Landscapes and Effective Fitness

Peter F. Stadler[¶,†,‡,*] and Christopher R. Stephens[§]

[*,¶]Lehrstuhl f. Bioinformatik, Institut für Informatik, Universität Leipzig, Kreuzstraße 7b, D-04103 Leipzig, Germany.
peter.stadler@bioinf.uni-leipzig.de

[†]Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria

[‡]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

[§]NNCP, Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, A. Postal 70-543, México D.F. 04510
stephens@nuclecu.unam.mx

[*]Address for correspondence

Abstract. The concept of a fitness landscape arose in theoretical biology, while that of effective fitness has its origin in evolutionary computation. Both have emerged as useful conceptual tools with which to understand the dynamics of evolutionary processes, especially in the presence of complex genotype-phenotype relations. In this contribution we attempt to provide a unified discussion of these two approaches, discussing both their advantages and disadvantages in the context of some simple models. We also discuss how fitness and effective fitness change under various transformations of the configuration space of the underlying genetic model, concentrating on coarse graining transformations and on a particular coordinate transformation that provides an appropriate basis for illuminating the structure and consequences of recombination.

1. Introduction

Evolution theory has as its cornerstone the concept of *fitness*. Fitness is traditionally defined as the relative reproductive success of a genotype as measured by survival, fecundity or other life history parameters. This definition is less than satisfactory, however, when one is confronted with the problem of actually measuring fitness in a given system. In its simplest form fitness is quantified as the relative number of fertile offspring produced by one genotype versus another [1] ("reproductive fitness"). However, there are many components that affect this number, such as the probability to survive to reproductive age [2] ("survival fitness" or "viability"), differences in mating success or in the average number of produced offspring ("fertility" [3]). In evolutionary computation, fitness is typically portrayed in terms of viability.

These two approaches to quantifying fitness are quite different. First of all, survival fitness is a property of an individual, in that it does not depend on other genotypes, even though the reproductive fitness function may reflect "environmental" effects. Reproductive fitness is much less straightforward than it might appear, since it depends on what counts as an offspring. As long as selection is the dominating genetic operator and mixing operators such as mutation and recombination can be neglected, all offspring will be identical copies of the parent; in this case survival fitness and reproductive fitness can be scaled to have the same numerical values. When mixing operators dominate, as in the case of neutral evolution, however, the two concepts can be very different. For instance, it is not even clear whether one should count any offspring or only identical copies. Notions of lineage fitness [4] and inclusive fitness [5, 6, 7], where the fitness of an allele or set of alleles is measured not only by its effect on an individual but also by its effect on related individuals that also possess it, have their origin in this question. Thus, in talking about fitness one must distinguish between individuals and populations, as selective values of genotypes are based on competition among genotypes, but may not adequately reflect their effect on the population, such as in the case of altruism.

Fitness landscapes [8] have played an important role in improving our understanding of evolutionary dynamics. Given the many different meanings and associated mathematical representations of fitness however, we need to understand exactly which fitness is being represented. Fitness landscapes are by construction a static concept that assign fitness values to the points of an underlying configuration space and hence do not represent (except in some approximation) situations where the represented fitness measure is dynamical. Irrespective, the ultimate goal is to understand population flows on the configuration space. Standard intuition views a fitness landscape as a rugged terrain where populations flow towards fitness peaks. Thus, natural selection can be viewed as a type of "hill climbing" on this topography. Although, a very powerful paradigm it behooves us to critically analyze under what circumstances this intuitive picture is valid and, in cases where it is not, to see if there exists an alternative conceptual framework.

Although the fundamental level is the "microscopic" genotypic one, fitness, and hence the corresponding fitness landscape, is usually taken as acting at the more "macroscopic" phenotypic one. The action of selection on the genotype is then determined via a genotype-phenotype map, which may or may not be simple (in biology it is almost inevitably exceedingly complex while in evolutionary computation it is almost always trivial). All genotypes corresponding to the same phenotype have the same survival fitness and hence, perhaps somewhat loosely, we may talk of a genotype-phenotype "symmetry" which can be broken "spontaneously" via finite size effects, i.e. neutral drift. The phenotype then, seen from the genotypic level, is a coarse grained effective degree of freedom as far as selection is concerned. Again, when we consider a simple model consisting of a small number of loci, this too corresponds to a coarse graining where we imagine having averaged over the effects of other loci, supposing that for the observables of interest only the loci under consideration are of relevance. The fitness landscape on this reduced space is posited as being static. That is part of the model. However, this is pure supposition. The nature of the coarse grained landscape should in principle be derived by coarse graining the landscape of the underlying more fundamental model. As we shall see, this inevitably, except in some approximation, leads to a coarse grained fitness landscape where the fitness of an individual depends on the state of the rest of the population.

Additionally, the classical fitness concept, and associated fitness landscape, do not take into account the important effect the mixing genetic operators may have in determining the complete reproductive success of an individual. In particular, the effect of these genetic operators, as we shall see below, can be such that population flows on the standard fitness landscape cannot be understood with any degree of intuition. In fact, the flows can be quite counterintuitive, leading to situations where populations flow *against* the fitness gradient. The mixing operators can also lead to directed flows on neutral networks due to an "induced" breaking of the genotype-phenotype symmetry and leading to a self-organization of the genotype-phenotype map [9, 10, 11, 12]. Such phenomenon, unlike the case of population flow due to positive selection cannot be naturally understood in terms of "hill climbing" on a standard fitness landscape. However, all these phenomena can be intuitively understood within the framework of a different paradigm - *effective fitness* [13, 14, 15, 16], albeit at a cost: effective fitness is not a constant quantity but rather depends on the state of the entire system and hence is intrinsically time dependent, see section 7. However, as we emphasized above, any coarse grained fitness landscape is dynamic, and only approximately static in some regimes. Effective fitness also provides a quantitative measure for the degree of symmetry breaking of the genotype-phenotype map by mixing operators.

2. Evolutionary Dynamics

Before passing to a discussion of fitness landscapes let us remind ourselves of the task at hand: to understand the dynamical evolution of a population of "types"[1], $x \subset X$, where X is the configuration space of types and has dimensionality $\dim(X)$. Since genetic variation is generated independently from the natural selection acting on it, the generic structure of an evolutionary model in discrete time can be written as

$$\mathbf{P}(t+1) = \mathfrak{S}\left(\mathbf{P}(t), \mathsf{f}\right) \circ \mathfrak{T}\left(\mathbf{P}(t), \mathsf{p}\right) \,, \tag{1}$$

where $\mathbf{P}(t)$ is the vector of type frequencies at generation t [17, 3]. As usual, $\circ$ denotes the Schur (Hadamard, component-wise) product of vectors. The transmission term $\mathfrak{T}(P(t), \mathsf{p})$ describes the probability of transforming one type into another one by mutation, recombination, or other genetic operators [18] and hence determines the structure on the set of all vectors of haplotype frequencies, which is a simplex of the form

$$\left\{ (P_1, \dots, P_{\dim(X)}) \middle| \sum_x P_x = 1, \; P_x \geq 0 \; \forall x \right\} \tag{2}$$

[1]which can either be quantitative phenotypic traits (such as weight or length of limb) or a discrete genetic structure

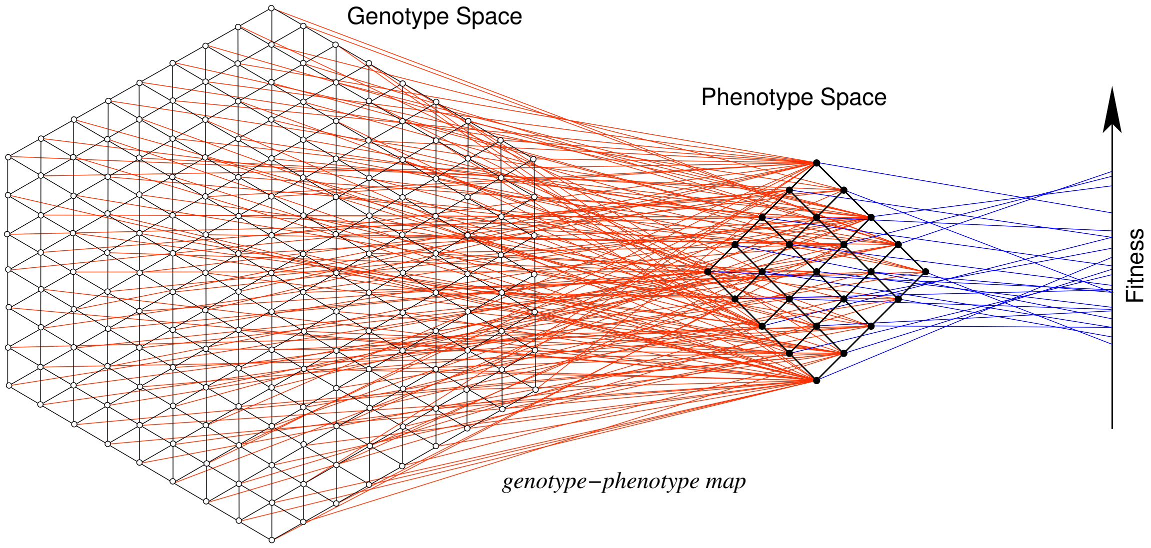

Figure 1. Fitness and Genotype-Phenotype map. Redrawn after [20]

where P_x is the frequency of type x. This structure can be understood in terms of certain classes of algebraic structures [19] that depend on the details of the transmission mechanism that is encoded by the parameters p. The term $\mathfrak{S}(\mathbf{P}, \mathsf{f})$ describes the selection forces acting on P. The parameters f determine the fitness function.

An important element in understanding the dynamics is a notion, $\mathcal{X}$, of neighborhood, nearness, distance, or accessibility on X. As we shall see, different genetic operators are often most naturally associated with different notions of nearness.

The mathematical representation of the different genetic operators is fairly simple. For selection we introduce a fitness function associated with types, $f_X : X \longrightarrow \mathbb{R}^+$. The real domain of f_X of course may be different from $\mathbb{R}^+$; e.g. the integers over a finite interval. It is conventional to consider fitness as a function of the phenotype, hence, if we wish to know how fitness acts at the level of genotypes one must define a phenotype-genotype map, $\phi : X \longrightarrow Y$, where X and Y are the genotype and phenotype configuration spaces respectively. The existence of ϕ and f_Y allows one to define on X a fitness function, $f_X = f_Y \bullet \phi$, induced by the action of ϕ and f_Y. In the case of selection only dynamics f_X gives a direct measure of the total reproductive success of a type, i.e. the total number of offspring of a type. However, in the presence of other operators it is more correctly interpreted as a measure of the viability (probability to survive to reproductive age) of a type and is but one component in determining the total reproductive success of a type.

Whenever the genotype-phenotype map is non-injective (many-to-one) the function f_Y will be degenerate, many genotypes corresponding to the same fitness value, and hence non-invertible, Fig. 1. This may also occur for other reasons. Thus, fitness defines an equivalence relation on X. The members of such an equivalence class are equivalent or *neutral* w.r.t. selection, i.e. they're all equally fit. In the language of physics one may refer to this equivalence of a set of genotypes under the action of reproductive selection as a "symmetry" between them. By definition, reproductive

GCGGGAAUAGCUCAGUUGGUAGAGCACGACCUUGCCAAGGUCGGGGUCGCGAGUUCGAGUCUCGUUUCCCGCUCCA

Figure 2. Three different coarse grainings of RNA structures. The sequence at the top forms the atomic structure on the right. The secondary structure on the left is defined by the Watson-Crick and GU base pairs that from when the molecule folds back onto itself. An intermediate level of description is the "ribbon-structure" in the middle. This ribbon structure corresponds also to the resolution of many models of proteins.

selection preserves this symmetry. A simple example of this would be the standard synonym symmetry of the genetic code. A similar artificial genotype/phenotype system is used in [21] for genetic programming.

To what extent a GP-map is many-to-one is of course a matter of resolution. In the case of RNA molecules and proteins, for instance, each structure that is specified at atomic resolution uniquely determines the sequence. In many cases, however, a coarse grained representation is more useful, e.g., for the comparison of protein folds or the analysis of conserved features in functional RNAs [22]. In the case of RNA a very natural and particularly useful level of description is that of secondary structure, Figure 2. Secondary structures, are routinely used to display, organize, and interpret experimental findings, they are oftentimes conserved over evolutionary times scales, and *in vitro* selection experiments with RNA more often than not yield families of selected sequences that share distinctive secondary structure features. Furthermore, secondary structures are folding intermediates in the sense that the secondary base pairs typically form before so-called tertiary contacts complete the formation of the three-dimensional structure. Different secondary structures typically lead to distinct 3D shapes although there are examples of RNA molecules with significantly different secondary structure which exhibit similar 3D structures and the same function [23].

Whether there exists a non-trivial genotype-phenotype map is therefore to a large extent problem dependent. The notion of a protein fold, for instance, necessarily implies a level of coarse graining since it makes no sense to say that "two different sequences have the same structure" at atomic resolution. Once we have fixed a level of resolution we implicitly assume that differences at finer scales are irrelevant for the dynamic processes at hand. Computational investigations of the GP-map of both RNA [24, 25, 26, 27, 28, 29] and proteins [30, 31, 32, 33] show that the coarse graining has far-reaching consequences. The coarse graining process defines a partition of the type space X into classes of genotypes with the same phenotype. Not surprisingly, the organization of these classes in genotype space is crucial for the dynamics. We will return to this topic again in section 5.

Mutation is conveniently defined via an operator $\mathbf{M}$ whose entries $\mathbf{M}_{xy}$ are the probabilities that a type y mutates to a type x. An often used representation for $\mathbf{M}_{xy}$ is

$$\mathbf{M}_{xy} = p^{d^H(x,y)}(1-p)^{N-d^H(x,y)}, \tag{3}$$

where $d^H(x, y)$ is the Hamming distance between types x and y and p is the mutation rate. Similarly, we may represent recombination via an operator $\mathbf{R}_{xyz}(m)$ where the first index represents the target (child) type and y and z are the parent types. The argument m refers to the fact that a recombination operator is not uniquely specified by the target and parents. For instance, the target type `1111` can be obtained from parents `1101` and `1011` by combining the first and second loci of the first parent with the third and fourth of the second. Alternatively, it may be obtained by taking the first and third loci from the second parent and the second and fourth from the first parent. A convenient way of representing many types of recombination operator is via a recombination "mask", m, which is an N-bit binary sequence wherein a `1` signifies take the allele from the corresponding locus of the first parent type while a zero signifies take it from the second parent type. For example, with the mask `1010` and parent types `1111` and `0000` the resultant offspring are `1010` and `0101`. There are 2^N possible masks.

Generically, one may write $\mathbf{R}_{xyz}(m)$ in the form $p_c(m)\mathcal{C}_{xyz}(m)$ where $p_c(m)$ is the probability to implement the mask m and $\mathcal{C}_{xyz}(m)$ depends purely on the configurations of the types x, y and z and the mask m. The explicit form of the $\mathcal{C}_{xyz}(m)$ depends on how we wish to model the set of possible recombination events. Often, it is used only to model "construction events", i.e., recombination events that lead to an increase in the number of type x, such as in the above example when the target type is `1010`. In this case $\mathcal{C}_{xyz}(m) = 1$ iff x can be formed from y and z via the mask m. Generically, this is very unlikely as the vast majority of recombination events do not lead to the desired target.

The structure of recombination operators has been investigated from an algebraic point of view in [19, 34, 35, 36] and more recently as in the context of generalized topologies [37, 38]. It is interesting to note that recombination operators are closely related to so-called "betweenness relations" and "transit function" in graph theory [39, 40].

We will not follow this line of investigation here and restrict ourselves to a matrix representation. For binary alleles, for a given x and m, $\mathcal{C}_{xyz}(m)$ is a (2^N-1)-dimensional square matrix[2]. However, only of the order of 2^N matrix elements are non-zero. Thus, the microscopic representation is very inefficient, there being very few ways of creating a given target type by recombination of parent types. The vast majority of recombination events are neutral in that they lead to no non-trivial interaction. These comments also hold for more complicated type representations, such as by variable-length strings or by trees, as in the case of Genetic Programming. The $\mathcal{C}_{xyz}(m)$ can also be used to model "destruction events", such as in the above example again but now taking as target type `1111`. Now, $\mathcal{C}_{xyz}(m) = -1$ iff $y = x$ or $z = x$ and x in the parental generation is destroyed via the action of the mask m. Both construction and

[2] We omit the trivial case where x is one or both of the parent types as in this case there is no net gain, hence the dimension is only $2^N - 1$.

destruction events, and hence all non-trivial recombination events that lead to losses or gains, can be modeled for a given target x via one 2^N-dimensional square matrix $\mathcal{C}_{xyz}(m)$, where $\mathcal{C}_{xxx}(m) = 0$, destruction events appear in the elements $\mathcal{C}_{xxz}(m)$ and $\mathcal{C}_{xyx}(m)$ and construction events in the rest of the matrix. Usually, $\mathbf{M}_{xy}$ and $\mathbf{R}_{xyz}$ possess the symmetry properties $\mathbf{M}_{xy} = \mathbf{M}_{yx}$ and $\mathbf{R}_{xyz} = \mathbf{R}_{xzy}$.

Note that both selection and mutation are essentially unary operators, taking only one input whereas recombination is binary taking both "parent" types as input. If we take a given ordering for the operators we may write a more explicit equation than that of equation (1). Taking the ordering

$$\text{selection} \to \text{recombination} \to \text{mutation}$$

we may write

$$P_x(t+1) = \sum_y \mathbf{M}_{xy} P_y^c(t) \tag{4}$$

where $P_x^c(t)$ is the proportion of type x after selection and recombination. Explicitly $P_x^c(t)$ is given by

$$P_x^c(t) = (1 - p_c) P_x'(t) + \sum_{m=1}^{2^N} \sum_{y,z} \mathbf{R}_{xyz}(m) P_y'(t) P_z'(t) \tag{5}$$

where $P_x'(t)$ is the proportion of type x after selection only and $p_c = \sum_m p_c(m)$ is the probability to recombine via any mask. Normally in biological applications $p_c = 1$. In the case of proportional selection $P_x' = (f_x/\bar{f}(t))P_x$, where

$$\bar{f}(t) = \sum_x f_x P_x(t) \tag{6}$$

is the average population fitness. By writing the above in this form $\mathbf{R}_{xyz}(m)$ models both non-trivial ($y \neq x$, $z \neq x$) and trivial ($y \neq x$, $z = x$; $y = x$, $z \neq x$; $y = x$, $z = x$) construction events. Once again, however, we emphasize that, for a given x, $\mathbf{R}_{xyz}(m)$ is a very sparse matrix and hence the microscopic representation is very inefficient in practice.

Note that, although we have written down (4) and (5) for the case of fixed-length representations and homologous recombination representable by masks, the same functional form is valid for more general type representations, such as variable-length chromosomes and trees, and also for more general recombination operators, such as sub-tree crossover. In the case of non-homologous recombination recombination distributions must be modeled by recombination "modes" which are more general than masks.

Usually, selection is taken to be the dominant operator and this is certainly the standard neo-Darwinian point of view. However, there are many important situations, such as neutral evolution, where this is not the case and hence any intuition gleaned from the selection dominant case, as we shall see, can be quite misleading.

For concreteness, we will frequently restrict attention to the case of a fixed number of genetic loci, N, binary alleles and haploid reproduction. In this case we may naturally represent X as a N-dimensional hypercube. A natural metric in this case is Hamming distance, types associated with adjacent vertices being Hamming distance one apart. The N loci form a complete orthonormal basis for the hypercube.

3. Coarse graining and Coordinate Transformations

Given that genetic dynamics takes place on a configuration space X there are two important questions to answer: first, for a given model, what is X?; and secondly, given an X what is the most appropriate "coordinate" system, $\mathcal{M}_X$, for X? Both of these questions are intimately related to what are the appropriate effective degrees of freedom of the model, as both the choice of X and $\mathcal{M}_X$ affect how easily the dynamics may be written in terms of them. We will generically consider three types of transformation: coarse grainings, "coordinate" transformations, and embeddings, concentrating mainly on the first two.

3.1. Coarse Graining. The generic dynamics discussed in the previous section, taking the case of haploids with binary alleles, is described by 2^N coupled, non-linear difference equations representing the microscopic degrees of freedom, i.e. the completely specified sequences themselves. In the absence of recombination, the equations are essentially linear and the problem reduces down to finding the eigenvalues and eigenvectors of the selection-mutation matrix. However, save in very simple problems, such as a linear fitness landscape, even this simpler problem is formidable. Recombination adds yet another layer of complexity. Naturally, in such problems one always wishes to find the correct effective degrees of freedom so as to be able to affect an effective reduction in the dimensionality of the problem. Such a reduction can be affected by an appropriate coarse graining.

We can formalize these considerations by introducing a general coarse graining operator, $\mathcal{R}(\eta,\eta')$, which coarse grains from the variable $\eta \in X_\eta$ to the variable $\eta' \in X_{\eta'} \subset X_\eta$. Thus, the action of $\mathcal{R}$ is a projection. Given two such coarse grainings we have

$$\mathcal{R}(\eta,\eta')P_\eta(t) = P_{\eta'}(t) \qquad\qquad \mathcal{R}(\eta,\eta'')P_\eta(t) = P_{\eta''}(t) \tag{7}$$

However, given that $\mathcal{R}(\eta',\eta'')P_{\eta'}(t) = P_{\eta''}(t)$ we deduce that

$$\mathcal{R}(\eta,\eta'') = \mathcal{R}(\eta,\eta')\mathcal{R}(\eta',\eta'') \tag{8}$$

i.e., the space of coarse grainings has a semi-group structure. This type of structure is well known in theoretical physics and is known, by an abuse of language, as the Renormalization Group, see e.g. [41]. The naturalness of a particular coarse graining transformation will be to a large extent determined by how the transformed dynamics looks.

Considering (1), then given that $\mathcal{R}(\eta,\eta')P_\eta(t) = P_{\eta'}(t)$ the dynamics under a coarse graining is governed by $\mathcal{R}(\eta,\eta')\mathfrak{S}\left(\mathbf{P}_\eta(t),\mathsf{f}\right) \circ \mathfrak{T}\left(\mathbf{P}_\eta(t),\mathsf{p}\right)$, where $\mathfrak{S}\left(\mathbf{P}_\eta(t),\mathsf{f}\right)$ and $\mathfrak{T}\left(\mathbf{P}_\eta(t),\mathsf{p}\right)$ are the dynamical operators associated with the variables η. If this can be written in the form $\mathfrak{S}\left(\mathbf{P}_{\eta'}(t),\mathsf{f}'\right) \circ \mathfrak{T}\left(\mathbf{P}_{\eta'}(t),\mathsf{p}'\right)$ with suitable "renormalizations", $\mathsf{f} \longrightarrow \mathsf{f}'$ and $\mathsf{p} \longrightarrow \mathsf{p}'$ of the model's parameters, then the dynamics is form covariant or invariant under this coarse graining. In the case of selection only, the coarse graining transforms the fitness

$$f_\eta \longrightarrow f_{\eta'} = \mathcal{R}(\eta,\eta')f_\eta = \sum_{\eta\in\eta'} f_\eta P_\eta(t) / \sum_{\eta\in\eta'} P_\eta(t). \tag{9}$$

An important point to note here is that, generically, a coarse graining gives rise to a *time dependent* coarse-grained fitness.

Of course, there are many types of coarse graining procedure all of which lead to a dimensional reduction. Such reductions can sometimes come about in a relatively "trivial" fashion, such as in the case of the genotype-phenotype map. The dynamics in invariant e.g. in this case of selection in the absence of mixing operators. In fact, it is strictly invariant not just form invariant, as there is no renormalization necessary of any parameter or variable and we have $f_{\eta'} = \mathcal{R}(\eta,\eta')f_\eta = f_\eta$, where, here, η' represents the phenotype and η the genotype. A concrete example is that of the Eigen model [42, 43], where the fitness landscape is degenerate for all genotypes except one, the master sequence. At the level of selection only, given that there are only two phenotypes, there is a reduction in the size of the configuration space from 2^N to 2, i.e. a reduction in the number of degrees of freedom from N to 1. However, if we include in the effect of mutation we see that there is an induced breaking of the genotype-phenotype symmetry due to the fact that sequences close to the master sequence in Hamming distance have more offspring than the equally fit genotypes that are further away. As we shall see, this has a ready interpretation in terms of "effective" fitness. In this case $\mathcal{R}(\eta,\eta')\mathfrak{S}(\mathbf{P}_\eta(t),\mathsf{f}) \circ \mathfrak{T}(\mathbf{P}_\eta(t),\mathsf{p}) \neq \mathfrak{S}(\mathbf{P}_{\eta'}(t),\mathsf{f}') \circ \mathfrak{T}(\mathbf{P}_{\eta'}(t),\mathsf{p}')$.

Another important class of coarse grainings is that corresponding to "schemata", which we will denote[3] by α, and which represent specific, fixed subsets of loci. The marginal probability, $P_\alpha(t)$, represents the relative frequency of the schema α at time t. A specific schema is determined by summing over all possible allele values at those loci that are not part of the schema. In this case the action of the coarse graining operator is: $\mathcal{R}(x,\alpha)P_x(t) = P_\alpha(t) = \sum_{x\in\alpha} P_x(t)$. One may denote a locus that has been summed over by a $*$. Thus, $\texttt{11}*$ represents the two chromosomes $\texttt{111}$ and $\texttt{110}$. The number of definite loci of the schema defines its order, N_2, while the distance between the outermost defining loci defines its length l. Thus, $*\texttt{11}**\texttt{0}**$ has $N_2 = 3$ and $l = 5$. Schemata have a simple geometric interpretation in the binary case, a particular schema being represented by an $(N - N_2)$-dimensional hyperplane in X which passes through the 2^{N-N_2} vertices that represent the loci that have been coarse grained. A schema partition then consists of 2^{N_2} of such 2^{N-N_2}-dimensional hyperplanes. Thus, there is an effective dimensional reduction from a 2^N-dimensional configuration space X to a 2^{N_2}-dimensional one, X_α, and a corresponding reduction in the number of degrees of freedom from N to N_2.

Unlike the simple case of the coarse graining to phenotype, in this case the coarse-grained, or schema, fitness is time-dependent, with the "renormalized" fitness now being a highly non-trivial function of the original microscopic fitness, being defined as $f_\alpha = \mathcal{R}(x,\alpha)f_x = \sum_{x\in\alpha} f_x P_x(t)$. It may be shown [44], in the case of a flat fitness landscape, that an analog of equation (4) is form invariant without renormalizations under the coarse graining to an arbitrary schema; recombination and mutation acting

[3]We will use letters from the beginning of the Greek alphabet to represent schemata.

only on the partition associated with the definite schema loci. Hence, in this case X_η is of dimension N and $X_{\eta'}$ of dimension N_2. We will see momentarily that there exists a wider form invariance if one passes to a transformed basis for the dynamics. Also, as there exist 3^N possible schemata a full schema basis is over-complete and non-orthonormal. However, the space of schemata is not the natural one for recombination, as we shall see.

3.2. Embeddings. In the case of embeddings one passes from a lower to a higher dimensional configuration space. An example would be that of passing from a representation where types are represented by variable-length strings, of up to maximum size N_m with binary alleles, to a fixed length representation of size N_m, by including a third allele value that specifies that there was no corresponding bit in the variable length case. The original configuration space would be of dimension $2(2^{N_m}-1)$, however, due to the addition of a third allele the dimension of the embedding space is 3^{N_m}. Of course, for these more general transformations development of the operators and the corresponding theory necessary to maintain syntactic correctness of the offspring is a totally open issue. In this case, one might be better off using the theory for variable length structures already developed in Genetic Programming.

3.3. Coordinate Transformations. Transformations of the coordinate system are a standard tool in the mathematical analysis of physical models. The goal is to choose a set of variables that is well-adapted to the internal structure of the model and hence simplifies the structure of the equations that have to be analyzed. For our purposes we will only need *linear* transformations which can be described in terms of matrices. We restrict our discussion of explicit examples to the case of binary strings, which we will denote by $x = (x_1, \dots, x_N)$ with $x_i = \pm 1$. The alternative encoding as a binary string $\tilde{x}$ is obtained by setting $\tilde{x}_i = 1$ iff $x_i = 1$ and $\tilde{x}_i = 0$ iff $x_i = -1$.

Let us start with the standard basis, which we have been using up to now, $\boldsymbol{\delta} = \{\delta_z | z \in X\}$. It consists of the δ-functions

$$\delta_z(x) = \begin{cases} 1 & \text{if} \quad x = z \\ 0 & \text{if} \quad x \neq z \end{cases} \tag{10}$$

As an example, the quasispecies model takes its usual form

$$\dot{P}_x(t+1) = \frac{1}{\bar{f}} \sum_y \mathbf{M}_{xy} f_y P_y(t) \tag{11}$$

in this coordinate system. The three alternative bases we will consider: the Walsh basis, the Taylor basis and the Building Block Basis are all related to the standard basis via linear transformations. In Figure 3 we show the explicit transformation matrices for the case of three loci.

Walsh (Fourier) Basis. Probably the most important alternative basis is the *Walsh basis* $\boldsymbol{\psi}$, consisting of *Walsh functions*

$$\psi_I(x) = 1/\sqrt{|X|} \prod_{j \in I} x_j \tag{12}$$

where I is a subset of $\{1,\ldots,N\}$. The *order* of a Walsh function is $|I|$, the number of loci that are multiplied. The Walsh functions are normed and orthogonal. The Walsh-transform, $\hat{f}$, of a function f is defined implicitly by

$$f(x) = \sum_I \hat{f}(I)\psi_I(x) \tag{13}$$

Multiplying with $\psi_K(x)$ and summing over all $x \in X$ we obtain

$$\sum_{x\in X} f(x)\psi_K(x) = \sum_I \hat{f}(I) \sum_{x\in X} \psi_I(x)\psi_K(x) = \sum_I \hat{f}(I)\delta_{IK} = \hat{f}(K) \tag{14}$$

In matrix form we obtain the compact expression $\hat{f} = \mathbf{\Psi} f$, where the matrix $\mathbf{\Psi}$ has the Walsh functions ψ_K as its rows. Conversely, we have $f = \mathbf{\Psi}^+ \hat{f}$. The most important property of the Walsh functions, at least for our purposes, is that they are eigenfunctions of the mutation operator (3)

$$\mathbf{M}\psi_I = (1 - 2|I|/N)\psi_I \tag{15}$$

The mutation operator is therefore diagonal in the Walsh basis, see [45, 46]. Eqn. (4) reads in these coordinates

$$\hat{P}_I(t+1) = \sum_x (\mathbf{\Psi M})_{Ix} P^c_x = \left(1 - 2\frac{|I|}{N}\right) \hat{P}^c_I \tag{16}$$

We can see that the Walsh basis will be particularly useful if the transformed selection-recombination term $\hat{P}^c_I$ also has a simple form. The Walsh transformation is a special case of the Fourier transformation on groups. A discussion of this connection in the context of landscapes can be found in [47]. Approximations to the recombination operator that become diagonal in the Fourier basis are described in [35, 48].

Taylor Series. While the standard basis $\boldsymbol{\delta}$ and the Walsh basis $\boldsymbol{\psi}$ are orthonormal, this is not necessarily the case in general. In [46], for instance, the *Taylor series* of a landscape on the Boolean hypercube is introduced in terms of the polynomials

$$\tau_I(x) = \prod_{i\in I} \tilde{x}_i \tag{17}$$

such that $f(x) = \sum_I \tilde{f}(I)\tau_I(x)$. Let us write $x \Subset I$ if $\tilde{x}_i = x_i = 1$ for all $i \in I$. We define the matrix $\mathbf{\Upsilon}$ by $\mathbf{\Upsilon}_{Ix} = 1$ if $x \Subset I$ and $\mathbf{\Upsilon}_{Ix} = 0$ otherwise, i.e., $\mathbf{\Upsilon}_{Ix} = \tau_I(x)$. Thus we can write the Taylor series expansion in the form $f = \mathbf{\Upsilon}\tilde{f}$, i.e., $\tilde{f} = \mathbf{\Upsilon}^{-1} f$. The multi-linear fitness model described in [49] is based on a similar idea.

The matrix $\mathbf{\Upsilon}$ is invertible [46] but it is neither normalized nor orthogonal. In other words, the basis functions $\tau_I(x)$ do not form an orthonormal basis. In fact, we have

$$\sum_{x\in X} \tau_I(x)\tau_k(x) = 2^{N-|I\cup K|} \tag{18}$$

since $\tau_I(x)\tau_k(x) = 1$ whenever $x_i = 1$ for all $i \in I \cup K$, and 0 otherwise.

The Building Block Basis. Coordinate transformations such as the Walsh transform are general. Here, however, we wish to consider a particular coordinate transformation that arises as an almost inevitable consequence of the action of recombination. We have already seen that representing recombination in terms of the fundamental, microscopic types is very inefficient due to the sparsity of the interaction matrix. This is an indication that individual sequences are not the natural effective degrees of freedom for recombination. So what are? To form the sequence `111` with a recombination mask `100` one can join strings `111`, `110`, `101`, and `100` with either `111` or `011`. In other words, for the first parent the second and third bit values are unimportant and for the second the first bit value is unimportant. Thus, it is natural to coarse grain over those sequences that give rise to the desired target for a given mask.

If one picks arbitrarily a vertex in X, associated with a type x, one may perform a linear coordinate transformation $\Lambda : X \to \tilde{X}$ to a basis consisting of all schemata that contain x. For instance, for two bits $X = \{\texttt{11}, \texttt{10}, \texttt{01}, \texttt{00}\}$, while $\tilde{X} = \{\texttt{11}, \texttt{1}*, *\texttt{1}, **\}$. The invertible matrix $\mathbf{\Lambda}$ is such that $\mathbf{\Lambda}_{\alpha x} = 1 \iff x \in \alpha$.

We denote the associated coordinate basis the Building Block basis (BBB)[4] in that one may think of the elements of this basis as the BBs[5] that are joined together by recombination to form the type x.[6] The BBB is complete but clearly not orthonormal. Note that the vertex x by construction is a fixed point of this transformation. Apart from the vertex x, the points in $\tilde{X}$, being schemata, correspond to higher dimensional objects in X. For instance, `1`$*$ and $*$`1` are one-planes in X while $**$ is the whole space. In the BBB one may transform (5) to find

$$\tilde{P}^c_\alpha(t+1) = (1-p_c)\tilde{P}'_\alpha(t) + \sum_{m=1}^{2^N} \sum_{\beta,\gamma} \tilde{\mathbf{R}}_{xyz}(m) \tilde{P}'_\beta(t) \tilde{P}'_\gamma(t) \tag{19}$$

where $\tilde{\mathbf{R}}_{\alpha\beta\gamma}(m) = \mathbf{\Lambda}_{\alpha x} \mathbf{R}_{xyz} \mathbf{\Lambda}^{-1}_{\beta j} \mathbf{\Lambda}^{-1}_{\gamma k}$.

One may ask what is the advantage of this new representation? In the original microscopic basis the properties and symmetries of $\mathbf{R}_{xyz}(m)$ are very hidden. However, this is not the case for $\tilde{\mathbf{R}}_{\alpha\beta\gamma}(m)$ which has the property that for a given mask only interactions between BBs that construct the target schema are non-zero, i.e., $\tilde{\mathbf{R}}_{\alpha\beta\gamma}(m) = 0$, unless γ corresponds to a schema which is the complement of β with respect to α. Furthermore, $\tilde{\mathbf{R}}_{\alpha\beta\gamma}(m) = 0$ unless β is equivalent to m, whereby equivalent we mean that for any `1` in the mask we have a `1` at the corresponding locus in β and for any 0 we have a $*$. These two important properties mean that the two summations over β and γ in (19) both disappear and we are left with only the sum over masks with an "interaction" constant $p_c(m)$ which depends only on the mask. For example, for two bits, if we choose as vertex `11`, then `11` may interact only with $**$, while `1`$*$ may interact only with $*$`1`.

[4] This basis is implicit in the work [50, 13, 51] but has only been considered in more detail recently [52].

[5] Note that these BBs are not the same as their well known counterparts in the theory of Genetic Algorithms [53] being dynamic, not static, objects. Neither are they necessarily short or fit.

[6] Given the arbitrariness of the choice of vertex there are in fact 2^N equivalent BBBs each transformable to any other by a permutation.

$\mathbf{\Lambda} =$

	111	110	101	011	100	010	001	000
111	1	0	0	0	0	0	0	0
11*	1	1	0	0	0	0	0	0
1*1	1	0	1	0	0	0	0	0
*11	1	0	0	1	0	0	0	0
1**	1	1	1	0	1	0	0	0
1	1	1	0	1	0	1	0	0
**1	1	0	1	1	0	0	1	0
***	1	1	1	1	1	1	1	1

$\mathbf{\Upsilon} =$

	ooo	loo	olo	ool	llo	lol	oll	lll
111	1	1	1	1	1	1	1	1
110	1	1	1	0	1	0	0	0
101	1	1	0	1	0	1	0	0
011	1	0	1	1	0	0	1	0
100	1	1	0	0	0	0	0	0
010	1	0	1	0	0	0	0	0
001	1	0	0	1	0	0	0	0
000	1	0	0	0	0	0	0	0

$\mathbf{\Psi} = \frac{1}{\sqrt{8}}$

	111	110	101	011	100	010	001	000
111	−1	+1	+1	+1	−1	−1	−1	+1
110	+1	−1	−1	−1	−1	−1	+1	+1
101	+1	−1	+1	−1	−1	+1	−1	+1
011	+1	−1	−1	+1	+1	−1	−1	+1
100	−1	−1	−1	+1	−1	+1	+1	+1
010	−1	−1	+1	−1	+1	−1	+1	+1
001	−1	+1	−1	−1	+1	+1	−1	+1
000	+1	+1	+1	+1	+1	+1	+1	+1

Figure 3. Examples of transformation matrices for $N = 3$.

In X this has the interesting interpretation that for a target schema α of dimensionality $(N-d)$ only geometric objects "dual" in the d-dimensional subspace of X that corresponds to α may interact. In other words, a k-dimensional object recombines only with a $(N-d-k)$-dimensional object. Additionally, a $(N-d)$-dimensional object may only be formed by the interaction of higher dimensional objects. In this sense interaction is via the geometric intersection of higher dimensional objects. For example, the point 11 can be formed by the intersection of the two lines 1* and *1. Similarly, 1111 can be formed via intersection of the three-plane 1*** with the line *111 or via the intersection of the two two-planes 11** and **11.

As mentioned, one of the primary advantages of the BBB representation is that the sums over j and k disappear thus obtaining

$$P_x^c(t) = (1-p_c)P_x'(t) + \sum_{m=1}^{2^N} p_c(m) P'_{x_m}(t) P'_{x_{\bar{m}}}(t) \tag{20}$$

where $P'_{x_m}(t)$ is the probability to select the BB x_m (note that the mask uniquely specifies which element, x_m, of the BBB to choose) and $P'_{x_{\bar{m}}}(t)$ is the probability to select the BB $x_{\bar{m}}$, which is uniquely specified as the complement of x_m in x. Both x_m and $x_{\bar{m}}$ are elements of the BBB associated with x. The above equation clearly shows that recombination is most naturally considered in terms of the BBB. In the standard basis there were of the order of 2^{2N} elements of $\mathbf{R}_{xyz}$ to be taken into account for a fixed x. In the BBB there is only one term. Of course, we must remember that the coarse grained averages of x_m and $x_{\bar{m}}$ contain 2^N terms, still, the reduction in complication is enormous. Thus, we see that recombination as an operator naturally introduces the idea of a coarse graining, the natural effective degrees of freedom associated with recombination being the BBs we have defined. This is an important point as it shows that evolution is acting in the presence of recombination most naturally at the level of populations, the BBs representing populations with a certain degree of "kinship" to the target type.

Inserting (20) in (4) we can try to solve for the dynamics. However, in order to do that we must know the time dependence of x_m and $x_{\bar{m}}$. Although the number of BB basis elements is 2^N we may generalize and consider the evolution of an arbitrary schema, α. To do this we need to sum with $\sum_{x\supset\alpha}$ on both sides of the equation (4). This can simply be done to obtain [50, 13, 51] again the form (4), where this time the index α runs only over the 2^{N_2} elements of the schema partition and where again $\mathbf{M}_{\alpha\beta} = p^{d^H(\alpha,\beta)}(1-p)^{N-d^H(\alpha,\beta)}$. In this case however $d^H(\alpha,\beta)$ is the Hamming distance between the two schemata. For instance, for types with three loci the schemata partition associated with the first and third loci is $\{1*1, 1*0, 0*1, 0*0\}$. In this case $d^H(1,2) = 1$ and $d^H(1,4) = 2$. $P^c_\alpha(t) = \sum_{x\supset\alpha} P^c_x(t)$ is the probability of finding the schema α after selection and recombination. Note the form invariance of the equation after coarse graining. To complete the transformation to schema dynamics we need the schema analog of (20). This also can be obtained by acting with $\sum_{x\supset\alpha}$ on both sides of the equation. One obtains

$$P^c_\alpha(t) = (1 - p_c N_\alpha) P'_\alpha(t) + \sum_{m\in\mathcal{M}_r} p_c(m) P'_{\alpha_m}(t) P'_{\alpha_{\bar{m}}}(t) \tag{21}$$

where α_m represents the part of the schema α inherited from the first parent and $\alpha_{\bar{m}}$ that part inherited from the second. $N(\alpha)$ is the number of recombination masks that affect α, relative to the total number of masks with $p_c(m) \neq 0$, the set of such masks being denoted by $\mathcal{M}_r$. Obviously, these quantities depend on the type of recombination implemented and on properties of the schema such as defining length. Note that the BBB naturally coarse grains here to the BBB appropriate for the schema α as opposed to the sequence x.

Thus, we see that the evolution equation for schemata is form invariant, there being only a simple multiplicative renormalization of the recombination probability p_c. This form invariance, first shown in [13], demonstrates that BB schemata in general are a preferred set of coarse grained variables and, more particularly, the BBB is a preferred basis in the presence of recombination. It has also been shown [44] that schemata, more generally, are the only coarse graining that leads to invariance in the presence of mutation and recombination.

Considering again the structure of (20) and (21) we see that variables associated with a certain degree of coarse graining are related to BB "precursors" at an earlier time, which in their turn ... etc. This hierarchical structure terminates at order-one BBs as these are unaffected by crossover. Thus, for example, the level one BB combinations of 111, i.e., BBs that lead directly upon recombination to 111 are: $\mathtt{11}*:**\mathtt{1}$, $\mathtt{1}*\mathtt{1}:*\mathtt{1}*$ and $\mathtt{1}**:*\mathtt{11}$. The level two BBs are $\mathtt{1}**$, $*\mathtt{1}*$ and $**\mathtt{1}$. Thus, a typical construction process is that BBs $\mathtt{1}**$ and $*\mathtt{1}*$ recombine at $t=t_1$ to form the BB $\mathtt{11}*$ which at some later time t_2 recombines with the BB $**\mathtt{1}$ to form the sequence 111.

In the case of recombination note also that the coarse graining operator associated with the BBs satisfies

$$\mathcal{R}(\eta,\eta') = \mathcal{R}(\eta^m,\eta'^m)\mathcal{R}(\eta^{\bar{m}},\eta'^{\bar{m}}) \tag{22}$$

where $\mathcal{R}(\eta^m,\eta'^m)$ represents the action of the coarse graining on the BB § while $\mathcal{R}(\eta^{\bar{m}},\eta'^{\bar{m}})$ represents the action on the BB $\bar{m}$.

4. Fitness Landscapes

The concept of a *fitness landscape* originated in the 1930s in theoretical biology [54, 55] as a means of visualizing evolutionary adaptation. A fitness landscape is a kind of "potential function" on which a population moves, Fig. 4. Implicit in this notion is both a *fitness function* f that assigns a fitness value to every possible type (or organism), and the arrangement of the set of types in some kind of abstract space that provides some notion of accessibility or reachability.

A significant part of modern population genetics is still based on these ideas. The basic ingredients are [3] as stated in the previous section: A set X of "types", a fitness function evaluating the types and a set of genetic operators that induce a dynamics on X. (Strictly speaking, this implies that a measure must be defined on the set of types. Since we will mostly be concerned with the case of a finite set X with the uniform measure we shall not be concerned with this complication.) A very general framework for multi-locus models is described in [49].

The mathematical theory of fitness landscapes starts with specifying the three basic ingredients of the *landscape* $(X,\mathcal{X},f)$:

(1) A set X of configurations,
(2) a notion $\mathcal{X}$ of neighborhood, nearness, distance, or accessibility on X, and
(3) a fitness function $f:X\to\mathbb{R}$.

The set X together with the "structure" $\mathcal{X}$ forms the configuration space. The definition of $\mathcal{X}$ is purposefully left vague at this point and will be made precise in the following section.

The intuition behind a fitness landscape is that fitness can be thought of as a "height function" orthogonal to X and hence higher fitness types are located at "peaks" and low fitness types in "valleys". The dynamics associated with selection at the population level can then be interpreted as "hill climbing" on this topography. This is a very intuitive and appealing paradigm allowing for a "geometrical" understanding

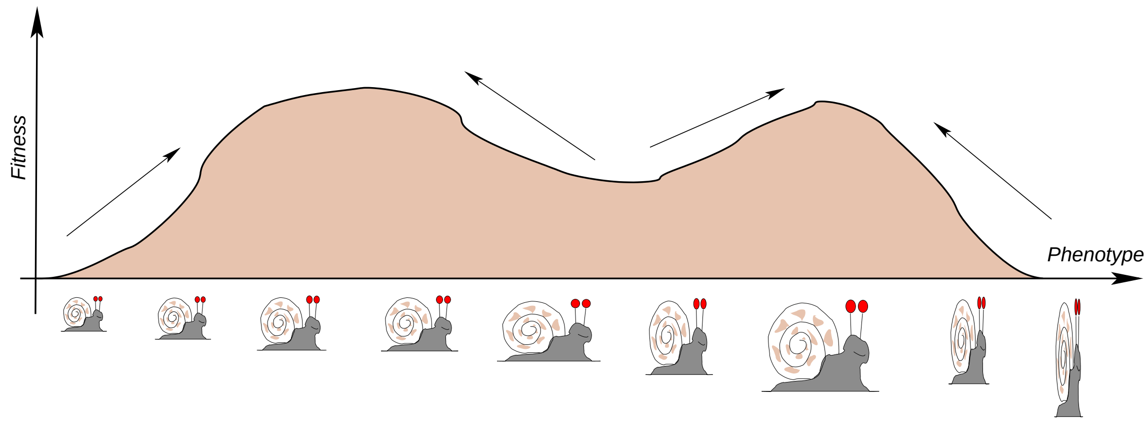

Figure 4. The fitness landscape paradigm.

of adaptation under selection. In fact, the average population fitness is a Lyapunov function in this case. As we have stressed however, generally it is only in the absence of other genetic operators that the fitness as determined by the landscape measures the full reproductive success of a type.

Mathematical constructions analogous to a fitness landscape arise naturally in many other areas of scientific study. For instance, in the physics of disordered systems where spin-glasses, for example, can be cast into the same form [56, 57]. Each spin configuration is assigned an energy by virtue of the Hamiltonian that specifies the model; the dynamic properties invoke a collection of transitions between configurations. In biophysics energy landscapes govern the folding of biopolymers, including proteins [58, 59, 60] and nucleic acids [28, 61]. Conceptually, there is a close connection with the *potential energy surfaces* (PES) of theoretical chemistry [62, 63]: As a consequence of the validity of the Born-Oppenheimer approximation, the PES provides the potential energy $U(\vec{R})$ of a molecule with n atoms as a function of its nuclear geometry $\vec{R} \in \mathbb{R}^{3n}$. Electoral Landscapes are used to explain party platform behavior in spatial voting models [64, 65]. In combinatorial optimization the fitness function is usually referred to as the *cost function*, and a *move-set* allows interchange of the elements of the *search space* [66]. The application of evolutionary models to combinatorial optimization problems has lead to the design *evolutionary algorithms* such as Genetic Algorithms, Evolution Strategies, and Genetic Programming [67, 68, 69, 70, 71].

Usually the viability fitness is taken to be static, though this is not a restriction. The corresponding fitness landscapes are thus static and hence have been principally studied by focusing on geometric properties such as smoothness, ruggedness, and neutrality. This static point of view has been the main focus of two recent reviews [8, 72], which are complementary to material presented here. Even in this much simplified setting however, there is no complete classification, or taxonomy, of landscapes. On the other hand, interesting and important classes of landscapes, such as the Nk models [70], elementary landscapes [73] (which include the p-spin models), and landscapes with tunable neutrality [74] (including short range spin glasses) are known and have been extensively studied. Such a classification, if it exists at all, analogous to the classification of phase transitions, catastrophes or critical points, would inevitably be based on statistical properties of the landscape and would be of great importance.

An important class of landscapes are those for which the coordinate transformations discussed in the previous section act as projectors. The constant function is an eigenfunction of the mutation operator (3). The projection of any landscape onto this eigenspace is the average fitness $\bar{f} = (1/|X|)\sum_x f_x$ of the landscape. A class of landscapes that is of particular interest are the so-called *elementary landscapes* for which $f - \bar{f}$ is an eigenfunction of $\mathbf{M}$. In this case we have in the Walsh basis $\hat{f}(I) = 0$ unless $I = \emptyset$ or $|I| = p$ for a fixed valued of the interaction order p, see [73, 75]. For a detailed account of this topic see [8]; we will return to this topic below in section 4.1.

4.1. Coarse Grained and Coordinate Transformed Fitness Landscapes. We have emphasized that there are many alternative representations to the microscopic one based on the types themselves. Given the utility of alternative representations it is important to consider how the corresponding transformed fitness landscapes look.

Coarse Grained Fitness Landscapes. Given a fitness landscape associated with variables η the coarse grained landscape associated with variables η', coarse grained relative to η, is given by

$$\mathcal{R}(\eta,\eta')f_\eta(t) = f_{\eta'}(t) = \sum_{\eta\in\eta'} f_\eta(t)P_\eta(t)/\sum_{\eta\in\eta'} P_\eta(t) \tag{23}$$

Notice the presence of the probability distribution $P_\eta(t)$ in the above. By introducing this probability distribution we are weighting the $f_\eta(t)$ by a factor that is intrinsically population dependent. The motivation for this is the following: consider the case of pure proportional selection, under a coarse graining

$$P_\eta(t+1) \longrightarrow \mathcal{R}(\eta,\eta')\frac{f_\eta(t)}{\bar{f}(t)}P_\eta(t) = \frac{\sum_{\eta\in\eta'} f_\eta(t)P_\eta(t)}{\bar{f}(t)} = \frac{f_{\eta'}(t)}{\bar{f}(t)}P_{\eta'}(t) \tag{24}$$

The evolution of η' can be understood either in terms of the microscopic η variables via the factor $\sum_{\eta\in\eta'} f_\eta(t)P_\eta(t)$ or in terms of the coarse grained fitness $f_{\eta'}(t)$. The latter preserves the natural interpretation of selection in that η'-types increase/decrease when their fitness is greater/less than average.

Note that even if f_η is time independent $f_{\eta'}$ will depend on time via an implicit dependence on the population distribution. A static coarse grained fitness landscape can be defined by coarse graining with respect to a particular *fixed* probability distribution. The most natural one is probably that associated with a random population. In this case $f_\eta \longrightarrow \sum_{\eta\in\eta'} f_\eta/\sum_{\eta\in\eta'} 1$. Hence, if f_η is time independent so is $f_{\eta'}$. Whether a statically or dynamically generated coarse grained fitness landscape is better depends on one's point of view. A static coarse graining has the advantage that the landscape, which is the underlying substrate on which the population dynamics takes place, is constant. However, shares with the standard fitness landscape the deficiencies that we will discuss in section 6. In cases where the coarse graining is associated with symmetries, such as is the case for the genotype-phenotype map, then the static and dynamic coarse grained landscapes are the same.

As emphasized previously, the most obvious example of a coarse grained landscape is that associated with the phenotype. Given that all genotypes associated with the same phenotype have the same fitness the phenotypic landscape besides being of lower dimension will also be more rugged, the genotypic landscape generically containing neutral networks, a topic we will discuss in more depth in section 5. A simple example of such is the Quasispecies model for three loci. The genotypic landscape is three dimensional and the phenotypic, one dimensional. The non-master sequence genotypes form a neutral network. Although a degenerate genotype-phenotype map is necessary to guarantee a neutral network it is not sufficient, as can be simply seen in the case of a landscape where fitness is additive and proportional to the number of alleles of a given type.

Another example of a coarse grained landscape is that associated with schemata. Once again, we may simply illustrate this with the Quasispecies model. If the master sequence is chosen to be `111` with fitness $(1+\delta)$ while the other genotypes have fitness 1 then concentrating on only the first two loci the static coarse grained schema fitnesses are: $f_{11*} = (1+\delta/2)$, $f_{10*} = f_{01*} = f_{00*} = 1$. In this case there is a "master schema". This coarse grained landscape would lead one to believe that the population should ascend the coarse grained fitness peak at $\mathtt{11}*$. The dynamic picture would take as coarse grained fitnesses: $f_{11*}(t) = 1 + (\delta P_{111}(t)/(P_{111}(t) + P_{110}(t)))$, $f_{10*} = f_{01*} = f_{00*} = 1$. If one were to observe population flows in this model one would see that the "master schema" has a certain reproductive success. However, its observed success would be better described by the dynamical rather than the static coarse grained fitness, i.e. its reproductive success would generally be observed to be greater than that associated with the static schema fitness.

Walsh Transformed Fitness Landscapes. The Walsh-transform of a landscape highlights global features, in particular correlations. It is well known that the total variation of the landscape is $\mathrm{var}[f] = \sum_{I\neq\emptyset} \hat{f}(I)^2$. The quantities

$$B_p = \frac{1}{\mathrm{var}[f]} \sum_{I:\,|I|=p} \hat{f}(I)^2 \tag{25}$$

measure the relative importance of the projections of the fitness function f onto the eigenspaces of the mutation operator $\mathbf{M}$. Elementary landscapes are thus characterized by $B'_p = 1$ for a particular value p' of p and $B_p = 0$ otherwise. These landscapes can be shown to exhibit particular geometric features. For example, all local minima have a fitness that is smaller than $\bar{f}$, while all local maxima have a larger fitness [76]. Furthermore, a version of Courant's nodal domain theorem limits the number of large mountain ranges and valleys [77]. Examples include quadratic spin glasses and certain combinatorial optimization problems, such as weight partitioning, graph matching, the traveling salesman problem, and graph coloring.

The distribution of B_p-values, termed the amplitude spectrum e.g. in [78], characterizes the correlation structure of a landscape, see Fig. 5 for an example. In particular, the correlation length ϱ of a landscape (defined along a sequence of randomly placed

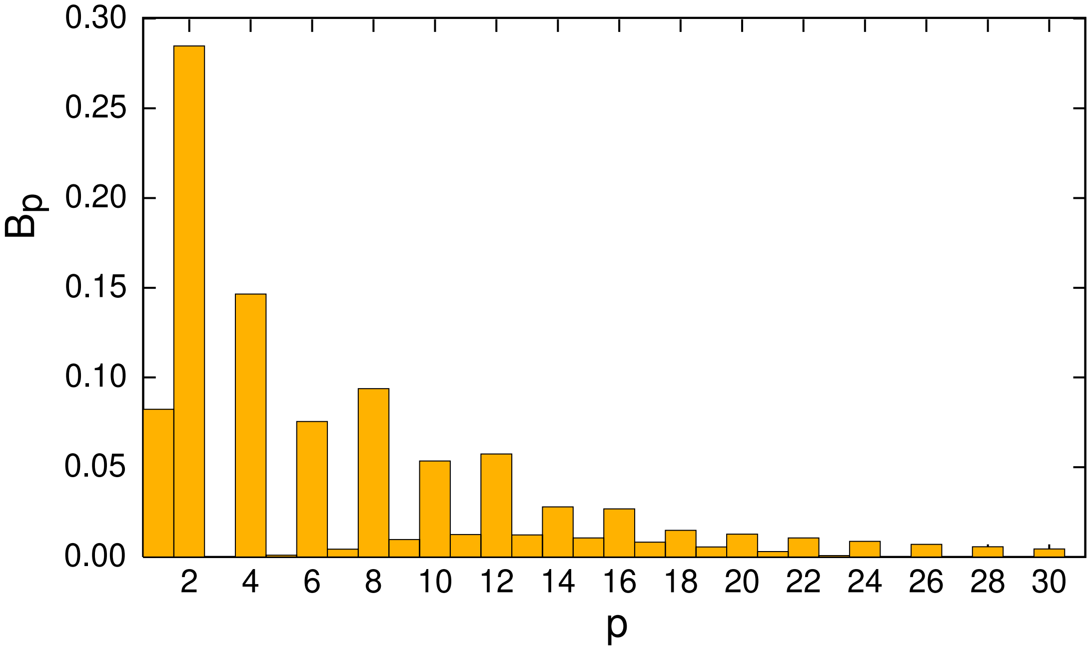

Figure 5. The estimated amplitude spectrum for a GC **what's GC** landscape with $N = 100$ under mutation [78]. The most striking feature of the amplitude spectrum of RNA landscapes is a strong difference between even and odd modes. This can easily be explained in terms of the physics underlying RNA folding: The major contribution of the folding energy comes from stacking of base pairs.

subsequent mutations) can be computed as

$$\varrho = \frac{N}{2} \sum_{p>0} \frac{B_p}{p} \tag{26}$$

It is interesting to note that many landscapes in physics and combinatorial optimization are defined in the Walsh basis. For instance, a quadratic spin glass Hamiltonian is usually written in the form $\mathcal{H}(\boldsymbol{\sigma}) = \sum_{i<j} J_{ij}\sigma_i\sigma_j$. One easily checks that $J_{ij} = \hat{\mathcal{H}}(I)/\sqrt{N}$ with $I = \{i, j\}$.

Landscapes in the BBB. Once again, one may take a static or dynamic point of view of the transformation. One uses the BBB transformation matrix Λ. As it is more natural to consider fitness averages rather than pure sums in the case of a static transformation we divide the matrix elements of Λ by $|\alpha| = 2^{N-N_2}$, the number of sequences that contribute to the BB α. For instance, for the three loci Quasispecies model the fitness landscape in the BBB is: $f_{111} = (1+\delta)$, $f_{11*} = f_{1*1} = f_{*11} = (1+\delta/2)$, $f_{1**} = f_{*1*} = f_{**1} = (1+\delta/4)$ and $f_{***} = \delta/8$. Interestingly, in this basis we see there is no neutral network, there being a steady fitness gradient in the direction of the master sequence corresponding to a ranking of the fitnesses of the master sequence's BBs. In the dynamic picture one replaces the factors of $\delta/|\beta|$, where $|\beta|$ is the number of sequences that contribute to the BB β, with corresponding factors: $\delta P_{111}(t)/\sum_{x\subset\beta} P_x(t)$. Once again, in the dynamic picture the height of each state in the landscape is a function of time due to the dependence on the population state.

5. Geometric Characterization of Landscapes: Ruggedness, Neutrality, and Isotropy

Having discussed the basic elements of fitness landscapes and various relevant transformations on them we pass now to the topic of how to characterize landscapes geometrically. Geometric characterization of static landscapes, which form a possible basis for a classification of landscapes, were the topic of recent reviews by one of the authors [8, 79, 72]; we therefore touch this topic only very briefly. It turns out that the ruggedness of landscape is intimately related to *epistasis* [80, 48], where epistasis is defined as the non-additive component of the fitness function. In the Walsh basis additive and epistatic interactions are separated: The values of $\hat{f}(I)$ with $|I| > 0$ describe the epistatic interactions. A straight forward measure of the strength of epistasis is therefore

$$\varepsilon = \frac{1}{\mathrm{var}[f]} \sum_{I:|I|>1} \hat{f}(I)^2 \,. \tag{27}$$

The ruggedness of a landscape is most easily quantified by measuring the correlation of fitness values between "neighboring" positions [81, 46, 82, 73]. It turns out that many of the paradigmatic combinatorial optimization problems, including the Traveling Salesman Problem, Graph Coloring, Graph Biparitioning, Graph Matching, and Number Partitioning, belong to a single class of highly correlated landscapes. Probably the most typical feature of this landscape is the existence of an exponentially large number of local optima and a strong correlation between the fitness of a local optimum and the size of its basin of attraction [83, 84].

Biologically relevant fitness landscape apparently are not of this type, however. Their correlation seem to be less pronounced [25], while on the other hand neutrality plays a crucial role [26, 85]. Neutrality, i.e., the existence of neighboring configurations with the same fitness, appears to be just a way of achieving a "smooth" landscape, though as we have seen in the case of the BBB neutrality depends intimately on how one defines a neighbor. Therefore it comes as a surprise, that ruggedness and neutrality are in fact independently tunable parameters [74].

In the case of RNA molecules at least, it is possible to explicitly compute the genotype-phenotype map and to investigate its global properties explicitly [26]:

(i) **Redundancy.** For sequence spaces of chain lengths $n \geq 10$ there are orders of magnitude more sequences than structures and hence, the map is many-to-one. This fact can be proved combinatorially. The GP-map of RNA is therefore highly redundant.

(ii) **Connected Neutral Networks.** Neutral networks of common structures are connected unless specific and readily recognizable special features of RNA structures require specific non-random distributions in sequence space. A random graph approach [86, 87] also predicts that the neutral networks are connected provided the average neutrality, $\bar{\lambda}$, exceeds a threshold that only depends on the nucleic acid alphabet.

(iv) **Shape Space Covering**. The neutral network $\mathcal{G}(\psi)$ is embedded in a compatible set $C(\psi)$ which includes all sequences that can form the structure ψ

at least as a suboptimal conformation. Note that $C(\psi)$ is determined only by the logic of base pairing in RNA, where only the six combinations AU, UA, GC, CG, GU, and UG can pair, Fig. 2. Computational studies, e.g. [26, 88, 89], showed that $\mathcal{G}(\psi)$ is well described in many respects by a random induced subgraph of the compatible part of sequence space $\mathcal{Q}[C(\psi)]$. More explicitly, the random graph model [86] assumes that the structure of $\mathcal{G}(\psi)$ is determined only by the fraction $\bar{\lambda}$ of neutral mutations when restricting the acceptable mutations to the sequences that are compatible with the structure ψ. On the other hand, one can prove rigorously that the compatible sets of any two phenotypes $C(\psi)$ and $C(\varphi)$ have a non-empty intersection [86]. As a result it is possible to define a spherical ball with a diameter $R_{\text{cov}} \ll N$ which contains on the average, for every common structure, at least one sequence that folds into it.

These predictions are in agreement with experimental findings. For instance, Schultes & Bartel [90] describe an RNA that simultaneously carries the properties of two different RNA folds and which is connected by a neutral path with the two sequences from which it was originally designed. Computational studies based on inverse folding by means of knowledge-based potentials [91] strongly indicate that all four features hold for protein spaces as well [30, 31]. Again, there is experimental evidence for the correctness of this picture, see e.g. [92, 93]. Empirical evidence for a large degree of *functional* neutrality in protein space is described e.g. by Wain-Hobson and co-workers [94].

Neutrality has a number of important impacts on the dynamical behavior of a population of replicating individuals. Most importantly, there is a diffusive motion of the population's "center of gravity" through sequence space [95]. The diffusion constant is related to population size n, per digit mutation rate p and the fraction λ of neutral neighbors [96]. A constant "rate of innovation" is reported in [97] for the landscapes in which all neutral networks come close together, as in the case of RNA. A population therefore evolves by producing mutants in the boundary of the neutral network of the currently dominating species. Short, intermittent episodes of selection interrupt the diffusive behavior when a fitter mutant invades the population and spreads through it.

A genotype-phenotype map that leads to neutral networks is thus an evolutionary advantage because it increases evolvability. In a non-neutral, spin-glass like landscape with many local optima the population will very quickly get stuck in a locally optimal configuration from which it can escape only via an advantageous mutant. In contrast to a population that freely diffuses on a neutral network, however, the exploration will decrease to almost 0 for a stuck population because it gets harder with time to produce new mutants that are further and further away from the trap.

A third aspect that is presumably independent of ruggedness and neutrality is the degree to which a landscape is isotropic or anisotropic, i.e., to what extent different local subsets are distinguishable. Surprisingly, little progress has been made in this direction, see however [98, 99].

There have been several attempts to identify properties of fitness functions that relate to the action of recombination operators. Probably the most import one is the concept of a *deceptive function* [100, 101, 102], see also [103, 79]. A landscape is deceptive (for a Genetic Algorithm) if the BBs for an optimal schema are less fit than the corresponding BBs for a non-optimal schema on the same partition. For example, if the optimum is `11` and the BB `1`$*$ is less fit than the BB `0`$*$. Deception has inevitably been portrayed in the static picture. However, its effects really need to be considered in the dynamical picture as whether or not it prevents the algorithm from finding the optimum depends sensitively on the population state. It is interesting to mention, however, that all fitness functions of practical interest are either closely related to an additive landscape or deceptive. A more fine-grained classification, which should reflect the vast differences in difficulty observed in practice, would be highly desirable.

6. Limitations of the Fitness Landscape Paradigm

Despite its wide range of applicability in different fields, the usefulness of the standard picture of fitness landscapes in understanding either biological evolution or evolutionary computation is limited to certain, precisely defined, situations. Whether or not the dynamics of equ.(1) is consistent with or determined by a fitness landscape depends on the particularities of the selection term. Setting

$$\mathfrak{S}(P, \mathsf{w}) = P \circ F(P, \mathsf{f})\,, \tag{28}$$

i.e., in component-wise notation, $\mathfrak{S}_x(P, \mathsf{w}) = F_x(P, \mathsf{f})P_x$, we have selection proportional to a *growth-rate function* F_x for each type x [104]. We suggest that the canonical *fitness landscapes* paradigm is only valid when

$$F_x(P, \mathsf{w}) \approx f_x / \sum_x f_x P_x \tag{29}$$

where f_x is a constant that is characteristic of the type x. Otherwise the fitness (growth rate) F_x of type x depends on, and changes with, the frequencies P_y of all other types y. Models of such co-evolution are sometimes viewed as "coupled dancing landscapes" [70] where a species x changes the landscape of species y, and y changes the landscape of x, at the same time scale at which both species adapt to their respective landscapes. We argue that in this case the standard landscape metaphor is not useful anymore and prefer to limit the use of the term landscape to situations in which: (1) fitness is characteristic of a type and (2) constant in time, at least approximately. More stringently, one might also require that (3) fitness is approximately a *Lyapunov function* for the dynamics of equ.(1), i.e., that fitness increases along evolutionary trajectories.

As we shall see however, effective fitness, is perfectly well-defined in the absence of a landscape in the above sense. The corresponding concept of an *effective fitness landscape* has intuitive value by restoring a notion of hill-climbing even in cases where the requirements for having a (static) fitness landscape are violated.

The requirement that fitness be a Lyapunov function of equation (1) is very restrictive. To see this we consider again Eigen's quasispecies model [42] described by equation (11). In the absence of mutation $\mathbf{M}_{xy}$ is the identity matrix. Now, consider a population consisting exclusively of individuals that maximize f_x. In the presence of mutation the average fitness of the population will decrease due to the introduction of disadvantageous mutants. Hence, the average population fitness $\overline{f}$ *cannot* be a Lyapunov function. The model however does have a Lyapunov function, namely the projection

$$L(P) = \sum_x P^*_x P_x \tag{30}$$

onto the stationary distribution P^* defined by the linear eigenvalue equation

$$\sum_y \mathbf{M}_{xy} f_y P^*_y = \overline{f} P^*_x \,, \tag{31}$$

see e.g. [43]. On the other hand, $\bar{f}$ is "sort of" a Lyapunov function for small mutation rates (off-diagonal elements of the matrix $\mathbf{M}$ are small) because in this case P^* is concentrated around the master sequence x_{ms} and hence $L(P) \approx P_{x_{\text{ms}}}$. The average population fitness, on the other hand, also tends to increase with the fraction of maximally fit individuals $P_{x_{\text{ms}}}$.

In the canonical quasispecies model with one fitness peak, i.e. one associated quasispecies, one may use as a measure of order in the population the relative concentration of the master sequence, $P_{\text{ms}}(t)$; and in particular in the long time limit, $P_{\text{ms}}(\infty)$, where a steady state population is reached — the quasi-species. In this case as is well known [42] $P_{\text{ms}}(\infty)$ monotonically decreases as a function of the mutation rate p until a critical rate, p_{cri}, associated with the error threshold, is reached beyond which $P_{\text{ms}}(\infty) = 1/2^N$. i.e. the master sequence is present in the same proportion as it would be in a purely random population.

It is important to realize that the fitness landscape is constant throughout this behavior. Thus, for $p = 0$, i.e. in the absence of mutations, the entire population climbs the fitness peak until $\bar{f}(t) \approx f_{\text{ms}}$, while for $p \geq p_{\text{cri}}$ the population is uniformly dispersed throughout the entire landscape. Clearly no intuition about these very different types of population flow endpoints can be gleaned from the structure of the landscape itself. In one limit selection dominates, in the other it has no effect. Whether or not the population will ascend a fitness peak due to selection depends crucially on the presence of another genetic operator — mutation. Thus, in the presence of mutation we see important regimes where the hill climbing paradigm fails completely. It is not difficult to construct many other relevant examples [15, 16] where the same phenomenon occurs.

The one-peak landscape also serves as a good example of induced symmetry breaking. In this landscape there are only two phenotypes, y_0 and y_{ms} where the phenotype y_0 corresponds to all genotypes save x_{ms}. Thus there is a (2^N-1)-dimensional degeneracy associated with the fitness landscape. Hence, in terms of fitness one would expect to find all genotypes but the master sequence in equal proportions. However, the steady state population – the quasi-species – exhibits a preference for those sub-optimal genotypes that are closer in Hamming distance to the optimum. Thus, the effect of the mutation operator is to lift the large landscape degeneracy. There still remains a residual symmetry associated with the fact that, at least starting with an homogeneous population, all genotypes of a given Hamming distance from the master sequence will have the same probability of reproductive success.

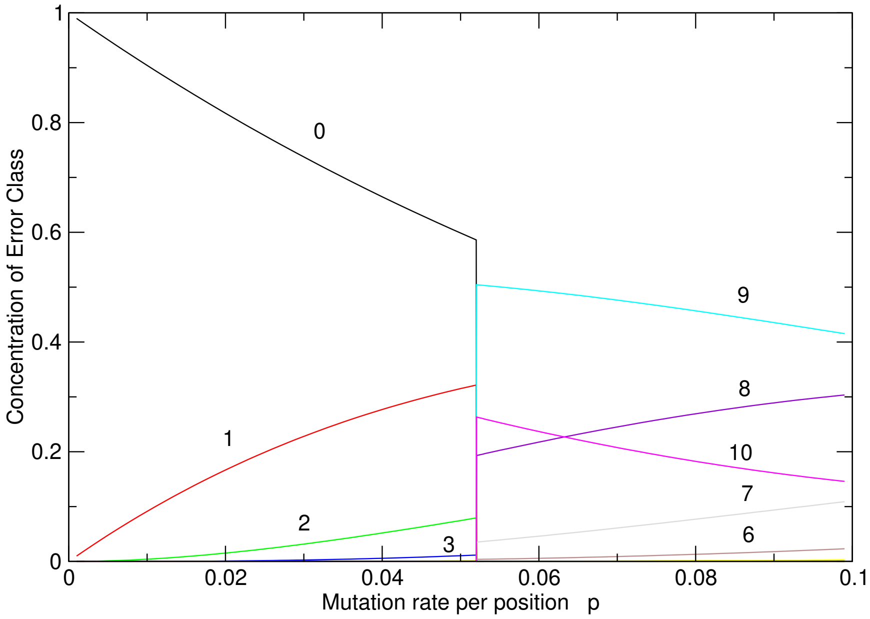

Figure 6. Stationary distribution P_x on a landscape with two peaks: $f(\vec{0}) = 1.0$, $f(\vec{1}) = 0.99$ and $f_x = 0.98$ for the one-error mutants of the all-1 string $\vec{1}$. We observe a sudden jump at $p \approx 0.52$ where the effective fitness of the peak at $\vec{1}$ exceeds the effective fitness of the absolutely fittest individual $\vec{0}$.

Similar phenomena occur in the case of a landscape with two fitness peaks, i.e. two possible quasispecies [105]. In this case it can be shown that the dominating type may switch with increasing mutation rate from the "individually fittest" x_{ms} to another type x^* with $f_{x^*} < f_{x_{\text{ms}}}$ that is surrounded by a broad plateau of relatively fit neighbors, see Fig. 6. i.e. that the quasispecies associated with the type x^* has more reproductive success than that associated with the master sequence. Once again this cannot be understood in the context of hill climbing on the fitness landscape.

A second example of this type is evolution towards robustness. In [106] it is shown that the stationary mutant distribution on a holey landscape [107] (consisting of a fit neutral network and a non-viable desert) depends on the local connectivity of the network. Once again, differences in reproductive success associated with different degrees of local connectivity cannot be understood in terms of hill climbing on the corresponding fitness landscape. Evidence for evolution towards robustness can be found e.g. in RNA viruses [108]. The massive effect of neutrality on evolutionary trajectories is described in some more detail in [109].

As a final example, we will now illustrate the the limitation of the standard landscape paradigm in the context of a more complex model [10], a variant of Kitano's neurogenetic model [110], that illustrates all of the above discussed weaknesses. In this type of model the chromosome encodes rules for cellular division and the phenotype is a 16-cell organism interpreted as a connectivity matrix for a feed-forward neural network (NN). The corresponding artificial ecological environment thus consists of a single species composed of NNs as individuals. A GA is then applied to the chromosomes present in the population at each epoch which induces a search of the connectivity matrix space determined by the structure of the NN. Environmental effects are included in the fitness function that measures the learning capacity of a particular individual.

Explicitly, a chromosome consists of eight blocks of four genes each one of which is a three bit structure. Both the blocks and the genes themselves are labeled from a to h. The reproduction process always begins with block a. Thus, the first four genes have a privileged role as they label the cells that are going to be reproduced in the second step of reproduction. As an example, consider the chromosome $baea.dcaa.defa.becd.aaea.aafh.haec.fgaa$. The two step reproduction process specified by this chromosome can be written as

$$a \longrightarrow \begin{pmatrix} b & a \\ e & a \end{pmatrix} \longrightarrow \begin{pmatrix} d & c & b & a \\ a & a & e & a \\ a & a & b & a \\ e & a & e & a \end{pmatrix} \leftrightarrow \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{32}$$

The genotype-phenotype map in this case is highly degenerate. For example, in the above we can change blocks c, e, f, g and h without changing the resulting phenotype. It is also a non-local function on the chromosomes since entries of block number one can target any one of the other blocks irrespective of their distance. To define a fitness function the learning speed of the NNs on a given test function was measured and a GA used to find the NN capable of learning this function with the smallest number of attempts.

Given the highly degenerate nature of the genotype-phenotype map one might expect to see an optimum phenotype emerge with a random distribution of the corresponding genotypes. However, this was not the case — certain genotypes were consistently preferred, thus indicating that the genotype-phenotype symmetry was broken. The reason for this is that although degenerate genotypes were equivalent in terms of reproductive selection the mixing genetic operators, mutation and recombination, broke the symmetry picking out the more robust or less brittle genotypes, i.e. those that were most likely to lead to other fit NNs.

Remarkably, the induced symmetry breaking in this context can be described in terms of the emergence of an "algorithmic language" [10]. The first word of the chromosome governs the grammar of the system by determining the syntax of the four-word combinations that make up the algorithm that generates the phenotype. The requirement that a phenotype learns well the test function implies that many four-word algorithms are "meaningless" in the sense that they lead to phenotypes of zero or very low fitness. Which words are chosen and their associated syntax is now crucially important, and one must confront the brittleness problem as mutation and recombination can convert a very fit phenotype into a very unfit one. It is precisely in this sense that a preferred language and grammar emerged with respect to which algorithms are written that lead to fit phenotypes. In particular robustness was an emergent property via a self-organization of the genotype-phenotype map, robust genotypes representing a given phenotype having more reproductive success than their less robust degenerate counterparts. As a simple illustration of this phenomenon consider the words "defunct" and "dead". Both have essentially the same meaning however, dead is much more robust to one letter mutations than defunct. For instance, dead can mutate to: read, bead, deed, dear etc. all of which have meaning. On the other hand it is difficult to find a one mutant neighbor of defunct that has meaning.

In the model at hand, it was found that the switching block, a, which can play a role as a word as well as a grammatical rule, rarely appeared as a word. This helped in making the system less susceptible to destructive mutations and hence more evolutionarily robust. Additionally, structures for this block that after mutation still formed fit phenotypes were seen to have more reproductive success. For instance, during the early part of the evolution first blocks of *daaa* and *baaa* were common. Both of these led to fit NNs with appropriate connectivity associated with the genes in the d and b blocks. However, these genotypes were susceptible to destructive mutations of the d or b genes in the first block. To counter this the system evolved a first block of type *daba*. Even though the resulting phenotype had no fitness advantage over its predecessors it led to greater reproductive success in that either of the d or b genes in the first block without creating an unfit phenotype.

7. Effective Fitness

In the previous section we discussed the advantages and disadvantages of the fitness landscape paradigm. Obviously, it would be of great benefit to have an alternative that includes the advantages and eliminates the disadvantages. In this section we introduce and discuss just such a paradigm: effective fitness.

There are several possibilities for such a quantity. For instance, based on the thermodynamic analogy [111] between an inhomogeneous two-dimensional spin model and a population evolving with respect to reproductive selection and mutation, one could define quite naturally the free energy per row as an effective measure of fitness. Here, however, we will use another definition, more directly related to the traditional idea of fitness in biology.

We define the effective fitness in the case of strings as [13, 50, 51, 14]

$$P_x(t+1) = \frac{f_x^{\mathrm{eff}}(t)}{\bar{f}(t)} P_x(t) \tag{33}$$

One may think of the effective fitness as representing the effect of all genetic operators in a single value for the viability or survival fitness. $f_x^{\mathrm{eff}}(t)$ is the fitness value at time t required to increase or decrease $P_x(t)$ by pure selection by the same amount as all the genetic operators combined in the context of a survival fitness f_x. If $f_x^{\mathrm{eff}}(t) > f_x(t)$ the effect of the genetic operators other than selection is to enhance the reproductive success of type x. Obviously, the converse is true when $f_x^{\mathrm{eff}}(t) < f_x(t)$. Here, we have defined effective fitness with proportional selection in mind as the selection operator. However, effective fitness is obviously applicable in a wider context: i.e. for a given selection operator, what fitness value in the selection only case leads to the same expected number of offspring of a given type as the case where several genetic operators are acting simultaneously.

Just as with the standard fitness one may represent mathematically the effective fitness very simply via a map: $f_X^{\text{eff}} : X \longrightarrow \mathbb{R}^+$, where as before we may work with phenotypes or genotypes and pass between the two representations using the genotype-phenotype map such that $f_X^{\text{eff}} = f_Y^{\text{eff}} \bullet \phi$ where X and Y are the genotype and phenotype configuration spaces respectively. Additionally, we may construct an effective fitness landscape by using f_X^{eff} as a height function above the space X. The chief difference here is that f_X^{eff} is almost inevitably time dependent hence the associated landscape is dynamical. It is also possible to effect the transformations discussed in section 3 on effective fitness thus leading to coarse grained or coordinate transformed effective fitness functions and landscapes.

Contrasting with the standard picture we may view population flows on a static landscape, but where the flows, as seen in the last section can be very counterintuitive, not related to a hill climbing intuition, or we may consider the flows on a dynamic substrate wherein the hill climbing paradigm can be seen to be restored. Although, working with a dynamic landscape might seem like a heavy price to pay, as we have emphasized, almost any coarse graining leads to a resultant fitness function which is time dependent anyway.

The exact functional form of the effective fitness obviously depends on the set of genetic operators involved. For the fairly general case of equation (4) in the case of proportional selection we have

$$f_x^{\text{eff}}(t) = \frac{\bar{f}(t)}{P_x(t)} \sum_y \mathbf{M}_{xy} P_y^c(t) \tag{34}$$

Note that it is time dependent and depends on the actual composition of the population. However, as we have emphasized, a static landscape in many ways is a rather artificial concept and only applicable in a certain approximate limit. In the limit $p \to 0$, $p_c \to 0$ (or in more generic circumstances when the strengths of operators other than reproductive selection $\to 0$) $f_x^{\text{eff}}(t) \to f_x$.

One is also naturally led to the idea of an effective selection coefficient,

$$s_x^{\text{eff}} = f_x^{\text{eff}}(t)/\bar{f}(t) - 1\,, \tag{35}$$

which directly measures effective selection pressure. If we think of s_x^{eff} as being approximately constant in the vicinity of time t_0, then $s_x^{\text{eff}}(t_0)$ gives us the exponential rate of increase or decrease of growth of the type x at time t_0. In the limit of a continuous time evolution the solution of the evolution equation (34) is

$$P_x(t) = P_x(0)\mathrm{e}^{\int_0^t s_{\text{eff}} dt'} \tag{36}$$

In the case of a finite population one must interpret differently the right hand side of the equation for effective fitness. One may say that effective fitness applies to a finite population if we interpret in equation (34) the sequence proportions on the RHS as being the actual proportions in the population rather than as expected values. In other words that the evolution equations relate the actual population proportions at time t to the expected value at time $t + 1$. However, this can lead to pathological behavior, such as $f_x^{\rm eff}(t) \to \infty$ when $P_x(t) = 0$ and $P_x(t+1) \neq 0$. Additionally, in this case the corresponding effective fitness landscape will change stochastically in time. This is not something intrinsic to the reproductive landscape, or the other operators themselves, but is a finite size effect arising from the limited stochastic sampling. For these reasons it is more appropriate to think of the effective landscape as being defined in the infinite population limit where its evolution is completely deterministic. In this case the effective fitness landscape will give information about the expected evolution of a population. Of course, the "experimental" effective fitness can be determined by measuring the proportion of individuals of a given type in two successive generations. This will not generically compare well with theoretical estimates for small population sizes due to stochastic sampling effects.

We do not wish to propagate the point of view that definition (34) is unique, nor even the best definition under all possible circumstances. For instance, in [13, 50] another natural definition was presented which follows from the split into those terms of the evolution equation that are linear in $P_x(t)$ and those source terms that are independent of it. For instance, in the case of selection and recombination we have

$$P_x(t+1) = \frac{f_x^{\prime\rm eff}(t)}{\bar{f}(t)} P_x(t) + j_x(t) \tag{37}$$

where

$$f_x^{\prime\rm eff}(t) = (1-p_c)\frac{f_x}{\bar{f}(t)} \quad \text{and} \quad j_x(t) = \sum_{m=1}^{2^N} p_c(m) P'_{x_m}(t) P'_{x_{\bar{m}}}(t) \tag{38}$$

The corresponding effective selection coefficient is

$$s'_{\rm eff} = (1-p_c)\,\frac{f_x}{\bar{f}(t)} - 1\,. \tag{39}$$

This effective fitness is intuitively more similar to that of [112, 113, 114, 115] in that it takes into account only the destructive effect of crossover. Nevertheless, it captures a rather natural division into terms that lead to a multiplicative renormalization of reproductive fitness (destruction terms) and those that lead to an additive renormalization (creation terms). In this decomposition formal results such as Geiringer's theorem follow in a most straightforward fashion [51][7].

The key element behind effective fitness, irrespective of its mathematical definition, is that population flows in the presence of operators other than pure reproductive selection are much more readily understood in terms of it. In fact, to go further, even in the case of pure reproductive selection, if one performs any sort of coarse graining and considers schemata rather than sequences, then population flow is more readily understood in terms of an effective landscape rather than the reproductive one. As an example, for the evolution of a particular order 1 schema in a population of sequences of N loci it is more natural to consider the *time dependent* one-dimensional landscape associated with the schema than the collective dynamics of the $2^{(N-1)}$ sequences that go up to make the 1-schema. The job of evolution at the end of the day is to produce

[7]Extensions of Geiringer's theorem to more general representations [116, 117], such as variable-length sequences, and with non-homologous operators have recently been found implementing the techniques we have outlined in this paper.

fit offspring which in their turn produce fit offspring which in their turn ... Thus, it is no use having an individual with high survival fitness that is associated with a high probability to mutate to a very low survival fitness individual.

In the case of the single peak quasispecies model the effective fitness of the master sequence, under selection and mutation only, is

$$f_{\mathrm{ms}}^{\mathrm{eff}}(t) = f_{\mathrm{ms}}(1-p)^N + f_1 \sum_{x \neq \mathrm{ms}} \frac{P_x(t)}{P_{\mathrm{ms}}(t)} p^{d_{\mathrm{ms},x}^H} (1-p)^{N-d_{\mathrm{ms},x}^H} \tag{40}$$

where f_0 and f_{ms} are the fitnesses of the master sequence and the other sequences respectively. In the limit $p \to p_{\mathrm{cri}}$ we have $f_{\mathrm{ms}}^{\mathrm{eff}}(t) \to \bar{f}(t)$. The effective fitness landscape thus becomes flat, thereby giving an intuitive explanation for the behavior in the vicinity of the error threshold. We can thus think of evolution as a hill-climbing process on an effective fitness landscape (which is time dependent). In this model mutation breaks the genotype-phenotype symmetry among the non-master sequences. Those sequences that are closer in Hamming distance to the master sequence have more reproductive success. Once again, this cannot be understood in terms of the fitness landscape as it is flat for the non-master sequences. The analog of equation (40) for non-master sequences shows us however that the effective fitness of sequences that are close to the master sequence is higher than that of distant sequences. Explicitly

$$\Delta f_{xy}^{\mathrm{eff}}(t) = (f_{\mathrm{ms}} - f_1) \frac{P_{\mathrm{ms}}(t)}{P_x(t) P_y(t)} \left(p^{d_{\mathrm{ms},x}^H} (1-p)^{N-d_{\mathrm{ms},x}^H P_y(t)} - p^{d_{\mathrm{ms},y}^H} (1-p)^{N-d_{\mathrm{ms},y}^H} P_x(t) \right) \tag{41}$$

where $\Delta f_{xy}^{\mathrm{eff}}(t)$ is the difference in effective fitness between the two genotypes x and y. The bigger is $(d_{\mathrm{ms},y}^H - d_{\mathrm{ms},x}^H)$ the greater is the total reproductive success of x versus y. Effective fitness in this sense is a direct measure of the strength of the induced breaking of the genotype-phenotype symmetry by the effect of mutation and hence offers both a qualitative and quantitative framework within which other more complex symmetry breaking phenomena such as bloat in Genetic Programming [118] and evolutionary robustness may be understood.

In Figure 7 we see a plot of the effective fitness for the master sequence of the Eigen model in the steady state as a function of p. In this case $f_{\mathrm{ms}} = 10$ and $f_1 = 1$.

For $p = 0$, $f_{\mathrm{ms}}^{\mathrm{eff}}(\infty) = f_{\mathrm{ms}}$ i.e. survival and effective fitness are the same. Note how the error threshold manifests itself in terms of the effective fitness — that at and above the threshold $f_{\mathrm{ms}}^{\mathrm{eff}}(t) \to \bar{f}(t) \approx f_1$. i.e. once again the effective fitness landscape will be flat. Thus, the effective fitness itself can serve as an order parameter to distinguish the selection dominated regime from the mutation dominated one, i.e. strong selection from weak selection. In the two-peak quasispecies model the fact that the quasispecies associated with the lower survival fitness can dominate is easily explainable using the effective fitness paradigm as the higher degree of robustness of the less fit quasispecies compensates for the fact that it is less viable, the net result being a higher value of effective fitness for this quasispecies relative to the other.

In the more subtle neurogenetic problem it is more much more difficult to think of calculating the effective fitness due to the fact that the fitness function has to be experimentally measured rather than being algorithmically determined. Additionally,

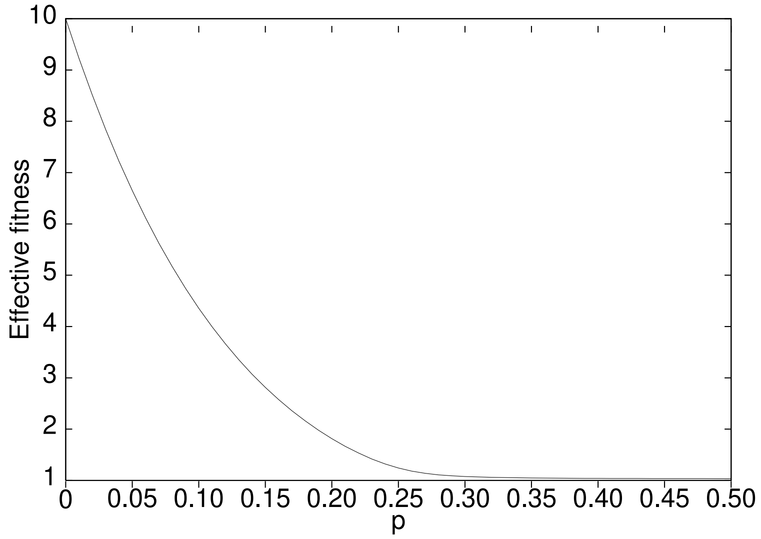

Figure 7. Graph of effective fitness of the master sequence versus p in the steady state limit of the single peak quasispecies landscape.

given that in this case the population size is far smaller than the size of the state space an experimental determination of the effective fitness would be very "noisy". However, it is clear that the example has a ready explanation in terms of effective fitness. Basically, in this context, one can think of the difference between effective fitness and reproductive fitness as being a measure of the evolutionary robustness of a genotype. Certain classes of genotypes are picked out from others of equal reproductive fitness because they are more robust in the presence of mutation and crossover and therefore enjoy, in the long term, a greater degree of reproductive success. Thus, genotypes that have block a both as a word and a grammatical rule have less reproductive success than those where a represents only a grammatical rule, even though both, in a given generation, lead to the same NN, i.e. the same phenotype. Thus, there is a population flow along an effective fitness gradient from genotypes where a plays a dual role to those where it is specialized.

8. Conclusions

Fitness is a central and fundamental paradigm in evolution theory. It has many different facets and its study is subject to many different pitfalls, both in concept and measurement. All would agree that it is a measure of reproductive success. However, to what level of description it applies, and as to how it is most appropriately measured, there is less consensus. In terms of level of description it is usually applied at the level of the individual, described by a genotype or phenotype. However, concepts, such as inclusive or lineage fitness consider fitness at the level of a population. Conceptually at least it is clear that if it can be applied at the level of the individual then, via a coarse graining, it may be applied at the level of a population, irrespective of whether the group exhibits close kinship, or is a looser agglomeration, such as a species or a family. At the level of molecular evolution these different levels of coarse graining can be even more abstract — such as passing from linear arrays of amino acids to more coarse grained secondary and tertiary protein structures. The mapping between these different levels of structure is most likely exceedingly complex. A paradigmatic example would be that of the genotype-phenotype map which can exhibit a very complex structure.

Considering measurement, we must ask how reproductive success should be measured and subsequently how this relates to characteristics of the thing who's fitness is being considered. We have emphasized that selection is only one contributing factor to reproductive success. Traditionally it has been considered to be the dominant force. However, it is clear that frequently, and in important contexts and regimes, this is not the case. Neutral evolution is one example. However, relatively speaking this is a somewhat trivial case as the subsequent dynamics is simple: random motion through the space of neutral types. When considering evolution under the effect of multiple genetic operators the fitness that traditionally appears in a mathematical representation of selection is survival fitness or viability. Its role and importance relative to other operators, such as mutation and recombination, must be carefully considered.

Fitness landscapes have played an important role in developing an intuitive picture of evolution as hill climbing on a rugged terrain. Although deducing what is the fitness landscape for a complex macroscopic evolutionary system is far from our capabilities it is important to emphasize that important classes of model landscapes can be studied as well as those associated with biologically relevant micro-systems, such as RNA. Hill climbing on a static fitness landscape is an adequate metaphor in the case of evolution dominated by selection. We have emphasized in this paper that when this is not the case, then the utility of the landscape concept is much diminished. We gave several examples where population flows could be understood neither in terms of the hill climbing paradigm nor in terms of a random diffusion process on the landscape, as in the case of neutral evolution. A particularly important class of such examples is that where the symmetry between different genotypes that correspond to the same phenotype is broken by the action of genetic operators such as mutation and recombination. In this case the system evolves along preferred directions in the neutral networks that are a consequence of the genotypic degeneracy. A simple example of this phenomenon is the increased reproductive success of error classes close to the master sequence in the case of the Eigen model. In the same model, an even more striking example is that of the error threshold. Above the error threshold the reproductive success of the master sequence is the same as that of any other sequence irrespective of the fitness advantage it has over other sequences, the latter affecting only the mutation rate characteristic of the error threshold not whether or not it exists.

We discussed the concept of effective fitness showing that in non-selection dominated regimes it is a more appropriate measure of reproductive success than viability, as it takes into account on a democratic footing all genetic operators. We showed that it offered both a qualitative and quantitative framework within which to intuitively understand population flows in non-selection dominated evolution where it restores the hill climbing paradigm, populations flowing up to effective fitness peaks. In the case of the error threshold we saw that the effective fitness of the master sequence above and beyond the critical mutation rate/error threshold was equal to that of any other sequence thus intuitively explaining in terms of flows on an effective fitness landscape why the population did not climb the master sequence peak.

We noted that effective fitness is almost always time dependent. Although this might seem like a significant disadvantage we emphasized that, at best, the concept of a *static* fitness landscape in the standard picture is an approximation. This can be understood most simple by realizing that almost any landscape represents a coarse grained model relative to an underlying more microscopic one. This is certainly true if one imagines a landscape for a macro-system. However, even at the microlevel a landscape for coarse grained protein secondary structures should properly be deduced from a corresponding landscape at the atomic level. The crucial point here, is that any coarse graining from a level where a landscape is either fundamentally or approximately static will take us to a coarse grained landscape that is dynamic, the time dependence being induced by the population dependence of the coarse graining. Whether this coarse grained landscape can be considered static or not is either a question of comparison between theory and experiment or must be addressed from a more detailed consideration of the coarse graining. Such coarse grained landscapes are certainly the appropriate arena for considering at a more fundamental level inclusive fitness and fitness of populations in general.

We also discussed how fitness and effective fitness, and their corresponding landscapes, transformed under different types of transformation of interest on the configuration space of the underlying genetic model. In particular we discussed the importance of coarse graining as this is of crucial importance in being able to pass to a reduced, more tractable model of reduced dimensionality associated with the appropriate effective degrees of freedom of the problem. In this context one might conjecture that the landscape should look less rugged in terms of the right effective degrees of freedom. We also discussed a coordinate basis that was shown to be particularly appropriate for recombination where the elements of the basis are the BBs (subsequences) associated with the construction of a given sequence. In the presence of recombination both fitness and effective fitness, and their corresponding landscapes, are most appropriately analyzed in terms of this basis.

Acknowledgments

Financial support from DGAPA project ES100201, Conacyt project 30422-E, and the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung, project 13545-MAT is gratefully acknowledged. CRS wishes to thank Riccardo Poli for many stimulating and productive discussions.

References

[1] Strickberger, M. W. *Evolution*. Jones and Bartlett, Boston MA, 1990.

[2] Hofbauer, J. and Sigmund, K. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge U.K., 1998.

[3] Bürger, R. *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester, UK, 2000.

[4] Jost, J. On the notion of fitness, or: the selfish ancestor, 2002. MPI MIS Leipzig, Internal Memorandum.

[5] Hamilton, W. D. The genetical evolution of social behaviour. I. *J. Theor. Biol.*, **7**:1–16, 1964.

[6] Grafen, A. How not to measure inclusive fitness. *Nature*, **298**:425, 1982.

[7] Creel, S. How to measure inclusive fitness. *Proc. Royal Soc. London Ser. B*, **241**:229–231, 1990.

[8] Reidys, C. M. and Stadler, P. F. Combinatorial landscapes. *SIAM Review*, **44**:3–54, 2002.

[9] Stephens, C. R. Effect of mutation and recombination on the genotype-phenotype map. In: Banzhaf, W., Daida, J., Eiben, A. E., Garzon, M. H., Honovar, V., Jakiela, M., and Smith, R. E., editors, *GECCO99*, pp. 1382–1390. Morgan-Kaufman, San Francisco, CA, 1999.

[10] Angeles, O., Stephens, C. R., and Waelbroeck, H. Emergence of algorithmic language in genetic systems. *Biosystems*, **47**:129–147, 1998.

[11] Mora, J., Stephens, C., Waelbroeck, H., and Zertuche, F. Symmetry breaking and adaptation: Evidence from a simple toy model of a viral neutralization epitope. *Biosystems*, **51**:1–14, 1999.

[12] Stephens, C. R., García Olmedo, I., Mora Vargas, J., and Waelbroeck, H. Self-adaptation in evolving systems. *Artificial Life*, **4**:183–201, 1998.

[13] Stephens, C. R. and Waelbroeck, H. Effective degrees of freedom in genetic algorithms. *Phys. Rev.*, **57**:3251–3264, 1998.

[14] Stephens, C. R. Effective fitness landscapes in evolution. In: Angeline, P., editor, *CEC99*, pp. 703–714. IEEE Press, 1999.

[15] Stephens, C. R. and Mora Vargas, J. Effective fitness as an alternative paradigm for evolutionary computation I: General formalism. *Genetic Prog. Evol. Machines*, **1**:363–378, 2000.

[16] Stephens, C. R. and Mora Vargas, J. Effective fitness as an alternative paradigm for evolutionary computation II: Examples and applications. *Genetic Prog. Evol. Machines*, **2**:7–32, 2001.

[17] Hofbauer, J. and Sigmund, K. *Dynamical Systems and the Theory of Evolution*. Cambridge University Press, Cambridge U.K., 1988.

[18] Altenberg, L. and Feldman, M. W. Selection, generalized transmission, and the evolution of modifier genes. I. The reduction principle. *Genetics*, **117**:559–572, 1987.

[19] Lyubich, Y. I. *Mathematical structures in population genetics*. Springer-Verlag, Berlin, 1992.

[20] Schuster, P. A testable genotype-phenotype map: Modeling evolution of RNA molecules. In: Lässig, M. and Valleriani, A., editors, *Biological Evolution and Statistical Physics*, pp. 56–83. Springer-Verlag, Berlin, 2002.

[21] Ferreira, C. Genetic representation and genetic neutrality in gene expression programming. *Adv. Complex Syst.*, **5**, 2002. In press.

[22] Schuster, P. and Stadler, P. F. Discrete models of biopolymers. In: J. Crabbe, A. Konopka, M. D., editor, *Handbook of Computational Chemistry and Biology*. Marcel Dekker, New York, 2002. In press.

[23] Uhlenbeck, O. C. A coat for all sequences. *Nature Struct. Biol.*, **5**:174–176, 1998.

[24] Fontana, W., Schnabl, W., and Schuster, P. Physical aspects of evolutionary optimization and adaption. *Phys. Rev. A*, **40**:3301–3321, 1989.

[25] Fontana, W., Stadler, P. F., Bornberg-Bauer, E. G., Griesmacher, T., Hofacker, I. L., Tacker, M., Tarazona, P., Weinberger, E. D., and Schuster, P. RNA folding landscapes and combinatory landscapes. *Phys. Rev. E*, **47**:2083–2099, 1993.

[26] Schuster, P., Fontana, W., Stadler, P. F., and Hofacker, I. L. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. Lond. B*, **255**:279–284, 1994.

[27] Schuster, P. Genotypes with phenotypes: Adventures in an RNA toy world. *Biophys. Chem.*, **66**:75–110, 1997.

[28] Flamm, C., Hofacker, I. L., and Stadler, P. F. RNA *in silico*: The computational biology of RNA secondary structures. *Adv. Complex Syst.*, **2**:65–90, 1999.

[29] Fontana, W. and Schuster, P. Continuity in evolution: On the nature of transitions. *Science*, **280**:1451–1455, 1998.

[30] Babajide, A., Hofacker, I. L., Sippl, M. J., and Stadler, P. F. Neutral networks in protein space: A computational study based on knowledge-based potentials of mean force. *Folding & Design*, **2**:261–269, 1997.

[31] Babajide, A., Farber, R., Hofacker, I. L., Inman, J., Lapedes, A. S., and Stadler, P. F. Exploring protein sequence space using knowledge based potentials. *J. Theor. Biol.*, **212**:35–46, 2001.

[32] Bastolla, U., Porto, M., Roman, H., and Vendruscolo, M. Connectivity of neutral networks, overdispersion, and structural conservation in protein evolution. *J. Mol. Evol.*, 2002. In press.

[33] Bastolla, U., Porto, M., Roman, H., and Vendruscolo, M. Lack of self-averaging in neutral evolution of proteins. *Phys. Rev. Lett.*, 2002. In press.

[34] Gitchoff, P. and Wagner, G. P. Recombination induced hypergraphs: A new approach to mutation-recombination isomorphism. *Complexity*, **2**:47–43, 1996.

[35] Stadler, P. F. and Wagner, G. P. The algebraic theory of recombination spaces. *Evol. Comp.*, **5**:241–275, 1998.

[36] Wagner, G. P. and Stadler, P. F. Complex adaptations and the structure of recombination spaces. In: Nehaniv, C. and Ito, M., editors, *Algebraic Engineering*, pp. 96–115. World Scientific, Singapore, 1999. Proceedings of the Conference on Semi-Groups and Algebraic Engineering, University of Aizu, Japan.

[37] Stadler, B. M. R., Stadler, P. F., Shpak, M., and Wagner, G. P. Recombination spaces, metrics, and pretopologies. *Z. Phys. Chem.*, **216**:217–234, 2002.

[38] Stadler, B. M. R. and Stadler, P. F. Generalized topological spaces in evolutionary theory and combinatorial chemistry. *J. Chem. Inf. Comput. Sci.*, **42**:577–585, 2002. Proceedings MCC 2001, Dubrovnik.

[39] Changat, M., Klavžar, S., and Mulder, H. M. The all-path transit function of a graph. *Czech. Math. J.*, **51**:439–448, 2001.

[40] Morgana, M. A. and Mulder, H. M. The induced path convexity, betweenness, and svelte graphs. *Discr. Math.*, **254**:349–370, 2002.

[41] Kadanoff, L. P. *Statistical Physics: Statics, Dynamics and Renormalization.* World Scientific, Singapore, 2000.

[42] Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, **10**:465–523, 1971.

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

[44] Vose, M. D. *The Simple Genetic Algorithm.* MIT Press, MA, 1999.

[45] Dress, A. W. M. and Rumschitzky, D. S. Evolution on sequence space and tensor products of representation spaces. *Acta Appl. Math.*, **11**:103–115, 1988.

[46] Weinberger, E. D. Fourier and Taylor series on fitness landscapes. *Biological Cybernetics*, **65**:321–330, 1991.

[47] Rockmore, D., Kostelec, P., Hordijk, W., and Stadler, P. F. Fast fourier transform for fitness landscapes. *Appl. Comput. Harmonic Anal.*, **12**:57–76, 2002.

[48] Stadler, P. F., Seitz, R., and Wagner, G. P. Evolvability of complex characters: Population dependent Fourier decomposition of fitness landscapes over recombination spaces. *Bull. Math. Biol.*, **62**:399–428, 2000.

[49] Hansen, T. F. and Wagner, G. P. Modelling genetic architecture: A multilinear theory of gene interaction. *Theor. Pop. Biol.*, **59**:61–86, 2001.

[50] Stephens, C. R. and Waelbroeck, H. Effective degrees of freedom of genetic algorithms and the block hypothesis. In: Bäck, T., editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pp. 31–41. Morgan Kaufmann, San Francisco, CA, 1997.

[51] Stephens, C. R. and Waelbroeck, H. Schemata evolution and building blocks. *Evol. Comp.*, **7**:109–124, 1999.

[52] Stephens, C. R. The renormalization group and the dynamics of genetic systems. *Acta Phys. Slov.*, **to be published**, 2003. Preprint: cond-mat/0210217.

[53] Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley, Reading MA, 1989.

[54] Wright, S. The roles of mutation, inbreeding, crossbreeeding and selection in evolution. In: Jones, D. F., editor, *Proceedings of the Sixth International Congress on Genetics*, vol. 1, pp. 356–366. 1932.

[55] Wright, S. "Surfaces" of selective value. *Proc. Nat. Acad. Sci. USA*, **58**:165–172, 1967.
[56] Binder, K. and Young, A. P. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Rev. Mod. Phys.*, **58**:801–976, 1986.
[57] Mézard, M., Parisi, G., and Virasoro, M. *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1987.
[58] Chan, H. S. and Dill, K. A. Sequence space soup. *J. Chem. Phys.*, **95**:3775–3787, 1991.
[59] Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yeo, D. P., Thomas, P. D., and Chan, H. S. Principles of protein folding: a perspective from simple exact models. *Prot. Sci.*, **4**:561–602, 1995.
[60] Onuchic, J. N., Nymeyer, H., Garcia, A. E., Chahine, J., and Socci, N. D. The energy landscape theory of protein folding: Insights into folding mechanisms and scenarios. *Adv. Protein Chem.*, **53**:87–152, 2000.
[61] Flamm, C., Fontana, W., Hofacker, I., and Schuster, P. RNA folding kinetics at elementary step resolution. *RNA*, **6**:325–338, 2000.
[62] Mezey, P. G. *Potential Energy Hypersurfaces*. Elsevier, Amsterdam, 1987.
[63] Heidrich, D., Kliesch, W., and Quapp, W. *Properties of Chemically Interesting Potential Energy Surfaces*, vol. 56 of *Lecture Notes in Chemistry*. Springer-Verlag, Berlin, 1991.
[64] Kollman, K., Miller, J. H., and Page, S. E. Adaptive parties in spatial elections. *Amer. Pol. Sci. Rev.*, **86**:929–937, 1992.
[65] Stadler, B. M. Adaptive platform dynamics in multi-party spatial voting. *Adv. Complex Syst.*, **2**:101–116, 1999.
[66] Garey, M. and Johnson, D. *Computers and Intractability. A Guide to the Theory of $\mathcal{NP}$ Completeness*. Freeman, San Francisco, 1979.
[67] Rechenberg, I. *Evolutionstrategie*. Frommann-Holzboog, Stuttgart, 1973.
[68] Koza, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1992.
[69] Holland, J. H. *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, MA, 1993.
[70] Kauffman, S. A. *The Origin of Order*. Oxford University Press, New York, Oxford, 1993.
[71] Fogel, D. B. *Evolutionary Computation*. IEEE Press, New York, 1995.
[72] Stadler, P. F. Fitness landscapes. In: Lässig, M. and Valleriani, A., editors, *Biological Evolution and Statistical Physics*, pp. 187–207. Springer-Verlag, Berlin, 2002.
[73] Stadler, P. F. Landscapes and their correlation functions. *J. Math. Chem.*, **20**:1–45, 1996.
[74] Reidys, C. M. and Stadler, P. F. Neutrality in fitness landscapes. *Appl. Math. & Comput.*, **117**:321–350, 2001.
[75] Stadler, P. F. and Tinhofer, G. Equitable partitions, coherent algebras and random walks: Applications to the correlation structure of landscapes. *MATCH*, **40**:215–261, 1999.
[76] Grover, L. K. Local search and the local structure of NP-complete problems. *Oper. Res. Lett.*, **12**:235–243, 1992.
[77] Davies, E. B., Gladwell, G. M. L., Leydold, J., and Stadler, P. F. Discrete nodal domain theorems. *Lin. Alg. Appl.*, **336**:51–60, 2001.
[78] Hordijk, W. and Stadler, P. F. Amplitude spectra of fitness landscapes. *Adv. Complex Systems*, **1**:39–66, 1998.
[79] Stadler, P. F. Spectral landscape theory. In: Crutchfield, J. P. and Schuster, P., editors, *Evolutionary Dynamics—Exploring the Interplay of Selection, Neutrality, Accident, and Function*. Oxford University Press, 2002. In press.
[80] Whitlock, M. C., Phillips, P. C., Moore, F. B., and Tonsor, S. J. Multiple fitness peaks and epistasis. *Ann. Rev. Ecol. System.*, **26**:601–629, 1995.
[81] Weinberger, E. D. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol. Cybern.*, **63**:325–336, 1990.
[82] Stadler, P. F. and Schnabl, W. The landscape of the travelling salesman problem. *Phys. Lett. A*, **161**:337–344, 1992.
[83] Ferreira, F. F., Fontanari, J. F., and Stadler, P. F. Landscape statistics of the low autocorrelated binary string problem. *J. Phys. A: Math. Gen.*, **33**:8635–8647, 2000.

[84] Flamm, C., Hofacker, I. L., Stadler, P. F., and Wolfinger, M. T. Barrier trees of degenerate landscapes. *Z. Phys. Chem.*, **216**:155–173, 2002.

[85] Stadler, P. F. Fitness landscapes arising from the sequence-structure maps of biopolymers. *J. Mol. Struct. (THEOCHEM)*, **463**:7–19, 1999.

[86] Reidys, C. M., Stadler, P. F., and Schuster, P. Generic properties of combinatory maps: Neural networks of RNA secondary structures. *Bull. Math. Biol.*, **59**:339–397, 1997.

[87] Reidys, C. M. Random induced subgraphs of generalized n-cubes. *Adv. Appl. Math.*, **19**:360–377, 1997.

[88] Grüner, W., Giegerich, R., Strothmann, D., Reidys, C. M., Weber, J., Hofacker, I. L., Stadler, P. F., and Schuster, P. Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. *Monath. Chem.*, **127**:355–374, 1996.

[89] Grüner, W., Giegerich, R., Strothmann, D., Reidys, C. M., Weber, J., Hofacker, I. L., Stadler, P. F., and Schuster, P. Analysis of RNA sequence structure maps by exhaustive enumeration. II. Structures of neutral networks and shape space covering. *Monath. Chem.*, **127**:375–389, 1996.

[90] Schultes, E. A. and Bartel, D. P. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science*, **289**:448–452, 2000.

[91] Sippl, M. J. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Computer-Aided Molec. Design*, **7**:473–501, 1993.

[92] Dalal, S., Balasubramanian, S., and Regan, L. Protein alchemy: Changing β-sheet into α-helix. *Nat. Struct. Biol.*, **4**:548–552, 1997.

[93] Keefe, A. D. and Szostak, J. W. Functional proteins from a random-sequence library. *Nature*, **410**:715–718, 2001.

[94] Martinez, M. A., Pezo, V., Marlière, P., and Wain-Hobson, S. Exploring the functional robustness of an enzyme by *in vitro* evolution. *EMBO J.*, **15**:1203–1210, 1996.

[95] Derrida, B. and Peliti, L. Evolution in a flat fitness landscape. *Bull. Math. Biol.*, **53**:355–382, 1991.

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

[97] Huynen, M. A. Exploring phenotype space through neutral evolution. *J. Mol. Evol.*, **43**:165–169, 1996.

[98] Stadler, P. F. and Grüner, W. Anisotropy in fitness landscapes. *J. Theor. Biol.*, **165**:373–388, 1993.

[99] Stadler, P. F. and Happel, R. Random field models for fitness landscapes. *J. Math. Biol.*, **38**:435–478, 1999.

[100] Goldberg, D. E. Genetic algorithms and Walsh functions. Part II: deceptiveness and its analysis. *Complex Systems*, **3**:153–176, 1989.

[101] Liepins, G. E. and Vose, M. D. Deceptiveness and genetic algorithm dynamics. In: Rawlins, G. J. E., editor, *Foundations of Genetic Algorithms*, pp. 36–50. Morgan Kaufmann, San Mateo, CA, 1991.

[102] Whitley, L. D. Fundamental principles of deception in genetic search. In: Rawlins, G., editor, *Foundations of Genetic Algorithms*, pp. 221–241. Morgan Kaufmann, San Mateo, CA, 1991.

[103] Naudts, B. and Kallel, L. Comparison of summary statistics of fitness landscapes. *IEEE Trans. Evol. Comp.*, 2000. To appear.

[104] Weibull, J. W. *Evolutionary Game Dynamics*. MIT Press, Cambridge MA, 1996.

[105] Swetina, J. and Schuster, P. Self-replication with errors – a model for polynucleotide replication. *Biophys.Chem.*, **16**:329–345, 1982.

[106] van Nimwegen, E., Crutchfield, J. P., and Huynen, M. A. Neutral evolution of mutational robustness. *Proc. Natl. Acad. Sci. USA*, **96**:9716–9720, 1999.

[107] Gavrilets, S. Evolution and speciation on holey adaptive landscapes. *Trends in Ecology and Evolution*, **12**:307–312, 1997.

[108] Wagner, A. and Stadler, P. F. Viral rna and evolved mutational robustness. *J. Exp. Zool./ MDE*, **285**:119–127, 1999.

[109] van Nimwegen, E. and Crutchfield, J. P. Metastable evolutionary dynamics: Crossing fitness barriers or escaping via neutral paths? *Bull. Math. Biol.*, **62**:799–848, 2000.

[110] Kitano, H. Neurogenetic learning: an integrated method of designing and training neural networks using genetic algorithms. *Physica D*, **75**:225–238, 1994.

[111] Leuthäusser, I. An exact correspondence between eigen's evolution model and a two-dimensional ising model. *J. Chem. Phys.*, **84**:1884, 1986.

[112] Nordin, P. and Banzhaf, W. Complexity compression and evolution. In: Eshelman, L., editor, *Genetic Algorithms: Proceedings of the Sixth International Conference (ICGA95)*, pp. 310–317. Morgan Kaufmann, Pittsburgh, PA, 1995.

[113] Nordin, P., Francone, F., and Banzhaf, W. Explicitly defined introns and destructive crossover in genetic programming. In: Rosca, J. P., editor, *Proceedings of the Workshop on Genetic Programming: From Theory to Real World Applications*, pp. 6–22. Tahoe City, CA, 1995.

[114] Goldberg, D. E. Genetic algorithms and Walsh functions: Part I. A gentle introduction. *Complex Systems*, **3**:123–152, 1989.

[115] Goldberg, D. E. Genetic algorithms and Walsh functions: Part II. Deception and its analysis. *Complex Systems*, **3**:153–171, 1989.

[116] Poli, R., Rowe, J. E., Stephens, C. R., and Wright, A. H. Allele diffusion in linear genetic programming and variable-length genetic algorithms with subtree crossover. In: *Genetic Programming, Proceedings of the 5th European Conference, EuroGP 2002, LNCS, Vol. 2278*, pp. 211–227. Springer-Verlag, 2002.

[117] Poli, R., Stephens, C. R., Rowe, J. E., and Wright, A. H. On the search biases of homologous crossover in linear genetic programming and variable-length genetic algorithms. In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2002)*, pp. 211–227. Morgan Kaufmann, 2002.

[118] Langdon, W. B. and Poli, R. *Foundations of Genetic Programming*. Springer Verlag, Berlin, New York, 2002.