be suitable in the evolutionary context? As noted earlier, this question does not have a general answer, but there are some insights to guide us. On a very intuitive level, the Neo-Darwinian paradigm of biological evolution suggests a natural decomposition of the evolutionary dynamics into a "selection" part and a "genetic diversification" part. Simply stated, the selection is thought of as an ordering force that installs information about the environment into the population by letting "adapted" individuals survive and reproduce, and letting maladapted individuals die. Selection acts on the level of the phenotypes, or even more abstractly, on the degree of the individuals' adaption to their environment. In contrast, genetic diversification is viewed as an *independent* and largely disordering force that acts on the level of the genotypes. Roughly speaking, one can argue that different genotypes with the same level of adaptation to the environment are treated symmetrically *on average* by the evol-

utionary dynamics. Selection by definition does not distinguish between equal fitness individuals. Since genetic diversification is randomizing to a certain extent, it does not distinguish between equal fitness-individuals on average.

3.2.1 Neutrality and the Macroscopic State Spaces

As noted, we consider cases where the fitness of individuals is a direct function of their genotype. In choosing a set of macroscopic variables we are guided by a single key feature of the fitness functions that are studied here. That key feature is that there are large *degeneracies* in the map from genotype to fitness. In other words, there are many more genotypes than distinct fitness values.

Additionally, these fitness functions typically give rise to *neutral subbasins*. Neutral subbasins are sets of iso-fitness genotypes that are mutually connected through paths of single genetic diversification steps—such as point mutations. The occurrence of such neutral subbasins is, to a certain extent, due to the high dimensionality of genotype spaces. The fact that each genotype has many single mutant neighbors makes it likely that at least one of its single mutant neighbors is a *neutral* neighbor—i.e. a genotype with the same fitness. Viewed in a slightly different way, the key feature of the fitness functions studied in this thesis is that they possess *local symmetries* with respect to fitness. That is, for any genotype there always are some local genetic diversification moves that leave the fitness unchanged. This ensures, in particular, that sets of iso-fitness genotypes form connected components (subbasins) under the local genetic operators. In this way, the genotype space decomposes into a relatively small set of neutral subbasins that are entangled with each other in complicated ways[2].

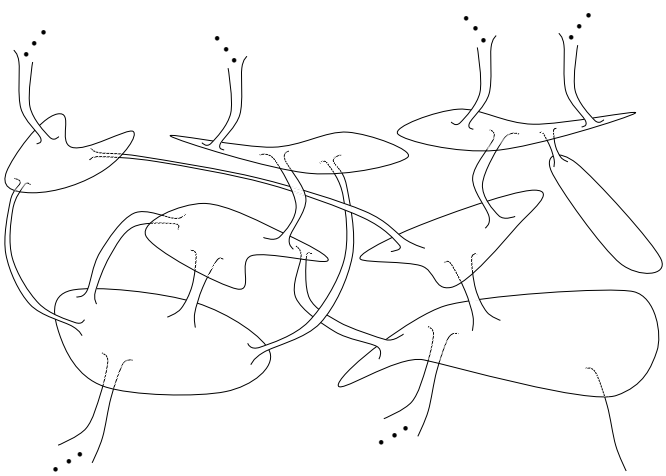

Figure 3.1: Caricature of a neutral subbasin architecture in genotype space.

[2] In the theory of molecular evolution, these neutral subbasins are generally referred to as *neutral networks*.

Figure 3.1 shows a caricature of such an architecture of neutral subbasins. The neutral subbasins are indicated as the larger volumes while the connections between them are indicated as tubes. The representation of the connections between the subbasins is somewhat misleading in that, given the high-dimensionality of genotype spaces, different subbasins are often nested inside other subbasins. The figures does, however, nicely convey the conceptual idea of genotype space as an architecture of entangled neutral subbasins.

In light of the above observations, it is natural to choose the neutral subbasins as macroscopic variables for studying the evolutionary dynamics. More specifically, we will take the proportions of the population in each of these neutral subbasins as macroscopic variables that describe the state of an evolving population. Assume that there are N neutral subbasins. In most cases that we consider, there is only one neutral subbasin for each fitness value. We then describe the state of the population at any time t by a fitness or neutral subbasin distribution $\vec{P}(t) = (P_1(t), \ldots, P_N(t))$, where the components $P_i(t)$ denote the *proportion* of the population in each of the neutral subbasins or fitness classes i at time t. The vector $\vec{P}(t)$ constitutes the set of macroscopic variables to which we will apply the maximum entropy techniques that were discussed in chapter 2.

When there are N different neutral subbasins in genotype space, the macroscopic state vector $\vec{P}$ has N components. Since the components P_i give the proportions of individuals in each neutral subbasin, $P_i \geq 0$, and we have the normalization condition

$$\sum_{i=1}^{N} P_i = 1. \tag{3.1}$$

The set of all such vectors $\vec{P}$ forms the *macroscopic state space* Λ

$$\Lambda = \{\vec{P} \in \mathbb{R}_+^N \,|\, \sum_{i=1}^{N} P_i = 1\}. \tag{3.2}$$

Strictly speaking, only infinite populations are allowed anywhere in this space Λ. For an infinite population, the proportions P_i can take on any value between 0 and 1. For a finite population of size M, in contrast, the values of P_i can only be multiples of $1/M$. That is, P_i can equal 0, $1/M$, $2/M$, and so on, but not any of the intermediate values. Thus, for finite populations, the state space is a discrete subset of Λ. In particular, the state space Λ_M for a population of size M is a lattice embedded in the state space Λ with a lattice spacing of $1/M$:

$$\Lambda_M = \{\vec{P} = \frac{\vec{n}}{M}, \vec{n} \in \mathbb{N}^N \,|\, \sum_{i=1}^{N} n_i = M\}. \tag{3.3}$$

The discrete nature of the finite-population state space plays an important role in the qualitative behavior of the evolutionary dynamics for finite populations.

3.3 Infinite-Population Dynamics

First, we will construct the evolutionary dynamics on the level of the macroscopic variables in the limit of infinite, or very large, populations. This large population-size limit

is analogous to the thermodynamic limit in statistical mechanics. We start by specifying explicitly the microscopic evolutionary dynamics in this limit. For an infinite population, the microscopic states $\mathcal{S}$ are given by density distributions over genotype space. In other words, each state $\mathcal{S}$ is a list of the relative proportions of all possible genotypes. For infinite populations, the dynamics on the level of these states $\mathcal{S}$ is *deterministic*. See, for instance, [32, 33] for a formulation of such deterministic dynamics for well mixed populations of self-replicating molecules.

In principle, the microscopic dynamics is determined explicitly in terms of parameters, such as selection coefficients, and mutation and crossover rates. The compound actions of selection and the genetic operators deterministically map each genotype distribution $\mathcal{S}$ to a genotype distribution $\mathcal{S}' = g(\mathcal{S})$ at the next generation. Similarly, we denote by $g_t(\mathcal{S})$ the microscopic genotype distribution at time t, given that the population had a genotype distribution $\mathcal{S}$ at $t = 0$. From this deterministic microscopic dynamics on the level of the genotypes, we construct the dynamics on the level of the fitness distribution $\vec{P}$. To this end, we will follow the maximum entropy method described in section 2.4.

We want to make a prediction for the fitness distribution $\vec{P}(t+1)$ given that we have information about the fitness distributions $\vec{P}(\tau)$, at all previous time steps, $\tau = 0, 1, \ldots, t$. Thus, our prediction for $\vec{P}(t+1)$ is based on the information that the fitness distribution was $\vec{P}(0)$ at time 0, $\vec{P}(1)$ at time 1, and so on. Note that we assume that the fitness distribution took on these values *exactly*, as opposed to only knowing the *average* values of the fitness distribution. The maximum entropy approach then tells us to assign equal weight to all initial genotype distributions $\mathcal{S}$ that are consistent with all "measured" macroscopic states $\vec{P}(0)$ through $\vec{P}(t)$. Denote by C the set of all initial genotype distributions that are consistent with the entire sequence of fitness distributions $\vec{P}(0)$, $\vec{P}(1)$, and so on. Then we have that the expected fitness distribution $\langle \vec{P}(t+1) \rangle$ at time $t+1$ is:

$$\langle \vec{P}(t+1) \rangle = \sum_{\mathcal{S} \in C} \frac{\vec{P}[g_{t+1}(\mathcal{S})]}{|C|}, \tag{3.4}$$

where by $\vec{P}[\mathcal{S}]$ we denote the fitness distribution of the genotype distribution $\mathcal{S}$, and $|C|$ is the size of the set of consistent genotype distributions. Obviously, the determining component here is the set C. This set generally depends on all previous fitness distributions $\vec{P}(t)$. In other words, we cannot generally predict the fitness distribution $\vec{P}(t+1)$ from the current fitness distribution $\vec{P}(t)$ alone.

3.3.1 Memoryless Approximation

From a mathematical point of view, we can only predict $\vec{P}(t+1)$ from the current state $\vec{P}(t)$ if the microscopic dynamics is exactly symmetric with respect to the fitness distribution in the following sense: If two microstates have the same fitness distribution, then they always lead to equal fitness-distributions at the next time step. Formally, if for all $\mathcal{S}$ and $\mathcal{S}'$ with $\vec{P}[\mathcal{S}] = \vec{P}[\mathcal{S}']$ we have that $\vec{P}[g(\mathcal{S})] = \vec{P}[g(\mathcal{S}')]$. In those cases, we really only need to know the fitness distribution $\vec{P}(t)$ to predict the distribution $\vec{P}(t+1)$. Such exact symmetries do not typically occur. However, we may still be able to

accurately predict the dynamics on the level of fitness distributions using a memoryless approximation.

To see how accurate prediction may still occur, we denote by C_1 the set of genotype distributions $\mathcal{S}$ that have the current fitness distribution $\vec{P}(t)$ as their fitness distribution, i.e. $\vec{P}[\mathcal{S}] = \vec{P}(t)$. Using only this current fitness distribution as our information on which to base our prediction, rather than *all* previous fitness distributions, we have

$$\langle \vec{P}(t+1) \rangle_1 = \sum_{\mathcal{S} \in C_1} \frac{\vec{P}[g(\mathcal{S})]}{|C_1|}. \tag{3.5}$$

The 1 in the subscript of the prediction $\langle \vec{P}(t+1) \rangle_1$ denotes that we have only used the current fitness distribution $\vec{P}(t)$ and the corresponding set of microscopic states C_1 in our prediction.

If, for most cases, this prediction $\langle \vec{P}(t+1) \rangle_1$ is close to the prediction $\langle \vec{P}(t+1) \rangle$ which is obtained by including all previous fitness distributions, one obtains accurate predictions for the dynamics on the level of fitness distributions by only taking the current fitness distribution into account. In other words, the values of the fitness distributions at previous times do not contain much information with respect to the future of the fitness distributions. In this thesis we use this memoryless approximation to the dynamics of the fitness distribution without explicitly attempting to prove that it leads to accurate predictions. That our memoryless approximation leads to accurate predictions simply follows from comparing these predictions with data obtained from simulations of the actual population dynamics.

In summary, at each time t we use the maximum entropy distribution over genotype distributions, conditioned on $\vec{P}(t)$ only, to predict $\vec{P}(t+1)$. The maximum entropy distribution is uniform over the set C_1 of genotype distributions that are consistent with the current fitness distribution $\vec{P}(t)$.

3.3.2 The Generation Operator

The memoryless maximum-entropy approximation to the dynamics of fitness distributions is implemented by constructing a *generation operator* $\mathbf{G}$ that takes the current fitness distribution $\vec{P}(t)$ and maps it to the fitness distribution $\vec{P}(t+1)$ at the next time step. Formally,

$$\vec{P}(t+1) = \mathbf{G}[\vec{P}(t)]. \tag{3.6}$$

We shall focus on simple evolutionary dynamics, which only involve selection and mutation. We decompose the generation operator $\mathbf{G}$ into a selection operator $\mathbf{S}$ and a mutation operator $\mathbf{M}$ that account for the effects of selection and mutation on the fitness distribution respectively.

The selection operator is easy to construct since its effects on the fitness distribution depend only on the current fitness distribution. One of the most common forms of selection is fitness-proportionate selection: the expected number of offspring that an individual with fitness f produces in the next generation is proportional to f. If we denote by f_i the fitness of the genotypes in neutral subbasin i and by P_i^{sel} the proportion

of individuals in subbasin i after selection, we have:

$$P_i^{\text{sel}} = \frac{f_i P_i}{\langle f \rangle} \equiv \left(\mathbf{S}[\vec{P}]\right)_i . \tag{3.7}$$

To calculate the effects of mutation on the fitness distribution we must explicitly use the maximum entropy method. We need to calculate the probabilities M_{ij} that a genotype in neutral subbasin j will mutate to a genotype in neutral subbasin i. As explained in the previous section, the maximum entropy distribution is uniform over all genotype distributions that are consistent with the fitness distribution. This is equivalent to assuming that a single individual in neutral subbasin j is equally likely to be any of the genotypes in neutral subbasin j. If we denote by V_j the set of genotypes in neutral subbasin j and by $T_\mu(s \to s')$ the probability that mutation transforms genotype s into genotype s', we have

$$M_{ij} = \sum_{s \in V_j, s' \in V_i} \frac{T_\mu(s \to s')}{|V_j|}. \tag{3.8}$$

In this thesis we consider genotypes that are bit strings of some fixed length L. If we denote by $d(s, s')$ the Hamming distance between genotypes s and s', the mutation probabilities for a uniform mutation rate μ per bit become

$$T_\mu(s \to s') = (1-\mu)^L \left(\frac{\mu}{1-\mu}\right)^{d(s,s')} . \tag{3.9}$$

Finally, the generation operator is the product of the selection and mutation operators. If P_j^{sel} is the proportion of the population in the neutral subbasin j after selection, then $\sum_j M_{ij} P_j^{\text{sel}}$ is the proportion of the population in the neutral subbasin i after selection and mutation. More formally, we have for the equations of motion of the fitness distribution $\vec{P}(t)$:

$$\vec{P}(t+1) = \mathbf{M} \cdot \left(\mathbf{S}[\vec{P}(t)]\right) \equiv \mathbf{G}[\vec{P}(t)]. \tag{3.10}$$

With these, the infinite-population dynamics has been explicitly constructed on the level of neutral subbasins or, equivalently, on the level of fitness distributions. For each current fitness distribution $\vec{P}$, acting with the generation operator $\mathbf{G}$ gives us the fitness distribution at the next generation.

Viewed in a slightly different way, we can focus on the *change* $d\vec{P} = \mathbf{G}[\vec{P}] - \vec{P}$ over one generation when the population currently has fitness distribution $\vec{P}$. By considering this change $d\vec{P}$ for each point $\vec{P}$ in the macroscopic state space, we get a sense of the "force" that is generating the flow of populations through the state space of fitness distributions.

An example of such a flow through the macroscopic state space is shown in figure 3.2. The figure illustrates the flow $d\vec{P}$ for a simple fitness function, over the space of all 2^{30} bit strings of length $L = 30$, that contains only 4 neutral subbasins. The subbasins are denoted 0, 1, 2, and 3 and have respective fitnesses $f_0 = 0$, $f_1 = 1$, $f_2 = 2$, and $f_3 = 3$. The three-dimensional state space Λ forms a simplex in 4 dimensions, by normalization equation (3.1). The component P_3 is determined by the others, i.e. $P_3 = 1 - P_0 - P_1 - P_2$. The population evolves under fitness proportionate selection and

a uniform mutation rate $\mu = 0.005$ per bit per reproduction. For a subset of the possible fitness distributions $\vec{P}$ in Λ the arrows show the change $d\vec{P}$ over one generation. In other words, the arrows point from $\vec{P}$ to $\mathbf{G}[\vec{P}]$.

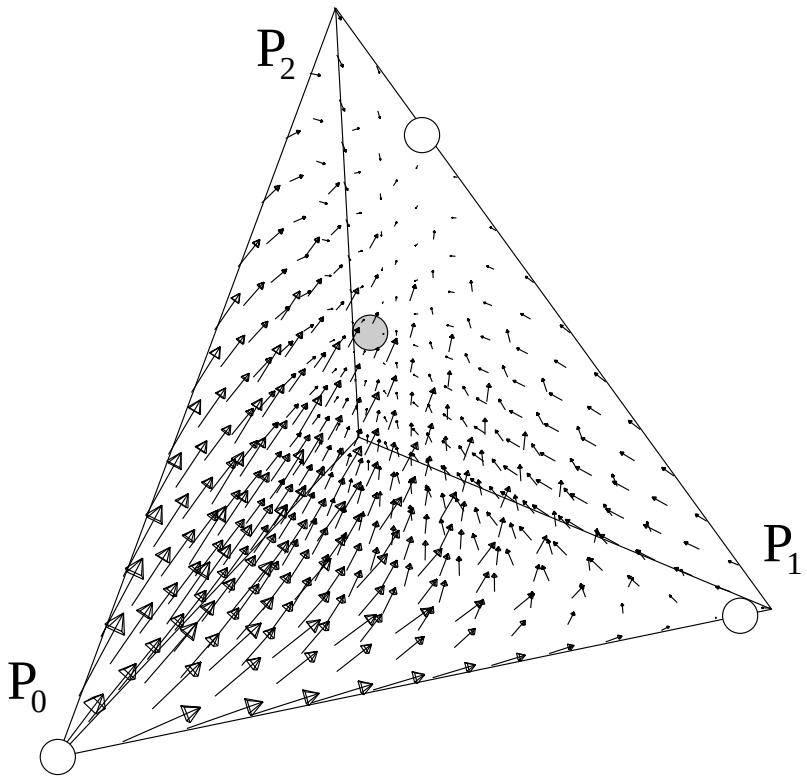

Figure 3.2: Fitness distribution flow $d\vec{P}$ in the state space simplex Λ for a fitness function which contains 4 neutral subbasins and for a population evolving under fitness proportionate selection and mutation. The arrows indicate $d\vec{P}$. Fixed points of the flow are shown as large balls. The grey ball corresponds to the stable, asymptotic fixed point in the interior of the simplex Λ. The white balls indicate the locations of the unstable fixed points that are outside the simplex. The latter do not represent valid populations, but nonetheless they affect the dynamics of allowed populations within the simplex by slowing down (short arrows) the flow near them. This figure was taken from [24].

Once these arrows are known for all $\vec{P}$, analyzing the infinite-population dynamics becomes a more or less standard problem in dynamical systems. All techniques that have been developed in dynamical systems theory [113] can then be brought to bear on the analysis of the population dynamics. For instance, we have also indicated as large balls in figure 3.2 the fixed points of the generation operator $\mathbf{G}$ where the flow $d\vec{P} = 0$. Of course, one can then perform linear perturbation analysis to determine the stability of these fixed points. The white balls indicate the locations of the unstable hyperbolic fixed points whereas the gray ball indicates the location of the asymptotically stable fixed

point. In fact, the fixed point corresponding to the gray ball is the only fixed point that is located *inside* the state space simplex Λ. The other fixed points are located just outside this simplex. The population can therefore never reach these fixed points. However, as we explain in the next section, these fixed points still play an important role in the finite-population dynamics.

For infinite populations, all possible genotypes are always present in the population. Additionally, there is always a nonzero probability for any genotype s' to be generated by any other genotype s, i.e. $T_\mu(s \to s') > 0$. Even though some of these transition probabilities may be extremely small, in an infinite population such transitions still occur infinitely often. For finite populations of any reasonable size, most of these very unlikely mutations will not occur and this may drastically alter the dynamics. In the following section we consider how the infinite-population dynamics constructed above can be used and altered to obtain the dynamics for finite populations.

3.4 Finite Population Dynamics

As pointed out above, we mainly focus on the dynamics of finite populations of a constant size in this thesis. For populations of size M the state space Λ_M is a lattice embedded in the infinite-population state space Λ with a lattice spacing, $1/M$, that is inversely proportional to the population size. Apart from the fact that the population can only take on points of the discrete lattice Λ_M at any point in time, the largest difference between the finite-population dynamics and the infinite-population dynamics is that the dynamics is no longer *deterministic* on the level of fitness distributions. In two different realizations of the dynamics, the same fitness distribution $\vec{P}(t)$ may give rise to different future trajectories through the space of fitness distributions.

There are two ways to deal with this situation. We could restrict ourselves to predicting the *average dynamics*, averaged over many realizations of the process, together with variances and maybe even including higher moments. This is the approach taken by Shapiro, Prügel-Bennett, and Rattray in describing the dynamics of genetic algorithms [116, 117, 118, 119]. They also make use of the maximum entropy methods from statistical mechanics and focus on "fitness" as a macroscopic variable as well. More specifically, they use the average evolution of the first cumulants of the fitness distribution as macroscopic variables. This approach works well when the dynamics in each of the realizations of the dynamics fluctuates around the *average* dynamics over many realizations. That is, if the average dynamics is *typical* for any realization of the process. In this thesis, a somewhat different approach is taken.

When the current fitness distribution $\vec{P}(t)$ is known, we simply restrict ourselves to predicting the *probability distribution* of the fitness distributions $\vec{P}(t+1)$ occurring at time $t+1$. That is, we do not assume that the average dynamics is necessarily representative for the typical dynamics in each particular realization. For almost all cases studied in this thesis, the dynamics varies between realizations in such a way, that the average dynamics is indeed not representative for any of the realizations.

Thus, we want to construct the probabilities $\Pr[\vec{Q}|\vec{P}]$ that the current fitness distribution $\vec{P} \in \Lambda_M$ leads to a distribution $\vec{Q} \in \Lambda_M$ at the next generation. The probability distribution $\Pr[\vec{Q}|\vec{P}]$ generally depends on the way selection is implemented. In this

thesis, we focus on fitness-proportionate selection in which a new generation of individuals is created by selecting M times, with replacement, a random individual from the current population. The probability for each individual to be selected is proportional to its fitness. After that, all M selected individuals are mutated. The mutated copy of each selected individual is placed in the next generation. Another way one can think of implementing fitness proportionate selection is that each individual in the current population creates a large number of copies of itself as "potential offspring". The number of potential offspring that each individuals creates is proportional to its fitness. From this large pool of potential offspring, M individuals are selected at random and then mutated. These mutated individuals then form the next generation. Fitness-proportionate selection is equivalent to selection in continuous-time models where an individuals' reproduction rate is proportional to its fitness, and a global dilution flux ensures that the population remains roughly constant in size—such as in the Eigen model of molecular evolution [32, 33].

It is easy to see that for fitness-proportionate selection, each of the M offspring in the next generation has a probability P_i^{sel} to be the offspring of an individual in the neutral subbasin i, with P_i^{sel} given by equation (3.7). Individuals that are offspring of an individual in neutral subbasin j have a probability M_{ij} to occur in the neutral subbasin i after mutation has taken place. Therefore, the probability that a randomly chosen individual in the next generation is type i is given by the ith component $\mathbf{G}_i[\vec{P}]$ of the generation operator acting on the current fitness distribution $\vec{P}$. Finally, since each of the M individuals in the next generation are the result of *independent* selection and mutation events, it follows that the distribution $\vec{Q}$ at the next time, is given by a multinomial sample of size M of the distribution $\mathbf{G}[\vec{P}]$. If we define $Q_i = n_i/M$ we have:

$$\Pr[\vec{Q}|\vec{P}] = M! \prod_{i=1}^{N} \frac{\left(\mathbf{G}_i[\vec{P}]\right)^{n_i}}{n_i!}. \tag{3.11}$$

Thus, the *expected* fitness distribution at the next generation is still $\mathbf{G}[\vec{P}]$. In the limit of infinite populations, this distribution is always exactly realized. For finite populations, however, the distribution $\mathbf{G}[\vec{P}]$ is typically not exactly realized: different distributions $\vec{Q}$ occur for different runs of the evolutionary dynamics. Moreover, It is generally not *possible* that the expected distribution $\mathbf{G}[\vec{P}]$ is realized, since the components of $\mathbf{G}[\vec{P}]$ are unlikely to be multiples of $1/M$.

The finite-population dynamics as constructed from the infinite-population dynamics $\vec{P} \to \mathbf{G}[\vec{P}]$ and the multinomial sampling over the lattice Λ_M are illustrated in figure 3.3. Allowed finite-population fitness distributions $\vec{Q} \in \Lambda_M$ are shown as the large dots. The arrow points from the current fitness distribution $\vec{P}$ to the expected fitness distribution $\mathbf{G}[\vec{P}]$ at the next generation. This expected fitness distribution is shown as a small dot and does not typically fall on one the allowed finite-population distributions of Λ_M. The bars over the large dots indicate the multinomial distribution $\Pr[\vec{Q}|\vec{P}]$. The variance of the multinomial distribution around the expected distribution $\mathbf{G}[\vec{P}]$ is proportional to $1/M$.

In analyzing the finite-population dynamics, we in general do not attempt to "iterate" the stochastic dynamics of equation (3.11) to obtain the stochastic population dynamics

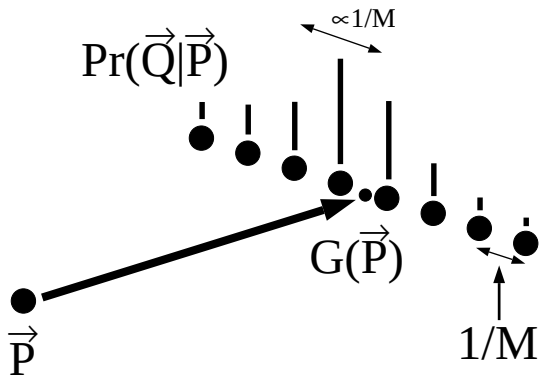

Figure 3.3: Illustration of the stochastic dynamics over one generation. The current fitness distribution $\vec{P}$ is mapped by the generation operator $\mathbf{G}$ to the *expected* fitness distribution $\mathbf{G}[\vec{P}]$ at the next generation, which is indicated by the small dot. The actual fitness distribution $\vec{Q}$ at the next generation is given by a multinomial sample of size M with distribution (3.11), which is indicated by the bars over the large dots. The large dots indicate allowed finite population fitness distributions $\vec{Q} \in \Lambda_M$. Note that the expected distribution $\mathbf{G}[\vec{P}]$ is typically not located on one of the points of Λ_M.

over arbitrary lengths of time. Formally, of course, this would give a rigorous view of the different dynamical "trajectories" that a finite population may follow with more or less probability. However, such an approach is not practical simply because it does not appear to be tractable analytically. Instead, we use the infinite-population dynamics to identify where in state space the "interesting" regions are and to get a rough sense of what regions of state space are likely to be visited by the finite population. We for instance find that the finite population dynamics spends most of its time close to unstable hyperbolic fixed points of $\mathbf{G}$ and short transition times in "tubes" connecting the regions close to these unstable hyperbolic fixed points. We then analyze the finite population dynamics more explicitly in those specific areas. In particular, we approximate the local finite-population dynamics in a region of state space using diffusion equation approximations analogous to those introduced in mathematical population genetics by Kimura [90].

3.5 Metastability and Phase Space Unfolding

Obvious candidates for the locations of interesting state space regions are the neighborhoods of the fixed points of the generation operator $\mathbf{G}$. It turns out, however, that for the dynamics studied in this thesis there is typically only a single fixed point of $\mathbf{G}$ located inside the state space Λ. This fixed point gives the asymptotically stable fitness distribution towards which the population evolves in the limit of long times. Since the other fixed points of $\mathbf{G}$ lie outside the state space, one would generally conclude that the population simply cannot reach these points. However, we also find that these fixed

points are typically located *very close* to the state space Λ. As can be seen from the small arrows in the neighborhood of the unstable fixed points (white balls) in figure 3.2, the flow can become very small in the neighborhood of these unstable fixed points.

This observation turns out to be of great importance for the qualitative dynamics of finite populations. The essential point is that a finite population can only take on fitness distributions that are points of the discrete lattice Λ_M. If the expected flow dP_i in direction i is small compared to the lattice spacing $1/M$, the population is most likely not to move in direction i. A large population can still be carried by a small flow, but if the population gets small—such that the lattice spacing $1/M$ is large with respect to the flow—the population stops moving, even if there is no fixed point locally. Only after a long time will the population make the "jump" to the next lattice point stochastically. If the flow at this lattice point is much larger, the population may then "take off", moving rapidly away from the neighborhood of small flow. In this way, finite populations induce metastability in the absence of fixed points. The fixed points outside Λ play a prominent role, since they indicate where the flow is smallest and, thus, where metastability is likely to occur for small populations.

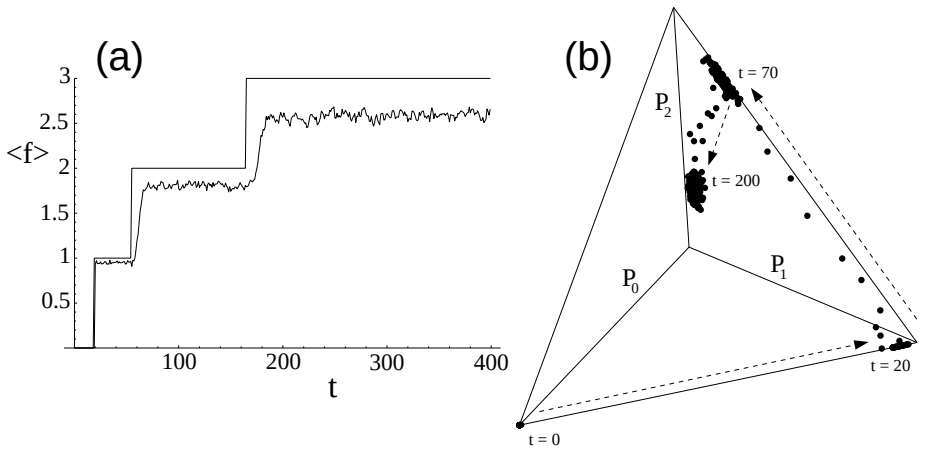

Figure 3.4: Illustration of the population dynamics on the level of average and best fitness (a), and fitness distributions (b) for the same fitness function and evolutionary parameters as in figure 3.2. In (b) the fitness distribution at each generation is indicated by a dot. The arrows indicate the direction of the flow of the fitness distribution over time. The times at which different metastable states are first reached are indicated as well. (Figure taken from [24])

This mechanism is illustrated in figure 3.4. The figure shows the dynamics of the average fitness and best fitness (a) and the fitness distribution (b) for a single run of the evolutionary population dynamics corresponding to the flow of figure 3.2 with a population size of $M = 250$. The evolution of average fitness in figure (a) shows four *epochs*, corresponding to time intervals of constant best fitness in the population. In figure 3.4(b) the fitness distribution at each generation is indicated by a dot. We see that up to generation $t = 20$ the fitness distribution is located in the lower left corner

($P_0 = 1$) of the simplex. Between $t = 20$ and $t = 60$ it fluctuates around an equilibrium in the lower right corner, which corresponds to one of the unstable fixed points in figure 3.2. At time $t = 60$ the fitness distribution suddenly starts to move upward and reaches a new equilibrium point in the top of the simplex around $t = 70$. This equilibrium fitness distribution corresponds to one of the unstable fixed points in figure 3.2 as well. The fitness distribution fluctuates around this point until approximately $t = 170$, at which point it starts to move downward to the asymptotically stable fixed point (the gray ball in figure 3.2). It reaches this fixed point around $t = 200$.

Note that the consecutive epochs are associated with increasing dimensionalities of their metastable fitness distributions. That is, the first metastable distribution occurs at $P_0 = 1$ and has dimensionality 0. The second metastable distribution occurs on the line $P_0 + P_1 = 1$ and therefore has dimensionality 1. That is, as the fitness distribution fluctuates around the fixed point during this epoch, it remains on the *line* $P_0 + P_1 = 1$. The third metastable distribution is located in the plane $P_0 + P_1 + P_2 = 1$, and the final asymptotically stable fitness distribution has dimensionality 3. In this way, the succession of metastable states through which a population evolves is associated with *unfolding dimensions* of the macroscopic state space.

3.5.1 Unfolding Dimensions

Intuitively, this type of qualitatively behavior is caused by the enormous variance in the relative sizes of the neutral subbasins in genotype space. Neutral subbasins that correspond to sequences of low fitness are typically large while neutral subbasins of genotypes with high fitness are small. The genotype space is dominated by genotypes of low fitness. When the evolution is seeded with individuals occurring at one or more randomly chosen genotypes, the population is most likely to contain low-fitness genotypes only. In the example of figure 3.4, for instance, the population initially only contains sequences of fitness zero. That is, the fitness distribution is $P_0 = 1$; a zero-dimensional fitness distribution. The population is located at the unstable fixed point in the lower left corner of figure 3.2. For populations that are not too large, the population remains in this corner for some period of time. The flow components dP_1 and dP_2 are not zero, but they are very small compared to $1/M$. In the language of neutral subbasins, the flows dP_1 and dP_2 are small since it is unlikely that any individual in the population will leave a mutant offspring of fitness 1 or 2 in the next generation. Of course, mutations induce the population to explore new parts of genotype space, but since the subbasin of fitness zero genotypes dominates genotype space, it generally takes many generations before an individual embarks on a sequence with fitness 1, 2, or 3. To be more precise, if $dP_1 = 0.05/M$, this can be interpreted as meaning that for a population of size M, on average 0.05 sequences of fitness 1 will be created in the next generation. It will thus take on the order of 20 generations before *one* sequence of fitness 1 is discovered.

When this has happened, the component P_1 jumps from from $P_1 = 0$ to $P_1 = 1/M$. Typically, selection then quickly expands the population of sequences with fitness 1 until an equilibrium between fitness-0 and fitness-1 sequences is established in the population. The fitness distribution is located on the line $P_0 + P_1 = 1$ near the unstable fixed point on the right in figure 3.2. This fitness distribution is 1-dimensional in the sense that it is described by the proportions of two components that sum to one. Through the

discovery of sequences of fitness 1 a new dimension has been added to the macroscopic state space. In figure 3.4, the population fluctuates around this metastable state in the time period between $t = 20$ and $t = 60$.

This scenario repeats itself. Under mutation, the population moves through the neutral subbasins with fitness 0 and 1, but since genotypes of fitness 2 are even more rare, they take a longer time to be discovered by mutation. Mutation has to move the population through most parts of the neutral subbasins in genotype space before a fitness-2 sequence is discovered. When a fitness-2 sequence is discovered it quickly spreads. The population moves into the plane $P_0 + P_1 + P_2 = 1$ and will stabilize in this plane around the location of the upper unstable fixed point in figure 3.2. This happens between $t = 60$ and $t = 70$ in figure 3.4. The dimensionality of the state space has become 2-dimensional at this point. Another dimension has been unfolded by the dynamics. Finally, when sequences of fitness 3 are discovered, the population moves to the asymptotic fixed point in the interior, indicated by the gray ball. This occurs between times $t = 170$ and $t = 200$ in figure 3.4.

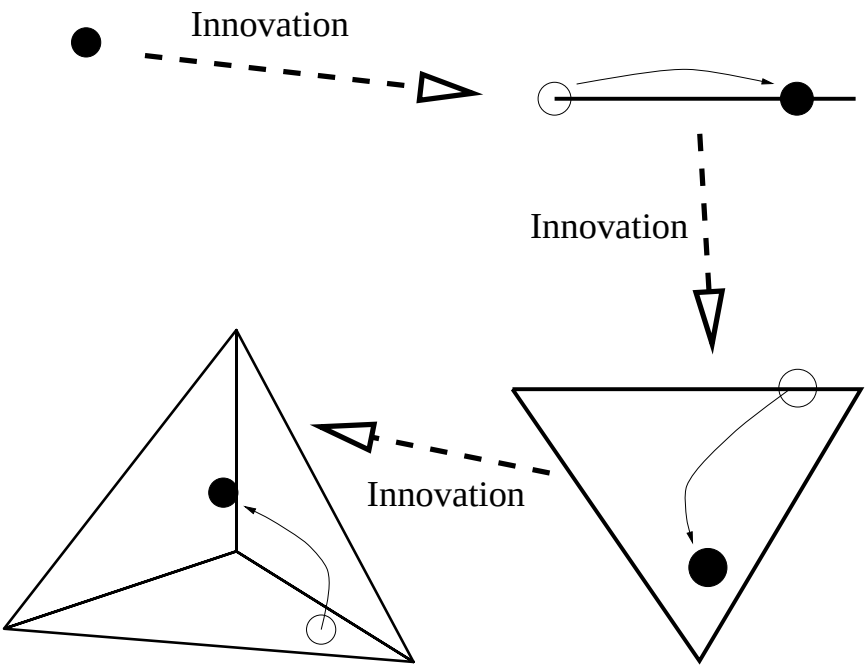

Figure 3.5: Illustration of the unfolding of macroscopic state space through epochal evolutionary dynamics. While, in genotype space, the population drifts through neutral subbasins of iso-fitness sequences, the distribution of fitness or phenotypes in the population is constant. Every time a new and better adapted phenotype is discovered that spreads through the population (an innovation), a new dimension is added to the macroscopic state space.

This scenario of macroscopic state space unfolding one dimension at a time is il-

lustrated schematically by figure 3.5. The population starts out with individuals of only 1 macroscopic type which is indicated as a dot. Once individuals of a new, and better adapted phenotype have been discovered, this fixed point becomes unstable and the population moves to a new equilibrium which involves a mixture of individuals of both the old and the new macroscopic type. The equilibrium distribution is now a point on a line. Further innovations to new macroscopic types move the population into a plane, and into three dimensional space. Each time a new macroscopic type is discovered the old fixed point becomes unstable and the population moves to a fixed point that contains more independent components. In this way, the macroscopic state of the population increases its dimensionality each time a better adapted macroscopic type is discovered. Of course, this unfolding can go on to successfully higher dimensions.

The scenario of incremental unfolding of macroscopic dimensions is potentially very general. In constant selective environments, the discovery of genotypes that confer a substantially higher fitness on individuals tends to be very rare. Often, such changes coincide with the discovery of new functionality on the level of the phenotypes or new adaptations to the environment. In this sense, these *innovations* add a new degree-of-freedom to the population dynamical system on the level of macroscopic states. This is reflected formally by the unfolding of a new dimension in state space.

It is important to note here that such new macroscopic dimensions are by no means uniquely predetermined. There may be many different macroscopic types with an increased adaptive value that may be accessed from the current state of the population. Depending on which innovative type is discovered first (an essentially stochastic process) different new macroscopic types may "freeze in". These occurrences are generally referred to as *frozen accidents*. The occurrence of frozen accidents is directly formalized in our analytical framework by the fact that at any point in time there may be many different but *mutually exclusive* dimensions that may be unfolded.

Moreover, once a new macroscopic dimension has unfolded, the stage may be set, so to speak, for further macroscopic dimensions that may now be unfolded. In other words, the potential unfolding of new macroscopic dimensions is contingent on the unfolding of previous macroscopic dimensions. This phenomenon, which we generally refer to as *historical contingency*, also fits naturally within our picture of unfolding macroscopic dimensions. The current macroscopic types that occur in the population determine what the genotypic potential for neutral variations is. This space of neutral variants determines which new macroscopic types may unfold next.

3.5.2 Unfolding and Phase Transitions

During a metastable period when the population state is located around a fixed point in the macroscopic state space, random genetic diversification mechanisms lead the population to explore phenotypically neutral variations of the current macroscopic states. This exploration of neutral variants continues until one of these variants turns out to have adaptive value. At the moment this happens, a symmetry of the dynamics is broken.

Remember that our statistical dynamics approach assumes a maximum entropy distribution over the subbasin of mutually neutral variants in genotype space. This is equivalent to assuming that the microscopic dynamics is *symmetric* with respect to all the neutral variants, as explained in section 3.3. While the population explores the space

of neutral variants, these neutral variants effectively act as if they are symmetric with respect to the dynamics on the level of the macrostates. Simply put, the dynamics on the level of the macroscopic states is not affected by the genetic exploration of neutral variants. To give a simple example: an individual's fitness may be independent of the content of letters 11 through 23 of its genotype, *except* when these 13 letters form the combination `AATGGTCATACGT`. In this case, the small segment of the genome changes from a disfunctional pseudogene into a gene with a novel and adaptive functionality. The dynamics *will* be effectively symmetric with respect to the content of letters 11 through 23 as long as no individual has hit the above "jackpot" combination. When this happens, the microscopic symmetry will be broken and a new macroscopic dimension will unfold.

The above discussion makes it clear that there is a strong connection between the concept of a phase transition from statistical mechanics and the unfolding of new macroscopic dimensions through evolutionary innovations: a microscopic symmetry of the dynamics is broken and a new macroscopic variable appears. The situation is, however, not entirely identical. In the statistical mechanical examples of chapter 2, a phase transition is induced by the change in an *external* control parameter. In the evolutionary case, the symmetry breaking occurs *dynamically* through a process that is endogenous to the system. Moreover, in the evolutionary process, a symmetry is based on a lack of information. The dynamics appears symmetric with respect to the different genotypic variants as long as all variants that have been explored have been of a macroscopic type already present in the current population. As long as the exploration of new genotypes through mutations only encounters neutral (or deleterious) genotypes, the dynamics on the level of macroscopic variables is invariant under this genotypic exploration—i.e. as if such genotypic variations form a symmetry with respect to the macroscopic dynamics. When the "jackpot" genetic combination is hit for the first time, the dynamics *discovers* that the symmetry was not complete: a very special genotypic combination did not give a neutral or deleterious variant, but gave something new on a macroscopic level. This transition is then accompanied by the appearence of a new "order parameter".

4

Statistical Dynamics of the Royal Road Genetic Algorithm

Erik van Nimwegen, James P. Crutchfield, and Melanie Mitchell
Theoretical Computer Science
220 (in press 1999).
Santa Fe Institute Working Paper 97-04-035

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.

4.1 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.,[12, 37, 64, 79, 99, 104, 120]). Epochal behavior has also been observed in natural phenomena such as the outbreak of epidemics [15], the progression of diseases such as cancer [8] and AIDS [16] 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.

4.1.1 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., [11, 28, 38, 49]). 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 [139] or the state of the population as a whole [109]. 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 [109] and [139] 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 [117, 118, 119]. 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 [32, 35]. 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 [90] that were originally developed in the context of statistical physics [45, 93, 115]. 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.

4.1.2 Organization of the Analysis

Our analysis of epochal evolution in a mutation-only genetic algorithm first appeared in [137]. The present work goes into considerably more depth. Section 4.2 introduces the simplified GA used. In section 4.3 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 4.4 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 4.5 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 4.6 is concerned with the finite-population dynamics and presents the main results. This section builds on the results from section 4.5 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 sec-

tion 4.6 are the eigenvalues and eigenvectors of the Jacobian of the generation operator that describes the dynamics in the limit of infinite populations. Section 4.7 discusses the average durations of the epochs and describes how the model breaks down in predicting these average durations. Section 4.8 discusses the results of our paper and looks ahead to future work.

4.2 A Simple Genetic Algorithm on the Royal Road Fitness Function

4.2.1 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

$$f(s) = \sum_{i=1}^{N} f_i \delta_{s_i,x_i}, \tag{4.1}$$

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 [105] and so could test a GA's ability to preferentially propagate useful genetic "building blocks". The Royal Road functions defined in [105] 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.

4.2.2 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 4.6.5 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.

4.3 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 4.1: 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

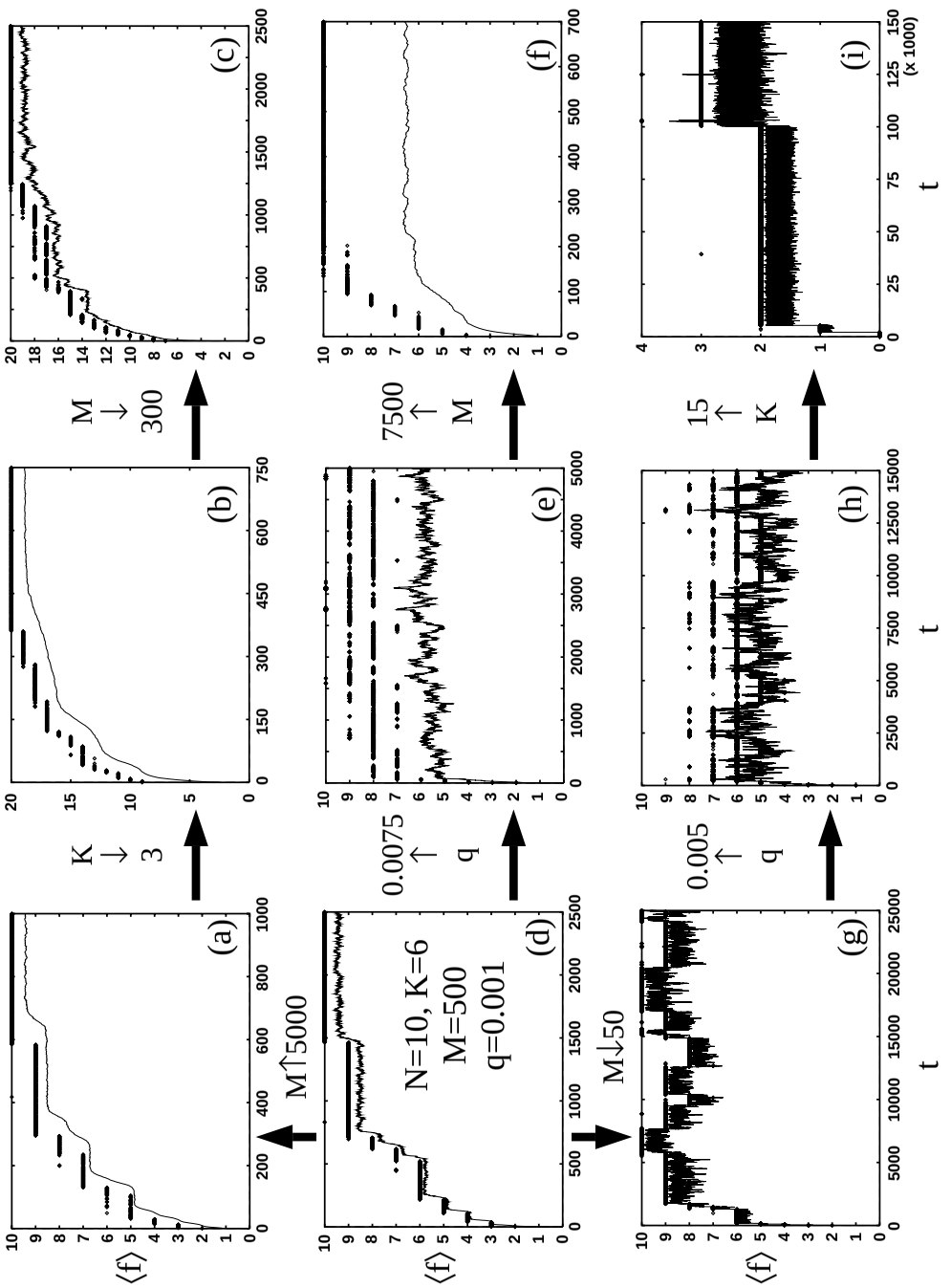

Figure 4.1: For caption see text.

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 4.1(a) through 4.1(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 4.1(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 4.1(d), selected as the canonical example of epochal evolution, was performed with $N = 10$ 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 4.1(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 4.1(a) and 4.1(g) in the left-hand column show the effect on the search behavior of changing the population size M. Run 4.1(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 4.1(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 4.1(a), 4.1(d), and 4.1(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 4.1(b), 4.1(e), and 4.1(h)—illustrate how change in a single parameter can cause fitness epochs to disappear. In run 4.1(b), for example, the block length was lowered from $K = 6$ to $K = 3$, otherwise the parameter settings are those of 4.1(a). Note that since we are keeping string length constant ($L = 60$) throughout the figure mosaic, in run 4.1(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 4.1(e) has a higher mutation rate ($q = 0.0075$) than that of the canonical run 4.1(d). Here too the fitness epochs have disappeared from the behavior, but in a markedly different way from that in going from run 4.1(a) to run 4.1(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 4.1(h) has a higher mutation rate ($q = 0.005$) than that in run 4.1(g) ($q = 0.001$). The behavior of run 4.1(h) is similar to that in run 4.1(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 4.1(e) the fitness fluctuates up and down on a time scale that is roughly constant. In run 4.1(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 4.1(d) to 4.1(e), we would like to understand why the fitness epochs disappear in going from 4.1(g) to 4.1(h) by increasing mutation and what mechanisms cause the different types of fluctuation behavior in 4.1(e) and 4.1(h).

Finally, the runs in the rightmost column—Figures 4.1(c), 4.1(f), and 4.1(i)—illustrate how change in another (different) single parameter can make epochal evolution dynamics reappear. In run 4.1(c) M was lowered to 300, compared with run 4.1(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 4.1(f) M was raised from 500 to 7500. Here the noise seen in run 4.1(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 4.1(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 4.1(i) the block size was increased to $K = 15$. Here the fitness epochs that disappeared in run 4.1(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 4.1(g), the bounded fluctuations seen in runs 4.1(e) and 4.1(h), and the relatively smooth dynamics of run 4.1(b) and run 4.1(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 4.1, 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.

4.4 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 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 4.5.1 we will give theoretical arguments in support of the maximum entropy

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

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,

$$\sum_{f=0}^{N} P_f = 1. \tag{4.2}$$

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

$$\langle f \rangle = \sum_{f=0}^{N} f P_f. \tag{4.3}$$

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

$$\Lambda_M = \{\vec{P} \,:\, P_f = \frac{n_f}{M},\ n_f \in \mathbb{N},\ \sum_{f=0}^{N} n_f = M\}, \tag{4.4}$$

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

$$|\Lambda_M| = \binom{M+N}{N}, \tag{4.5}$$

which is on the order of M^N states for $N \ll M$. The above set Λ_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:

$$\Lambda_\infty = \{\vec{P} \in \mathbb{R}^{N+1} \,:\, \sum_{f=0}^{N} P_f = 1,\ P_f \geq 0\}. \tag{4.6}$$

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.

4.5 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 $\mathbf{G}$ 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 $\mathbf{G}$ on the state $\vec{P}(t)$ of the population at time t:

$$\vec{P}(t+1) = \mathbf{G}[\vec{P}(t)], \tag{4.7}$$

where the operator $\mathbf{G}$ formally can be written as the product,

$$\mathbf{G} = \mathbf{M} \cdot \mathbf{S}, \tag{4.8}$$

of the selection operator $\mathbf{S}$ and the mutation operator $\mathbf{M}$.

This construction is similar in spirit to the generation operators in [139], 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 [117], 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 4.7 is analogous to a discrete-time version of the differential equations of the Eigen model of molecular evolution [32, 34]. 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 $\mathbf{G}$ 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 $\mathbf{G}$. Once we have this composite expression for $\mathbf{G}$, we turn to explicitly solving for the dynamics.

4.5.1 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:

$$A = \frac{1}{2^K - 1} \sum_{d=1}^{K} \binom{K}{d} q^d (1-q)^{K-d} = \frac{1 - (1-q)^K}{2^K - 1}. \tag{4.9}$$

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 4.9 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:

$$A_s \geq A. \tag{4.10}$$

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

$$A_s \leq q(1-q)^{K-1}. \tag{4.11}$$

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.

4.5.2 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

$$D = 1 - (1-q)^K. \tag{4.12}$$

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

$$\mathbf{M}_{ij} = \sum_{k=0}^{N-j} \sum_{l=0}^{j} \delta_{j+k-l,i} \binom{N-j}{k} \binom{j}{l} A^k (1-A)^{N-j-k} D^l (1-D)^{j-l}. \tag{4.13}$$

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:

$$P_i^m = \sum_{j=0}^{N} \mathbf{M}_{ij} P_j, \tag{4.14}$$

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

$$\sum_{i=0}^{N} \mathbf{M}_{ij} = 1. \tag{4.15}$$

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.

4.5.3 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:

$$P_i^s = c\, i\, P_i, \tag{4.16}$$

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:

$$\sum_{i=0}^{N} P_i^s = 1. \tag{4.17}$$

Therefore, we have

$$c = \left[\sum_{i=0}^{N} i P_i \right]^{-1} = \frac{1}{\langle f \rangle}, \tag{4.18}$$

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

$$\mathbf{S}_{ij} = \frac{i\delta_{ij}}{\langle f \rangle}. \tag{4.19}$$

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.

[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.

4.5.4 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:

$$\mathbf{G}_{ij} = \sum_{k=0}^{N} \mathbf{M}_{ik}\mathbf{S}_{kj}. \tag{4.20}$$

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

$$\mathbf{S} = \frac{1}{\langle f\rangle}\tilde{\mathbf{S}}, \tag{4.21}$$

giving

$$\tilde{\mathbf{G}} = \mathbf{M}\cdot\tilde{\mathbf{S}}. \tag{4.22}$$

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,

$$\vec{P}(t) = \mathbf{G}^t[\vec{P}(0)]. \tag{4.23}$$

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

$$\vec{P}_i(t) = \sum_j C[t,\vec{P}(0)]\tilde{\mathbf{G}}^t_{ij}P_j(0). \tag{4.24}$$

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

$$C[t,\vec{P}(0)] = \left[\sum_{i,j}\tilde{\mathbf{G}}^t_{ij}P_j(0)\right]^{-1}. \tag{4.25}$$

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

$$P_i(0) = \binom{N}{i}2^{-Ki}\left(1-2^{-K}\right)^{N-i}. \tag{4.26}$$

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

[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.

$\vec{V}^i$ and their corresponding eigenvalues g_i. These obey

$$\tilde{\mathbf{G}} \cdot \vec{V}^i = g_i \vec{V}^i, \tag{4.27}$$

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

$$\sum_{i=0}^{N} V_i^j = 1. \tag{4.28}$$

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

$$\mathbf{G}'_{ij} = (\mathbf{R}^{-1} \cdot \tilde{\mathbf{G}} \cdot \mathbf{R})_{ij} = g_i \delta_{ij}. \tag{4.29}$$

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

$$\sum_{i=0}^{N} \mathbf{R}_{ij} = \sum_{i=0}^{N} V_i^j = 1. \tag{4.30}$$

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

$$\tilde{\mathbf{G}}^t_{ij} = \sum_{k=0}^{N} \mathbf{R}_{ik} g_k^t \mathbf{R}^{-1}_{kj}. \tag{4.31}$$

This allows us to solve for $\vec{P}(t)$ in equation 4.24, obtaining

$$\vec{P}_i(t) = C[t, \vec{P}(0)] \sum_{k,j} \mathbf{R}_{ik} g_k^t \mathbf{R}^{-1}_{kj} P_j(0). \tag{4.32}$$

Equation 4.32 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 4.32 more transparent by moving to the basis of $\tilde{\mathbf{G}}$'s eigenvectors. First, we write the fitness distributions in this basis as

$$\alpha_i(t) = \sum_{j=0}^{N} \mathbf{R}^{-1}_{ij} P_j(t). \tag{4.33}$$

Note that the $\alpha_i(t)$ are normalized. We further simplify the expression for $C[t, \vec{P}(0)]$ using equations 4.31 and 4.33 in equation 4.25:

$$C[t, \vec{P}(0)] = \left[\sum_{i,j,k=0}^{N} \mathbf{R}_{ik} g_k^t \mathbf{R}^{-1}_{kj} P_j(0) \right]^{-1} = \left[\sum_{k=0}^{N} g_k^t \alpha_k(0) \right]^{-1}. \tag{4.34}$$

Transforming the equations of motion (equation 4.32) to the basis of the eigenvectors $\vec{V}^i$ we find that the fitness distribution is given by

$$\alpha_i(t) = \frac{g_i^t \alpha_i(0)}{\sum_{j=0}^{N} g_j^t \alpha_j(0)}. \tag{4.35}$$

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]

$$\langle f \rangle = \sum_i i P_i = \sum_i g_i \alpha_i. \tag{4.36}$$

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:

$$\langle f(t) \rangle = \frac{\sum_i g_i^{t+1} \alpha_i(0)}{\sum_j g_j^t \alpha_j(0)}. \tag{4.37}$$

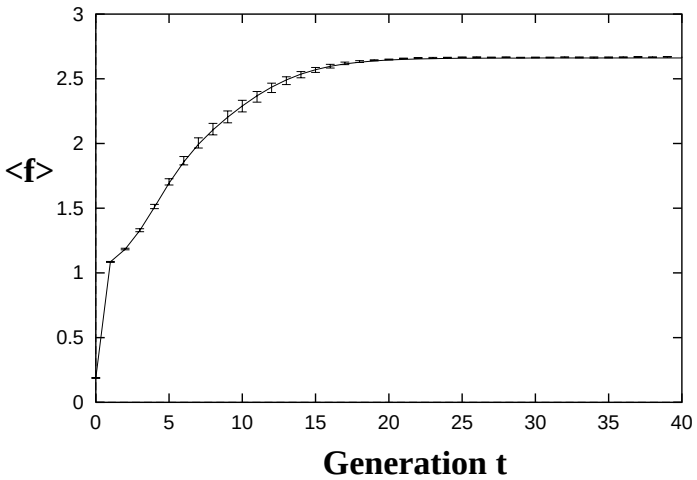

Figure 4.2: Average fitness $\langle f(t) \rangle$ over time t averaged over 20 GA runs along with the theoretical prediction (solid line) for infinite population obtained from equation 4.37. The error-bars show ± 2 standard deviations from the average fitness over the 20 runs. The parameters here are $N = 3$ and $K = 4$ with $q = 0.01$ and a population size of 10^4.

Figure 4.2 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 D 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

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

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 4.35) and (ii) the average fitness $\langle f(t) \rangle$ in the population (equation 4.37).

4.5.5 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

$$\mathbf{G}(\vec{P}) = \frac{\tilde{\mathbf{G}}}{\langle f \rangle} \cdot \vec{P} \, ; \tag{4.38}$$

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

$$\tilde{\mathbf{G}} \cdot \vec{P} = \langle f \rangle \vec{P}. \tag{4.39}$$

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 4.15), we find that *all* eigenvectors $\vec{V}^i$ of $\tilde{\mathbf{G}}$ fulfill the above restriction (equation 4.39). First, using equation 4.27, note that

$$\sum_{i,j} \tilde{\mathbf{G}}_{ij} V_j^k = \sum_i g_k V_i^k = g_k. \tag{4.40}$$

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

$$\sum_{i,j} \tilde{\mathbf{G}}_{ij} V_j^k = \sum_{i,j} \mathbf{M}_{ij} j V_j^k = \sum_j j V_j^k = \langle f \rangle. \tag{4.41}$$

Thus, we see that

$$g_k = \langle f \rangle \tag{4.42}$$

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.

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

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 C 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 Λ_∞ 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 4.37, 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,

$$\lim_{t\to\infty} \langle f(t) \rangle = g_N. \tag{4.43}$$

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

$$\lim_{t\to\infty} \vec{P}(t) = \vec{V}^N. \tag{4.44}$$

Globally, the map $\mathbf{G}$ has one fixed point lying inside the state space Λ_∞ 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 [32]. 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 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.

4.5.6 GA Dynamics as a Flow in Fitness Distribution Space

With equation 4.32 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 4.2. It does not exhibit the fitness epochs illustrated in figure 4.1. 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

[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 [79] for a discussion of the implications of this *not* being the case.

dynamics as a flow in the space of fitness distributions Λ_∞. As a simple example, figure 4.3 illustrates the dynamics for the parameter settings $N = 3$, $K = 4$, and $q = 0.01$. The state space Λ_∞ can be projected into 3-dimensional Euclidean space for this case. Figure 4.3 has P_0 on the x-axis, P_1 on the y-axis, and P_2 on the z-axis. Of course, we have $P_3 = 1 - P_0 - P_1 - P_2$.

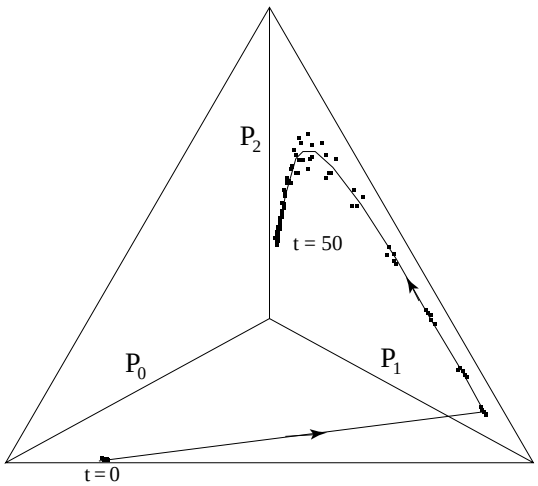

Figure 4.3: Infinite population dynamics as a trajectory (solid line) in Λ_∞ for $N = 3$, $K = 4$, and $q = 0.01$, together with the results (squares) of five runs with these parameters and $M = 10^4$. Cf. figure 4.2, which plots the average fitness dynamics.

Figure 4.3 shows the theoretical infinite population trajectory of the dynamics in Λ_∞ 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 4.2.) 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 4.4 shows $d\vec{P}$ in the simplex for the same parameter settings as in figure 4.3. We now see how the trajectory as shown in figure 4.3 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

[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.

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.

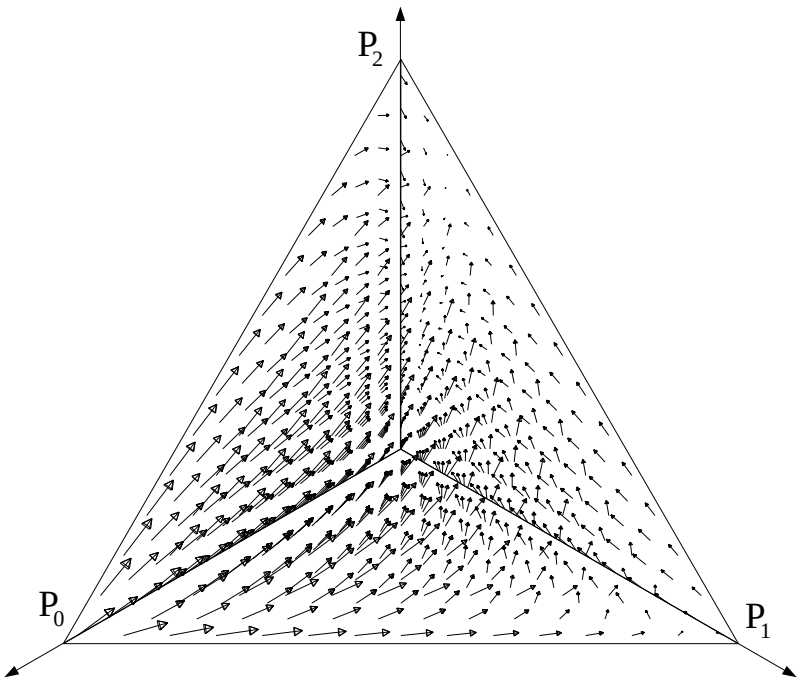

Figure 4.4: Fitness distribution flow $d\vec{P}$ in the simplex for $N = 3$, $K = 4$, and $q = 0.01$.

4.6 Finite Population Dynamics

As we saw in section 4.4, the fitness distribution state space for a finite population is given by a rectangular lattice Λ_M over the $N+1$ dimensional simplex Λ_∞ 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 transition probability $\Pr[\vec{P}^n \to \vec{P}^m]$ 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 $\mathbf{G}(\vec{P}^n)$:

$$\Pr[\vec{P}^n \to \vec{P}^m] = M! \prod_{i=0}^{N} \frac{[\mathbf{G}_i(\vec{P}^n)]^{m_i}}{m_i!}, \tag{4.45}$$

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 D this distribution is used to prove that for large populations the finite population dynamics approaches the infinite population dynamics arbitrarily closely.

4.6.1 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 Λ_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

$$|\langle dP_3 \rangle| \approx 6.5 * 10^{-4} \ll \frac{1}{M} = 0.01. \tag{4.46}$$

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 4.5 and 4.6 illustrate this behavior. Figure 4.5 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 4.6 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

[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 4.6.3.

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.

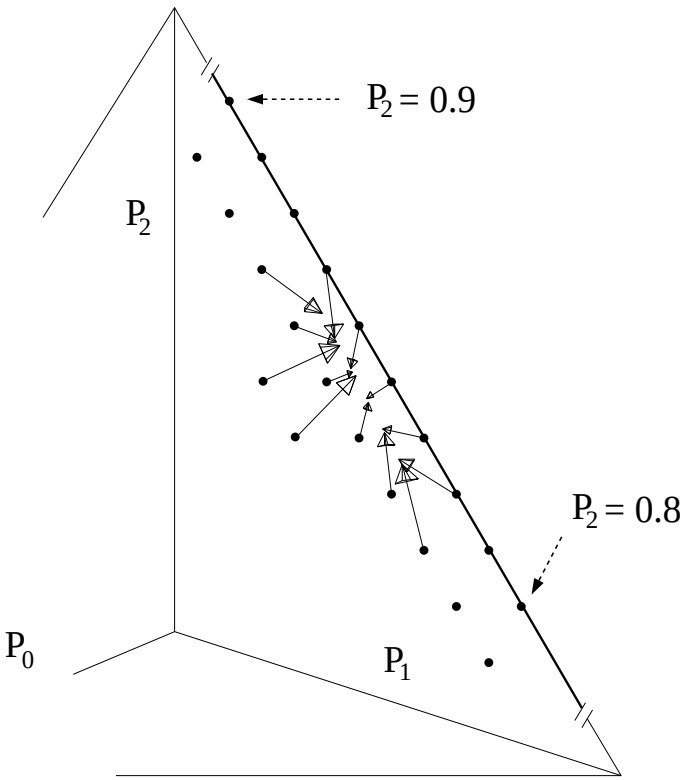

Figure 4.5: Face view of the flow $\langle d\vec{P} \rangle$ within the surface $P_3 = 0$. The flow of fitness distributions centers around an area where $P_2 \approx 0.85$, $P_1 \approx 0.14$ and $P_0 \approx 0.01$. This region is shown substantially enlarged. Parameters are set to $N = 3$, $K = 4$, $q = 0.01$, and $M = 100$. The dots indicate the allowed finite populations on a portion of Λ_M.

We would like to have a way to determine *where* metastable regions like the one shown in figure 4.6 are to be found in the state space. In section 4.5.5 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.

4.6.2 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

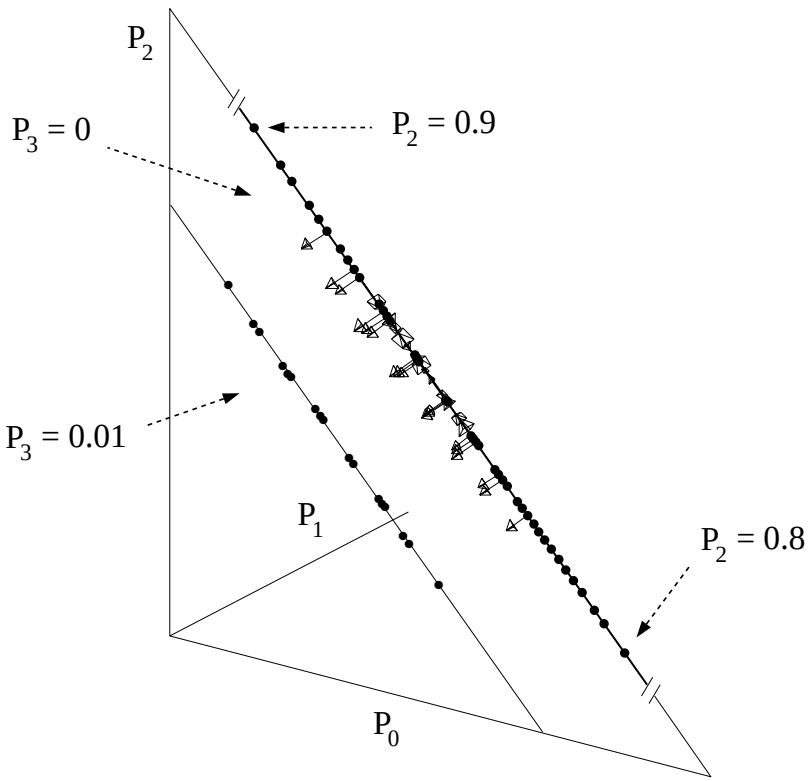

Figure 4.6: Side view of the simplex perpendicular to the surfaces $P_3 = 0$ and $P_3 = 0.01$, which are shown as the upper and lower diagonal solid lines, respectively. The dots indicate the allowed finite populations in Λ_M. The expected change $\langle dP_3 \rangle$ (arrows) is so small that the dynamics 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 $M = 100$.

that will be explained in subsequent sections.

Figures 4.7(a) through 4.7(i) show the same runs as those in figures 4.1(a) through 4.1(i) together with the theoretical predictions of the fitness levels (horizontal lines) at which we expect the 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 4.7: 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 4.1(a) through 4.1(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 4.6.4 we calculate simple analytical expressions for these epoch levels.

In runs 4.7(a), 4.7(d), and 4.7(g) (first column of figure 4.7), 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 4.7(g) need to be explained.

For the runs in the middle column (figures 4.7(b), 4.7(e), and 4.7(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 4.7(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 4.5; epochs do not occur. But we must explain why the epochs appear in run 4.7(a) but disappear in run 4.7(b) as a result of decreasing the block size K to 3.

For run 4.7(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

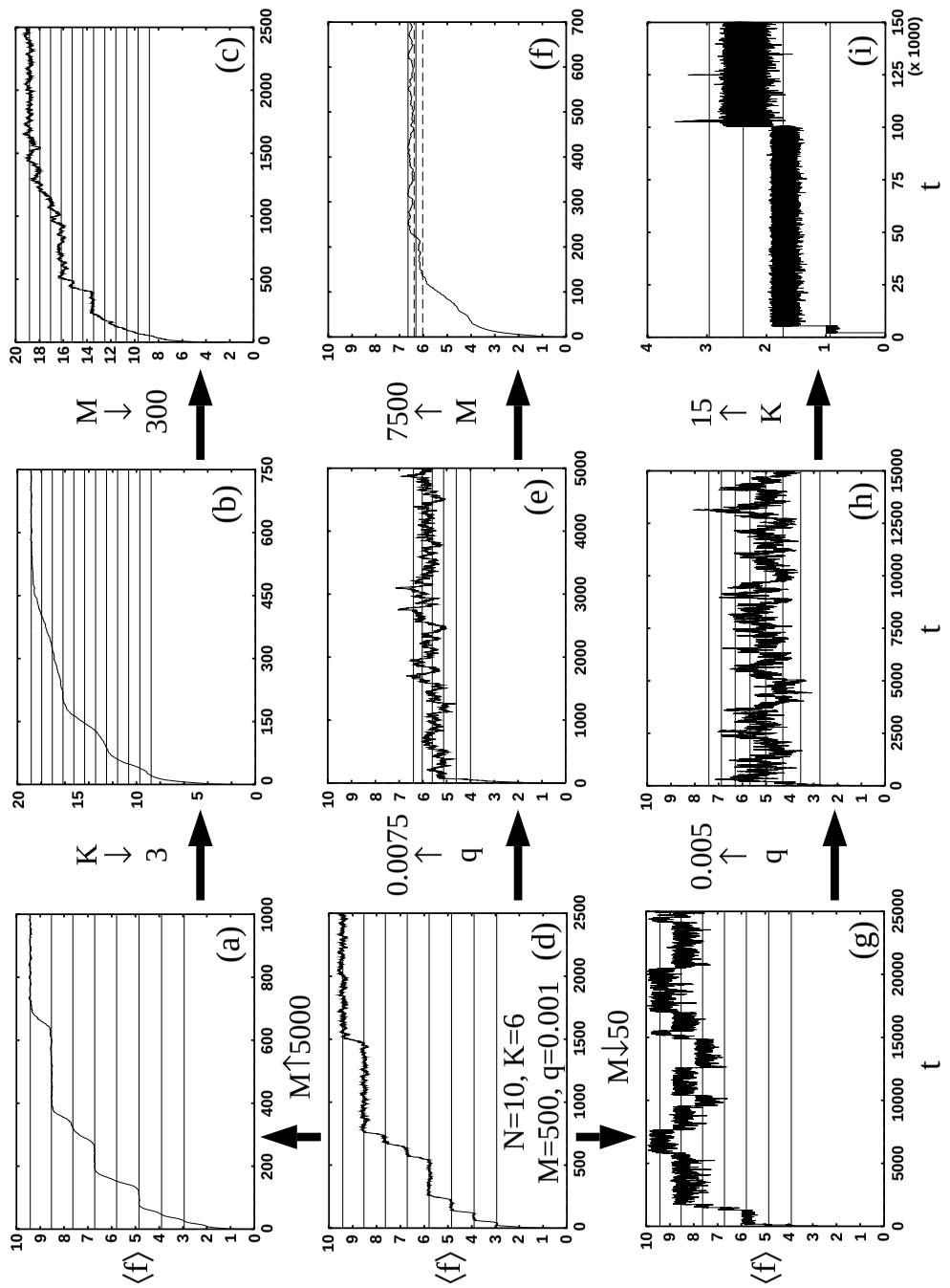

Figure 4.7: For caption see text.

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 4.7(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 4.7(e).

The theoretical epoch fitnesses plotted in run 4.7(c) correctly predict the levels of the epochs that can be distinguished in this case as well as the asymptotic fitness level.

In run 4.7(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 4.5.1 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 4.7(f). These upper and lower bounds still give quite accurate predictions for the observed fitness levels.

Finally, in run 4.7(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.

4.6.3 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 Λ_∞. 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 Λ_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

$$\sum_{i=0}^{n} P_i = 1, P_i \geq 0. \tag{4.47}$$

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 $\mathbf{G}^n$:

$$\mathbf{G}^n_{ij} = \begin{cases} \mathbf{G}_{ij} & \text{if } i,j \leq n \\ 0 & \text{otherwise.} \end{cases} \tag{4.48}$$

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,

$$\tilde{\mathbf{G}}^n \cdot \vec{P}^n = e_n \vec{P}^n, \tag{4.49}$$

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.

4.6.4 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

$$\tilde{\mathbf{G}}_{ij} = j\left[\delta_{ij}\left(1 - q(A_1(N-j) + Kj)\right) + \delta_{(i-1)j} A_1 (N-j) q + \delta_{(i+1)j} K j q\right], \tag{4.50}$$

plus correction terms of order $\mathcal{O}(q^2)$ and $A_1 q$ is the probability A to align a block to first-order in q (from equation 4.9):

$$A = A_1 q + \mathcal{O}(q^2) = \frac{K}{2^K - 1} q + \mathcal{O}(q^2) \tag{4.51}$$

Formally, we can split equation 4.50 into a term that is zeroth order in q and a term that is first order in q:

$$\tilde{\mathbf{G}} = \tilde{\mathbf{H}}^{\mathbf{0}} + q\tilde{\mathbf{H}}^{\mathbf{1}} + \mathcal{O}(q^2), \tag{4.52}$$

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:

$$P^n_i \equiv \delta_{in} + q\Delta^1_i \tag{4.53}$$

and

$$e_n = n + qe^1_n. \tag{4.54}$$

Using the eigenvalue equation

$$\tilde{\mathbf{G}} \cdot \vec{P}^n = e_n \vec{P}^n, \tag{4.55}$$

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

$$P^n_n = (1 - n^2Kq) \text{ and } P^n_{n-1} = n^2Kq. \tag{4.56}$$

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

$$e_n = n - \left[n^2K + n(N-n)A_1\right] q. \tag{4.57}$$

For the average fitness

$$f_n = \sum_{i=0}^{n} iP^n_i, \tag{4.58}$$

we obtain

$$f_n = n - n^2Kq. \tag{4.59}$$

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:

$$P^n_n \to \frac{n}{f_n} P^n_n \to P^n_n \frac{n}{f_n}(1-q)^{nK}, \tag{4.60}$$

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:

$$f_n = n(1-q)^{nK}. \tag{4.61}$$

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 4.59 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 [33, 101], 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., [101], 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 [79] 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 [20]. In the evolutionary setting the simplices unfold; in the particle case they collapse, due to annihilation.

4.6.5 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 4.8 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 4.8 it is clear that the same theoretical predictions for the

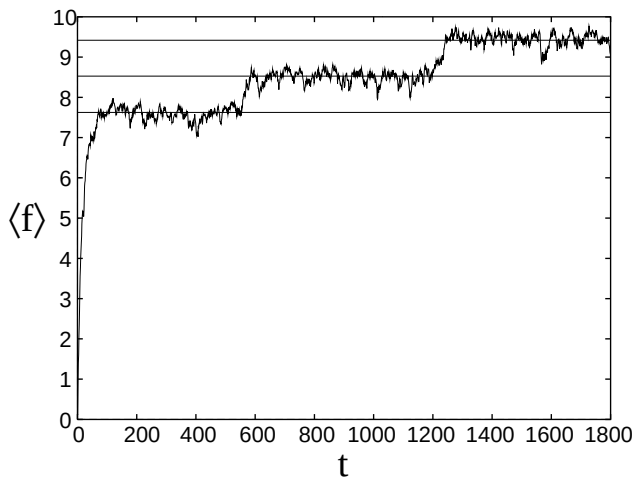

Figure 4.8: A run of the GA with single-point crossover, displaying the average fitness (solid curve) as a function of generation, together with the theoretically predicted epoch fitness levels (solid lines) for the 8^{th}, 9^{th}, and 10^{th} epochs. The parameters for this run are $N = 10$, $K = 6$, $q = 0.001$, and $M = 100$. Cf. figures 4.1(d) and 4.7(d).

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.

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.

4.6.6 Stable and Unstable Manifolds

It is clear from figure 4.1 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 Λ_∞ 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:

$$\mathbf{G}(\vec{V}+\vec{\epsilon}) = \vec{V} + \mathbf{DG}\cdot\vec{\epsilon}, \tag{4.62}$$

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

$$\mathbf{DG}_{ij}(\vec{V}^n) = \left[\frac{\partial \mathbf{G}_i(\vec{P})}{\partial P_j}\right]_{\vec{P}=\vec{V}^n} = \frac{\tilde{\mathbf{G}}_{ij} - jV_i^n}{g_n}. \tag{4.63}$$

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 4.62, since the deviations grow and shrink for eigenvalues greater or less than 1, respectively. By inspection of equation 4.63, 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 $\vec{V}^n$:

$$\vec{U}^i = \vec{V}^i - \vec{V}^n. \tag{4.64}$$

Substituting these vectors into the eigenvalue equation, we find that

$$\mathbf{DG}(\vec{V}^n)\cdot\vec{U}^i = \frac{g_i}{g_n}\,\vec{U}^i. \tag{4.65}$$

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

$$\lambda_i^n = \frac{g_i}{g_n}. \tag{4.66}$$

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 λ_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 4.61 to obtain analytical expressions for the eigenvalues of $\mathbf{DG}$ in the n^{th} fitness epoch. These are

$$\lambda_i^n = \frac{i}{n}(1-q)^{(i-n)K}\,, 0 \le i \le N\,. \tag{4.67}$$

[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.

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 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:

$$\Pr(f < n) = \Pr(P_i = 0,\ i \geq n). \tag{4.68}$$

We then have

$$\Pr(f < n) = \left[\sum_{i=0}^{n-1} \binom{N}{i} 2^{-Ki} \left(1 - 2^{-K}\right)^{N-i}\right]^M. \tag{4.69}$$

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 λ_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 4.67 we see that for larger q the eigenvalues λ_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 λ_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 4.1(d)). We will discuss the size of the fitness fluctuations in more detail in section 4.6.8 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 4.1(a) and 4.1(d)). Following up this line of qualitative analysis, in the next section we estimate the innovation times quantitatively from the values of λ_i^n.

Recapitulating, we have argued that the main qualitative features of epochal evolution can be understood in terms of the eigenvalues λ_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.

4.6.7 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 4.1 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 $\vec{P}^{n+1}$:

$$\vec{P}(t) = (1 - \alpha(t))\vec{P}^n + \alpha(t)\vec{P}^{n+1}, \tag{4.70}$$

with the initial condition $\alpha(0) = 1/M$. One should interpret α 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 4.70 and using the fact that the epochs correspond to eigenvectors of the generation operator, we find for the evolution of $\alpha(t)$ that

$$\alpha(t+1) = \frac{f_{n+1}}{f_{n+1}\alpha(t) + (1 - \alpha(t))f_n}\alpha(t). \tag{4.71}$$

Assuming that $\alpha(t)$ changes slowly and smoothly as a function of time, we can deduce

a differential equation from the above equation:

$$\frac{d\alpha}{dt} \approx \alpha(t+1) - \alpha(t) = \gamma_n \alpha \frac{1-\alpha}{\gamma_n \alpha + 1}, \tag{4.72}$$

where

$$\gamma_n = \frac{f_{n+1} - f_n}{f_n} \tag{4.73}$$

is the relative increase in fitness from the n^{th} to the $(n+1)^{st}$ epoch. It is possible to solve equation 4.72 analytically to obtain t as a function of α:

$$t = \frac{1}{\gamma_n} \left[\log(M\alpha) - (1+\gamma_n) \log \frac{M(1-\alpha)}{M-1} \right], \tag{4.74}$$

where we have used the boundary condition $\alpha(0) = 1/M$. In general it is impossible to invert equation 4.74 analytically to obtain α as a function of t. Nonetheless, we can use equation 4.74 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 α has reached the level $1 - 1/M$. Since we treat α 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

$$t_n = \frac{2+\gamma_n}{\gamma_n} \log[M-1]. \tag{4.75}$$

This equation has a simple interpretation. γ_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 γ_n to first order in the mutation rate q and setting $\log[M-1] \approx \log[M]$ we find that

$$t_n = [1 + 2n + 2Kn(n+1)q] \log M + \mathcal{O}(q^2). \tag{4.76}$$

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 4.72 by neglecting the small term $\gamma_n \alpha$ in the denominator.[10] Equation 4.72 then turns into the well known logistic growth equation, which can be solved analytically:

$$\alpha(t) = \frac{\exp(\gamma_n t)}{\exp(\gamma_n t) + M - 1}. \tag{4.77}$$

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

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

$$\langle f(t) \rangle = f_n + (f_{n+1} - f_n)\alpha(t). \tag{4.78}$$

Figures 4.9(a) and 4.9(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 4.9(a) and 4.9(b) give some examples of the empirically observed innovations from runs of the GA. Figure 4.9(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 4.9(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 4.9(b). The predicted innovation durations in the above cases are $t_3 \approx 32$ and $t_8 \approx 98$, respectively.

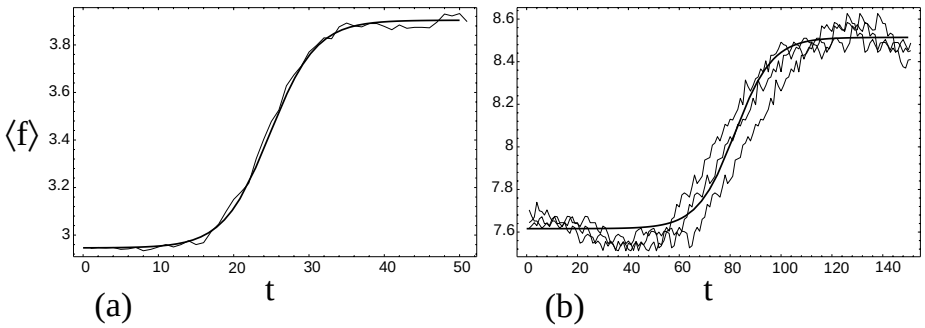

Figure 4.9: Innovation between the third and fourth (a) and eight and ninth (b) epochs for GA runs with parameters $N = 10$, $K = 6$, $q = 0.001$, and $M = 500$. (See figure 4.1(d).) The thick lines give the theoretical predictions of the innovation curves. The thinner lines show examples of these innovations from runs of the GA.

4.6.8 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 eigenvalues λ_i^n of epoch n and the fluctuations due to the finite population sampling occur as a Gaussian-noise diffusion. The idea of

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

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 [90]. This type of stochastic model assumes that gene frequencies, such as P_i or ϵ_i, can be approximated by real variables and that the gene frequencies change slowly over a single time step. (See [58] 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 Λ_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 $\vec{P}^n$:

$$\vec{P} = \vec{P}^n + \sum_{i=0}^{n-1} \epsilon_i \vec{U}^i, \tag{4.79}$$

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

$$\langle \vec{P}' \rangle = \mathbf{G}(\vec{P}) = \vec{P}^n + \sum_{i=0}^{n-1} \frac{f_i}{f_n} \epsilon_i \vec{U}^i. \tag{4.80}$$

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 $\vec{U}^i$:

$$\langle d\epsilon_i \rangle = \left(\frac{f_i}{f_n} - 1 \right) \epsilon_i. \tag{4.81}$$

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

$$\mu_i = \frac{f_n - f_i}{f_n}. \tag{4.82}$$

Then equation 4.81 takes on a simple form:

$$\langle d\epsilon_i \rangle = -\mu_i \epsilon_i. \tag{4.83}$$

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

Equation 4.83 gives the *expected* change of ϵ_i around the epoch center $\vec{P}^n$. Of course, there will be fluctuations in this change due to finite-size sampling. In equation 4.45 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:

$$\langle dP_i dP_j \rangle = \frac{P_i(\delta_{ij} - P_j)}{M}. \tag{4.84}$$

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 $\vec{U}^i$:

$$\epsilon_i = \sum_{i=0}^{n-1} \mathbf{R}_{ij}^{-1} P_j. \tag{4.85}$$

Note that the sum goes from 0 to $n-1$ and not up to n, since the component ϵ_n is not independent of the others. (It is determined from the requirement that the fitness distribution vector be normalized.) Using equations 4.84 and 4.85 the second moments of the change in the fluctuations are given by

$$\langle d\epsilon_i d\epsilon_j \rangle = \sum_{k,m=0}^{n-1} \mathbf{R}_{ik}^{-1} \mathbf{R}_{jm}^{-1} \frac{P_k(\delta_{km} - P_m)}{M}. \tag{4.86}$$

Assuming that the fluctuations ϵ_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

$$\langle d\epsilon_i d\epsilon_j \rangle = \sum_{k,m=0}^{n-1} \mathbf{R}_{ik}^{-1} \mathbf{R}_{jm}^{-1} \frac{P_k^n(\delta_{km} - P_m^n)}{M} \equiv \frac{\mathbf{B}_{ij}}{M}, \tag{4.87}$$

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 4.83 and 4.87 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 4.83 and the diffusion term given by equation 4.87. This is

$$\frac{\partial \Pr[\vec{\epsilon}, t]}{\partial t} = -\sum_i \frac{\partial \langle d\epsilon_i \rangle \Pr[\vec{\epsilon}, t]}{\partial \epsilon_i} + \frac{1}{2} \sum_{i,j} \frac{\partial^2 \langle d\epsilon_i d\epsilon_j \rangle \Pr[\vec{\epsilon}, t]}{\partial \epsilon_i \partial \epsilon_j}. \tag{4.88}$$

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 $\langle d\epsilon_i d\epsilon_j \rangle$ [136]. This is given by

$$\Pr[\vec{\epsilon}\,] = \frac{1}{\sqrt{(2\pi)^n Det[\mathbf{C}]}} \exp\left[-\frac{1}{2} \sum_{i,j=0}^{n-1} \epsilon_i \mathbf{C}_{ij}^{-1} \epsilon_j \right], \tag{4.89}$$

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

$$\mathbf{C}_{ij} = \langle \epsilon_i \epsilon_j \rangle = \frac{\mathbf{B}_{ij}}{M(\mu_i + \mu_j)}. \tag{4.90}$$

Of course, the means are all 0:

$$\langle \epsilon_i \rangle = 0. \tag{4.91}$$

Using equation 4.89 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

$$\mathrm{Var}[f] = \frac{Kn^3 q}{2M}. \tag{4.92}$$

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

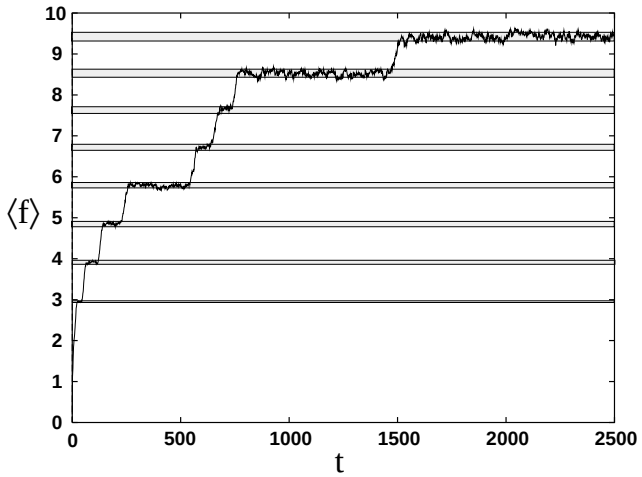

Figure 4.10: Predicted size of the epoch fitness fluctuations to two standard deviations, given by the grey bands. The parameters for this run are $N = 10$, $K = 6$, $q = 0.001$, and $M = 500$. This is the same run as was plotted in figure 4.1(d).

Figures 4.10, 4.11, and 4.12 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 4.10 shows the fitness fluctuations for our canonical parameter settings $N = 10$, $K = 6$, $q = 0.001$, and $M = 500$. Figure 4.11 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 4.12 has the same parameter settings as figure 4.10, except for a lower population size of $M = 50$. As predicted, the fitness fluctuations increase by a factor of approximately $\sqrt{10}$.

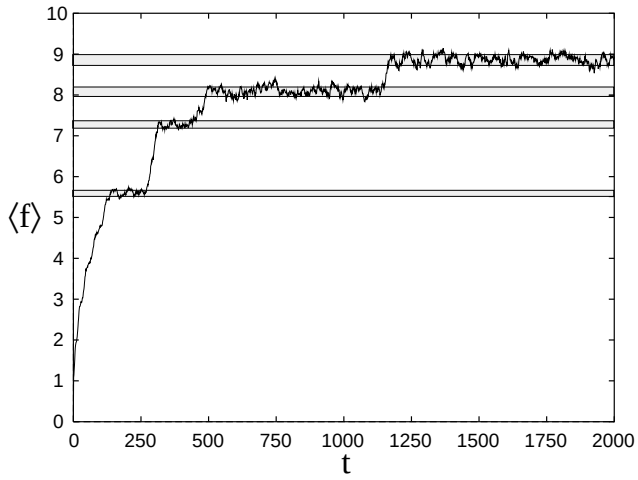

Figure 4.11: Predicted size of the fitness fluctuations during the epochs, plotted up to two standard deviations using the grey bands. The parameters for this run are $N = 10$, $K = 6$, $q = 0.002$, and $M = 500$. Cf. figure 4.10.

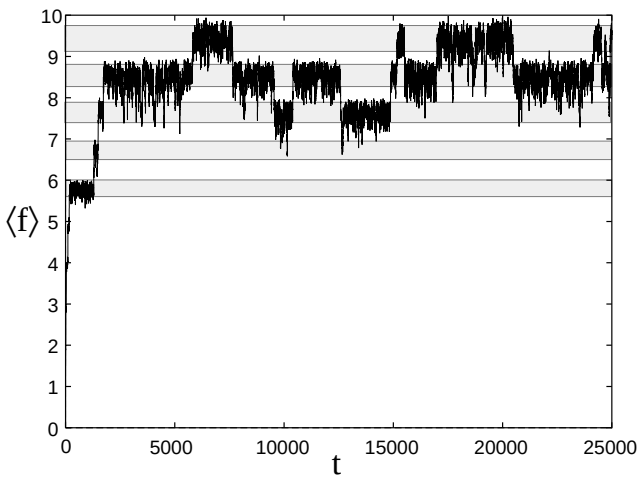

Figure 4.12: Predicted size of the fitness fluctuations during the epochs, plotted up to two standard deviations using the grey bands. The parameters for this run are $N = 10$, $K = 6$, $q = 0.001$, and $M = 50$. This is the same run as was plotted in figure 4.1(g).)

4.6.9 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:

$$P_n^n = (1 - n^2 Kq) + \mathcal{O}(q^2). \tag{4.93}$$

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 4.1(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 4.56, the probability $\Pr(f \neq n)$ that none of the strings will be in class n is given by

$$\Pr(f \neq n) = (1 - P_n^n)^M \approx (n^2 Kq)^M. \tag{4.94}$$

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 4.12, 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 ϵ_n in the component P_n of the fitness distribution near the epoch center. That is, let

$$P_n = P_n^n + \epsilon_n. \tag{4.95}$$

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

$$\epsilon_n = -P_n^n \sum_{i=0}^{n-1} \epsilon_i. \tag{4.96}$$

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

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

$$\langle P'_n \rangle = P_n^n - P_n^n \sum_{i=0}^{n-1} \frac{f_i}{f_n} \epsilon_i \equiv P_n^n + \lambda \epsilon_n, \tag{4.97}$$

where we have defined λ as the factor by which the fluctuation ϵ_n is scaled down. Of course, in general λ depends on the particular distribution of the fluctuations ϵ_i over the different directions. In the previous section we saw that the fluctuations ϵ_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

$$\lambda = \frac{\sum_{i=0}^{n-1} f_i \langle \epsilon_i^2 \rangle}{f_n \sum_{j=0}^{n-1} \langle \epsilon_j^2 \rangle}. \tag{4.98}$$

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

$$\langle d\epsilon_n \rangle = -(1-\lambda)\epsilon_n \equiv -\mu \epsilon_n, \tag{4.99}$$

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

$$\langle (d\epsilon_n)^2 \rangle = \langle (dP_n)^2 \rangle \approx \frac{P_n^n (1 - P_n^n)}{M}. \tag{4.100}$$

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 [54]). 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

$$T(0) = \frac{M P_n^n}{1 - P_n^n} + \frac{\pi}{2\mu} \mathrm{erfi}\left[\sqrt{\frac{M\mu P_n^n}{1 - P_n^n}} \right] \mathrm{erf}\left[\sqrt{\frac{M\mu(1 - P_n^n)}{P_n^n}} \right], \tag{4.101}$$

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 [92].

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

Table 4.1: 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 4.1(g).

Table 4.1 shows the average times $T_n(0)$, starting from the epoch center $\epsilon = 0$, for some of the epochs of figure 4.1(g) with $N = 10$, $K = 6$, $q = 0.001$, and $M = 50$.

Comparing these average destabilization times with figure 4.1(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, nicely explain the occurrence of the intermittent behavior seen around epochs 9 and 10 in this run.

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

Table 4.2: 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 4.1(h).

Table 4.2 shows the average destabilization times for some of the epochs of figure 4.1(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 4.1(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 4.1(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 4.1(e)) we find that epoch 7 is still stable for 48×10^3 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 4.1(e).

Destabilization is very closely related to the so-called "error threshold" of self-replicating molecules in the theory of molecular evolution [133]. 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 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.

4.7 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 [44].

4.7.1 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:

$$\Pr[in] = \sum_{i=0}^{n} \mathbf{G}_i(\vec{P}^n) = \sum_{i=0}^{n} \mathbf{G}_i^n(\vec{P}^n) = \sum_i \frac{e_n}{f_n} \vec{P}_i^n = \frac{e_n}{f_n} \,. \tag{4.102}$$

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 ith 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

$$\Pr[out] = 1 - \left[\frac{e_n}{f_n}\right]^M. \tag{4.103}$$

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 τ_n until the population jumps out of the epoch n subsimplex is then given by

$$\tau_n = \frac{1}{Pr[out]} = \frac{f_n^M}{f_n^M - e_n^M}. \tag{4.104}$$

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:

$$\tau_n = \frac{1}{M(N-n)A}. \tag{4.105}$$

Thus, the expected number of generations τ_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.

4.7.2 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 4.6.7 we see that the expected change $\langle dP_{n+1} \rangle$ per time step is given by

$$\langle dP_{n+1} \rangle = \frac{f_{n+1} - f_n}{f_n} P_{n+1} = \gamma_n P_{n+1}, \tag{4.106}$$

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

$$\langle (dP_{n+1})^2 \rangle = \frac{P_{n+1}(1 - P_{n+1})}{M}. \tag{4.107}$$

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:

$$\frac{\partial \pi(p,t)}{\partial t} = \langle dP_{n+1} \rangle \frac{\partial \pi(p,t)}{\partial p} + \frac{\langle (dP_{n+1})^2 \rangle}{2} \frac{\partial^2 \pi(p,t)}{\partial p^2}, \tag{4.108}$$

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 [89] in the context of the drift of the frequency of a certain genotype in a population. The solution is

$$\pi(p) = \frac{\int_0^p G(x)dx}{\int_0^{P_{n+1}^{n+1}} G(x)dx}, \tag{4.109}$$

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

$$G(x) = (1-x)^{2M\gamma_n}. \tag{4.110}$$

Performing the integral, we obtain

$$\pi_n \equiv \pi(\frac{1}{M}) = \frac{1-\left(1-\frac{1}{M}\right)^{2M\gamma_n+1}}{1-\left(1-P_{n+1}^{n+1}\right)^{2M\gamma_n+1}} \approx 1 - e^{-2\gamma_n}, \tag{4.111}$$

where we have set the initial proportion $p = 1/M$. The approximation on the right-hand side holds only for large population sizes. Equation 4.111 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

$$T_n = \frac{\tau_n}{\pi_n} = \frac{f_n^M}{(f_n^M - e_n^M)(1-\exp(-2\gamma_n))}. \tag{4.112}$$

For small q or large K this becomes

$$T_n = \frac{1}{M(N-n)A\left[1-\exp(-2/n)\right]}. \tag{4.113}$$

As we found in section 4.6.8, the fitness fluctuates around f_n during the n^{th} epoch in an approximately Gaussian way with the standard deviation given by

$$\sigma_n = \sqrt{\frac{Kn^3q}{2M}}. \tag{4.114}$$

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

$$T_n(f) = \frac{T_n}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{f-f_n}{\sigma_n}\right)^2\right]. \tag{4.115}$$

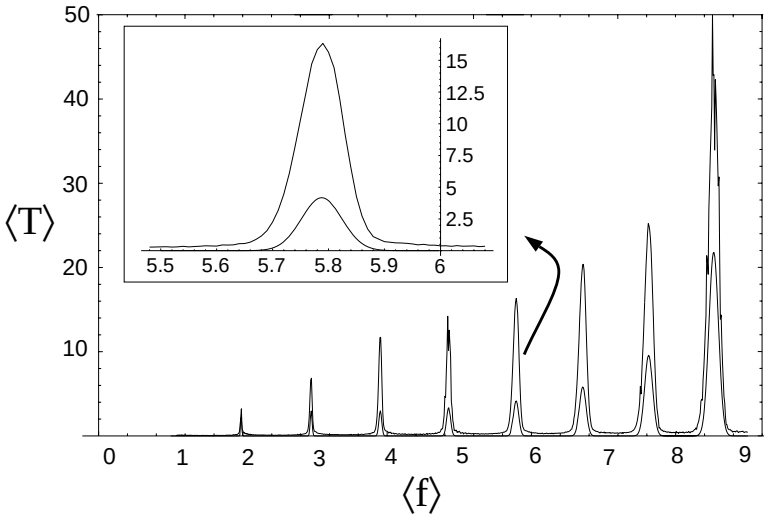

Figure 4.13: Empirical (upper curve) and theoretical (lower curve) fitness histograms. The horizontal axis shows the average fitness and the vertical axis shows the average number of time steps the population has that average fitness during a GA run, averaged over 500 runs. The parameters for this run 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.

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 4.13 shows the results of such an experiment for the parameter setting $N = 10$, $K = 6$, $M = 500$, and $q = 0.001$ of run 4.1(d) together with the theoretical predictions for the peaks at these parameter settings. The inset plot shows a magnification of the peak at the 6^{th} epoch. The upper curves are from the experiment and the lower curves plot the theoretical predictions (equation 4.115). 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 4.14 shows the results for the parameter setting $N = 20$, $K = 3$, $M = 300$, and $q = 0.001$ of run 4.1(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.

[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 $\langle dx \rangle / \langle dx^2 \rangle$.

Nonetheless, it is instructive to compare the above histogram with run 4.1(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 4.14 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 4.14 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 4.1(c). Averaging the average fitness at each time step from a large number of runs like that in figure 4.1(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.

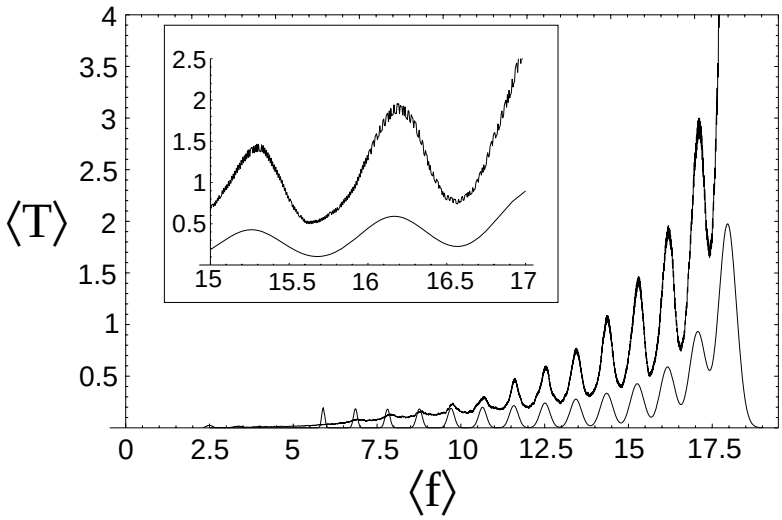

Figure 4.14: Empirical (upper curve) and theoretical (lower curve) fitness histograms. The horizontal axis shows the average fitness and the vertical axis shows the average number of time steps the population has that average fitness during the run, averaged over 500 runs. The parameters for this run 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 4.14 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 4.1(c). In obtaining the theoretical curve for figure 4.14 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 4.6.5 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 [70], 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.

4.8 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 $\mathbf{G}$. 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 4.3 presents an overview of the major analytical results we obtained for the different dynamical quantities, for small mutation rates q.

4.8.1 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 σ_n^2 of an epoch's average fitness fluctuation. The fluctuation amplitudes (σ_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.

$f_n = n(1-q)^{nK}$ (4.61)	Epoch fitness
$\sigma_n^2 = Kn^3q/2M$ (4.92)	Epoch fitness fluctuations
$P_n^n = 1 - n^2Kq$ and $P_{n-1}^n = n^2Kq$ (4.56)	Epoch population
$\lambda_i^n = f_i/f_n = i(1-q)^{(i-n)K}/n$ (4.67)	Epoch stability
$T_n = [M(N-n)A(1-\exp(-2/n))]^{-1}$ (4.113)	Epoch duration
$t_n = [1 + 2n + 2Kn(n+1)q]\log M$ (4.76)	Innovation duration

Table 4.3: 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.

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-1}^n of strings in the $(n-1)^{th}$ fitness class also grows with the same coefficient.

The fourth line gives the eigenvalues λ_i^n 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 $\mathbf{G}$ 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 $\mathbf{G}$. 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).

4.8.2 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 $\mathbf{G}$ 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 [64] 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 [149], 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" [114]. 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 [33]. 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 [91]. 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 4.6.4, 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.

4.8.3 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 [21]. 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 [105].) 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 [22, 27], 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.

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.

5

Optimizing Epochal Evolutionary Search: Population-Size Independent Theory

Erik van Nimwegen and James P. Crutchfield
Computer Methods in Applied Mechanics and Engineering,
special issue on Evolutionary and Genetic Algorithms
in Computational Mechanics and Engineering,
D. Goldberg and K. Deb, editors (in press 1999)
Santa Fe Institute Working Paper 98-06-046.

Epochal dynamics, in which long periods of stasis in population fitness are punctuated by sudden innovations, is a common behavior in both natural and artificial evolutionary processes. We use a recent quantitative mathematical analysis of epochal evolution to estimate, as a function of population size and mutation rate, the average number of fitness function evaluations to reach the global optimum. This is then used to derive estimates of and bounds on evolutionary parameters that minimize search effort.

5.1 Engineering Evolutionary Search

Evolutionary search refers to a class of stochastic optimization techniques—loosely based on processes believed to operate in biological evolution—that have been applied successfully to a variety of different problems; see, for example, Refs. [4, 11, 18, 28, 38, 49, 60, 94, 103] and references therein. Unfortunately, the mechanisms constraining and driving the population dynamics of evolutionary search on a given problem are not well understood.

In mathematical terms, evolutionary search algorithms are nonlinear stochastic population-based dynamical systems. The complicated dynamics exhibited by such systems has been appreciated for decades in the field of mathematical population genetics [69]. For example, the effects on evolutionary behavior of the rate of genetic variation, the population size, and the function to be optimized typically cannot be analyzed separately; there are strong, nonlinear interactions between them. These complications make an empirical approach to the question of whether and how to use evolutionary search problematic in general. The work presented here continues an attempt to unify and extend theoretical work that has been done in the areas of evolutionary search theory,

molecular evolution theory, and mathematical population genetics. The goal is to obtain a more general and quantitative understanding of the emergent mechanisms that control the population dynamics of evolutionary search and other population-based dynamical systems.

Here we focus on a class of population dynamics that we call *epochal evolution*. In epochal evolution, long periods of stasis in the average fitness of the population are punctuated by rapid innovations to higher fitness. In our analysis of this behavior [137] and chapter 4, we view large degeneracies in the genotype-fitness mapping to be the main source of the epochal nature of the evolutionary dynamics. For every genotype, we assume that there is a range of genotypic change that mutation can induce, without affecting phenotype or fitness. In this way, the space of genotypes is broken into strongly and weakly connected sets with respect to fitness. Sets of equal fitness genotypes are strongly connected, while sets of different fitness are only weakly connected to each other. Epochs occur because members of the evolving population have to search through most of the space of neutral variants before higher fitness genotypes are discovered. This structural view of the search space motivates several simplifying assumptions that lead to the class of fitness functions we analyze in the following. Stated in the simplest possible terms, all of the resulting population dynamical behavior derives from the interplay of this architecture, the infinite-population dynamics, and the stochasticity and discretization arising from finite-population sampling.

Considering our choice of focus, one might object that important details of real biological evolution, on the one hand, or of alternative evolutionary search algorithms, on the other hand, are not described by the resulting class of epochal evolutionary dynamical systems. Our response is simple: One must start somewhere. The bottom line is that the results and their predictive power justify the approach. Moreover, along the way we come to appreciate a number of fundamental trade-offs and basic mechanisms that drive and inhibit epochal evolutionary search.

Our results show that a detailed dynamical understanding, announced in Ref. [137] and expanded in chapter 4, can be turned to a very practical advantage. Specifically, we determine how to set mutation rate to reach, in the fewest steps, the global optimum in a wide class of fitness functions that give rise to epochal evolution. Moreover, the resulting detailed analytic understanding of the behavior identifies the mechanisms causing the dynamical instabilities that emerge when setting parameters for optimal, efficient search.

As we outline at the end, our analysis provides several insights that are useful knowledge for engineers applying evolutionary search to even more complicated optimization problems—and, for that matter, for the theory of evolutionary dynamics in biology—whenever genetic neutrality and degenerate genotype-to-fitness maps occur. Specifically, the results apply to cases in which epochal evolution is observed in the population dynamics.

5.2 Landscape Architecture

The fitness functions characteristic of problems that evolutionary search or (say) simulated annealing are being used for in practice are very complicated, almost by definition. On the one hand, detailed knowledge of the fitness function implies that one does not

need to run an optimization method to find high fitness solutions. On the other hand, assuming no knowledge at all leads to a completely random set of fitness functions for which it is known that *any* optimization algorithm performs as well on average as random search [146]. Not surprisingly, reality is a middle ground.

Therefore, our strategy to understand the workings of evolutionary search algorithms is to assume some specific structure in the fitness function that is germane to epochal evolution and to assume that, beyond this, there is no other structure. That is, apart from the structure we impose, the fitness function is as simple as can be.

There is a concomitant and compelling biological motivation for our focus on epochal dynamics. This is the common occurrence in natural evolutionary systems of "punctuated equilibria"—a process first introduced to describe sudden morphological changes in the paleontological record [64]. Similar behavior has been recently observed experimentally in bacterial colonies [37] and in simulations of the evolution of tRNA secondary structures [47]. This class of behavior appears general enough to also occur in artificial evolutionary systems, such as evolving cellular automata [22, 104] and populations of competing self-replicating computer programs [1].

How are we to think of the mechanisms that cause this evolutionary behavior? The evolutionary biologist Wright introduced the notion of "adaptive landscapes" to describe the (local) stochastic adaptation of populations to themselves and to environmental fluctuations and constraints [150]. This geographical metaphor has had a powerful influence on thinking about natural and artificial evolutionary processes. The basic picture is that the evolving population stochastically crawls along a surface determined, perhaps dynamically, by the fitness of individuals, moving to peaks and very occasionally hopping across fitness "valleys" to nearby, and hopefully higher fitness, peaks.

More recently, it has been assumed that the typical fitness functions of combinatorial optimization and biological evolution can be modeled as "rugged landscapes" [88, 100]. These are functions with wildly fluctuating fitnesses even at the smallest scales of single-point mutations. Moreover, it is assumed that these "landscapes" possess a large number of local optima. With this picture in mind, the common interpretation of punctuated equilibria in evolving populations is that of a population being "stuck" at a local peak in the landscape, until a rare mutant crosses a valley of relatively low fitness to a higher peak.

At the same time, an increasing appreciation has developed, in contrast to this rugged landscape view, that there are substantial degeneracies in the genotype-to-phenotype and the phenotype-to-fitness mappings. The crucial role played by these degeneracies has found important applications in molecular evolution; e.g. see Ref. [48]. When these degeneracies are operating, different genotypes in the population fall into a relatively small number of distinct fitness classes of genotypes with approximately equal fitness. Moreover, due to the high dimensionality of genotype spaces, sets of genotypes with approximately equal fitness tend to form components in genotype space that are connected by paths made of single-mutation steps. Note that, due to intrinsic or even exogenous effects (e.g. environmental), in addition there simply may not exist a deterministic "fitness" value for each genotype. In this case, fluctuations can induce variation in fitness such that genotypes with neighboring similar average fitness values are not distinct at the level of selection. Thus, genotype-to-fitness degeneracies also occur when there is noise in the fitness evaluation of individuals.

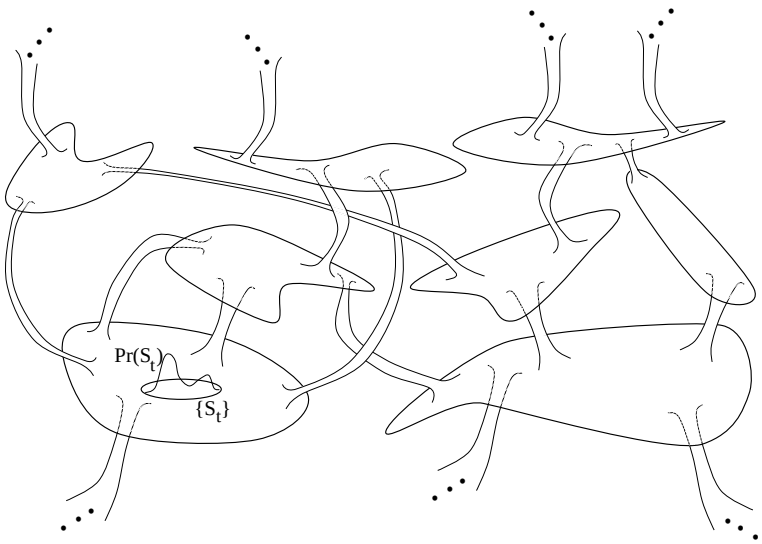

Figure 5.1: Subbasin and portal architecture underlying epochal evolution. Roughly speaking, a population diffuses in the subbasins (large sets) until a portal (a tube) to a higher fitness subbasin is found.

When these biological facts are taken into account we end up with an alternative view to both Wright's "adaptive landscapes" and the more recent "rugged landscapes". That is, the genotype space decomposes into a set of *neutral networks* or *subbasins* of approximately isofitness genotypes that are entangled with each other in a complicated and largely unstructured fashion; see Fig. 5.1. Within each neutral network selection is effectively disabled and neutral evolution dominates. This leads to a rather different interpretation of the processes underlying punctuated equilibria. Instead of the population being stuck at a local optimum in genotype space as suggested by the "landscape" models, the population drifts randomly over a neutral network of isofitness genotypes until a connection or *portal* to a neutral net of higher fitness is discovered [79].

Some of the first steps in understanding the biological consequences of neutral evolution, in single neutral networks, were taken by Kimura in the 1960's using stochastic process methods adapted from statistical physics [91]. Despite the early progress in neutral evolution, a number of fundamental problems remain only partly solved [31]. Although we will analyze neutral evolution in the following, we also emphasize the global architectural structure that connects the neutral networks and drives and constrains epochal evolutionary search. That is, the search space view of entangled neutral networks provides a natural and quite general architecture for studying epochal evolution.

This intuitive view of biologically plausible genotype-space architectures—as a relatively small number of connected neutral nets—is the one that we adopt in the following analysis. We formalize it by making several more specific (simplifying) assumptions

about the fitness function. First, we assume that there are $N+1$ different neutral nets, with fitnesses $1, 2, \ldots, N+1$. Second, we assume that the higher the fitness, the smaller the number of isofitness genotypes contained in the neutral net. That is, there are fewer genotypes of high fitness than low. More specifically, we assume that the proportion of genotype space that is occupied by genotypes of fitness n scales as 2^{-Kn}, where K is a measure of the rate at which the proportion of higher fitness genotypes decreases with fitness level. Finally, we assume that genotypes with fitness $n+1$ can be reached by a single point mutation from at least one genotype with fitness n. Specifically, we assume that the set of genotypes with fitness $n+1$ is a subspace of the set of genotypes with fitness n. The resulting architecture is a modified version of the general subbasin-portal structure of Fig. 5.1 and is illustrated in Fig. 5.2; after Ref. [20].

Why assume that genotypes of higher fitness are nested inside those of lower fitness? We believe that this assumption is consonant, by definition, with the very idea of using evolutionary search for optimization. Imagine, on the contrary, that genotypes of fitness $n+1$ are more likely to be close to genotypes with fitness $n-1$ than to those of fitness n. It then seems strange to have selection preferably replicate genotypes of fitness n over genotypes with fitness $n-1$. One result is that this leads to an increased (ineffective) search effort centered around genotypes of fitness n. Therefore, designing a search algorithm to select the current best genotypes implicitly assumes that genotypes of higher fitness tend to be found close to genotypes of the current best fitness.

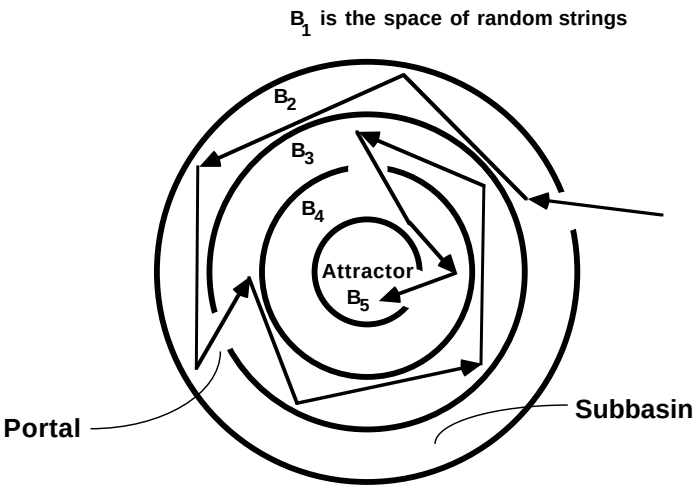

Figure 5.2: The dimensional hierarchy of subbasins and portals for the Royal Staircase fitness functions.

Note that in epochal dynamics there is a natural separation of time scales. Selection operates in short burst after which equilibrium on the level of fitness is established in the population. On a much slower time scale, as the members of the population diffuse through subbasins of isofitness genotypes, search occurs until a rare higher fitness genotype is discovered. In this way we shift our view away from the geographic metaphor of evolutionary search "crawling" along a fixed and static (smooth or rugged) "land-

scape" to the view of a diffusion process constrained by the subbasin-portal architecture induced by degeneracies in the genotype-to-phenotype and phenotype-to-fitness mappings. Moreover, as will become more apparent, our approach is not simply a shift in architecture, but it also focuses on the dynamics of populations as they move through the subbasins to find portals to higher fitness. A side benefit is that our approach is not limited to evolutionary processes for which a potential energy function exists, as the "landscape" models are.

5.3 The Royal Staircase Fitness Function

Under the assumptions specified in the previous paragraphs the class of fitness functions, referred to as the *Royal Staircase*, delineated is equivalent to the following specification:

1. Genomes are represented by binary strings $s = s_1 s_2 \cdots s_L, s_i \in \{0, 1\}$, of length $L = NK$.

2. Reading the genome from left to right, the number $I(s)$ of consecutive 1s is counted.

3. The fitness $f(s)$ of string s with I consecutive ones, followed by a zero, is $f(s) = 1 + \lfloor I(s)/K \rfloor$. The fitness is thus an integer between 1 and $N + 1$.

Note that the (single) global optimum is the genome $s = 1^L$; namely, the string of all 1s.

From this it is easy to see that we have chosen N (consecutive) sets of K bits to represent the different fitness classes. These sets we call *blocks*. The first block consists of the first K bits on the left, i.e. $s_1 \cdots s_K$. The second block consists of bits $s_{K+1} \cdots s_{2K}$ and so on. For each of these blocks there is one *aligned* configuration consisting of K 1s and $2^K - 1$ *unaligned* configurations. If the first block is unaligned, the string obtains fitness 1. If the first block is aligned and the second unaligned, it obtains fitness 2. If the first two blocks are aligned and the third unaligned, it obtains fitness 3, and so on up to the globally optimal string with all aligned blocks and fitness $N + 1$.

Without affecting the evolutionary dynamics or the underlying architecture of genotype space, we could have chosen a different aligned-block pattern than the K-1s configuration. In fact, we could have chosen different aligned configurations for the different blocks without affecting the dynamics. Furthermore, since we will not be analyzing position-dependent operators—such as, crossover—we could have chosen the locations of the bits of each block to be anywhere in the genome without affecting the dynamics. The only constraint, other than the block's ordering, is that we have N disjoint sets of K bits.

Notice that the proportion ρ_n of strings with fitness n is given by:

$$\rho_n = 2^{-K(n-1)} \left(1 - 2^{-K}\right) . \tag{5.1}$$

The net result is that the Royal Staircase fitness functions implement the intuitive idea that increasing fitness is obtained by setting more and more bits correctly. One can only set correct bit values in groups of K bits at a time and in blocks from left to right. A

genome's fitness is proportional to the number of blocks it has set correctly. Finally, note that due to the modularity of the subbasin-portal architecture, and of the resulting theory we present below, the restriction to uniform block size and equal fitness contributions of each block could also be lifted.

The Royal Staircase fitness functions thus realize our view of the underlying architecture as sets of isofitness genomes that occur in nested neutral networks of smaller and smaller volume; as shown in Fig. 5.2. Despite their simplicity, which is quite helpful for our population dynamics analyses, the preceding series of observations show that the Royal Staircase fitness functions describe a large class of genotype space architectures for epochal evolution.

5.4 The Genetic Algorithm

For our analysis of evolutionary search we have chosen a simplified form of a genetic algorithm (GA) that does not include crossover and that uses fitness-proportionate selection. The GA is defined by the following steps.

1. Generate a population of M bit strings of length $L = NK$ with uniform probability over the space of L-bit strings.
2. Evaluate the fitness of all strings in the population.
3. Stop the algorithm, noting the generation number $t_{\rm opt}$, if a string with optimal fitness $N + 1$ occurs in the population. Else, proceed.
4. Create a new population of M strings by selecting, with replacement and in proportion to fitness, strings from the current population.
5. Mutate each bit in each string of the new population with probability q.
6. Go to step 2.

The total number E of fitness function evaluations is $E = Mt_{\rm opt}$. We are interested in the average number E of evaluations per GA run required over a large number R of runs.

The main motivation for leaving out crossover is that this greatly simplifies our analysis. The benefit is that we can make detailed quantitative predictions of the GA's behavior. Moreover, we believe that from the point of view of optimization, the addition of crossover to the genetic operators only marginally improves the efficiency of the search for fitness functions that typically show epochal evolution. We comment on this later in section 5.8. Additional discussion and supporting evidence for this claim can be found in section 4.6.5 and in Ref. [104].

The selection method we use is simple fitness-proportionate selection. There is a wide variety of selection mechanisms used and studied in the evolutionary search literature [61]. In fact, many practitioners prefer other selection schemes over simple fitness-proportionate selection. Our use of fitness-proportionate selection is motivated by two observations. First, it is the most appropriate selection scheme for the analysis of biological systems, in which we are also interested. It is, therefore, the first selection

scheme that we study. Second, our analysis is most straightforward and the results easiest to understand for fitness-proportionate selection. Fortunately, our analyses can be readily extended to other selection schemes, such as truncation (elite) selection, ranking schemes, and tournament selection. We will report on these extensions of our methods elsewhere.

Notice that our GA effectively has two parameters: the mutation rate q and the population size M. A given search problem, then, is specified by the fitness function in terms of N and K. The central goal of the following analysis is to find those settings of M and q that minimize the number E of fitness function evaluations for given N and K.

5.5 Observed Population Dynamics

The typical dynamics of a population evolving on a landscape of connected neutral networks, such as defined above, alternates between long periods of stasis in the population's average fitness (*epochs*) and sudden increases in the average fitness (*innovations*). (See, for example, Fig. 1 of chapter 4.) As was first pointed out in the context of molecular evolution in Ref. [79], through neutral mutations, the best individuals in the population diffuse over the neutral network (*subbasin*) of isofitness genotypes until one of them discovers a connection (*portal*) to a neutral network of even higher fitness. The fraction of individuals on this new network then grows rapidly, reaching an equilibrium after which the new subset of most-fit individuals diffuses again over the new neutral network. In addition to the increasing attention paid to this type of epochal evolution in the theoretical biology community [47, 51, 77, 108, 50, 142], recently there has also been an increased interest by evolutionary search theorists [6, 70].

The GA just defined is the same as that studied in our earlier analyses [137] and chapter 4. Also, the Royal Staircase fitness function defined above is very similar to the *Royal Road* fitness function that we used there. It should not come as a surprise, therefore, that qualitatively the GA's experimentally observed behavior is very similar to that reported in Ref. [137] and chapter 4. Moreover, most of the theory developed there for epochal evolutionary dynamics carries over to the Royal Staircase class of fitness functions.

We now briefly recount the experimentally observed behavior of typical Royal Staircase GA runs. The reader is referred to chapter 4 for a detailed discussion of the dynamical regimes this type of GA exhibits under a range of different parameter settings.

Figure 5.3 shows the GA's behavior with four different parameter settings. The vertical axes show the best fitness in the population (upper lines) and the average fitness in the population (lower lines) as a function of the number of generations. Each figure is produced from a single GA run. In all of these runs the average fitness $\langle f \rangle$ in the population goes through stepwise changes early in the run, alternating epochs of stasis with sudden innovations in fitness. Later in the run, the average fitness has higher fluctuations that tend to obscure the epochal nature of the dynamics. Notice also that $\langle f \rangle$ roughly tracks the epochal behavior of the best fitness in the population. Notice, too, that the best fitness often shows a series of "spikes", corresponding to the discovery of a higher fitness string that is soon lost. Eventually, one of these discoveries *fixates* in the population. Finally, for each of the four settings of N and K we have chosen the

values of M and q such that the total number E of fitness function evaluations to reach the global optimum for the first time is roughly minimal. Thus, the four plots illustrate the GA's typical dynamics close to optimal (q, M)-parameter settings—the analysis for which begins in the next section.

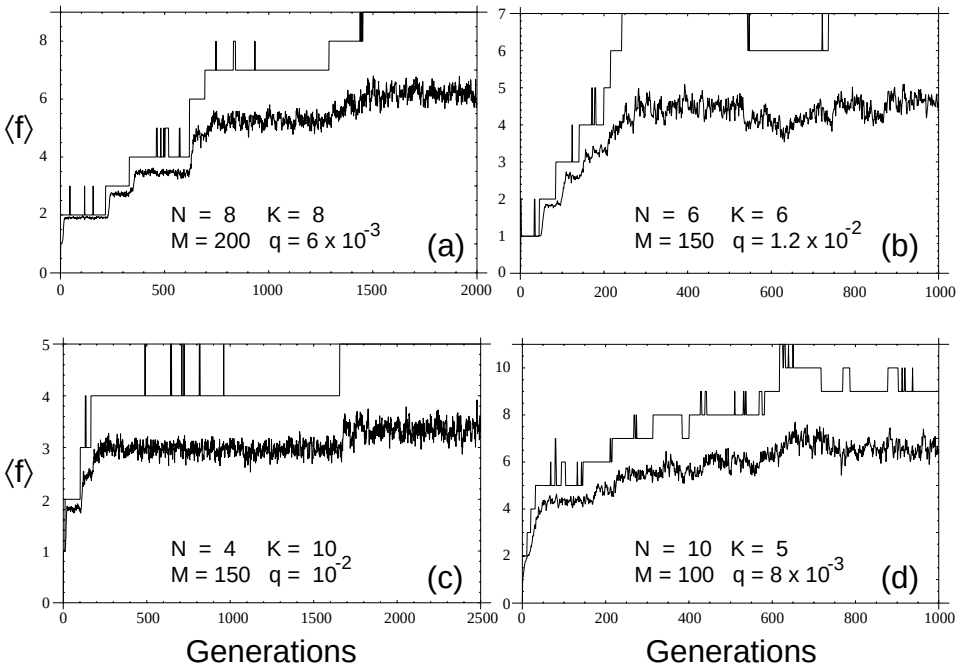

Figure 5.3: Examples of the Royal Staircase GA population dynamics with different parameter settings. The four plots show best fitness in the population (upper lines) and average fitness in the population (lower lines) as a function of time, measured in generations. The fitness function and GA parameters are given in each plot. In each case we have chosen q and M in the neighborhood of their optimal settings (see later) for each of the four values of N and K.

There is a large range, almost a factor of 10, in times to reach the global optimum across the runs. Thus, there can be a strong parameter dependence in search times. Moreover, the *standard deviation* of the total number E of fitness function evaluations is roughly as large as the average E. That is, there are large run-to-run variations in the time to reach the global optimum, even with the parameters held constant. This is true for all parameter settings with which we experimented, of which only a few are reported here.

Figure 5.3(a) plots the results of a GA run with $N = 8$ blocks of $K = 8$ bits each, a mutation rate of $q = 0.006$, and a population size of $M = 200$. During the epochs, the best fitness in the population jumps up and down several times before it finally jumps up and the new more-fit genotype stabilizes in the population. This transition is reflected in the average fitness also starting to move upward. In this particular run, it took the GA

approximately 2.9×10^5 fitness function evaluations to reach the global optimum for the first time. Over 250 runs the GA takes on average 5×10^5 fitness function evaluations to reach the global optimum for these parameters. The inherent large per-run variation means in this case that some runs take less than 10^5 function evaluations and that others take many more than 10^6.

Figure 5.3(b) plots a run with $N = 6$ blocks of length $K = 6$ bits, a mutation rate of $q = 0.012$, and a population size of $M = 150$. The GA reached the global optimum after approximately 3.6×10^4 fitness function evaluations. On average, the GA uses approximately 5×10^4 fitness function evaluations to reach the global fitness optimum. Notice that somewhere after generation 500 the global optimum is lost again from the population. It turns out that this is a typical feature of the GA's behavior for parameter settings close to those that give minimal E. The global fitness optimum often only occurs in relatively short bursts after which it is lost again from the population. Notice also that there is only a small difference in $\langle f \rangle$ depending whether the best fitness is either 6 or 7.

Figure 5.3(c) shows a run for a small number ($N = 4$) of large ($K = 10$) blocks. The mutation rate is $q = 0.01$ and the population size is again $M = 150$. As in all three other runs we see that the average fitness goes through epochs punctuated by rapid increases of average fitness. We also see that the best fitness in the population jumps up several times before the population fixates on strings with higher fitness. The GA takes about 2×10^5 fitness function evaluations on average to reach the global optimum for these parameter settings. In this run, the GA just happened to have taken about 2.5×10^5 fitness function evaluations.

Finally, Fig. 5.3(d) shows a run with a large number ($N = 10$) of smaller ($K = 5$) blocks. The mutation rate is $q = 0.008$ and the population size is $M = 100$. Notice that in this run, the best fitness in the population alternates several times between fitnesses 7, 8, and 9 before it reaches the global fitness optimum of 11. After it has reached the global optimum for several time steps, the global optimum disappears again and the best fitness in the population alternates between 9 and 10 from then on. It is notable that this intermittent behavior of the best fitness is barely discernible in the behavior of the average fitness. It appears to be lost in the "noise" of the average fitness fluctuations. The GA takes about 10^5 fitness function evaluations on average at these parameter settings to reach the global optimum; while in this particular run the GA took only 6.3×10^4 fitness function evaluations.

Again, we stress that there are large fluctuations in the total number of fitness evaluations to reach the global optimum between runs. One tentative conclusion is that, if one is only going to use a GA for a few runs on a specific problem, there is a relatively large operating regime in parameter space for which the GA's performance is statistically equivalent. On the one hand, the large fluctuations in the GA's search dynamics make it hard to predict for a single run how long it is going to take to reach the global optimum. On the other hand, as we will see below, this prediction is not very sensitive to the precise parameter settings.

5.6 Statistical Dynamics of Evolutionary Search

In Refs [137] and chapter 4 we developed the statistical dynamics of genetic algorithms to analyze the behavioral regimes of a GA searching the Royal Road fitness function. The analysis here builds on those results. Due to the strong similarities we will only briefly introduce this analytical approach to evolutionary dynamics. The reader is referred to chapter 4 for an extensive and detailed exposition. There the reader also will find a review of the connections and similarities of our work with the alternative methodologies for GA theory developed by Vose and collaborators [109, 139, 138], by Prügel-Bennett, Rattray, and Shapiro[117, 118, 119, 116], in the theory of molecular evolution [32, 33], and in mathematical population genetics.

From a microscopic point of view, the state of an evolving population is only fully described when a list $\mathcal{S}$ of all genotypes with their frequencies of occurrence in the population is given. On the microscopic level, the evolutionary dynamics is implemented via the conditional transition probabilities $\Pr(\mathcal{S}'|\mathcal{S})$ that the population at the next generation will be the microscopic state $\mathcal{S}'$; see Refs. [69] and [109]. For any reasonable genetic representation, however, there will be an enormous number of these microscopic states $\mathcal{S}$ and their transition probabilities. This makes it almost impossible to quantitatively study the dynamics at this microscopic level.

More practically, a full description of the dynamics on the level of microscopic states $\mathcal{S}$ is neither useful nor typically of interest. One is much more likely to be concerned with relatively coarse statistics of the dynamics, such as the evolution of the best and average fitness in the population or the waiting times for evolution to produce a genotype of a certain quality. The result is that quantitative mathematical analysis faces the task of finding a description of the evolutionary dynamics that is simple enough to be tractable numerically or analytically and that, moreover, facilitates the predicting the quantities of interest to a practitioner.

With these issues in mind, we specify the state of the population at any time by some relatively small set of *macroscopic* variables. Since this set of variables intentionally ignores vast amounts of microscopic detail, it is generally impossible to exactly describe the GA's dynamics in terms of these macroscopic variables. To achieve the benefits of a coarser description, however, we assume that, given the state of the macroscopic variables, the population has equal probabilities to be in any of the microscopic states consistent with the specified macroscopic state. This *maximum entropy* assumption lies at the heart of statistical mechanics in physics.

We assume in addition that in the limit of infinite population size, the resulting equations of motion for the macroscopic variables become closed. That is, for infinite populations, we assume that we can exactly predict the state of the macroscopic variables at the next generation, given the present state of the macroscopic variables. This limit is analogous to the thermodynamic limit in statistical mechanics. The corresponding assumption is analogous to *self-averaging* of the dynamics in this limit.

The key, and as yet unspecified step, in developing such a thermodynamic model of evolutionary dynamics is to find an appropriate set of macroscopic variables that satisfies the above assumptions. In practice this is difficult. Ultimately, the suitability of a set of macroscopic variables has to be verified by comparing theoretical predictions with experimental measurements. In choosing such a set of macroscopic variables one is

guided by our knowledge of the fitness function and the search's genetic operators.

To see how this choice is made imagine that strings can have only two possible values for fitness, f_A and f_B. Assume also that under mutation all strings of type A are equally likely to turn into type-B strings and that all strings of type B have equal probability to turn into strings of type A. In this situation, it is easy to see that we can take the macroscopic variables to be the relative proportions of A strings and B strings in the population. Any additional microscopic detail, such as the number of 1s in the strings, is not required or relevant to the evolutionary dynamics. Neither selection nor mutation distinguish different strings within the sets A and B on the level of the proportions of A's and B's they produce in the next generation.

Similarly, our approach describes the state of the population at any time only by the distribution of *fitness* in the population. That is, we group strings into equivalence classes of equal fitness and assume that, on the level of the fitness distribution, the dynamics treats all strings within a fitness class as equal. At the macroscopic (fitness) level of the dynamics, we know that a string of fitness n has the first $n-1$ blocks aligned and the nth block in one of the $2^K - 1$ other unaligned configurations. The maximum entropy assumption specifies that for all strings of fitness n, the nth block is *equally likely* to be in any of the $2^K - 1$ unaligned configurations and that each of the blocks $n+1$ through N are equally likely to be in any of their possible 2^K block configurations.

Various reasons suggest that the maximum entropy approximation will not be valid in practice. For example, the convergence due to finite-population sampling makes it hard to believe that all unaligned block configurations in all strings are random and *independent*. For large populations, fortunately, the assumption is compelling. In fact, in the limit of very large population sizes, typically $M > 2^L$, the GA's dynamics on the level of fitness distributions accurately captures the fitness distribution dynamics found experimentally (see chapter 4). In any case, we will solve explicitly for the fitness distribution dynamics in the limit of infinite populations using our maximum entropy assumption and then show how this solution can be used to solve for the finite-population dynamics.

The essence of our statistical dynamics approach then is to describe the population state at any time during a GA run by a relatively small number of macroscopic variables—variables that in the limit of infinite populations self-consistently describe the dynamics at their own level. After obtaining this infinite population dynamics explicitly, we then use it to solve for the GA's dynamical behaviors with finite populations.

Employing the maximum entropy principle and focusing on fitness distributions is also found in an alternative statistical mechanics approach to GA dynamics developed by Prügel-Bennett, Rattray, and Shapiro [117, 118, 119]. In their approach, however, maximum entropy is assumed with respect to the ensemble of entire GA *runs*. Specifically, in their analysis the *average* dynamics, averaged over many runs, of the first few cumulants of the fitness distribution are predicted theoretically. In more recent work, the model has been used to calculate ensemble cumulants, such as variance of the average dynamics over an ensemble of runs [116]. These macroscopic variables are generally unsuited, however, to capture the epochal behavior illustrated in Fig. 5.3 of the previous section. In contrast, our statistical dynamics approach applies maximum entropy only to the population's *current* state, given its current fitness distribution. The result is that for finite populations we do not assume, as we do for infinite populations, that the GA

dynamics is self-averaging. That is, two runs, with equal fitness distributions $\vec{P}$ at time t, are not assumed to have the same future macroscopic behavior. They are assumed only to have the same probability distribution of possible futures.

5.6.1 Generation Operator

The macroscopic state of the population is determined by its fitness distribution, denoted by a vector $\vec{P} = \{P_1, P_2, \ldots, P_{N+1}\}$, where the components $0 \leq P_n \leq 1$ are the proportion of individuals in the population with fitness $n = 1, 2, \ldots, N+1$. We refer to $\vec{P}$ as the *phenotypic quasispecies*, following its analog in molecular evolution theory [32, 33, 34]. Since $\vec{P}$ is a distribution, we have the normalization condition:

$$\sum_{n=1}^{N+1} P_n = 1. \tag{5.2}$$

The average fitness $\langle f \rangle$ of the population is given by:

$$\langle f \rangle = \sum_{n=1}^{N+1} n P_n. \tag{5.3}$$

In the limit of infinite populations and under our maximum entropy assumption, we can construct a generation operator $\mathbf{G}$ that maps the current fitness distribution $\vec{P}(t)$ deterministically into the fitness distribution $\vec{P}(t+1)$ at the next time step; that is,

$$\vec{P}(t+1) = \mathbf{G}[\vec{P}(t)] \; . \tag{5.4}$$

The operator $\mathbf{G}$ consists of a selection operator $\mathbf{S}$ and a mutation operator $\mathbf{M}$:

$$\mathbf{G} = \mathbf{M} \cdot \mathbf{S}. \tag{5.5}$$

The selection operator encodes the fitness-level effect of selection on the population; and the mutation operator, the fitness-level effect of mutation. Below we construct these operators for our GA and the Royal Staircase fitness function explicitly. For now we note that the infinite population dynamics can be obtained by iteratively applying the operator $\mathbf{G}$ to the initial fitness distribution $\vec{P}(0)$. Thus, the macroscopic equations of motion are formally given by

$$\vec{P}(t) = \mathbf{G}^{(t)}[\vec{P}(0)] \; . \tag{5.6}$$

Recalling Eq. (5.1) it is easy to see that the initial fitness distribution $\vec{P}(0)$ is given by:

$$P_n(0) = 2^{-K(n-1)} \left(1 - 2^{-K}\right) \; , \; 1 \leq n \leq N \; , \tag{5.7}$$

and

$$P_{N+1}(0) = 2^{-KN}. \tag{5.8}$$

As shown in Ref. [137] and chapter 4, despite $\mathbf{G}$'s nonlinearity, it can be easily linearized such that the tth iterate $\mathbf{G}^{(t)}$ can be directly obtained by solving for the eigenvalues and eigenvectors of $\mathbf{G}$.

For very large populations ($M > 2^L$) the dynamics of the fitness distribution obtained from GA simulation experiments is accurately predicted by iteration of the operator $\mathbf{G}$. (See Ref. [137] and chapter 4.) It is noteworthy, though, that this "infinite" population dynamics is qualitatively very different from the behavior shown in Fig. 5.3. For large populations strings of *all* fitnesses are present in the initial population and the average fitness increases smoothly and monotonically to an asymptote over a small number of generations. (This limit is tantamount to an exhaustive search, requiring as it does $\mathcal{O}(2^L)$ fitness function evaluations.) Despite this seeming lack of utility, in the next section we show how to use the infinite-population dynamics given by $\mathbf{G}$ to describe the finite-population behavior.

5.6.2 Finite-Population Dynamics

There are two important differences between the infinite-population dynamics and that with finite populations. The first is that with finite populations the components P_n cannot take on arbitrary values between 0 and 1. Since the number of individuals with fitness n in the population is necessarily integer, the values of P_n are quantized to multiples of $1/M$. The space of allowed finite-population fitness distributions thus turns into a regular lattice in $N+1$ dimensions with a lattice spacing of $1/M$ within the simplex specified by normalization (Eq. (5.2)). Second, the dynamics of the fitness distribution is no longer deterministic. In general, we can only determine the conditional probabilities $\Pr[\vec{Q}|\vec{P}]$ that a certain fitness distribution $\vec{P}$ leads to another $\vec{Q}$ in the next generation.

Fortunately, the probabilities $\Pr[\vec{Q}|\vec{P}]$ are simply given by a multinomial distribution with mean $\mathbf{G}[\vec{P}]$, which in turn is the result of the action of the infinite-population dynamics. This can be understood in the following way. In creating the population for the next generation individuals are selected, copied, and mutated, M times from the *same* population $\vec{P}$. This means that for each of the M selections there are equal probabilities q_n to produce an individual of fitness n. The probabilities q_n give the *expected* proportions of fitness n strings in the next generation. The actual proportions Q_n of individuals with fitness n in the next generation are thus given by a multinomial sample of size M with from the distribution of expected proportions q_n. Moreover, since for an infinite population the expected proportions q_n are equal to the actual proportions Q_n, the probabilities q_n are equal to the components of the infinite population distribution $\mathbf{G}[\vec{P}]$.

Putting these observations together, if we write $Q_n = m_n/M$, with $0 \leq m_n \leq M$ being integers that sum to M, we have:

$$\Pr[\vec{Q}|\vec{P}] = M! \prod_{n=1}^{N+1} \frac{\left(\mathbf{G}_n[\vec{P}]\right)^{m_n}}{m_n!} \; . \tag{5.9}$$

In mathematical population genetics, such multinomial sampling Markov models are known as Wright-Fisher models [69, pp. 66-70]. We see that for any finite-population fitness distribution $\vec{P}$ the operator $\mathbf{G}$ still gives the GA's *average* dynamics over one time step, since the *expected* fitness distribution at the next time step is $\mathbf{G}[\vec{P}]$. Note that the components $\mathbf{G}_n[\vec{P}]$ need not be multiples of $1/M$. Therefore the *actual* fitness

distribution $\vec{Q}$ at the next time step is not $\mathbf{G}[\vec{P}]$, but is instead one of the lattice points of the finite-population state space. Since the variance around the expected distribution $\mathbf{G}[\vec{P}]$ is proportional to $1/M$, $\vec{Q}$ is likely to be close to $\mathbf{G}[\vec{P}]$.

5.6.3 Epochal Dynamics

For finite populations, the expected change $\langle d\vec{P} \rangle$ in the fitness distribution over one generation is given by:

$$\langle d\vec{P} \rangle = \mathbf{G}[\vec{P}] - \vec{P}. \tag{5.10}$$

Assuming that some component $\langle dP_i \rangle$ is much smaller than $1/M$, the actual change in component P_i is likely to be $dP_i = 0$ for a long succession of generations. That is, if the size of the *flow* $\langle dP_i \rangle$ in some direction i is much smaller than the lattice spacing $(1/M)$ for the finite population, we expect the fitness distribution to not change in direction (fitness) i. In Ref. [137] and chapter 4 we showed this is the mechanism that causes epochal dynamics for finite populations. More formally, each epoch n of the dynamics corresponds to the population being restricted to a region in the n-dimensional lower-fitness subspace of fitnesses 1 to n of the macroscopic state space. Stasis occurs because the flow out of this subspace is much smaller than the finite-population induced lattice spacing.

As Fig. 5.3 illustrates, each epoch in the average fitness is associated with a constant value of the best fitness in the population. More detailed experiments reveal that not only is $\langle f \rangle$ constant on average during the epochs, in fact, up to fluctuations, the entire fitness distribution $\vec{P}$ is constant during the epochs. We denote by $\vec{P}^n$ the average fitness distribution during the generations when n is the highest fitness in the population. As was shown in chapter 4, each epoch fitness distribution $\vec{P}^n$ is the unique fixed point of the operator $\mathbf{G}$ restricted to the n-dimensional subspace of strings with $1 \leq f \leq n$. That is, if $\mathbf{G}^n$ is the projection of the operator $\mathbf{G}$ onto the n-dimensional subspace of fitnesses from 1 up to n, we have:

$$\mathbf{G}^n[\vec{P}^n] = \vec{P}^n \ . \tag{5.11}$$

The average fitness f_n in epoch n is then given by:

$$f_n = \sum_{j=1}^{n} j P_j^n. \tag{5.12}$$

Thus, the fitness distribution $\vec{P}^n$ during epoch n is obtained by finding the fixed points of $\mathbf{G}$ restricted to the first n dimensions of the fitness distribution space. We will construct the operators $\mathbf{G}^n$ explicitly below for our GA and solve analytically for the epoch fitness distributions $\vec{P}^n$ as a function of n, K, and q.

The global dynamics can be viewed as an incremental discovery of successively more dimensions of the fitness distribution space. In most realistic settings, it is typically the case that population sizes M are much smaller than 2^L. Initially, then, only strings of low fitness are present in the initial population, i.e. see Eq. (5.7). The population then stabilizes on the epoch fitness distribution $\vec{P}^n$ corresponding to the best fitness n in the

initial population. The fitness distribution fluctuates around $\vec{P}^n$ until a string of fitness $n+1$ is discovered and spreads through the population. The population then settles into fitness distribution $\vec{P}^{n+1}$ until a string of fitness $n+2$ is discovered, and so on, until the global optimum at fitness $N+1$ is found. In this way, the global dynamics can be seen as stochastically hopping between the different epoch distributions $\vec{P}^n$.

Whenever mutation creates a string of fitness $n+1$, this string may either disappear before it spreads (seen as the isolated jumps in best fitness in Fig. 5.3) or it may spread, leading the population to fitness distribution $\vec{P}^{n+1}$. We have been calling the latter process an *innovation*. Fig. 5.3 also showed that it is possible for the population to fall from epoch n (say) down to epoch $n-1$. This happens when, due to fluctuations, all individuals of fitness n are lost from the population. We refer to this as a *destabilization* of epoch n. For some parameter settings, such as shown in Figs. 5.3(a) and 5.3(c), this is very rare. The time for the GA to reach the global optimum is mainly determined by the time it takes to discover strings of fitness $n+1$ in each epoch n. For other parameter settings, however, such as in Figs. 5.3(b) and 5.3(d), the destabilizations play an important role in how the GA reaches the global optimum. In these regimes, destabilization must be taken into account in calculating search times. This task is accomplished in the next chapter.

5.6.4 Selection Operator

We now explicitly construct the generation operator $\mathbf{G}$ for the limit of infinite population size by constructing the selection operator $\mathbf{S}$ and mutation operator $\mathbf{M}$. First, we consider the effect of selection on the fitness distribution. Since we are using fitness-proportionate selection, the proportion P_i^s of strings with fitness i after selection is proportional to i and to the fraction P_i of strings with fitness i before selection; that is,

$$P_i^s = c\, i\, P_i \; . \tag{5.13}$$

The constant c can be determined by demanding that the distribution remains normalized:

$$1 = \sum_{i=1}^{n} P_i^s = c \sum_{i=1}^{N+1} i P_i \; . \tag{5.14}$$

Since the average fitness $\langle f \rangle$ of the population is given by:

$$\langle f \rangle = \sum_{i=1}^{N+1} i P_i \; , \tag{5.15}$$

we have

$$P_i^s = \frac{i P_i}{\langle f \rangle} \; . \tag{5.16}$$

We can thus define a (diagonal) operator $\mathbf{S}$ that works on a fitness distribution $\vec{P}$ and produces the fitness distribution $\vec{P}^s$ after selection by:

$$\left(\mathbf{S} \cdot \vec{P} \right)_i = \sum_{j=1}^{N+1} \frac{\delta_{ij} j}{\langle f \rangle} P_j \; . \tag{5.17}$$

Notice that this operator is nonlinear since, by Eq. (5.15), the average fitness $\langle f \rangle$ is a function of the distribution $\vec{P}$ on which the operator acts.

5.6.5 Mutation Operator

The components of the mutation operator $\mathbf{M}$ give the probabilities $\mathbf{M}_{ij}$ that a string of fitness j is turned into a string with fitness i under mutation. These components naturally fall into three categories. First, consider the components $\mathbf{M}_{ij}$ with $i < j$. These strings are obtained if mutation leaves the first $i-1$ blocks of the string unaltered and disrupts the ith block in the string. The effect of mutation on the blocks $i+1$ through N is immaterial for this transition. Multiplying the probabilities that the preceding $i-1$ blocks remain aligned and that the ith block becomes unaligned we have:

$$\mathbf{M}_{ij} = (1-q)^{(i-1)K} \left(1 - (1-q)^K\right), \quad i < j \; . \tag{5.18}$$

Second, the diagonal components $\mathbf{M}_{jj}$ are obtained when mutation leaves the first $j-1$ blocks unaltered and does *not* mutate the jth block to be aligned. The maximum entropy assumption says that the jth block is random and so the probability P_a that mutation will change the unaligned jth block to an aligned block is given by:

$$P_a = \frac{1-(1-q)^K}{2^K - 1} \; . \tag{5.19}$$

This is the probability that at least one mutation will occur in the block times the probability that the mutated block will be in the correct configuration. Thus the diagonal components are given by:

$$\mathbf{M}_{jj} = (1-q)^{(j-1)K}(1-P_a) \; . \tag{5.20}$$

Finally, we calculate the probabilities for increasing-fitness transitions $\mathbf{M}_{ij}$ with $i > j$. Depending on the approximation used, these transition probabilities might depend on the value n of the current best fitness in the population. As we show below in section 5.8, the reason for this is that all individuals in the population during epoch n are relatively recent descendants of strings in the highest fitness class n. Therefore, the blocks j through $n-1$ in strings with fitness $j < n$ during epoch n tend to reflect memory of the fact that they are relatively recent descendants of a string with fitness n. This will of course influence the probabilities $\mathbf{M}_{ij}$ with $i > j$ differently in each epoch n. It turns out, however, that for the theoretical predictions of our chosen observables—such as, epoch durations and total number of fitness function evaluations—the maximum entropy assumption gives results that are very similar to results obtained with epoch-dependent approximations.

Using the maximum entropy assumption, we obtain epoch-independent transition probabilities $\mathbf{M}_{ij}$ by assuming that all the blocks j through N are random. The jth block is equally likely to be in any of $2^K - 1$ unaligned configurations. All succeeding blocks are equally likely to be in any one of the 2^K configurations, including the aligned one. In order for a transition to occur from state j to i, first of all the $j-1$ aligned blocks have to remain aligned, then the jth block has to become aligned through the mutation.

The latter has probability P_a. Furthermore, the following $i-j-1$ blocks all have to be aligned. Finally, the ith block has to be unaligned. Putting these together, we find that:

$$\begin{aligned}\mathbf{M}_{ij} &= (1-q)^{(j-1)K} P_a \\ &\quad \left(\frac{1}{2^K}\right)^{i-j-1} \left(1-\frac{1}{2^K}\right) \ , \ i > j \ . \end{aligned} \tag{5.21}$$

As a technical aside, note that for the case $i = N+1$ the last factor does not appear.

The generation operator $\mathbf{G}$ is obtained by taking the product of the selection operator $\mathbf{S}$ with the mutation operator $\mathbf{M}$:

$$\mathbf{G} = \mathbf{M} \cdot \mathbf{S}. \tag{5.22}$$

The restricted generation operators $\mathbf{G}^n$ are obtained by setting all components with $i > n$ or $j > n$ to zero, where the component indices of the mutation and selection operators run from 1 to n.

5.7 Quasispecies Distributions and Epoch Fitness Levels

In this section, we solve for the metastable fitness distribution $\vec{P}^n$ during each epoch n. In section 5.9 we will use the $\vec{P}^n$ to calculate the waiting times for an innovation from epoch n to $n+1$ to occur.

The quasispecies fitness distribution $\vec{P}^n$ is given by a fixed point of the operator $\mathbf{G}^n$. To obtain this fixed point we linearize the generation operator by taking out the factor $\langle f \rangle$, thereby defining a new operator $\tilde{\mathbf{G}}^n$ via:

$$\mathbf{G}^n = \frac{1}{\langle f \rangle} \tilde{\mathbf{G}}^n. \tag{5.23}$$

The operator $\tilde{\mathbf{G}}^n$ is just an ordinary (linear) matrix operator and the quasispecies fitness distribution is nothing other than the principal eigenvector of this matrix. The principal eigenvalue f_n of $\tilde{\mathbf{G}}^n$ is the average fitness of the quasispecies distribution. In this way, obtaining the quasispecies distribution reduces to calculating the principal eigenvector of the matrix $\tilde{\mathbf{G}}^n$. Again the reader is referred to chapter 4.

As in Ref. [137] and chapter 4, the local stability of the epochs can be analyzed by calculating the eigenvalues and eigenvectors of the Jacobian matrix $\mathbf{DG}^n$ around each epoch fitness distribution $\vec{P}^n$. The components $\mathbf{DG}^n_{ij}$ of the Jacobian around epoch n are given by:

$$\mathbf{DG}^n_{ij} = \left[\frac{\partial \mathbf{G}_i(\vec{P})}{\partial P_j}\right]_{\vec{P}=\vec{P}^n} = \frac{\tilde{\mathbf{G}}^n_{ij} - j\vec{P}^n_j}{f_n} \ . \tag{5.24}$$

Just as in chapter 4, the eigenvectors $\vec{U}^i$ of the Jacobian are given by $\vec{U}^i = \vec{P}^i - \vec{P}^n$, with corresponding eigenvalues $e^n_i = f_i/f_n$. Thus, the spectra of the Jacobian matrices are simply determined by the spectrum of the generation operator $\tilde{\mathbf{G}}^n$. The eigenvalues e^n_i determine the bulk of the GA's behavior. Since the matrix $\tilde{\mathbf{G}}^n$ is generally of modest size, i.e. its dimension is determined by the number of blocks N, we can easily obtain

numerical solutions for the epoch fitnesses f_n and the epoch quasispecies distributions $\vec{P}^n$. At the same time one also obtains the eigenvalues e_i^n and eigenvectors $\vec{U}^i$ of the Jacobian.

For a clearer understanding of the functional dependence of the epoch fitness distributions on the GA's parameters, however, we will now develop analytical approximations to the epoch fitness levels f_n and quasispecies distributions $\vec{P}^n$.

In order to explicitly determine the form of the quasispecies distribution $\vec{P}^n$ during epoch n we will approximate the matrix $\tilde{\mathbf{G}}^n$. As we saw in section 5.6.5, the components $\mathbf{M}_{ij}$ (and so of $\tilde{\mathbf{G}}$) naturally fall into three categories. Those with $i < j$, those with $i > j$, and those on the diagonal $i = j$. Components with $i > j$ involve at least one block becoming aligned through mutation. For block lengths K that are not too small, these components are generally much smaller than those only involving the destruction of aligned blocks or for which there is no change in the blocks. We therefore approximate $\tilde{\mathbf{G}}^n$ by neglecting terms proportional to P_a. The distribution $\vec{P}^n$ is then simply the result of a balance between selection expanding the higher fitness strings and mutation disrupting blocks, whereby high fitness strings turn into lower fitness strings.

Under this approximation we have for the components of $\tilde{\mathbf{G}}^n$:

$$\tilde{\mathbf{G}}^n_{ij} = j(1-q)^{(i-1)K}(1-(1-q)^K), \quad i < j \; , \tag{5.25}$$

and

$$\tilde{\mathbf{G}}^n_{jj} = j(1-q)^{(j-1)K} \; . \tag{5.26}$$

The components with $i > j$ are now all zero. The result is that $\tilde{\mathbf{G}}^n$ is upper triangular. As is well known in general matrix theory, the eigenvalues of an upper triangular matrix are given by its diagonal components. Therefore, the average fitness f_n in epoch n, which is given by the largest eigenvalue, is equal to the largest diagonal component of $\tilde{\mathbf{G}}^n$. That is,

$$f_n = n(1-q)^{(n-1)K} \; . \tag{5.27}$$

Notice that the matrix elements only depend on q and K through the effective parameter λ defined by:

$$\lambda = (1-q)^K \; . \tag{5.28}$$

λ is just the probability that a block will not be mutated.

The principal eigenvector $\vec{P}^n$ is the positive and normalized solution of the equation:

$$\sum_{j=1}^{n} \left(\tilde{\mathbf{G}}^n_{ij} - f_n \delta_{ij} \right) P^n_j = 0 \; . \tag{5.29}$$

Since the components of $\tilde{\mathbf{G}}^n$ depend on λ in such a direct way, we can analytically solve for this eigenvector. We find that the quasispecies components are given by:

$$P^n_i = \frac{(1-\lambda)n\lambda^{n-1-i}}{n\lambda^{n-1-i} - i} \prod_{j=1}^{i-1} \frac{n\lambda^{n-j} - j}{n\lambda^{n-1-j} - j} \; . \tag{5.30}$$

For the component P_n^n this reduces to

$$P_n^n = \prod_{j=1}^{n-1} \frac{n\lambda^{n-j} - j}{n\lambda^{n-1-j} - j} \; . \tag{5.31}$$

The above equation can be re-expressed in terms of the epoch fitness levels f_j:

$$P_n^n = \lambda^{n-1} \prod_{j=1}^{n-1} \frac{f_n - f_j}{f_n - \lambda f_j} \; . \tag{5.32}$$

Note that it is straightforward to increase the accuracy of our analytical approximations by including terms proportional to P_a in the matrix $\tilde{\mathbf{G}}^n$ and then treating these terms as a perturbation to the upper triangular matrix. Using standard perturbation theory, one can then obtain corrections due to block alignments. We will not pursue this here, however, since the current approximation is accurate enough for the optimization analysis.

5.8 Quasispecies Genealogies and Crossover's Role

Before continuing on to solve this problem, we digress slightly at this point for two reasons. First, we claimed in a previous section that *all* individuals in the population during epoch n are descendants of strings with fitness n. We will demonstrate this now by considering the genealogies of strings occurring in a quasispecies population $\vec{P}^n$. Second, since the argument is quite general and only depends on the effects of selection, it has important implications for population structure in metastable states (such as fitness epochs) and, more specifically, for the role of crossover in epochal evolutionary search.

For the nth epoch, let the set of "suboptimal" strings be all those with fitness lower than n; their proportion is simply $1 - P_n^n$. This proportion is constant on average during epoch n. Over one generation, the suboptimal strings in the next generation will be either descendants of suboptimal strings in the current generation or mutant descendants of "optimal" strings with fitness n. Let r denote the average number of offspring per suboptimal individual. The fact that the total proportion of suboptimal strings remains constant gives us the equation:

$$1 - P_n^n = r(1 - P_n^n) + \frac{(1 - \mathbf{M}_{nn})n}{f_n} P_n^n \; . \tag{5.33}$$

The last term is the proportion of individuals of fitness n that are selected and mutate to suboptimal strings. From this we can solve for r and find:

$$r = 1 - \frac{(1 - \mathbf{M}_{nn})nP_n^n}{f_n(1 - P_n^n)} = \frac{1 - \lambda^{1-n}P_n^n}{1 - P_n^n} \; , \tag{5.34}$$

where we have used the previous analytical approximations to $n\mathbf{M}_{nn}$ and f_n, Eqs. (5.26) and (5.27), in the last equality.

We see that in every generation only a fraction $r < 1$ of the suboptimal individuals are descendants of suboptimal individuals in the previous generation. This means that after t generations, only a fraction r^t of the suboptimal individuals are the terminations of lineages that solely consist of suboptimal strings. There is a fraction of $1 - r^t$ strings that have one or more ancestors with fitness n in the t preceding generations. After a certain number of generations t_c the fraction r^{t_c} becomes so small that less than one individual (on average) has only suboptimal ancestors. That is, after approximately t_c generations in epoch n, the whole quasispecies will consist of strings that are descendants of a string with fitness n somewhere in the past. Explicitly, we find that

$$t_c = \frac{\log\left[M(1 - P_n^n)\right]}{\log\left[\frac{1-\lambda^{1-n}P_n^n}{1-P_n^n}\right]} \,. \tag{5.35}$$

As expected, t_c is proportional to the logarithm of the total number $M(1 - P_n^n)$ of suboptimal strings in the quasispecies.

The above result can be generalized to the case in which the suboptimal strings are defined to be all those with fitnesses 1 to $n - i$. One can then calculate the time until all strings in these classes are descendants of strings with fitness $n - i + 1$ to n to find that these lower classes are taken over even faster by descendants of strings with fitness n.

The preceding result is significant for a conceptual understanding of the structure of epoch populations. All suboptimal individuals in the population have an ancestor of optimal fitness that is a relatively small number of generations in the recent past. The result is that in genotype space all suboptimal individuals are always relatively close to some individual of optimal fitness. The suboptimal individuals never wander far from the individuals of optimal fitness. More precisely, the average duration of suboptimal lineages is $1/(1 - r)$ generations. That is, all suboptimal individuals are typically found within a Hamming distance of $Lq/(1 - r)$ from optimal-fitness individuals.

The individuals with optimal fitness, however, *can* wander through genotype space as long as they do not leave the neutral network of optimal fitness strings—those with fitness n in epoch n. If the population is to traverse large regions of genotype space in order to discover a string of fitness larger than n, it must do so along this neutral network. In short, this is the reason we believe the existence of neutral networks, consisting of approximately equal fitness strings that percolate through large parts of the genotype space, is so important for evolutionary search; cf. Ref. [79]. If strings of fitness n were to form a small island in a sea of lower fitness strings that are at relatively large Hamming distances from islands with fitness higher than n, then there would be little chance that a suboptimal fitness mutant will drift far enough from the island of fitness n strings to discover another island with higher fitness.

This result—that all strings in the metastable population are relatively recent descendants of strings in the highest fitness class—generalizes to other selection schemes such as elite selection, rank selection, and tournament selection. Furthermore, this result also provides some insight into the effects of adding crossover to evolutionary search algorithms on problems that show epochal evolution. Let's assume that we add crossover to our current GA; see also the discussion of similar crossover experiments in section 4.6.5. The initial population still has a distribution of fitnesses given by Eq. (5.7). The best fitness in the initial population might be (say) 3, corresponding to the first two

blocks being aligned and the third block unaligned. It is unlikely that crossovers will lead quickly to strings of fitness 4. Although the initial population will have strings with the 3rd block aligned, these strings are very unlikely to also have the first two blocks aligned. This means that these strings have low fitness and so are unlikely to be selected as the parent of a crossover event. Moreover, relatively quickly, the entire population will become descendants of strings with fitness 3 that, by definition, have the third block unaligned. Crossover events will thus almost never lead to the creation of strings of higher fitness; at least not by combining building blocks [74].

The positive contribution of crossover is that an aligned block may be formed from two parents each with an unaligned 3rd block if the crossover point falls within the 3rd block and if the resulting complementary subblocks are themselves aligned. The negative effect is that crossover may also combine lower fitness strings with higher fitness strings so as to produce two lower fitness offspring. Thus, with nothing else said or added, one concludes that the effect of crossover in our GA is marginal. Experiments with single-point crossover confirm this. The global optimum is found somewhat more quickly, but the improvement in search time is very small.

These arguments are specific to the Royal Staircase (and also Royal Road) classes of fitness function. However, for search dynamics exhibiting epochal evolution it is generally the case that the population structure is a cloud of strings *localized* around strings of the current best fitness during the epochs. Therefore, the effect of crossover will mostly be to increase the amount of mixing *within* the quasispecies cloud. That is, during the epochs crossover acts as a local mixing operator very much as mutation does. Note that this is not a result of our fitness function demanding that blocks be found consecutively.

In fact, by the same arguments as sketched above, crossover only marginally improves the search on Royal Road functions (chapter 4), which were specifically *designed* as fitness functions for which crossover should be most beneficial [105]. As was shown in Ref. [134], for simple genetic algorithms, the population size must be set unreasonably high in order for crossover to produce globally optimal strings before the population has converged. In order for recombination operators to be useful on these types of building block functions, more elaborate genetic algorithms—such as, messy GAs [63, 62]—must be constructed. However, as shown below, building-block-based fitness functions, such as the Royal Staircase and the Royal Road, can be efficiently searched with mutation-only genetic algorithms, provided that the mutation rate is set correctly.

5.9 Mutation Rate Optimization

In the previous sections we argued that the GA's behavior can be viewed as stochastically hopping from epoch to epoch until the search discovers a string with fitness $N + 1$. Assuming for the moment that the total time to reach this global optimum is dominated by the time the GA spends in the epochs, we will derive a way to tune the mutation rate q such that the time the GA spends in an epoch is minimized.

During epoch n no string in the population has the nth block aligned. Thus, the main contribution to the waiting time in epoch n is given by the time it takes for a mutant with the nth block set correctly to appear. As we have seen, the probability P_a to mutate a

single block such that it becomes aligned is given by:

$$P_a = \frac{1-(1-q)^K}{2^K-1} = \frac{1-\lambda}{2^K-1}, \tag{5.36}$$

where again λ is the probability that a block will not be mutated at all. Obviously, the higher P_a, the more likely mutation is to produce a new aligned block. Every generation each individual in the population has a probability P_a of aligning the nth block. Aligning the nth block only creates a string of fitness $n+1$ when all the $n-1$ blocks to its left are aligned as well. That is, only the fraction P_n^n of the population with fitness n will produce a fitness $n+1$ string by aligning the nth block. Therefore, the probability C_{n+1} that a string of fitness $n+1$ will be created over one generation is given by:

$$C_{n+1} \approx M P_n^n P_a \; . \tag{5.37}$$

Our claim is that, as a first approximation, maximizing C_{n+1} minimizes the number of generations the population spends in epoch n.

In section 5.7 we derived an analytic approximation, Eq. (5.31), to the proportion P_n^n of individuals in the highest fitness class during epoch n as a function of $\lambda = (1-q)^K$. Since P_a also only depends on q through λ and λ is monotonic in q, we can maximize C_{n+1} as a function of λ instead of q. Ignoring proportionality constants, the function to maximize is:

$$C_{n+1} \propto (1-\lambda) P_n^n(\lambda) \; . \tag{5.38}$$

Using Eq. (5.31) for the dependence of P_n^n on λ and differentiating the above function C_{n+1} with respect to λ, we find that the optimum λ_o satisfies:

$$\frac{n(1-\lambda_o)}{\lambda_o}\left[\sum_{i=1}^{n-1} \frac{i\lambda_o^i}{n\lambda_o^i - n + i} - \frac{(i-1)\lambda_o^{i-1}}{n\lambda_o^{i-1} - n + i}\right] = 1 \; . \tag{5.39}$$

The solution of this equation gives the optimum $\lambda_o(n)$ which is only a function of the epoch number n. This is an important observation, since it means that the optimal value of the mutation rate q_o takes the following general form as a function of n and K:

$$q_o = 1 - \sqrt[K]{\lambda_o(n)} \; . \tag{5.40}$$

Once we solve for the function $\lambda_o(n)$, we can immediately obtain the dependence of q_o on K using Eq. (5.40).

In this calculation we assumed that the waiting time for *discovering* a higher fitness string dominated the time spent in an epoch. This amounts to assuming that as soon as a string of fitness $n+1$ is created, copies of this string spread through the population and the population stabilizes onto epoch $n+1$. In fact, it is quite likely that the string with fitness $n+1$ will be lost through a deleterious mutation in the aligned blocks or via sampling, before it gets a chance to spread in the population. Recall the transitory jumps in the best fitness seen in Fig. 5.3. In chapter 4 we used a diffusion equation approximation to calculate the probability π_n that a string with fitness $n+1$ will spread. We found that to a good approximation it is given by:

$$\pi_n = \frac{1-\left(1-\frac{1}{M}\right)^{2M\gamma_n+1}}{1-\left(1-P_{n+1}^{n+1}\right)^{2M\gamma_n+1}} \approx 1 - e^{-2\gamma_n} \; , \tag{5.41}$$

where

$$\gamma_n = \frac{f_{n+1} - f_n}{f_n}, \tag{5.42}$$

and the last approximation in Eq. (5.41) holds for relatively large population sizes ($P_n^n M \gg 1$). Equation (5.41) is an adaptation of diffusion equation results first obtained by Kimura [89]. Notice that the spreading probability π_n only depends on the relative difference of the average fitness in epoch $n+1$ and epoch n. Using Eq. (5.27) for f_n we find:

$$\gamma_n = \left(1 + \frac{1}{n}\right)\lambda - 1 \; . \tag{5.43}$$

Thus, we find that $\pi_n(\lambda)$ is approximately given by:

$$\pi_n(\lambda) = 1 - e^{-2\left(1+\frac{1}{n}\right)\lambda+2} \; . \tag{5.44}$$

The probability C'_{n+1} to create a string of fitness $n+1$ that spreads through the population is given by:

$$C'_{n+1} = C_{n+1}\pi_n(\lambda) \; . \tag{5.45}$$

Thus, taking the spreading probability π_n into account, we want to maximize C'_{n+1} in order to minimize the time spent in epoch n. Note that, also in this case, the dependence on q and K enters only through λ. For each n there is an optimal value of λ from which the optimal mutation rate can be determined as a function of K.

The optimal value λ_o in this case is the solution of:

$$\begin{aligned} \frac{n(1-\lambda_o)}{\lambda_o} & \left[\sum_{i=1}^{n-1} \frac{i\lambda_o^i}{n\lambda_o^i - n + i} - \frac{(i-1)\lambda_o^{i-1}}{n\lambda_o^{i-1} - n + i}\right] \\ + \quad (1-\lambda_o) & \frac{2\left(1+\frac{1}{n}\right)e^{-2\left(1+\frac{1}{n}\right)\lambda_o+2}}{1 - e^{-2\left(1+\frac{1}{n}\right)\lambda_o+2}} = 1. \end{aligned} \tag{5.46}$$

Numerically, the solution $\lambda_o(n)$ is well approximated by:

$$\lambda_o(n) = 1 - \frac{1}{3n^{1.175}} \; , \tag{5.47}$$

as shown in Fig. 5.4, which plots $1-\lambda_o$ as a function of n. The solid line is the numerical solution obtained from Eq. (5.46); the dashed line is the approximation Eq. (5.47).

For large n, using Eqs. (5.40) and (5.47) we can approximate the optimal mutation rate by:

$$q_o = \frac{1}{3(nK)n^{0.175}} \; . \tag{5.48}$$

Thus, the optimal mutation rate drops as a power-law in both n and K. This implies that, generally for the Royal Staircase fitness functions, the mutation rate should decrease as a GA run progresses so that the search will find the global optimum as quickly as possible. We pursue this idea more precisely elsewhere by considering an adaptive mutation rate scheme for the GA. In the following, the mutation rate is assumed to be constant during the runs.

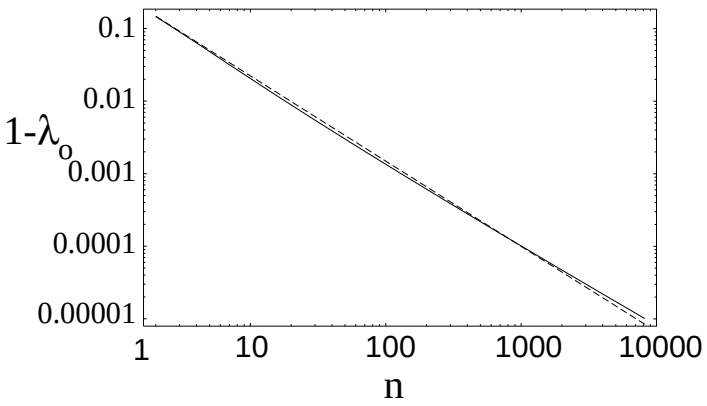

Figure 5.4: Optimal λ as a function of n. The vertical axis shows $1-\lambda_0$ on a logarithmic scale. The horizontal axis plots n on a logarithmic scale. The solid line is the numerical solution of Eq. (5.46) and the dashed line is the approximation given by Eq. (5.47).

5.10 Fitness Function Evaluations

In the preceding sections we derived an expression for the probability C'_{n+1} to create, over one generation in epoch n, a string of fitness $n+1$ that will stabilize by spreading through the population. From this we now estimate the total number E of fitness function evaluations the GA uses on average before an optimal string of fitness $N+1$ is found. As a first approximation, we assume that the GA visits all epochs, that the time spent in innovations between them is negligible, and that epochs are always *stable*. By epoch stability we mean that it is highly unlikely that strings with the current highest fitness will disappear from the population through a fluctuation, once such strings have spread. These assumptions appear to hold for the parameters of Figs. 5.3(a) and 5.3(c). They may hold even for the parameters of Fig. 5.3(b), but they probably do not for Fig. 5.3(d). For the parameters of Fig. 5.3(d), we see that the later epochs ($n = 8$, 9, and 10) easily destabilize a number of times before the global optimum is found. We develop an extension of the analysis that addresses this more complicated behavior in the next chapter.

The average number T_n of generations that the population spends in epoch n is simply $1/C'_{n+1}$—the inverse of the probability that a string of fitness $n+1$ will be discovered and spread through the population. For a population of size M, the number of fitness function evaluations per generation is M, so that the total number E_n of fitness function evaluations in epoch n is given by $T_n M$. More explicitly, we have from this and Eqs. (5.37) and (5.45) that:

$$E_n = \frac{2^K - 1}{(1-\lambda)P_n^n \pi_n} \,. \tag{5.49}$$

This says that, given our approximations, the total number of fitness function evaluations in each epoch is independent of the population size M. The epoch lengths, measured

in generations, are inversely proportional to M, while the number of fitness function evaluations per generation is M. Substituting our analytical expression for P_n^n, Eq. (5.31), we have:

$$E_n(\lambda) = \frac{2^K - 1}{(1-\lambda)\pi_n(\lambda)} \prod_{i=1}^{n-1} \frac{n\lambda^{n-i-1} - i}{n\lambda^{n-i} - i} \; . \tag{5.50}$$

The total number of fitness function evaluations $E(\lambda)$ to reach the global optimum is simply given by the sum of $E_n(\lambda)$ over all epochs n from 1 to N:

$$E(\lambda) = \sum_{n=1}^{N} \frac{2^K - 1}{(1-\lambda)\pi_n(\lambda)} \prod_{i=1}^{n-1} \frac{n\lambda^{n-i-1} - i}{n\lambda^{n-i} - i} \; . \tag{5.51}$$

The optimal mutation rate for an entire run is obtained by minimizing the above expression for E with respect to λ. In general, this minimum E_o can only be determined numerically. However, since the last epoch N is in general by far the longest epoch, a rough approximation to the minimum of E can be obtained by taking $\lambda = \lambda_o(N)$, i.e. setting the mutation rate optimally for the last epoch. Substituting Eq. (5.47) into Eq. (5.51) we find, numerically, that the minimum number of fitness function evaluations E_o as a function of N and K is approximately given by:

$$E_o \approx 2.2 \; N^{3.1} 2^K \; . \tag{5.52}$$

Somewhat surprisingly perhaps, the block length only enters in the factor 2^K. This factor 2^K is simply the number of block states and is also exactly the ratio between the volume of the neutral nets with fitness n and $n+1$. That is, the factor 2^K accounts for the amount of neutral net volume that has to be searched before a higher fitness string is discovered. Note that in E_o the factor that accounts for the genotypic structure scales roughly as the cube of the number of blocks, *independent* of the block length K. Although this search-time scaling for our mutation-only GA seems quite efficient, it must be compared to simple random-mutation hill climbing algorithms [105] for which it can be shown that the optimum is found on average in $\approx N^2 2^K$ fitness function evaluations.

Note that we set $\pi_N = 1$ in Eq. (5.51), since by definition the GA stops at the *first* occurrence of the globally optimum string. For the last epoch N it is only necessary to create a string of fitness $N+1$, it does not need to spread through the population.

5.11 Theory versus Experiment

We now turn to examining the dependency of E on the mutation rate q in the neighborhood of the optimum. Figure 5.5 compares Eq. (5.51) to the number (solid lines) of fitness function evaluations estimated from 250 GA runs for the four different settings of N and K of Fig. 5.3. Each plot shows E as a function of mutation rate q for a set of different population sizes M. In addition, each of the four plots in Fig. 5.5 shows, as a dashed line, the population-size independent theoretical approximation to the total number E of fitness function evaluations as a function of q.

Figure 5.5(a) has $N = 8$ blocks of $K = 8$ bits. The dependence of E on mutation rate is shown over the range from $q = 0.001$ to $q = 0.012$. The optimal mutation

rate occurs somewhere around $q_o = 0.006$. Each solid line shows the experimental dependence of E on q for a fixed population size. At high mutation rates, the smaller population sizes have higher E. At high mutation rates, the set of solid lines in Fig. 5.5 show, from left to right, $E(q)$ for population sizes $M = 150$, $M = 200$, $M = 250$, $M = 300$, and $M = 350$, respectively.

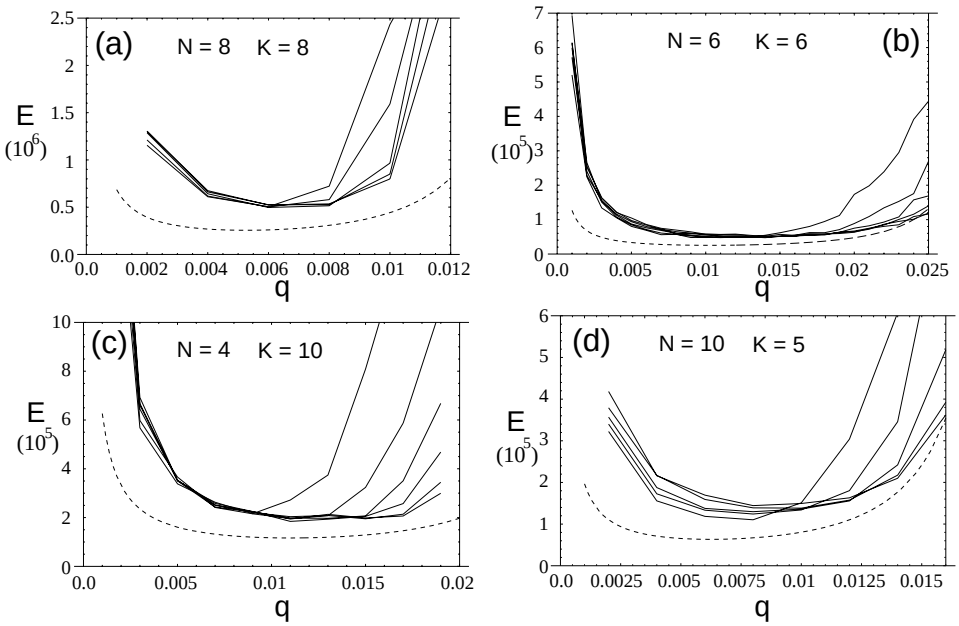

Figure 5.5: Comparison of experimental results (solid lines) and theoretical predictions (dashed lines) of the total number E of function evaluations to reach the global optimal as a function of the mutation rate q, for four different fitness functions as determined by N and K. The fitness function parameters are the same as those used in Fig. 5.3. The dashed lines give the theoretical predictions according to Eq. (5.51). The solid lines are the results of experiments with different population sizes M and show the experimentally estimated E as a function of q for a constant population size M. Each data point on the solid lines gives E averaged over 250 GA runs. For low mutation rates, the solid lines approximately overlap, indicating that E is approximately independent of M in this regime. For large mutation rates, small population sizes have larger E than large populations. Plot (a) shows, from left to right, population sizes $M = 150$, $M = 200$, $M = 250$, $M = 300$, and $M = 350$; plot (b) $M = 60$, $M = 90$, $M = 120$, $M = 150$, $M = 180$, and $M = 210$; plot (c) $M = 50$, $M = 80$, $M = 110$, $M = 140$, $M = 170$, and $M = 200$; and finally, plot (d) $M = 100$, $M = 150$, $M = 200$, $M = 250$, and $M = 300$.

Figure 5.5(b) has parameters $N = 6$ and $K = 6$. $E(q)$ is shown over the range $q = 0.001$ to $q = 0.025$. Again, at large mutation rates the smaller population sizes have higher $E(q)$. The solid lines from top to bottom in the high mutation rate regime

show $E(q)$ for population sizes $M = 60$, $M = 90$, $M = 120$, $M = 150$, $M = 180$, and $M = 210$, respectively. The optimal mutation rate occurs near $q_o = 0.012$.

Figure 5.5(c) has $N = 4$ blocks of length $K = 10$ as parameter settings. $E(q)$ is shown over the range $q = 0.001$ to $q = 0.02$, with the optimum occurring around $q_o = 0.011$. The solid lines show the experimental $E(q)$ for population sizes $M = 50$, $M = 80$, $M = 110$, $M = 140$, and $M = 200$.

Finally, Fig. 5.5(d) has $N = 10$ blocks of length $K = 5$ bits. The horizontal axis ranges from $q = 0.001$ to $q = 0.016$ with the optimal mutation rate occurring around $q_o = 0.008$. The population sizes are, from top to bottom at high q, $M = 100$, $M = 150$, $M = 200$, $M = 250$, and $M = 300$, respectively.

Note that for figures 5.5(b), 5.5(c), and 5.5(d) the tick marks on the vertical axis have a scale of 10^5 fitness function evaluations, while Fig. 5.5(a) uses a scale of 10^6 fitness function evaluations. Figure 5.5 shows that for all these different parameter settings, the theory, which is independent of population size M, reasonably predicts the location of the optimal mutation rate q_o. It also predicts moderately well the shape of the curve around the optimum.

It is notable that the theory consistently underestimates $E(q)$. Furthermore, it underestimates $E(q)$ more in the low mutation rate regime than in the high. We believe that both of these offsets can be explained by the effects of finite-population sampling. We assumed that all unaligned blocks in members of the current highest fitness class are random and *independent* of each other. This last assumption does not hold in general [31]. Due to finite-population sampling and the resulting tendency of the population to converge, strings in the highest fitness class are correlated with each other. This means that if one individual in the highest fitness class has it's nth unaligned block at a large Hamming distance from the desired configuration (with that block aligned), then many individuals in the highest fitness class also tend to have their nth block at large Hamming distances from the desired configuration. This genetic correlation among the individuals increases the epoch durations. Moreover, this effect is more severe for small mutation rates which, along with small population sizes, increase population convergence.

It turns out that this effect is difficult to analyze quantitatively. In spite of this it appears, as Fig. 5.5 shows, that for low mutation rates on up to the optimal mutation rate, the total number E of fitness function evaluations is indeed approximately independent of population size M. Moreover, the theory still accurately predicts the location of the optimal mutation rate q_o. Also, although the exact value of E is underestimated, the largest deviation in E from the experimental minimum is 47% for the parameters of Fig. 5.5(a), whereas the minimal deviation is 37% for the parameters of Fig. 5.5(c). Thus, the theory predicts the correct order of magnitude for the minimal number of fitness function evaluations.

As was noted above, the experimental curves for different population sizes appear to collapse onto each other in the low mutation rate regime. As the mutation rate increases, the solid curves break off this common curve one by one and do so delayed in proportion to population size. As the mutation rate increases it appears that progressively larger population sizes are necessary to maintain the search's efficiency.

Of course, this effect is not explained by our population-size independent theory. Intuitively, however, it can be understood as follows. As the mutation rate increases beyond the optimum, the mean proportion P_N^N of individuals with fitness N (during the

last epoch) decreases quite rapidly. When P_N^N is on the order of $1/M$, on average there are very few members with fitness N in the population. These fitness N strings can be easily lost due to a sampling fluctuation. When this happens, the epoch destabilizes, as can be seen occurring in Fig. 5.3(d). E starts rising rapidly for these small population sizes, since epoch N is likely to destabilize several times before an optimal string is found. For larger population sizes, however, the epochs are still stable.

We analyze this destabilization dynamics in detail in the next chapter. Roughly speaking, the population size can be optimized by determining P_N^N for the optimal mutation rate and then by lowering the population size as far as possible without destabilizing epoch N. Note also that there is a critical mutation rate q_c beyond which epoch N is absolutely unstable, even for *infinite* population size. This critical mutation rate is analogous to the error threshold in the theory of molecular evolution [33, 133]. For mutation rates beyond this error threshold, the optimum genome essentially will never be found. The error threshold q_c is also calculated explicitly in the following chapter. The resulting analytical theory ends up predicting the entire *search efficiency surface* $E(q, M)$.

5.12 Conclusion and Future Analyses

In summary, our findings here are the following.

1. At fixed population size M, there is a smooth cost function $E(q)$ as a function of mutation rate q. It has a *single* and *shallow* minimum q_o.

2. The optimal parameter settings roughly occur in the regime where the highest epochs $N-1$, N, and $N+1$ are marginally stable; see Fig. 5.3.

3. At the optimal parameter settings the average number of fitness function evaluations is $\mathcal{O}(N^{3.1}2^K)$. This is less efficient than a simple mutational hill climber, $\mathcal{O}(N^2 2^K)$.

4. For lower mutation rates than q_o the total number of fitness function evaluations $E(q)$ grows steadily and becomes almost independent of the population size M.

5. Crossover's role in reducing search time is marginal due to population convergence during the epochs.

6. There is a mutational error threshold in q that bounds the upper limit of the GA's efficient search regime. Above the threshold, which is population-size independent, suboptimal strings of fitness N cannot stabilize in the population. For small populations, this error threshold occurs at lower values of q.

7. The function $E(q)$ appears to be concave. That is, for any two parameters q_1 and q_2, the straight line connecting these two points is everywhere above the curve $E(q)$. We conjecture that this is always the case for mutation-only genetic algorithms with a unique global optimum. This feature is useful in that a steepest-descent algorithm applied to parameter q will lead to the unique optimum q_o.

Synopsizing the results this way, we anticipate some of the results developed for the population-size dependent theory of chapter 6.

Having come to this point one must ask, What relevance do these results have for a practitioner applying evolutionary search to complicated optimization problems? First of all, our analysis applies to problems showing epochal evolution. When epochal evolution occurs for a realistic search problem it may either be a result of the population getting stuck on local fitness optima or it may be a result of large neutral subspaces through which the population diffuses. From the practitioner's standpoint, it is important to distinguish these two cases, since their search dynamics are very different. A population stuck on a local optimum can be identified by the fact that the population forms a cloud around a *single genotype* in genotype space. Therefore, the genotypic content of the population is constant on average. In contrast, for the second case with diffusion over neutral networks, only the *average fitness* is constant over time while the population cloud of genotypes *drifts* in an erratic way through genotype space. It should therefore be relatively straightforward for the practitioner to determine what *kind* of epochal evolution is occurring by monitoring these features.

On the one hand, if the population is stuck at a local optimum, it is most often beneficial to decrease the selection pressure or to increase the mutation rate until the population begins to "move" again through genotype space. Basically, in this situation the population has already discovered the highest fitness genotype in the region of genotype space around the local optimum. No higher fitness strings are likely to be discovered unless the population starts to move through genotype space.

On the other hand, if the population is diffusing over a neutral net of roughly equal fitness strings, it is straightforward to determine an optimal mutation rate experimentally. Our theory indicates that the optimal mutation rate occurs when the product $qP_n^n(q)$ of the proportion of individuals P_n^n on the highest fitness neutral net and the mutation rate q is maximal. In short, the optimal mutation rate occurs when the total number of mutations in individuals of the current highest fitness class is maximal. After the optimal mutation rate is determined (holding population size constant), one can lower the population size until the product $P_n^n M$ is so small that further decreases in M compromise the stability of current highest fitness individuals. Since the proportion P_n^n can be measured experimentally, the procedure just outlined allows for the maximization of the product qP_n^n for complicated search problems.

If the product qP_n^n happens to be too hard to obtain experimentally, a rough estimate of the optimal parameter settings can still be obtained by noting that the optimal parameter settings typically occur when the current highest fitness members in the population are only marginally stable. That is, the mutation rate is as high as can be during each epoch without destroying all individuals of the highest fitness class. (This behavioral regime is analyzed in more detail in the next chapter.)

Finally, we note that the analysis presented here can be adapted in a straightforward manner to other selection schemes—such as, elite and tournament selection. Results for these will be reported elsewhere.

Acknowledgements

The authors thank Melanie Mitchell for helpful discussions. This work was supported at the Santa Fe Institute by NSF Grant IRI-9705830, ONR grant N00014-95-1-0524 and by Sandia National Laboratory contract AU-4978.

6

Optimizing Epochal Evolutionary Search: Population-Size Dependent Theory

Erik van Nimwegen and James P. Crutchfield
Machine Learning Journal,
(1999, accepted for publication)
Santa Fe Institute Working Paper 98-10-090.

Epochal dynamics, in which long periods of stasis in an evolving population are punctuated by a sudden burst of change, is a common behavior in both natural and artificial evolutionary processes. We analyze the population dynamics for a class of fitness functions that exhibit epochal behavior using a mathematical framework developed recently, which incorporates techniques from the fields of mathematical population genetics, molecular evolution theory, and statistical mechanics. Our analysis predicts the total number of fitness function evaluations to reach the global optimum as a function of mutation rate, population size, and the parameters specifying the fitness function. This allows us to determine the optimal evolutionary parameter settings for this class of fitness functions.

We identify a generalized error threshold that smoothly bounds the two-dimensional regime of mutation rates and population sizes for which epochal evolutionary search operates most efficiently. Specifically, we analyze the dynamics of epoch *destabilization* under finite-population sampling fluctuations and show how the evolutionary parameters effectively introduce a coarse graining of the fitness function. More generally, we find that the optimal parameter settings for epochal evolutionary search correspond to behavioral regimes in which the consecutive epochs are marginally stable against the sampling fluctuations. Our results suggest that in order to achieve optimal search, one should set evolutionary parameters such that the coarse graining of the fitness function induced by the sampling fluctuations is just large enough to hide local optima by rendering them unstable.

6.1 Designing Evolutionary Search

Evolutionary search algorithms are a class of stochastic optimization procedures inspired by biological evolution; e.g., see Refs. [4, 60, 94, 103]: A population of candidate solutions evolves under selection and random "genetic" diversification operators. Evolutionary search algorithms have been successfully applied to a diverse variety of optimization

problems; see, for example Refs. [11, 18, 28, 38, 49] and references therein. Unfortunately, and in spite of a fair amount of theoretical investigation, the mechanisms constraining and driving the dynamics of evolutionary search on a given problem are often not well understood.

There are very natural difficulties that are responsible for this situation. In mathematical terms evolutionary search algorithms are population-based discrete stochastic nonlinear dynamical systems. In general, the constituents of the search problem, such as the structure of the fitness function, selection, finite-population fluctuations, and genetic operators, interact in complicated ways to produce a rich variety of dynamical behaviors that cannot be easily understood in terms of the constituents individually. These complications make a strictly empirical approach to the question of whether and how to use evolutionary search problematic.

The wide range of behaviors exhibited by nonlinear population-based dynamical systems have been appreciated for decades in the field of mathematical population genetics. Unfortunately, this appreciation has not led to a quantitative predictive theory that is applicable to the problems of evolutionary search; something desired, if not required, for the engineering use of this stochastic search method.

We believe that a general, predictive theory of the dynamics of evolutionary search can be built incrementally, starting with a quantitative analytical understanding of specific problems and then generalizing to more complex situations. In this vein, the work presented here continues an attempt to unify and extend theoretical work in the areas of evolutionary search theory, molecular evolution theory, and mathematical population genetics. Our strategy is to focus on a class of problems that, despite their simplicity, exhibit some of the rich behaviors encountered in the dynamics of evolutionary search algorithms. Using analytical tools from statistical mechanics, dynamical systems theory, and the above mentioned fields we developed a detailed and quantitative understanding of the search dynamics for a class of problems that exhibit epochal evolution. On the one hand, as we show here, this allows us to analytically predict optimal parameter settings for this class of problems. On the other hand, the detailed understanding of the behavior for this class of problems provides valuable insights into the emergent mechanisms that control the dynamics in more general settings of evolutionary search and in other population-based dynamical systems.

In a previous paper Refs [137] and chapters 4 and 5, we analyzed in some detail the metastable population dynamics of what we call *epochal evolution*. In epochal evolution, long periods of stasis in the average fitness of the population are punctuated by rapid innovations to higher fitness. This *punctuated equilibrium* behavior is a common occurrence in both natural and artificial evolutionary processes, see for instance Refs. [1, 22, 37, 47, 64, 104].

For populations evolving under a static fitness function, which is typically the case in evolutionary search, it has been commonly assumed that *local optima* in the fitness "landscape" are responsible for metastability in the population dynamics. The geographic metaphor of a population crawling up the slopes of a fitness or adaptive landscape was originally introduced by the evolutionary biologist Sewall Wright [150]. More recently, it has been assumed that the typical fitness functions of combinatorial optimization and biological evolution can be modeled as "rugged landscapes" [88, 100]. Rugged landscapes are fitness functions with wildly fluctuating fitnesses even at the smallest

scales of single-point mutations. It is natural to assume that such rugged landscapes possess a large number of local optima. With this picture in mind, one would explain punctuated equilibria as the result of the population getting "pinned" at a local optimum in the landscape, until a rare mutant crosses a valley of low fitness to a different, higher local optimum.

In contrast, there has been an increasing realization in recent years that the large degeneracies that occur in biological fitness functions play an important role in evolutionary dynamics [79]. Such degeneracies have also been observed in evolutionary search problems [22] and typically occur when there is redundancy in the genetic representation (*genotype*) of candidate solutions to a combinatorial optimization problem. When these degeneracies are operating, the set of all genotypes breaks into a relatively small number of distinct fitness classes of genotypes with approximately equal fitness. Moreover, due to the high dimensionality of genotype spaces, sets of genotypes with approximately equal fitness tend to form simply connected components—the members of which can be reached via paths made of single-mutation steps. Such components are generally referred to as *neutral networks* in molecular evolution theory, see Refs. [47, 79, 77, 50, 142]. Epochal behavior occurs in evolution under these fitness functions because members of the evolving population must search through most of the network of neutral variants before a connection to a neighboring network of higher fitness is discovered; during this time the average fitness is constant, up to fluctuations.

In our analysis of epochal evolution (chapters 4 and 5), we view large degeneracies in the genotype-to-fitness mapping as the main source of the epochal nature of the evolutionary dynamics. We have constructed a wide class of fitness functions that realize our view of genotype space decomposing into a relatively small collection of entangled neutral networks and analyzed the resulting evolutionary population dynamics. In chapter 5, we showed how this detailed dynamical understanding can be turned to practical advantage by analytically determining the mutation rates to reach, in the fewest number of fitness function evaluations, the global optimum in this class of fitness functions. Here we recount our basic analytical approach and extend it to incorporate population size dependent dynamical effects. As will be explained below, population-size effects enter primarily through the dependence of the stability of an epoch's metastable population on finite-population sampling fluctuations. The result is a more general and accurate theory that analytically predicts the total number of fitness function evaluations needed on average for the algorithm to discover the global optimum of the fitness function as a function of both mutation rate and population size.

In addition, we develop a detailed understanding of the operating regime in parameter space for which the search is performed most efficiently. We believe this will provide useful guidance on how to set search algorithm parameters for more complex problems. In particular, our theory explains how optimal search occurs in the parameter regime where metastable populations are only marginally stable. The results raise the general question of whether it is desirable for optimal search to run in dynamical regimes that are a balance of stability and instability. More specifically, we show how the interplay of mutation, selection, and finite population sampling fluctuations effectively induce a coarse graining of the fitness function. Based on this, we conjecture that optimal search occurs when the level of this coarse graining is just enough to hide local optima by rendering them unstable.

6.2 Royal Staircase Fitness Functions

Choosing a class of fitness functions, whose population dynamics one wishes to analyze, is a delicate compromise between generality, mathematical tractability, and the degree to which the class is representative of problems often encountered in evolutionary search. A detailed knowledge of the fitness function is very *atypical* of evolutionary search problems. If one knew the fitness function in detail, one would not have to run an evolutionary search algorithm to find high-fitness solutions in the first place. The other extreme of assuming complete generality, however, cannot lead to enlightening results either, since averaged over all problems, all optimization algorithms perform equally well (or badly); see Ref. [146]. We thus focus on a specific subset of fitness functions, somewhere between these extremes, that we believe at least have ingredients typically encountered in evolutionary search problems and that exhibit dynamical behaviors widely observed in both natural and artificial evolutionary processes.

As explained in the previous section, we focus on fitness functions that induce a collection of entangled neutral networks: genotype space decomposes into a set of (large) networks of isofitness genotypes that are connected via point mutation steps. Consequently, the number of different fitness values that genotypes can take is much smaller than the number of different genotypes. We also assume that higher-fitness networks are smaller in volume than low-fitness networks. Finally, we assume that from any neutral network there exist connections to higher-fitness networks such that, taken as a whole, the fitness landscape has no local optima other than the global optimum.

Under these assumptions, genotype space takes on a particular type of architecture: *subbasins* of the neutral networks are connected by *portals* leading between them and so to higher or lower fitness. Stated in the simplest terms possible, the evolutionary population dynamics then becomes a type of diffusion constrained by this architecture. For example, individuals in a population diffuse over neutral networks until a portal to a network of higher fitness is discovered and the population moves onto this network.

Viewed from a somewhat different perspective, such neutral network architectures in genotype space may be induced from any fitness function by coarse graining the fitness values into a small number of fitness classes. Under such coarse graining genotypes whose fitnesses fall into the same fitness class are treated as mutually neutral under selection. For instance, neutral networks in NK fitness landscapes have been constructed in this way [6, 108]. As we show below, explicit constructions may not be necessary: the evolutionary parameters themselves effectively induce a coarse graining of the fitness function. To some extent this justifies our grouping fitness values into a relatively small number of fitness classes. Moreover, we will argue that an optimal setting of evolutionary parameters for efficient search is achieved when, in effect, these parameters induce the "right" coarse graining of the fitness function.

In order to model the evolutionary behavior associated with neutral network architectures, we defined the class of *Royal Staircase* fitness functions that capture the essential elements sketched above (see chapter 5). Importantly, this class of fitness functions is simple enough to admit a fairly detailed quantitative mathematical analysis of the associated epochal evolutionary dynamics.

The Royal Staircase fitness functions are defined as follows.

1. Genotypes are specified by binary strings $s = s_1 s_2 \cdots s_L, s_i \in \{0, 1\}$, of length $L = NK$.

2. Reading the genotype from left to right, the number $I(s)$ of consecutive 1s is counted.

3. The fitness $f(s)$ of genotype s with $O(s)$ consecutive ones, followed by a zero, is $f(s) = 1 + \lfloor O(s)/K \rfloor$. The fitness is thus an integer between 1 and $N + 1$.

Four observations are in order with regard to this definition.

1. The fitness function has two parameters, the number N of blocks and the number K of bits per block. Fixing them determines a particular optimization problem.

2. There is a single global optimum: the genotype $s = 1^L$—namely, the string of all 1s—with fitness $f(s) = N + 1$.

3. The proportion ρ_n of genotype space filled by strings of fitness n is given by:

$$\rho_n = 2^{-K(n-1)} \left(1 - 2^{-K}\right), \tag{6.1}$$

   for $n \leq N$. Thus, high-fitness strings are exponentially more rare than low-fitness strings.

4. For each block of K bits, the all-1s pattern is the one that confers increased fitness on a string. Without loss of generality, any of the other $2^K - 1$ configurations could have been chosen as the "correct" configuration, including different patterns for each of the N blocks. Furthermore, since the GA here does not use crossover, arbitrary permutations of the L bits in the fitness function definition leave the evolutionary dynamics unchanged.

By implementing the architecture of neutral networks in this way, high-fitness neutral networks are nested inside lower fitness networks. Higher fitness strings are rarer since they require more bits in the genotype to be set "correctly". Each step upward in fitness is associated with setting an additional K bits in the genotype correctly. One can only set correct bit values in sets of K bits at a time, creating an *aligned* block, and in blocks from left to right. A genotype's fitness is proportional to the number of such aligned blocks. Since the $(n+1)$st block only confers fitness when all n previous blocks are aligned as well, there is contingency between blocks.

Using the same analysis as presented below one can analyze more complex cases in which different blocks have different numbers of bits and networks are entangled in more complicated ways than the simple nesting chosen here. However, the main conclusions of our analysis can be more transparently presented using this relatively simple Royal Staircase class. The reader is referred to Ref. [24] for an outline of the application of our analysis to a broad class of more complex fitness functions.

6.3 The Genetic Algorithm

For our analysis of evolutionary search we have chosen a simplified form of a genetic algorithm (GA) that does not include crossover and that uses fitness-proportionate selection. The GA is defined by the following steps.

1. Generate a population of M bit strings of length $L = NK$ with uniform probability over the space of L-bit strings.
2. Evaluate the fitness of all strings in the population.
3. Stop, noting the generation number $t_{\rm opt}$, if a string with optimal fitness $N + 1$ occurs in the population. Else, proceed.
4. Create a new population of M strings by selecting, with replacement and in proportion to fitness, strings from the current population.
5. Mutate, i.e. change, each bit in each string of the new population with probability q.
6. Go to step 2.

When the algorithm terminates there have been $E = Mt_{\rm opt}$ fitness function evaluations.

Notice that this algorithm omits the often used crossover operator. The main reason for excluding crossover is that it greatly simplifies the analysis. However, with respect to epochal evolution, the addition of crossover does not significantly alter or improve the evolutionary search behavior. We will provide some arguments for this claim below. For a more detailed discussion of crossover's lack of effectiveness in improving *epochal* evolutionary search, the reader is referred to chapter 5.

Our GA effectively has two parameters: the mutation rate q and the population size M. A given optimization problem is specified by the fitness function in terms of N and K. Stated most prosaically, then, the central goal of the following analysis is to find those settings of M and q that minimize the average number $\langle E \rangle$ of fitness function queries for given N and K required to discover the global optimum genotype of fitness $N + 1$. Our approach is to develop analytical expressions for E as a function of N, K, M, and q and then to study the *search-effort surface* $E(q, M)$ at fixed N and K. Before beginning the analysis, however, it is helpful to develop an appreciation of the basic dynamical phenomenology of evolutionary search on this class of fitness functions. Then we will be in a position to lay out the evolutionary equations of motion and analyze them.

6.4 Observed Population Dynamics

The typical behavior of a population evolving under a fitness function that induces connected neutral networks, such as defined above, alternates between long periods (*epochs*) of stasis in the population's average fitness and sudden increases (*innovations*) in the average fitness. We now briefly recount the experimentally observed behavior of typical Royal Staircase GA runs in which the parameters q and M are set close to their optimal

setting. The reader is referred to chapter 4 for a detailed discussion of the other dynamical behaviors this type of GA exhibits over a range of different parameter regimes.

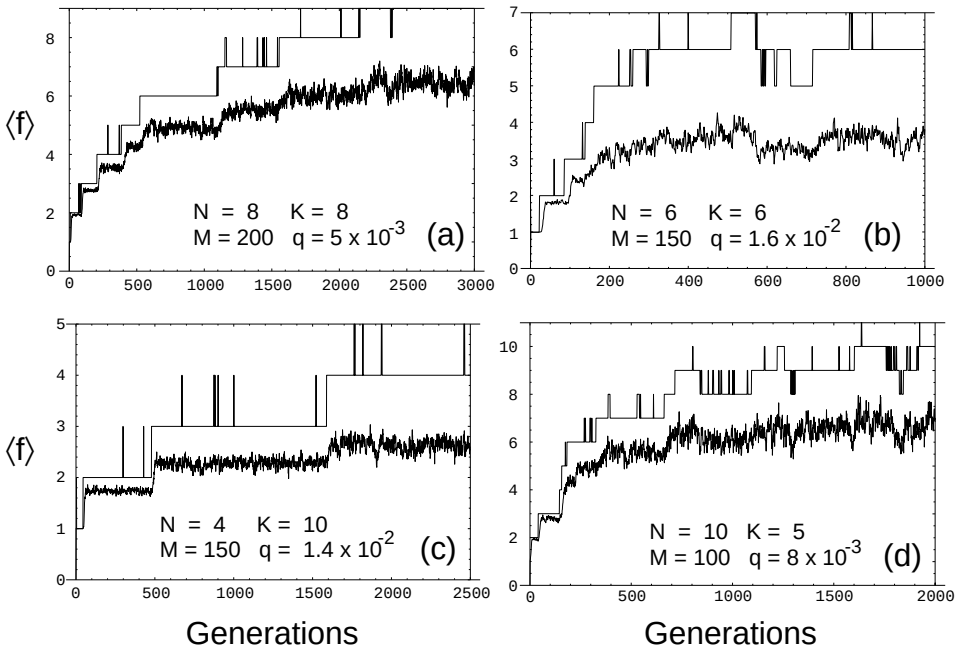

Figure 6.1: Examples of the Royal Staircase GA population dynamics with different parameter settings. The four plots show best fitness in the population (upper lines) and average fitness in the population (lower lines) as a function of time, measured in generations. The fitness function and GA parameters are given in each plot. In each case we have chosen q and M in the neighborhood of their optimal settings (see later) for each of the four values of N and K.

Figure 6.1 illustrates the GA's behavior at four different parameter settings. Each individual figure plots the best fitness in the population (upper lines) and the average fitness $\langle f \rangle$ in the population (lower lines) as a function of the number of generations. Each plot is produced from a single GA run. In all of these runs the average fitness $\langle f \rangle$ in the population goes through stepwise changes early in the run, alternating epochs of stasis with sudden innovations in fitness. Later in each run, especially for those in Figs. 6.1(b) and 6.1(d), $\langle f \rangle$ tends to have higher fluctuations and the epochal nature of the dynamics becomes unclear.

In the GA runs the population starts out with strings that only have relatively low fitness, say fitness n. (In all four plots of Fig. 6.1 we have $n = 1$.) Selection and mutation then establish an equilibrium in the population until a string aligns the nth block and descendants of this string with fitness $n+1$ spread through the population. A new equilibrium is then established until a string of fitness $n+2$ is discovered and so on, until finally, a string of fitness $N+1$ is discovered. In the figure we let the runs continue past the generation at which the global optimum was first discovered. Except for the run

in Fig. 6.1(a), the strings of fitness $N + 1$ do not manage to stabilize themselves in the population.

Notice that the behavior of the average fitness $\langle f \rangle$ roughly tracks the epochal behavior of the best fitness in the population. Every time a newly discovered higher-fitness string has spread through the population, $\langle f \rangle$ reaches a new, higher equilibrium value around which it fluctuates. As a run progresses to higher epochs, $\langle f \rangle$ tends to have higher fluctuations and the epochal nature of the dynamics is obscured. This is a result of the fact that for the highest epochs the difference between $\langle f \rangle$ in consecutive epochs is smaller than the average fitness fluctuations induced by the finite-population sampling; see chapter 4 for an analytical treatment of this particular phenomenon.

Notice, too, that often the best fitness shows a series of brief jumps to higher fitness during an epoch. When this occurs strings of higher fitness are discovered but, rather than spreading through the population, are lost within a few generations.

For each of the four settings of N and K we have chosen values of q and M such that the average total number $\langle E \rangle$ of fitness function evaluations to reach the global optimum for the first time is minimal. Thus, the four plots illustrate the GA's typical dynamics close to optimal (q, M)-parameter settings.

Despite what appears at first blush to be relatively small variations in fitness function and GA parameters, there is a large range, almost a factor of 10, in times to reach the global optimum across the runs. One concludes that there can be a strong parameter dependence in search times. It also turns out that the standard deviation σ of the mean total number $\langle E \rangle$ of fitness function evaluations is of the same order as $\langle E \rangle$. (See Table 6.1.) Thus, there are large run-to-run variations in the time to reach the global optimum. This is true for all parameter settings with which we experimented, of which only a few are reported here.

Having addressed the commonalities between runs, we now turn to additional features that each illustrates. Figure 6.1(a) shows the results of a GA run with $N = 8$ blocks of $K = 8$ bits each, a mutation rate of $q = 0.005$, and a population size of $M = 200$. During the later epochs, the best fitness in the population hops up and down several times before it finally jumps up and the new more-fit strings stabilize in the population. In this particular run, it took the GA approximately 3.4×10^5 fitness function evaluations (1700 generations) to discover the global optimum for the first time. Over 500 runs, the GA takes on average 5.3×10^5 fitness function evaluations to reach the global optimum for these parameters. The inherent large per-run variation means in this case that some runs take less than 10^5 function evaluations and that others take many more than 10^6. As our analysis will make clear, these large run-to-run variations are endogenous to the GA dynamics and cannot be reduced by changes in the parameters q and M. In fact, the run-to-run variations are minimal where the mean $\langle E \rangle$ itself is minimal.

Figure 6.1(b) plots a run with $N = 6$ blocks of length $K = 6$ bits, a mutation rate of $q = 0.016$, and a population size of $M = 150$. The GA discovered the global optimum after approximately 4.8×10^4 fitness function evaluations (325 generations). For these parameters, the GA uses approximately 5.5×10^4 fitness function evaluations on average to reach the global fitness optimum. Notice that the global optimum is only consistently present in the population between generations 530 generation 570. After that, the global optimum is lost again until after generation 800. As we will show, this is a typical feature of the GA's behavior for parameter settings close to those that give

minimal $\langle E \rangle$. The global fitness optimum often only occurs in relatively short bursts after which it is lost again from the population. Notice also that there is only a small difference in $\langle f \rangle$ depending whether the best fitness is either 6 or 7 (the optimum).

Figure 6.1(c) shows a run for a small number ($N = 4$) of large ($K = 10$) blocks. The mutation rate is $q = 0.014$ and the population size is again $M = 150$. As in the three other runs we see that $\langle f \rangle$ goes through epochs punctuated by rapid increases in $\langle f \rangle$. We also see that the best fitness in the population typically jumps several times before the population fixes on a higher-fitness string. The GA takes about 1.9×10^5 fitness function evaluations on average to discover the global optimum for these parameter settings. In this run, the GA first discovered the global optimum after 2.7×10^5 fitness function evaluations. Notice that the optimum never stabilized in the population.

Finally, Fig. 6.1(d) shows a run with a large number ($N = 10$) of relatively small ($K = 5$) blocks. The mutation rate is $q = 0.008$ and the population size is $M = 100$. Notice that in this run, the best fitness in the population alternates several times between fitnesses 8, 9, and 10 before it reaches (fleetingly) the global fitness optimum of 11. Quickly after it has discovered the global optimum, it disappears again and the best fitness in the population largely alternates between 9 and 10 from then on. It is notable that this intermittent behavior of the best fitness is barely discernible in the behavior of $\langle f \rangle$. It appears to be lost in the "noise" of the average fitness fluctuations. The GA takes about 1.2×10^5 fitness function evaluations on average at these parameter settings to reach the global optimum; while in this particular run the GA took 1.6×10^5 fitness function evaluations (1640 generations) to briefly reach the optimum for the first time.

N	K	M	q	$\langle E \rangle$	σ
8	8	200	0.005	5.3×10^5	2.1×10^5
6	6	150	0.016	5.5×10^4	3.0×10^4
4	10	150	0.014	1.9×10^5	1.0×10^5
10	5	100	0.008	1.2×10^5	4.9×10^4

Table 6.1: Mean $\langle E \rangle$ and standard deviation σ of the expected number of fitness function evaluations for the Royal Staircase fitness functions and GA parameters shown in the runs of Fig. 6.1. The estimates were made from 500 GA runs and so the standard error in our estimates for $\langle E \rangle$ are the values of σ divided by $\sqrt{500}$.

6.5 Statistical Dynamics of Evolutionary Search

In Refs [137] and chapter 4 we developed the statistical dynamics of genetic algorithms to analyze the behavioral regimes of a GA searching the Royal Road fitness functions, which are closely related to the Royal Staircase fitness functions that we study here. The analysis here builds on those results and, additionally, is a direct extension of the optimization analysis and calculations in chapter 5. We briefly review the essential points from these previous papers. We refer the reader to chapter 4 for a detailed description of the similarities and differences of our theoretical approach with other theoretical approaches

such as the work by Prugel-Bennett, Rattray, and Shapiro, Refs. [117, 118, 119], the diffusion equation methods from mathematical population genetics developed by Kimura, Refs. [90] and [91], and the quasispecies theory of molecular evolution, Ref. [33].

6.5.1 Macrostate Space

Formally, the state of a population in an evolutionary search algorithm is only specified when the frequency of occurrence of each of the 2^L genotypes is given. Since 2^L is typically very large, the dimension of the corresponding microscopic state space is very large as well. One immediate consequence is that the evolutionary dynamic, on this level, is given by a stochastic (Markovian) operator of size (at least) $\mathcal{O}(2^L \times 2^L)$. Generally, using such a microscopic description makes analytical and quantitative predictions of the GA's behavior unwieldy. Moreover, since the practitioner is generally interested in the dynamics of some more macroscopic statistics, such as best and average fitness, a microscopic description is uninformative unless an appropriate projection onto the desired macroscopic statistic is found.

With these difficulties in mind, we choose to describe the macroscopic state of the population by its fitness distribution, denoted by a vector $\vec{P}$, where the components $0 \leq P_f \leq 1$ are the proportions of individuals in the population with fitness $f = 1, 2, \ldots, N+1$. We refer to $\vec{P}$ as the *phenotypic quasispecies*, following its analog in molecular evolution theory; see Refs. [32, 33, 34]. Since $\vec{P}$ is a distribution, it is normalized:

$$\sum_{f=1}^{N+1} P_f = 1. \tag{6.2}$$

The average fitness $\langle f \rangle$ of the population is given by:

$$\langle f \rangle = \sum_{f=1}^{N+1} f P_f. \tag{6.3}$$

6.5.2 The Evolutionary Dynamic

The fitness distribution $\vec{P}$ does not uniquely specify the microscopic state of the population. That is, there are many microstates (genotype distributions) with the same fitness distribution. An essential ingredient of the statistical dynamics approach is to assume a maximum entropy distribution of microstates conditioned on the macroscopic fitness distribution. Note that our approach shares a focus on fitness distributions and maximum entropy methods with that of Prugel-Bennett, Rattray, and Shapiro, Refs. [117, 118, 119]. In our case, the maximum entropy assumption entails that, given a fitness distribution $\vec{P}(t)$ at generation t, each microscopic population state with this fitness distribution is equally likely to occur.

A few comments on this maximum entropy method are in order. For the maximum entropy assumption to be useful it is not strictly necessary that the population takes on all genotype distributions equally often over the ensemble of instances for which a given fitness distribution occurs. (In fact, it is not difficult to find counterexamples for which

this is almost certainly false—and false for several reasons.) In order for the method to work we only require that the dynamics on the level of the fitness distributions, as calculated using the maximum entropy assumption, corresponds to the actual dynamics of the fitness distribution. That is, as long as the deviations of the actual genotype distributions from the maximum entropy distribution does not introduce a "bias" on the level of fitness distributions, the predictions will not be affected. Deciding whether or not the maximum entropy assumption works follows from comparing the theoretical predictions to data from simulations. In the case that the maximum entropy assumption *does* break down, it simply points out that additional macroscopic variables are needed to describe the macroscopic dynamics in which we are interested. For instance, in the following ultimately we are only interested in the dynamics of the best fitness in the population. However, taking the best fitness as the only variable describing the population and then introducing the maximum entropy assumption leads to unacceptably poor theoretical predictions. Rather we need to use the entire fitness distribution.

Given the maximum entropy assumption on the level of fitness distributions, we can construct a generation operator $\mathbf{G}$ that acts on the current fitness distribution and gives the *expected* fitness distribution of the population at the next time step. (See $\vec{P}(t) \rightarrow \mathbf{G}[\vec{P}(t)]$ illustrated in Fig. 6.2.) In the limit of infinite populations, which is similar to the thermodynamic limit in statistical mechanics, the fluctuations due to the finite size of the population are damped out, and this expected distribution is always exactly realized at the next generation. That is, the operator $\mathbf{G}$ maps the current fitness distribution $\vec{P}(t)$ deterministically to the fitness distribution $\vec{P}(t+1)$ at the next time step;

$$\vec{P}(t+1) = \mathbf{G}[\vec{P}(t)] \; . \tag{6.4}$$

Simulations indicate that for very large populations ($M \gtrapprox 2^L$) the dynamics on the level of fitness distributions is indeed deterministic and given by the above equation; thereby justifying the maximum entropy assumption at least in this infinite-population limit.

The operator $\mathbf{G}$ consists of a selection operator $\mathbf{S}$ and a mutation operator $\mathbf{M}$:

$$\mathbf{G} = \mathbf{M} \cdot \mathbf{S}. \tag{6.5}$$

The selection operator encodes the fitness-level effect of selection on the population; and the mutation operator, the fitness-level effect of mutation. Appendices E.1 and E.2 review the construction of these operators for our GA and the Royal Staircase fitness functions.

For now, we note that the infinite-population dynamics can be obtained by iteratively applying the operator $\mathbf{G}$ to the initial fitness distribution $\vec{P}(0)$. Thus, the solutions to the macroscopic equations of motion, in the limit of infinite populations, are formally given by

$$\vec{P}(t) = \mathbf{G}^{(t)}[\vec{P}(0)] \; . \tag{6.6}$$

Recalling Eq. (6.1), it is easy to see that the initial fitness distribution $\vec{P}(0)$ is given by:

$$P_n(0) = 2^{-K(n-1)} \left(1 - 2^{-K}\right) \; , \; 1 \leq n \leq N \; , \tag{6.7}$$

and

$$P_{N+1}(0) = 2^{-KN}. \tag{6.8}$$

As shown in Ref. [137] and chapter 4, the equations of motion of Eq. (6.6) can be linearized in a straightforward manner by introducing a linearized generator operator $\tilde{\mathbf{G}}$. The tth iterate $\mathbf{G}^{(t)}$ can then be directly obtained by solving for the eigenvalues and eigenvectors of the linearized version $\tilde{\mathbf{G}}$.

For large ($M \gtrapprox 2^L$) and infinite populations the dynamics of the fitness distribution is qualitatively very different from the behavior shown in Fig. 6.1: $\langle f \rangle$ increases smoothly and monotonically to an asymptote over a small number of generations. That is, there are no epochs. The reason is that for an infinite population, all genotypes are present in the initial population. Instead of the evolutionary dynamics *discovering* fitter strings over time, it essentially only expands the proportion of globally optimal strings already present in the initial population at $t = 0$. In spite of the qualitatively different dynamics for large populations, the (infinite population) operator $\mathbf{G}$ is the essential ingredient for describing the finite-population dynamics with its epochal dynamics as well, as we will now discuss.

6.5.3 Finite-Population Sampling

There are two important differences between the infinite-population dynamics and that with finite populations. The first is that with finite populations the components P_n cannot take on continuous values between 0 and 1. Since the number of individuals with fitness n in the population is necessarily an integer, the values of P_n are quantized in multiples of $1/M$. Thus, the space of allowed finite-population fitness distributions turns into a regular lattice in $N+1$ dimensions with a lattice spacing of $1/M$ within the simplex specified by the normalization Eq. (6.2).

Second, due to the sampling of members in the finite population, the dynamics of the fitness distribution is no longer deterministic. In general, we can only determine the conditional probabilities $\Pr[\vec{Q}|\vec{P}]$ that a given fitness distribution $\vec{P}$ leads to another $\vec{Q} = (Q_1, \ldots, Q_{N+1})$ in the next generation. These probabilities $\Pr[\vec{Q}|\vec{P}]$ are given by a multinomial distribution with mean $\mathbf{G}[\vec{P}]$:

$$\Pr[\vec{Q}|\vec{P}] = M! \prod_{n=1}^{N+1} \frac{\left(\mathbf{G}_n[\vec{P}]\right)^{m_n}}{m_n!} , \tag{6.9}$$

where the components Q_i are multiples of $1/M$: $Q_i = m_i/M$, with integers $0 \leq m_i \leq M$.

Equation (6.9) can be most easily understood as follows. The population for the next generation is created by selecting, copying, and mutating M times in the same way from the current population $\vec{P}$. This implies that each of the M individuals in the next generation has equal and independent probabilities q_i to be of fitness i. These probabilities q_i also give the *expected* proportions q_i of individuals with fitness i in the next generation. The *actual* proportions Q_i of individuals with fitness i in the next generation are then given by a multinomial sample of size M from the distribution of expected proportions q_i. Since in the limit $M \to \infty$ of infinite populations we have that the expected proportions equal the actual proportions, we necessarily have that $q_i = \mathbf{G}_i[\vec{P}]$. In other words, the probabilities $\Pr[\vec{Q}|\vec{P}]$ are given by a multinomial sampling distribution of sample

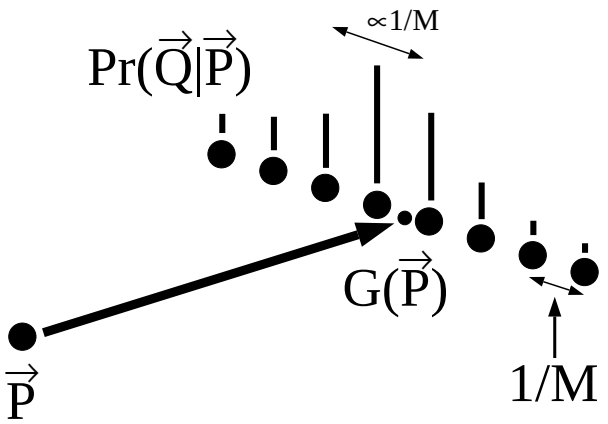

Figure 6.2: Illustration of the stochastic dynamics involved in going from one generation to the next with a finite population. The arrow points from the current distribution $\vec{P}$ to the expected next generation distribution $\mathbf{G}[\vec{P}]$. The large dots indicate points of the lattice of allowed finite population fitness distributions. The columns above these lattice points indicate the size of $\Pr[\vec{Q}|\vec{P}]$. The width of this distribution is inversely proportional to the population size M. Note that the expected distribution $\mathbf{G}[\vec{P}]$ (small dot) does not occur at a lattice point.

size M with mean $\mathbf{G}[\vec{P}]$; just as Eq. (6.9) expresses. In mathematical population genetics, such multinomial sampling Markov models are known as Wright-Fisher models see Ref. [69, pp. 66-70] and also Ref. [39].

Thus, for any finite-population fitness distribution $\vec{P}$ the (infinite population) operator $\mathbf{G}$ still gives the GA's *average* dynamics over one time step. Note that the components $\mathbf{G}_i[\vec{P}]$ need not be multiples of $1/M$. Therefore, the *actual* fitness distribution $\vec{Q}$ at the next time step is not $\mathbf{G}[\vec{P}]$, but is instead one of the allowed lattice points in the finite-population state space. Since the variance around the expected distribution $\mathbf{G}[\vec{P}]$ is proportional to $1/M$, $\vec{Q}$ tends to be one of the lattice points close to $\mathbf{G}[\vec{P}]$. This finite-population dynamics is illustrated in Fig. 6.2.

6.5.4 Epochal Dynamics

We will now discuss how the epochal behavior of the dynamics for a finite population comes about within the mathematical framework presented above.

For finite populations, the expected change $\langle d\vec{P} \rangle$ in the fitness distribution over one generation is given by:

$$\langle d\vec{P} \rangle = \mathbf{G}[\vec{P}] - \vec{P} \; . \tag{6.10}$$

Assuming that some component $\langle dP_i \rangle$ and its variance are much smaller than $1/M$, the actual change in component P_i is likely to be $dP_i = 0$ for a long succession of generations. That is, if the size of the *flow* $\langle dP_i \rangle$ in some direction i is much smaller than the lattice spacing $(1/M)$ for the finite population, we expect the fitness distribution

to not change in direction (fitness) i. In Ref. [137] and chapter 4 we showed that, from a mathematical point of view, this is the mechanism by which finite populations cause epochal dynamics.

For the Royal Staircase fitness functions, we have that whenever fitness n is the highest in the population, such that $P_i = 0$ for all $i > n$, the rate at which higher-fitness strings are discovered is very small: $\langle dP_i \rangle \ll 1/M$ for all $i > n$, provided that the population size M is not too large. A period of stasis (an evolutionary *epoch*) thus corresponds to the time the population spends before it discovers a string of fitness higher than n. More formally, each epoch n corresponds to the population being restricted to a region in the n-dimensional lower-fitness subspace consisting of fitnesses 1 to n of the macroscopic state space. Stasis occurs because the flow out of this subspace is much smaller than the finite-population induced lattice spacing.

As the experimental runs of Fig. 6.1 illustrated, each epoch in the average fitness is associated with a (typically) constant value of the best fitness in the population. More detailed experiments reveal that not only is $\langle f \rangle$ constant on average during the epochs, in fact the entire fitness distribution $\vec{P}$ fluctuates in an approximately Gaussian way around some constant fitness distribution $\vec{P}^n$ during the epoch n—the generations when n is the highest fitness in the population.

As was shown in chapter4 , each epoch fitness distribution $\vec{P}^n$ is the unique fixed point of the operator $\mathbf{G}$ restricted to the n-dimensional subspace of strings with $1 \leq f \leq n$. That is, if $\mathbf{G}^n$ is the projection of the operator $\mathbf{G}$ onto the n-dimensional subspace of fitnesses from 1 up to n, then we have:

$$\mathbf{G}^n[\vec{P}^n] = \vec{P}^n \ . \tag{6.11}$$

Intuitively, the operator $\mathbf{G}^n$ gives the average change in the fitness distribution *conditioned* on no strings of fitness higher than n being generated. The epoch distribution is then the fixed point of this operator. The uniqueness and construction of these fixed points will be discussed further below. By Eq. (6.3), then, the average fitness f_n in epoch n is given by:

$$f_n = \sum_{j=1}^{n} j P_j^n. \tag{6.12}$$

To summarize at this point, the statistical dynamics analysis is tantamount to the following qualitative picture. The global dynamics can be viewed as an incremental discovery of successively more (macroscopic) dimensions of the fitness distribution space. Initially, only strings of low fitness are present in the initial population. The population stabilizes on the epoch fitness distribution $\vec{P}^n$ corresponding to the best fitness n in the initial population. The fitness distribution fluctuates around the n-dimensional vector $\vec{P}^n$ until a string of fitness $n+1$ is discovered and spreads through the population. The population then settles into the $(n+1)$-dimensional fitness distribution $\vec{P}^{n+1}$ until a string of fitness $n+2$ is discovered, and so on, until the global optimum at fitness $N+1$ is found. In this way, the global dynamics can be seen as stochastically hopping between the different epoch distributions $\vec{P}^n$, unfolding a new macroscopic dimension of the fitness distribution space each time a higher-fitness string is discovered.

Whenever mutation creates a string of fitness $n+1$, this string may either disappear before it spreads, seen as the transient jumps in best fitness in Fig. 6.1, or it may spread,

leading the population to fitness distribution $\vec{P}^{n+1}$. We call the latter process an *innovation*. Through an innovation, a new (macroscopic) dimension of fitness distribution space becomes stable.

Fig. 6.1 also showed that it is possible for the population to fall from epoch n (say) down to epoch $n-1$. This happens when, due to fluctuations, all individuals of fitness n are lost from the population. We refer to this as a *destabilization* of epoch n. Through a destabilization, a dimension can, so to speak, collapse. For some parameter settings, such as shown in Figs. 6.1(a) and 6.1(c), this is very rare. In these cases, the time for the GA to reach the global optimum is mainly determined by the time it takes to discover strings of fitness $n+1$ in each epoch n. For other parameter settings, however, such as in Figs. 6.1(b) and 6.1(d), the destabilizations play an important role in how the GA reaches the global optimum. In these regimes, destabilization must be taken into account in calculating search times. This is especially important in the current setting since, as we will show, the optimized GA often operates in this type of marginally stable parameter regime where later epochs destabilize quite easily.

6.6 Quasispecies Distributions and Epoch Fitness Levels

During epoch n the quasispecies fitness distribution $\vec{P}^n$ is given by a fixed point of the (projected) operator $\mathbf{G}^n$. To obtain this fixed point we linearize the generation operator by taking out the factor $\langle f \rangle$, thereby defining a new operator $\tilde{\mathbf{G}}^n$ via:

$$\mathbf{G}^n = \frac{1}{\langle f \rangle} \tilde{\mathbf{G}}^n \ , \tag{6.13}$$

where $\langle f \rangle$ is the average fitness of the fitness distribution that $\mathbf{G}^n$ acts upon; see App. E.1. The operator $\tilde{\mathbf{G}}^n$ is just an ordinary (linear) matrix operator and therefore, the fixed point equation (6.11) simply becomes an eigenvector equation. Since all components of $\tilde{\mathbf{G}}^n$ are positive, this fixed point is unique, from the positive matrix theorem by Perron [53]. The quasispecies fitness distribution $\vec{P}^n$ is given by the principal eigenvector of the matrix $\tilde{\mathbf{G}}^n$ (normalized in probability). From Eq. (6.13) it also follows that the principal eigenvalue f_n of $\tilde{\mathbf{G}}^n$ equals the average fitness of the quasispecies distribution. In this way, obtaining the quasispecies distribution $\vec{P}^n$ reduces to calculating the principal eigenvector of the matrix $\tilde{\mathbf{G}}^n$; see App. E.3.

The matrices $\tilde{\mathbf{G}}^n$ are generally of modest size: i.e., their dimension is smaller than the number of blocks N and substantially smaller than the dimension of genotype space. Due to this we can easily obtain numerical solutions for the epoch fitnesses f_n and the epoch quasispecies distributions $\vec{P}^n$. For a clearer understanding of the functional dependence of the epoch fitness distributions on the GA's parameters, however, App. E.3 recounts analytical approximations to the epoch fitness levels f_n and quasispecies distributions $\vec{P}^n$.

The result is that the average fitness f_n in epoch n, which is given by the largest eigenvalue, is equal to the largest diagonal component of the analytical approximation to $\tilde{\mathbf{G}}^n$ derived in App. E.3. That is,

$$f_n = n(1-q)^{(n-1)K} \ . \tag{6.14}$$

The epoch quasispecies is given by:

$$P_i^n = \frac{(1-\lambda)n\lambda^{n-1-i}}{n\lambda^{n-1-i} - i} \prod_{j=1}^{i-1} \frac{n\lambda^{n-j} - j}{n\lambda^{n-1-j} - j} \,, \tag{6.15}$$

where $\lambda = (1-q)^K$ is the probability that a block will undergo no mutations. For the following, we are actually interested in the most-fit quasispecies component P_n^n in epoch n. For this component, Eq. (6.15) reduces to

$$P_n^n = \lambda^{n-1} \prod_{j=1}^{n-1} \frac{f_n - f_j}{f_n - \lambda f_j} \,, \tag{6.16}$$

where we have expressed the result in terms of the epoch fitness levels $f_j = j\lambda^{j-1}$.

6.7 Mutation Rate Optimization

In the previous sections we argued that the GA's behavior can be viewed as stochastically hopping from one epoch to the next—when the search discovers a string with increased fitness that spreads in the population. Assuming that the total time to reach this global optimum is dominated by the time the GA spends in the epochs, chapter 5 developed a way to tune the mutation rate q such that the time the GA spends in an epoch is minimized. We briefly review this here before moving on to the more general theory that includes population-size effects and epoch destabilization.

To move from epoch n to epoch $n+1$, a string of fitness $n+1$ has to be discovered and spread through the population. During epoch n, the population is in a metastable state where it fluctuates around a constant fitness distribution $\vec{P}^n$. To a good approximation, we can assume that in each generation there is an equal and independent probability that epoch n will end by creating a fitness $n+1$ string that spreads through the population. Note that this immediately implies that the distribution of epoch times is geometric for each individual epoch.

The creation of a fitness $n+1$ string is most likely to occur through a string of fitness n mutating its nth block to the correct configuration of all 1s. Optimizing the mutation rate now amounts to finding a balance between two opposing effects of varying mutation rate. On the one hand, when the mutation rate is increased, the average number of mutations in the unaligned blocks goes up, thereby increasing the probability of creating newly aligned blocks. On the other hand, due to the increased number of deleterious mutations, the equilibrium proportions P_n^n of individuals in the highest fitness class during each epoch n decreases and so the number of individuals that are likely to discover a string of fitness $n+1$ decreases.

In chapter 5 we derived an expression for the probability C_{n+1} to create, over one generation in epoch n, a string of fitness $n+1$ that will stabilize by spreading through the population. This is given by

$$C_{n+1} = MP_n^n P_a \pi_n(\lambda) \,, \tag{6.17}$$

where $P_a = (1-\lambda)/(2^K-1)$ is the probability of aligning a block (see App. E.2) and $\pi_n(\lambda)$ is the probability that a string of fitness $n+1$ will spread, as opposed to being lost through a fluctuation or a deleterious mutation. This spreading probability π_n can be calculated using a diffusion-equation approximation similar to the ones developed in population genetics by Kimura [90]. In chapter 4 we showed how to adapt this diffusion-equation method to present type of problem. We found that the spreading probability π_n largely depends on the relative average fitness difference of epoch $n+1$ over epoch n. Denoting this difference as

$$\begin{aligned} \gamma_n &= \frac{f_{n+1}-f_n}{f_n} \\ &= \left(1+\frac{1}{n}\right)\lambda - 1 \,, \end{aligned} \tag{6.18}$$

where we have used Eq. (6.14), one finds:

$$\pi_n(\lambda) = \frac{1-\left(1-\frac{1}{M}\right)^{2M\gamma_n+1}}{1-\left(1-P_{n+1}^{n+1}\right)^{2M\gamma_n+1}} \,. \tag{6.19}$$

If $P_{n+1}^{n+1} \gg 1/M$, this reduces to a population-size independent estimate of the spreading probability

$$\pi_n \approx 1 - e^{-2\gamma_n} \,. \tag{6.20}$$

If one were to allow for changing mutation rates between epochs, one would minimize the time spent in each epoch by maximizing C_{n+1} of Eq. (6.17) using Eqs. (6.16), (6.18), and (6.20). Note that C_{n+1} depends on q only through λ. The optimal mutation rate in each epoch n is determined by estimating the optimal value λ_o of λ for each n. We found that λ_o is well approximated (chapter 5) by

$$\lambda_o(n) \approx 1 - \frac{1}{3n^{1.175}} \,. \tag{6.21}$$

For large n this gives the optimal mutation rate as

$$q_o \approx \frac{1}{3Kn^{1.175}} \,,\; n \gg 1 \,. \tag{6.22}$$

Thus, the optimal mutation rate drops as a power-law in both n and K. This implies that if one is allowed to adapt the mutation rate during the run, the mutation rate should decrease as a GA run progresses so that the search will find the global optimum as quickly as possible.

We now turn to the simpler problem of optimizing mutation rate for the case of a *constant* mutation rate throughout a GA run. In chapter 5 we used Eq. (6.17) to estimate the total number E of fitness function evaluations the GA uses on average before an optimal string of fitness $N+1$ is found. As a first approximation, we assumed that the GA visits all epochs, that the time spent in innovations between them is negligible, and that epochs are *always* stable. The epoch stability assumption entails that it is assumed to be highly unlikely that strings with the current highest fitness will disappear from

the population through a fluctuation, once such strings have spread. These assumptions appear to hold for the parameters of Figs. 6.1(a) and 6.1(c). They may hold even for the parameters of Fig. 6.1(b), but they most likely do not for Fig. 6.1(d). For the parameters of Fig. 6.1(d), we see that the later epochs ($n = 9$ and 10) easily destabilize a number of times before the global optimum is found. Although we will develop a generalization that addresses this more complicated behavior in the next sections, it is useful to work through the optimization of mutation rate under the stability assumption first.

The average number T_n of generations that the population spends in epoch n is simply $1/C_{n+1}$, the inverse of the probability that a string of fitness $n+1$ will be discovered and spread through the population. For a population of size M, the number of fitness function evaluations per generation is M, so that the total (average) number E_n of fitness function evaluations in epoch n is given by MT_n. More explicitly, we have:

$$E_n = \left(P_n^n P_a \pi_n\right)^{-1} . \tag{6.23}$$

That is, the total number of fitness function evaluations in each epoch is independent of the population size M. This is due to two facts, given our approximations. First, the epoch lengths, measured in generations, are inversely proportional to M, while the number of fitness function evaluations per generation is M. Second, since for stable epochs $P_n^n \gg 1/M$, the probability π_n is also independent of population size M; recall Eq. (6.20).

The total number of fitness function evaluations $E(\lambda)$ to reach the global optimum is simply given by substituting into Eq. (6.23) our analytical expressions for P_n^n and π_n, Eqs. (6.16) and (6.20), respectively, and then summing $E_n(\lambda)$ over all epochs n from 1 to N. We then have:

$$E(\lambda) = \sum_{n=1}^{N} \frac{1}{P_a \pi_n(\lambda)} \prod_{i=1}^{n-1} \frac{n\lambda^{n-i-1} - i}{n\lambda^{n-i} - i} . \tag{6.24}$$

Note that in the above equation we set $\pi_N = 1$ by definition because the algorithm terminates as soon as a string of fitness $N+1$ is found. That is, strings of fitness $N+1$ need not spread through the population, they just need to be discovered once. The optimal mutation rate for an entire run is then obtained by minimizing Eq. (6.24) with respect to λ.

Figure 6.3 shows for $N = 4$ blocks of length $K = 10$ bits the dependence of the average total number $E(q)$ of fitness function evaluations on the mutation rate q. The dashed line is the theoretical prediction of Eq. (6.24); while the solid lines show the experimentally estimated values of $\langle E \rangle$ for four different population sizes. Each experimental data point is an estimate obtained from 250 GA runs. Figure 6.3 illustrates in a compact form the findings of chapter 5, which can be summarized as follows.

1. At fixed population size M, there is a smooth cost function $E(q)$ as a function of mutation rate q. It has a *single* and *shallow* minimum q_o, which is accurately predicted by the theory.

2. The curve $E(q)$ is everywhere concave.

3. The theory slightly underestimates the experimentally obtained $\langle E \rangle$.

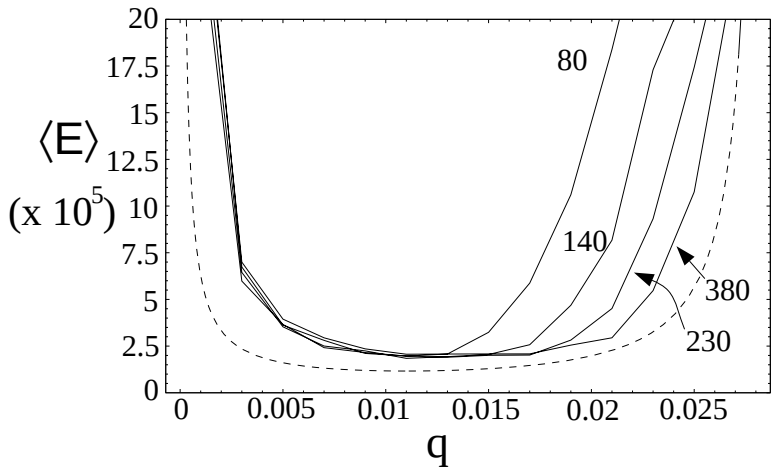

Figure 6.3: Average total number $\langle E \rangle$ of fitness function evaluations as a function of mutation rate q, from the theory (dashed), Eq. (6.24), and from experimental estimates (solid). The fitness function parameter settings are $N = 4$ blocks of length $K = 10$ bits. The mutation rate runs from $q = 0.001$ to $q = 0.028$. Experimental data points are estimates over 250 runs each. The experimental curves show four different population sizes; $M = 80$, $M = 140$, $M = 230$, and $M = 380$.

4. The optimal mutation rate q_o roughly occurs in the regime (between $q = 0.01$ and $q = 0.015$) where the highest epochs are marginally stable; see Fig. 6.1.

5. For mutation rates lower than q_o the experimentally estimated total number of fitness function evaluations $\langle E \rangle$ grows steadily and becomes almost independent of the population size M. (This is where the experimental curves in Fig. 6.3 overlap.) For mutation rates larger than q_o the total number of fitness function evaluations does depend on M, which is not explained by the preceding theory.

6. There is a mutational error threshold in q that bounds the upper limit in q for which the GA can discover the optimum *at all*. Above the threshold, which is population-size independent, suboptimal strings of fitness N cannot stabilize in the population, even for very large population sizes. This error threshold is also correctly predicted by the theory. It occurs around $q_c = 0.028$ for $N = 4$ and $K = 10$.

Before embarking on the population-size dependent analysis, it is useful to make a few comments about the role of crossover if it had been included in our GA. As explained in more detail in chapter 5, during epoch n *all* individuals in the population are relatively recent descendants from a string of fitness n. More precisely, strings with fitness $i < n$, have on average less than 1 offspring with fitness i. Therefore, lineages entirely consisting of suboptimal strings tend to be short lived. This, in turn, implies that all individuals are relatively recent descendants of a fitness n string. This means that all strings in the population share the genetic content of most of their blocks. That is,

they differ only by a relatively small number of mutations. In particular, *all* strings in the population are unlikely to have their nth block aligned. This implies that it is almost impossible for a crossover event to create a string of fitness $n+1$. Such an event can only occur when strings of fitness n are crossed over, the crossover point falls *within* the nth block, *and* the corresponding subblocks form a new aligned block. In general, the contribution of such beneficial events is marginal, especially taking into account the deleterious effects that crossover also produces, when high- and low-fitness parents combine to form two low-fitness offspring. Thus, the role of crossover is epochal evolution is marginal.

6.8 Epoch Destabilization: Population-Size Dependence

We now extend the above analysis to account for E's dependence on population size. This not only improves the parameter-optimization theory, but also leads us to consider a number of issues and mechanisms that shed additional light on how GAs work near their optimal parameter settings. Since it appears that optimal parameter settings often lead the GA to run in a behavioral regime were the population dynamics is marginally stable in the higher epochs, we consider how epoch destabilization dynamics affects the time to discover the global optimum.

We saw in Figs. 6.1(b) and 6.1(d) that, around the optimal parameter settings, the best fitness in the population can show intermittent behavior. Apparently, fluctuations sometimes cause an epoch's current best strings (of fitness n) in the population to disappear. The best fitness then drops to $n-1$. Often, strings of fitness n are rediscovered later on. What happens is that for these higher epochs, the fluctuations in the proportion P_n of strings with fitness n becomes comparable to the average P_n^n of this proportion. That is, a fluctuation may bring P_n very close to 0, and so all strings of fitness n may be lost. As we will see, the probability of a destabilization is sensitive to the population size M, introducing population-size dependence in the average total number $\langle E \rangle$ of fitness function evaluations.

As we just noted, the theory for $E(q)$ used in chapter 5 assumed that destabilizations do not occur, leading to a population-size independent theory. However, as is clear from Fig. 6.1(d), it is possible that epoch n destabilizes several times to epoch $n-1$ before the population moves to epoch $n+1$, and this will considerably alter E. For example, if during epoch n the population is 3 times as likely to destabilize to epoch $n-1$ compared to innovating to epoch $n+1$, then we expect epoch n to destabilize three times on average before moving to epoch $n+1$. Assuming that epoch $n-1$ is stable, this means that epoch n has to be rediscovered 3 times on average before epoch $n+1$ is discovered. This will effectively increase the time spend in epoch n by 3 times the average number of generations spent in epoch $n-1$.

To make these ideas precise we introduce a Markov chain model to describe the "hopping" up and down between the epochs. The Markov chain has $N+1$ states, each representing an epoch. In every generation there are probabilities p_n^+ to innovate from epoch n to epoch $n+1$ and p_n^- to destabilize, falling from epoch n to epoch $n-1$. The globally optimum state $N+1$ is an absorbing state. Starting from epoch 1 we calculate the expected number T of generations for the population to reach the absorbing state for

the first time.

The innovation probabilities p_n^+ to move from epoch n to $n+1$ are just given by the C_{n+1} of Eq. (6.17):

$$p_n^+ = C_{n+1} = \frac{M}{E_n}, \tag{6.25}$$

where E_n is given by the approximation of Eq. (6.23). Note that when MP_n^n approaches 1 the spreading probability π_n, as given by Eq. (6.19), becomes population-size dependent as well, and we use Eq. (6.19) rather than Eq. (6.20). To obtain the destabilization probabilities p_n^- we assume that in each generation the population has an equal and independent probability to destabilize to epoch $n-1$. This probability is given by the inverse of the average time until a destabilization occurs.

To calculate the average time D_n that the population spends in epoch n before it destabilizes we have to analyze the dynamics of fluctuations in the proportion P_n of individuals with fitness n. As was explained in detail in chapter 4, this can be analyzed most easily using a diffusion-equation approximation. (See Ref. [90] for an introduction to using diffusion equations in these settings.)

In short, we introduce the deviation $x(t)$ from the mean proportion P_n^n of individuals with fitness n by defining $P_n(t) = P_n^n + x(t)$. We then calculate the expected change $\langle \delta x \rangle = \langle x(t+1) \rangle - x(t)$ and the second moment $\langle (\delta x)^2 \rangle$ of the expected change in x. The dynamics of the probability distribution $\Pr(x,t)$ that the deviation will be x at time t is then approximated by the Fokker-Planck equation:

$$\frac{\partial}{\partial t}\Pr(x,t) = \frac{\partial}{\partial x}\langle \delta x \rangle \Pr(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\langle (\delta x)^2 \rangle \Pr(x,t) \; . \tag{6.26}$$

Once we have expressions for $\langle \delta x \rangle$ and $\langle (\delta x)^2 \rangle$ we can use these diffusion equations to solve for such statistics as the variance $\mathrm{Var}(P_n)$ of the proportion of highest fitness individuals during epoch n and the average time D_n for P_n to reach values smaller than $1/M$—the time at which all strings of fitness n are lost. See, for instance, Ref. [54] for the details of such calculations.

For small x, the values $\langle (\delta x)^2 \rangle$ can be approximated by the values of $\langle (\delta x)^2 \rangle$ when x is *zero*. When $x = 0$, we have that $P_n = P_n^n$, and the *variance* in P_n over one generation is simply given by the variance due to the multinomial sampling of Eq. (6.9):

$$\langle (\delta x)^2 \rangle = \frac{P_n^n (1 - P_n^n)}{M} \; . \tag{6.27}$$

This is the approximation we will use for $\langle (\delta x)^2 \rangle$ in Eq. (6.26).

The calculation of $\langle \delta x \rangle$ is somewhat more involved and will not be presented here (see chapter 4) in detail, because of space constraints. In general we find the intuitive result that the deviation x is expected to be scaled down by a constant factor each generation:

$$\langle \delta x \rangle = -\mu_n x \; . \tag{6.28}$$

The scale factor μ_n is a function of the epoch distributions $\vec{P}^i$ and epoch fitnesses f_i. Said simply, the fluctuations in P_n are the result of fluctuations in the directions of all lower lying epochs $\vec{P}^i$ with $i < n$. The fluctuations in the direction of epoch i are scaled

down at an average rate $(f_n - f_i)/f_n$ per generation. Thus, fluctuations in the direction of low epochs are scaled down most rapidly. The quantity μ_n is then a weighted sum:

$$\mu_n = \frac{\sum_{i=1}^{n-1}(f_n - f_i)B_i}{f_n \sum_{i=1}^{n-1} B_i} , \tag{6.29}$$

where B_i is a measure of how much of the multinomial sampling fluctuations occur in the direction of epoch i. The expected fluctuations in components i and j due to multinomial sampling are given by

$$\langle dP_i dP_j \rangle = \frac{P_i^n(\delta_{ij} - P_j^n)}{M}, \tag{6.30}$$

when $x = 0$. We calculate the weights B_i by calculating the overlap of these fluctuations with the epoch distributions $\vec{P}^i$ for each $i < n$. These can be calculated by introducing a matrix $\mathbf{R}$ that contains the epoch distributions in its columns:

$$\mathbf{R}_{ij} = P_i^j , \tag{6.31}$$

The overlaps B_i are then calculated using the inverse of $\mathbf{R}$:

$$B_i = \frac{1}{M} \sum_{k,m=1}^{n-1} \mathbf{R}_{ik}^{-1}\mathbf{R}_{im}^{-1}P_k^n \left(\delta_{km} - \vec{P}_m^n\right). \tag{6.32}$$

Again, the reader is referred to chapter 4 for the details of this procedure. Generally, μ_n decreases monotonically as a function of n since fluctuations in the proportion P_n^n of individuals in the highest-fitness class n decay more slowly for higher epochs.

Continuing at this level of summary, the variance is simply given by $\mathrm{Var}(P_n) = P_n^n(1 - P_n^n)/(M\mu_n)$, and the average time until destabilization is approximately given by

$$D_n = \frac{MP_n^n}{1 - P_n^n} + \frac{\pi}{2\mu_n}\mathrm{erfi}\left[\sqrt{\frac{M\mu_n P_n^n}{1 - P_n^n}}\right]\mathrm{erf}\left[\sqrt{\frac{M\mu_n(1 - P_n^n)}{P_n^n}}\right] , \tag{6.33}$$

where $\mathrm{erf}(x)$ is the error function and $\mathrm{erfi}(x) = \mathrm{erf}(ix)/i$ is the imaginary error function.

Notice that the argument of erfi(x), $\sqrt{M\mu_n P_n^n/(1 - P_n^n)}$, is the ratio between the mean proportion P_n^n and standard deviation of the number of individuals with fitness n. The function $\mathrm{erfi}(x)$ is a very rapidly growing function of its argument: $\mathrm{erfi}(x) \approx \exp(x^2)/x$ for x larger than 1. Therefore, $\sqrt{M\mu_n P_n^n/(1 - P_n^n)}$ being either smaller (larger) than 1 is a reasonable criterion for the instability (stability) of an epoch. When the standard deviation of P_n is (much) smaller than the mean P_n^n, epochs are stable for a very long time; while they become unstable very quickly as the ratio of the standard deviation and the mean approaches 1.

The above formula Eq. (6.33) is analogous to error thresholds in the theory of molecular evolution. Generally, error thresholds denote the boundary in parameter space between a regime where a certain high fitness string, or an equivalence class of high

fitness strings, is stable in the population and a regime where it is unstable. In the case of a single high-fitness *master sequence* one speaks of a genotypic error threshold; see Refs. [2, 33, 110, 133]. In the case of an equivalence class of high-fitness strings, one speaks of a *phenotypic* error threshold; see Refs. [79, 50].

A sharply defined error threshold generally only occurs in the limit of infinite populations and infinite string length [98], but extensions to finite population cases have been studied in Refs. [2, 110, 50]. In Ref. [50], for example, the occurrence of a finite-population phenotypic error threshold was defined by the equality of the standard deviation and the mean of the number of individuals of the highest fitness class. This definition is in accord with Eq. (6.33) as we explained above.

The average time until destabilization is thus given by D_n of Eq. (6.33), and so the average probability per generation for a destabilization to occur is simply its inverse:

$$p_n^- = \frac{1}{D_n} \, . \tag{6.34}$$

Finally, note that the probability to remain in epoch n is $1 - p_n^+ - p_n^-$.

We now have expressions for all of the Markov chain's transition probabilities. With these it is straightforward to calculate the average number T of generations before the GA discovers the global optimum for the first time. This is done by calculating the average time for the Markov chain to reach its absorbing state, starting from epoch 1. Following, for instance, Sec. 7.4 of Ref. [54]) the result is

$$T = \sum_{n=1}^{N} \phi_n \sum_{k=1}^{n} \frac{1}{p_k^+ \phi_k} \, , \tag{6.35}$$

where ϕ_n is defined as:

$$\phi_n = \prod_{k=2}^{n} \frac{p_k^-}{p_k^+}, \; n \geq 2 \, , \tag{6.36}$$

and

$$\phi_1 = 1 \, . \tag{6.37}$$

Since Eq. (6.35) gives the average number T of generations, the average number of fitness function evaluations $E(q, M)$ is given by:

$$\begin{aligned} E(q,M) &= MT \\ &= E_N + E_{N-1}\left(1 + \frac{E_N}{MD_N}\right) \\ &+ E_{N-2}\left(1 + \frac{E_{N-1}}{MD_{N-1}}\left(1 + \frac{E_N}{MD_N}\right)\right) \\ &+ \cdots \, , \end{aligned} \tag{6.38}$$

where the E_n are given by Eq. (6.23) and where the last equality is obtained by rewriting the sums in Eq. (6.35). As epochs become arbitrarily stable ($D_n \to \infty$), the terms with D_n in the denominator go to zero, and Eq. (6.38) reduces to Eq. (6.24), as it should.

6.9 Theory versus Experiment

We can now compare this population-size dependent approximation, Eq. (6.38), with the experimentally measured dependence on M of the average total number $\langle E \rangle$ of fitness function evaluations. Figure 6.4 shows the dependence of $\langle E \rangle$ on the population size M for two different parameter settings of N and K and for a set of mutation rates q.

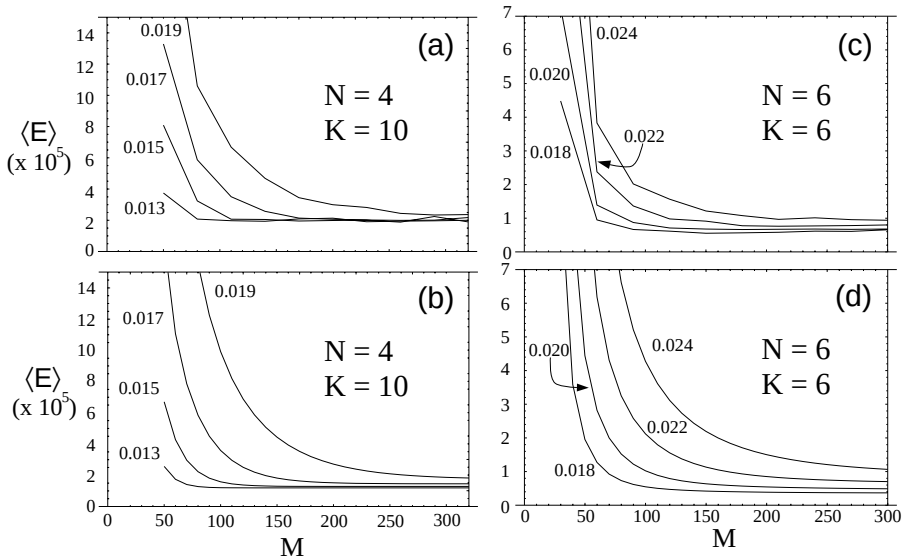

Figure 6.4: Average total number $\langle E \rangle$ of fitness function evaluations as a function of the population size M for two different fitness function parameters and four mutation rates each, both experimentally (Figs. (a) and (c), top row) and theoretically (Figs. (b) and (d), bottom row). In each figure each solid line gives $E(M)$ for a different mutation rate. Each experimental data point is an average over 250 GA runs. Figures (a) and (b) have $N = 4$ blocks of length $K = 10$. The upper figure (a) shows the experimentally estimated $E(M)$ as a function of M for the mutation rates $q \in \{0.013, 0.015, 0.017, 0.019\}$. The lower figure (b) shows the theoretical results, as given by Eq. (6.38), for the same parameter settings. In both, the population size ranges from $M = 50$ to $M = 320$ on the horizontal axis. Figures (c) and (d) have $N = 6$ blocks of length $K = 6$. Figure (c) shows the experimental averages and figure (d) the theoretical predictions for the same parameter settings. The population sizes on the horizontal axis run from $M = 30$ to $M = 300$. The mutation rates shown in (c) and (d) are $q \in \{0.018, 0.02, 0.022, 0.024\}$.

The upper figures, Figs. 6.4(a) and 6.4(c), give the dependence of the experimentally estimated $\langle E \rangle$ on the population size M. The lower figures, Figs. 6.4(b) and 6.4(d), give the theoretical predictions from Eq. (6.38). The upper left figure, Fig. 6.4(a), shows $\langle E \rangle$ as a function of M for $N = 4$ blocks of length $K = 10$ for four different mutation rates: $q \in \{0.013, 0.015, 0.017, 0.019\}$. The population size ranges from $M = 50$ to $M = 320$. The total number of fitness function evaluations on the ver-

tical axis ranges from $\langle E \rangle = 0$ to $\langle E \rangle = 15 \times 10^5$. Each data point was obtained as an average over 250 GA runs. Figure 6.4(b) shows the theoretical predictions for the same parameter settings. Figure 6.4(c) gives the experimental estimates for $N = 6$ blocks of length $K = 6$, over the range $M = 30$ to $M = 300$, for four mutation rates: $q \in \{0.018, 0.02, 0.022, 0.024\}$. The total number of fitness function evaluations on the vertical axis range from $\langle E \rangle = 0$ to $\langle E \rangle = 7 \times 10^5$. Figure 6.4(d) shows the theoretical predictions for the same range of M and the same four mutation rates.

We see that as the population size becomes "too small" destabilizations make the total number of fitness function evaluations increase rapidly. The higher the mutation rate, the higher the population size at which the sharp increase in $\langle E \rangle$ occurs. These qualitative effects are captured accurately by the theoretical predictions from Eq. (6.38). Although our analysis involves several approximations (e.g. as in the calculations of D_n), the theory does quantitatively capture the population-size dependence well, both with respect to the predicted absolute number of fitness function evaluations and the shape of the curves as a function of M for the different mutation rates. From Figs. 6.4(c) and 6.4(d) it seems that the theory overestimates the growth of $\langle E \rangle$ for the larger mutation rates as the population size decreases. Still, the theory correctly captures the sharp increase of $\langle E \rangle$ around a population size of $M = 50$.

As the population size increases beyond approximately $M = 200$, we find experimentally that the average total number of fitness function evaluations $\langle E \rangle$ starts rising very slowly as a function of M. This effect is not captured by our analysis. It is also barely discernible in Figs. 6.4(a) and 6.4(c). We believe that the slow increase of $\langle E \rangle$ for large population sizes derives from two sources.

First, by the maximum entropy assumption, our theory assumes that all individuals in the highest fitness class are genetically *independent*, apart from the sharing of their aligned blocks. Under that assumption, the average number of fitness function evaluations to discover a string of fitness $n+1$ in epoch n is independent of M. A population of size $2M$ is assumed to take half as many generations to discover a higher fitness string as a population of size M. This is not true in general. The sampling during the selection process introduces genetic correlations in the individuals of the highest fitness class. Due to these correlations, the MP_n^n strings of fitness n are not searching for a higher fitness string *independently* and therefore the probability (per generation) to discover a higher fitness string grows somewhat slower than linearly with M. Since the number of fitness function evaluations *does* grow linearly with M, the correlation effect leads to a slow growth of $\langle E \rangle$ with M. Unfortunately, this effect is very hard to address analytically and quantitatively.

The second reason for the increase of E with increasing population size comes from the time the population spends in the short innovations between the different epochs. During the innovation, the single mutant of fitness $n+1$ amplifies in the population until its epoch equilibrium value P_{n+1}^{n+1} is reached. Up to now, we have neglected these innovation periods. Generally, they only contribute marginally to E. In chapter 4 we calculated the approximate number of generations g_n that the population spends in the innovation from epoch n to epoch $n+1$ and found that:

$$g_n = \frac{2 + \gamma_n}{\gamma_n} \log [M] , \tag{6.39}$$

where γ_n is the fitness differential given by Eq. (6.18). That is, the number of generations taken is proportional to the logarithm of the population size, and grows with decreasing fitness differentials γ_n. Using this, we find for the total number of fitness function evaluations I that the GA expends in the innovations

$$I = M \log [M] \sum_{n=1}^{N-1} \frac{2 + \gamma_n}{\gamma_n} \, . \tag{6.40}$$

Notice that this number grows as $M \log [M]$. This is the second source for the increase of E with M. Since the terms in the above sum are generally much smaller than E_n, the contribution of I only leads to a slow increase.

6.10 Search-Effort Surface and Generalized Error Thresholds

We summarize our theoretical and experimental findings for the entire *search-effort surface* $E(q, M)$ of the average total number of fitness function evaluations in Fig. 6.5.

The figure shows the average total number $E(q, M)$ of fitness function evaluations for $N = 4$ blocks of length $K = 10$ bits; the same fitness function as used in Figs. 6.1(c), 6.3, 6.4(a), and 6.4(b). The top plot shows the theoretical predictions, which now include the innovation-time correction from Eq. (6.40); the bottom, the experimental estimates. The horizontal axis ranges from a population size of $M = 1$ ($M = 20$, experimental) to a population size of $M = 380$ with steps of $\Delta M = 1$ ($\Delta M = 30$, experimental). The vertical axis runs from a mutation rate of $q = 0.001$ to $q = 0.029$ with steps of $\Delta q = 0.00025$ in the theoretical plot and $\Delta q = 0.002$ in the experimental. The experimental search-effort surface is thus an interpolation between 195 data points on an equally spaced lattice of parameter settings. Each experimental data point is an average over 250 GA runs. The contours range from $E(q, M) = 0$ to $E(q, M) = 2 \times 10^6$ with each contour representing a multiple of 10^5. Note that the lowest values of E lie between 10^5 and 2×10^5. Lighter gray scale corresponds to smaller values of $E(q, M)$.

The first observations that can be made from Fig. 6.5 were already implied in the data from Fig. 6.3 and Fig. 6.4. First, the theory correctly predicts the relatively large region in parameter space where the GA searches most efficiently. Second, the theory predicts the location of the optimal parameter settings, indicated by the dot in the upper plot of Fig. 6.5 approximately correctly. The optimum occurs for somewhat higher population size in the experiments, as indicated by the dot in the lower plot of Fig. 6.5. Due to the large variance in E from run to run (recall Table 6.1) and the rather small differences in the experimental values of $\langle E \rangle$ near this regime, however, it is hard to infer from the experimental data exactly where the optimal population size occurs. Third, the theory underestimates the absolute magnitude of $E(q, M)$ somewhat. Fourth, at small mutation rates the theory underestimates the increase of $E(q, M)$ for decreasing q (moving down vertically in Fig. 6.5). Apart from this, though, the plots illustrate the general shape of the search-effort surface $E(q, M)$, and indicate that the theory accurately captures this shape.

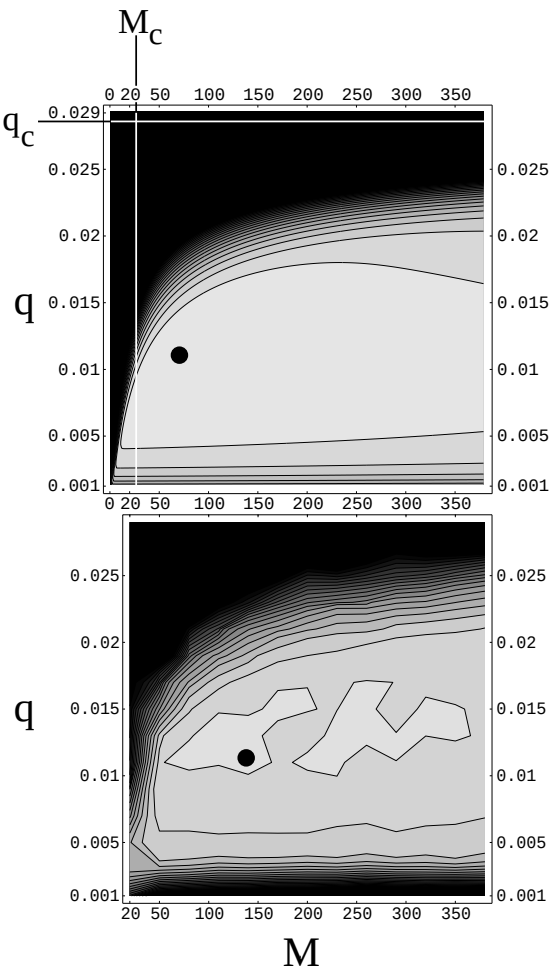

Figure 6.5: Contour plots of the search-effort surface $E(q, M)$ of the average total number of fitness function evaluations for the theory (upper), Eqs. (6.38) and (6.40), and for experimental estimates (lower). The parameter settings are $N = 4$ blocks of length $K = 10$ bits. The population size M runs from $M = 1$ to $M = 380$ on the horizontal axis on the upper plot and from $M = 20$ to $M = 380$ on the lower. The mutation rate runs from $q = 0.001$ to $q = 0.029$ on the vertical. The contours are plotted over the range $E(q, M) = 0$ to $E(q, M) = 2 \times 10^6$ with a contour at each multiple of 10^5. The experimental surface was interpolated from 195 equally spaced data points, 13 increments of $\Delta M = 30$ on the horizontal axis by 15 increments of $\Delta q = 0.002$ on the vertical. The theoretical surface was interpolated over a grid using $\Delta M = 1$ and $\Delta q = 0.00025$. The optimal theoretical parameter setting, $(q_o, M_o) = (0.011, 60)$, and the estimated optimal experimental parameter setting, $(q_o, M_o) = (0.011, 140)$, are marked in their respective plots with a dot.

The GA runs efficiently for a relatively large area of parameter space around the optimal setting (q_o, M_o). Moving away from this optimal setting horizontally (changing M) increases $E(q, M)$ only slowly at first. For decreasing M one reaches a "wall" relatively quickly around $M = 30$. For population sizes lower than $M = 30$, the higher epochs become so dynamically unstable that it is difficult for the population to reach the global optimum string *at all*. In contrast, moving in the opposite direction, increasing population size, $E(q, M)$ increases slowly over a relatively large range of M. Thus, choosing the population size too small is far more deleterious than setting it too large.

As one moves away from the optimal setting vertically (changing q) the increase of $E(q, M)$ is also slow at first. Eventually, as the plots make clear, increasing q one reaches the same "wall" as encountered in lowering M. This occurs at $q \approx 0.02$ in Fig. 6.5. For larger mutation rates the higher epochs become too unstable in this case as well, and the population is barely able to reach the global optimum.

The wall in (q, M)-space is the two-dimensional analogue of a phenomenon known as the *error threshold* in the theory of molecular evolution. As pointed out in Sec. 6.8, in our case error thresholds form the boundary between parameter regimes where epochs are stable and unstable. Here, the *error boundary* delimits a regime in parameter space where the optimum is discovered relatively quickly from a regime, in the black upper-left corners of the plots, where the population essentially never finds the optimum. For too high mutation rates or too low population sizes, selection is not strong enough to maintain high fitness strings—in our case those close to the global optimum—in the population against sampling fluctuations and deleterious mutations. Strings of fitness N will not stabilize in the population but will almost always be immediately lost, making the discovery of the global optimum string of fitness $N + 1$ extremely unlikely.

Note that the error boundary rolls over with increasing M in the upper-left corner of the plots. It bends all the way over to the right, eventually running horizontally, thereby determining a population-size *independent* error threshold. For our parameter settings this occurs around $q \approx 0.028$. Thus, beyond a critical mutation rate of $q_c \approx 0.028$ the population almost never discovers the global optimum, even for very large populations.

The value of this horizontal asymptote q_c can be roughly approximated by calculating for which mutation rate q_c epoch N has exactly the same average fitness as epoch $N - 1$; i.e. find q_c such that $f_N \approx f_{N-1}$. For those parameters, the population is under no selective pressure to move from epoch $N - 1$ to epoch N. Thus, strings of fitness N will generally not spread in the population. Using our analytic approximations, we find that the critical mutation rate q_c is simply given by:

$$q_c = 1 - \sqrt[\kappa]{\frac{N-1}{N}} \; . \tag{6.41}$$

For the parameters of Fig. 6.5 this gives $q_c = 0.0284$. This asymptote is indicated there by the horizontal line in the top plot.

Similarly, below a critical population size M_c, it is also practically impossible to reach the global optimum, even for low mutation rates. This M_c can also be roughly approximated by calculating the population size for which the sampling noise is equal to the fitness differential between the last two epochs. Formally, when $1/\sqrt{M}$ becomes

equal to γ_{N-1}. We then find:

$$M_c = \left(\frac{N-1}{N\lambda - N + 1} \right)^2 . \tag{6.42}$$

For the parameters of Fig. 6.5 this gives $M_c \approx 27$ around (the optimum) $q = 0.011$. This threshold estimate is indicated by the vertical line in Fig. 6.5.

Further, notice that for small mutation rates, at the bottom of each plot in Fig. 6.5, the contours run almost horizontally. That is, for small mutation rates relative to the optimum mutation rate q_o, the total number of fitness function evaluations $E(q, M)$ is insensitive to the population size M. Decreasing the mutation rate too far below the optimum rate increases $E(q, M)$ quite rapidly. According to our theoretical predictions it increases roughly as $1/q$ with decreasing q. The experimental data indicate that this is a slight underestimation. In fact, $E(q, M)$ appears to increase as $1/q^\alpha$ where the exponent α lies somewhere between 1 and 2.

Globally, the theoretical analysis and empirical evidence indicate that the search-effort surface $E(q, M)$ is everywhere concave. That is, for any two points (q_1, M_1) and (q_2, M_2), the straight line connecting these two points is everywhere above the surface $E(q, M)$. We believe that this is always the case for mutation-only genetic algorithms with a static fitness function that has a unique global optimum. This feature is useful in the sense that a steepest descent algorithm on the level of the GA *parameters* q and M will always lead to the unique optimum (q_o, M_o).

Finally, it is important to emphasize once more that there are large run-to-run fluctuations in the total number of fitness evaluations to reach the global optimum. (Recall Table 6.1.) Theoretically, each epoch has an exponentially distributed length since there is an equal and independent innovation probability of leaving it at each generation. The standard deviation of an exponential distribution is equal to its mean. Since the total time $E(q, M)$ is dominated by the last epochs, the total time $E(q, M)$ has a standard deviation close to its mean.

One conclusion from this is that, if one is only going to use a GA for a few runs on a specific problem, there is a large range in parameter space for which the GA's performance is statistically equivalent. In this sense, fluctuations lead to a large "sweet spot" of GA parameters. On the other hand, these large fluctuations reflect the fact that individual GA runs do not reliably discover the global optimum within a fixed number of fitness function evaluations. Notice that this is not a feature of the parameter settings or the analysis that we perform but a feature of the GA dynamics itself.

6.11 Conclusions

We derived explicit analytical approximations to the total number of fitness function evaluations that a GA takes on average to discover the global optimum as a function of both mutation rate and population size. The class of fitness functions so analyzed describes a general subbasin-portal architecture in genotype space. The GA's dynamics on this class of fitness functions consists of alternating periods of stasis (epochs) in the fitness distribution of the population, with short bursts of change (innovations) to higher

average fitness. During the epochs the most-fit individuals in the population diffuse over neutral networks of iso-fitness strings until a portal to a network of higher fitness is discovered. Then descendants of this higher-fitness string spread through the population.

The time to discover these portals depends both on the fraction of the population that is located on the highest neutral net in equilibrium and the speed at which these population members diffuse over the network. Although increasing the mutation rate increases the diffusion rate of individuals on the highest neutral network, it also increases the rate of deleterious mutations that cause these members to "fall off" the highest-fitness network. The mutation rate is optimized when these two effects are balanced so as to maximize the total amount of explored volume on the neutral network per generation. The optimal mutation rate, as given by Eq. (6.22), is dependent on the neutrality degree (the local branching rate) of the highest-fitness network and on the fitnesses of the lower lying neutral networks onto which the mutants are likely to fall.

With respect to optimizing population size, we found that the optimal population size occurs when the highest epochs are just barely stable. That is, given the optimal mutation rate, the population size should be tuned such that only a few individuals are located on the highest-fitness neutral network. The population size should be large enough such that it is relatively unlikely that all the individuals disappear through a deleterious fluctuation, but not much larger than that. In particular, if the population is much larger, so that many individuals are located on the highest-fitness network, then the sampling dynamics causes these individuals to correlate genetically. Due to this genetic correlation, individuals on the highest-fitness network do not independently explore the neutral network. This leads, in turn, to a deterioration of the search algorithm's efficiency. Therefore, the population size should be as low as possible without completely destabilizing the last epochs. Given this, one cannot help but wonder how general the association of efficient search and marginal stability is.

6.11.1 Genetic Algorithms versus Hill Climbers

It would appear that the GA wastes computational resources when maintaining a population quasispecies that contains many suboptimal fitness members; that is, those that are not likely to discover higher-fitness strings. This is precisely the reason that the GA performs so much more poorly than a simple hill climbing algorithm on this particular set of fitness functions, as shown in Ref. [105]. To be more specific, let's compare the GA at its optimal parameter settings with a Random Mutation Hill Climber, which performs a random bit flip at each time step and accepts this change when no lowering of fitness occurs. When the fitness of the string is n, mutations in blocks 1 through $n-1$ are always deleterious, and mutations in blocks $n+1$ through N are always neutral. Only 1 in every N mutations occurs in block n. Roughly 2^K mutations have to occur in a block before it is aligned for the first time. Thus, aligning block n takes the random mutation hill climber roughly $N2^K$ time steps on average, independent of n. In other words, the hill climber spends $N2^K$ time steps on average in each "epoch". Since N blocks have to be aligned in total, the random mutation hill climber takes roughly $N^2 2^K$ time steps to reach the global optimum.

In contrast, by numerically determining the optimal parameter settings from Eq. (6.38) we find that at its optimal parameter settings, the GA takes approximately $E \approx$

$2.2N^{3.1}2^K$ fitness function evaluations to reach the global optimum. That is, roughly a factor N more than the random mutation hill climber. This factor N is the result of the many suboptimal fitness individuals that are maintained in the population. The deleterious mutations together with the nature of the selection mechanism drive up the fraction of lower-fitness individuals in the quasispecies, and this fraction in the population plays no role in the search for higher-fitness strings.

If we allowed ourselves to tune the selection strength, we could have tuned selection so high that only the most-fit individuals would ever be selected. In this "infinite selection" limit, we would have that *only* strings of fitness n would be selected during epoch n. It is easy to calculate the optimal parameter settings for this regime. With $\lambda = (1-q)^K$, we have that the probability $P_{\rm disc}$ for a fitness n string to turn into a fitness $n+1$ string is $P_{\rm disc} \approx \lambda^{n-1}(1-\lambda)/2^K$. This probability is maximal for $\lambda = (1-1/n)$. Using this as optimal parameter settings, the average number of fitness function evaluations during epoch n becomes

$$E_n \approx n2^K \left(1 - \frac{1}{n}\right)^{1-n}. \tag{6.43}$$

For n not too small, the last factor is roughly equal to e. We can then sum over n from 1 to N and have for the total time $E \approx eN(N+1)2^K/2$. In short, in the limit of infinite selection strength, both hill climbing algorithms and the GA have a scaling of the total number of fitness function evaluations $E \approx N^2 2^K$ to leading order.

6.11.2 Coarse Graining the Fitness Function

For the fitness functions we analyzed, setting the selection strength infinitely high thus turns out to be the best strategy. Of course, this is largely the result of the fact that none of the fitness functions in our set has local optima. However, it is a common belief that fitness functions typical of combinatorial optimization problems possess many local optima. In general, tuning the selection strength infinitely high causes the population to become "pinned" on the tiniest of local peaks. Thus, this cannot be a good general strategy. With this in mind, let us step back from our detailed analysis and consider the broader implications of our results.

We believe that the results point to an interesting interplay between *neutrality*, *local optima*, and *marginal stability*—an interplay that is potentially quite general. Neutrality refers to the phenomenon that, for any genotype, there are always some single mutant neighbors that have the same fitness. When neutrality is present, a population does not become pinned to any particular point or island in genotype space, but instead has the possibility of diffusing through genotype space. In the Royal Staircase functions used here, this neutrality was explicitly built in from the start. Epochs corresponded to times during which the population diffused over the current highest-fitness network in search for a connection to higher-fitness networks.

We found that the GA searches most efficiently when population size and mutation rate are set such that these epochs are *marginally stable*. That is, the GA dynamics is as "stochastic" as possible without destabilizing the current and later epochs. Strings of the current highest fitness are (only slightly) preferentially reproduced over strings with

lower fitness, and the population size is just large enough to protect these highest-fitness strings from disappearing through deleterious sampling fluctuations. Thus, for fitness functions consisting of interwoven neutral networks, our analysis shows that *marginal stability* of the highest fitness strings corresponds to optimal search.

The role of marginal stability in more general cases, involving fitness functions with many local optima, can be best understood by perturbing away from our Royal Staircase class. Assume, for instance, that we add small fitness fluctuations to the fitnesses of each genotype. That is, a genotype that previously had fitness n, now receives fitness $n + \epsilon$ where ϵ is some small (random) number. These fluctuations are likely to induce many local optima in the fitness landscape. The random mutation hill climber will easily become pinned on this type of "landscape". However, the GA, in the regime of optimal parameter settings, will hardly be affected in its dynamics. The reason for this is that selection simply does not "notice" these small fitness differences. In the optimal parameter regime, strings of fitness n are barely distinguished from strings with fitness $n-1$, let alone from strings of fitness $n + \epsilon$. That is, fitness differentials between strings have to be above some minimal size to be "noticed" by the dynamics.

This intuitive idea, which is closely related to the *nearly neutral* theories of molecular evolution of Refs. [111, 112], can be made more precise. Let us assume that the current most-fit strings in the population have a fitness f and that strings with fitness f typically have around d *defining bits*—i.e., these bits are deleterious when changed. Under the assumption that mutations from low- to high-fitness strings are negligible, this leads to an average fitness of $\langle f \rangle = f(1-q)^d$ in the population. Assume that the population discovers on one or more strings of higher fitness $f + \delta f$, which have an *additional* b defining bits. If strings of fitness $f + \delta f$ were to stabilize, the average fitness would become $\langle f \rangle = (f + \delta f)(1-q)^{d+b}$. However, such strings can only stabilize when the relative difference between these two average fitnesses is larger than the finite-population sampling fluctuations; and these are of order $1/\sqrt{M}$. Thus, for the higher-fitness strings to be noticed by the dynamics, the condition

$$\delta f \geq \frac{f}{(1-q)^b} \left(\frac{1}{\sqrt{M}} + 1 - (1-q)^b \right) \tag{6.44}$$

must be met. Below this fitness differential, strings of fitness $f + \delta f$ are effectively neutral with respect to strings with fitness f. The net result is that the search parameters, such as q and M, determine a *coarse graining* of fitness levels, where strings in the band of fitness between f and $f + \delta f$ are treated as having equal fitness. This is important since it shows that even for fitness functions that do not have explicit neutrality, neutrality may still be effectively *induced* by the dynamics.

By tuning the evolutionary parameters q and M one effectively coarse grains the fitness "landscape", as if one were squinting at it. What we found in the preceding analysis is that for the Royal Staircase fitness functions, search was optimal when the staircase fitness steps were just discernible. It seems intuitive then that fitness functions containing many local optima, but that by definition have no neutrality, might still be searched efficiently by a genetic algorithm provided that there exists a level of coarse graining which turns the "rugged fitness landscape" into an architecture of interwoven neutral networks. These would be fitness functions that, at the smallest scales, possess

many local optima, but that at some well chosen scale their local optima are hidden by the induced coarse graining. To put it somewhat differently, a fitness function may be efficiently searched by a GA if there is some scale of coarse graining at which higher-fitness strings "point the way" to strings of even higher fitness.

With these observations in mind, our analysis suggests that the question of what fitness functions can be efficiently searched by evolutionary methods translates into the question of what fitness functions can be coarse grained so as to turn into neutral network architectures.

Acknowledgments

The authors thank the referees for helpful suggestions. This work was supported at the Santa Fe Institute by NSF Grant IRI-9705830, ONR grant N00014-95-1-0524, and Sandia National Laboratory contract AU-4978.

7

Neutral Evolution of Mutational Robustness

Erik van Nimwegen, James P. Crutchfield, and Martijn Huynen
Proceedings of the National Academy of Science U.S.A.
96 (1999) 9716-9720.
Santa Fe Institute Working Paper 99-03-021.

We introduce and analyze a general model of a population evolving over a network of selectively neutral genotypes. We show that the population's limit distribution on the neutral network is solely determined by the network topology and given by the principal eigenvector of the network's adjacency matrix. Moreover, the average number of neutral mutant neighbors per individual is given by the matrix spectral radius. This quantifies the extent to which populations evolve mutational robustness—the insensitivity of the phenotype to mutations—and so reduce genetic load. Since the average neutrality is independent of evolutionary parameters—such as, mutation rate, population size, and selective advantage—one can infer global statistics of neutral network topology using simple population data available from *in vitro* or *in vivo* evolution. Populations evolving on neutral networks of RNA secondary structures show excellent agreement with our theoretical predictions.

7.1 Introduction

Kimura's contention that a majority of genotypic change in evolution is selectively neutral [91] has gained renewed attention with the recent analysis of evolutionary optimization methods [22, 137] and the discovery of *neutral networks* in genotype-phenotype models for RNA secondary structure [48, 65, 126] and protein structure [3, 9]. It was found that collections of mutually neutral genotypes, which are connected via single mutational steps, form extended networks that permeate large regions of genotype space. Intuitively, a large degeneracy in genotype-phenotype maps, when combined with the high connectivity of (high-dimensional) genotype spaces, readily leads to such extended neutral networks. This intuition is now supported by recent theoretical results [6, 56, 122, 123].

In *in vitro* evolution of ribozymes, mutations responsible for an increase in fitness are only a small minority of the total number of accepted mutations [147]. This indicates that, even in adaptive evolution, the majority of point mutations is neutral. The fact that only a minority of loci is conserved in sequences evolved from a single ancestor similarly

indicates a high degeneracy in ribozymal genotype-phenotype maps [96]. Neutrality is also implicated in experiments where RNA sequences evolve a given structure starting from a range of different initial genotypes [36]. More generally, neutrality in RNA and protein genotype-phenotype maps is indicated by the observation that their structures are much better conserved during evolution than their sequences [66, 78].

Given the presence of neutral networks that preserve structure or function in sequence space, one asks, How does an evolving population distribute itself over a neutral network? Can we detect and analyze structural properties of neutral networks from data on biological or *in vitro* populations? To what extent does a population evolve toward highly connected parts of the network, resulting in sequences that are relatively insensitive to mutations? Such *mutational robustness* has been observed in biological RNA structures [82, 140] and in simulations of the evolution of RNA secondary structure [81]. However, an analytical understanding of the phenomenon, the underlying mechanisms, and their dependence on evolutionary parameters—such as, mutation rate, population size, selection advantage, and the topology of the neutral network—has up to now not been available.

Here we develop a dynamical model for the evolution of populations on neutral networks and show analytically that, for biologically relevant population sizes and mutation rates, a population's distribution over a neutral network is determined solely by the network's topology. Consequently, one can infer important structural information about neutral networks from data on evolving populations, even without specific knowledge of the evolutionary parameters. Simulations of the evolution of a population of RNA sequences, evolving on a neutral network defined with respect to secondary structure, confirm our theoretical predictions and illustrate their application to inferring network topology.

7.2 Modeling Neutrality

We assume that genotype space contains a neutral network of high, but equal fitness genotypes on which the majority of a population is concentrated and that the neighboring parts of genotype space consist of genotypes with markedly lower fitness. The genotype space consists of all sequences of length L over a finite alphabet $\mathcal{A}$ of A symbols. The neutral network on which the population moves can be most naturally regarded as a graph G embedded in this genotype space. The vertex set of G consists of all genotypes that are on the neutral network; denote its size by $|G|$. Two vertices are connected by an edge if and only if they differ by a single point mutation.

We will investigate the dynamics of a population evolving on this neutral network and analyze the dependence of several population statistics on the topology of the graph G. With these results, we will then show how measuring various population statistics enables one to infer G's structural properties.

For the evolutionary process, we assume a discrete-generation selection-mutation dynamics with constant population size M. Individuals on the neutral network G have a fitness σ. Individuals outside the neutral network have fitnesses that are considerably smaller than σ. Under the approximations we use, the exact fitness values for genotypes off G turn out to be immaterial. Each generation, M individuals are selected with re-

placement and with probability proportional to fitness and then mutated with probability μ. These individuals form the next generation.

This dynamical system is a discrete-time version of Eigen's molecular evolution model [32]. Our analysis can be directly translated to the continuous-time equations for the Eigen model. The results remain essentially unchanged.

Although our analysis can be extended to more complicated mutation schemes, we will assume that only single point mutations can occur at each reproduction of an individual. With probability μ one of the L symbols is chosen with uniform probability and is mutated to one of the $A-1$ other symbols. Thus, under a mutation, a genotype s moves with uniform probability to one of the $L(A-1)$ neighboring points in genotype space.

For the results presented below to hold, it is not necessary that all genotypes in G have exactly the same fitness. As in any model of neutral evolution [59, 91], it is sufficient to assume that the fitness differentials between distinct genotypes in G are smaller than the reciprocal $1/M$ of the population size. Additionally, we assume that the fitness differentials between genotypes in G and genotypes outside G are much larger than $1/M$. These assumptions break down when there is a continuum of fitness differentials between genotypes or in the case of very small population size, which readily allows the spreading of mildly deleterious mutations [106].

7.2.1 Infinite-Population Solution

The first step is to solve for the asymptotic distribution of the population over the neutral network G in the limit of very large population size.

Once the (infinite) population has come to equilibrium, there will be a constant proportion P of the population located on the network G and a constant average fitness $\langle f \rangle$ in the population. Under selection the proportion of individuals on the neutral network increases from P to $\sigma P/\langle f \rangle$. Under mutation a proportion $\langle \nu \rangle$ of these individuals remains on the network, while a proportion $1-\langle \nu \rangle$ falls off the neutral network to lower fitness. At the same time, a proportion Q of individuals located outside G mutate *onto* the network so that an equal proportion P ends up on G in the next generation. Thus, at equilibrium, we have a balance equation:

$$P = \frac{\sigma}{\langle f \rangle} \langle \nu \rangle P + Q \; . \tag{7.1}$$

In general, the contribution of Q is negligible. As mentioned above, we assume that the fitness σ of the network genotypes is substantially larger than the fitnesses of those off the neutral network and that the mutation rate is small enough so that the bulk of the population is located on the neutral network. Moreover, since their fitnesses are smaller than the average fitness $\langle f \rangle$, only a fraction of the individuals outside the network G produces offspring for the next generation. Of this fraction, only a small fraction mutates *onto* the neutral network G. Therefore, we neglect the term Q in Eq. (7.1) and obtain:

$$\frac{\sigma}{\langle f \rangle} \langle \nu \rangle = 1 \; . \tag{7.2}$$

This expresses the balance between selection expanding the population on the network and deleterious mutations reducing it by moving individuals off.

Equation (7.2) can also be phrased in terms of the *genetic load* $\mathcal{L}$, defined as the relative distance of the average fitness from the optimum fitness in the population: $\mathcal{L} = (\sigma - \langle f \rangle)/\sigma$. $\mathcal{L}$ measures the selection pressure that the population is experiencing. According to Eq. (7.2), in the presence of neutrality $\mathcal{L}$ is simply equal to the proportion $1 - \langle \nu \rangle$ of offspring that falls off the network G. Thus, Eq. (7.2) states that $\mathcal{L}$ is equal to the proportion of deleterious mutations per generation, in accordance with Haldane's original result [67].

Under mutation an individual located at genotype s of G with vertex degree d_s (the number of neutral mutant neighbors) has probability

$$\nu_s = 1 - \mu \left(1 - \frac{d_s}{(A-1)L} \right) \tag{7.3}$$

to remain on the neutral network G. If asymptotically a fraction P_s of the population is located at genotype s, then $\langle \nu \rangle$ is simply the average of ν_s over the asymptotic distribution on the network: $\langle \nu \rangle = \sum_{s \in G} \nu_s P_s / P$. As Eq. (7.3) shows, the average $\langle \nu \rangle$ is simply related to the *population neutrality* $\langle d \rangle = \sum_{s \in G} d_s P_s / P$. Moreover, using Eq. (7.2) we can directly relate the population neutrality $\langle d \rangle$ to the average fitness $\langle f \rangle$:

$$\langle d \rangle = L(A-1) \left[1 - \frac{\sigma - \langle f \rangle}{\mu \sigma} \right]. \tag{7.4}$$

Despite our not specifying the details of G's topology, nor giving the fitness values of the genotypes lying off the neutral network, one can relate the population neutrality $\langle d \rangle$ of the individuals on the neutral network directly to the average fitness $\langle f \rangle$ in the population. It may seem surprising that this is possible at all. Since the population consists partly of sequences off the neutral network, one expects that the average fitness is determined in part by the fitnesses of these sequences. However, under the assumption that *back mutations* from low-fitness genotypes off the neutral network onto G are negligible, the fitnesses of sequences outside G only influence the total proportion P of individuals on the network, but not the average fitness in the population.

Equation (7.4) shows that the population neutrality $\langle d \rangle$ can be inferred from the average fitness and other parameters—such as, mutation rate. However, as we will now show, the population neutrality $\langle d \rangle$ can also be obtained independently, from knowledge of the topology of G alone.

The asymptotic equilibrium proportions $\{P_s\}$ of the population at network nodes s are the solutions of the simultaneous equations:

$$P_s = (1-\mu) \frac{\sigma}{\langle f \rangle} P_s + \frac{\mu}{L(A-1)} \sum_{t \in [s]_G} \frac{\sigma}{\langle f \rangle} P_t, \tag{7.5}$$

where $[s]_G$ is the set of neighbors of s that are also on the network G. Using Eq. (7.4), Eq. (7.5) can be rewritten as an eigenvector equation:

$$\langle d \rangle P_s = \left(\mathbf{G} \cdot \vec{P} \right)_s, \tag{7.6}$$

where $\mathbf{G}$ is the adjacency matrix of the graph G:

$$\mathbf{G}_{st} = \begin{cases} 1 & t \in [s]_G, \\ 0 & \text{otherwise.} \end{cases} \tag{7.7}$$

Since $\mathbf{G}$ is nonnegative and the neutral network G is connected, the adjacency matrix is irreducible. Therefore, the theorems of Frobenius-Perron for nonnegative irreducible matrices apply [53]. These imply that the proportions P_s of the limit distribution on the network are given by the principal eigenvector of the graph adjacency matrix $\mathbf{G}$. Moreover, the population neutrality is equal to $\mathbf{G}$'s spectral radius ρ: $\langle d \rangle = \rho$. In this way, one concludes that asymptotically the population neutrality $\langle d \rangle$ is independent of evolutionary parameters (μ, L, σ) and of the fitness values of the genotypes outside the neutral network. It is a function *only* of the neutral network topology as determined by the adjacency matrix $\mathbf{G}$.

In genetic load terminology, our results imply that

$$\mathcal{L} = \mu \left(1 - \frac{\rho}{L(A-1)} \right) . \tag{7.8}$$

We see that Haldane's result ($\mathcal{L} = \mu$) [67] is recovered in the absence of neutrality ($\rho = 0$). In the presence of neutrality, the genetic load is reduced by a factor that only depends on the spectral radius ρ of the network's adjacency matrix.

The fortunate circumstance that the population neutrality depends only on G's topology, allows us to consider several practical consequences. Note that knowledge of μ, σ, and $\langle f \rangle$ allows one to infer a dominant feature of G's topology, namely, the spectral radius ρ of its adjacency matrix. In *in vitro* evolution experiments in which biomolecules are evolved (say) to bind a particular ligand [135], by measuring the proportion $\langle \nu \rangle$ of functional molecules that remain functional after mutation, we can now infer the spectral radius ρ of their neutral network. In other situations, such as in the bacterial evolution experiments of Ref. [37], it might be more natural to measure the average fitness $\langle f \rangle$ of an evolving population and then use Eq. (7.4) to infer the population neutrality $\langle d \rangle$ of viable genotypes in sequence space.

7.2.2 Blind and Myopic Random Neutral Walks

In the foregoing we solved for the asymptotic average neutrality $\langle d \rangle$ of an infinite population under selection and mutation dynamics and showed that it was uniquely determined by the topology of the neutral network G. To put this result in perspective, we now compare the population neutrality $\langle d \rangle$ with the effective neutralities observed under two different kinds of random walk over G. The results illustrate informative extremes of how network topology determines the population dynamics on neutral networks and affects the evolution of robustness.

The first kind of random walk that we consider is generally referred to as a *blind ant* random walk. An ant starts out on a randomly chosen node of G. Each time step it chooses one of its $L(A-1)$ neighbors at random. If the chosen neighbor is on G, the ant steps to this node, otherwise it remains at the current node for another time step. It is easy to show that this random walk asymptotically spends equal amounts of time at all of G's nodes [75]. Therefore, the *network neutrality* $\bar{d}$ of the nodes visited under this type of random walk is simply given by:

$$\bar{d} = \sum_{s \in G} \frac{d_s}{|G|} . \tag{7.9}$$

It is instructive to compare this with the effective neutrality observed under another random walk, called the *myopic ant*. An ant again starts at a random node $s \in G$. Each time step, the ant determines the set $[s]_G$ of network neighbors of s and then steps to one at random. Under this random walk, the asymptotic proportion P_s of time spent at node s is proportional to the node degree d_s [75]. It turns out that the *myopic neutrality* $\widehat{d}$ seen by this ant can be expressed in terms of the mean $\bar{d}$ and variance $\mathrm{Var}(d)$ of node degrees over G:

$$\widehat{d} = \bar{d} + \frac{\mathrm{Var}(d)}{\bar{d}} \,. \tag{7.10}$$

The network and myopic neutralities, $\bar{d}$ and $\widehat{d}$, are thus directly given in terms of *local* statistics of the distribution of vertex degrees, while the population neutrality $\langle d \rangle$ is given by ρ, the spectral radius of G's adjacency matrix. The latter is an essentially *global* property of G.

7.3 Mutational Robustness

With these cases in mind, we now consider how different network topologies are reflected by these neutralities. In prototype models of populations evolving on neutral networks, the networks are often assumed to be or are approximated as regular graphs [51, 121, 123, 137] and chapter 4 of this thesis. If the graph G is, in fact, regular, each node has the same degree d and, obviously, one has $\langle d \rangle = \bar{d} = \widehat{d} = d$.

In more realistic neutral networks, one expects G's neutralities to vary over the network. When this occurs, the population neutrality is typically larger than the network neutrality: $\langle d \rangle = \rho > \bar{d}$. Their difference quantifies precisely the extent to which a population seeks out the most connected areas of the neutral network. Thus, a population will evolve a *mutational robustness* that is larger than if the population were to spread uniformly over the neutral network. Additionally, the mutational robustness tends to increase during the transient phase in which the population relaxes toward its asymptotic distribution.

Assume, for instance, that initially the population is located entirely off the neutral network G at lower fitness sequences. At some time, a genotype $s \in G$ is discovered by the population. To a rough approximation, one can assume that the probability of genotype s being discovered first is proportional to the number of neighbors, $L(A-1) - d_s$, that s has *outside* the neutral network. Therefore, the population neutrality $\langle d_0 \rangle$ when the population first enters the neutral network G is approximately given by:

$$\langle d_0 \rangle = \bar{d} - \frac{\mathrm{Var}(d)}{L(A-1) - \bar{d}} \,. \tag{7.11}$$

We define the *excess robustness* r to be the relative increase in neutrality between this initial neutrality and the (asymptotic) population neutrality:

$$r \equiv \frac{\langle d \rangle - \langle d_0 \rangle}{\langle d_0 \rangle} \,. \tag{7.12}$$

For networks that are sparse, i.e. $\bar{d} \ll L(A-1)$, this is well approximated by $r \approx (\langle d \rangle - \bar{d})/\bar{d}$. Note that, while r is defined in terms of population statistics, the preceding

results have shown that this robustness is only a function of G's topology and should thus be considered a property of the network.

7.4 Finite-Population Effects

Our analysis of the population distribution on the neutral network G assumed an infinite population. For finite populations, it is well known that sampling fluctuations converge a population and this raises a question: To what extent does the asymptotic distribution P_s still describe the distribution over the network for small populations? As a finite population diffuses over a neutral network [79], one might hope that the time average of the distribution over G is still given by P_s. Indeed, the simulation results shown below indicate that for moderately large population sizes, this is the case. However, a simple argument shows that this cannot be true for arbitrarily small populations.

Assume that the population size M is so small that the product of mutation rate and population size is much smaller than 1; i.e. $M\mu \ll 1$. In this limit the population, at any point in time, will be completely converged onto a single genotype s on the neutral network G. With probability $M\mu$ a single mutant will be produced at each generation. Such a mutant is equally likely to be one of the $L(A-1)$ neighbors of s. If this mutant is not on G, it will quickly disappear due to selection. However, if the mutant is on the neutral network, there is a probability $1/M$ that it will take over the population. When this happens, the population will effectively have taken a random-walk step on the network, of exactly the kind followed by the blind ant. Therefore, for $M\mu \ll 1$, the population neutrality will be equal to the network neutrality: $\langle d \rangle = \bar{d}$. In this regime, $r \approx 0$ and excess mutational robustness will not emerge through evolution.

The extent to which the population neutrality $\langle d \rangle$ approaches its infinite population value ρ is determined by the extent to which it is not converged by sampling fluctuations. In neutral evolution, population convergence is generally only a function of the product $M\mu$ [31, 90, 148]. Thus, as the product $M\mu$ ranges from values much smaller than 1 to values much larger than 1, we predict that the population neutrality $\langle d \rangle$ shifts from the network neutrality $\bar{d}$ to the infinite-population neutrality, given by $\mathbf{G}$'s spectral radius ρ.

7.5 RNA Evolution on Structurally Neutral Networks

The evolution of RNA molecules in a simulated flow reactor provides an excellent arena in which to test the theoretical predictions of evolved mutational robustness. The replication rates (fitnesses) were chosen to be a function only of the secondary structures of the RNA molecules. The secondary structure of RNA is an essential aspect of its phenotype, as documented by its conservation in evolution [66] and the convergent *in vitro* evolution toward a similar secondary structure when selecting for a specific function [36]. RNA secondary structure prediction based on free energy minimization is a standard tool in experimental biology and has been shown to be reliable, especially when the minimum free energy structure is thermodynamically well defined [80]. RNA secondary structures were determined with the Vienna Package [71], which uses the free energies from Ref. [141]. Free energies of dangling ends were set to 0.

The neutral network G on which the population evolves consists of all RNA molecules of length $L = 18$ that fold into a particular *target structure*. A target structure (Fig. 7.1) was selected that contains sufficient variation in its neutrality to test the theory, yet is not so large as to preclude an exhaustive analysis of its neutral network topology.

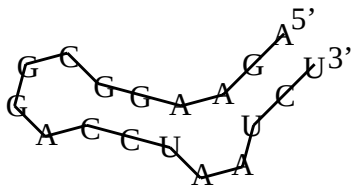

Figure 7.1: The target RNA secondary structure.

Using only single point mutations per replication, purine-pyrimidine base pairs {G-C, G-U, A-U} can mutate into each other, but not into pyrimidine-purine {C-G, U-G, U-A} base pairs. The target structure contains 6 base pairs which can each be taken from one or the other of these two sets. Thus, the approximately 2×10^8 sequences that are consistent with the target's base pairs separate into $2^6 = 64$ disjoint sets. Of these, we analyzed the set in which all base pairs were of the purine-pyrimidine type and found that it contains two neutral networks of $51,028$ and $5,169$ sequences that fold into the target structure. Simulations were performed on the largest of the two. The exhaustive enumeration of this network showed that it has a network neutrality of $\bar{d} = 12.0$ with standard deviation $\sqrt{Var(d)} \approx 3.4$, a maximum neutrality of $d_s = 24$, and a minimum of $d_s = 1$. The spectral radius of the network's 51028×51028 adjacency matrix is $\rho \approx 15.7$. The theory predicts that, when $M\mu \gg 1$, the population neutrality should converge to this value.

The simulated flow reactor contains a population of replicating and mutating RNA sequences [32, 46]. The replication rate of a molecule depends on whether its calculated minimum free energy structure equals that of the target: Sequences that fold into the target structure replicate on average once per time unit, while all other sequences replicate once per 10^4 time units on average. During replication the progeny of a sequence has probability μ of a single point mutation. Selection pressure in the flow reactor is induced by an adaptive dilution flow that keeps the total RNA population fluctuating around a constant capacity M.

Evolution was seeded from various starting sequences with either a relatively high or a relatively low neutrality. Independent of the starting point, the population neutrality converged to the predicted value, as shown in Fig. 7.2.

Subsequently, we tested the finite-population effects on the population's average neutrality at several different mutation rates. Figure 7.3 shows the dependence of the asymptotic average population neutrality on population size M and mutation rate μ. As expected, the population neutrality depends only on the product $M\mu$ of population size and mutation rate. For small $M\mu$ the population neutrality increases with increasing $M\mu$, until $M\mu \approx 500$ where it saturates at the predicted value of $\langle d \rangle \approx 15.7$. Since small populations do not form a stationary distribution over the neutral net, but diffuse over it [79], the average population neutrality at each generation may fluctuate consid-

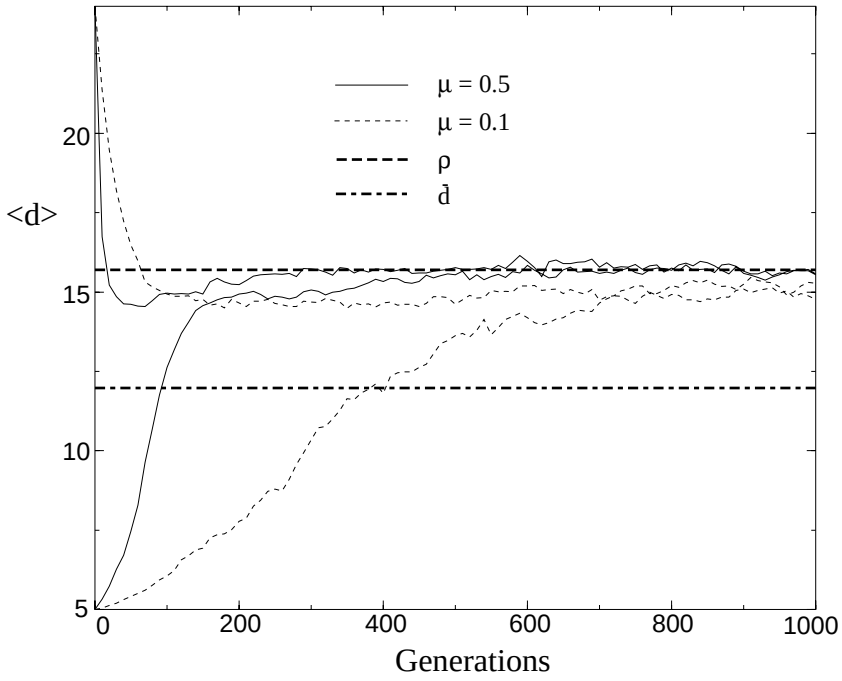

Figure 7.2: The evolution of RNA mutational robustness: convergence of the population's average neutrality to the theoretical value, $\langle d \rangle = \rho \approx 15.7$, predicted by $\mathbf{G}$'s spectral radius (upper dashed line). The network's average neutrality $\bar{d}$ is the lower dashed line. Simulations used a population size of $M = 10^4$ and mutation rates of $\mu = 0.5$ and $\mu = 0.1$ per sequence. They were started at sequences with either a relatively large number of neutral neighbors ($d_s = 24$) (upper curves for each mutation rate) or a small number ($d_s = 5$) (lower curves).

erably for small populations. Theoretically, sampling fluctuations in the proportions of individuals at different nodes of the network scale inversely with the square root of the population size. We therefore expect the fluctuations in population neutrality to scale as the inverse square root of the population size as well. This was indeed observed in our simulations.

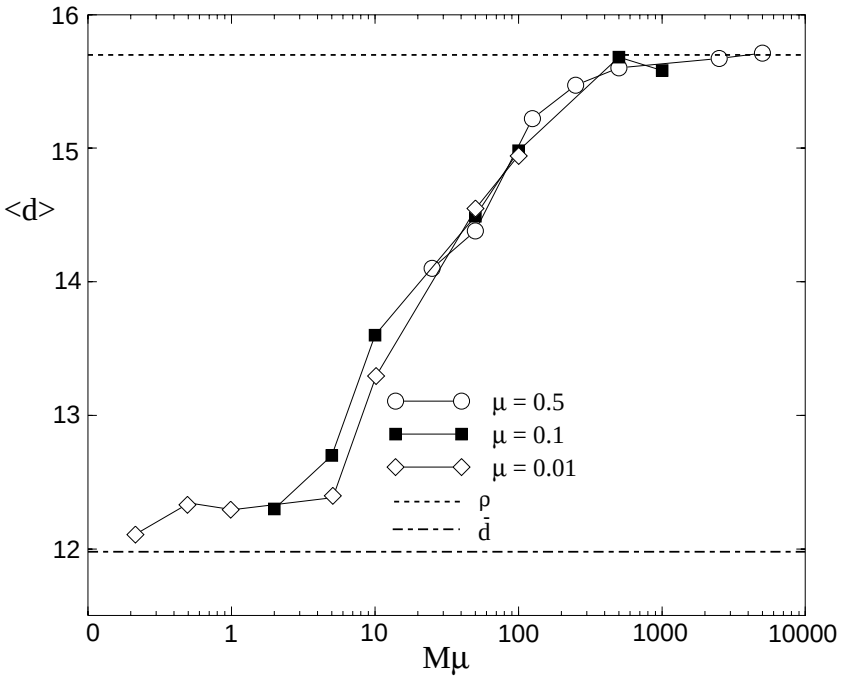

Figure 7.3: Dependence of the population neutrality on mutation rate μ and population size M. Simulations used three mutation rates, $\mu \in \{0.5, 0.1, 0.01\}$, and a range of population sizes, $M \in \{10000, 5000, 1000, 500, 250, 100, 50, 20\}$. The results show that the evolved neutrality in the population depends on the product $M\mu$ of population size and mutation rate. Neutrality increases with increasing $M\mu$ and saturates when $M\mu > 500$. When $M\mu < 1$ population neutrality approximates G's average neutrality $\bar{d} \approx 12.0$. When $M\mu > 500$ population neutrality converges to the spectral radius of the network's adjacency matrix, $\rho \approx 15.7$.

Finally, the fact that $r \approx 0.31$ for this neutral network shows that under selection and mutation, a population will evolve a mutational robustness that is 31 percent higher than if it were to spread randomly over the network.

7.6 Conclusions

We have shown that, under neutral evolution, a population does not move over a neutral network in an entirely random fashion, but tends to concentrate at highly connected

parts of the network, resulting in phenotypes that are relatively robust against mutations. Moreover, the average number of point mutations that leave the phenotype unaltered is given by the spectral radius of the neutral network's adjacency matrix. Thus, our theory provides an analytical foundation for the intuitive notion that evolution selects genotypes that are mutationally robust and reduce genetic load.

Perhaps surprisingly, the tendency to evolve toward highly connected parts of the network is independent of evolutionary parameters—such as, mutation rate, selection advantage, and population size (as long as $M\mu \gg 1$)—and is solely determined by the network's topology. One consequence is that one can infer properties of the neutral network's topology from simple population statistics.

Simulations with neutral networks of RNA secondary structures confirm the theoretical results and show that even for moderate population sizes, the population neutrality converges to the infinite-population prediction. Typical sizes of *in vitro* populations are such that the data obtained from experiments are expected to accord with the infinite-population results derived here. It seems possible then to devise *in vitro* experiments that, using the results outlined above, would allow one to obtain information about the topological structure of neutral networks of biomolecules with similar functionality.

We will present the extension of our theory to cases with multiple-mutation events per reproduction elsewhere. We will also report on analytical results for a variety of network topologies that we have studied.

Finally, here we focused only on the asymptotic distribution of the population on the neutral network. But how did the population attain this equilibrium? The transient relaxation dynamics, such as that shown in Fig. 7.2, can be analyzed in terms of the non-principal eigenvectors and eigenvalues of the adjacency matrix $\mathbf{G}$. Since the topology of a graph is almost entirely determined by the eigensystem of its adjacency matrix, one should in principle be able to infer the complete structure of the neutral network from accurate measurements of the transient population dynamics.

Acknowledgments We thank the participants of the Santa Fe Institute workshop on Evolutionary Dynamics for stimulating this work, which was supported in part at SFI by NSF under grant IRI-9705830, Sandia National Laboratory, and the Keck Foundation. M.H. gratefully acknowledges support from a fellowship of the Royal Netherlands Academy of Arts and Sciences. We thank Sergey Gavrilets in particular for pointing us to the connections of our results with genetic load.

8

Metastable Evolutionary Dynamics: Crossing Fitness Barriers or Escaping via Neutral Paths?

Erik van Nimwegen and James P. Crutchfield
submitted to **Bulletin of Mathematical Biology**,
Santa Fe Institute Working Paper 99-07-041.

We analytically study the dynamics of evolving populations that exhibit metastability on the level of phenotype or fitness. In constant selective environments, such metastable behavior is caused by two qualitatively different mechanisms. One the one hand, populations may become pinned at a local fitness optimum, being separated from higher-fitness genotypes by a *fitness barrier* of low-fitness genotypes. On the other hand, the population may only be metastable on the level of phenotype or fitness while, at the same time, diffusing over *neutral networks* of selectively neutral genotypes. Metastability occurs in this case because the population is separated from higher-fitness genotypes by an *entropy barrier*: The population must explore large portions of these neutral networks before it discovers a rare connection to fitter phenotypes.

We derive analytical expressions for the barrier crossing times in both the fitness barrier and entropy barrier regime. In contrast with "landscape" evolutionary models, we show that the waiting times to reach higher fitness depend strongly on the width of a fitness barrier and much less on its height. The analysis further shows that crossing entropy barriers is faster by orders of magnitude than fitness barrier crossing. Thus, when populations are trapped in a metastable phenotypic state, they are most likely to escape by crossing an entropy barrier, along a neutral path in genotype space. If no such escape route along a neutral path exists, a population is most likely to cross a fitness barrier where the barrier is *narrowest*, rather than where the barrier is shallowest.

8.1 Introduction

For populations evolving under selection, mutation, and a static fitness function, there are two main mechanisms thought to be responsible for the occurrence of dynamical

metastability—a behavior commonly observed in natural and artificial evolutionary processes [1, 7, 22, 37, 47, 108] and called *punctuated equilibria* in paleobiology [64]. First, a population may become trapped around a local optimum in the fitness "landscape" until a rare mutant crosses a *fitness barrier* to a higher nearby peak. Second, more recently it has been proposed [47, 79, 137] that populations may evolve neutrally, drifting randomly over *neutral networks* of isofitness genotypes in genotype space, until a rare single-point mutant connection is found to another neutral network of higher fitness. In this case, the population must cross an *entropy barrier* by visiting a large volume of the neutral network before it discovers a path to higher fitness.

To understand the relative roles of these two mechanisms in evolutionary metastability, in the following we study the dynamics of a population evolving under simple fitness functions that contain a single fitness barrier of tunable height and width. In order for the population to escape its current metastable state and so reach higher fitness, it must create a genotype that is separated from the current fittest genotypes in the population by a *valley* of lower-fitness genotypes. The *height* of the fitness barrier measures the relative selective difference between the current fittest genotypes and the lower-fitness genotypes in the intervening valley. Its *width* denotes the number of point mutations the current fittest genotypes must undergo to cross the valley of low fitness genotypes. We derive explicit analytical predictions for the barrier crossing times as a function of population size, mutation rate, and barrier height and width. The scaling of the fitness-barrier crossing time as a function of these parameters shows that the waiting time to reach higher fitness depends crucially on the width of the barrier and much less on the barrier height.

This contrasts with the scaling of the barrier crossing time for a particle diffusing in a double-well potential—a model proposed previously for populations crossing a fitness barrier [95, 107]. For such stochastic processes, it is well known that the waiting time scales exponentially with the barrier height [54]. In the population dynamics that we analyze here, we find that the waiting time scales approximately exponential with barrier width and only as a power law of the logarithm of barrier height. In addition, the waiting time scales roughly as a power law in both population size and mutation rate.

When the barrier height is lowered below a critical height, the fitness barrier turns into an entropy barrier. We show that, in general, neutral evolution via crossing entropy barriers is faster by orders of magnitude than fitness barrier crossing. Additionally, we show that the waiting time for crossing entropy barriers exhibits anomalous scaling with population size and mutation rate.

Finally, we extend our analysis to a class of more complicated fitness functions that contain a network of tunable fitness and entropy barriers. We show that the theory still accurately predicts fitness- and entropy-barrier crossing times in these more complicated cases.

The general conclusion drawn from our analysis is that, when populations are trapped in a metastable phenotypic state, they are most likely to escape this metastability by crossing an entropy barrier. That is, the escape to a new phenotype occurs along a neutral path in genotype space. If no such neutral path exists, then the population is most likely to cross a fitness barrier at the place where the barrier is *narrowest*.

8.1.1 Evolutionary Pathways and Metastability

The notion of an *adaptive landscape*, first introduced by Wright [149], has had a large impact on our appreciation of the mechanisms that control how populations evolve in static environments. The intuitive idea is that a population moves up the slopes of its fitness "landscape" just as a physical system moves down the slope of its potential-energy surface. Once this analogy has been accepted, it is natural to borrow many of the qualitative results on the dynamics of physical systems to account for the dynamics of evolving populations. For instance, it has become common to assume that an evolving population can be modeled by a single uphill walker in a "rugged" fitness landscape [88, 100].

There are, however, seemingly different kinds of evolutionary behavior than incremental adaptation via "landscape" crawling. For example, metastability or punctuated equilibrium of phenotypic traits in an evolving population appears to be a common occurrence in biological evolution [37, 64] as well as in models of natural and artificial evolution [1, 22, 47]. As just pointed out, for simple cases where populations evolve in a relatively constant environment, there are two main mechanisms that have been proposed to account for this type of metastable behavior.

The first and most commonly accepted explanation was already implicit in Wright's *shifting balance* theory [150]. A population moves up the slope of its fitness "landscape" until it reaches a local optimum, where it stabilizes. The population is pictured as a cloud in genotype space focused around this local optimum. The population remains in this state until a rare sequence of mutants crosses a *valley* of low fitness towards a higher fitness peak. In this view, metastability is the result of *fitness barriers* that separate local optima in genotype space.

This mechanism for metastability is very reminiscent of that found in physical systems. Metastability occurs there because local energy minima in state space are separated by potential energy barriers, which impede the immediate transition between the minima. A physical system generally moves through its state space along trajectories that lower its energy. Once it reaches a local minimum it tends to stay there. However, when such a system is subject to thermal fluctuations, through a sequence of chance events it can eventually be pushed over a barrier that separates the current local minimum from another. When this transition occurs, it turns out that the system moves quickly to the new local minimum.

Mathematically, barrier crossing processes in physical systems are most often described as diffusion in a potential field, where the potential represents the energy "landscape". These processes have been extensively studied and the basic quantitative results are widely known [52, 54, 136]. For example, barrier crossing times increase exponentially with the *height* of the barrier and inverse exponentially with the fluctuation amplitude, as measured by temperature.

In light of the physical metaphor for evolving populations, it is not surprising that the dynamics of populations crossing fitness barriers has been modeled using a class of diffusion equations analogous to those used to describe thermally driven systems in a potential [95, 107]. In this approach, the dynamics of the average fitness of the population is modeled as diffusion over the "fitness landscape", thermal fluctuations are replaced by random genetic mutations and drift, and the population size, which controls

sampling stochasticity, plays the role of inverse temperature. As a direct consequence, it was found that fitness-barrier crossing times scale exponentially with population size in these models. Note that it is assumed in this approach that the population *as a whole* must cross the fitness barrier.

In the following, we show that the analogy with the physical situation and, in particular, the translation of results from there are misleading for the understanding of the evolutionary dynamics. For example, a direct analysis of the population dynamics reveals that for most parameter ranges, the time to cross a fitness barrier scales very differently for populations evolving under selection and mutation. For example, the waiting time is determined by how long it takes to generate a *rare sequence of mutants* that crosses the fitness barrier, as opposed to how long it takes the population *as a whole* to cross the fitness barrier.

This brings us to the second main mechanism for metastability—one that has been put forward more recently [7, 47, 79, 108, 137]. The second mechanism derives from the observation that large sets of mutually fitness-neutral genotypes are interconnected along single-point mutation paths. That is, sets of isofitness genotypes form extended *neutral networks* under single-point mutations in genotype space.

In this alternative scenario, a population displays a constant distribution of phenotypes for some period while, at the same time, individuals in the population diffuse over a neutral network in genotype space. That is, despite phenotypic metastability, there is no *genotypic* stasis during this period. The phenotype distribution remains metastable until, via diffusion over the neutral network, a member of the population discovers a genotypic connection to a higher-fitness neutral network.

When this mechanism operates, metastability is the result of an *entropy barrier*, as we call it. The population must spread over or search large parts of the neutral network before it finds a connection to a higher-fitness network. One envisages the population moving randomly through a genotypic labyrinth of common phenotypes with only a single or relatively few exits to fitter phenotypes.

8.1.2 Overview

In the following, we analyze and compare the population dynamics of crossing such fitness and entropy barriers with the goals of elucidating the basic mechanisms responsible for each, calculating the scaling forms for the evolutionary times scales associated with each, and understanding their relative importance when both can operate simultaneously.

Section 8.2 defines the basic evolutionary model.

Section 8.3 introduces a tunable fitness function that models the simplest case in which to study both types of barrier crossing. It consists of a single local optimum, with a valley, and a single portal (target genotype) in genotype space. By tuning the height of the local optimum one can change the fitness barrier into an entropy barrier. We analyze this basic model as a branching process, calculating the statistics of lineages of individuals in the fitness valley. Comparison of the theoretical predictions for the fitness-barrier crossing times with data obtained from simulations shows that the theory accurately predicts these fitness-barrier crossing times for a wide range of parameters. We also derive several simple scaling relations for the fitness-barrier crossing times appropriate to different parameter regimes.

Section 8.4 first determines the barrier heights at which the fitness-barrier regime shifts over into an entropic one. After this, we discuss the population dynamics of crossing entropy barriers, providing rough scaling relations for the barrier crossing times in this regime. Comparison of these results with the scaling relations for fitness-barrier crossing shows that entropy-barrier crossing proceeds markedly faster than crossing fitness barriers.

Section 8.5 extends our analysis to a set of much more complicated fitness functions—a class called the *Royal Staircase with Ditches*. These fitness functions are closely related to the Royal Road [137] and chapter 4, and Royal Staircase (chapters 5 and 6) fitness functions that we studied earlier, which consist of a sequence or a network of entropy barriers only. The Royal Staircase with Ditches generalizes this class of fitness functions to one that possesses multiple fitness and entropy barriers of variable width, height, and volume. We adapt the theoretical analysis using our statistical dynamics approach of chapter 4 to deal with these more complicated, but more realistic cases. Comparison of the theoretically predicted and experimentally obtained barrier crossing times again shows that the theory accurately predicts the barrier crossing times in these more complicated situations as well.

Finally, Sec. 8.6 presents our conclusions and discusses the general picture of metastable population dynamics that emerges from our analyses.

8.2 Evolutionary Dynamics

We consider a simple evolutionary dynamics of selection and mutation with a constant population size M. An individual's genotype consists of a binary sequence of on-or-off genes. We consider the simple case in which the fitness of an individual is determined by its genotype only. The genotype-to-phenotype and phenotype-to-fitness maps are collapsed into a direct determination of a genotype's fitness. Selection and reproduction are assumed to take place in discrete generations, with mutation occurring at reproduction. The exact evolutionary dynamics is defined as follows.

- A *population* consists of M binary sequences of a fixed size L.
- A *fitness* f_s is associated with each of the 2^L possible genotypes $s \in \mathcal{A}^L$, where $\mathcal{A} = \{0, 1\}$.
- Every generation M individuals in the current population are sampled with replacement and with a probability proportional to their fitness. Thus, the expected number of offspring for an individual with genotype s is $f_s/\langle f \rangle$, where $\langle f \rangle$ is the current average fitness of the population.
- Once the M individuals have been selected, each bit in each individual is mutated (flipped) with probability μ, the *mutation rate*.

In this basic model there are effectively two evolutionary parameters: the mutation rate μ and the population size M.

Several aspects of the basic model—such as, discrete generations and fixed population size—were mainly chosen for analytical convenience. The discrete-generation

assumption can be lifted, leading to a continuous-time model, without affecting the results presented below. As for the assumption of fixed population size, the analysis can be adapted in a straightforward manner to address (say) fluctuating or exponentially growing populations.

Models including genetic recombination are notoriously more difficult to analyze mathematically. Despite our interest in the effects of recombination, it is not included here largely for this reason. Moreover, for wide parameter ranges in the neutral and piecewise-neutral evolutionary processes we consider, it appears that recombination need not be a dominant mechanism. For example, sections 4.6.5 and 5.8 show that recombination often does not significantly affect population dynamics in these cases.

8.3 Crossing a Single Barrier

We first consider the simple case of a single barrier for the population to cross. Of the 2^L genotypes, there is one with fitness $\sigma > 1$ that we refer to as the *peak* genotype Π. Then there are $2^L - 2$ genotypes with fitness 1 that we refer to as *valley* genotypes. Finally, there is a single *portal* genotype Ω at a Hamming distance w from the fitness-σ peak genotype Π. We view the portal genotype as giving access to higher-fitness genotypes—genotypes whose details are unimportant, since in this section we only analyze the dynamics up to the portal's first discovery.

The variable w tunes the fitness barrier's *width* and the variable σ its *height*. The height σ also indicates a peak individual's *selective advantage* over those in the valley, as measured by the relative difference $\sigma - 1$ of their expected number of offspring. Figure 8.1 illustrates the basic setup.

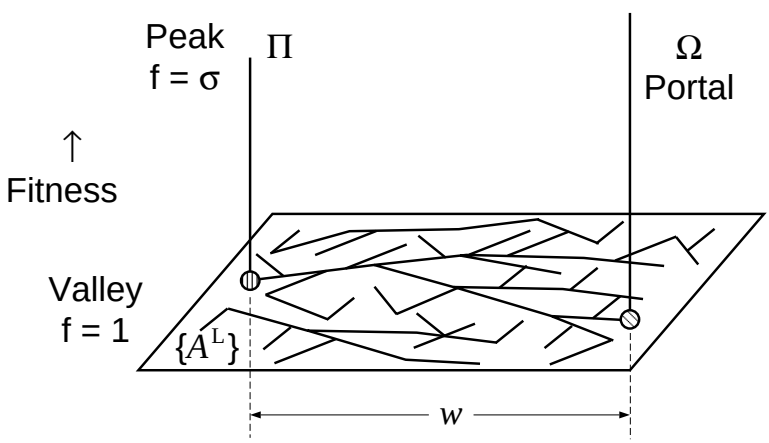

Figure 8.1: Evolution from the peak genotype Π to the higher-fitness portal genotype Ω via low-fitness valley genotypes. The selective advantage of the peak individuals over those in the valley is controlled by the peak height σ. The portal and peak genotypes are a Hamming (mutational) distance w apart. The domain is the hypercube $\mathcal{A}^L$ of all length-L genotypes.

At time $t = 0$ the population starts with all M genotypes located at the peak Π. We then evolve the population under selection and mutation, as described in the previous section, until the portal genotype Ω occurs in the population for the first time. (Hence, the portal's fitness is not relevant.) This defines one evolutionary *run*. We record the time t at which the portal is discovered. We are interested in the average discovery time $\langle t \rangle$, averaged over an ensemble of such runs. We are particularly interested in the scaling of this *barrier crossing time* $\langle t \rangle$ as a function of the evolutionary parameters M and μ, as well as the barrier parameters σ and w.

Let's briefly review in simple language the evolutionary dynamics before launching into the mathematical analysis. In the parameter regime with $\sigma \gg 1$, where the peak fitness is considerably larger than the valley fitness, and with the mutation rate μ not too high, the bulk of the population remains at the peak. That is, the population is a *quasispecies* cloud, centered around the peak genotype Π [33]. For such parameter regimes, the barrier is clearly a fitness barrier: the waiting time $\langle t \rangle$ is determined by the time it takes to create a rare sequence of mutant genotypes that crosses the valley between the peak and the portal.

However, as $\sigma \to 1^+$, the fitness barrier transforms into an entropy barrier. For $\sigma = 1$ there is no fitness difference between peak and valley genotypes and the entire population simply diffuses through genotype space until the portal is discovered. As we will see below, the entropic regime sets in rather suddenly at a value of σ_c somewhat above $\sigma = 1$. As we show, this transition is the well known *error threshold* of molecular evolution theory [32]. At $\sigma = \sigma_c$, the value of which depends on the population size M and mutation rate μ, the subpopulation on the peak becomes unstable in the sense that all individuals on the peak may be lost through a fluctuation. More precisely, the waiting time for such a fluctuation to occur becomes short in comparison to the fitness-barrier crossing time. When this fluctuation has occurred, there is no longer a restoring "force" that keeps the population concentrated around the peak genotype. The population as a whole diffuses randomly through the valley as if the genotypes were all fitness neutral. While our analysis accurately predicts the barrier crossing times in the fitness-barrier regime, it is notable that beyond the error threshold, in the entropic regime, only order of magnitude predictions can be obtained using the current analytical tools.

Calculating the barrier crossing time proceeds in three stages. First, in Sec. 8.3.1 we determine the population's quasispecies distribution, defined as the average proportions of individuals located on the peak and in the valley during the metastable state. From this, one directly calculates the average fitness in the population. Second, in Sec. 8.3.2 we consider the genealogy statistics of individuals in the valley. In the fitness-barrier regime, genealogies in the valley are generally short-lived and are all seeded by mutants of the peak genotype. We approximate the evolution of valley genealogies as a branching process and use this representation to calculate the average barrier crossing time. Third, with this analysis complete, Sec. 8.5.5 then addresses the transition from the fitness-barrier regime to the entropic one.

8.3.1 Metastable Quasispecies

Each evolutionary run, the population starts out concentrated at the peak genotype Π. After a relaxation phase, assumed to be short compared to the barrier crossing time, there

will be roughly constant proportions of the population on the peak and in the valley. We now calculate the equilibrium proportion P_Π of peak individuals and the population's average fitness $\langle f \rangle$, after this relaxation phase.

To first approximation, one can neglect *back mutations* from valley individuals back into the peak genotype. First of all, if $\sigma \gg 1$, selection keeps the bulk of the population on the peak. Additionally, valley individuals produce fewer offspring than peak individuals and they are unlikely—with a probability $1/L$ at most—to move back onto the peak when they mutate. In this regime, the quasispecies distribution is largely the result of a balance between selection expanding the peak population by a factor of $\sigma/\langle f \rangle$ and deleterious mutations moving them into the valley with probability $1-(1-\mu)^L$. The result is that we have a balance equation for the proportion P_Π of peak individuals given by

$$P_\Pi = \frac{\sigma}{\langle f \rangle}(1-\mu)^L P_\Pi \; . \tag{8.1}$$

From this we immediately have that

$$\langle f \rangle = \sigma(1-\mu)^L \; . \tag{8.2}$$

Since we also have that $\langle f \rangle = \sigma P_\Pi + 1 \cdot (1 - P_\Pi)$, we can determine the proportion of peak individuals to be:

$$P_\Pi = \frac{\langle f \rangle - 1}{\sigma - 1} \; . \tag{8.3}$$

For parameters where $\langle f \rangle = \sigma(1-\mu)^L \gg 1$, Eqs. (8.2) and (8.3) give quite accurate predictions for the average fitness and the proportion of individuals on the peak.

In cases where $\langle f \rangle$ is close to 1, a substantial proportion of the population is located in the valley and back mutations from the valley onto the peak must be taken into account. To do this, we introduce the quasispecies Hamming distance distribution $\vec{P} = (P_0, \ldots, P_i, \ldots, P_L)$, where P_i is the proportion of individuals located at Hamming distance i from Π. Thus, $P_0 = P_\Pi$ indicates the proportion on the peak. Under selection, the distribution $\vec{P}$ changes according to:

$$P_i^{\text{sel}} = \frac{(\sigma-1)\delta_{i0}+1}{\langle f \rangle} P_i \; , \tag{8.4}$$

where $\delta_{ij} = 1$, if $i = j$, and $\delta_{ij} = 0$, otherwise. We can write this formally as the result of an operator acting on $\vec{P}$:

$$\vec{P}^{\text{sel}} = \frac{\left(\mathbf{S} \cdot \vec{P}\right)_i}{\langle f \rangle} \; , \tag{8.5}$$

where

$$S_{ij} = \left[(\sigma-1)\delta_{i0}+1\right]\delta_{ij} \; , \tag{8.6}$$

defines the *selection operator* $\mathbf{S}$.

Next, we consider the transition probabilities M_{ij} that under mutation a genotype at Hamming distance j from the peak moves to a genotype at Hamming distance i from

the peak. We have that:

$$\mathrm{M}_{ij} = \sum_{u=0}^{L-j} \sum_{d=0}^{j} \delta_{j+u-d,i} \binom{L-j}{u} \binom{j}{d} \times \mu^{u+d} (1-\mu)^{L-u-d} \,. \tag{8.7}$$

That is, M_{ij} is the sum of the probabilities of all possible ways to mutate u of the $L-j$ bits shared with Π and d of the j bits that differ, such that $j+u-d=i$. Equation (8.7) defines the *mutation operator* $\mathbf{M}$.

We can now introduce the *generation operator* $\mathbf{G} = \mathbf{M} \cdot \mathbf{S}$. The equilibrium quasispecies distribution $\vec{P}$ is a solution of the equation

$$\vec{P} = \frac{\mathbf{G} \cdot \vec{P}}{\langle f \rangle} \,. \tag{8.8}$$

In this way, the quasispecies distribution is given by the principal eigenvector, normalized in probability, of the matrix $\mathbf{G}$; while the average fitness $\langle f \rangle$ is given by $\mathbf{G}$'s principal eigenvalue. Note that this is conventional quasispecies theory [33], apart from the facts that we have grouped the quasispecies members into Hamming-distance classes and that we consider discrete generations, rather than continuous time.

8.3.2 Valley Lineages

Under the approximation that back mutations from the valley onto the peak can be neglected, a roughly constant proportion $1 - P_\Pi$ of valley individuals is maintained by a roughly constant influx of mutants from the peak. Every generation, some peak individuals leave mutant offspring in the valley. Additionally, each valley individual leaves on average a fraction $1/\langle f \rangle$ offspring in the next generation, as can be seen from Eq. (8.4). This means that the fraction of valley individuals, for which *all* of its t ancestors in the previous t generations were valley individuals as well, is only $1/\langle f \rangle^t$. For $\langle f \rangle \gg 1$, this implies in turn that whenever a peak individual seeds a new lineage of valley individuals by leaving a mutant offspring in the valley, this lineage is unlikely to persist for a large number of generations. In other words, lineages composed of valley individuals are short lived.

Intuitively, the idea is that the preferred selection of peak individuals leads to a "surplus" of peak offspring that spills into the valley through mutations. Each mutant offspring of a peak individual forms the root of a relatively small, i.e., short-lived, genealogical tree of valley individuals. The barrier crossing time is determined essentially by the waiting time until one of these *genealogical bushes* produces a descendant that discovers the portal. These processes are illustrated in Fig. 8.2.

We will analyze the evolution of a valley lineage as a branching process [68]. The probability O_n that one particular valley individual leaves n offspring in the next generation is given by a binomial distribution. This is well approximated by a Poisson distribution as follows:

$$O_n = \binom{M}{n} \left(\frac{1}{M\langle f \rangle} \right)^n \left(1 - \frac{1}{M\langle f \rangle} \right)^{M-n} \approx \frac{1}{n!} \left(\frac{1}{\langle f \rangle} \right)^n e^{-1/\langle f \rangle} \,. \tag{8.9}$$

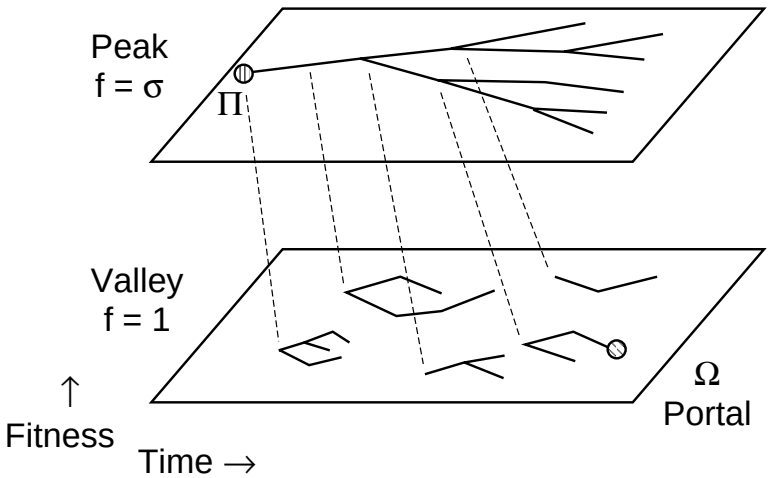

Figure 8.2: Genealogies during fitness-barrier crossing. An example genealogical tree is sketched for peak individuals (above); they have fitness σ and are copies of the peak genotype Π. The valley individuals (below) at lower fitness occur in genealogies that are seeded (dashed lines) from peak individuals. These genealogies are relatively short-lived bushes. Evolution continues until the time at which one of the valley bushes discovers the portal genotype Ω.

To a good approximation, we may treat the evolution of each valley lineage as independent of the other valley lineages. Under this approximation, each valley individual independently has a distribution of offspring given by Eq. (8.9). Of course, under fixed population size, the independence assumption may break down when valley lineages dominate the population.

We now calculate the probability that a valley lineage produces a descendant that discovers the portal Ω before the lineage goes extinct. Let a valley lineage be founded by an ancestor in the valley that is located at Hamming distance j from Ω. We denote by $\bar{p}_j(t)$ the probability that t generations from now, *none* of this ancestor's descendants will have discovered the portal. This probability can be determined recursively, in terms of the probabilities $\bar{p}_i(t-1)$ as follows,

$$\begin{aligned} \bar{p}_j(t) &= O_0 + O_1 \sum_{i=1}^{L} \bar{p}_i(t-1)\mathrm{M}_{ij} \\ &+ O_2 \sum_{i,k=1}^{L} \bar{p}_i(t-1)\mathrm{M}_{ij}\bar{p}_k(t-1)\mathrm{M}_{kj} + \dots \ . \end{aligned} \tag{8.10}$$

The first term in the above equation corresponds to the ancestor having no offspring. This, of course, ensures that the portal will not be discovered t generations from now, since leaving zero offspring implies that the genealogy goes extinct immediately. The second term corresponds to the ancestor having one offspring, at Hamming distance i

from the portal, that will *not* give rise to discovery of the portal. That is, since this offspring itself forms the ancestor of a new valley lineage, $\bar{p}_i(t-1)$ gives the probability that none of *its* descendants discovers the portal within the next $t-1$ generations. The third term corresponds to the ancestor having two offspring, one at distance i from the portal and one at distance k, neither of which give rise to the discovery of Ω. The higher-order terms in Eq. (8.10) correspond to the ancestor having 3, 4, and more offspring.

Recall that the mutation operator M_{ij}, as defined in Eq. (8.7), gave the probability to go from Hamming distance j to distance i from the *peak* under mutation. M_{ij} appears above in Eq. (8.10) with a different, but equivalent, meaning: M_{ij} there gives the probability to go from a Hamming distance j to a distance i from the *portal* under mutation. This use of M_{ij} appears repeatedly in the following.

Using Eq. (8.9) we can sum the series in Eq. (8.10), obtaining:

$$\bar{p}_j(t) = e^{\left([\bar{\mathbf{p}}(t-1)\cdot\mathbf{M}]_j - 1\right)/\langle f\rangle} , \tag{8.11}$$

where $\bar{p}(t) = (\bar{p}_1(t), \bar{p}_2(t), \ldots, \bar{p}_L(t))$ and the vector notation denotes the sum

$$[\bar{p}(t-1)\cdot\mathbf{M}]_j = \sum_{i=1}^{L} \bar{p}_i(t-1)\mathrm{M}_{ij} . \tag{8.12}$$

For $\langle f\rangle \geq 1$, a valley genealogy eventually either discovers Ω or goes extinct; see, for instance, Ref. [68]. Letting $t \to \infty$ in Eq. (8.11), we obtain a set of nonlinear equations for the asymptotic probabilities $\bar{p}_j$ that a genealogical bush, whose founder started at Hamming distance j from Ω, goes extinct before any of its descendants discovers the portal. These are given by

$$\bar{p}_j = e^{\left([\bar{p}\cdot\mathbf{M}]_j - 1\right)/\langle f\rangle} . \tag{8.13}$$

Equations (8.13) appear to be unsolvable in closed analytical form. Their solutions may be numerically approximated in a straightforward manner; for instance, by simply iterating Eq. (8.11). However, in the regime where μ is small and $\langle f\rangle$ is not too close to 1, the probabilities $\bar{p}_i$ are generally very close to 1. In this regime, one may expand Eq. (8.13) to first order around $\bar{p}_i = 1$. To do this, we introduce the probabilities $\epsilon_i = 1 - \bar{p}_i$ that the portal *does* get discovered by the lineage before it goes extinct. To first order in ϵ_i we obtain from Eq. (8.13) the equations given by

$$\epsilon_j = 1 - e^{-\mathrm{M}_{0j}/\langle f\rangle}\left(1 - \frac{[\epsilon\cdot\mathbf{M}]_j}{\langle f\rangle}\right) , \tag{8.14}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_L)$. These equations can be easily inverted, yielding:

$$\epsilon_j = \sum_{i=1}^{L}\left(1 - e^{-\mathrm{M}_{0i}/\langle f\rangle}\right)(\mathbf{I} - \mathbf{R})^{-1}_{ij} , \tag{8.15}$$

where $\mathbf{I}$ is the identity matrix and the matrix $\mathbf{R}$ has components

$$\mathrm{R}_{ij} = \frac{\mathrm{M}_{ij}}{\langle f\rangle} e^{-\mathrm{M}_{0j}/\langle f\rangle} . \tag{8.16}$$

Note that the indices i and j in the matrices run from 1 to L, corresponding to ancestors at Hamming distances between 1 and L from the portal. Note also that, by definition, $\epsilon_0 = 1$.

To first order, Eqs. (8.15) give the probabilities ϵ_j that a valley lineage, founded by an ancestor at a Hamming distance j from Ω, discovers the portal before the lineage goes extinct. Now to calculate the barrier crossing time we just have to determine the number of new valley lineages that are founded per generation.

8.3.3 Crossing the Fitness Barrier

Every generation, M individuals are selected in proportion to their fitness. Each such selection may lead to the seeding of a new lineage in the valley. The probability $P^{\rm not}$ that a selection will *not* lead to the founding of a new valley lineage is given by:

$$P^{\rm not} = \frac{1 - P_\Pi}{\langle f \rangle} + \frac{\sigma}{\langle f \rangle} P_\Pi (1-\mu)^L \ . \tag{8.17}$$

The first term corresponds to selecting a valley individual, that by definition is already on a lineage. The second term corresponds to a peak individual being selected and reproducing without mutation, leaving an offspring on the peak.

The probability $P_j^{\rm seed}$ that a new lineage will be seeded in the valley and at a Hamming distance j from Ω is given by

$$P_j^{\rm seed} = \frac{\sigma}{\langle f \rangle} P_\Pi \left[{\rm M}_{jw} - (1-\mu)^L \delta_{jw} \right] \ , \tag{8.18}$$

where w is the Hamming distance between the peak and the portal. The first factor, $\sigma P_\Pi / \langle f \rangle$, gives the probability that a peak individual will be selected. The term ${\rm M}_{jw}$ is the probability that under mutation this peak individual moves from Hamming distance w to distance j from the portal. For $j \neq w$ this always corresponds to a new lineage in the valley. For $j = w$, we must discount for the probability that the peak individual did not undergo any mutations at all. This is given by the term $(1-\mu)^L \delta_{jw}$.

Putting these together, we find the probability $\bar{P}^{\rm sel}$, that a selection does *not* seed a lineage leading to the portal, is given by

$$\bar{P}^{\rm sel} = P^{\rm not} + \sum_{j=1}^{L} (1 - \epsilon_j) P_j^{\rm seed}. \tag{8.19}$$

Using Eqs. (8.13), (8.17), and (8.18) and the identity $\langle f \rangle = 1 + (\sigma - 1) P_\Pi$, Eq. (8.19) can be rewritten as

$$\bar{P}^{\rm sel} = 1 + \frac{\sigma(\langle f \rangle - 1)}{(\sigma - 1)\langle f \rangle} \left[(1-\mu)^L \epsilon_w + \langle f \rangle \log(1 - \epsilon_w) \right] \ . \tag{8.20}$$

Expanding the logarithm to first order in ϵ_w, and using the approximation in Eq. (8.2) for $\langle f \rangle$, we obtain the simple expression

$$\bar{P}^{\rm sel} = 1 - \epsilon_w (\langle f \rangle - 1). \tag{8.21}$$

The probability that none of the M selections from the current generation seeds a lineage that discovers the portal is simply $\left(\bar{P}^{\rm sel}\right)^M$. By our assumption of a roughly constant quasispecies distribution, this probability is constant. Thus, the expected number $\langle t\rangle$ of generations until a lineage will be seeded that discovers the portal is given by

$$\langle t\rangle = \frac{1}{1-\left(\bar{P}^{\rm sel}\right)^M} \approx \frac{1}{M(\langle f\rangle - 1)\epsilon_w} \; . \tag{8.22}$$

where ϵ_w is given by Eqs. (8.15) and (8.16).

Equation (8.22) constitutes our theoretical prediction for the average barrier crossing time $\langle t\rangle$ as a function of the population size M, the fitness differential σ between the peak and the valley, the mutation rate μ, the string length L, and the width w of the fitness barrier. To obtain it, we made several approximations. We assumed that σ was large enough and μ small enough such that $\langle f\rangle$ was substantially larger than 1. Under those assumptions, lineages in the valley are short lived, the total number of individuals in the valley will be small with respect to M, and the probabilities ϵ_i will be small. This justifies our leading-order expansion for $\langle t\rangle$.

8.3.4 Additional Time in Valley Bushes

Equation (8.22) gives the average number $\langle t\rangle$ of generations until a lineage is founded that discovers the portal. The actual average time until the portal is discovered is somewhat longer, since the lineage that finds the portal itself takes a certain average number of generations to discover the portal. Specifically, there is an additional average time, that we denote by $\langle dt\rangle$, between the *founding* of the first lineage that discovers the portal and the actual discovery of Ω.

We can directly approximate this correction term $\langle dt\rangle$ when the ϵ_i are small. As we will see below, generally $\langle dt\rangle \ll \langle t\rangle$ in the parameter regime where our approximations are valid. This makes the effect of including the correction term $\langle dt\rangle$ rather small in these parameter regimes. However, as we approach the parameter regime where the ϵ_i become large, the average number of generations $\langle t\rangle$ until the founding of the lineage that discovers the portal becomes comparable to the average number of generations $\langle dt\rangle$ that it takes this lineage to actually discover the portal. In this (limited) parameter regime, including the correction term $\langle dt\rangle$ leads to a significant improvement of our theoretical predictions.

Paralleling the development of the Eq. (8.13), we start by expanding Eq. (8.11) to first order in $\epsilon_j(t)$; the probability that the lineage starting at distance j has discovered the portal by time t. We find that

$$\epsilon_j(t) = 1 - e^{-{\rm M}_{0j}/\langle f\rangle}\left(1 - \frac{[\epsilon(t-1)\cdot\mathbf{M}]_j}{\langle f\rangle}\right) \; . \tag{8.23}$$

The expected additional time $\langle dt_j\rangle$, given that the lineage started at a Hamming distance j from Ω and *conditioned* on the lineage discovering the portal, is formally given by

$$\langle dt_j\rangle = \sum_{t=1}^{\infty} t\,\frac{\epsilon_j(t)-\epsilon_j(t-1)}{\epsilon_j} = \sum_{t=0}^{\infty}\frac{\epsilon_j-\epsilon_j(t)}{\epsilon_j} \; , \tag{8.24}$$

where the asymptotic ϵ_j is given by Eqs. (8.15) and (8.16). Using Eq. (8.23) and the boundary conditions $\epsilon_j(0) = 0$, the above sum gives:

$$\langle dt_j \rangle = \frac{1}{\epsilon_j} \sum_{i=1}^{L} \left(1 - e^{-\mathrm{M}_{0i}/\langle f \rangle}\right) (\mathbf{I} - \mathbf{R})^{-2}_{ij} , \tag{8.25}$$

where the matrix $\mathbf{R}$ is again defined by Eq. (8.16).

In order to obtain $\langle dt \rangle$ we have to *weigh* each of the times $\langle dt_j \rangle$ with a factor c_j corresponding to the relative proportion of times that a lineage starting at Hamming distance j discovers the portal. That is, averaged over an ensemble of runs, c_j is the proportion of times that the portal was discovered by a lineage that started at Hamming distance j. The weight c_j should be proportional to both the probability ϵ_j and the rate of creating of lineages at Hamming distance j from the portal. We have that

$$c_j = \frac{\epsilon_j \left(\mathrm{M}_{jw} - (1-\mu)^L \delta_{jw}\right)}{\sum_k \epsilon_k \left(\mathrm{M}_{kw} - (1-\mu)^L \delta_{kw}\right)} , \; j = 0, 1, \ldots, L , \tag{8.26}$$

where the factors in parentheses are similar to those found in Eq. (8.18). It should be noted that here the indices run from 0 to L and not from 1 to L, since the portal may also be discovered by a jump mutation directly from the peak.

Combining Eqs. (8.25) and (8.26) and using Eq. (8.15), we find that the average length of the valley bush that discovers the portal is

$$\langle dt \rangle = \frac{\left[\epsilon \cdot (\mathbf{I} - \mathbf{R}) \cdot \left(\mathbf{M} - \mathbf{I}(1-\mu)^L\right)\right]_w}{\left[\epsilon \cdot (\mathbf{M} - \mathbf{I}(1-\mu)^L)\right]_w + \mathrm{M}_{0w}} , \tag{8.27}$$

where, again, ϵ is given by its components in Eqs. (8.15) and (8.16). The indices in the vector notation now run from 1 to L.

Adding the correction term $\langle dt \rangle$ to $\langle t \rangle$ as given by Eq. (8.22) improves our theoretical predictions especially in the regime where the ϵ_i become large. However, we still expect the approximations leading to the above equations for $\langle t \rangle$ and $\langle dt \rangle$ to break down when $\langle f \rangle \to 1^+$.

8.3.5 Theory versus Simulation

We simulated an evolving population using a fitness function consisting of a single barrier, as described in Secs. 8.2 and 8.3, for a wide range of parameter settings to quantitatively test our theoretical predictions. Results for several parameter regimes are shown in Fig. 8.3, where the simulation results are plotted using dashed lines and the theoretical predictions are plotted with solid lines. Each data point on the dashed lines was obtained by averaging over 250 runs with equal parameter settings. The theoretical predictions are shown as pairs of solid lines, where the lower solid line in each pair shows the predictions from Eq. (8.22) and the upper solid line shows Eq. (8.22) plus the correction term of Eq. (8.27). Note that for most parameter ranges the difference between the two solid lines is so small as to be undetectable.

Figures 8.3(a) and 8.3(b) show the average barrier crossing time $\langle t \rangle$ as a function of the logarithm $\log(\sigma)$ of the barrier height. Additionally, both $\langle t \rangle$ and $\log(\sigma)$ are plotted

using a logarithmic scale. The shapes of the curves correspond to the dependencies of $\log\langle t\rangle$ on $\log(\log\sigma)$. Portions of curves that are straight lines thus indicate a scaling of the form $\langle t\rangle \propto (\log\sigma)^s$, with s the slope of the straight portion. Note that $\langle t\rangle$ ranges over 5 orders of magnitude, from 10 to 10^6, in both Figs. 8.3(a) and 8.3(b). We see that the theory accurately predicts the simulation results for barrier heights that are not too small.

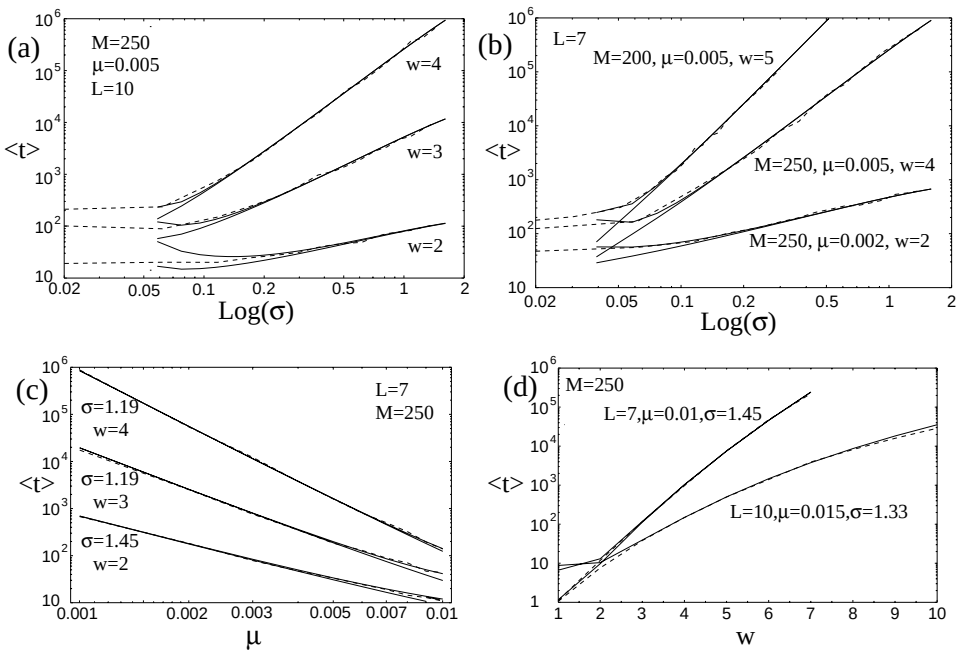

Figure 8.3: Barrier-crossing times $\langle t\rangle$ as a function of barrier height σ, mutation rate μ, and barrier width w, for a variety of parameter settings. The simulation results are plotted using dashed lines. Each point on each dashed line is an estimate of $\langle t\rangle$ averaged over 250 simulation runs. The theoretical predictions are shown as pairs of solid lines: The lower of each pair gives the theoretical predictions of Eq. (8.22), while the higher has the additional correction term of Eq. (8.27) added. Note that, except for the horizontal axis in Fig. (d), all axes use a logarithmic scale. General parameter settings are indicated at the top of each plot, while parameters specific to the different runs are indicated next to the their lines.

In Fig. 8.3(a) the theory starts deviating from the experimental data around $\log(\sigma) \approx 0.06$ for the upper two curves and around $\log(\sigma) \approx 0.15$ for the lowest. These values of σ correspond to selective advantages $\sigma - 1$ of the peak of a little over 6 and 16 percent, respectively. Notice that the upper two experimental curves are almost horizontal for small values of σ up to $\log(\sigma) \approx 0.06$, after which they trend upwards becoming almost linear. As we will show below, it turns out that the location of this crossover is found at the *finite-population error threshold* that separates the entropy-barrier regime from the fitness-barrier regime. That is, for the parameters $M = 250$, $\mu = 0.005$, and $L = 10$, the

critical value σ_c below which the population dynamics acts effectively as if there were no fitness peak at all occurs around $\log(\sigma_c) \approx 0.06$. The same phenomenon is observed in the two upper curves of Fig. 8.3(b): the crossover occurs around $\log(\sigma_c) \approx 0.05$. Note that the correction terms $\langle dt \rangle$ extend the parameter region over which the theory provides accurate predictions approximately up to the finite-population error threshold.

Above the error threshold, for values of σ in the fitness-barrier regime, the curves appear nearly linear. This indicates that the barrier crossing times scale with powers of the logarithm of the barrier height σ: $\langle t \rangle \propto (\log \sigma)^s$, where s is the line's slope. Thus, the barrier crossing time increases relatively slowly as a function of the barrier height. One further observes that the barrier crossing times are not only longer for wider barriers (larger values of w), but that the slopes of the curves are larger as well. That is, for large widths w the barrier crossing time increases faster as a function of σ than for low values of w.

Figure 8.3(c) shows the barrier crossing time $\langle t \rangle$ as a function of the mutation rate μ, for three different values of the barrier width w and two different values of the barrier height σ. The population size is $M = 250$ and the genotypes length $L = 7$ for all three curves. On the logarithmic scales, the curves again look approximately linear, indicating that the barrier crossing time scales as a power law in the mutation rate μ: $\langle t \rangle \propto \mu^s$, where s is the slope. We again see that for wider barriers, the waiting times are both larger and vary more rapidly with μ. The theory predicts the simulation results quite accurately over the entire range. Only for large mutation rates ($\mu \approx 0.01$) do the theoretical predictions with and without the correction term of Eq. (8.27) differ significantly. In this regime the theoretical and experimental values start to differ slightly as well, although the predictions are still accurate. It is notable in the two lower curve families, with barrier widths $w = 2$ and $w = 3$, that the correction term $\langle dt \rangle$ improves the theoretical predictions for high mutation rates.

Finally, Fig. 8.3(d) shows the barrier crossing time $\langle t \rangle$ as a function of the barrier width w. Only the barrier crossing time $\langle t \rangle$ is shown on a logarithmic scale, so that any linear dependence indicates an exponential scaling: $\langle t \rangle \propto 10^{sw}$, where s is the slope. Again, the theory accurately predicts the barrier crossing time. The fact that the curves are not linear and bend downwards shows that $\langle t \rangle$ grows more slowly than exponential with barrier width; although it still increases rapidly as a function of w. In fact, we chose large values of the mutation rate μ in these plots ($\mu = 0.01$ and $\mu = 0.015$) to ensure that the barrier crossing time is still in a reasonably bounded range up to large barrier widths. For smaller mutation rates, the barrier crossing times become so large as to make it impossible to perform an adequate number of simulation runs. For the case $w = 1$, the correction term $\langle dt \rangle$ leads to an overestimation of $\langle t \rangle$. Note, however, that for $w = 1$ there is effectively no fitness barrier; the portal is a mutant neighbor of the peak genotype and so valley bushes are essentially nonexistent.

In summary, Figs. 8.3(a)-(d) show that the theoretical predictions of Eq. (8.22), possibly including the correction term of Eq. (8.27), accurately predict the average barrier crossing times estimated over a wide range of parameters from simulations of an evolving population. The theory breaks down, as expected, when the barrier height σ becomes small ($\sigma \approx 1$) and this is illustrated on the left-hand sides of Figs. 8.3(a) and 8.3(b). In this low-σ regime, which sets in suddenly as a function of σ, $\langle t \rangle$ becomes almost independent of σ. Roughly speaking, the selection pressure is too small to keep the

population concentrated around the peak, and the population randomly diffuses through the valley until it discovers the portal. In this regime, the barrier is in effect not a fitness barrier, but an entropy barrier.

In Sec. 8.4 we will analyze the location of the finite-population error threshold that separates the fitness and entropy barrier regimes and discuss the entropy-barrier crossing population dynamics. In the next subsection, though, we first discuss the scaling of the fitness-barrier crossing time $\langle t \rangle$ with the different parameters σ, w, μ, and M.

8.3.6 Scaling of the Barrier Crossing Time

In Fig. 8.3 we saw, by varying one parameter at a time, that the barrier crossing time scaled as a power law in the logarithm of the barrier height $\log(\sigma)$, as a power law in mutation rate μ, and somewhat slower than exponential in the barrier width w. Analytically extracting these scalings from Eq. (8.22) is quite challenging and incomplete at this time. Empirically, though, we found that the barrier crossing time can be fit quite accurately, in the regime where σ is not too small (above the error threshold), to a scaling function with the following form

$$\langle t \rangle \propto \frac{1}{w! M \mu} \left(\frac{\log(\sigma)}{\mu} \right)^{w-1} [\log(\sigma)]^{-\gamma - \delta \log(\mu)} \ , \tag{8.28}$$

where γ and δ are (constant) scaling exponents. For *both* the genotype lengths ($L = 7$ and $L = 10$) for which we have detailed data, we found that $\gamma \approx 0.75$ and $\delta \approx 0.1$.

This empirical scaling law confirms that, in fact, the barrier crossing time $\langle t \rangle$ scales as a power law in both $\log(\sigma)$ and μ. We see, in particular, that the dependence on the mutation rate μ scales roughly inversely with μ^w and the dependence on $\log(\sigma)$ scales roughly as $[\log(\sigma)]^{w-2}$. Furthermore, we see that $\langle t \rangle$ scales as $e^{cw}/w!$, with c a constant, when only the barrier width w is varied. The scaling with w is thus by far the *most rapid* and therefore dominant scaling. That is, widening the barrier increases the waiting time $\langle t \rangle$ much more than increasing the height of the barrier or decreasing the mutation rate.

These empirically observed scaling behaviors can be elucidated using a simple analytical argument. To this end we employ several simplifications. First, we assume that the major contributions to the probability of barrier crossing come from terms with the *minimal* number of mutations. That is, for barriers of width w, at least w mutations must occur in a peak individual in order to discover the portal. Thus, we assume that contributions from "paths" between peak and portal that involve more than w mutations are negligible. This for instance implies that we neglect the contributions from lineages founded at Hamming distances w through L from Ω. Furthermore, we assume that valley lineages are unlikely to be founded more than 1 mutation away from the peak. Putting these together, the dominant contribution to the barrier crossing probability comes from lineages that are founded at a Hamming distance $w - 1$ from the portal. Note that if we set $P_\Pi = 1$ for simplicity, each generation approximately

$$\frac{w \mu M}{1 - \mu} \tag{8.29}$$

such lineages are founded.

We will now estimate the probability that a lineage, starting at Hamming distance $w-1$ from the portal, discovers the portal exactly t generations after its founding. We approximate the valley genealogies by assuming that each valley individual can only have zero or one offspring each generation. This implies that a valley lineage consists of a single line of individuals; i.e., lineages do not branch. The probability that such a lineage persists for at least t time steps is $1/\langle f \rangle^t$. At $t=0$, the lineage has $w-1$ bits set incorrectly, and $L-w+1$ bits set correctly. In order for the lineage to discover the portal *exactly* at time t, it will have to mutate its bits such that, at time t and for the first time, the $w-1$ "incorrect" bits will all have been flipped to the correct state and all the $L-w+1$ correct bits are left undisturbed. Thus, between time 0 and t, the $w-1$ incorrect bits have to be mutated exactly once, while the correct bits have been undisturbed. Since we are calculating the probability for the portal to be discovered exactly at time t, one of the $w-1$ bits has to flip at time t, while the other $w-2$ might flip at any prior time. This gives $(w-1)\,t^{w-2}$ possibilities for contributing flips. All other bits have to remain unflipped for all time steps.

Thus, the probability $P_t^{\rm find}$ that a lineage finds the portal exactly at time t is approximately given by:

$$P_t^{\rm find} = (w-1)\,t^{w-2}\mu^{w-1}(1-\mu)^{Lt-w+1}\left(\frac{1}{\langle f\rangle}\right)^t \,, \tag{8.30}$$

where the last factor gives the probability that the lineage survives until time t. Using Eq. (8.2) and summing Eq. (8.30) over t we find:

$$\begin{aligned} P^{\rm find} &= \sum_{t=0}^{\infty} P_t^{\rm find} \\ &= (w-1)\left(\frac{\mu}{1-\mu}\right)^{w-1}\sum_{t=0}^{\infty} t^{w-2}\left(\frac{1}{\sigma}\right)^t \\ &= (w-1)\left(\frac{\mu}{1-\mu}\right)^{w-1}{\rm Li}_{2-w}\left(\frac{1}{\sigma}\right) \,, \end{aligned} \tag{8.31}$$

where ${\rm Li}_n(x)$ is the poly-logarithm function: essentially defined by the sum in the second line above. It is more insightful to approximate the sum with an integral. We then obtain

$$P^{\rm find} = (w-1)\left(\frac{\mu}{1-\mu}\right)^{w-1}\int_0^\infty t^{w-2}e^{-\log(\sigma)t}dt = (w-1)!\left(\frac{\mu}{\log(\sigma)(1-\mu)}\right)^{w-1} \tag{8.32}$$

Recall that the rate at which lineages at Hamming distance $w-1$ are being created is $w\mu M/(1-\mu)$. Using this and noting that the barrier crossing time is inversely proportional to $P^{\rm find}$, we obtain a scaling of the form

$$\langle t\rangle \propto \frac{1}{w!M\mu}\left(\frac{\log(\sigma)}{\mu}\right)^{w-1} \,, \tag{8.33}$$

where we have neglected the factor $(1-\mu)^w \approx 1$. The scaling relation of Eq. (8.33) recovers most of the empirically determined scaling behavior in Eq. (8.28).

The dominant scaling with μ and $\log(\sigma)$ can be understood as follows. The average time that a lineage spends in the valley before going extinct is roughly $1/\log(\sigma)$. Thus, $\mu/\log(\sigma)$ gives the average number of mutations that a lineage in the valley undergoes before it goes extinct. Since this number is generally much smaller than 1, it can be interpreted as the probability of having a single mutation in a valley lineage. The probability of having $w-1$ mutations is then of course $(\mu/\log(\sigma))^{w-1}$. There is an additional factor $1/\mu$ from the rate at which valley lineages are being created at Hamming distance $w-1$. Ref. [143] also argues, along somewhat different lines, that the barrier crossing time should have a power-law dependence on mutation rate: $\langle t\rangle \propto \mu^{-w}$.

The correction factors—those with scaling exponents γ and δ in Eq. (8.28)— probably arise from the fact that lineages are not simple unbranching lines of descendants, as we have assumed, but are more complicated tree-like genealogies.

The factor $w!$ in Eq. (8.33) counts the number of distinct paths of minimal length between the peak and the portal. Curiously, it appears from the scaling formulas that when w gets very large, the barrier crossing time starts to *decrease* again. Applying Stirling's approximation to the factorial function in Eq. (8.33) indicates that $\langle t\rangle$ has a maximum around $w\mu = \log(\sigma)$. Although this may initially seem strange, it does make sense, since as we will now argue, fitness barriers for which $w\mu > \log(\sigma)$ do not exist.

If there are $w!$ independent paths between peak and portal, this implies that there are w independent directions from the peak into the valley. In other words, at least w bits of the peak genotype can undergo deleterious mutations. As we will see in Sec. 8.4 below, the error threshold at which peak individuals becomes unstable in the population occurs near

$$\sigma(1-\mu)^L = 1\ . \tag{8.34}$$

To first order in μ, this is equivalent to $\log(\sigma) = L\mu$. That is, if $L\mu > \log(\sigma)$, peak individuals will be lost from the population, and the population will start diffusing randomly through genotype space. Obviously, the genotype length has to be longer than the barrier width $L > w$. Therefore, $w\mu > \log(\sigma)$ implies that the genotype length L is so large that it is impossible to stabilize the peak individuals. In other words, fitness barriers with $w\mu > \log(\sigma)$ simply do not exist.

Finally, it should be noted that as $\log(\sigma) \to \infty$, the barrier crossing time goes to a finite asymptote and *not* to infinity. Since valley lineages have probability zero to reproduce in this limit, the asymptotic barrier crossing time is given by the (finite) waiting time for a "long jump" in which a peak individual undergoes w mutations at once.

The main consequence of the scaling relations just derived, is that if the population is located on a fitness "plateau" in genotype space, surrounded by different valleys on all sides, then it will most likely escape from the plateau via the valley with the smallest *width* and *not* via the valley path with the smallest depth. One concludes that high barriers can be passed relatively easily, as long as they are narrow; while wide barriers take a very long time to cross, even if they are shallow.

We should emphasize that this situation is very different from the scaling of barrier crossing times generally encountered in physics or, for that matter, in evolutionary models that literally interpret the "landscape" metaphor as leading to stochastic gradient dynamics on a fitness "potential". In these settings, the system's state space has an energy function defined on it that acts as a potential field. In the absence of any noise, the

system is assumed to follow the gradient (downward) of the energy "landscape". In the presence of noise, the system can deviate from its gradient path, but movement against the gradient is unlikely in proportion to its deviation from the local gradient. The barrier crossing times then depend mainly on the barrier *height*, and they scale exponentially with this barrier height [54].

For example, imagine an energy barrier that consists of a steep slope upwards, followed by a long plateau and then a steep slope leading downward on the other side. The initial steep ascent from the valley onto the plateau is very unlikely since it involves moving against a steep gradient. However, after this unlikely step has been established, the system can cross the long plateau to the other side relatively easily, since it does not involve moving against an energy gradient. Thus, the width of the energy barrier is almost immaterial, while the barrier height is the defining impediment, since it determines the extent to which movement against the gradient must occur.

The situation is entirely different for fitness "landscapes" in which an evolving population moves. For an evolving population, making a large jump in fitness is not unlikely at all. One mutation in one individual can do the trick. However, since some individuals remain at the peak, the individuals in the valley are continuously in competition with these higher-fitness peak individuals. An absolute fitness scale is set by these peak individuals. It is therefore survival at low fitness—compared to the most-fit individuals in the population—for an extended period of time that is unlikely. And this is why the time it takes to move across the plateau is the key parameter—which, of course, is controlled by the barrier width.

The preceding discussion should make it clear, once again, that this analogy—between a population evolving over a fitness "landscape" and a physical system moving over its energy "landscape" in state space—is problematic: at best it may lead one to the wrong intuitions; at worst the basic physical results simply do not describe evolutionary behavior.

8.4 The Entropy Barrier Regime

Figures 8.3(a) and 8.3(b) showed that below a critical barrier height σ_c, where the dashed lines began to run horizontally, the barrier crossing time became effectively independent of σ. We also saw that the theory breaks down for $\sigma < \sigma_c$. In this regime, all peak individuals are quickly lost and the population diffuses through the valley until the portal is discovered. The theoretical calculations, in contrast, assumed the population was located at the peak and that short lineages were continuously spawned in the valley. It is no surprise then that the predictions break down in this regime. Since the population dynamics is dominated by diffusing through the valley's fitness-neutral volume, we refer to this as the *entropy-barrier regime*. Before discussing the barrier crossing times in this entropic regime, we first estimate the error threshold's location σ_c as a function of μ, M, and L.

8.4.1 Error Thresholds

As a first, population-size independent approximation one might guess that the error threshold occurs when the average number of peak-offspring produced by a single peak individual in a population of valley individuals is 1. From Eq. (8.2) this leads to an estimated critical barrier height σ_c of

$$\sigma_c = (1-\mu)^{-L} \; . \tag{8.35}$$

This equation is the standard error-threshold result in molecular quasispecies theory [32, 33]. For the parameters $L = 10$ and $\mu = 0.005$ of Fig. 8.3(a), this leads to $\log(\sigma_c) \approx 0.05$, or $\sigma_c \approx 1.05$. As seen from the figure, though, the entropic regime extends to somewhat higher peak fitness; as far as $\log(\sigma) \approx 0.06$. This deviation is due to finite-population sampling effects and to neglecting back mutations, which become important as the error threshold is approached.

For finite populations, the error threshold can be defined most naturally as those parameter values for which the mean proportion P_Π of peak individuals equals the variance, due to finite population fluctuations, in P_Π. That is, the criterion for reaching the error threshold is

$$(P_\Pi)^2 = \mathrm{Var}(P_\Pi) \; . \tag{8.36}$$

The intuition behind this definition is as follows: Since the proportion of peak individuals fluctuates, eventually a large fluctuation will occur that leads to the loss of all peak individuals. As was shown in chapter 4, however, the waiting time for such a destabilization to occur increases exponentially with the ratio $(P_\Pi)^2/\mathrm{Var}(P_\Pi)$. Only when $(P_\Pi)^2 < \mathrm{Var}(P_\Pi)$ do such destabilizations occur relatively frequently. For $(P_\Pi)^2 > \mathrm{Var}(P_\Pi)$ the fluctuations are small enough so that the proportion of peak individuals typically does not vanish. Therefore, it is natural to use Eq. (8.36) to delineate the regimes with "unstable" and "stable" peak populations and so to distinguish between the fitness-barrier and entropy-barrier regimes in the population dynamics.

Finite-population error thresholds may also be defined in alternative ways; cf. Ref. [110]. Typically, one finds that, although the conceptual motivations differ, the quantitative parameter values for which the different error thresholds occur are quite similar.

The variance $\mathrm{Var}(P_\Pi)$ can be most easily calculated using diffusion-equation methods. For an introduction to these techniques in the context of mathematical population genetics, see for instance Ref. [90]. To begin, we assume that, due to sampling fluctuations, at some particular time t the actual proportion of peak individuals is not P_Π but instead is $P(t) = P_\Pi + x(t)$. That is, the proportion $P(t)$ of individuals on the peak deviates $x(t)$ from its equilibrium value P_Π. We focus on the dynamics of the deviation $x(t)$. At the next generation, the expected deviation $\langle x(t+1)\rangle$ is

$$\langle x(t+1)\rangle = \frac{x(t)}{\langle f\rangle} \; . \tag{8.37}$$

Thus, the *expected* change $\langle \delta x\rangle$ in the deviation is given by

$$\langle \delta x\rangle = \frac{1-\langle f\rangle}{\langle f\rangle}\, x \equiv -\gamma\, x \; , \tag{8.38}$$

where we have defined γ by the last equality. γ measures the average rate at which fluctuations around the quasispecies equilibrium distribution are damped. The second moment $\langle(\delta x)^2\rangle$ of the change δx is approximately given by the variance of the binomial-sampling distribution. One finds that

$$\langle(\delta x)^2\rangle = \frac{1}{M}\left(P_\Pi + \frac{x}{\langle f\rangle}\right)\left(1 - P_\Pi - \frac{x}{\langle f\rangle}\right). \tag{8.39}$$

A Fokker-Planck diffusion equation approximation determines the temporal evolution of distribution $\Pr(x,t)$ of $x(t)$ via

$$\frac{\partial}{\partial t}\Pr(x,t) = -\frac{\partial}{\partial x}\langle\delta x\rangle\Pr(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\langle(\delta x)^2\rangle\Pr(x,t)\ , \tag{8.40}$$

where $\langle\delta x\rangle$ from Eq. (8.38) gives the drift term and $\langle(\delta x)^2\rangle$ from Eq. (8.39) the diffusion term. Solving for the limit distribution $\Pr(x)$ for x yields

$$\Pr(x) = C\left(P_\Pi + \frac{x}{\langle f\rangle}\right)^{2M\langle f\rangle(\langle f\rangle-1)P_\Pi}\left(1 - P_\Pi - \frac{x}{\langle f\rangle}\right)^{2M\langle f\rangle(\langle f\rangle-1)(1-P_\Pi)}. \tag{8.41}$$

Here C is a normalization constant that ensures $\Pr(x)$ is normalized on the interval $x \in [-P_\Pi, 1-P_\Pi]$. If we expand the fluctuations to second-order around $x = 0$, the distribution becomes a Gaussian given by

$$\Pr(x) = \tilde{C}e^{-\frac{M\gamma}{P_\Pi(1-P_\Pi)}x^2}, \tag{8.42}$$

where $\tilde{C}$ is again a normalization constant. From this distribution of fluctuations one directly reads off the variance $\mathrm{Var}(P_\Pi)$, finding that

$$\mathrm{Var}(P_\Pi) = \frac{P_\Pi(1-P_\Pi)}{2M\gamma} = \frac{\langle f\rangle\left(\sigma - \langle f\rangle\right)}{2M(\sigma-1)^2}\ , \tag{8.43}$$

where we used Eqs. (8.3) and (8.38) to arrive at the last line.

As noted before, we define the finite-population error threshold by those parameter values for which $(P_\Pi)^2 = \mathrm{Var}(P_\Pi)$. Using Eq. (8.43) leads to the error-threshold parameter constraints given by

$$\frac{2M(\langle f\rangle - 1)^2}{\langle f\rangle(\sigma - \langle f\rangle)} = 1. \tag{8.44}$$

If we substitute the parameter values $\mu = 0.005$, $M = 250$, and $L = 10$ of Fig. 8.3(a) and use Eq. (8.2) for $\langle f\rangle$, we find the error threshold at $\log(\sigma_c) \approx 0.059$. This agrees quite well with the location at which the experimental curves start bending upwards with increasing peak height.

8.4.2 The "Landscape" Regime

As we have pointed out previously, the scaling relations derived in Sec. 8.3.6 contrast strongly with those based on "landscape" models in which the population *as a whole*

diffuses through the fitness landscape [95, 107]. For those models, the barrier crossing time scales exponentially with population size and barrier height. It turns out that this scaling behavior—appropriate to the "landscape" regime—can be reconciled with the scaling formulas derived in Sec. 8.3.6 by closer inspection of Eq. (8.44).

As noted above, the average *destabilization time* for a fluctuation to occur that makes all peak individuals disappear from the population scales exponentially in the ratio $\mathrm{Var}(P_\Pi)/(P_\Pi)^2$ given by Eq. (8.44). Thus, Eq. (8.44) shows that this destabilization time increases exponentially with population size. For cases where $\langle f \rangle \gg 1$ and for reasonable population sizes, the destabilization time is so large that the barrier crossing time is determined by how long it takes a rare mutant to cross the fitness valley.

Close to the finite-population error-threshold ($\langle f \rangle \approx 1$), however, it might be the case that the time to create such a rare sequence of mutants is long in comparison to the destabilization time. In this situation, the barrier crossing time is essentially given by the destabilization time: As soon as all peak individuals are lost, the population diffuses through the valley and quickly discovers the portal. Thus, in the very restricted "landscape" parameter regime just around the error-threshold, the barrier crossing time is determined by the destabilization time and *does* scale exponentially with population size and barrier height.

Beyond the error threshold—that is, for smaller populations, larger mutation rates, smaller barrier heights, or longer genotypes—the peak readily becomes unoccupied. In this regime, the barrier crossing time becomes almost independent of barrier height σ. The barrier to be crossed is then no longer a fitness barrier. Instead, it has become an entropy barrier. The population must search through almost all of the valley until the portal is discovered. Thus, only for parameters near the boundary between the fitness and entropic regime does the barrier crossing time scale in accordance with the "landscape" models.

8.4.3 Time Scales in the Entropic Regime

The population dynamics in the entropic regime beyond the error threshold is modeled most directly by considering an entirely flat (constant) fitness function; in particular, one in which all genotypes have fitness 1 and containing a single portal Ω. The population starts out concentrated on a genotype at Hamming distance w from Ω and evolves under selection and mutation until the portal genotype is discovered for the first time. Denote this average entropy-barrier crossing time by τ.

The calculation of the entropy-barrier crossing time appears less analytically tractable than the calculation of the fitness-barrier crossing time. The main difficulty arises from the sampling of individuals at each generation, combined with the global constraint of a fixed population size. Due to this sampling dynamics, subtle genetic correlations emerge between the individuals. Although some of the aspects of the correlation statistics have been derived analytically [31], the entropy-barrier crossing time τ depends in a complicated, and not yet well understood, way on these correlations. We will discuss the difficulties with calculating entropic barrier crossing time by deriving several simple approximations and discussing why they fail to provide accurate quantitative predictions.

First, one can approximate the neutral evolution just defined by assuming that each individual in the population has exactly one offspring. In this case, the population effect-

ively consists of M independent *random walkers* that diffuse through genotype space. Since each individual has only one offspring one can identify its genealogy with a single evolving genotype that mutates each bit with probability μ at each generation. Since $\mu L \ll 1$ in general, this genotype effectively performs a random walk in the hypercube, where random walk steps are made at a rate of one step per $1/(L\mu)$ generations on average.

The average time τ_1 a *single* random walker takes to discover Ω is given by:

$$\tau_1 = \sum_{i=1}^{L} (\mathbf{I} - \mathbf{M})^{-1}_{iw} \ , \tag{8.45}$$

where the matrix indices run from Hamming distance 1 through L. τ_1 determines an upper bound for the entire population's barrier crossing time. For parameter settings in the fixation regime, where $ML\mu \ll 1$, sampling fluctuations cause the population to converge onto M copies of a single genotype. As is well known from the theory of neutral evolution [91], this set of identical genotypes performs a random walk through the genotype space at the same rate as a single individual. Thus, in this limit, τ_1 gives a reasonable prediction for the entropy-barrier crossing time. However, for Fig. 8.3(a)'s parameter settings ($M = 250$, $\mu = 0.005$, and $L = 10$) that give $ML\mu = 12.5$, we find that $\tau_1 \approx 23000$, almost independent of valley width w. Of course, this grossly overestimates the observed barrier crossing times, which vary from $\langle t \rangle \approx 25$ for $w = 2$ to $\langle t \rangle \approx 227$ for $w = 4$.

For M independent random walkers, one might simply assume that the waiting time would be roughly a factor M slower, i.e. $\tau_M = \tau_1/M$. Unfortunately, this leads to $\tau_M \approx 93$ which overestimates the observed time for $w = 2$ and underestimates $\langle t \rangle$ for $w = 4$.

The precise probability $\bar{p}_w(t)$, that *none* of M independent random walkers starting at a Hamming distance w have found the portal by time t, is given by:

$$\bar{p}_w(t) = \left(\sum_{i=1}^{L} \mathrm{M}^t_{iw} \right)^M \ . \tag{8.46}$$

From this, one estimates the average entropy-barrier crossing time τ to be:

$$\tau_M = \sum_{t=1}^{\infty} t \left[\bar{p}_w(t-1) - \bar{p}_w(t) \right] = \sum_{t=0}^{\infty} \bar{p}_w(t) \ . \tag{8.47}$$

For Fig. 8.3(a)'s parameters, Eq. (8.47) gives $\tau \approx 15$, 58, and 117 for barrier widths $w = 2, 3,$ and 4, respectively. These values underestimate each observed waiting time by almost a factor of 2. Apparently, sampling fluctuations cause the population to explore the genotype space less rapidly than *independent* random walkers do. As already noted above, the reason for this is that sampling convergence leads different individuals to evolve genetic correlations to some degree.

One way to think about this is to investigate genealogies. Ref. [31] showed that the probability $\Pr(t)$ for two randomly chosen individuals in the current population to have had a common ancestor less than t generations ago is approximately given by

$$\Pr(t) \approx 1 - e^{-t/M} \ . \tag{8.48}$$

This means that, on average, a pair of individuals has only undergone $ML\mu$ mutations each since the time $t \approx M$ they descended from a common ancestor. When $M\mu$ is not much larger than 1, this implies that two individuals are more strongly correlated genetically than random genotypes. Due to this, it is easy to see, at least qualitatively, that the entropy-barrier crossing time is longer than that predicted for independent random walkers. The correlation, or clustering, of individuals in genotype space leads the population to explore the valley's neutral volume at a slower rate. Thus, the predictions obtained by assuming M random walkers, as given by Eqs. (8.46) and (8.47), are lower bounds to the actual waiting times.

It turns out that the upper (Eq. (8.45)) and lower (Eq. (8.47)) estimates do not tightly bound the actual waiting times $\langle t \rangle$. They may differ by several orders of magnitude. Fortunately, the lower bound obtained from Eqs. (8.46) and (8.47) typically produces reasonable order of magnitude estimates for parameter regimes in which $ML\mu > 1$. This order of magnitude estimate gives the following scaling relation for the entropy-barrier crossing time

$$\tau \approx \frac{2^L}{ML\mu} \,. \tag{8.49}$$

8.4.4 Anomalous Scaling

The order of magnitude estimate given by Eq. (8.49) predicts that the the entropy-barrier crossing time τ scales inversely with both μ and M. This scaling is, of course, exactly one's intuitive expectation: the rate at which the genotype space is explored is proportional to both mutation rate μ and population size M. M individuals cover M times as much genotypic "ground" as one individual. Individuals that "move" twice as fast, cover twice as much ground as well. And so, the waiting time should be inversely proportional to both M and μ, which set the exploration rate.

In light of this, it is interesting that data from simulations shows that the entropy-barrier crossing time τ scales as a power law in both μ and M, but *not* with exponents equal to -1, as the preceding simple argument suggests. To be clearer on this point, Fig. 8.4 illustrates the observed scaling behavior of the entropy-barrier crossing time as a function of M and μ.

The solid lines plot the data obtained from simulations while the dashed lines show scaling (power-law) functions that were fitted to the experimental data. All axes use logarithmic scales. All simulations were performed with genotypes of length $L = 10$ bits. In all of the runs, at time $t = 0$ all individuals start at Hamming distance $w = 5$ from the portal. Figure 8.4(a) shows τ's dependence on M for three different values of μ. The approximately straight lines show that the entropy-barrier crossing time depends roughly as a power law on the population size M:

$$\tau \propto \frac{1}{M^\alpha} \,. \tag{8.50}$$

Of course, the scaling exponent α may itself depend on μ.

Similarly, Fig. 8.4(b) shows the dependence of τ on μ for two different values of M. In this case too, the curves appear well approximated by a straight line, indicating that

for fixed M the dependence on μ is roughly given by

$$\tau \propto \frac{1}{\mu^{\beta}} , \tag{8.51}$$

where β may again depend on the population size M. In Table 8.1 the exponents of the estimated dashed lines in Figs. 8.4(a) and 8.4(b) are given, along with their estimated errors.

μ	0.002	0.005	0.008
α	0.740 ± 0.01	0.744 ± 0.02	0.761 ± 0.03
M	50	250	
β	1.292 ± 0.008	1.365 ± 0.014	

Table 8.1: Estimated exponents α and β as defined by Eqs. (8.50) and (8.51).

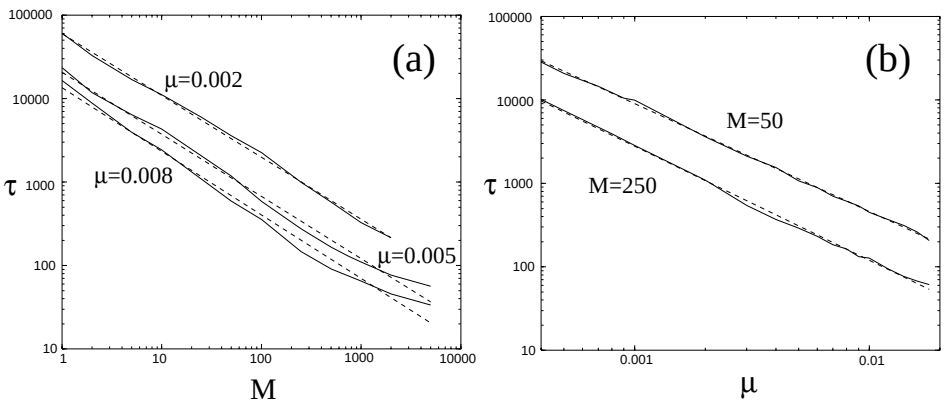

Figure 8.4: Entropy-barrier crossing time τ as a function of population size M and mutation rate μ. The solid lines are data obtained from simulations, while the dashed lines show the estimated scaling functions. All experiments were done with genotypes of length $L = 10$ and a barrier width of $w = 5$. All axes are shown on logarithmic scales. In Fig. (a) τ is shown as a function of population size M for three values of the mutation rate μ. In Fig. (b) τ is shown as a function of the mutation rate μ for two values of population size M. The approximately straight lines show that τ scales as a power law in M when μ is kept constant and, vice versa, as a power law in μ when M is kept constant. Table 8.1 lists the estimated exponents for the power laws. These were used to plot the dashed lines.

The values of the exponents α for different μ and β for different M are very close to each other; $\alpha \approx 3/4$ and $\beta \approx 4/3$. It is clear, however, that they are not *constants*: α does depend on μ and β on M. Note that the estimates for α are all below 1, while those for β are above 1. Thus, doubling the population size decreases τ by *less* than a factor

of two, while doubling the mutation rate decreases τ with *more* than a factor of two. Intuitively, what is happening is that, due to the clustering in the population, doubling the population size does not lead to a doubling of the exploration rate. That is, some of the "added" members in the larger population will simply occur at genotypes where other members of the population are already located. Thus, they do not contribute to additional novel exploration. In contrast, doubling the mutation rate not only doubles the rate of movement (diffusion) of individuals in the population, it additionally decreases the clustering and so reduces genetic correlations. Due to the combination of these two effects, the entropy-barrier crossing time decreases more than a factor two.

Of course, one would like to predict these anomalous exponents. In principle, they should be calculable from knowledge of the clustering structure of the population at different values of M and μ. For example, if we view the population as a blob or collection of blobs in genotype space, one would like to know how many distinct genotypes, on average, are neighboring one or more individuals of the population. Roughly speaking, we would like to know the average "surface area" of the population in genotype space. Knowledge of this statistic would then supply us with the average probability that a mutation leads to a genotype not currently present in the population. This, in turn, quantifies the population's rate of exploring novel genotypes, while taking into account genetic correlations. Although several statistics of these genotype blobs were calculated in Ref. [31], we have at present not been able to adapt these results to infer the necessary type of statistics just outlined. The analytical prediction of the scaling exponents α and β thus awaits further progress. For the present, we will use our order of magnitude estimate in Eq. (8.49) to compare the entropy-barrier crossing time with the fitness-barrier crossing times.

In summary, we analyzed the fitness- and entropy-barrier crossing times for the simplest (single peak) landscapes in which both types of barrier occur. Our results are summarized by the scaling relations of Eqs. (8.28) and (8.49). In the following sections, we apply the preceding analysis to more complicated fitness functions that contain multiple fitness and entropy barriers.

8.5 Traversing Complex Fitness Functions

Up to this point, to make analytical progress we focused on fitness functions that were intentionally simple: a single portal and a single peak in genotype space. Despite this, the analysis of barrier crossing times just developed can be extended with relative ease to more complicated evolutionary processes. To illustrate this extension of the theory, we now introduce a class of more complicated fitness functions that contain multiple fitness and entropy barriers of tunable width and height. That is, in this class of fitness functions, the population may have to cross both a fitness *and* entropy barrier to escape from its metastable state. Since the relative sizes of both these types of barriers can be tuned, we can explicitly compare the time scales for crossing fitness and entropy barriers within the same evolutionary process.

8.5.1 The Royal Staircase with Ditches

The class of fitness functions which we call the *Royal Staircase with Ditches* is closely related to the *Royal Staircase* and *Royal Road* fitness functions that we have analyzed previously, see [137, 24] and chapters 4, 5, and 6. Those fitness functions did not contain fitness barriers but instead consisted of a series of entropy barriers. The function class of Royal Staircases with Ditches generalizes these fitness functions and is defined as follows:

- Genotypes consist of bit sequences of length L, interpreted as N blocks of K bits each: $L = NK$.

- The blocks are ordered, not in the sense that they correspond to particular positions in the genotype, but only in the sense that they are indexed 1 through N. Note that since our evolutionary process does not include recombination, the population dynamics is invariant under arbitrary permutations of a genotype's bits. For convenience, we order the blocks from left to right in the genotypes. That is, bits 1 through K belong to the first block, bits $K+1$ through $2K$ belong to the second block, and so on.

- The 2^K possible configurations of the K-bit blocks each are divided into three classes.

  1. Type-A blocks consist of a configuration with K ones:

  $$A = \underbrace{111\cdots111}_{K} \ . \tag{8.52}$$

  2. Type-B blocks consist of a configuration with $K-w$ ones and w zeros:

  $$B = \underbrace{111\cdots111}_{K-w}\overbrace{000\cdots000}^{w} \ . \tag{8.53}$$

  As will become clear, the parameter w controls the *width* of the barriers.

  3. All other $2^K - 2$ configurations are denoted as Type-$*$ blocks.

- A genotype s with blocks 1 through $n-2$ of type B and block $n-1$ of type A receives fitness $f(s) = n$. These genotypes have the structure:

  $$\underbrace{BB\cdots BB}_{n-2}A\overbrace{**\cdots**}^{N-n+1} \ . \tag{8.54}$$

  Note that the configurations of blocks n through N are immaterial (denoted $*$) when the first $n-1$ blocks occur in the above genotype configuration.

- Genotypes s with blocks 1 through $n-2$ of type B , block $n-1$ of type $*$, and block n of type A receive fitness $f(s) = n - h$. These genotypes have the structure:

$$\underbrace{BB\cdots BB}_{n-2} * A \overbrace{**\cdots**}^{N-n}, \tag{8.55}$$

  Again, the configurations of blocks $n+1$ through N are immaterial. The parameter h controls the *height* of the fitness barrier.

The Royal Staircase fitness functions that we studied in chapters 5 and 6 are a special case ($w = 0$) of the Royal Staircase with Ditches class of fitness functions. For the special case $w = 0$ there are no fitness barriers and a genotype has fitness n when the first $n-1$ blocks are $A = B$ (all 1s) types and the nth block is set to any of the $2^K - 1$ other configurations.

Setting $w = 1$ produces a somewhat degenerate case that we will not consider.

For values of $w \geq 2$, there is a genuine fitness barrier of width w bits. For instance, consider the case where, at some point in time, the highest fitness in the population is $f = 4$. This corresponds to genotypes that have first and second blocks of type B and a third block of type A. In this case, a *portal* genotype Ω_n, corresponding to a fitness of 5, is obtained when the fourth block is set to type A and the third block is changed from type A to type B. Genotypes with fitness 4 can mutate their fourth block, until it becomes type A, without changing their fitness. That is, the fourth block may be changed into type A along a *neutral path* and setting the 4th block correctly corresponds to crossing an entropy barrier. However, after that, the third block needs to be changed from type A to type B. All intermediate type $*$ blocks give genotypes a reduced fitness of $f = 4 - h$. We call these *ditch* genotypes, since they are located in a lower-fitness region in genotype space that separates genotypes with fitness 4 from genotypes with fitness 5.

8.5.2 Evolutionary Dynamics

We evolve populations on the Royal Staircase with Ditches under a simple selection and mutation dynamics similar to the one outlined in Sec. 8.2. This consists of the following steps.

1. At time $t = 0$ a population of M random binary-allele genotypes (bit sequences) of length L is created. These M individuals constitute the initial population.

2. The fitness of all M individuals is determined, using the function defined in the previous section.

3. M individuals are sampled from the population, with replacement, and with probability proportional to their fitness. That is, the population undergoes fitness-proportional selection in discrete generations.

4. Each bit in each of the M selected individuals is mutated with a probability μ. The M individuals thus obtained form the new generation.

5. The procedure is repeated from Step 2.

We evolve the population according to the above dynamics until genotypes of optimal fitness have been discovered and the population appears to have reached a stable average fitness. During each run, we estimate a number of statistics—such as, the average time until individuals of a certain fitness appear for the first time.

8.5.3 Observed Population Dynamics

The population dynamics under Royal Staircase with Ditches functions is qualitatively very similar to that under the Royal Road and Royal Staircase fitness functions. Samples of this typical behavior are shown in Figs. 8.5(a)-(d).

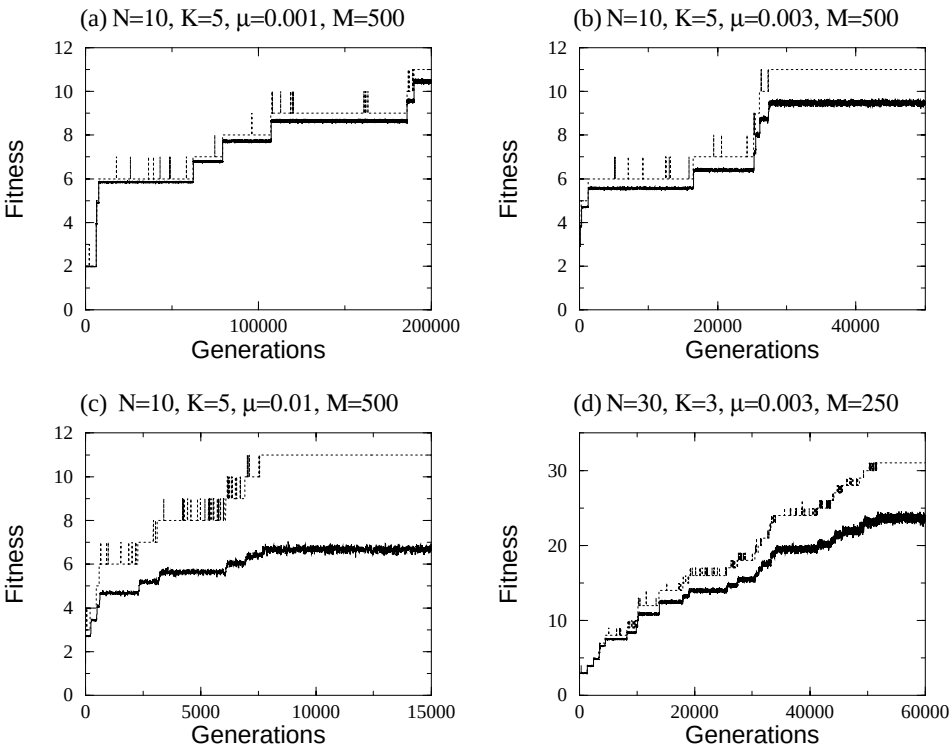

Figure 8.5: Four runs of the Royal Road with Ditches population dynamics with four different parameter settings. The upper dashed lines plot best fitness and the lower solid lines, average fitness $\langle f \rangle$ in the population as a function of time, measured in generations. The parameter settings for each run are indicated above each figure. All fitness functions have barriers with a width of $w = 2$ and a height of $h = 1$. The values of the average fitnesses were obtained by taking a running average over 10 generations. This reduces the relatively large fluctuations in average fitness between successive generations.

The plots there show the average $\langle f \rangle$ (lower, solid lines) and best (upper, dashed lines) fitnesses in the population over time for four single runs with four different parameter settings. The parameter settings for each run are indicated above each figure, except for the barrier widths w and barrier heights h that were used. All runs used barriers of widths $w = 2$ and heights $h = 1$.

The qualitative dynamics follows the typical alternation of long periods of stasis (*epochs*) in the average population fitness and short bursts of *innovation* to higher average fitness; a class of evolutionary dynamics that we call *epochal evolution* [137]. At the beginning of a run the average and best fitness are low. This simply reflects the fact that high-fitness genotypes are very rare in genotype space and therefore do not occur in an initial random population. A balance is quickly established between selection and mutation that leads to a roughly constant average fitness in the population. This period of stasis we call an *epoch*. After some time, a mutant may cross the fitness barrier and discover a *portal*, i.e., a higher-fitness genotype. Relatively frequently, this high-fitness mutant is lost through sampling fluctuations or deleterious mutations. Such events are seen as isolated spikes in the best fitness in Fig. 8.5. Eventually, one of these beneficial mutants spreads through the population—an *innovation* occurs. At this point the average fitness increases, until a new equilibrium between selection and mutation is established.

Although many properties of epochal evolution can be treated analytically (chapter 4), here we will focus solely on the *epoch times*. These are the average times between the start of a given epoch and the start of the next.

Average epoch times can be obtained from simulation data by tracing backward in time the behavior of the population's best fitness. The population dynamics runs until genotypes of the highest possible fitness $N + 1$ have established themselves in the population for an extended period of time. (For certain parameter settings—specifically, those beyond the error threshold—genotypes with the highest fitness may not stabilize in the population at all. For such parameter settings, an epoch time can be effectively infinite. We will not consider such parameter settings explicitly here.) From there we trace backwards the *last* time t_n that fitness n was the highest in the population, for all values of n. The differences $t_n - t_{n-1}$ give the epoch times τ_n. Some fitness levels may not occur during a run. For instance, in Fig. 8.5(a) fitness $f = 1$ never occurs, since the population starts out with genotypes of fitness 2. The epoch time τ_1 is therefore 0 for this particular run. Average epoch times are then estimated by averaging the epoch time τ_n over an ensemble of runs. In the following we calculate analytical approximations to these epoch times τ_n, using the results from the preceding development.

8.5.4 Epoch Quasispecies and The Statistical Dynamics Approach

In order to approximate epoch times analytically, we need to determine the average proportion of the population at highest fitness during each epoch. This is the equivalent of P_{II} in Sec. 8.3. From this we can estimate the rate of creation of genealogies in the *ditch* between the neutral networks of two successive epochs. Then we need to calculate the average population fitness during each epoch to determine the average life time of these ditch genealogies. For the fitness functions studied in the Sec. 8.3 these quantities were relatively straightforward to calculate. There, individuals were either on the peak or in the valley, at a certain distance from the portal. For the Royal Staircase with Ditches

fitness functions the situation is more complicated.

In principle, one can calculate the current equivalent of P_Π by representing the population as a distribution of *genotypes* and calculating metastable genotype distributions for each epoch. This is typically done in population genetics models [39] and in the standard quasispecies models of molecular evolution [33]. However, since genotype spaces are typically very large, an analytical treatment that explicitly takes into account finite-population effects is generally infeasible within this genotypic representation. To address this problem, we introduced an alternative approach that we call *statistical dynamics* [24, 137] and chapter 4. There one chooses a relatively small number of *macroscopic variables* with which to describe the population at any given time. Other degrees of freedom are then averaged out using a *maximum entropy* method similar to the Gibbs method from statistical mechanics. We will use this approach below and simply refer the reader to Refs. [24] and chapter 4 for more extensive treatment of statistical dynamics and a discussion of the relation of this approach to standard quasispecies theory, mathematical population genetics, and other theories from the field of evolutionary computation [117].

We represent the population state at any point in time by a *fitness distribution*. That is, instead of describing the population by the relative frequencies of all genotypes in the population, we describe it by the relative frequencies of different fitness values. Of course, a given fitness distribution does not uniquely specify a population of genotypes. In order to construct the dynamics on the level of fitness distributions, we must "average out" the additional genotypic degrees of freedom somehow. We do this by assuming, at each generation, a maximum entropy distribution of genotypes given the distribution of fitness. In the Royal Staircase with Ditches fitness functions, this translates into assuming that each string with a given fitness f is equally likely to be *any* of the genotypes with fitness f. That is, a genotype with fitness n will have its first $n-1$ blocks each in a specific state. Blocks n through N are assumed to occur in any of their 2^K possible states with equal probability. Similarly, for a "ditch" genotype with fitness $n-h$, there are $n-1$ blocks in fixed genotypic states and one block of type $*$. We assume this $*$-block occurs in any of its 2^K-2 possible bit configurations with equal probability. Similarly, we assume that blocks $n+1$ through N are equally likely to be in any of their 2^K configurations. That is, we assign equal probability to all genotype distributions that are consistent with the given fitness distribution. Since we know the dynamics on the genotype level, we can construct the expected dynamics on the level of fitness distributions under these assumptions.

Formally, we represent the population at time t as a vector $\vec{P}(t)$ with components $P_n(t), n = 1, 2, \ldots, N+1$ that denote the proportion of fitness-n individuals in the population and with components $P_{n,*}(t)$ that denote the proportion of fitness $n-h$ individuals in the ditch between fitness n and fitness $n+1$ genotypes. (Note that in the cases where h is an integer the distribution $\vec{P}(t)$ is not simply a fitness distribution, since it distinguishes ditch individuals from nonditch individuals at the same fitness.)

We then construct a generation operator $\mathbf{G}$, similar to the one in section 8.3.1, that acts on a fitness distribution $\vec{P}(t)$ and returns the *expected* fitness distribution $\langle \vec{P}(t+1) \rangle$

at the next generation. That is, the dynamics at the level of fitness is to be given by

$$\langle \vec{P}(t+1) \rangle = \frac{\mathbf{G} \cdot \vec{P}(t)}{\langle f \rangle} \ , \tag{8.56}$$

where $\langle f \rangle$ is the average fitness in the population.

During epoch n, in which the best fitness occurring in the population is n, there will be a roughly constant fitness distribution $\vec{P}^n$. These vectors $\vec{P}^n$ are the solutions of the fixed point equations $\langle \vec{P}(t+1) \rangle = \vec{P}(t)$ determined by Eq. (8.56). Once the operator $\mathbf{G}$ is constructed, the epoch distribution can be calculated quite easily. More specifically, in Refs. [137] and chapter 4 we showed that the metastable fitness distribution that occurs during epoch n is determined by projecting the operator $\mathbf{G}$ onto all dimensions with fitness smaller than or equal to n and calculating the principal eigenvector of this projected operator. Determining the epoch-n quasispecies reduces, in this way, to finding the principal eigenvectors of the matrix $\mathbf{G}$ restricted to components with fitness lower than or equal to n. In App. F the generation operator $\mathbf{G}$ for the Royal Staircase with Ditches is constructed explicitly, and analytical approximations to the epoch quasispecies distributions $\vec{P}^n$ are calculated as well. Since the expressions we find for the fitness distributions $\vec{P}^n$ are rather cumbersome and we will not give them here.

8.5.5 Crossing the Fitness Barrier

With the analytical expressions for the epoch fitness distributions $\vec{P}^n$ in hand, we can calculate the expected epoch times τ_n. An epoch ends via an innovation—a process that proceeds in two stages. First, a portal genotype of fitness $n+1$ is created. And second, this beneficial mutant spreads through the population, rather than being lost.

In order to calculate the average time until a portal genotype is discovered we calculate the probability P^{seed} that a *single selection plus mutation* from the current population seeds a new lineage that discovers the portal. That is, either by creating a new lineage in the ditch that discovers the portal or else by producing a jump mutation that becomes a portal genotype at once.

This calculation is very similar to that in Sec. 8.3. First of all, the portal is unlikely to be discovered by anything other than either a jump mutation from a fitness-n individual or by a mutation from a ditch genotype. Moreover, ditch lineages are very unlikely to be seeded by anything other than mutants of fitness-n genotypes. Therefore, we can write P^{seed} as

$$P^{\text{seed}} = \frac{n P_n^n (1-\mu)^{(n-2)K}}{f_n 2^K} \sum_{i=0}^{K} \epsilon_i \left(\mathrm{M}_{iw} - (1-\mu)^K \delta_{iw} \right) \ . \tag{8.57}$$

The first factor $n P_n^n / f_n$ gives the probability that a fitness n individual is selected. For the offspring of this individual to end up in the ditch, its nth block should be type A (thereby contributing the factor 2^{-K}) and its first $n-2$ blocks should be left undisturbed by mutation (corresponding to the factor $(1-\mu)^{(n-2)K}$). The terms within the sum give the probability that a valley lineage is seeded by mutation of the $(n-1)$st block at Hamming distance i from the portal. Finally, the factors ϵ_i give the probabilities

that a lineage, starting at Hamming distance i from the portal, discovers the portal. They are analogous to the ϵ_i of Sec. 8.3.2. Note that the term $\epsilon_0 \mathrm{M}_{0w} = 1 \cdot \mu^w (1-\mu)^{K-w}$ corresponds to a jump mutation from a fitness-n genotype directly to a portal configuration. Again, changes to blocks $n+1$ through N are immaterial.

The ϵ_i are calculated paralleling the development in Sec. 8.3.2. The only differences are that the genotype length L is now the block length K here, since we are only interested in mutations in the $(n-1)$st block, and that the average number of offspring is slightly modified. The average number of *ditch* offspring r that an individual in the ditch produces on average is given by

$$r = \frac{(n-d)(1-\mu)^{(n-1)K}}{f_n} = \frac{n-d}{n} \,. \tag{8.58}$$

The expressions for the ϵ_i in this case can be obtained by replacing the factor $1/\langle f \rangle$ in the equations in Sec. 8.3.2 by r.

Eq. (8.57) can be further simplified and we eventually obtain the result that

$$P^{\text{seed}} = -\frac{P_n^n}{2^K} \left[\epsilon_w + \frac{n}{(1-\mu)^K (n-d)} \log(1-\epsilon_w) \right] \,, \tag{8.59}$$

where ϵ_w is given by Eqs. (8.15) and (8.16) and P_n^n is determined as outlined in App. F.

Thus, once we have calculated ϵ_w using the results of Sec. 8.3.2 and calculated P_n^n using the derivation described in App. F, we find that the average number g_n of epoch-n generations until a genotype of fitness $n+1$ is created is given by

$$g_n = \frac{1}{1-(1-P^{\text{seed}})^M} . \tag{8.60}$$

In this, we have neglected the correction term $\langle dt \rangle$ of Sec. 8.3.4 for the time between the seeding of the ditch lineage that leads to the discovery of the portal and the actual time at which the portal is discovered.

We now must calculate the second stage of epoch termination; namely, the probability π_n that a newly discovered mutant of fitness $n+1$ spreads through the population. In chapter 4 we showed how a diffusion equation approach [89, 90] can be used to estimate this probability. Applying this here, the probability π_n that a mutant of fitness $n+1$ will spread through the population is given by:

$$\pi_n = 1 - \exp\left(-2 \frac{(n+1)(1-\mu)^K}{n} + 2 \right) \,. \tag{8.61}$$

A mutant must be discovered $1/\pi_n$ times on average before it stabilizes and spreads through the population. Thus, g_n/π_n gives the average number of generations until a portal genotype is discovered that spreads through the population—an innovation.

Finally, we have to account for the possibility that epoch n does not occur at all during a run. Only a fraction $P_e(n)$ of the runs contain epoch n, since, if a higher-fitness genotype occurs in the initial random population, epoch n will be skipped. The proportion $P_e(n)$ is simply the probability that *no* genotypes of fitness greater than n

occur in the initial population. This is given by

$$P_e(n) = \left[\sum_{i=0}^{n-1} \frac{1}{2^{Ki}} \left(1 - \frac{1}{2^K} \right) \right]^M . \tag{8.62}$$

Putting Eqs. (8.60), (8.61), and (8.62) together, the theoretical predictions for the epoch times τ_n become

$$\tau_n = P_e(n) \frac{g_n}{\pi_n} . \tag{8.63}$$

8.5.6 Theoretical and Experimental Epoch Times

We tested the predictions of Eq. (8.63) against experimentally obtained average epoch times for the same parameter settings used in Fig. 8.5. The results are shown in Fig. 8.6. Epoch times τ_n are shown as a function of epoch number n. The dashed lines give the theoretical predictions of Eq. (8.63) and the solid lines, the experimentally estimated averages. The data in Fig. 8.6(a) is an average over 150 evolutionary runs. All other plots are averages over 250 runs. The fluctuations in the experimental lines indicate that there is a large variance of epoch times between runs and that, therefore, the raw data is rather noisy. Despite this, the figure demonstrates that the theory estimates epoch times quite accurately.

First, these results show that the analysis presented in Sec. 8.3 can be usefully applied to more complicated fitness functions that posses many fitness and entropy barriers. For simplicity in the present case, we restricted ourselves to fitness functions with barriers of equal height and width. However, since our statistical dynamics methods analyze epochs in a piecewise manner, the theoretical results easily extend to more general cases with barriers of variable width and height. The essential ingredient of the analysis is still the genealogy statistics of valley lineages in the ditch that connects to portal genotypes. The only additional ingredients required by the analysis for more complicated cases are (i) the rate of creation of new lineages in the ditch and (ii) the distribution of Hamming distances to the portal at which these lineages are seeded. In the case of the Royal Staircase with Ditches, this was largely determined by the proportion P_n^n of fitness n individuals during each epoch n. Apart from this determining factor, the actual crossing of fitness barriers is still governed by the scaling relations presented in Sec. 8.3.6. In particular, the qualitative remarks at the end of Sec. 8.3.6 still hold in these more general settings. Epoch times grow very rapidly with barrier width and quite slowly with barrier height.

Second, and perhaps of more immediate interest, the general shape of the curves in Fig. 8.6 reveals several novel population dynamical phenomena. For the lower mutation-rate runs (Figs. 8.6(a), 8.6(b), and 8.6(d)), the epoch times show a distinct break between the relatively small τ_1 and the much larger τ_2. This jump simply reflects the fact that for these population sizes, it is very likely that individuals with fitness 2 occur in the initial random population. Due to this, epoch 1 is skipped most of the time. However, individuals of fitness 3 are unlikely to occur in the initial population—the occurrence of a fitness 3 individual is roughly one in 10^3—so that the second epoch is not skipped. In Fig. 8.6(a) and 8.6(b), the epoch time curve reaches a maximum after this quick jump

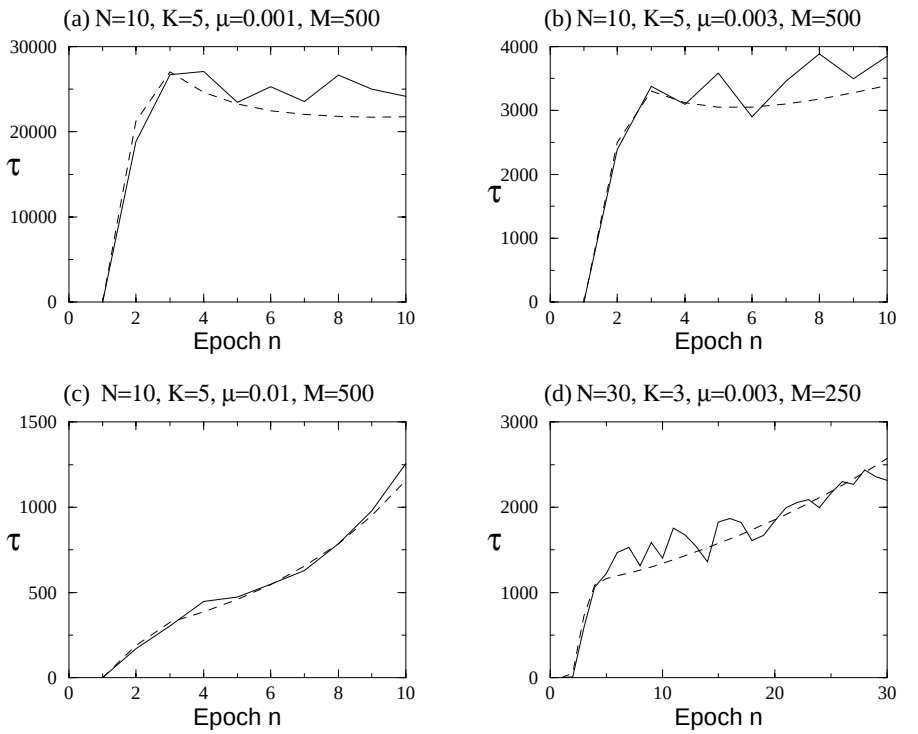

Figure 8.6: Experimentally estimated (solid lines) and theoretically predicted (dashed lines) epoch times τ_n for the four different parameter settings of Fig. 8.5 as a function of epoch number n. All fitness functions have ditches of width $w = 2$ and height $h = 1$. The experimental epoch times are an average over 150 runs for Fig. (a) and over 250 runs for Figs. (b), (c), and (d).

and then drops off slightly. In Fig. 8.6(a) with mutation rate $\mu = 0.001$ the times seem to just reach a minimum around the last epoch, number 10. The curves in Fig. 8.6(b) reach a minimum epoch time somewhere around epoch 5 and then start increasing again. The behavior typically occurs for a range of parameter values with low mutation rates and with block sizes K that are not too small. For instance, the theoretical analysis predicts that if there were more than 10 blocks in the case of Fig. 8.6(a), then the curve would start to rise after epoch 10.

Qualitatively, this behavior can be understood as follows. Since the barriers all have equal *height* h, the barriers at later epochs are relatively more shallow than those visited during early epochs. From the analysis in the last section recall that the average number of offspring that ditch individuals produce in the ditch is $r = (n-h)/n$. As n increases, this number approaches 1 from below. That is, ditch lineages survive longer for later epochs. This effect, of course, decreases epoch times, and this causes the initial decrease of epoch times in Figs. 8.6(a) and 8.6(b). However, the ditch lineages are seeded by mutants of fitness-n individuals. For later epochs, fitness n individuals have many

more bits, $(n-1)K$, that need to be set to specific configurations. They are, therefore, more likely to undergo deleterious mutations. The proportion P_n^n of fitness-n individuals during epoch n thus decreases as a function of n. This tends to increase the epoch times since smaller P_n^n implies a lower rate of seeding of ditch lineages. Moreover, the probability π_n that a genotype of fitness $n+1$, once found, spreads through the population decreases as n increases as well. Eventually, these effects start to dominate, and epoch times start rising again. In fact, at a certain point, for large n, the probability π_n may become so small that fitness-$n+1$ individuals cannot be stabilized in the population at all. In this regime, the dynamics again reaches the well known error threshold: the population dynamics cannot store the necessary nK bits of information that define epoch $n+1$.

In Figs. 8.6(c) and 8.6(d), there is no initial decrease of epoch times: the epoch times increase monotonically as a function of n. In this case, from the start of the runs the decrease of π_n and P_n^n with n dominates the effect of the relatively shallower ditches.

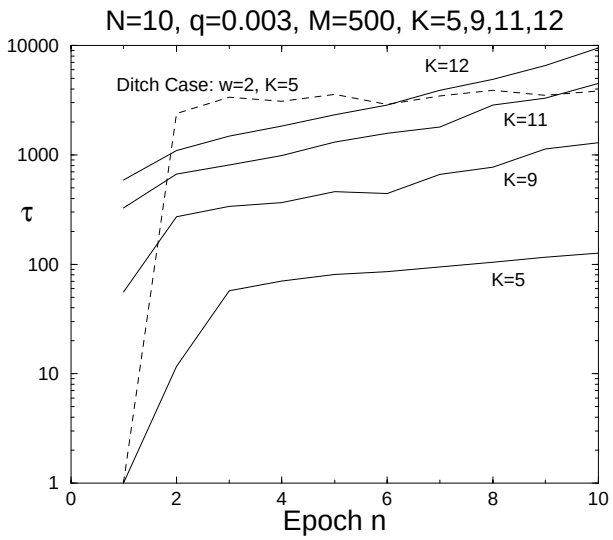

Figure 8.7: Comparison of the entropy- and fitness-barrier crossing times for the Royal Staircase with Ditches fitness function. The dashed line gives the epoch times for the case of *minimal width* ($w = 2$) ditches of height $h = 1$. This data is the same as that in Fig. 8.6(b). The solid lines plot the entropy-barrier epoch times for a fitness function without ditches ($w = 0$) for several different block lengths: $K = 5$, 9, 11, and 12. All other parameters are identical to the ditch-case parameters: i.e., $N = 10$ blocks, a mutation rate of $\mu = 0.003$, and a population size of $M = 500$. The vertical axis is shown on logarithmic scale.

Finally, we compare these epoch times, for ditches of height $h = 1$ and width $w = 2$, with the epoch times for the entropy-barrier case ($w = 0$) at different block lengths K. This comparison is shown in Fig. 8.7. Epoch times τ are shown on a logarithmic scale. The dashed line shows the experimentally obtained data for the ditch case with $N = 10$

blocks of length $K = 5$, a (minimal) ditch width of $w = 2$, and height of $h = 1$, a population size of $M = 500$, and a mutation rate of $\mu = 0.003$; as used in Fig. 8.6(b). The solid lines show experimentally estimated epoch times for the neutral case ($w = 0$), for several different block lengths. All other parameters are the same.

In comparing the solid line for $K = 5$ with the dashed line we see that the introduction of this minimal ditch increases the epoch times by *factors* ranging from 50 to 250. In order to obtain roughly comparable epoch times for the neutral case $w = 0$, one has to increase the block length to as high as $K = 12$. This demonstrates how much more rapidly entropy barriers are crossed than fitness barriers. In the time that it takes the population to cross a fitness barrier of width $w = 2$, the neutrally diffusing population will have crossed an entropy barrier of an additional $12 - 5 = 7$ bits. Thus, the neutrally diffusing population explores roughly $2^7 = 128$ times as many different neutral configurations in the time that it takes to cross the 2-bit fitness barrier. As we will see below, the difference in time scale for crossing fitness and entropy barriers grows much larger for more realistic (lower) mutation rates and (longer) sequences.

8.6 Conclusions

We analyzed in detail the barrier crossing dynamics of a population evolving under selection and mutation in a constant selective environment. Barriers of two distinct types exist: fitness barriers and entropy barriers. Fitness barriers occur in a fitness "landscape" when genotypes of higher fitness than currently present in the population are separated from the current most-fit genotypes by valleys of lower fitness. In order for such barriers to be crossed, a rare sequence of mutants must cross the valley of low-fitness genotypes. The second type of barrier, entropy barriers, occurs when genotypes of current best-fitness form large *neutral networks* in genotype space that only have a small number of connections (*portals*) to genotypes of higher fitness. Evolving populations diffuse at random through these neutral subbasins under selection and mutation. Since connections to higher fitness genotypes are rare, the population must search and spread over large parts of the neutral network, before higher-fitness genotype portals are discovered.

We will now qualitatively and quantitatively summarize the general picture that has emerged from our analysis of barrier crossing dynamics. The first important observation to be made is that there is a large qualitative difference between the genealogies of individuals of current best fitness and those with suboptimal fitness. All suboptimal individuals in the population are relatively recent descendants of genotypes with the current highest fitness. In other words, individuals with suboptimal fitness only give rise to genealogical bushes—genealogies of finite, typically short, length. Individuals with the current highest fitness are the only ones that give rise to lineages of potentially infinite length.

More formally, let us denote by Λ the neutral network of genotypes whose fitness equals that of the current highest-fitness individuals. The subpopulation of the current population that is on the neutral network Λ then effectively acts as a source for the whole population's descendants. Genealogies of lower-fitness individuals outside of Λ go extinct relatively quickly and are replaced by new genealogies that are mutant descendants of individuals on the neutral network Λ. Therefore, only lineages of individuals on the

network Λ can "travel" long distances through genotype space. The subpopulation of genotypes in Λ diffuses randomly through Λ, eventually visiting almost all genotypes in Λ and all of their single mutant neighbors.

At the same time, the excess reproduction of individuals in Λ, combined with deleterious mutations, creates individuals in the lower-fitness valleys or ditches around Λ. These individuals then give rise to genealogies of valley individuals that, in turn, probe at random the genotype space surrounding Λ. However, since these valley genealogies typically go extinct quite rapidly, individuals in the valley never travel far from the neutral network Λ. That is, only those parts of the valleys that are in the immediate neighborhood of the neutral network Λ will be quickly explored. Valley genealogies that cross a wide valley are very rare.

Therefore, on the shortest time scale, the population explores the neutral network Λ and its neighborhood, as shown in Fig. 8.2. If there are any higher-fitness genotypes in the immediate neighborhood of the network Λ, the population is most likely to escape from the metastable state (epoch) by crossing this entropy barrier. If the neutral network Λ is completely surrounded by valleys of lower fitness, the population will eventually escape from the metastable state when a mutant crosses one of the valleys that surrounds Λ, i.e., by crossing a fitness barrier. Since the waiting time for such a fitness barrier crossing increases very rapidly with the width of the barrier, it is most likely that the escape will occur along one of the narrowest valleys that connects Λ to higher fitness genotypes. Although shallow valleys are more easily crossed than deep valleys, the effect of barrier height is relatively small compared to the effect of barrier width.

This basic picture of the metastable population dynamics is captured more quantitatively by the scaling relations of Eqs. (8.33) and (8.49) for the fitness-barrier and entropy-barrier crossing times, respectively. By equating these, we get a rough comparison of the trade offs in the crossing of entropy versus fitness barriers. To make the comparison for more general cases, we replace the factor 2^L in the entropy-barrier scaling form, Eq. (8.49), with V_Λ to denote the volume of the neutral network Λ. We also replace L by the average *neutrality* ν—i.e., the average number of neutral neighbors—of genotypes in Λ. This extends the entropy-barrier scaling form to more general neutral network topologies than simple binary hypercubes. We then obtain

$$V_\Lambda = \frac{\nu}{w!}\left(\frac{\log(\sigma)}{\mu}\right)^{w-1} . \tag{8.64}$$

This is an estimate of the volume V_Λ of the network Λ that is explored in the time it takes the population to cross a fitness barrier of height σ and width w.

Recall that the factor $w!$ denotes the number of different paths leading across the fitness barrier and that $\mu/\log(\sigma)$ gives the average number of mutations valley lineages undergo before they go extinct. The relative size of μ and $\log(\sigma)$, together with the barrier width w, are the decisive quantities. Only when $\mu/\log(\sigma)$ approaches 1 can fitness barriers be crossed relatively quickly. However, since μ is typically very small, even in cases where the highest-fitness genotypes have only a small fitness advantage over valley genotypes—i.e., $\sigma = 1 + \epsilon$—the determining ratio μ/ϵ may still be small.

As an illustration, assume that each genotype in Λ has an average of $\nu = 60$ neutral neighbors, that the mutation rate is $\mu = 10^{-6}$, and that genotypes in Λ are on average 1

percent more fit than genotypes in the valley. Substituting these values into Eq. (8.64) for a barrier of width $w = 3$, we find $V_\Lambda \approx 10^9$. Thus, before this fitness barrier will be crossed, on the order of 10^9 genotypes on the neutral network Λ will have been explored. Of course, these numbers should not be taken as quantitative predictions, they are only rough order of magnitude estimates. However, they do indicate the difference in time scales at which entropy and fitness barriers are crossed.

In cases where the fitness advantage ϵ becomes so small as to be comparable in size to μ, the fitness barrier will have effectively turned into an entropy barrier. That is, for such small fitness advantages, selection is not able to stabilize the population within Λ and the population will freely diffuse through the valleys. In short, before ϵ becomes so small as to be comparable to μ, the population will have already crossed the error threshold. This can be seen, for instance, by taking the logarithm on both sides in Eq. (8.35).

As we discussed earlier, the picture of the barrier crossing dynamics that emerges from our analysis contrasts strongly with the common view that this process is analogous to the dynamics of energy-barrier crossing in physical systems. In those systems, the stochastic dynamics follows the *local* gradient of the energy landscape. This local character of the dynamics causes the barrier *height* to be the main determinant of the barrier crossing time. Generally, the evolutionary population dynamics cannot be described as following a local gradient. The current highest fitness of the individuals located on the neutral network Λ sets an absolute fitness scale against which valley individuals must compete. Therefore, valley lineages are unlikely to survive for many generations and so cannot undergo many mutations before going extinct. This mechanism causes the barrier width to be the main determinant of the fitness-barrier crossing time in population dynamics.

The preceding results demonstrate the important role of *neutral networks* in genotype space for evolutionary dynamics: where a population goes in genotype space is largely determined by the available neutral paths.

Acknowledgments

The authors thank Paulien Hogeweg and Dan McShea for helpful comments on the manuscript. This work was supported in part by NSF grant IRI-9705830, Sandia National Laboratory, and the Keck Foundation. EvN's investigations were also supported by the Priority Program Nonlinear Systems of the Netherlands Organization for Scientific Research (NWO).

9

Conclusions

At this point in time, the Darwinian theory of evolution is the sole theoretical paradigm that is potentially capable of explaining the diversity and complexity of the biological world around us. The mathematical formalization and analysis of evolutionary dynamics is, however, still at an early point in its development. As the genotypic content, and the gene and metabolic networks of simple organisms are being deciphered at an increasing pace, there is a growing need for mathematical approaches that are capable of describing evolutionary dynamics in a quantitative way. This thesis provides both conceptual and technical contributions towards this general goal.

On an abstract level, the thesis illustrates how a *maximum entropy* methodology may be used to infer the evolutionary dynamics on some macroscopic level of interest from underlying microscopic equations of motion on the level of genotypes. We defined a series of simple evolutionary systems and studied their emergent dynamical behaviors on the level of fitness distributions. In particular, we studied evolutionary dynamics in populations of individuals that evolve under selection and mutation in a constant selective environment. Individuals are represented by their genotypes, and a fixed fitness value is assigned to each possible genotype. Within this framework, we studied evolutionary systems that gave rise to what we have called *epochal evolution*.

In epochal evolution, the state of a population on the level of fitness or phenotypes alternates between long periods of stasis (epochs), and sudden transitions (innovations) to a different quasi-stationary state. In the majority of the evolutionary systems that we studied, this epochal behavior is the result of *entropy barriers* in the fitness function. Entropy barriers generally occur in highly degenerate genotype-to-fitness mappings—i.e. when many genotypes give rise to the same phenotype or fitness. In these situations, the genotype space may be conceived as decomposing into a number of *neutral subbasins*: sets of iso-fitness genotypes that are connected via single mutational steps. Since neutral subbasins of high fitness are generally much smaller than subbasins of low fitness, populations may wander around lower-fitness subbasins for long times before embarking on a connection (portal) to a subbasin of higher-fitness. That is, the population must cross an *entropy barrier* before it moves to a higher-fitness state. During this time, the fitness or phenotype distribution typically fluctuates around a stationary distribution.

The precise quantitative behavior of this type of epochal evolution is the main topic of this thesis. We have called our analytical approach *Statistical Dynamics* since we are studying dynamics on a macroscopic level by a statistical inference method similar to the one used in statistical mechanics. Chapters 2 and 3 gave a conceptual introduction to

this general technique and showed how the *maximum entropy* methodology can be used in an evolutionary context. It appears that the maximum entropy methodology has the potential to be successfully applied to much more realistic evolutionary systems than the ones studied in this thesis. In fact, if one accepts that the actual problem of understanding evolutionary behavior reduces to inferring higher-level behavior from underlying rules, it becomes clear that this problem can *only* be approached through the maximum entropy methodology. It is the only unbiased way of constructing a higher-level description of a system's behavior from its underlying microscopic dynamics.

9.1 Summary of Results

Chapter 4 presented the first and most thorough application of our statistical dynamics analysis to a model evolutionary system that shows epochal evolution. We studied the dynamical behavior of a population evolving under selection and mutation on a class of fitness functions that are called *Royal Road* functions. The main important feature of these fitness functions, with respect to the evolutionary dynamics that they generate, is that they contain a series of entropy barriers. Using the statistical dynamics approach we analyzed, in particular, how changes in parameters give rise to a variety of qualitatively distinct dynamical behaviors for a population evolving on these fitness functions.

The main results of the analysis can be summarized as follows: We first constructed a generation operator $\mathbf{G}$ that describes the evolutionary dynamics on the macroscopic level of fitness distributions in the limit of large populations. We then showed how this generation operator also controls the dynamics in the finite-population case. From this, we demonstrated that the locations in fitness distribution space where the metastable epochs occur, and the values of their average fitnesses, are given by the *spectrum* of this generation operator.

Next, we analyzed the dynamics in and between different epochs by linearizing the dynamics around each epoch fitness distribution. The spectra of the Jacobian matrices of $\mathbf{G}$ around these epoch locations are also given in terms of the spectrum of the generation operator $\mathbf{G}$. In particular, the stability of an epoch is determined by the average fitness in the population that is associated with it, relative to the average fitnesses associated with other epochs. Innovations from one epoch to the next take on the shape of logistic growth curves on the level of fitness. The fluctuations in the fitness distribution during the epochs are approximately Gaussian. The latter are calculated using a diffusion equation approach.

The stability of these epochs can be calculated in a similar way. We derived the average waiting time before an epoch *destabilizes*. An epoch destabilizes when *all* individuals of the current highest fitness are lost through a sampling fluctuation. This behavior occurs typically at parameter settings for which consecutive epochs have only slightly different average fitnesses associated with them, and when population sizes are small. Our destabilization results are directly related to error-thresholds in the theory of molecular evolution and they predict precisely the evolutionary behavior for finite populations around an error-threshold.

Finally, chapter 4 provided some approximations of the average times that the population spends in each epoch. Comparison with experimental results from computer

simulations showed that these approximations underestimate the actual epoch times. We discussed why our current theory fails to exactly predict this statistic. On a more qualitative level, the analysis in this chapter showed and explained how under simple changes of the evolutionary parameters, the evolutionary dynamics on the same class of fitness functions may show radically different qualitative dynamical behaviors.

Chapters 5 and 6 explored how the statistical dynamics analysis of the preceding chapter can be applied to optimize epochal evolutionary search. That is, in these chapters we focused on the evolutionary dynamics from a search-theoretic point of view and analyzed how to set evolutionary parameters such that the time spend in each epoch is minimal. The fitness functions in these chapters are closely related to the Royal Road functions studied in chapter 4 and are called *Royal Staircase* fitness functions.

Chapter 5 focused mainly on optimizing the mutation rate. The result was that an optimal mutation rate is obtained by balancing two opposite effects of mutations. On the one hand, increasing the mutation rate increases the exploration rate of populations through the neutral subbasins. On the other hand, increasing the mutation rate also increases the rate of deleterious mutations to inferior subbasins, which results in turn in smaller populations in the subbasin of current best fitness. The optimal mutation rate occurs when these two effects are balanced such that the *total explored volume* in the neutral subbasin per generation is maximized.

In chapter 5 we also went into more detail regarding the role of genetic crossover in epochal evolution. To this end, we derived statistics on genealogies of individuals in different fitness classes during an epoch. The main result was that *all* individuals in the population have a relatively recent ancestor from the neutral subbasin with the current highest fitness. This implies that all individuals in the population are genotypically closely related to individuals in the subbasin of current highest fitness. In other words, none of the individuals in the population ever wanders far from the subbasin of current highest-fitness genotypes. We also derived the scaling of the minimal number of fitness function evaluations to reach the global optimum and compared this scaling with the scaling for simple hill-climbing algorithms.

Chapter 6 analyzed in addition optimizing the population size for the class of Royal Staircase fitness functions. The dominant population size effect turned out to involve the destabilization of epochs. When the population size gets too small, the current highest fitness strings can easily disappear from the population through sampling fluctuations, making it hard for the population to discover genotypes of even higher fitness. The analysis showed that the search is optimal when each epoch is only *marginally stable*. That is, in order to reach the global optimum as fast as possible, the population size should be turned so low as to make each epoch close to destabilizing. A two-dimensional *search effort surface* was also presented that shows the average number of fitness function evaluations to reach the global optimum as a function of both population size and mutation rate. This surface succinctly illustrated the generalized error-threshold that bounds the area of efficient search in the space of evolutionary parameters. Finally, this chapter also discussed how the parameters of the evolutionary search induce an effective coarse graining of the fitness function, and how this fact may be essential in determining which kind of fitness functions can be efficiently searched by an evolutionary algorithm. We conjectured that fitness functions that can be efficiently searched by an evolutionary search algorithm have the property that there is a level of coarse graining at which their

local optima disappear and all genotypes are in the basin of attraction of the global optimum.

In chapter 7 we then focused on the dynamics of a population within a single neutral subbasin or *neutral network*, as it is generally referred to in the theory of molecular evolution. Up to this chapter, the neutral subbasins in the fitness functions we studied were all of a simple type: a region in genotype space where each genotype has a constant number of single-mutant neighbors within the same subbasin. In this chapter, we studied the evolutionary dynamics for subbasins with arbitrary topology, i.e. the subbasin may be an arbitrary graph in genotype space. The main result is that the population spreads through the subbasin in a manner that, in the limit of long times, only depends on the *topology* of the subbasin or neutral network and is *independent* of evolutionary parameters, such as mutation rate and population size. Additionally, we showed that the population tends to evolve towards those parts of the subbasin that have the highest *robustness* under mutations. That is, the population evolves towards parts of the subbasin where many of the single-mutant neighbors are in the same neutral subbasin. The extent to which this mutational robustness evolves is thus found to be a function of the subbasin topology only. In this way, one can classify subbasin topologies by the amount of mutational robustness that they generate in evolving populations. We also indicated how *in vitro* experiments with evolving populations allow one to infer the structure of neutral subbasins of biomolecules from simple population statistics.

Finally, in chapter 8 we focused on the dynamics of crossing fitness barriers and compared it with the crossing of entropy barriers that was studied in most of the preceding chapters. First, we studied the dynamics of the population crossing a single fitness barrier. That is, we studied the evolutionary dynamics in a fitness function that has a single local optimum of tunable height and a single *portal* genotype at some tunable distance from the local optimum. We analyzed the average waiting time before a rare mutant reaches the portal genotype by crossing the intermediate fitness valley. Using a branching process approach, we derived accurate approximations to the fitness-barrier crossing times as a function of mutation rate, population size, and the height and width of the fitness barrier. The main result was that the barrier crossing time scales only very slowly with barrier height (as a power of its logarithm) and very fast with barrier width (almost exponentially). The barrier crossing time also scales as a power-law in both population size and mutation rate. We noted that this scaling of the fitness-barrier crossing time contrasts strongly with the typical scaling found for barrier crossing in physical systems—such as, the crossing of an energy barrier. The scaling results illustrate particularly clearly why the *fitness landscape* metaphor, that is so often used in the theory of evolution, can be very misleading.

After that, we showed that a fitness barrier turns into an entropy barrier rather suddenly as the evolutionary parameters reach the error threshold. We derived accurate predictions for the location of the finite-population error threshold by using a diffusion equation approximation of the sampling fluctuations in the quasispecies distribution. We then studied the barrier-crossing dynamics in the entropy-barrier regime. The main result here was that the entropy-barrier crossing time shows interesting anomalous scaling with both population size and mutation rate. In particular, the entropy barrier crossing time increases faster than $1/\mu$ with increasing mutation rate μ and slower than $1/M$ with increasing population size M. Unfortunately, at this point, we only have a qualitative

understanding of this anomalous scaling.

In the second half of chapter 8 we then combined the fitness-barrier crossing results with the statistical dynamics results of previous chapters to derive epoch durations for more complicated fitness functions that possess both fitness and entropy barriers. This wide class of fitness functions we have called the class of Royal Staircase with Ditches. Comparison of our theoretical predictions with data from simulations showed that our theory accurately predicts epoch times in this class of fitness functions. Additionally, the analysis showed that the crossing of entropy barriers proceeds on much faster time scales than the crossing of fitness barriers. The chapter ended with a general comparison of the scaling of fitness- and entropy-barrier crossing times.

9.2 Integrative Unfolding View

Having summarized the main results like this, the discussion below presents a more integrative and qualitative picture of the understanding of evolutionary dynamics that has emerged through this work and speculates rather freely on possible directions of future work.

Intuitively, evolutionary systems seem to be set apart from other kinds of dynamical systems in that their dynamics appears *open-ended*. The evolutionary process takes place in an ever changing environment that, to a certain extent, is generated by the evolutionary process *itself*. Each time new organizational structures have appeared through an evolutionary innovation, these structures seem to consequently act as a substrate for further evolutionary innovations—see [130] for a discussion of such major evolutionary innovations. This feature lends the evolutionary process its open-ended character.

In contrast, the typical formalization of dynamical systems by defining a state space, and equations of motion in this state space, seems to preclude such an open-ended character. By defining a fixed state space for the evolutionary system at the outset, the range of its possible dynamical behaviors is predetermined. The evolutionary dynamics is so to speak "put in a box" from which it can never escape. However, the tendency to move beyond its current organizational structures is precisely the feature that the evolutionary dynamics exhibits, intuitively speaking: eventually new structures appear that make the process break out of its current level of description and that move it to a new level of organization at which essentially new types of dynamics take place. In this sense, defining a state space at the outset seems fundamentally ill-conceived for a general description of evolutionary dynamics.

The problem with this observation is, however, that there seems no mathematical alternative for formalizing dynamics. It is one thing to observe that the ordinary formalization of dynamical systems will not suffice, but another to provide a workable alternative. It is hard to imagine a mathematical formalization of a dynamical system that does *not* involve some state space and equations of motion. Moreover, rejecting altogether a description of evolutionary systems in terms of state spaces and equations of motion seems logically inconsistent with the fact that the fundamental laws of physics *are* defined as equations of motion in a state space. In principle, then, biological systems should at some level be describable by the laws of physics and therefore, by equations of motion in a fixed state space—which is, admittedly, very large and complicated. Thus,

we are faced with the problem that evolutionary systems are in principle described, at some level, by equations of motion in a fixed state space while, at the same time, such a description seems fundamentally at odds with the nature of evolutionary processes.

The way out of this of this difficulty is in the addition "at some level" in the previous sentence. We tend to think of the evolutionary dynamics of biological systems on a relatively high level of representation, far removed from underlying physical laws. This situation is not conceptually different from the situation found in physics. As explained in chapter 2, many of the common physical laws are defined at much higher levels of representation than the underlying fundamental laws—i.e. Ohm's law is defined in terms of current and resistance, while the fundamental law of quantum electro-dynamics is defined in terms of quantum fields. This should make it clear that by moving the representation of the dynamics to a much higher level, the apparent nature of the dynamics may be fundamentally altered.[1] Similarly, the nature of evolutionary dynamics as open-ended is not a feature of the underlying microscopic dynamics, but only an *emergent* feature on a higher level of representation. Thus, what is needed to understand evolutionary dynamics is a general methodology that starts with the underlying microscopic equations of motion and infers the dynamical behavior on a higher level of representation. Using such a methodology, one would *infer* that the dynamics has such qualitatively different features on a higher level of representation.

The analysis of simple evolutionary systems in this thesis shows how such a situation may come about in general. At the microscopic level, evolving populations move through a relatively simple state space of genotype distributions. The microscopic equations of motion that describe this movement are the result of selection and mutation at the level of genotypes. However, we did not study the dynamics on this microscopic level of genotypes, but instead described the dynamics on the level of fitness distributions. On this level, the dynamics is *not* described by equations of motion in a fixed state space. Instead, the dynamics shows a process of evolutionary *unfolding* of its own macroscopic state space. Below, I will first summarize how this picture of the dynamics emerges for the systems studied in this thesis and will then discuss how such a view may be applicable to more general situations.

As already noted, the key ingredient of the fitness functions studied in this thesis is that they divide the genotype space into a relatively small number of neutral subbasins. Additionally, subbasins of high-fitness genotypes are much smaller than subbasins of low-fitness genotypes. Because of this, a population that is randomly distributed in genotype space is likely to contain genotypes in lower-fitness subbasins only. Such a population will have a fitness distribution that contains few nonzero components—i.e., those corresponding to the large lower-fitness subbasins. On the level of fitness distributions, one only needs a few macroscopic variables to describe the population's state. As was explained in detail in the previous chapters, the population diffuses through these low-fitness subbasins while the fitness distribution fluctuates around a fixed point in its macroscopic state space. This diffusion proceeds until an individual in the population discovers a rare *portal* to a subbasin of higher fitness. When this has occurred, a new macroscopic degree of freedom is created in the fitness distribution. That is, a new

[1] Another example is the emergent irreversible macroscopic behavior in thermodynamics from underlying reversible laws.

macroscopic dimension *unfolds*.

As members of the population diffuse through the neutral subbasins, their genotypes are undergoing stepwise genetic modifications that are selectively neutral. They randomly move through different combinations of the values in their microscopic variables—somewhat akin to a burglar trying to open a safe by dialing different combinations at random. A *portal* genotype, then, corresponds to a rare combination of the microscopic variables that confers new functionality and fitness. The evolutionary process then capitalizes on this "jackpot" combination and it will be frozen into its macroscopic state.

Viewed in this way, it should be clear that this picture can be extended to much more general settings. In particular, we do not have to imagine that the microscopic state space and equations of motion are all given from the outset. Instead, we may think of the microscopic dynamics as occurring in a background of a *sea of microscopic variables*. Evolutionary innovations then correspond to the *crystallization* of macroscopic variables in this microscopic sea. This metaphor regards the macroscopic state of the "frozen" functionality-conferring combinations of microscopic variables that the evolutionary process has capitalized on as *static*, while it regards the remaining sea of microscopic variables as a *fluid* of degrees of freedom that are not yet frozen into functionality-conferring combinations.

During each epoch, the macroscopic state of the population occurs in some simple dynamical attractor, such as a fixed point. This macroscopic state of the population is the result of a historical sequence of evolutionary innovations that occurred through the discovery of "jackpot" combinations in the microscopic variables. Thus, one may envision this macroscopic state as a "crystal", corresponding to the combinations of microscopic variables that have "crystallized" in the past. Pursuing this metaphor, the macroscopic crystal then defines a *surface* in the sea of microscopic variables on which new functionality-conferring combinations of microscopic variables may crystallize. The current macroscopic state of the population thus effectively determines which regions of the microscopic variable-sea can be reached by the evolutionary process. That is, it defines an *envelop* of microscopic variables through which the population may diffuse—for instance, under the influence of random genetic modifications.

In this way, the evolutionary dynamics may be studied *epoch-wise*. Each epoch is defined by a macroscopic state. This macroscopic state then defines an envelop of microscopic variables through which the population may diffuse. The dynamics on the microscopic level defines how the population diffuses through this envelop. An epoch ends when the population has embarked on a portal within this envelop of microscopic variables. This portal combination of microscopic variables then crystallizes on the macroscopic state, adding a new branch to the macroscopic crystal. This new macroscopic state defines a new epoch, and the process repeats itself—but with a new envelop of microscopic variables and a new set of portals in this envelop.

The key step in this procedure is, of course, the construction of the microscopic state space that envelops the current epoch. This step will be discussed shortly. The following paragraphs clarify how this view of epochal evolutionary dynamics naturally accommodates such concepts as *frozen accidents* and *structural contingency* that are often used in qualitative reasoning about evolutionary processes.

Frozen accidents refer to persistent phenotypic features that have been selected at

random from a range of structurally different alternatives by evolutionary events in the past. One imagines that through essentially random events, one phenotypic alternative gains a small advantage over others, which is then amplified up to the point where this feature is "frozen in". In the view of epochal evolutionary dynamics sketched above, frozen accidents simply occur when the envelop of microscopic variables, through which the population diffuses during an epoch, contains structurally different portals. Which of these portals gets discovered first is essentially random, since the diffusion dynamics through the envelop of microscopic variables is itself random. However, once a portal is discovered, and a new macroscopic dimension has unfolded, it may be unlikely to collapse. In this way, the random diffusion determines which of a set of mutually exclusive new macroscopic dimensions may unfold.

Within our view, *structural contingency*—the dependency of future evolutionary events on current structural constraints—corresponds simply to the fact that the current set of macroscopic variables determines what regions of the microscopic sea of variables may be reached. The possibilities for future innovations are therefore dependent on the current macroscopic state as well. This brings us back to the problem of determining, based on the current macroscopic variables, which regions in the sea of microscopic variables envelop the current population.

9.3 Neutrality in the Sea of Microscopic Variables

The work in this thesis has shown that *neutrality* is an important player in determining which regions in the microscopic state space may be visited given the current macroscopic state of the population. As shown in chapter 5, all individuals in the evolving population have a relatively recent ancestor in the subbasin of current highest fitness. In this sense, the subbasin of the current highest fittest is the genealogical *source* for all individuals in the population. On the microscopic level of genotypes this means that every individual in the population is closely genetically related to an individual from the highest-fitness subbasin. In other words, only the parts of genotype space that immediately neighbor the subbasin of current highest-fitness can be easily visited by the population. Genotypes far removed from the subbasin of current highest fitness cannot be reached by the evolving population. This was shown more quantitatively in chapter 8, where we saw that the crossing of fitness barriers takes place on much longer time scales than the crossing of entropy barriers—i.e., the exploration of the current highest-fitness subbasin. In particular, the fitness-barrier crossing times become astronomically large for portal genotypes that are far removed from the current highest-fitness subbasin. Thus, for the simple evolutionary systems that we have studied, the envelope in the sea of microscopic variables that the population may explore is essentially determined by the space of possible *neutral* genetic variations.

It appears that this may be more generally the case. As long as the selective environment is relatively constant, selection restricts the movement of the population in genotype space to regions where individuals can go without lowering fitness. To construct the sea of microscopic variables that envelops the current epoch, we essentially have to construct the microscopic state space of selectively neutral genetic variations on the current highest-fitness genotypes. In more general cases, these microscopic state

spaces may be much more complex. For example, genotypes do not need to have fixed size: gene duplication and deletion events may lead genotypes to grow and shrink. It is easy to imagine that a gene duplication event on a particular genotype tends to leave the genotype in its neutral subbasin. These duplicated genes may then undergo further neutral mutations, until they embark on a functional gene that introduces novel functionality for the organism.

In summary, we envision epochal evolutionary dynamics in more general cases as still governed by a set of macroscopic variables that describe the evolutionary attractor on a higher level while, at a microscopic level, evolution is exploring neutral variations in a microscopic state space. At the same time, short-lived genealogies of deleterious mutants are spawned in the areas of the microscopic state space that neighbor the space of neutral variants. These lower-fitness mutants never wander very far from the subbasin of neutral variants. Innovations occur when the exploration of the space of neutral variants has embarked on a variant that confers new functionality. The evolutionary dynamics then capitalizes on these new phenotypic variants, a new macroscopic dimension is unfolded in the macroscopic state space, and the population dynamics stabilizes on a new macroscopic attractor. This new attractor then forms the substrate that determines the basins of neutral variations and, thereby, the potential paths to further innovations.

In future work I hope to extend these ideas to more realistic situations in which the sea of microscopic variables consists of the underlying *physical* degrees of freedom in the environment. It is tempting to speculate that evolutionary innovations can then be understood in terms of the discovery, by the evolutionary process, of novel interactions with the physical microscopic degrees of freedom. An evolutionary innovation would, for instance, consist in the discovery of a way to channel energy that is stored in certain degrees of freedom in the environment to "functional" degrees of freedom in the organism—for instance, when an energy flow in the environment is channeled towards the production of ATP molecules, as opposed to exciting vibrational degrees of freedom of a cell membrane. From the organism's point of view, such an innovation would correspond to the transformation of *heat* into *work*.

A

Information theory

In this appendix we will review the simplest concepts and definitions of information theory. An extensive and excellent textbook on information theory is [19]. Some inspiration for the presentation below was provided by [40].

A.1 Axiomatic Definition of Entropy

We want to construct a function $H(p_1, p_2, \ldots, p_n)$ that takes some probability distribution $\{p_i\}$ and assigns to it a real number which somehow quantifies the amount of "uncertainty" which is contained in this distribution $\{p_i\}$. As was first shown by Shannon [127], one only needs to assume a few reasonable properties for this entropy function to determine it uniquely up to a multiplicative constant. These assumptions are

1. The function $H(p_1, p_2, \ldots, p_n)$ is continuous in all its variables.
2. The entropy of the uniform distribution on n variables

$$h(n) = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right), \tag{A.1}$$

   is monotonically increasing with n.
3. If there are several ways of interpreting the same distribution, then calculating the entropy in correspondingly different ways should all lead to the same answer.

The first assumption (or axiom) formalizes the intuition that a tiny change in probabilities should result in only a small change in the uncertainty. Or more formally, that one can always make the changes in the p_i so small that the change in uncertainty becomes very small as well.

The second assumption seems almost trivial. It says that if two experiments both have a range of equally probable outcomes, then the highest uncertainty is associated with the experiment that has the *largest* range of outcomes. For instance, the uncertainty in the outcome of throwing a dice is larger than the uncertainty in the outcome of flipping a coin. It seems hard to imagine an uncertainty function that would not obey this assumption.

The third assumption is the vaguest and strongest of the three. It can be illustrated by some simple examples. It is often the case that the same experiment and distribution

can be interpreted in different ways. If one throws two coins, there are four possible outcomes, but these outcomes can be naturally partitioned in different ways. For instance, one can first look at the probabilities p_0, p_1, and p_2 that there will be zero, one, or two heads thrown, and then, in the case where there is one head, further give the probabilities that the first was heads and the second tails and vice versa. Another way of partitioning would be to say that there are probabilities p_h and p_t that the first coin is heads or tails, and then there are conditional probabilities that the other coin will be heads or tails given the outcome of the first (since we don't want to exclude "correlated" coins).

Let's assume that some experiment has four possible outcomes a, b, c, and d and that they have corresponding probabilities p_a, p_b, p_c, and p_d. The entropy function assigns some entropy $H(p_a, p_b, p_c, p_d)$ to this probability distribution. We can now partition the four possible outcomes in two classes x and y with probabilities p_x and p_y of occurrence. We'll assume that a and b fall in class x and c and d fall in class y. Demanding that, viewed in this partitioned way, the uncertainty is still the same, we obtain the functional equation:

$$H(p_a, p_b, p_c, p_d) = H(p_x, p_y) + p_x H\left(\frac{p_a}{p_x}, \frac{p_b}{p_x}\right) + p_y H\left(\frac{p_c}{p_y}, \frac{p_d}{p_y}\right). \tag{A.2}$$

Probabilities such as p_a/p_x give the conditional probabilities, such as in this case, the probability that a occurs given that x has occurred. Notice that in writing the contribution of the uncertainty of either x or y occurring and the conditional uncertainties of a, b, c, and d occurring given the occurrence of either x or y, we have implicitly assumed that these uncertainties *add* to give the full uncertainty. One could think of this as an additional assumption. It implies, for instance, that if an experiment is repeated, that the total uncertainty for both experiments is double the uncertainty of a single experiment (provided they are independent). This seems reasonable as well.

Assuming that the uncertainty should be invariant under any repartitioning of the possibilities, one obtains generalized versions of the equation (A.2). From these, and the first two assumptions, one can derive the functional form for the entropy function [87, 127]. To give the general flavor of this derivation, assume that we have an experiment with a large number of N outcomes that all have a probability $1/N$ of occurring. We can now divide these N outcome into m sets, with n_i elements in each set i. The probabilities p_i of the outcome falling in set i are of course given by $p_i = n_i/N$. We then have

$$h(N) = H(p_1, p_2, \ldots, p_m) + \sum_{i=1}^{m} p_i h(n_i), \tag{A.3}$$

where $h(n)$ is again the entropy for a uniform probability distribution over n possible outcomes. If we choose all n_i equal, i.e. $n_i = n = N/m$ we obtain

$$h(mn) = h(m) + h(n). \tag{A.4}$$

Obviously, this equation is satisfied by $h(n) = k\log(n)$, where k is an arbitrary constant. Using this form and moving the second term on the right hand side of equation (A.3) to the left, we obtain the general functional form of the entropy function

$$H(p_1, p_2, \ldots, p_m) = -k\sum_{i=1}^{m} p_i \log(p_i), \tag{A.5}$$

where one can think of the constant k as determining the base of the logarithm. In information theory, one often takes the logarithm base two, so that the entropy is measured in *bits*.

The above argument is of course no proof that the form (A.5) is unique. It only tries to clarify where the specific form (A.5) comes from. To show that this form is unique, one needs to use the first two axioms as well. The reader is again referred to [87, 127]. The entropy H is nonnegative, i.e. $H \geq 0$. It is zero only for a distribution that has all p_i are zero except for one. It is maximal for the uniform distribution $p_i = 1/m$. Finally, note that by continuity, the contribution $p_i \log(p_i)$ of a zero probability term $p_i = 0$ is set to zero.

A.2 Average Number of Yes/No Questions

Apart from being determined by a small set of basic assumptions, as shown in the previous section, entropy can also be regarded as the average number of yes/no questions that need to be answered to uniquely determine the outcome of an experiment which has some probability distribution associated with it. This can be shown using the following construction.

Assume that an experiment can have n outcomes with probabilities p_i, for each outcome i. We divide the n possibilities in two sets x_0 and x_1, in such a way that both have a total probability of around one half. We then ask which of the two sets x_0 and x_1 the outcome belongs to. Say that this is set x_0. We then divide set x_0 into two sets $x_{0,0}$ and $x_{0,1}$ that again both have probabilities as close as possible to one half. Again we ask which set the outcome belongs to, further reducing the possibilities. We repeat this process until we are left with a unique outcome. It can be shown that, averaged over many repititions of this question and answering game, the number of questions that we have to ask before getting to a unique outcome lies between the entropy and the entropy plus one:

$$H \leq \langle \#questions \rangle \leq H + 1, \tag{A.6}$$

where the entropy is calculated in bits.

This further strengthens our understanding of entropy as a measure of uncertainty. It is important to realize that it is, to some extent, the *unique* measure of uncertainty. Because of this, it is an invaluable tool for statistical inference as discussed in chapter 2. Because uncertainties of distributions can be compared on a single scale, we can prevent ourselves from making unwarranted assumptions in our inferences by using the distribution with the largest entropy.

A.3 Mutual Information

The functional form (A.5), together with the rules of probability theory in principle define entropy in any setting where there is a discrete set of possibilities. For clarity, we show some extensions of the form (A.5) in cases where the possibilities are partitioned in a certain way.

If we partition the outcomes in possibilities for two variables x and y, that take on the values $\{x_i\}$ and $\{y_i\}$, then the *joint entropy* for the variables x and y becomes

$$H(x,y) = -\sum_{i,j} p(x_i, y_j) \log\left[p(x_i, y_j)\right]. \tag{A.7}$$

Where $p(x_i, y_j)$ is the joint probability of both x_i and y_j occurring[1].

We can also define *conditional* entropies which give the uncertainty in the outcome of variable x given that we know the outcome of the variable y. If $p(x_i|y_j) = p(x_i, y_j)/p(y_j)$ denotes the probability that x_i occurs given that y_j has occurred, then the conditional entropy of x given y is

$$H(x|y) = -\sum_{i,j} p(x_i, y_j) \log\left[p(x_i|y_j)\right]. \tag{A.8}$$

Notice that in both the joint as in the conditional case we have for simplicity used the notations $H(x,y)$ and $H(x|y)$ that suggest that the entropy is a function of the variables. It is, of course, a function of the *probability distribution* over these variables. The meaning of the conditional entropy is that in repeated experiments, the uncertainty left in the variable x after obtaining information about the variable y is given by $H(x|y)$. If x is completely determined by y we have $H(x|y) = 0$. In general we have that $H(x) \geq H(x|y)$, since knowledge of y's value cannot possible *increase* the uncertainty in the value of x.

Finally, we define the *mutual information* $I(x,y)$ between two variables x and y:

$$I(x,y) = \sum_{i,j} p(x_i, y_j) \log\left[\frac{p(x_i, y_j)}{p(x_i)p(y_j)}\right]. \tag{A.9}$$

Notice that $I(x,y)$ is zero when x and y are independent. The meaning of the mutual information becomes clearer if one considers the following equalities, that can all be obtained by simply writing out the definitions of the corresponding quantities:

$$\begin{aligned} I(x,y) &= H(x) - H(x|y) \\ &= H(y) - H(y|x) \\ &= H(x) + H(y) - H(x,y). \end{aligned} \tag{A.10}$$

The mutual information is thus equal to

1. The reduction in the entropy of x given knowledge of y.
2. The reduction in the entropy of y given knowledge of x.
3. The difference between the sum of the entropies of x and y separately and their joint entropy.

These equalities justify the name "mutual information" for the quantity defined by equation (A.9). Mutual information between two variables is only zero when knowledge of one does not reduce the uncertainty in the value of the other variable.

[1] We will set $k = 1$ from now on

B

The Method of Lagrange Multipliers

Constructing a maximum entropy distribution given knowledge of a few macroscopic variables is often mathematically equivalent to finding the extremal point of a function $f(\vec{x})$ in a subspace of a high-dimensional space. The subspace in this case refers to the space of probability distributions that fulfill the constraints set by the knowledge of the macroscopic variables. The points $\vec{x}$ refer to probability distributions, and the function $f(\vec{x})$ refers to the entropy function on the distribution $\vec{x}$.

This mathematical problem is often most easily solved by using the method of Lagrange multipliers, which is explained below.

B.1 Derivation

Consider an n-dimensional space over which a function $f(\vec{x})$ is defined, with $\vec{x}$ an n-dimensional vector. Within this space, we want to restrict ourselves to points $\vec{x}$ for which the following set of constraints is fulfilled:

$$g_i(\vec{x}) = g_i, \tag{B.1}$$

for all i ranging from 1 to m. The number of constraints m is of course generally smaller than n, since otherwise certain constraints may contradict one another. We denote by Γ the subspace of all points $\vec{x}$ that fulfill the above set of constraints (B.1). Within this subspace Γ, we want to find a point $\vec{y}$ for which the function $f(\vec{y})$ is at an extremum. That is, moving an infinitesimal distance away from $\vec{y}$ in *any* direction, that does not take us out of the subspace Γ, should leave $f(\vec{y})$ unchanged.

Consider a direction $\vec{v}$ and a point $\vec{x}$ within the subspace Γ for which

$$\vec{v} \cdot \nabla g_i(\vec{x}) = 0. \tag{B.2}$$

This implies that direction $\vec{v}$ is *parallel* to a surface of constant g_i. If the above equation is satisfied for all constraints i, it implies that moving a small distance in direction $\vec{v}$ leaves us within the subspace Γ. There are $n - m$ independent directions $\vec{v}$ for which equation (B.2) holds. If we additionally have that for all these directions $\vec{v}$

$$\vec{v} \cdot \nabla f(\vec{x}) = 0, \tag{B.3}$$

then the point $\vec{x}$ is the desired extremal point of $f(\vec{x})$.

Consider now the function $F(\vec{x})$, defined as

$$F(\vec{x}) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i g_i(\vec{x}), \tag{B.4}$$

where the parameters λ_i are as yet undetermined *Lagrangian multipliers*. We now focus on all combinations of parameters λ_i and points $\vec{x}$ for which

$$\frac{\partial F(\vec{x})}{\partial x_j} = 0, \tag{B.5}$$

for all j from 1 to n. It is easy to see that this implies that

$$\vec{v} \cdot \nabla F(\vec{x}) = 0, \tag{B.6}$$

for *any* vector $\vec{v}$ pointing either within or orthogonal to the subspace Γ. In particular, for all $n - m$ independent vectors $\vec{v}$ pointing within the subspace, all combinations of $\vec{x}$ and the λ_i that satisfy equation (B.6) will automatically satisfy equation (B.3). That is, for points $\vec{x}$ and λ_i for which (B.5) holds, we immediate have the $f(\vec{x})$ is stationary in all directions that leave us within the subspace Γ.

Notice that there are n independent equations in (B.5), whereas the function $F(\vec{x})$ has $n + m$ independent unspecified parameters. Generally, what one does it to solve the n equations (B.5) for the n independent parameters of the point $\vec{x}$ in terms of the λ_i. After that, those m parameters λ_i can then be fixed by using the m constraint equations (B.1).

B.2 Example

As an example, most directly related to statistical mechanics, we determine the maximum entropy distribution $\vec{p}$ for a given average of a given variable h. The vector $\vec{p}$ is a distribution, with p_i the probability of event i. The entropy function $f(\vec{p})$ is given by:

$$f(\vec{p}) = -\sum_{i=1}^{n} p_i \log(p_i). \tag{B.7}$$

The constraint functions $g_1(\vec{p})$ and $g_2(\vec{p})$ in this case are that the distribution is normalized

$$g_1(\vec{p}) \equiv \sum_{i=1}^{n} p_i = 1, \tag{B.8}$$

and that the mean of the variable h is given by $\langle h \rangle$:

$$g_2(\vec{p}) \equiv \sum_{i=1}^{n} h_i p_i = \langle h \rangle. \tag{B.9}$$

Substituting these definitions into equations (B.4) and (B.5), we obtain

$$-\log(p_i) - 1 + \lambda_1 + \lambda_2 h_i = 0, \tag{B.10}$$

for all i. Thus, the solution of $\vec{p}$ in terms of the λ_i is

$$p_i = e^{\lambda_1 - 1 + \lambda_2 h_i}. \tag{B.11}$$

Now we just have to determine λ_1 and λ_2 from the constraints. Using the normalization condition, we can first express λ_1 in terms of λ_2:

$$e^{\lambda_1 - 1} = \left(\sum_{i=1}^{n} e^{\lambda_2 h_i} \right)^{-1}. \tag{B.12}$$

Finally, the parameter λ_2 is then a solution of the equation

$$\frac{\sum_i h_i e^{\lambda_2 h_i}}{\sum_j e^{\lambda_2 h_j}} = \langle h \rangle. \tag{B.13}$$

Solving this equation gives us the unique maximum entropy distribution $\vec{p}$ conditioned on the mean $\langle h \rangle$ of the function h.

As an aside, note that by defining the function

$$Z(\lambda_2) = \sum_{i=1}^{n} e^{\lambda_2 h_i}, \tag{B.14}$$

the equation for λ_2 becomes

$$\frac{\partial \log [Z(\lambda_2)]}{\partial \lambda_2} = \langle h \rangle. \tag{B.15}$$

This form is often easier to solve in practice. The function $Z(\lambda_2)$ is generally called a *partition function*. If we take the indices i to represent states of a physical system and we take the functions h_i to be the energies of the different states i, then equation (B.11) gives the well known *Boltzmann distribution*. The parameter $\lambda_2 = -\beta$ is equal to the *inverse temperature* in that case. That is, the Boltzmann distribution is nothing else then that distribution over physical states that maximizes the entropy of the system given a certain fixed average energy.

C

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 4.5.5. 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

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

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

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

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

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

We can now turn the eigenvector equation

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

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

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

and into an equation for the zeroth component P_0,

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

Since $\mathbf{H}$ is a positive definite matrix, Perron's theorem applies: $\mathbf{H}$ has a unique positive definite eigenvector, the eigenvalue of which is larger than all other eigenvalues. This means that $\mathbf{H}$ 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 C.6 that P_0 is positive. Specifically, we have

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

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 $\mathbf{H}$ have at least one negative component, the above eigenvector $\vec{P}$ is the unique positive definite eigenvector of $\mathbf{G}$ and therefore $\vec{P}$ is the only eigenvector of $\mathbf{G}$ that can be interpreted as a fitness distribution.

D

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 [109]. (Useful mathematical background can be found in [42].) Note that we will prove that the finite population dynamics converges towards the infinite population dynamics as given by $\mathbf{G}$, *provided* that $\mathbf{G}$ 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

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

At time $t = 0$ the fitness distribution for the finite population is taken to be the infinite population distribution $\vec{I}(0)$. Using equation 4.45 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

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

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 γ, any component $\vec{P}_i(t)$ of the fitness distributions $\vec{P}(t)$ stays within β of the infinite population trajectory. Specifically, we want to prove that for sufficiently large populations one has

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

where the $\epsilon(t)$ are uniformly smaller than the chosen bound β 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 D.3 into conditional probabilities:

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

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

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

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

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

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

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

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 D.6,

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

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

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

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

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

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

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

Furthermore, we have that

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

Now, if we define

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

then we find that

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

Looking over a series of time steps, if we define

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

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

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

Requiring this probability to be greater than γ we obtain

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

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

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

where

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

Concluding, we see that if we choose M such that

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

then, for all $t \leq T$, the $\delta_i(t)$ are smaller than β with probability greater than γ. Since the right-hand side in D.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 $T \to \infty$ there is no finite population M for which the deviations $\delta_i(t)$ remain arbitrarily small for arbitrary time.

E

Generation Operator and Quasispecies for the Royal Staircase

E.1 Selection Operator

Since the GA uses fitness-proportionate selection, the proportion P_i^s of strings with fitness i after selection is proportional to i and to the fraction P_i of strings with fitness i before selection; that is, $P_i^s = c\ i\ P_i$. The constant c can be determined by demanding that the distribution remains normalized. Since the average fitness $\langle f \rangle$ of the population is given by Eq. (6.3), we have $P_i^s = iP_i/\langle f \rangle$. In this way, we define a (diagonal) operator $\mathbf{S}$ that acts on a fitness distribution $\vec{P}$ and produces the fitness distribution $\vec{P}^s$ after selection by:

$$\left(\mathbf{S} \cdot \vec{P}\right)_i = \sum_{j=1}^{N+1} \frac{\delta_{ij} j}{\langle f \rangle} P_j \ . \tag{E.1}$$

Notice that this operator is nonlinear since the average fitness $\langle f \rangle$ is a function of the distribution $\vec{P}$ on which the operator acts.

E.2 Mutation Operator

The component $\mathbf{M}_{ij}$ of the mutation operator gives the probability that a string of fitness j is turned into a string with fitness i under mutation.

First, consider the components $\mathbf{M}_{ij}$ with $i < j$. These strings are obtained if mutation leaves the first $i-1$ blocks of the string unaltered and disrupts the ith block in the string. Multiplying the probabilities that the preceding $i-1$ blocks remain aligned and that the ith block becomes unaligned we have:

$$\mathbf{M}_{ij} = (1-q)^{(i-1)K} \left(1 - (1-q)^K\right), \quad i < j \ . \tag{E.2}$$

The diagonal components $\mathbf{M}_{jj}$ are obtained when mutation leaves the first $j-1$ blocks unaltered and does *not* mutate the jth block to be aligned. The maximum entropy assumption says that the jth block is random and so the probability P_a that mutation will change the unaligned jth block to an aligned block is given by:

$$P_a = \frac{1 - (1-q)^K}{2^K - 1} \ . \tag{E.3}$$

This is the probability that at least one mutation will occur in the block times the probability that the mutated block will be in the aligned configuration. Thus, the diagonal components are given by:

$$\mathbf{M}_{jj} = (1-q)^{(j-1)K}(1-P_a). \tag{E.4}$$

Finally, we calculate the probabilities for increasing-fitness transitions $\mathbf{M}_{ij}$ with $i > j$. These transition probabilities depend on the states of the unaligned blocks $j-1$ through i. Under the maximum entropy assumption all these blocks are random. The jth block is equally likely to be in any of $2^K - 1$ unaligned configurations. All succeeding blocks are equally likely to be in any one of the 2^K configurations, including the aligned one. In order for a transition to occur from state j to i, all the first $j-1$ aligned blocks have to remain aligned, then the jth block has to become aligned through the mutation. The latter has probability P_a. Furthermore, the following $i-j-1$ blocks all have to be aligned. Finally, the ith block has to be unaligned. Putting these together, we find that:

$$\mathbf{M}_{ij} = (1-q)^{(j-1)K} P_a \left(\frac{1}{2^K}\right)^{i-j-1} \left(1 - \frac{1}{2^K}\right) \ , \ i > j \ . \tag{E.5}$$

The last factor does not appear for the special case of the global optimum, $i = N+1$, since there is no $(N+1)$st block.

E.3 Epoch Fitnesses and Quasispecies

The generation operator $\mathbf{G}$ is given by the product of the mutation and selection operators derived above; i.e. $\mathbf{G} = \mathbf{M} \cdot \mathbf{S}$. The operators $\mathbf{G}^n$ are defined as the projection of the operator $\mathbf{G}$ onto the first n dimensions of the fitness distribution space. Formally:

$$\mathbf{G}^n_i[\vec{P}] = \mathbf{G}_i[\vec{P}], \quad i \leq n, \tag{E.6}$$

and, of course, the components with $i > n$ are zero.

The epoch quasispecies $\vec{P}^n$ is a fixed point of the operator $\mathbf{G}^n$. As in Sec. 5.7, we take out the factor $\langle f \rangle$ to obtain the matrix $\tilde{\mathbf{G}}^n$. The epoch quasispecies is now simply the principal eigenvector of the matrix $\tilde{\mathbf{G}}^n$ and this can be easily obtained numerically.

However, in order to obtain analytically the form of the quasispecies distribution $\vec{P}^n$ during epoch n we approximate the matrix $\tilde{\mathbf{G}}^n$. As shown in App. E.2, the components $\mathbf{M}_{ij}$ (and so of $\tilde{\mathbf{G}}^n$) naturally fall into three categories. Those with $i < j$, those with $i > j$, and those on the diagonal $i = j$. Components with $i > j$ involve at least one block becoming aligned through mutation. These terms are generally much smaller than the terms that only involve the destruction of aligned blocks or for which there is no change in the blocks. We therefore approximate $\tilde{\mathbf{G}}^n$ by neglecting terms proportional to the rate of aligned-block creation—what was called P_a in App. E.2. Under this approximation for the components of $\tilde{\mathbf{G}}^n$, we have:

$$\tilde{\mathbf{G}}^n_{ij} = j(1-q)^{(i-1)K}(1-(1-q)^K), \quad i < j \ , \tag{E.7}$$

and

$$\tilde{\mathbf{G}}^n_{jj} = j(1-q)^{(j-1)K} \ . \tag{E.8}$$

The components with $i > j$ are now all zero.

Note first that all components of $\tilde{\mathbf{G}}^n$ only depend on q and K through $\lambda \equiv (1-q)^K$, the probability that an aligned block is not destroyed by mutation. Note further that in this approximation $\tilde{\mathbf{G}}^n$ is upper triangular. As is well known in matrix theory, the eigenvalues of an upper triangular matrix are given by its diagonal components. Therefore, the average fitness f_n in epoch n, which is given by the largest eigenvalue, is equal to the largest diagonal component $\tilde{\mathbf{G}}^n$. That is,

$$f_n = n(1-q)^{(n-1)K} = n\lambda^{n-1} \; . \tag{E.9}$$

The principal eigenvector $\vec{P}^n$ is the solution of the equation:

$$\sum_{j=1}^{n} \left(\tilde{\mathbf{G}}^n_{ij} - f_n \delta_{ij} \right) P^n_j = 0 \; . \tag{E.10}$$

Since the components of $\tilde{\mathbf{G}}^n$ depend on λ in such a simple way, we can analytically solve for this eigenvector; finding that the quasispecies components are given by:

$$P^n_i = \frac{(1-\lambda) n \lambda^{n-1-i}}{n\lambda^{n-1-i} - i} \prod_{j=1}^{i-1} \frac{n\lambda^{n-j} - j}{n\lambda^{n-1-j} - j} \; . \tag{E.11}$$

For the component P^n_n this reduces to

$$P^n_n = \prod_{j=1}^{n-1} \frac{n\lambda^{n-j} - j}{n\lambda^{n-1-j} - j} \; . \tag{E.12}$$

The above equation can be re-expressed in terms of the epoch fitness levels f_j:

$$P^n_n = \lambda^{n-1} \prod_{j=1}^{n-1} \frac{f_n - f_j}{f_n - \lambda f_j} \; . \tag{E.13}$$

F

Epoch Quasispecies for the Royal Staircase with Ditches

In this appendix, we construct the *generation operator* $\mathbf{G}$ for the class of Royal Staircase with Ditches fitness functions, and analytically approximate the metastable quasispecies distributions $\vec{P}^n$.

As before, the operator $\mathbf{G}$ decomposes into a part $\mathbf{S}$ that encodes the expected effects of selection on the population fitness distribution and a mutation operator $\mathbf{M}$ that encodes the expected effects of mutation.

The selection operator $\mathbf{S}$ is easy to construct since selection depends only on the population's fitness distribution. After selection, we have for the nonditch genotypes that:

$$P_n^{\text{sel}} \equiv \frac{\left(\mathbf{S} \cdot \vec{P}\right)_n}{\langle f \rangle} = \frac{n}{\langle f \rangle} P_n \ , \tag{F.1}$$

and for the ditch genotypes that

$$P_{n,*}^{\text{sel}} = \frac{n-h}{\langle f \rangle} P_{n,*} \ . \tag{F.2}$$

Construction of the mutation operator $\mathbf{M}$ is more involved, but straightforwardly follows chapter 4. Formally, the probability M_{ij} that a genotype of type j turns into a genotype of type i under mutation is given by a sum over all genotypes of type i and j, weighted by the probability of mutating from one genotype to the other. When we refer to the "type" of a genotype we distinguish genotypes only based on their location in an epoch's neutral network (n) or intervening ditch ($n, *$). We denote by Λ_i the set of all genotypes of type i and by $d(s, s')$ the Hamming distance between the genotypes s and s'. Formally, we then have:

$$\mathrm{M}_{ij} = \sum_{s \in \Lambda_i} \sum_{s' \in \Lambda_j} \frac{\mu^{d(s,s')}(1-\mu)^{L-d(s,s')}}{|\Lambda_j|} \ , \tag{F.3}$$

where $|\Lambda_j|$ is the size of the set Λ_j. That is, we assume that a type j genotype is equally likely to be any of its possible $|\Lambda_j|$ genotypes. We then sum over all possible genotypes of type i to which j can mutate. The generation operator $\mathbf{G}$ is then, as before, the product of the selection operator, Eqs. (F.1) and (F.2), and the resulting mutation operator: $\mathbf{G} = \mathbf{M} \cdot \mathbf{S}$.

We now list expressions for the different types of $\mathbf{G}$'s components. In order to simplify the formulas, we first define

$$\alpha = \left(\frac{\mu}{1-\mu}\right)^w , \tag{F.4}$$

and the probability λ to not mutate a block, which is given by

$$\lambda = (1-\mu)^K . \tag{F.5}$$

Using this notation, the probability to mutate a block from type A to type B, for instance, becomes $\Pr(A \to B) = \alpha\lambda$. Components G_{ij} with $i < j$ become

$$G_{ij} = \alpha\lambda^{i-1} j , \tag{F.6}$$

for $i \neq 1$, and

$$G_{1j} = \left(1 - \lambda^{j-1} - \alpha\frac{\lambda - \lambda^{j-1}}{1-\lambda}\right) j , \tag{F.7}$$

for $i = 1$. The diagonal components are given by

$$G_{jj} = \lambda^{j-1} j . \tag{F.8}$$

For the ditch genotypes, the diagonal components are:

$$G_{(j-h)(j-h)} = \lambda^{j-1}(j-h) . \tag{F.9}$$

Components describing the transitions from ditch genotypes to lower-fitness nonditch genotypes $i < j$ are given by:

$$G_{i(j-h)} = \alpha\lambda^{i-1}(j-h) . \tag{F.10}$$

Finally, mutations from nonditch genotypes to lower lying ditch genotype come in three varieties. First, for $i < j-2$ we have

$$G_{(i-h)j} = \alpha\lambda^{i-1}(1-\lambda-\lambda\alpha)j . \tag{F.11}$$

For $i = j-1$ we have

$$G_{(j-1-h)j} = \lambda^{j-2}(1-\lambda-\alpha\lambda)j . \tag{F.12}$$

And for the ditch lying immediately below, $i = j$, we have

$$G_{(j-h)j} = \frac{1}{2^K}\lambda^{j-2}(1-\lambda-\alpha\lambda)j . \tag{F.13}$$

The quasispecies fitness distribution during epoch n is given by the principal eigenvector of the matrix $\mathbf{G}$ restricted to the components with fitness lower than or equal to n; see Refs. [137] and chapter 4. This eigenvector can be obtained numerically, using the above formulas for the components G_{ij}.

An analytic approximation to $\vec{P}^n$ can be obtained by considering mutations only from each fitness j (or $j-h$) to equal or lower fitness. This approximation is justified by the fact that fitness-increasing mutations are very rare compared to fitness-decreasing mutations. Under that approximation, the matrix $\mathbf{G}$ becomes upper triangular. For upper triangular matrices, the components $\vec{P}_i^n$ of the principal eigenvector $\vec{P}^n$ obey the equations:

$$P_i^n = \frac{\sum_{j>i}^n G_{ij} P_j^n}{G_{nn} - G_{ii}} \,, \tag{F.14}$$

and, since $\vec{P}^n$ is a distribution, we additionally have the normalization condition given by

$$\sum_{i=1}^{n} P_i^n = 1 \,. \tag{F.15}$$

Note that the sums in the above equations involve both ditch genotypes and nonditch genotypes. Their precise ordering with respect to fitness depends, of course, on the barrier height h. For instance, for $0 < h < 1$, the fitness $n-h$ genotypes in the ditch between fitness-n and fitness-$(n+1)$ genotypes lie between fitness $n-1$ and n.

With the analytic expressions for the components of $\mathbf{G}$ in hand, and by using Eq. (F.14), we can express all P_i^n for $i < n$ as a function of P_n^n. For instance, assuming $h < 1$, such that the genotypes of fitness $n-h$ are the second highest-fitness genotypes in the population, we have

$$P_{n,*}^n = \frac{n(1-\lambda-\alpha\lambda)}{2^K h\lambda} P_n^n \,. \tag{F.16}$$

Using this, we have for genotypes of fitness $n-1$ that

$$P_{n-1}^n = \frac{n\left(\alpha(n-h)(1-\lambda-\alpha\lambda) + 2^K\lambda\right)}{(1-n(1-\lambda))\,\lambda 2^K} P_n^n \,. \tag{F.17}$$

When all components P_i^n with $i < n$ have been expressed in terms of P_n^n in this way, one finally uses the normalization condition Eq. (F.15) to determine P_n^n.

This procedure leads in a straightforward way to a relatively accurate analytical expression for the epoch fitness distributions. The expressions, however, are rather cumbersome and we will not list them all here. In general, they depend on the size of h since the ordering of the fitness values of the different types of genotypes depends on h.

G

Publications

In chronological order:

E. van Nimwegen, J. P. Crutchfield, and M. Mitchell
Finite populations induce metastability in evolutionary search
Physics Letters A 229 (1997) 144-150
Santa Fe Institute Working Paper 96-08-054

M. A. Huynen and E. van Nimwegen
The frequency distribution of gene family sizes in complete genomes
Molecular Biology and Evolution 15 (1998): 583-589
Santa Fe Institute Working Paper 97-03-025.

E. van Nimwegen, J. P. Crutchfield, and M. Mitchell
Statistical Dynamics of the Royal Road Genetic Algorithm
Theoretical Computer Science 220 (1999, in press)
Santa Fe Institute Working Paper 97-04-035

M. Oprea, E. van Nimwegen, and A. Perelson.
Dynamics of one-pass germinal center models:
Implications for affinity maturation
Bulletin of Mathematical Biology (1999, accepted for publication)
Santa Fe Institute Working Paper 98-05-043

E. van Nimwegen and J. P. Crutchfield
Optimizing Epochal Evolutionary Search: Population-Size Independent Theory
Computer Methods in Applied Mechanics and Engineering, special issue on
Evolutionary and Genetic Algorithms in Computational Mechanics and Engineering,
D. Goldberg and K. Deb, editors (1999, in press)
Santa Fe Institute Working Paper 98-06-046

E. van Nimwegen and J. P. Crutchfield
Optimizing Epochal Evolutionary Search: Population-Size Dependent Theory
Machine Learning Journal (1999, accepted for publication)
Santa Fe Institute Working Paper 98-10-090.

J. P. Crutchfield and E. van Nimwegen
The Evolutionary Unfolding of Complexity
In *Evolution as Computation*,
L. F. Landweber, E. Winfree, R. Lipton and S. Freeland, editors
Lecture Notes in Computer Science, Springer-Verlag, New York
Santa Fe Institute Working Paper 99-02-015.

E. van Nimwegen, J. P. Crutchfield, and M. Huynen
Neutral Evolution of Mutational Robustness
Proceedings of the National Academy of Science U.S.A. **96** (1999) 9716-9720.
Santa Fe Institute Working Paper 99-03-021.

E. van Nimwegen and J. P. Crutchfield
Metastable Evolutionary Dynamics:
Crossing Fitness Barriers or Escaping via Neutral Paths?
submitted to *Bulletin of Mathematical Biology*,
Santa Fe Institute Working Paper 99-07-041.

H

Curriculum Vitae

The author was born on November 5 1970 in Amsterdam, the Netherlands. Between 1983 and 1989 he attended the Vossius Gymnasium in Amsterdam, obtaining his Gymnasium β diploma *cum laude* in June 1989. In September 1989, he commenced his studies at the physics and astronomy faculty of the University of Amsterdam (now fused with the faculty of mathematics and computer science).

He started a master's thesis project in June 1994 with Prof. Dr. R.H. Dijkgraaf on quantum fields in curved spaces and geometric entropy. In August 1995 he finished his studies in Amsterdam, obtaining his "doctoraal" diploma in theoretical physics.

From Septembert 1995 until present he has been working at the Santa Fe Institute in Santa Fe, New Mexico, U.S.A. as a graduate student with Prof. J.P. Crutchfield as his local supervisor, and Prof. P. Hogeweg of Utrecht University, the Netherlands as his official promotor. From October through December 1997 and April through June 1999, he visited and worked at the theoretical biology and bioinformatics group of Prof. P. Hogeweg at Utrecht University.

As of October 1 1999, he will be working as a postdoctoral researcher at the Santa Fe Institute, continuing his work with Prof. J.P. Crutchfield.

I

Samenvatting

Darwin's evolutietheorie kan in zijn essentie worden samengevat in termen van de volgende observaties. Ten eerste is er een doorlopend proces van organismen die zich vermenigvuldigen en vervangen worden door opeenvolgende generaties van nakomelingen. Ten tweede laten niet alle organismen hetzelfde aantal nakomelingen achter. In het algemeen zal het aantal nakomelingen dat een organisme achterlaat samenhangen met de mate waarin het is aangepast aan de omstandigheden waaronder het leeft. Bovendien zal die "fitness" tot op zekere hoogte over worden gedragen op zijn nakomelingen, omdat deze veel eigenschappen met hun voorzaten gemeen hebben. Ten derde worden er af en toe ook nakomelingen geboren die *nieuwe* eigenschappen hebben die geen van hun voorouders bezaten. De combinatie van deze observaties maakt het aannemelijk dat, gegeven genoeg tijd, de populatie meer en meer zal bestaan uit organismen die in hoge mate zijn aangepast aan hun omgeving. Varianten die niet goed aangepast zijn hebben in het algemeen minder nakomelingen, en zullen dus simpelweg verdwijnen uit de populatie. Dit, in een notedop, is Darwin's evolutietheorie.

Deze evolutietheorie verschaft een intuïtive en elegante verklaring voor de diversiteit en complexiteit van de biologische wereld om ons heen. Om een historische voorbeeld te noemen: Verschillende soorten vinken op de Galapagos eilanden vertonen een zeer nauwe verwantschap maar hebben verschillende snavelvormen. Het lijkt uitermate aannemelijk dat deze vinken allen afstammen van dezelfde soort "voorouder vink". Observatie leert bovendien dat deze vinken verschillende soorten voedsel eten en dat ieder van de snavelvormen aangepast is aan het soort voedsel dat de desbetreffende vink eet. De Darwiniaanse evolutietheorie verklaart deze verschijnselen op zo'n elegante wijze dat enig andere rationele verklaring moeilijk lijkt voor te stellen. Zo zijn er talloze andere voorbeelden in de biologie waar Darwiniaanse evolutietheorie op spectaculair overtuigende wijze wordt aangetoond. Gegeven de hoeveelheid biologische feiten dat kan worden verheldert met zo'n minimaal aantal ingrediënten, mag Darwin's evolutietheorie wel beschouwd worden als één van de meest significante doorbraken in de geschiedenis van de wetenschap.

In wetenschappelijke zin zijn we er daarmee helaas nog niet. De evolutietheorie is een heel ander soort *theorie* dan bijvoorbeeld de theorie van de zwaartekracht. Zwaartekracht is, zoals iedere theorie in de natuurkunde, een kwantitatieve en voorspellende theorie. Gegeven een appel van een zeker gewicht, die van een zekere hoogte uit een boom valt, kunnen we precies zeggen hoe hard die appel terecht zal komen op de neus van iemand die onder een boom ligt te slapen. Tot dergelijke voorspellingen is de evolutie-

theorie echter niet in staat. De theorie geeft kwalitatief begrip, en kan gebruikt worden in beschrijvingen, maar het geeft geen wiskundige procedure om specifieke kwantitatieve vragen te beantwoorden. Om een hedendaags voorbeeld te noemen: Antibiotica spelen een belangrijke rol voor de volksgezondheid doordat ze vele in het verleden dodelijke bacteriële infecties doelmatig kunnen bestrijden. Recentelijk hebben sommige van deze bacterieën echter een zorgbarende mate van *resistentie* ontwikkeld. Door evolutionaire modificaties zijn er varianten ontstaan die immuun zijn tegen veel vormen van antibiotica. Het zou dus van geweldig belang zijn als we een vraag als: "Hoe lang zal het gemiddeld duren voordat bacterieën van het soort x resistent zullen zijn geworden tegen een nieuw antibioticum y?" konden beantwoorden.

In zijn huidige vorm kan Darwin's evolutietheorie dit soort vragen echter niet beantwoorden. Het is misschien begrijpelijk dat in het verleden onderzoekers niet eens op het idee kwamen om een dergelijke specifieke vraag te stellen, maar de laatste decennia is de informatie over biologische systemen zo explosief toegenomen (van verschillende eencellige organismen is bijvoorbeeld de gehele DNA sequentie al bekend) dat veel theoretici zich serieus beginnen af te vragen hoe een wiskundige versie van de evolutietheorie kan worden ontworpen die dit soort specifieke vragen *wel* zou kunnen beantwoorden. Het onderzoek in dit proefschrift tracht een bijdrage te geven aan de formulering van zo'n wiskundige theorie van evolutie.

Concreet bestaat het proefschrift voornamelijk uit de bestuderingen van een aantal wiskundige modellen van evoluerende populaties. We beginnen met het definiëren van een precieze wiskundige representatie van de evoluerende populatie op het niveau van individuele organismen. Meer precies; we geven organismen weer met een *genoom* (zoals hun DNA sequentie) en bepalen voor ieder genoom wat de "fitness" is. Dat wil zeggen, voor ieder genoom leggen we een reproductiesnelheid vast voor organismen met dat genoom. Verder kunnen bij iedere reproductie letters in het genoom met een bepaalde kans *muteren*. Nakomelingen kunnen dus met een bepaalde kans een net even andere DNA sequentie hebben dan hun ouder. Nadat we de dynamica van het model zo precies hebben vastgelegd bestuderen we "wat er gebeurt" met betrekking tot bijvoorbeeld de gemiddelde fitness in de populatie over de tijd. We rekenen dan bijvoorbeeld uit hoe lang het gemiddeld duurt voordat organismen met een bepaalde fitness verschijnen.

Conceptueel is de bijdrage van deze studies tweeledig. Aan de ene kant kiezen we natuurlijk de modellen zo dat hun gedrag in zekere zin iets vertelt over het gedrag van werkelijke biologische systemen (in dit geval *zeer* eenvoudige biologische systemen). Doordat we het gedrag van deze modellen tot in detail kunnen beschrijven en voorspellen, menen we ook iets te kunnen zeggen over de te verwachten gedragingen van de werkelijke biologische systemen die lijken op onze modellen. Ik kom straks nog terug op dit facet van de studie. Aan de andere kant adresseert de analyse een meer algemeen probleem in de formele studie van evolutie.

Dit probleem komt erop neer dat de vragen die ons interesseren (hoe lang duur het voordat de bacterie resistent zal worden?) zich op een veel hoger niveau afspelen dan het niveau waarop men precieze wiskundige vergelijkingen kan neerschrijven voor de dynamica van het systeem (bijvoorbeeld op het niveau waarop de verschillende moleculen in en rond een bacterie met elkaar interacteren). Op het laagste niveau bestaan alle organismen uit moleculen en hun gedrag wordt dus op dit laagste niveau beschreven door de wetten van de natuur- en scheikunde. Alleen op een dergelijk laag niveau van

moleculen en chemische reacties kunnen we de dynamica van het biologische systeem redelijk precies vastleggen in termen van de natuurwetten. Dit is echter niet het niveau waarop zich de voor ons relevante vragen afspelen. Wat de moleculen in een bacterie en zijn omgeving precies doen interesseert ons niet, we willen dingen weten als: Hoeveel schade zal dit type bacterie aanrichten als het ons infecteert? Zal het evolueren naar meer of minder schadelijke vormen, en hoe snel zal deze evolutie plaatsvinden?

De Maximale Entropie Methode

De vraag is dus: hoe kan een analyse ons brengen van de precieze wiskundige representatie van de dynamica op een of ander laag niveau naar het beantwoorden van deze vragen op een veel hoger niveau? Als de formules die de dynamica op het laagste niveau beschrijven simpel en doorzichtig zijn, is het in principe mogelijk om wiskundig precies uit te rekenen wat er op een hoger niveau zal gebeuren: de wiskundige vergelijkingen kunnen dan precies opgelost worden. Helaas is dit in de praktijk bijna altijd onmogelijk. Er zijn enorme hoeveelheden chemische reacties gaande op het laagste niveau, die enorme aantallen ingewikkelde moleculen behelsen, en het is eenvoudigweg uitgesloten dat de wiskundige vergelijkingen die hiermee gepaard gaan allemaal opgelost zouden kunnen worden. We hebben dus een meer flexibele wiskundige methode nodig die het gedrag op een hoger niveau kan "afleiden" uit de dynamica op het laagste niveau zonder dat alle wiskundige vergelijkingen op dit niveau hoeven te worden opgelost.

Gelukkig bestaat er een algemene methode die precies het bovenstaande bewerkstelligd. Deze methode heet de *maximale entropie* methode en heeft zijn oorsprong in de statistische mechanica. In plaats van alle wiskundige vergelijkingen en variabelen precies door te rekenen worden bijna alle details van de precieze wiskundige beschrijving statistisch *uitgemiddeld*, totdat er vergelijkingen overblijven op het hogere niveau waarin we geïnteresseerd zijn. Deze methode wordt van origine toegepast in de natuurkunde op systemen zoals gassen en vloeistoffen. Alle precieze bewegingen en botsingen van de moleculen in een gas worden uitgemiddeld totdat er een beschrijving van het gas overblijft op het (macroscopische) niveau van variabelen zoals druk, volume, en temperatuur. De maximale entropie methode bestaat erin dat dit uitmiddelen op een zo *eerlijk* mogelijke manier moet plaatsvinden. Dat wil zeggen, we berekenen een statistisch gemiddelde over alle mogelijke bewegingen en botsingen die de moleculen in het gas zouden *kunnen* uitvoeren. Hoewel het te ver voert om op deze plaats uit te leggen, is het belangrijk dat de "eerlijkheid" van deze uitmiddelingsprocedure geen kwestie is van *smaak* of iets dergelijks. Zodra we vastleggen welke informatie over de dynamica op het laagste niveau we willen betrekken in de berekening, en welke grootheden op een hoger niveau we willen voorspellen is de uitmiddelingsprocedure in principe uniek wiskundig bepaald.

De maximale entropie methode is zo algemeen en flexibel dat het, behalve op gassen en vloeistoffen, kan worden toegepast op alle mogelijke systemen, dus ook op biologische systemen. De technische details van deze uitmiddelgingsprocedure zijn echter nog niet goed wiskundig uitgewerkt voor de meeste systemen die buiten het onmiddelijke domein van studie in de statistische mechanica vallen. De meer algemene bijdrage van dit proefschrift is om deze methode precies uit te werken, en zijn succes te demonstreren,

voor een aantal relatief eenvoudige evoluerende systemen. We beginnen steeds met de wiskundige beschrijving van onze modellen op het laagste niveau van reproducerende individuen, en middelen deze vergelijkingen uit om een beschrijving te krijgen van het systeem op het niveau van de gemiddelde en beste *fitness* in de populatie. Vergelijking van onze wiskundige resultaten met data verkregen met behulp van computersimulaties van deze systemen laat zien dat de methode met succes de dynamica op het niveau van fitness kan voorspellen voor problemen die te ingewikkeld zijn om direct op het laagste niveau op te lossen.

Evolutionaire Modellen

Aangezien we het evolutionaire gedrag van gemiddelde en beste fitness in onze modellen precies kunnen voorspellen, kunnen we zoals gezegd ook meer specifieke conclusies trekken voor de dynamica van simpele evolutionaire systemen. Om te kunnen bespreken welke realistische evolutionaire systemen beschreven kunnen worden door onze modellen, en om concreter te kunnen beschrijven welke conclusies voor evolutionair gedrag kunnen worden getrokken uit onze studies, moeten we eerst het type evolutionaire modellen dat in dit proefschrift bestudeerd wordt wat meer in detail introduceren.

In onze modellen kan ieder individu in de evoluerende populatie geïdentificeerd worden met zijn *genotype*: een sequentie van symbolen. Het biologisch meest relevante voorbeeld van genotypen zijn DNA moleculen. Een DNA molecuul bestaat uit een sequentie van de vier symbolen A, C, G, en T die corresponderen met de nucleotiden waaruit het DNA molecuul is opgebouwd. In onze modellen gebruiken we voornamelijk *binaire* sequenties van een vaste lengte L. Dat wil zeggen, een genotype is een sequentie van L symbolen, waarin ieder symbool een 0 of 1 kan zijn. Er zijn dus 2^L mogelijke genotypen in totaal.

Omdat mutaties in genotypen zeldzaam zijn, ondergaat een sequentie bijna altijd maar één mutatie tegelijk. Zo'n *puntmutatie* verandert één van de L symbolen van een 0 in een 1 of vice versa. In deze zin kunnen sequenties die maar op één plek van elkaar verschillen beschouwd worden als genetische "buren". Dat wil zeggen, twee genotypen vormen genetische buren als de ene door een puntmutatie in de andere kan overgaan. Men kan zich de totale verzameling van 2^L genotypen dan voorstellen als een hoogdimensionaal rooster. Ieder genotype komt overeen met een punt op het rooster, en roosterverbindingen verbinden paren van genotypen die maar op één plek van elkaar verschillen. Ieder genotype heeft dus L verschillende buren op het rooster. Dit rooster wordt ook wel de *genotype-ruimte* genoemd.

Aangezien in onze modellen het gedrag, de overlevingskansen, en reproductiesnelheid van een individu alleen afhangt van zijn genotype, kunnen we zoals gezegd ieder individu in principe identificeren met zijn genotype. We stellen ons dan voor dat ieder individu in de populatie zich op één van de roosterpunten van de genotype-ruimte bevindt. Wanneer een individu reproduceert zonder mutaties verschijnt er nieuw individu op hetzelfde roosterpunt. Als bij de reproductie een puntmutatie plaatsvindt verschijnt het nieuwe individu op één van de belendende roosterpunten.

Verder heeft ieder van de 2^L genotypen een voorafbepaalde fitness die overeenkomt met de reproductiesnelheid van individuen met dat genotype. Tenslotte worden er in-

dividuen willekeurig uit de populatie verwijdert om de totale populatiegrootte constant te houden. Deze combinatie van ingrediënten leidt er dan in het algemeen toe dat in gebieden van de genotype-ruimte waar fitness relatief hoog is het aantal individuen zal toenemen terwijl in gebieden met relatief lage fitness het aantal individuen zal afnemen.

Kortom, in onze modellen worden de evoluerende individuen weergegeven als genotypen in een vaste genotype-ruimte waar ieder genotype een voorafbepaalde constante fitness heeft. Genetische modificaties vinden plaats door mutaties van de genotypen tijdens reproductie. Dergelijk relatief eenvoudige situaties zijn in de praktijk vanzelfsprekend vrij zeldzaam. Er zijn echter wel een aantal experimenteel bestudeerde voorbeelden. Zo zijn er bijvoorbeeld experimenten waar in een chemisch reactievat verschillende RNA moleculen reproduceren. De reproductiesnelheid van ieder RNA molecuul is in deze experimenten alleen een functie van zijn sequentie. Tijdens reproductie kunnen er ook mutaties in deze sequentie plaatsvinden. Een dergelijk systeeem is dus een goed voorbeeld van het soort evolutionaire modellen dat we hier bestuderen. Ook zijn er vergelijkbare experimenten waarin biomoleculen artificieel worden gereproduceerd, gemuteerd, en geselecteerd. De biomoleculen worden bijvoorbeeld geselecteerd voor een bepaalde chemische eigenschap zoals hun capaciteit om een verbinding aan te gaan met een bepaald ander molecuul. Ingewikkelder systemen die nog wel te vergelijken zijn met de modellen die wij bestuderen zijn experimenten waarin E-coli bacterieën in een constante omgeving evolueren. Dat wil zeggen, E-coli bacterieën reproduceren in een reactievat waar op gezette tijden "voedsel" aan wordt toegevoegd, en dat periodiek wordt verdund om het totale aantal bacteriën ruwweg constant te houden.

Ten slotte zijn onze modellen ook voorbeelden van *evolutionaire algorithmen*. Evolutionaire algorithmen is een verzamelnaam voor computer algorithmen die met behulp van gesimuleerde evolutie ingewikkelde computationele problemen trachten op te lossen. Een populatie van voorbeeld oplossingen voor een bepaald probleem wordt in een computer gerepresenteerd als een verzameling genotypen, bijvoorbeeld als binaire sequenties. De populatie van voorbeeld-oplossingen wordt dan getest op hun capaciteit om het computationele probleem efficient op te lossen. Goede oplossingen worden in het computergeheugen "gereproduceerd" terwijl slechte uit het geheugen worden gewist. De voorbeeld-oplossingen ondergaan ook random genetische modificaties, zoals mutaties, tijdens hun reproductie. Dit proces van selectie en random genetische modificatie wordt dan vele generaties lang herhaalt. De hoop is dat dit evolutionaire proces uiteindelijk leidt tot een populatie met oplossingen van zeer hoge kwaliteit.

Een concreet voorbeeld is het zogenaamde *handelsreiziger probleem*. Een handelsreiziger moet een groot aantal steden bezoeken en wil zo min mogelijk kilometers afleggen in totaal. Hij zoekt dus naar de kortste route die alle steden die hij moet bezoeken eenmaal aandoet. Aangezien de totale lengte van de route alleen afhangt van de volgorde waarin hij de steden bezoekt, kan iedere route gerepresenteerd worden als een specifieke volgorde van de steden. Een evolutionair algorithme voor dit probleem zou nu bijvoorbeeld zijn: Creëer in het computergeheugen een populatie van willekeurige routes. Bereken de lengte van ieder van deze routes. Wis langere routes uit het geheugen en vervang ze met kopieën van kortere routes. Verwissel hier en daar random twee of meer steden in de volgorde van sommige routes. Bereken de lengtes van deze nieuwe generatie routes, reproduceer de goede en wis de slechte, enzovoorts voor een groot aantal generaties.

In veel gevallen vinden dit soort evolutionaire algorithmen goede oplossingen relatief snel, maar lang niet altijd. Gebruikers van evolutionaire algorithmen moeten vaak lang experimenteren met parameters zoals de mutatiefrequenties, populatiegroottes, en selectiedruk (welk percentage van de langere routes wordt gewist en welk percentage wordt gereproduceerd) voordat het algorithme bevredigend werkt. Soms lukt het helemaal niet om bevredigende oplossingen te evolueren voor een bepaald probleem. Er bestaat dus behoefte aan een theorie die kan voorspellen welk soort problemen efficient opgelost kan worden met behulp van evolutionaire algorithmen en hoe de parameters van het algorithme ingesteld zouden moeten worden om goede oplossingen zo snel mogelijk te vinden. De studies in dit proefschrift dragen direct bij aan een dergelijke theorie. In hoofdstukken 5 en 6, bijvoorbeeld, berekenen wij expliciet de optimale mutatiefrequenties en populatiegroottes voor een klasse van optimalisatieproblemen. De nadruk van de studies in dit proefschrift ligt echter niet specifiek op de optimilasitie eigenschappen van evolutionaire algorithmen maar meer algemeen op de analyse van kwalitatief gedrag dat veel evolutionaire systemen vertonen.

Gefaseerde Evolutie

Veel evolutionaire systemen vertonen een kwalitatief gedrag dat we "gefaseerde evolutie" hebben genoemd.[1] In gefaseerde evolutie doorloopt de evolutie een reeks van scherp gescheiden fases. Ieder van die fases vormt een soort van evolutionair *tijdperk* gedurende welke bepaalde eigenschappen van de individuen in de populatie redelijk constant blijven. Deze tijdperken komen dan plotseling ten einde, de populatie ondergaat snelle veranderingen, en de evolutie komt terecht in een nieuwe fase van schijnbare stabiliteit.

Zoals gezegd vertonen veel evolutionaire systemen dit soort gefaseerd gedrag. Onderzoek van fossielen lijkt bijvoorbeeld uit te wijzen dat evolutionaire veranderingen in organismen niet geleidelijk maar in plotselinge "sprongen" plaatsvinden. De lichaamsbouw van organismen blijft soms constant voor miljoenen jaren en verandert dan zeer plotseling. In de paleontologie wordt dit verschijnsel *punctuated equilibrium* genoemd. Ook in de experimenten met E-coli bacterieën die ik zonet noemde vind dit soort gefaseerd evolutionair gedrag plaats. De gemiddelde grootte en reproductiesnelheid van de E-coli bacterieën blijft lange tijden constant en verandert dan plotseling. In computersimulaties van evolutionaire processen, en in evolutionaire algorithmen komt dit gedrag ook zeer regelmatig voor. De gesimuleerde evolutie van RNA moleculen vertoont bijvoorbeeld ook dit gedrag.

Vaak hangen de plotselinge veranderingen, die de populatie voeren van het ene tijdperk naar het volgende, samen met het plotseling verschijnen van individuen die eigenschappen vertonen die nog niet eerder zijn voorgekomen in de populatie, en die zo succesvol zijn dat ze zich snel in de populatie verspreiden. Wij noemen deze plotselinge veranderingen *innovaties*.

Het is goed om hier op te merken dat het gefaseerde gedrag van stabiele tijdperken en plotselinge innovaties niet noodzakerlijkerwijs plaatsvindt op het niveau van de

[1] Dit is een niet helemaal geslaagde vertaling van het Engelse "epochal evolution" dat het vertoonde gedrag beter samenvat.

precieze toestand van de populatie maar alleen met betrekking tot bepaalde eigenschappen. In het voorbeeld van de fossielen kunnen we alleen vaststellen dat de skeletten van de organismen in kwestie dit gedrag vertonen. Het is dus heel goed mogelijk dat de DNA sequenties van deze organismen wel degelijk voortdurend aan geleidelijke veranderingen onderhevig zijn, maar dat gedurende een tijdperk de lichaamsbouw niet door deze genetische veranderingen beïnvloed wordt. Hetzelfde geldt voor de evoluerende E-coli bacterieën. De experimenten tonen dat hun reproductiesnelheid en celgrootte constant blijven gedurende de verschillende tijdperken, maar we weten niet wat er ondertussen plaatsvindt op het niveau van hun DNA sequenties. Gefaseerde evolutie is dus een verschijnsel dat plaatsvindt met betrekking tot bepaalde eigenschappen in de populatie zoals bijvoorbeeld lichaamsbouw of celgrootte. In de modellen in dit proefschrift vindt gefaseerde evolutie plaats op het niveau van de gemiddelde fitness in de populatie. Gedurende een tijdperk fluctueert de gemiddelde fitness in de populatie rond een vaste waarde. Een innovatie vindt plaats wanneer er plotseling individuen verschijnen met hogere fitness dan tevoren. Doordat nakomelingen van deze individuen zich snel in de populatie verspreiden groeit de gemiddelde fitness dan snel. Na een korte periode van verandering stabiliseert de gemiddelde fitness weer, waarna de evolutie zich in een nieuw tijdperk van stabiliteit bevindt.

Dit proefschrift concentreert zich voor het grootste gedeelte op een gedetailleerde wiskundige analyse van dit gefaseerde evolutionaire gedrag. Met behulp van de maximale entropie methode kunnen we voor onze specifieke modellen exact voorspellen bij wat voor waarden van gemiddelde fitness tijdperken zullen plaatsvinden, hoe groot de fluctuaties in fitness zijn gedurend ieder tijdperk, wanneer deze tijdperken instabiel kunnen worden, hoe snel de innovaties de populatie van het ene tijdperk naar het volgende brengen, hoe lang de tijdperken gemiddeld duren, enzovoorts. Deze analyse en resultaten kunnen met name gevonden worden in hoofdstuk 4. Deze samenvatting is echter niet de plaats om uitgebreid in te gaan op de precieze analyse en resultaten. In plaats daarvan volgt nu een korte bespreking van het meer kwalitatieve beeld van gefaseerde evolutie dat naar voren komt uit de studies in dit proefschrift.

Neutraliteit

In onze modellen correspondeert een evolutionair *tijdperk* met een periode waarin de hoogste fitness die voorkomt in de populatie constant is. Met ander woorden, gedurende een tijdperk hebben alle individuen in de populatie een fitness f of lager, en de evolutie blijkt dus voor langere tijd niet in staat om genotypen te produceren die een fitness hoger dan f hebben. Dat de het evolutionaire proces gedurende geruime tijd niet in staat is om genotypen met fitness hoger dan f te produceren kan het gevolg zijn van twee verschillende mechanismen. In de eerste plaats zou het zo kunnen zijn dat de populatie door bepaalde evolutionaire krachten uit de buurt van genotypen met fitness hoger dan f wordt gehouden. Het zou daarentegen ook zo kunnen zijn dat de populatie niet weerhouden wordt van de genotypen met fitness hoger dan f, maar dat de populatie simpelweg moeite heeft om deze genotypen te *vinden*. Laten we deze twee mogelijke mechanismen in meer detail bekijken.

De eerstgenoemde mogelijkheid, namelijk dat de populatie weerhouden wordt van

genotypen met fitness hoger dan f, is de mogelijkheid die door theoretici gewoonlijk in verband wordt gebracht met gefaseerde evolutie. Het wordt vrij algemeen aangenomen dat voor typische evolutionaire systemen er veel *locale optima* in de genotype-ruimte zijn. Een locaal optimum is een genotype dat fitter is dan al zijn genetische buren. Alle mogelijke puntmutaties op dit locaal optimale genotype leiden dus tot een verlaging van fitness. Enige paragrafen geleden bespraken we het algemene beeld van het evolutionaire proces als dat het aantal individuen in gebieden van genotype-ruimte met relatief hoge fitness groeit terwijl het aantal individuen in gebieden met relatief lage fitness afneemt. Als de populatie terecht is gekomen in een gebied rond een locaal optimum zal het evolutionaire proces de populatie dus concentreren op en rond dit locale optimum, omdat dit het meest fitte genotype in de omgeving is. Op deze wijze kan de populatie dus vast komen te zitten in een gebied in genotype-ruimte rond een locaal optimum. Dit zorgt er dan vanzelfsprekend voor dat bepaalde eigenschappen in de populatie, zoals gemiddelde fitness, voor een lange tijd constant kunnen blijven. Met andere woorden, een tijdperk correspondeert met een periode in de evolutie gedurende welke de populatie geconcentreerd is rond een locaal optimum van fitness f in genotype-ruimte. Dit tijdperk komt pas ten einde wanneer een zeer onwaarschijnlijke reeks van mutaties een relatief ver verwijderd genotype genereert buiten de onmiddelijke omgeving van het locale optimum, dat bovendien een fitness hoger dan f heeft. De populatie wordt weerhouden van het vinden van genotypen met fitness hoger dan f, niet omdat deze zo moeilijk te vinden zouden zijn, maar omdat individuen in de populatie deze genotypen niet kunnen *bereiken*; de selectie houdt de populatie geconcentreerd in genotype-ruimte rond het locaal optimale genotype met fitness f. Binnen dit mechanisme correspondeert een innovatie in de gefaseerde evolutie dus met het verspringen van de populatie van het gebied rond één locaal optimum naar een gebied rond een ander locaal optimum door middel van een zeer zeldzame reeks van mutaties. Merk ook op dat in dit mechanisme voor gefaseerde evolutie de populatie stabiel is op het niveau van genotypen gedurende een tijdperk en dus niet alleen op het niveau van een bepaalde eigenschap zoals fitness.

In de door ons bestudeerde modellen wordt de gefaseerde evolutie voornamelijk veroorzaakt door het mechanisme dat ik als tweede noemde: De populatie kan in principe genotypen met fitness hoger dan f makkelijk bereiken, maar deze genotypen zijn moeilijk om te vinden. Centraal in dit alternatieve mechanisme staat het begrip *neutraliteit*. Een genotype bezit neutraliteit als het één of meer genetische buren heeft die selectief neutraal zijn. Dat wil zeggen: het genotype heeft onmiddelijke buren in genotype-ruimte die dezelfde fitness hebben als hijzelf. In de modellen die we bestuderen in dit proefschrift bezitten bijna alle genotypen neutraliteit in meer of mindere mate. Deze simpele eigenschap heeft belangrijke gevolgen. Een individu met een genotype dat neutraliteit bezit kan altijd een gemuteerde nakomeling produceren die even fit is als hijzelf. Deze mogelijkheid voor *neutrale* variatie voorkomt dat de populatie ooit vast komt te zitten in een bepaald gebied van de genotype-ruimte. De populatie zal dus voortdurend over neutrale paden in de genotype-ruimte blijven bewegen. Een tijdperk ontstaat nu wanneer de populatie beweegt over neutrale paden van genotypen met fitness f, en het lang duurt voordat de populatie een genotype van hogere fitness tegenkomt op zijn reis over deze neutrale paden. Merk op dat in dit mechanisme alleen de gemiddelde en beste fitness constant blijft gedurende een tijdperk, maar dat de populatie zich voortdurend blijft bewegen door de genotype-ruimte.

Men zou de verschillen tussen de twee mechanismen als volgt in metaforen kunnen omschrijven: In het eerste geval zit de populatie als het ware opgesloten in een kleine omheinde ruimte. Om uit deze ruimte te ontsnappen moet tenminste één individu in de populatie erin slagen om over de omheining te klimmen. Als de omheining "hoog" is kan het geruime tijd duren voordat een individu hierin slaagt. In het tweede geval zit de populatie niet opgesloten, maar bevindt het zich als het ware in een groot doolhof. Individuen blijven zich voortdurend door de gangen van dit doolhof bewegen maar het kan geruime tijd duren voordat een individu erin slaagt om een "uitgang" te vinden.

Nu dat we deze twee mogelijke mechanismen voor gefaseerde evolutie hebben geïdentificeerd moeten we ons afvragen welke van deze twee het meest van belang is voor realistische evolutionaire systemen. Wij geloven dat de studies in dit proefschrift overtuigende argumenten aanvoeren dat het "doolhof mechanisme" waarschijnlijk een veel grotere rol speelt in evolutie dan het "omheinings mechanisme". Deze argumenten volgen uit een combinatie van observaties waarvan sommige volgen uit het recente werk van anderen en andere uit dit proefschrift naar voren komen.

Zoals gezegd speelt *neutraliteit* een centrale rol in het doolhof mechanisme; als genotype in het algemeen *geen* neutrale buren zouden hebben, zouden er nooit doolhoven van neutrale paden in de genotype-ruimte kunnen bestaan. De eerste observatie ter ondersteuning van het "doolhof mechanisme" is dat het al tientallen jaren bekend is dat neutraliteit veel voorkomt in de genotypen van realistische biologische systemen. De mathematisch populatie-geneticus Kimura stelde als eerste voor dat het overgrote deel van de mutaties die worden waargenomen in het DNA van organismen *neutrale* mutaties zijn. Dat wil zeggen, mutaties die geen gevolgen voor het phenotype of de fitness van het organisme hebben. Dat neutraliteit veel voorkomt is dus wel een geaccepteerd moleculair biologisch feit. Bovendien is er recentelijk uitgebreid onderzoek gedaan naar de "structuur" van biomoleculen zoals RNA. Ieder RNA molecuul bestaat uit een sequentie van de basen G, C, U, en A. Er zijn dus 4^L mogelijke RNA moleculen van lengte L. Uit onderzoek blijkt echter dat deze moleculen in een veel kleiner aantal klassen vallen met betrekking tot hun chemische eigenschappen. Dat wil zeggen, er zijn enorme aantallen RNA moleculen die allemaal ruwweg dezelfde chemische eigenschappen hebben (en dus dezelfde functies kunnen vervullen in een organisme). Verder blijkt ook uit dit onderzoek dat de verzamelingen van RNA moleculen met dezelfde eigenschappen precies zulke neutrale doolhoven vormen in de genotype-ruimte. Deze doolhoven worden ook wel *neutrale netwerken* genoemd. Nog recenter onderzoek suggereert dat deze neutrale netwerken niet alleen voorkomen voor RNA moleculen maar bijvoorbeeld ook voor eiwit moleculen. Uit studies van fitness functies in evolutionaire algorithme problemen komt de laatste tijd ook steeds meer naar voren dat neutraliteit en de daarbij horende neutrale netwerken in genotype-ruimte een veel voorkomend verschijnsel zijn. Vanuit een meer abstract gezichtspunt zou men kunnen zeggen dat neutrale netwerken bijna *onvermijdelijk* zijn wanneer er maar voldoende redundantie is in de genotypen. Zodra er veel genotypen zijn die dezelfde functionele eigenschappen hebben is het waarschijnlijk dat deze verzamelingen genotypen neutrale netwerken vormen.

Tenslotte moet men bedenken dat evolutie vaak zelf in staat is om neutraliteit te *creëren*. Behalve puntmutaties kunnen genotypen nog vele andere genetische modificaties ondergaan. Stukjes DNA kunnen worden gekopieerd, uitgeknipt, omgedraaid, ergens anders in de sequentie weer teruggeplakt, enzovoorts. Wanneer een stukje DNA

wordt gekopieerd en ingevoegd in de bestaande DNA sequentie is dit over het algemeen nauwelijks van nadelige gevolgen voor een organisme; het bezit nu simpelweg een extra kopie van dit stukje DNA. Dit betekent echter dat willekeurige mutaties op dit nieuwe stukje DNA bijna altijd neutraal zullen zijn. Het organisme heeft voor zijn functioneren alleen het origineel nodig, en veranderingen in de kopie hebben geen nadelige gevolgen. Door een stukje DNA te kopieren en in te voegen heeft het organisme als het ware een stukje genetische code gemaakt waarmee het kan experimenteren. Het lijkt al met al dus waarschijnlijk dat neutraliteit in genotype-ruimten de regel is in biologische systemen.

De meer directe bijdrage van dit proefschrift in de argumentatie dat het "doolhof mechanisme" van belang is voor evolutie is dat de studies aantonen dat neutrale evolutie door een doolhof enorm veel sneller verloopt dan wanneer de populatie vast is komen te zitten binnen een "omheining". In hoofdstuk 8 berekenen we expliciet dat in de tijd die een populatie nodig heeft om over zijn omheining te "klimmen" de neutrale evolutie al doolhoven van enorme afmetingen doorzocht heeft. We bekijken situaties waarin de populatie op twee manieren kan ontsnappen aan zijn tijdperk van constante fitness. In de genotype-ruimte zijn er twee soorten wegen naar hogere fitness. De eerste klasse van wegen bestaat uit wegen die erg kort zijn maar leiden langs genotypen met relatief lage fitness. Een individu dat langs dit soort wegen ontsnapt is analoog aan een individu dat "over de omheining klimt". De tweede klasse wegen zijn *omwegen*; wegen die in hun geheel over neutrale paden voeren maar die uitermate lang zijn. Individuen die langs deze wegen ontsnappen zijn analoog aan individuen die de uitweg uit een doolhof vinden. De wiskundige analyse toont aan dat, zelfs voor relatief "lage" omheiningen, het enorm lang kan duren voordat één individu erin slaagt om te ontsnappen langs een pad van relatief lage fitness genotypen. In diezelfde tijd heeft de populatie al enorme hoeveelheden neutrale paden afgezocht. Kortom, de analyse toont aan dat in de meeste situaties ontsnappen langs een weg met relatief lage fitness genotypen uitermate onwaarschijnlijk is. Als er ook maar één uitweg bestaat uit het doolhof van neutrale paden is het zeer waarschijnlijk dat de populatie deze uitweg eerder ontdekt.

Mutationele Robuustheid

Nu dat aannemelijk is gemaakt dat in gefaseerde evolutie de evolutie langs neutrale paden in genotype-ruimte veruit het belangrijkst is, is het vanzelfsprekend om de vraag te stellen: Hoe verspreidt de populatie zich door het doolhof van neutrale paden?

Het is goed om eerst in herinnering te roepen dat de individuen zelf natuurlijk niet *bewegen* over de neutrale paden. De "beweging" is het gevolg van individuen die reproduceren en een gemuteerde nakomeling nalaten in één van hun genetische buur-genotypen. Deze mutaties vinden totaal willekeurig plaats. Men zou dus misschien verwachten dat de populatie zich uiteindelijk geheel uniform over het neutrale netwerk verspreidt. De analyse in hoofdstuk 7 laat echter zien dat dit niet het geval is.

Het blijkt dat de populatie zich op den duur het meest zal concentreren in die gebieden van de genotype-ruimte waar relatief veel neutrale paden elkaar kruizen. Dit verschijnsel is enigszins vergelijkbaar met het verschijnsel dat, zelfs als iedereen willekeurig door de stad zou fietsen, bepaalde centrale kruispunten toch drukker zouden zijn dan andere, simpelweg omdat er meer wegen naar deze centrale kruispunten leiden.

Bovendien hebben we aangetoond dat dit verschijnsel alleen afhangt van de *topologie* van het netwerk van neutrale paden. Dat wil zeggen, het hangt alleen af van de "infrastructuur" van neutrale wegen in genotype-ruimte en niet van het aantal individuen in de populatie, de mutatiefrequentie, of de selectiedruk.

Gebieden in genotype-ruimte waar veel neutrale paden elkaar kruizen corresponderen met genotypen die veel neutrale buren hebben, i.e. genotypen met hoge neutraliteit. Het interessante is nu dat genotypen met hoge neutraliteit precies de genotypen zijn die relatief robuust zijn onder mutaties. Hoe hoger de neutraliteit van een genotype, hoe groter de kans dat zijn fitness onveranderd blijft wanneer het een puntmutatie ondergaat. In die zin is een genotype met veel neutrale buren dus relatief robuust tegen mutaties. Bekeken vanuit het oogpunt van mutationele robuustheid tonen de resultaten in hoofdstuk 7 dat de populatie, zonder ergens door "gestuurd" te worden, vanzelf zal evolueren naar die genotypen op het neutrale netwerk die het meest robuust zijn tegen mutaties. Deze evolutie van mutationele robuustheid is verschillende malen waargenomen in experimenten met evoluerende populaties en de theorie in hoofdstuk 7 geeft dus een rigoreuze theoretische verklaring voor dit verschijnsel.

Conclusie

Tenslotte vat ik het beeld dat is ontstaan uit ons onderzoek van evolutionaire processen samen, en speculeer over de mogelijke generalisaties van deze ideeën.

Evolutie vindt vaak plaats in fases of *tijdperken*. Gedurende een tijdperk is de macroscopische "toestand" van de populatie constant. De functionele karakteristieken van de organismen, hun gemiddelde fitness, of andere gemiddelde eigenschappen van de individuen veranderen niet gedurende deze periode. Dit betekent echter niet dat het evolutionaire proces tot stilstand is gekomen. Wat ook de karakteristieken van de populatie mogen zijn, er zijn altijd mogelijkheden om op genetisch niveau aan de individuen te "sleutelen", zonder dat hun fitness hier onder hoeft te lijden. Mutaties en andere genetische modificaties zorgen er voor dat er ook daadwerkelijk genetisch aan de individuen gesleuteld zal blijven worden. Daarom vinden er op het laagste niveau van de genetische samenstelling van de populatie voortdurend *neutrale* genetische modificaties plaats. De populatie begeeft zich als een dronken wandelaar willekeurig door de ruimte van neutrale variaties op het huidige type. Door middel van neutrale mutaties probeert de populatie als het ware alle mogelijke variaties uit die niet leiden tot een reductie in fitness. Na verloop van tijd kan een individu in de populatie dan plotseling tegen een genetische combinatie "aanlopen" die tot nieuwe functionaliteit en een verbetering in fitness leidt. Er vindt dan een *innovatie* plaats en na een korte periode zal het overgrote deel van de populatie bestaan uit individuen die deze nieuwe functionaliteit met zich meedragen.

Wanneer de populatie op deze manier in een nieuw tijdperk is beland herhaalt dit proces zich in principe. Terwijl de functionele karakteristieken en gemiddelde fitness van de organismen constant blijven probeert de populatie voortdurend neutrale genetische variaties uit. Het is goed om op te merken dat het scala van *mogelijkheden* tot neutrale variatie in principe bepaald wordt door de huidige vorm van de individuen. Simpel gezegd heeft een populatie van E-coli bacterieën hele andere mogelijkheden voor neutrale

genetische variatie dan bijvoorbeeld een populatie kevers, een populatie muggen, of een populatie van het HIV virus. In meer abstracte zin kunnen we dus stellen dat de huidige vorm van de populatie samen met de omgeving bepaald welke *ruimte* van mogelijkheden er is voor neutrale variatie. Dit leidt er dan weer toe dat de innovaties die "verborgen" liggen binnen deze neutrale ruimte in zekere zin ook zijn bepaald door de huidige toestand van de populatie. In evolutietheorie wordt dit algemene verschijnsel aangeduid als de *structurele beperkingen* van de evolutionaire mogelijkheden. Dit verschijnsel is dus op een natuurlijke wijze ingebed in onze theorie.

Het moge verder ook duidelijk zijn dat er niet altijd één unieke innovatiemogelijkheid is binnen een tijdperk. Het is veel vaker het geval dat er structureel verschillende "ontsnappingsmogelijkheden" zijn uit een tijdperk. Om een voorbeeld te geven: Als een chemisch proces binnen een organisme niet efficient verloopt omdat molecuul X voortdurend de interacties tussen moleculen Y en Z verstoort, zou een evolutionaire oplossing kunnen bestaan in het creëren van nieuwe moleculen die molecuul X afbreken. Een oplossing zou echter ook kunnen zijn om de moleculen Y en/of Z zo te veranderen dat X hen niet langer in hun interactie verstoord. In sommige gevallen zou het zo kunnen zijn dat het zuiver een kwestie van toeval is welke van de twee oplossingen de evolutie als eerste ontdenkt. Echter, wanneer één van deze twee oplossingen is ontdekt zal dit leiden tot een innovatie en zit de populatie voor zijn verdere toekomst "vast" aan deze oplossing, hetgeen weer de verdere toekomstige mogelijkheden in zekere zin bepaald. Dit kwalitatieve verschijnsel wordt in evolutietheorie een *ingevroren ongeluk* genoemd. Ingevroren ongelukken vinden dus ook een natuurlijke plaats binnen onze theorie.

Al de theoretische modellen die bestudeerd worden in dit proefschrift hebben een zekere mate van artificialiteit: Zij bestuderen evoluerende populaties in omgevingen die wij per fiat in een computer geschapen hebben. In de toekomst hoop ik de theorie toe te passen in meer realistische situaties waar de omgeving de fysische werkelijkheid beschrijft waarin organismen nu eenmaal evolueren. In deze context zou men over de evoluerende populatie kunnen denken als een verzameling machientjes die de energie, die voortdurend door de zon naar de aarde wordt gestraald, kunnen verwerken en gebruiken om kopieën van zichzelf te creëren. De fitness van zo'n machientje hangt af van de efficientie waarmee het energiestromen in zijn omgeving omzet in kopieën van zichzelf. Gedurende een tijdperk wordt er door neutrale evolutie voortdurend random gesleuteld aan deze machientjes zonder hun functionaliteit te verlagen. Een innovatie vindt dan plaats wanneer één zo'n machientje toevallig tegen een combinatie van neutrale verandering "aanloopt" die het efficienter maakt in het verwerken en omzetten van de energiestromen in zijn omgeving tot kopieën van zichzelf. Vanuit dit oogpunt bezien zou het weleens zo kunnen zijn dat de statistische mechanica, die historisch gezien is ontstaan uit de bestudering van de efficientie van stoommachines, relevanter is voor het begrip van de biologische evolutie dan voorheen is vermoed.

Bibliography

[1] C. Adami. Self-organized criticality in living systems. *Phys. Lett. A*, 203:29–32, 1995.

[2] D. Alves and J. F. Fontanari. Error threshold in finite populations. *Physical Review E*, 57:7008–7013, 1998.

[3] A. Babajide, I. L. Hofacker, M. J. Sippl, and P. F. Stadler. Neutral networks in protein space: A computational study based on knowledge-based potentials of mean force. *Folding Design*, 2:261–269, 1997.

[4] T. Back. *Evolutionary algorithms in theory and practice: Evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press, New York, 1996.

[5] R. Balian. Statistical mechanics and the maximum entropy method. In Peter Grassberger and Jean-Pierre Nadal, editors, *From statistical physics to statistical inference and back*, volume 428 of *NATO ASI series*, pages 11–44, 1992.

[6] L. Barnett. Tangled webs: Evolutionary dynamics on fitness landscapes with neutrality. Master's thesis, School of Cognitive Sciences, University of East Sussex, Brighton, 1997. http://www.cogs.susx.ac.uk/ lab/adapt/nnbib.html.

[7] L. Barnett. Ruggedness and neutrality: the NKp family of fitness landscapes. In *ALIFE VI*, 1998. available at: http://www.cogs.susx.ac.uk /users/lionelb/.

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

[9] U. Bastolla, H. E. Roman, and M. Vendruscolo. Neutral evolution of model proteins: Diffusion in sequence space and overdispersion. Technical report, LANL preprint archive: cond-mat/9811267, 1998.

[10] E. Batschelet, E. Domingo, and C. Weissmann. The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene*, 1:27–32, 1976.

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

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

[13] M. C. Boerlijst. *Self-Sructuring: a substrate for evolution*. PhD thesis, Utrecht University, 1994.

[14] L. Boltzmann. *Lectures on gas theory*. University of California press, 1964. Originally published as *Vorlesungen über Gastheorie* by J.A. Barth Leipzig, Part I 1896, Part II 1898.

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

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

[17] M. Bulmer. *Theoretical evolutionary ecology*. Sinauer Associates, Sunderland, Massachusetts, 1994.

[18] L. Chambers, editor. *Practical handbook of genetic algorithms*, Boca Raton, 1995. CRC Press.

[19] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley and Sons, 1991.

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

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

[22] J. P. Crutchfield and M. Mitchell. The evolution of emergent computation. *Proc. Natl. Acad. Sci. U.S.A.*, 92:10742–10746, 1995.

[23] J. P. Crutchfield and C. R. Shalizi. Thermodynamic depth of causal states: Objective complexity via minimal representations. *Physical Review E*, 59:275–283, 1999.

[24] J. P. Crutchfield and E. van Nimwegen. The evolutionary unfolding of complexity. In L. F. Landweber, E. Winfree, R. Lipton, and S. Freeland, editors, *Evolution as Computation*, Lecture Notes in Computer Science, to appear, New York, 1999. Springer-Verlag. Santa Fe Institute Technical Report 99-02-015; adap-org/9903001.

[25] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Physical Review Letters*, 63:105, 1989.

[26] C. R. Darwin. *On the origin of species by means of natural selection*. J. Murray, London, 1859.

[27] 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.

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

[29] R. Dawkins. *The selfish gene*. Oxford University Press, 1976.

[30] S. R. de Groot and P. Mazur. *Non-equilibrium Thermodynamics*. Dover Publications, New York, 1984. Originally published by North-Holland publishing company, Amsterdam, 1962.

[31] B. Derrida and L. Peliti. Evolution in a flat fitness landscape. *Bull. Math. Bio.*, 53(3):355–382, 1991.

[32] M. Eigen. Self-organization of matter and the evolution of biological macromolecules. *Naturwissen.*, 58:465–523, 1971.

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

[34] M. Eigen and P. Schuster. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissen.*, 64:541–565, 1977.

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

[36] E. H. Ekland and D. P. Bartel. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature*, 383:192, 1996.

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

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

[39] W. J. Ewens. *Mathematical population genetics*, volume 9 of *Biomathematics*. Springer-Verlag, 1979.

[40] D. Feldman. Information theory, excess entropy, and computational mechanics. http://hornacek.coa.edu/dave/Tutorial/index.html, 1998.

[41] D. P. Feldman and J. P. Crutchfield. Discovering noncritical organization: Statistical mechanical, information theoretic, and computational views of patterns in one-dimensional spin systems. *J. Stat. Phys.*, 1998, in press. SFI Working Paper 98-04-026.

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

[43] R. P. Feynman. *Feynman Lectures on Physics*. Addison-Wesley, 1970.

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

[45] A. D. Fokker. *Ann. Phys. (Leipzig)*, 43(310), 1914.

[46] W. Fontana, W. Schnabl, and P. Schuster. Physical aspects of evolutionary optimization and adaptation. *Phys. Rev. A*, 40:3301–3321, 1989.

[47] W. Fontana and P. Schuster. Continuity in evolution: On the nature of transitions. *Science*, 280:1451–5, 1998.

[48] W. Fontana, P. F. Stadler, E. G. Bornberg-Bauer, T. Griesmacher, I. L. Hofacker, M. Tacker, P. Tarazona, E. D. Weinberger, and P. Schuster. RNA folding and combinatory landscapes. *Phys. Rev. E*, 47:2083–2099, 1992.

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

[50] C. M. Reidys C. V. Forst and P. Schuster. Replication and mutation on neutral networks of RNA secondary structures. *Bull. Math. Biol.*, submitted, 1998. SFI Working Paper 98-04-036.

[51] C. V. Forst, C. Reidys, and J. Weber. Evolutionary dynamics and optimizations: Neutral networks as model landscapes for RNA secondary-structure folding landscape. In F. Moran, A. Moreno, J. Merelo, and P. Chacon, editors, *Advances in Artificial Life*, volume 929 of *Lecture Notes in Artificial Intelligence*. Springer, 1995. SFI preprint 95-20-094.

[52] H. Frauenfelder, editor. *Landscape Paradigms in Physics and Biology. Concepts, Structures and Dynamics (Papers originating from the 16th Annual International Conference of the Center for Nonlinear Studies. Los Alamos, NM, USA, 13-17 May 1996)*, Amsterdam, 1997. Elsevier Science. Published as a special issue of Physica D **107** no. 2–4 (1997).

[53] F. R. Gantmacher. *Applications of the theory of matrices*. Interscience Publishers, New York, 1959.

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

[55] A. J. M. Garrett. Irreversibility, probability and entropy. In Peter Grassberger and Jean-Pierre Nadal, editors, *From statistical physics to statistical inference and back*, volume 428 of *NATO ASI series*, pages 45–76, 1992.

[56] S. Gavrilets and J. Gravner. Percolation on the fitness hypercube and the evolution of reproductive isolation. *J. Theor. Biol.*, 184:51–64, 1997.

[57] J. W. Gibbs. *Elementary principles in statistical mechanics*. Longmans Green and company, New York, 1928.

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

[59] J. H. Gillespie. *The Causes of Molecular Evolution*. Oxford University Press, New York, 1991.

[60] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA, 1989.

[61] D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. In *Foundations Of Genetic Algorithms*, pages 69–93, San Mateo California, 1991. Morgan Kaufman.

[62] D. E. Goldberg, K. Deb, and B. Korb. Messy genetic algorithms revisited: Studies in mixed size and scale. *Complex Systems*, 4(4):415–444, 1990.

[63] D. E. Goldberg, B. Korb, and K. Deb. Messy genetic algorithms: Motivation, analysis, and first results. *Complex Systems*, 3:493–530, 1990.

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

[65] W. Grüner, R. Giegerich, D. Strothmann, C. Reidys, J. Weber, I. L. Hofacker, P. F. Stadler, and P. Schuster. Analysis of RNA sequence structure maps by exhaustive enumeration: I. neutral networks; II. structure of neutral networks and shape space covering. *Monatsh. Chem.*, 127:355–389, 1996.

[66] R. Gutell, A. Power, G. Z. Hertz, E. J. Putz, and G. D. Stormo. Identifying constraints on the higher-order structure of RNA: Continued development and application of comparative sequence analysis methods. *Nucleic Acids Res.*, 20:5785–5795, 1993.

[67] J. B. S. Haldane. The effect of variation on fitness. *Amer. Natur.*, 71:337–349, 1937.

[68] T. E. Harris. *The theory of branching processes*. Dover publications, New York, 1989.

[69] D. L. Hartl and A. G. Clark. *Principles of population genetics*. Sinauer associates, second edition, 1989.

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

[71] I. L. Hofacker, W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, 125(2):167–188, 1994.

[72] J. Hofbauer and K. Sigmund. *Evolutionary games and population dynamics*. Cambridge University Press, 1998.

[73] P. Hogeweg. Multilevel evolution: replicators and the evolution of diversity. *Physica D*, 75:275–291, 1994.

[74] J. H. Holland. *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, MA, 1992. Second edition (First edition, 1975).

[75] B. D. Hughes. *Random Walks and Random Environments*, volume II, chapter 7. Clarendon Press, Oxford, 1996.

[76] M. Huynen. *Evolutionary dynamics and pattern generation in the sequence and secondary structure of RNA*. PhD thesis, Universiteit Utrecht, The Netherlands, 1993.

[77] M. Huynen. Exploring phenotype space through neutral evolution. *J. of Mol. Evol.*, 43:165–169, 1996.

[78] M. Huynen, T. Doerks, F. Eisenhaber, C. Orengo, S. Sunyaev, Y. Yuan, and P. Bork. Homology-based fold predictions for *Mycoplasma genitalium* proteins. *J. Mol. Biol.*, 280:323–326, 1998.

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

[80] M. A. Huynen, R. Gutell, and D. Konings. Assessing the reliability of RNA folding using statistical mechanics. *J. Mol. Biol.*, 267:1104–12, 1997.

[81] M. A. Huynen and P. Hogeweg. Pattern generation in molecular evolution: Exploitation of the variation in RNA landscapes. *J. Mol. Evol.*, 39:71–79, 1994.

[82] M. A. Huynen, D. Konings, and P. Hogeweg. Multiple coding and the evolutionary properties of RNA secondary structure. *J. Theor. Biol.*, 165:251–267, 1993.

[83] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.

[84] E. T. Jaynes. *Papers on probability, statistics, and statistical physics*, volume 158 of *Synthese library*. D. Reidel publishing company, Dordrecht Holland, 1983. R.D. Rosenkrantz, editor.

[85] E. T. Jaynes. The evolution of Carnot's principle. In *Maximum Entropy and Bayesian Methods in Science and Engineering*. Kluwer Academic Publishers, Dordrecht Holland, 1988.

[86] E. T. Jaynes. Clearing up mysteries - the original goal. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods*, pages 1–27. Kluwer Academic Publishers, 1989. also available at http://bayes.wustl.edu/etj/node1.html.

[87] E. T. Jaynes. *Probability theory - the logic of science*. available at: ftp://bayes.wustl.edu/pub/Jaynes/book.probability.theory, 1996.

[88] S. A. Kauffman and S. Levin. Towards a general theory of adaptive walks in rugged fitness landscapes. *J. Theo. Bio.*, 128:11–45, 1987.

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

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

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

[92] 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.

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

[94] J. R. Koza. *Genetic programming: On the programming of computers by means of natural selection.* MIT Press, Cambridge, MA, 1992.

[95] R. Lande. Expected time for random genetic drift of a population between stable phenotype states. *Proc. Natl. Acad. Sci. USA*, 82:7641–7645, 1985.

[96] L. F. Landweber and I. D. Pokrovskaya. Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: The spandrels of RNA evolution. *Proc. Natl. Acad. Sci. USA*, 96:173–178, 1999.

[97] R. E. Lenski and M. Travisano. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences USA*, 91(15):6808–6814, 1994.

[98] I. Leutäusser. Statistical mechanics of Eigen's evolution model. *J. Stat. Phys.*, 48:343–360, 1987.

[99] 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.

[100] C. A. Macken and A. S. Perelson. Protein evolution in rugged fitness landscapes. *Proc. Nat. Acad. Sci. USA*, 86:6191–6195, 1989.

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

[102] D. R. Mills, F. R. Kramer, and S. Spiegelman. Complete nucleotide sequence of a replicating rna molecule. *Science*, 180:916–927, 1973.

[103] M. Mitchell. *An Introduction to Genetic Algorithms.* MIT Press, Cambridge, MA, 1996.

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

[105] 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, 1993. Morgan Kauffman.

[106] N. A. Moran. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. USA*, 93:2873–2878, 1996.

[107] C. M. Newman, J. E. Cohen, and C. Kipnis. Neo-darwinian evolution implies punctuated equilibrium. *Nature*, 315:400–401, 1985.

[108] M. Newman and R. Engelhardt. Effect of neutral selection on the evolution of molecular species. *Proc. R. Soc. London B.*, 256:1333–1338, 1998.

[109] A. E. Nix and M. D. Vose. Modeling genetic algorithms with Markov chains. *Ann. Math. Art. Intel.*, 5, 1991.

[110] M. Nowak and P. Schuster. Error thresholds of replication in finite populations, mutation frequencies and the onset of Muller's ratchet. *J. Theo. Biol.*, 137:375–395, 1989.

[111] T. Ohta. Slightly deleterious mutant substitutions in evolution. *Nature*, 247:96–98, 1973.

[112] T. Ohta and J. H. Gillespie. Development of neutral and nearly neutral theories. *Theo. Pop. Bio.*, 49:128–142, 1996.

[113] E. Ott. *Chaos in Dynamical Systems*. Cambridge University Press, 1993.

[114] 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*. Addison-Wesley, 1989.

[115] M. Planck. *Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl.*, page 324, 1917.

[116] A. Prügel-Bennett. Modelling evolving populations. *J. Theo. Bio.*, 185:81–95, 1997.

[117] A. Prügel-Bennett and J. L. Shapiro. Analysis of genetic algorithms using statistical mechanics. *Phys. Rev. Lett.*, 72(9):1305–1309, 1994.

[118] 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.

[119] 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.

[120] 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.

[121] C. M. Reidys. *Neutral networks of RNA secondary structures*. PhD thesis, Mathematics Faculty. Friedrich Schiller Universität Jena., 1995.

[122] C. M. Reidys and S. Fraser. Evolution on random structures. Technical Report 95-11-087, Santa Fe Institute, 1995.

[123] C. M. Reidys, P. F. Stadler, and P. Schuster. Generic properties of combinatory maps—Neutral networks of RNA secondary structures. *Bull. Math. Biol.*, 59:339–397, 1997.

[124] M. Ridley. *Evolution*. Blackwell scientific publications, 1993.

[125] Erwin Schrödinger. *What is life?* Cambridge University press, reprint edition, 1992.

[126] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. (London) B*, 255:279–284, 1994.

[127] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27, 1948.

[128] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Champaign-Urbana, 1962.

[129] J. Maynard Smith. *Evolution and the theory of games*. Cambridge University Press, 1982.

[130] J. Maynard Smith and E. Szathmáry. *The major transitions in evolution*. Oxford University Press, 1997.

[131] R. F. Streater. *Statistical dynamics, a stochastic approach to nonequilibrium thermodynamics*. Imperial college press, 1995.

[132] R. F. Streater. Statistical dynamics and information geometry. In Hanna Nencka and Jean-Pierre Bourguignon, editors, *Geomertry and Nature*, volume 203 of *Contemporary mathematics*, 1997.

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

[134] D. Thierens and D. E. Goldberg. Mixing in genetic algorithms. In S. Forrest, editor, *International Conference on Genetic Algorithms 5*, pages 38–45. Morgan Kaufman, 1993.

[135] C. Türk. Using the SELEX combinatorial chemistry process to find high affinity nucleic acid ligands to target molecules. *Methods Mol. Biol.*, 67:219–230, 1997.

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

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

[138] M. D. Vose. Modeling simple genetic algorithms. In L. D. Whitley, editor, *Foundations of Genetic Algorithms 2*, San Mateo, CA, 1993. Morgan Kauffman.

[139] M. D. Vose and G. E. Liepins. Punctuated equilibria in genetic search. *Complex Systems*, 5:31–44, 1991.

[140] A. Wagner and P. F. Stadler. Viral RNA and evolved mutational robustness. Technical Report 99-02-010, Santa Fe Institute, 1999.

[141] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, and M. Zuker. Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA*, 91:9218–9222, 1994.

[142] J. Weber. *Dynamics of Neutral Evolution. A case study on RNA secondary structures.* PhD thesis, Biologisch-Pharmazeutischen Fakultät der Friedrich Schiller-Universität Jena, 1996. http://www.tbi.univie.ac.at/ papers/PhD_theses.html.

[143] G. Weisbuch. *Complex Systems Dynamics: An Introduction to Automata Networks*, volume 2 of *Santa Fe Institute Studies in the Sciences of Complexity, Lecture Notes*. Addison-Wesley, Reading, Massachusetts, 1991.

[144] G.C. Williams. *Adaptation and Natural Selection: A critique of some current evolutionary thought*. Princeton university press, 1966.

[145] L. Wittgenstein. *Zettel, fragment number 608*. University of California press, Berkeley and Los Angeles, California, 1967. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by G.E.M. Anscombe.

[146] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Trans. Evol. Comp.*, 1:67–82, 1997.

[147] M. C. Wright and G. F. Joyce. Continuous in vitro evolution of catalytic function. *Science*, 276:614–617, 1997.

[148] S. Wright. Evolution in Mendelian populations. *Genetics*, 16:97–159, 1931.

[149] 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.

[150] S. Wright. Character change, speciation, and the higher taxa. *Evolution*, 36:427–43, 1982.

Nawoord

Het afronden van een promotie-onderzoek is een "mijlpaal", zoals dat heet. Dat het ooit nog zover heeft mogen komen! Aan mijn educationele loopbaan komt dan nu toch echt een einde. Meer diploma's vallen er niet te halen. Het zou echter een vergissing zijn te denken dat ik mij de afgelopen jaren in een woestijndorp heb opgesloten met als doel om uiteindelijk een diploma in ontvangst te mogen nemen. Ik ben de afgelopen vier jaar in de gelukkige omstandigheid geweest mij volledig te mogen wijden aan de studie van precies die onderwerpen die ik zelf het meest interessant en relevant vindt. In die zin is dit proefschrift evenzozeer een uitdrukking van mijn persoonlijke interesses als van het gedane onderzoek.

Er is een vrij berucht debat in de psychologie dat het *nature-nurture* debat wordt genoemd. Dit debat draait om de vraag of het gedrag en uiteidelijke succes van individuen voornamelijk afhankelijk is van de kwaliteit van de genen waarmee ze geboren worden, of veeleer afhankelijk is van de opvoeding die ze genoten hebben. Hoe het ook zij, het staat in ieder geval vast dat, als ik er in geslaagd mocht zijn een aardig proefschrift te schrijven, ik dit voornamelijk aan mijn ouders te danken heb. Zij zijn verantwoordelijk voor zowel mijn *nature* als *nurture*. Voor mijn genen durf ik niet te spreken. Wat betreft de *nurture* ben ik er van overtuigd ben dat een open geest, een onbevooroordeelde blik, een gezonde dosis zelfvertrouwen, scepsis, en vooral een onbezwaard gemoed, onontbeerlijk zijn voor het doen van wetenschappelijk onderzoek. Pap, mam, dat bovenal heb ik aan jullie te danken. Over alle steun die ik altijd van jullie heb ontvangen hoef ik het dan niet eens te hebben.

Most of the work on this thesis was done at the Santa Fe Institute in New Mexico. I feel privileged to have been able to work there. I can hardly imagine a more stimulating environment to do scientific research. Many of the locals and visitors at SFI have helped shape my ideas about evolution and its mathematical formalization. I thank you all... you know who you are. My co-author Martijn Huynen deserves special mention: Keep those good ideas coming in the future. I further thank Lee Altenberg, Walter Fontana, Sergey Gavrilets, Melanie Mitchell, Mark Newman, Richard Palmer, and Jon Shapiro for many useful comments and discussions. Of course, I reserve some special thanks to my advisor Jim Crutchfield. I have enjoyed our collaboration very much. Apart from helping me with many things, you always managed to maintain a stimulating and constructive spirit towards our work. It has become hard to judge to what extent my views on the important issues are influenced by you, or just happen to be similar to yours. I hope that that can be considered a compliment.

I am very grateful to the members of my reading comitee: Bernard Derrida, Peter Schuster, and Jon Shapiro. I imagine that it is not a very thankful job to have to plough through 300 pages and not even get to vent specific complaints in the end.

Terug in Nederland wil ik alle leden van de vakgroep theoretische biologie en bioinformatica bedanken voor de plezierige twee periodes in Utrecht waaraan jullie in hoge mate hebben bijgedragen. Tussen twee haakjes: Ik geloof niet dat het woord gezellig onvertaalbaar is. Het komt gewoon niet zoveel voor buiten Nederland.

Paulien, ik ben veel dank aan jou verschuldigd. Zonder jou was er niets van mijn ondernemingen terecht gekomen; vanaf het moment dat je instemde mijn begeleider te zijn (zonder me zelfs maar te kennen!) tot op de laatste lootjes waar je me behoedt hebt voor de grootste blunders. Het meest heb ik echter plezier beleefd aan onze urenlange discussies over evolutionaire dynamica.

And finally my paranimf for being my best friend, and my sweet paranimfette, for being my lady.