Mutation-selection models solved exactly with methods of statistical mechanics

Ellen Baake*
Zoologisches Institut
Universität München
Luisenstr. 14
D-80333 München
Germany
email: baake@zi.biologie.uni-muenchen.de
phone: (+49) 89 5902 327
FAX: (+49) 89 5902 474

Holger Wagner
Max-Planck-Institut für Biophysikalische Chemie
Am Faßberg 11
D-37077 Göttingen
Germany
email: hwagner1@gwdg.de
phone: (+49) 551-2011306
FAX: (+49) 551-2011089

July 19, 2001

*address for proofs.

Summary

We reconsider deterministic models of mutation and selection acting on populations of sequences, or, equivalently, multilocus systems with complete linkage. Exact analytical results concerning such systems are few, and we present recent and new ones obtained with the help of methods from quantum statistical mechanics.

We consider a continuous-time model for an infinite population of haploids (or diploids without dominance), with N sites each, two states per site, symmetric mutation, and arbitrary fitness function. We show that this model is exactly equivalent to a so-called Ising quantum chain. In this picture, fitness corresponds to the interaction energy of spins, and mutation to a temperature-like parameter. The highly elaborate methods of statistical mechanics allow one to find exact solutions for nontrivial examples. These include quadratic fitness functions, as well as 'Onsager's landscape'. The latter is a fitness function which captures some essential features of molecular evolution, like neutrality, compensatory mutations and flat ridges. We investigate the mean number of mutations, the mutation load, and the variance in fitness under mutation-selection balance. This also yields some insight into the 'error threshold' phenomenon, which occurs in some, but not all, examples.

1 Introduction

Models of mutation and selection as the only evolutionary forces (i.e. acting on a population of infinite size, without recombination and migration) have been considered in very different contexts, depending on what the unit of selection is supposed to be. In classical population genetics, there is a long tradition of one-locus K-allele models, as well as multilocus models with complete linkage; for review, see Bürger (1998) or Bürger (2000). Where absence of recombination is adequate for one-locus models, this is not so in the multilocus case; but even here, complete linkage serves as an important reference case for the evaluation of recombination effects.

A molecular variant of one-locus K-alleles models is obtained when the alleles are identified with sequences, and an appropriate mutation model is formulated. Deterministic mutation-selection models based on sequences (as opposed to the stochastic models for finite populations, which will not be our concern here) were first considered in the context of prebiotic evolution, and became known under the name of sequence space models (for review, see Eigen *et al.*, 1989). More recently, similar models have also been considered in molecular population genetics, e.g. O'Brien (1985); Li (1987). Due to the sequential structure of the 'alleles' involved, these models are, at the same time, equivalent to multilocus models with complete linkage. (Actually, sequence space models have arisen and evolved independently and only later (Higgs, 1994; Baake, 1995) been identified with mutation-selection models of population genetics. As a consequence, a special terminology was created (divergent evolution), and many results were duplicated (convergent evolution)).

Depending on the context, different mutation models and fitness functions have been used. In multilocus models, mutation is usually assumed unidirectional at every site (from wild-type to mutant). In the sequence context, on the other hand, symmetric mutation at every nucleotide site is more adequate. As to the fitness function, one common mode of selection is directional (we will not consider stabilizing selection here), with a unique fittest genotype and fitness a decreasing function of the number of deleterious mutations. In multilocus models, this is often a quadratic or Gaussian (cf. Kimura & Maruyama, 1966; Charlesworth, 1990), or a step function (truncation selection, e.g. Kondrashov 1988). The prototype fitness function of sequence space models has a single sharp peak, i.e. one type has a selective advantage over all others, which are equally unfit. This is an extreme case of truncation selection which seems highly artificial from the biological point of view, see the discussion in Charlesworth (1990).

The questions addressed with both types of models are fairly similar. Most of them are related to aspects of *mutation-selection balance*: How many mutations (relative to the wildtype) do individuals carry on average? What is the mutation load, i.e. the loss in fitness caused by the production of maladapted individuals due to mutation, under various fitness schemes? How large is the genetic variation within the population?

The interest in the mutation load lies in its significance for the evolution of mutation rates, genome sizes, and recombination. In spite of decades of experimental research, it is still not clear whether the mutation load is a small nuisance, or a weighty factor in nature; for review, see Crow (1993) and Kondrashov (1998). From the theoretical point of view, one corner stone is the well-known Haldane Principle: If organisms are haploid or diploid without dominance, reproduce without recombination, and mutation is unidirectional at every site, the mutation load at equilibrium equals the genomic mutation rate, independently of the fitness function; for other mutation schemes, this is still true to first order in the mutation rate (Kimura & Maruyama, 1966; Bürger, 1998). However, this is no longer true in diploids with dominance, or with recombination. The effect of epistasis in this case was analyzed in the classical paper by Kimura & Maruyama (1966) and, more recently, by Charlesworth (1990) and Higgs (1994). For quadratic fitness functions and various breeding systems, the general picture emerged that positive (i.e. diminishing returns) epistasis enhances the mutation load (with respect to the case without epistasis), whereas negative (i.e. synergistic) epistasis decreases it.

In the context of sequence space models, populations are usually haploid, and much attention has been paid to the mutation load in the disguise of the error threshold phenomenon. This may be generally, but vaguely, circumscribed as a critical mutation rate beyond which mutation can no longer be controlled by selection and leads to genetic degeneration. More specifically, it was originally described for the sharply-peaked landscape (see Fig. 1) and defined as a critical mutation rate above which the fittest genotype is lost from the population (see Eigen *et al.* (1989) for a review). In this setting, an inverse relationship between mutation rate and the (maximum) genome size is predicted. Such a relationship is, indeed, observed over a large variety of organisms, see the recent survey by Drake *et al.* (1998). Further, a threshold mutation rate above which the population goes extinct has been observed in mutagenesis experiments with certain RNA viruses (but not, or not so far, for organisms with larger genomes), see Holland *et al.* (1990), and Domingo & Holland (1997) for a review. Such observations have been taken as evidence for the existence of error thresholds, e.g. Maynard Smith & Szathmáry (1995). But the argument relies on the unnatural fitness function assumed, and may therefore be irrelevant (e.g. Charlesworth, 1990).

This was corroborated by Charlesworth (1990) and Higgs (1994) who noted that, in line with the results on mutation loads, error thresholds may occur for certain fitness functions with positive epistasis, but are absent if epistasis is negative. A similar observation was made by Wiehe (1997). Clearly, the results obtained, as well as the conclusions drawn (on the relevance of mutation loads, or error thresholds), hinge on the assumptions on the fitness functions. Since fitness landscapes are hard to access experimentally (but see (Whitlock *et al.*, 1995) for current knowledge and approaches), an important contribution of theory would be to give answers as general as possible, for large families of fitness functions. As noted already, the sharply-peaked and quadratic (or the closely related Gaussian) fitness functions have been prominent examples. They all come from what we would like to call the permutation-invariant family, where fitness is a function of the number of deleterious mutations, independently of their position in the genome. Usually, this function is decreasing, which rules out the existence of compensatory mutations in the sense that an additional mutation can never recover fitness.

Even these well-established fitness functions are hard to treat. Exact solutions are rare; one relies on approximate or numerical treatment. The main reason is that the usual tools from dynamical systems theory are inadequate to handle the large numbers of types involved. Recently, methods of statistical physics (which, by definition, deals with large numbers of particles) have been employed to arrive at *exact* solutions of certain mutation-selection models (Baake *et al.*, 1997; Wagner *et al.*, 1998). The purpose of the present paper is to make this toolbox (originally published in condensed form for the physics community) available to a population genetic readership, and to extend the results. We shall illustrate these methods by tackling the quadratic fitness functions, as well as a novel fitness landscape which is outside the permutation invariant family, and allows for compensatory mutations. For simplicity, we shall restrict ourselves to fitness schemes without dominance. We shall focus on symmetric mutation since it has attracted much recent interest (e.g. McVean & Charlesworth, 2000) due to its relevance for molecular population genetics, but is less well explored than the (simpler) unidirectional case. More general fitness functions as well as mutation models (including asymmetric mutation) will be tackled in a forthcoming publication.

The article is organized as follows. In Section 2, we shall discuss the mutation-selection model, along with our choice of fitness functions. To avoid duplication, we shall discuss multilocus and sequence space models in a unified framework, subsuming both types of models under a common terminology. In Section 3, we shall establish the equivalence between the mutation-selection model and a model of quantum statistical mechanics. A short interlude, Section 4, is concerned with the population averages ('observables') which we employ to characterize mutation-selection balance. Section 5 is devoted to the limit of an infinite number of sites. This is performed with a scaling unconventional in the population genetics literature and provides the key to much of our analysis. In Section 6, the method is applied to the fitness functions described in Section 2, where most of our effort will go into exact solutions of the quadratic fitness scheme. The results will be discussed in Section 7.

2 The mutation-selection model

2.1 General considerations

Haploid genotypes are identified with linear arrangements of N sites. Each site i is equipped with a variable s_i which may take values from the set $\{-1,+1\}$. A configuration (or type) may be denoted by $\boldsymbol{s} = s_1 s_2 \dots s_N \in \{-1,+1\}^N$. In the multilocus context, 'sites' are identified with 'loci', 'variables' s_i with 'alleles', and '+' (or $+1$) and '−' (or -1) with 'wildtype' and 'mutant', respectively. This way, $\boldsymbol{s}$ is the configuration of one gamete; the corresponding gametic space was introduced and visualized by Wright (1932). In the molecular context, configurations are DNA (or RNA) sequences, and we assume the lumping of nucleotides into purines (+) and pyrimidines ($-$), which is common in molecular evolution, cf. Swofford *et al.* (1995) or Li (1997). The set $\{-1,+1\}^N$ is sometimes called sequence space. (A four-letter alphabet is tackled in a forthcoming paper (Hermisson *et al.*, 2000)). Although genetic information is finite in principle, the limit $N \to \infty$ is often appropriate thanks to the large number of sites.

Let us now consider an infinite population of haploids, or diploids without dominance, which experience mutation and selection, but no recombination. We shall assume overlapping generations in continuous time t, and use the parallel (or decoupled) mutation-selection model (see, for example, Crow & Kimura (1970), or Hofbauer (1985)) to describe the dynamics of relative frequencies $p_{\boldsymbol{s}}$ of types $\boldsymbol{s}$, that is

$$\dot{p}_{\boldsymbol{s}}(t) = \big(R_{\boldsymbol{s}} - \bar{R}(t)\big)p_{\boldsymbol{s}}(t) + \sum_{\boldsymbol{s}'} m_{\boldsymbol{s}\boldsymbol{s}'} p_{\boldsymbol{s}'}(t)\,, \quad \boldsymbol{s} \in \{-1,+1\}^N\,. \tag{1}$$

Here, $R_{\boldsymbol{s}}$ denotes the Malthusian fitness of type $\boldsymbol{s}$, and $\bar{R}(t) := \sum_{\boldsymbol{s}} R_{\boldsymbol{s}} p_{\boldsymbol{s}}(t)$ is the mean fitness of the population (note that $\sum_{\boldsymbol{s}} p_{\boldsymbol{s}}(t) \equiv 1$). Further, $m_{\boldsymbol{s}'\boldsymbol{s}} = m_{\boldsymbol{s}' \leftarrow \boldsymbol{s}} \geq 0$ is the rate at which $\boldsymbol{s}$ mutates to $\boldsymbol{s}'$ for $\boldsymbol{s}' \neq \boldsymbol{s}$, and $m_{\boldsymbol{s}\boldsymbol{s}} = -\sum_{\boldsymbol{s}' \neq \boldsymbol{s}} m_{\boldsymbol{s}'\boldsymbol{s}}$. That is, the matrix $(m_{\boldsymbol{s}'\boldsymbol{s}})$ is a Markov generator (Rozanov, 1977). Finally, the dot denotes derivative with respect to time. If, instead of haploid genotypes, one has diploid ones, $\boldsymbol{s}\boldsymbol{s}'$, with genotypic fitnesses $R_{\boldsymbol{s}\boldsymbol{s}'}$, absence of dominance (in fitness) is assumed, i.e. $R_{\boldsymbol{s}\boldsymbol{s}'} = \frac{1}{2}(R_{\boldsymbol{s}\boldsymbol{s}} + R_{\boldsymbol{s}'\boldsymbol{s}'})$, and the $R_{\boldsymbol{s}}$ are defined as $R_{\boldsymbol{s}} := \frac{1}{2} R_{\boldsymbol{s}\boldsymbol{s}}$.

It is well-known (Thompson & McBride, 1974; Moran, 1976) that the transformation

$$x_{\boldsymbol{s}}(0) := p_{\boldsymbol{s}}(0), \quad x_{\boldsymbol{s}}(t) := p_{\boldsymbol{s}}(t) \exp\Big(\int_0^t \bar{R}(\tau) d\tau\Big)\,, \tag{2}$$

turns Eq. (1) into a linear system of differential equations, namely

$$\dot{x}_{\boldsymbol{s}}(t) = R_{\boldsymbol{s}} x_{\boldsymbol{s}}(t) + \sum_{\boldsymbol{s}'} m_{\boldsymbol{s}\boldsymbol{s}'} x_{\boldsymbol{s}'}(t) \quad \text{or} \quad \dot{\boldsymbol{x}}(t) = (\mathcal{M} + \mathcal{R})\, \boldsymbol{x}(t)\,, \tag{3}$$

where, with a suitable enumeration of types, $\boldsymbol{x}$ may be understood as the vector of *absolute* frequencies, from which the *relative* frequencies $p_{\boldsymbol{s}}$ may be retrieved through normalization, $p_{\boldsymbol{s}}(t) = x_{\boldsymbol{s}}(t) / \big(\sum_{\boldsymbol{s}'} x_{\boldsymbol{s}'}(t)\big)$. $\mathcal{M}$ is the mutation matrix with entries $m_{\boldsymbol{s}\boldsymbol{s}'}$, and $\mathcal{R}$ is the diagonal matrix which holds the $R_{\boldsymbol{s}}$. The latter may be understood as a reproduction matrix (we avoid the term 'fitness matrix' here since this is reserved for the fitness values of diploid genotypes). Due to the usual argument of positive invariance (Amann, 1990, Thm. 16.5), solutions of (1) cannot leave the positive cone, i.e., $x_{\boldsymbol{s}}(t) \geq 0$ for all times if $x_{\boldsymbol{s}}(0) \geq 0$ for all $\boldsymbol{s}$.

The general solution of the linear system (3) reads

$$\boldsymbol{x}(t) = \exp\big(t(\mathcal{M} + \mathcal{R})\big)\boldsymbol{x}(0)\,. \tag{4}$$

However, Eq. (4) remains a somewhat formal expression as long as the matrix exponential is not evaluated explicitly, which is rarely possible for large numbers of types. Since the matrix obtained from $\mathcal{M}$ by setting its diagonal entries to zero will be assumed to be primitive (i.e. some power of it has strictly positive entries only), the system will converge towards an equilibrium (mutation-selection balance). This is a consequence of the Perron-Frobenius theorem: The equilibrium is given by the eigenvector belonging to the largest eigenvalue of $\mathcal{M} + \mathcal{R}$ if supplied with positive sign and correct normalization, cf. Thompson & McBride (1974); Moran (1976).

In the sequence space context, the *coupled* mutation-selection equation,

$$\dot{p}_{\boldsymbol{s}}(t) = \left(\sum_{\boldsymbol{s}'} v_{\boldsymbol{s}\boldsymbol{s}'} R_{\boldsymbol{s}'} p_{\boldsymbol{s}'}(t) \right) - \bar{R}(t) p_{\boldsymbol{s}}(t) \,, \tag{5}$$

has been more popular so far; it was studied in the classical context by Akin (1979), Hadeler (1981), and others. Here, $v_{\boldsymbol{s}\boldsymbol{s}'} := v_{\boldsymbol{s} \leftarrow \boldsymbol{s}'}$ is the probability of $\boldsymbol{s}'$ producing offspring of type $\boldsymbol{s}$ on the occasion of a reproduction event ($\sum_{\boldsymbol{s}} v_{\boldsymbol{s}\boldsymbol{s}'} = 1$). Unlike the parallel model, where reproduction and mutation proceed independently of each other, mutations are coupled to reproduction here. As a consequence, the number of mutation events is proportional to time in the parallel version, but proportional to the number of reproduction events (and, hence, 'generations') in the coupled one. With an appropriate choice of parameters, however, both versions display very similar behaviour. This is due to the fact that the parallel version emerges from the coupled one in the limit of weak selection and mutation, as shown by Hofbauer (1985). In a similar way, the mutation-selection model with *discrete* generations converges to the parallel version in the limit of short generation times (see the Appendix and Fig. 4); this is equivalent to weak selection and mutation through a rescaling of time.

2.2 Mutation model and fitness functions

Let us now specify mutation rates and fitness functions. We start with mutation because the choice is straightforward and dealt with quickly. We shall restrict ourselves to symmetric mutation at every site. If, in the parallel model (1), every site mutates independently at rate μ, the elements of the mutation matrix read

$$m_{\boldsymbol{s}'\boldsymbol{s}} = \begin{cases} \mu, & D(\boldsymbol{s}, \boldsymbol{s}') = 1 \\ -N\mu, & \boldsymbol{s} = \boldsymbol{s}' \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $D(\boldsymbol{s}', \boldsymbol{s})$ is the *Hamming* (or *mutational*) *distance* of $\boldsymbol{s}$ and $\boldsymbol{s}'$, i.e. the number of sites where $\boldsymbol{s}$ and $\boldsymbol{s}'$ differ. Clearly, this mutation matrix (plus $N\mu$ times the identity) is primitive. Note that we have specified a fixed mutation rate per site, which implies that the overall mutation rate increases linearly with N, in line with the molecular mutation mechanism. However, constant genomic mutation rate (as often assumed in multilocus theory) may be considered as well by keeping $N\mu$ constant and reinterpreting μ accordingly. We shall discuss the general issue of scaling in a separate section, after we have introduced the various fitness functions.

Due to its intuitive appeal, we shall sometimes use the notion of *fitness landscape* in the way introduced by Kauffman & Levin (1987) as synonymous with 'fitness function', i.e. for the mapping from *(geno)types* into *(individual) fitness values*; this is not to be confused with the mapping from *population composition* into *mean fitness*, as often referred to as an *adaptive landscape* in the sense of Wright (1932). We shall be concerned with two fitness landscapes, the first of which is from the permutation-invariant family.

1. Quadratic fitness functions: Let '$+ + + + + \ldots +$' be the reference (mutation free) type and k the number of '$-$' sites, i.e. the number of sites which carry mutations, or the mutational distance from the reference genotype. We shall consider the family of quadratic fitness functions

$$R_k = -(ak + \frac{c}{2N} k^2) \tag{7}$$

with parameters a and c, where R_k may be understood as deficit in fitness of a type with k mutations, relative to the reference type (note that Malthusian fitnesses need only be specified up to an additive constant, since this cancels out of Eq. (1); in particular, Malthusian fitnesses do not suffer from being negative).

Quadratic fitness functions provide a good fit to the mutation accumulation data in *Drosophila* (Crow & Simmons, 1983), and have been used, for example, by Kimura & Maruyama (1966) to study the effect of epistasis on mutation-selection balance; so has the corresponding Gaussian fitness function as used by Charlesworth (1990). $c > 0$ implies synergistic (or negative) epistasis, whereas $c < 0$ means diminishing returns (or positive) epistasis. (We shall, in what follows, use the terms of 'positive' and 'negative' epistasis rather than 'diminishing returns' and 'synergistic', since the former do not depend on whether deleterious or beneficial mutations are considered. Note that *positive curvature* of the fitness function ($c < 0$) is identified with *positive epistasis* because it induces *positive linkage disequilibria*, and vice versa. For a very transparent dissection of these notions, see Phillips *et al.* (2000)). With $c = 0$, one has additive fitness, which corresponds to independent action of sites, i.e. absence of epistasis. If Wrightian instead of Malthusian fitness is considered, one obtains the more familiar multiplicative fitness scheme, $W_k = \exp(R_k) = (1-\alpha)^k$, where $\alpha := 1 - \exp(-a)$ with a from (7), and generations of unit length have been implied (see also Eq. (81) in the Appendix).

Our definition (7) differs from previous ones (Kimura & Maruyama, 1966; Charlesworth, 1990) by the factor $1/2N$ in the scaling of the quadratic term. While, for any finite N, this may just be understood as the redefinition of the quadratic coefficient, this point is important to get the scaling right in the limit $N \to \infty$. Again, this issue is deferred to the section on scaling; let us only note here that our choice ensures that, for a given *fraction* $d := k/N$ of mutated sites, fitness scales linearly with N, since R_k may be written as $R_d = -(ad + cd^2/2)N$.

Some concerns have been raised about quadratic fitness functions since, with certain combinations of a and c, fitness first decreases and then rises again with increasing k, which is an artifact introduced by the symmetry of the fitness function (Charlesworth, 1990). Various remedies are possible. In Charlesworth (1990) and Hermisson *et al.* (2000), the fitness function is cut off explicitly, i.e., R_k is replaced by its minimum value for k beyond the position of the minimum; it was observed that the results are very similar whether or not this measure is taken (Charlesworth, 1990; Hermisson *et al.*, 2000), at least for most combinations of the parameters. We shall adopt a somewhat different point of view here which is based on the limit $N \to \infty$. We shall see that, in this limit, there are two branches of the equilibrium solution whenever the fitness function has a secondary maximum – each branch is derived from one of the maxima (and, in fact, there is a third branch derived from the minimum between them). For $N \to \infty$, the branch derived from the larger fitness peak is globally attracting, irrespective of the height (or shape) of a possible secondary peak. To a very good approximation, this also holds for large N; in fact, even the tiniest symmetry breaking suffices to depopulate the 'lower' branch in numerical simulations (cf. Fig. 1). To make this plausible, consider the situation without mutation. Then, the whole population will finally arrive at the higher peak, no matter how minute its selective advantage, provided the fittest genotype is at all present initially. With mutation, some leakage will occur to the second peak, but it will become less pronounced with increasing distance of the peaks, which is of order N for the quadratic fitness function. Hence, the lower peak will be effectively empty for large N. By cutting off the fitness function, one therefore makes sure that the population is located near the primary maximum (at the wildtype). Alternatively, this may be realized by just 'picking' the corresponding branch and ignoring an artificial second one if present, whether or not the secondary maximum to which it belongs is lower than the primary one.

2. Onsager's landscape: All landscapes discussed so far are in the permutation-invariant family. They have been standard landscapes in population genetics since large simplifications result from this property. However, permutation invariance cannot seriously be considered a reasonable biological property. First of all, the fact that a permutation-invariant function can fit the data does not imply that the function is a good description of how genes interact. This was pointed out by Phillips *et al.* (2000), who further demonstrated that many other models, from within and without the permutation-invariant class, fit the data equally well. We go one step further here and also consider compensatory mutations, which, by construction, can never occur in a permutation-invariant model, as long as the fitness function is monotonic. As a first attempt to introduce compensatory mutations, we borrow from physics what we would like to call 'Onsager's landscape'. Here, $++++\cdots++$ is the fittest type, and the fitness deficit is 2γ times the number of 'domain walls' (i.e. changes in sign) in the configuration, e.g.

$$\begin{array}{cc} \boldsymbol{s} & R_{\boldsymbol{s}} \\ ++++++++ & 0 \\ +--+++++ & -4\gamma \\ +--++--+ & -8\gamma \\ \vdots & \vdots \end{array} \tag{8}$$

Note that the fitness values are invariant under reversal of all signs; i.e. the landscape has a 'mirror image' with $----\cdots--$ as a second fitness maximum. As with quadratic landscapes, however, we shall just ignore the solution branch derived from the secondary maximum, thus effecting a cutoff of the landscape beyond its minimum. This is a valid procedure for reasons similar to those discussed for the quadratic fitness function, which we shall cast in more mathematical terms here. For finite N, the two symmetric fitness peaks will be equally populated due to the uniqueness and symmetry of the Perron-Frobenius eigenvector. For $N \to \infty$, however, this eigenvector becomes asymptotically degenerate (the difference between the leading and the second-largest eigenvalue is of order $1/N$). Since the eigenvector corresponding to the second-largest eigenvalue is antisymmetric (i.e., its components change sign under reversal of signs in the configuration), any small symmetry breaking (of the landscape, or of the the initial conditions away from $u = 0$) are sufficient to restrict the population to one of the branches. This is a traditional key to the properties of the physical system (Yang (1952); see also Wagner (1998)). At the same time, it provides the justification to ignore the solution branch derived from the second maximum, thus effecting the (biologically reasonable) cutoff of the landscape beyond its minimum.

The interactions underlying this fitness function are those of the well-known Ising model (see the Appendix), more precisely, of the version solved by Onsager (see Thompson (1972)), hence the name of this setting. As with the quadratic landscape, there is no justification for Onsager's landscape at the level of genetic interactions. However, it bears a few properties which may be considered typical of molecular evolution, namely

a) *a large degree of neutrality,* i.e. the presence of configurations with the same fitness (e.g. $+++++--+++$ is selectively equivalent to $+++++++-+++$). For configurations not too far from the peak, the degree of neutrality increases with the distance from the peak, for simple combinatorial reasons.

b) *path structure*: the number of paths uphill dwindles the closer one gets to the peak (e.g. there are many ways uphill from $+-+-+-+-+-$ but only one from $+++++++-+++$).

c) *flat ridges*, i.e., regions from where immediate improvement is only possible after intermediate steps which do not change fitness. For example, the type $+++----+++$ allows deterioration in several ways, e.g. $+-+----+++$ and $+++-+--+++$, but to achieve an improvement, intermediate steps must be accomplished, e.g. $++++---+++$, $+++++--+++$ and $++++++-+++$ before, finally, $++++++++++$ may be reached. There are no strict local maxima, however, that is, climbing downhill is never required to access a higher peak. (The relevance of strict local maxima in molecular evolution is under debate. Judging from models with random assignment of fitness values to genotypes, it seems to be characteristic of high dimensions that there are flat ridges connecting most local peaks, compare Gavrilets (1997). Similarly, strict local maxima have been found to be rare for most combinations of parameters in the so-called multiple quantitative traits model (Taylor & Higgs, 2000), where every site randomly affects a number of traits.)

d) *compensatory mutations:* Fitness is not a monotonic function of the distance from the reference genotype; for example, improvement of $++-+-+++++$ is possible by introducing one further mutation, to arrive at $++---+++++$.

Whereas properties a) and b) are shared with the quadratic landscape, c) and d) rely on the lack of permutation invariance. It goes without saying that Onsager's landscape is an unrealistic toy model and has too high a degree of symmetry. But, just as with the popular quadratic fitness function, it is the symmetries which make exact solutions possible.

Onsager's landscape may be considered as displaying positive epistasis, where we use the following definition (which generalizes the situation for permutation-invariant models): Let $\hat{R}_k$ be the average fitness of types with k '$-$' sites, and the average is taken over all possible configurations (not over a population). In line with the definition in Phillips *et al.* (2000), epistasis is then positive (negative), if $\hat{R}_k$ is convex (concave). With increasing distance from the maximum of Onsager's landscape, a larger fraction of additional mutations either do no further harm, or even act in a compensatory manner. As a consequence, the average fitness clearly decreases less than linearly as a function of distance from the fittest type. This verbal argument will be made precise in Section 6, where we shall come back to the collection of fitness functions and work out solutions for them.

3 Mutation-selection model and quantum chain

In 1986, Leuthäusser put forward an exact equivalence between a *coupled* mutation-selection model in *discrete time*, and the Ising model of *classical* statistical mechanics, but exploiting it proved difficult. This equivalence, along with its limitations and the relationship with the present approach, is discussed in the Appendix. We shall, instead, work with the equivalence between the mutation-reproduction matrix of the *parallel* model in *continuous time*, and the Hamiltonian of a so-called *Ising quantum chain*, which was established by Baake *et al.* (1997). It is related to the classical Ising model described in the Appendix, but now the spins are quantum-mechanical objects. Let us anticipate that the equivalence is a mathematical one, and we shall, at no stage, require the physical concepts of quantum mechanics; as a consequence, we need not worry about its mysteries, either.

First of all, we need some vector space structure. Let us start with $N = 1$, i.e. a single site, which may be $+$ or $-$. The composition of a population may then be represented as a vector in $\mathbb{R}^2$, $\boldsymbol{p} = (p_+, p_-)^T = p_+\boldsymbol{e}_+ + p_-\boldsymbol{e}_-$, where $\boldsymbol{e}_+ := (1,0)^T$ and $\boldsymbol{e}_- = (0,1)^T$ are the canonical unit vectors which belong to the homogeneous (i.e., monomorphic) populations, and T denotes transpose. Since $p_+ + p_- = 1$, $\boldsymbol{p}$ is restricted to the unit simplex. We now introduce single-site *mutation* and *reporting operators*,

$$\sigma^x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \sigma^z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{9}$$

σ^x and σ^z are known as *Pauli's matrices* in quantum mechanics; they flip or measure a spin, respectively. Let us consider their action on the genotype. One verifies immediately that

$$\sigma^x \boldsymbol{e}_\pm = \boldsymbol{e}_\mp, \quad \sigma^z \boldsymbol{e}_\pm = \pm\boldsymbol{e}_\pm. \tag{10}$$

Clearly, σ^x *flips* the site variable, i.e. results in a mutation. The matrix σ^z, on the other hand, does not change the variable but reports its sign, which may be used to assign selective differences: $\eta\mathbf{1} + \eta_1\sigma^z$, $\mathbf{1}$ the identity matrix, assigns fitness $\eta + \eta_1$ to a $+$ site and $\eta - \eta_1$ to a $-$ site. (The subscript 1 refers to the single site considered here, in order to be consistent with the more general indexing to follow.) Thus, the linearized parallel mutation-selection equation for a single site (i.e. for two types) may be written as

$$\dot{\boldsymbol{x}}(t) = (\mathcal{M} + \mathcal{R})\boldsymbol{x} = \big(\mu\sigma^x + \eta_1\sigma^z + (\eta - \mu)\mathbf{1}\big)\boldsymbol{x}(t). \tag{11}$$

Note that only fitness differences ($= 2\eta_1$) will matter for the final, normalized result. The relative strengths of mutation and selection determine the direction of the vector $\boldsymbol{p}$, where μ 'draws' the population into the $45°$ direction, and η_1 draws into the horizontal one; see Fig. 2. This is where the physical picture comes in: a quantum-mechanical spin experiencing the competing action of a transversal field of strength μ in x direction, and a longitudinal field of strength η_1 in z direction (hence the symbols for the Pauli matrices), is described by the energy operator (or *Hamiltonian*) $\mathcal{H} = -(\mu\sigma^x + \eta_1\sigma^z)$, which differs from our $\mathcal{M} + \mathcal{R}$ only by a sign, and a trivial multiple of the identity. Hence, the biological population corresponds to the quantum-mechanical spin. The strength of the transversal field (μ) is known as a *temperature-like parameter* in physics, in the sense that it counteracts the orientation order of the spins with respect to the z direction. This also fits our intuition of mutation as a randomizing force.

Now, for N sites, every type $\boldsymbol{s} = s_1 \dots s_N$ may be identified with a unit vector $\boldsymbol{e}_{\boldsymbol{s}}$ in $\mathbb{R}^{2^N}$:

$$\boldsymbol{e}_{\boldsymbol{s}} := \boldsymbol{e}_{s_1} \otimes \boldsymbol{e}_{s_2} \otimes \dots \otimes \boldsymbol{e}_{s_N}, \tag{12}$$

where '$\otimes$' symbolizes the tensor (or Kronecker) product; readers unfamiliar with this formalism may consult Horn & Johnson (1991, Ch. 4.2) as a very readable introduction. Correspondingly, $\boldsymbol{p}$ and $\boldsymbol{x}$ may be understood as vectors $\boldsymbol{p} = \sum_{\boldsymbol{s}} p_{\boldsymbol{s}}\boldsymbol{e}_{\boldsymbol{s}}$, and $\boldsymbol{x} = \sum_{\boldsymbol{s}} x_{\boldsymbol{s}}\boldsymbol{e}_{\boldsymbol{s}}$. We now require a mutation operator which introduces a mutation at site i while leaving all other sites unchanged. Bearing in mind that tensor products of operators act on the sites independently, the required operator is

$$\sigma_i^x := \underbrace{\mathbf{1} \otimes \dots \otimes \mathbf{1}}_{i-1 \text{ copies}} \otimes \sigma^x \otimes \underbrace{\mathbf{1} \otimes \dots \otimes \mathbf{1}}_{N-i \text{ copies}}, \tag{13}$$

i.e. σ_i^x flips the i'th site and leaves all others unchanged. Hence, we can write down the overall mutation operator as

$$\mathcal{M} = \mu \sum_{i=1}^{N} (\sigma_i^x - \mathbf{1}). \tag{14}$$

Note that, by abuse of notation, we use $\mathbf{1}$ for the identity on $\mathbb{R}^2$ and $\mathbb{R}^{2^N}$ alike, since there is no danger of confusion. It is easily verified that the entries of $\mathcal{M}$ are, indeed, $m_{\boldsymbol{s}'\boldsymbol{s}}$ as given in (6).

Likewise, the operator reporting the sign at site i reads $\sigma_i^z := \mathbf{1} \otimes \ldots \otimes \mathbf{1} \otimes \sigma^z \otimes \mathbf{1} \otimes \ldots \otimes \mathbf{1}$ with σ^z at the i'th position. Let us now construct reproduction operators from single-site reporters. In the absence of epistasis, every site has its individual contribution to fitness, independently of all other sites. Thus, Malthusian fitness is additive, and $R_{\boldsymbol{s}}$ simply reads $R_{\boldsymbol{s}} = \eta + \sum_{i=1}^N \eta_i s_i$, with constants η and η_i. Put differently, the reproduction matrix may be written as

$$\mathcal{R} = \eta \mathbf{1} + \sum_{i=1}^{N} \eta_i \sigma_i^z . \tag{15}$$

This time, the population may be described by a *linear arrangement* of directions (spins), one for each site – hence the name *quantum chain*. Sites are independent of each other; each one experiences the competing action of a longitudinal field of strength η_i, and a transversal field of strength μ, cf. Fig. 2.

If epistasis is included, sites interact. Interaction may involve two, three, or even more sites. It is easy to see that any fitness landscape may be represented in terms of multiple-site interactions, i.e. as

$$R_{\boldsymbol{s}} = \eta + \sum_i \eta_i s_i + \sum_{i<j} \eta_{ij} s_i s_j + \sum_{i<j<k} \eta_{ijk} s_i s_j s_k + (\text{terms up to } N^{\text{th}} \text{ order}) . \tag{16}$$

This is known as the Walsh function expansion (Stadler *et al.*, 2000). Alternatively, the subscripts may be collected into index sets I, and Eq. (16) may be written more compactly as $R_{\boldsymbol{s}} = \sum_I \eta_I \prod_{i \in I} s_i$, where I ranges over all subsets of $\{1, 2, \ldots, N\}$, and $\eta = \eta_\varnothing$.

Let us now recast this into operator (or matrix) form. Interaction between sites $i \in I$ is described by the product $\prod_{i \in I} \sigma_i^z$, which has σ^z's at all sites in I (note that tensor products multiply sitewise; as a consequence, the effect of $\prod_{i \in I} \sigma_i^z$ is to multiply the signs reported for sites in I). Therefore, the operator version of the fitness function (16) reads

$$\mathcal{R} = \sum_I \eta_I \prod_{i \in I} \sigma_i^z . \tag{17}$$

In the quantum chain picture, the new terms represent interactions of the spins with each other, as opposed to interactions with external fields; see Fig. 2.

Summarizing, the mutation-reproduction matrix is equivalent to the Hamiltonian of an Ising quantum chain, with a transversal field which corresponds to mutation, a collection of longitudinal fields which correspond to the additive part of fitness, and arbitrary spin-spin interaction which corresponds to the epistatic part of fitness. More precisely, $\mathcal{H} = -(\mathcal{M} + \mathcal{R})$; therefore, the smallest eigenvalue of the Hamiltonian (the ground-state energy) corresponds to the Perron-Frobenius (PF) eigenvalue of the mutation-reproduction matrix (the equilibrium mean fitness).

Often, however, fitness functions are not, in the first place, defined in terms of such interactions, since these are not known (see our discussion in Section 2). We must then determine the interactions from a fitness function given otherwise. It is easy to see that η_i is the average fitness contribution of site i, averaged over all configurations; η_{ij} gives the average contribution of the correlation between i and j, and, in general,

$$\eta_I = \frac{1}{2^N} \sum_{\boldsymbol{s}} \Big(\prod_{i \in I} s_i \Big) R_{\boldsymbol{s}} . \tag{18}$$

Note that these coefficients may be understood as averages over the configuration space as opposed to averages over the current population; in particular, $\eta = \eta_\varnothing$ is the fitness of a population of random configurations, i.e., the fitness of a population at mutation equilibrium. We shall often choose $\eta_\varnothing = 0$ as a convenient reference point of the fitness scale.

The composition of the fitness function into interaction coefficients according to (18) is closely related to the decomposition into selection coefficients introduced by Barton & Turelli (1991). The latter are based on Wrightian fitness and result from an expansion around any convenient reference point (see Turelli & Barton, 1994), e.g., the current composition of the population, and may therefore change with time. In contrast, our point of expansion is a random population (i.e., equidistributed in the type space, $p_{\boldsymbol{s}} = 1/2^N$), wherefore the η_I are constants.

4 Population averages

With large numbers of types as involved here, one aims at a few averaged quantities rather than the complete distribution of genotypes, like the mutation load, the mean number of mutations, and the variance of these quantities. We shall subsume them under the name *observables* because they are, at least in principle, accessible to experiment, although their measurement may be a formidable task in practice. The *mutation load* is $L(t) := R_{\max} - \bar{R}(t)$, where $R_{\max}$ is the largest fitness value. Its equilibrium value is obtained from the physical picture in a straightforward way, since $\lim_{t\to\infty} \bar{R}(t) = \lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathcal{M} + \mathcal{R}$, and $\lambda_{\max}$ corresponds to the ground state energy of the spin chain (up to a sign). (From now on, we shall suppress the notation of time dependence where there is no danger of confusion, and mention explicitly if equilibrium values, i.e., $t \to \infty$, are referred to.)

Unlike the mean fitness, the mean mutational distance from the reference type has no direct counterpart in the physical picture. The closest relative is the *surplus* (or excess) of '+' sites, $\bar{U} := \sum_{\boldsymbol{s}} U_{\boldsymbol{s}} p_{\boldsymbol{s}}$, where $U_{\boldsymbol{s}} := \sum_{i=1}^N s_i$. $\bar{U}$ corresponds to the magnetization of a *classical* Ising system, but differs from the corresponding quantum-mechanical quantity (for more on this issue, see Section 6.2). We shall use $\bar{U}$ to characterize the genetic composition of a population, but it is, of course, trivially converted into the mean number of mutations: If $+ + + \cdots +$ is the fittest type (which will be assumed without loss of generality), the mean mutational distance is $\bar{D} = (1/2)(N - \bar{U})$.

As to the variability of the population, we shall concentrate on the *variance in fitness*,

$$V_R := \sum_{\boldsymbol{s}} (R_{\boldsymbol{s}} - \bar{R})^2 p_{\boldsymbol{s}} \, ; \tag{19}$$

after all, this is what selection acts on. Before embarking on specific fitness functions, let us state a fairly general result here. Assuming that the fitness landscape is homogeneous in the sense that only n-site interactions occur, i.e.

$$R_{\boldsymbol{s}} = \sum_{|I|=n} \eta_I \prod_{i \in I} s_i \tag{20}$$

for a certain n, one has

$$V_R = 2\mu n \bar{R} \tag{21}$$

at mutation-selection balance. To see this, consider the change in mean fitness under selection and mutation according to (1):

$$\begin{aligned} \frac{d}{dt}\bar{R} = \sum_{\boldsymbol{s}} R_{\boldsymbol{s}} \dot{p}_{\boldsymbol{s}} &= \sum_{\boldsymbol{s}} R_{\boldsymbol{s}} \big((R_{\boldsymbol{s}} - \bar{R}) p_{\boldsymbol{s}} + \sum_{\boldsymbol{s}'} m_{\boldsymbol{s}\boldsymbol{s}'} p_{\boldsymbol{s}'} \big) \\ &= V_R + \sum_{\boldsymbol{s}} R_{\boldsymbol{s}} \sum_{\boldsymbol{s}'} m_{\boldsymbol{s}\boldsymbol{s}'} p_{\boldsymbol{s}'} \,. \end{aligned} \tag{22}$$

Here, the first term represents the increase of mean fitness due to selection, as familiar from Fisher's Fundamental Theorem. The second term is the loss in fitness due to mutation. At equilibrium, therefore, $V_R = -\sum_{\boldsymbol{s}} R_{\boldsymbol{s}} \sum_{\boldsymbol{s}'} m_{\boldsymbol{s}\boldsymbol{s}'} p_{\boldsymbol{s}'}$. To evaluate the right-hand side, we use (16), (6), and (20) to obtain

$$\begin{aligned} \sum_{\boldsymbol{s}} R_{\boldsymbol{s}} m_{\boldsymbol{s}\boldsymbol{s}'} &= \sum_{\boldsymbol{s}} m_{\boldsymbol{s}\boldsymbol{s}'} \sum_{I} \eta_I \prod_{i \in I} s_i \\ &= \mu \Big(\sum_{j=1}^{N} \sum_{I} (-1)^{|\{j\} \cap I|} \eta_I \prod_{i \in I} s_i' \Big) - N\mu \Big(\sum_{I} \eta_I \prod_{i \in I} s_i' \Big) \\ &= -2\mu \sum_{j=1}^{N} \sum_{I \ni j} \eta_I \prod_{i \in I} s_i' = -2n\mu R_{\boldsymbol{s}'}. \end{aligned} \tag{23}$$

The second step reflects the fact that mutation at site j of sequence $\boldsymbol{s}'$, by 'flipping' s_j', changes the sign of those fitness contributions $\eta_I \prod_{i \in I} s_i'$ for which I contains j (consequently, summation in (23) is over all sets I which contain j). Due to homogeneity, every set I may be 'hit' by n different mutations, which is reflected in the last step. With (23), the loss in fitness due to mutation becomes

$$\sum_{\boldsymbol{s}} R_{\boldsymbol{s}} \sum_{\boldsymbol{s}'} m_{\boldsymbol{s}\boldsymbol{s}'} p_{\boldsymbol{s}'} = -2n\mu \bar{R} \,. \tag{24}$$

At equilibrium, the loss in fitness due to mutation must be balanced by the gain in fitness due to selection, hence (21) holds.

The homogeneity assumption holds for a number of fitness landscapes, of which we shall meet representatives shortly. In general, however, it is a restrictive assumption. Nevertheless, it provides some flavour of how genetic variation is created by interaction of sites. (Note that, in the trivial case $n = 0$, one has $R_{\boldsymbol{s}} \equiv \eta$ and $V_R = 0$, i.e. neutral evolution. For $n > 0$, the homogeneity assumption enforces $\eta = 0$, which is no restriction since only fitness differences matter.)

5 Scaling and limits

Often, the limit $N \to \infty$ is employed (and will be used here), for two reasons. First of all, it is a way to make life simpler (and a good one at that, because in genetic applications, 100 is usually closer to infinity than to one), or to make a solution possible at all. But even if a finite-size solution is available, the limit is often required to characterize the behaviour in a clear-cut way (e.g. to produce a sharp transition between two regimes instead of just a very steep one).

The limit may be performed in various ways, and it is crucial to get the scaling of the model parameters right. This is reminiscent of the diffusion approximation (cf. Ewens, 1979) to models of mutation, selection, and drift in finite populations, where the strengths of mutation and selection must scale inversely with population size to ensure that the differential equation is a valid approximation of the finite system. Similarly, in our infinite population model, scaling of mutation rates and fitness values should be chosen such that, for $N \to \infty$, the solution of the differential equation converges to a well-defined limiting form. To be more precise, the solution itself (i.e. the probability distribution of genotypes) cannot be expected to converge, since one can't find a space in which this might be possible. What can converge (at best) is the distribution of quantitative properties of genotypes (such as the surplus per site).

The options for the scaling are *extensive* versus *intensive*, where *extensive* scaling means that quantities like the mutation rate, or fitness, or any of the observables, are constant *per site* (and, hence, proportional to N), whereas *intensive* scaling implies that a quantity is independent of N. A standard approach for mutation-selection models is the *infinite sites limit* $N \to \infty$ under $N\mu = \text{const}$ as introduced by Kimura & Maruyama (1966), where it is further assumed that the deleterious effect of a mutation is independent of N; i.e., mutation is intensive, and fitness is extensive. In this limit, the number of mutated sites per type is approximately Poisson distributed, and the *mutation load* scales *intensively*. For additive fitness, the Poisson distribution is even exact, see the very clear recent article by Johnson (1999), and references therein.

The limit provides meaningful results as long as the mutation rate is so low that only a vanishing fraction of all sites will carry mutations, as is often assumed in population genetics. It breaks down, however, when mutations become so frequent that a sizeable fraction of the sites is affected (which also implies that sites may be hit more than once). Since we set out to study mutation-selection models without any restriction to the mutation rates, we need a limiting procedure which is more generally applicable. For this purpose, we borrow from the physical literature the wisdom that, for systems of statistical mechanics, *extensive* scaling of the Hamiltonian (i.e. of mutation rates and fitness) is required for a meaningful result in the thermodynamic limit; for quantum chain examples, see Lieb *et al.* (1961).

To translate this into the biological context, consider a mapping from our type space into a compact interval I of $\mathbb{R}$ which associates with every element $\boldsymbol{s} \in \{-1,1\}^N$ a value $y_{\boldsymbol{s}} \in I$. Given $\boldsymbol{p}(t)$, the time evolution of the probability distribution on $\{-1,1\}^N$, the mapping $\boldsymbol{s} \mapsto y_{\boldsymbol{s}}$ induces a (time-dependent) probability measure $\omega_N(y,t)$ on I. Let further $f(y)$ be any continuous function on I. We then demand, as our convergence criterion, that $\omega_N(y,t)$ converges towards a limiting measure $\omega(y,t)$ in the weak sense (i.e. convergence in distribution; cf. Bauer 1996), namely $\int f(y) d\omega_N(y,t) \stackrel{N\to\infty}{\longrightarrow} \int f(y) d\omega(y,t)$, with *uniform* (as opposed to pointwise) convergence in t, i.e.

$$\lim_{N\to\infty} \sup_{t\in\mathbb{R}_0^+} \left| \int f(y) d\omega_N(y,t) - \int f(y) d\omega(y,t) \right| = 0. \tag{25}$$

Here, $y_{\boldsymbol{s}}$ and $f(y)$ may be arranged such that $\int f(y) d\omega_N(y,t)$ represents any *intensively-scaled observable* (i.e., the mean fitness, the mean surplus, or the variance in fitness, *per site*). As an example, for the quadratic landscape, one might choose $y_{\boldsymbol{s}} := U_{\boldsymbol{s}}/N$ with values in $[-1,1]$, and $f(y)$ as the mean fitness per site of a sequence with $U_{\boldsymbol{s}}/N = y$. For later use, note that, for permutation-invariant initial conditions, the alternative choice $y_{\boldsymbol{s}} = s_1$ is equivalent. For, due to the symmetry, the population averages of $U_{\boldsymbol{s}}/N$ and s_1 are the same and equal to the mean surplus per site: $\bar{u} := (1/N)\sum_{\boldsymbol{s}} U_{\boldsymbol{s}} p_{\boldsymbol{s}} = \sum_{\boldsymbol{s}} s_1 p_{\boldsymbol{s}}$. For Onsager's landscape, one could choose as $y_{\boldsymbol{s}}$ the fraction of domain walls in the sequence, and as $f(y)$ the mean fitness of a sequence with a fraction y of domain walls. For both landscapes, $\int f(y) d\omega(y,t)$ then gives the time evolution of the mean fitness per site. Eq. (25) then implies that the time evolution of the observable converges to a well-defined limiting form. It can be shown (Wagner, 1998; Wagner *et al.*, 1998) that the only meaningful choice that achieves (25) is extensive scaling of both mutation and fitness, and Eq. (25) indeed holds for the fitness landscapes considered here, provided the initial conditions are compatible with the symmetries of the landscape. However, (25) does not hold for the sharply-peaked landscape, which does not have a reasonable scaling limit – another of its weaknesses.

Using, instead, intensive scaling for both mutation and fitness is equivalent to a rescaling of (continuous) time, i.e. dt is replaced by Ndt, see the discussion in Wagner *et al.* (1998); in the limit, then, evolution becomes infinitely slow. This reflects the fact that, whereas the absolute number of mutational steps per time unit remains constant in this limit, the loss or gain in fitness they bring about vanishes. As a consequence, even though the resulting model has the same equilibrium properties as the all-extensive one, this equilibrium will never be reached. This is often overlooked since only few investigations consider dynamical aspects at all. The time evolution of the mean fitness per site in Onsager's landscape is illustrated in Fig. 5, for extensive and intensive scaling. These figures are modified versions of those given in Wagner (1998) and Wagner *et al.* (1998), where explicit expressions are also derived. Finally, the choice of extensive mutation together with intensive fitness scaling (or vice versa) is not an option either since, in this case, changing N changes the relative strengths of mutation and selection. For $N \to \infty$, the system is dominated solely by mutation, which is not what one aims at with a mutation-selection model.

Convergence of the time evolution according to Eq. (25) is immediately clear for additive fitness and all-extensive scaling, where sites are independent, and neither the surplus nor the mean fitness per site depend on N at all, provided initial conditions are the same at all sites. (If there are any doubts about this argument, they will be removed in the next Section.) But, even for general fitness landscapes, it can be made plausible that the all-extensive scaling is the only one consistent with similarity of the time evolution in systems with different N. To see this, consider the dependence on N of Eq. (22) for the time evolution of the mean fitness, $\bar{R}$. Without mutation, $(d/dt)\bar{R}(t) = V_R(t)$; this elementary version of Fisher's Fundamental Theorem holds for *any given model*, independently of any scaling considerations. But we further demand that $\bar{R}$ and V_R scale in the same way in a *family of models* with different N (this may be considered a refined Fisher Property, which mimicks the Fundamental Theorem for the model family): If both quantities scale with N^α ($\alpha > 0$), then $(d/dt)(\bar{R}/N^\alpha)$ is of order one, i.e., the time scale of $\bar{R}/N^\alpha$ is independent of N, as required for similarity of the models within the family. Since a change of the fitness scale by a factor N^α changes $\bar{R}$ by N^α, but V_R by $N^{2\alpha}$, there is only one scaling compatible with our refined of Fisher Property.

Unfortunately, however, it is difficult to determine the scaling of V_R for a given fitness function. But we can check for consistency with mutation. If mutation is present, one has $\frac{d}{dt}\bar{R} = V_R + \sum_s R_s \sum_{s'} m_{ss'} p_{s'}$, see Eq. (22). Clearly, the second term on the right-hand side (the change in fitness due to mutation) must also have the same scaling as $\bar{R}$, and as V_R. Since the mixed scalings, i.e. extensive mutation and intensive fitness and vice versa, have already been excluded (since they imply a change in the relative strengths of mutation and selection), we have to check if one of the two remaining options is consistent. Eq. (24) shows that intensive scaling of mutation and fitness makes the change in fitness due to mutation scale with $1/N$; but $\bar{R}$ is of order one, hence this scaling is inconsistent. For extensive mutation and extensive fitness, on the other hand, both $\bar{R}$ and $\sum_s R_s \sum_{s'} m_{ss'} p_{s'}$ scale with N. Therefore, this is the only consistent candidate for scaling. It then remains to be checked (possibly numerically) whether convergence of the dynamics actually occurs for a given model; this is the case for all fitness functions considered in this article.

If this scaling is to be interpreted in the biological picture, it implies that μ is constant per site, as anticipated in Eq. (6), and the mutation rate *per genome*, $G := N\mu$, increases linearly with N. As to (Malthusian) fitness, extensive scaling implies that the fitness effects of single mutations remain constant, and the difference between the largest and the smallest fitness value, $R_{\max} - R_{\min}$, increases linearly with N. It is, however, important to stress that it is not required to ascribe any biological meaning to the scaling chosen, as long as we do not aim at comparisons of species (or sequences) with different N. In particular, there is no point in worrying about mean fitness becoming infinite for $N \to \infty$. It must be kept in mind that any real sequence is finite, and the limit is a technical process which aims at an approximation of the finite system by an infinite one with very similar properties. This is much as with the aforementioned diffusion approximation to models of finite populations, where selective differences and mutation rates are required to scale inversely with the population size, although there is no reason to assume that these forces are weaker in larger populations.

Recapitulating, with *extensive scaling* for *mutation* and *fitness*, the *intensively scaled observables*, namely *per site* fitness, surplus, and variance of fitness, will converge in the limit $N \to \infty$. We shall therefore work with the quantities $\bar{r} := \bar{R}/N$, $\ell = L/N$, $\bar{u} := \bar{U}/N$, $\bar{d} := \bar{D}/N$, and $v_R := V_R/N$, as well as the corresponding quantities for $N \to \infty$. (Note that we use capital letters for extensive quantities and small ones for the corresponding intensive ones.)

The surplus $\bar{u}$ may be interpreted as mean agreement with $+++\ldots+$ (which may be chosen to be the fittest sequence without loss of generality), averaged over all sites of the configuration, and all members of the population. Specifically, $\bar{u} = 1$ for a pure 'wildtype' population, and $\bar{u} = 0$ for a population of random configurations.

We have, so far, mainly talked about average quantities. Let us anticipate here that the underlying *distributions* for r (the fitness per site) and u (the surplus per site) will turn out to be delta distributions, i.e., point measures in the limit $N \to \infty$. This is not surprising in view of the following example: Consider N independent sites which read $+$ and $-$ with probability p and $1-p$, respectively. Then, N^+, the number of $+$ sites, is binomially distributed with expectation Np and variance $Np(1-p)$; on the other hand, the *fraction* of $+$ sites, N^+/N, has variance $p(1-p)/N$. Consequently, the distribution of the fraction of $+$ sites converges (in the weak sense) to a delta distribution located at p. Put differently, in the topology underlying the notion of weak convergence (i.e. probability measures $\omega(x)$ and $\omega'(x)$ are different if, for some continuous function $f(x)$, $\int f(x) d\omega(x) \neq \int f(x) d\omega'(x)$, cf. Bauer 1996), the limiting distribution cannot be distinguished from a delta distribution. Hence, no information about the variability of the population may be gained from this limiting distribution alone. However, $(1/N)\mathrm{Var}(N^+) = p(1-p)$ for our binomially-distributed N^+. In the same vein, Eq. (21) provides us with independent information about the variance through $v_R = V_R/N$, the *variance per site* in fitness; note that this is different from the variance in *fitness per site*, which would read V_R/N^2. According to Eq. (21), there is positive variance as long as the mean fitness is positive (for homogeneous landscapes).

6 Worked examples

6.1 Additive fitness

Let us, as a warm-up exercise which does not require any limiting procedure, practice the use of mutation-reproduction matrices with tensor-product structure. To this end, we reconsider the case of additive fitness, as also tackled by O'Brien (1985). Here, $R_k = -ak$, or, equivalently (i.e. up to the irrelevant constant term) and to comply with homogeneity, $R_k = a(N/2 - k)$. In site notation, this may be written as $R_{\boldsymbol{s}} = \alpha \sum_i s_i$ (where $\alpha = a/2$), and thus $\mathcal{R} = \alpha \sum_{i=1}^N \sigma_i^z$. The operator $\sum_{i=1}^N \sigma_i^z$ adds up the variables at the sites, i.e. it assigns the value $U_{\boldsymbol{s}}$ to configuration $\boldsymbol{s}$. Hence, the mutation-reproduction operator may be written as

$$\mathcal{H} = \mathcal{M} + \mathcal{R} = \mu \sum_i (\sigma_i^x - \mathbf{1}) + \alpha \sum_i \sigma_i^z = \sum_i H_i, \tag{26}$$

where $H_i := \mathbf{1} \otimes \ldots \otimes \mathbf{1} \otimes \mathfrak{h} \otimes \mathbf{1} \otimes \ldots \otimes \mathbf{1}$ with the *one-site Hamiltonian*

$$\mathfrak{h} := \mu(\sigma^x - \mathbf{1}) + \alpha\sigma^z \tag{27}$$

in the i'th place. (Here and in what follows, we shall work with the definition $\mathcal{H} = \mathcal{M} + \mathcal{R}$ instead of the physical convention $\mathcal{H} = -(\mathcal{M} + \mathcal{R})$.) The one-site Hamiltonian $\mathfrak{h}$ has eigenvalues $\lambda_{1,2} = \pm\sqrt{\alpha^2 + \mu^2} - \mu$ and corresponding eigenvectors $\boldsymbol{v}_1 = (\mu, -\alpha + \sqrt{\alpha^2 + \mu^2})^T$ and $\boldsymbol{v}_2 = (\mu, -\alpha - \sqrt{\alpha^2 + \mu^2})^T$. Recalling the spectral properties of tensor products (if an $n \times n$ matrix A has eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, and an $m \times m$ matrix B has eigenvalues $\gamma_1, \ldots, \gamma_m$ and corresponding eigenvectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m$, then $A \otimes B$ has eigenvectors $\boldsymbol{v}_i \otimes \boldsymbol{w}_j$ with corresponding eigenvalues $\lambda_i \gamma_j$, $i = 1, \ldots, n$, $j = 1, \ldots, m$), one first observes that H_i has eigenvalues $\lambda_{1,2}$ with multiplicity 2^{N-1} each. Further, since the H_i mutually commute, i.e. $[H_i, H_j] = H_i H_j - H_j H_i = 0$, they may be diagonalized simultaneously, and the spectrum of their sum composed from that of the one-site Hamiltonian. More precisely, $\mathcal{H}$ has eigenvectors $\boldsymbol{v}_{i_1} \otimes \boldsymbol{v}_{i_2} \otimes \ldots \otimes \boldsymbol{v}_{i_N}$ with corresponding eigenvalues $\lambda_{i_1} + \ldots + \lambda_{i_N}$, where i_j is either 1 or 2 for every site j. In particular, the PF eigenvector of $\mathcal{H}$ is $\bigotimes^N \boldsymbol{v}_1$, with eigenvector $\lambda_{\max} = N(\sqrt{\alpha^2 + \mu^2} - \mu)$. (A similar structure pertains in the transfer matrix of the corresponding quasispecies model, as exploited by Rumschitzky (1987).) The corresponding stationary distribution is obtained by replacing $\boldsymbol{v}_1$ by $\tilde{\boldsymbol{v}}_1 := \boldsymbol{v}_1 / \|\boldsymbol{v}_1\|_1$, where $\|.\|_1$ denotes the one-norm. Note that the tensor product structure expresses (statistical) independence of sites (i.e. the sites are in linkage equilibrium), with the components of $\tilde{\boldsymbol{v}}_1$ holding the marginal probabilities of $+$ and $-$ at every site. Hence, at equilibrium, the surplus per site reads

$$\bar{u} = -h + \sqrt{1 + h^2} = \frac{1}{h + \sqrt{1 + h^2}}, \tag{28}$$

and the mean fitness per site is

$$\bar{r} = \lambda_1 = \alpha\bar{u} \tag{29}$$

where we have set $h := \mu/\alpha$. Further, since the landscape is homogeneous with $n = 1$, the variance in fitness (again per site) is $v_R = 2\mu\alpha\bar{u}$, according to Eq. (21). Finally, the mutation load reads $\ell = \alpha - \lambda_1 = \mu + \alpha - \sqrt{\alpha^2 + \mu^2} = \mu - \mu^2/(2\alpha) + O(\mu^3)$. Not surprisingly, due to back mutations, this is less than the value μ from Haldane's Principle.

Even if we only give the equilibrium properties here, it is clear that knowledge of the eigensystem is complete, and allows explicit expressions for the time evolution of the system via (4), given any initial value. Note that, obviously, the solution procedure goes through if sites have different effects on fitness, i.e. $R_{\boldsymbol{s}} = \sum_i \alpha_i s_i$, but we will not spell this out here (the corresponding situation with unidirectional mutation has been tackled by Johnson (1999) in the infinite sites limit).

The observables are illustrated in Fig. 6. Not surprisingly, both fitness and surplus decline smoothly with increasing mutation rate, but never vanish; this is accompanied by a steady increase in variance in fitness. For small μ, the results are in agreement with the infinite sites limit which predicts the number of mutated sites, N^-, is Poisson distributed with expectation (and variance) G/s, where $G = N\mu$ is the overall mutation rate, and $s = 2\alpha$ is the fitness effect per mutation (cf. Johnson, 1999). As a consequence, the mutation load is G, and the variance in fitness equals the variance of the mutation load, namely $s^2 Var(N^-) = Gs$. For larger μ, deviations from the linear increase (in mutation load, and variance) become obvious – they are due to the fact that a sizeable fraction of the sites then carries mutations, so the Poisson approximation to the binomial distribution breaks down.

6.2 Quadratic fitness functions

We are now ready to embark on quadratic fitness landscapes with epistasis. Again, instead of $R_k = -(ak + \frac{c}{2N}k^2)$, fitness may equivalently be expressed as a function of $U_{\boldsymbol{s}}$, i.e. $R_{\boldsymbol{s}} = \alpha U_{\boldsymbol{s}} + \frac{\gamma}{2N}U_{\boldsymbol{s}}^2$, where $\alpha = a/2 - c/4$, $\gamma = -c/4$, and an additive constant is again suppressed. Recalling that $\sum_i \sigma_i^z$ measures $U_{\boldsymbol{s}}$ of configuration $\boldsymbol{s}$, and, hence, $(\sum_i \sigma_i^z)^2 = \sum_{i,j} \sigma_i^z \sigma_j^z$ measures $U_{\boldsymbol{s}}^2$, we obtain the mutation-reproduction matrix

$$\mathcal{H} = \mathcal{M} + \mathcal{R} = \mu \sum_i (\sigma_i^x - \mathbf{1}) + \alpha \sum_i \sigma_i^z + \frac{\gamma}{2N} \sum_{i,j} \sigma_i^z \sigma_j^z \,. \tag{30}$$

For $\gamma \neq 0$, sites are, of course, no longer independent. In fact, (30) shows that every site interacts with every other site, with the same interaction strength $\gamma/(2N)$ – hence the name *mean field* for such a setting in physics. What we need to solve the model is the theory of quantum mean field systems, which is highly developed (Gerisch, 1993; Raggio & Werner, 1989, 1991). We shall see that, in the limit $N \to \infty$, the largest eigenvalue (but not the corresponding eigenvector) may again be obtained through a one-site Hamiltonian, which acts equally and independently at every site. In contrast to the additive case, however, this is anything but obvious, and the theory is fairly subtle, so that we can make it plausible, but cannot prove it rigorously here; we must, instead, refer to the mathematical literature. To be more precise, we shall follow two parallel approaches. The first one uses the quantum-chain formalism and is mathematically rigorous, but must be accepted without proof here. Since it may seem difficult to the uninitiated, we shall use a second route with a more heuristic calculation, which uses elementary tools only and arrives at the same result. Let us set out in the first direction.

a) *Using the quantum statistical formalism*

In the quantum-mechanical formalism, the classical probability measure $\omega(x)$ (with $\int d\omega(x) = 1$), which associates with a real-valued, continuous function $f(x)$ an expectation $\int f(x) d\omega(x)$, is replaced by a so-called *density matrix* ϱ (which is self-adjoint, has nonnegative spectrum, and $\mathrm{tr}(\varrho) = 1$) which associates with an operator $\mathcal{O}$ a real (or complex) expectation $\varphi(\mathcal{O})$ via

$$\varphi(\mathcal{O}) := \mathrm{tr}(\mathcal{O}\varrho). \tag{31}$$

For a very readable introduction, see Balian (1991). Classical (discrete) probability distributions are contained as the special case of diagonal ϱ (the elements representing the probabilities) and diagonal $\mathcal{O}$ (representing the values of the random variable). Weak convergence of a probability measure then generalizes to weak convergence of a sequence φ_N, i.e. $\varphi_N(\mathcal{O}) \stackrel{N\to\infty}{\longrightarrow} \varphi_\infty(\mathcal{O})$ for all $\mathcal{O}$, or

$$\mathrm{tr}(\mathcal{O}\varrho_N) \stackrel{N\to\infty}{\longrightarrow} \mathrm{tr}(\mathcal{O}\varrho_\infty). \tag{32}$$

Here, $\mathcal{O}$ must refer to a fixed set of sites (i.e. independent of N), and be compatible with the symmetries of the system (cf. Section 5); for definiteness, think of $\mathcal{O} = \sigma_1^z$.

To establish the connection with the evolution model, we now identify the density matrix with the (normalized) time-evolution operator,

$$\varrho_N(t) := \frac{\exp(t\mathcal{H}_N)}{\mathrm{tr}\big(\exp(t\mathcal{H}_N)\big)}, \tag{33}$$

where we have renamed $\mathcal{H}$ by $\mathcal{H}_N$ to emphasize the dependence on system size. (Our time variable corresponds to minus the inverse temperature, i.e. $t = -\beta := -1/(kT)$, of the physical system, where T is the absolute temperature, and k is Boltzmann's constant, cf. the Appendix; note the formal similarity with the Boltzmann distribution, Eqs. (76) and (77). Indeed, (32) is its quantum-mechanical generalization, where the Hamiltonian measures the energy of the system. Note that, in our system, we have two notions related to temperature: time corresponds to the *inverse temperature*, whereas the mutation rate corresponds to the *temperature-like parameter* which we have met in (11).

It may now be shown (Gerisch, 1993; Raggio & Werner, 1989, 1991) that (32) holds if ϱ_∞ is chosen to be

$$\varrho_\infty := \otimes^\infty \frac{\exp(t\mathfrak{h})}{\mathrm{tr}\big(\exp(t\mathfrak{h})\big)}. \tag{34}$$

Here, $\otimes^\infty$ is a sloppy but self-explaining shorthand notation indicating that the limiting distribution is a product measure. As before, $\mathfrak{h}$ is the one-site Hamiltonian

$$\mathfrak{h} = \mu\sigma^x + (\alpha + \gamma m)\sigma^z, \tag{35}$$

where the parameter m must be determined such that

$$m(t) = \frac{\mathrm{tr}\big(\sigma^z \exp(t\mathfrak{h})\big)}{\mathrm{tr}\big(\exp(t\mathfrak{h})\big)}. \tag{36}$$

Eq. (36) is the so-called *self-consistency equation*. It determines m, the *quantum-mechanical magnetization*, i.e. the expectation of the operator σ^z (which measures the spin) of a system governed by the Hamiltonian $\mathfrak{h}$. The quantity m relies on a different notion of probability than does $\bar{u}$. As an example, consider a single-site system with PF eigenvector $\boldsymbol{v} = (v_1, v_2)^T$ of the Hamiltonian $\mathfrak{h}$; then, at equilibrium, one has $m = (v_1^2 - v_2^2)/\|\boldsymbol{v}\|_2$ as opposed to $\bar{u} = (v_1 - v_2)/\|\boldsymbol{v}\|_1$, the *classical magnetization*, which we have termed surplus to avoid ambiguities. Geometrically, quantum-mechanical probabilites are represented by unit vectors in the two-norm (with components $v_i/\|\boldsymbol{v}\|_2$), whereas classical probabilities are represented by unit vectors in the one-norm (with components $v_i/\|\boldsymbol{v}\|_1$.) For more on this subject, see Baake *et al.* 1998, and a forthcoming publication.

We now intend to use the above to determine the behaviour of the system at equilibrium, $t \to \infty$. To this end, we need a little interlude to establish that the PF eigenvector is permutation invariant. Indeed, this is true due to the uniqueness of the PF eigenvector, together with the invariance of the Hamiltonian under permutations of sites, and may be seen as follows: Define $\mathcal{P}$ as the operator which symmetrizes any given configuration, i.e. $\mathcal{P} := (1/N!) \sum_{g \in S_N} g$, where S_N is the set of possible permutations of N sites. Clearly, $\mathcal{P}$ is idempotent, i.e. $\mathcal{P}^2 = \mathcal{P}$. Further, since both $\mathcal{M}$ and $\mathcal{R}$ are permutation invariant, one has $\mathcal{M}\mathcal{P} = \mathcal{P}\mathcal{M}$, $\mathcal{R}\mathcal{P} = \mathcal{P}\mathcal{R}$, and hence

$$\exp\big(t(\mathcal{M}+\mathcal{R})\big)\mathcal{P}\boldsymbol{x}(0) = \exp\big(t(\mathcal{M}+\mathcal{R})\big)\mathcal{P}^2\boldsymbol{x}(0) = \mathcal{P}\exp\big(t(\mathcal{M}+\mathcal{R})\big)\mathcal{P}\boldsymbol{x}(0). \tag{37}$$

This reflects the plausible fact that a population, once symmetrized through the action of $\mathcal{P}$ (that is, with $+$ and $-$ distributed evenly across sites), remains symmetric under the time evolution defined by $\exp\big(t(\mathcal{M}+\mathcal{R})\big)$. Now, uniqueness of the PF eigenvector tells us that convergence towards it is independent of the initial conditions (if they are in the positive cone excluding the origin), and will hold for a symmetric initial vector in particular. Since we know that its time evolution will never leave the symmetric sector, the PF eigenvector must be permutation invariant itself. Since we only seek to determine the equilibrium behaviour here, we shall, therefore, assume permutation-invariance for all times to simplify matters.

After this interlude, we now proceed to determine the leading eigenvalue of $\mathcal{H}_N$ (or the mean fitness, respectively) to leading order in N. To this end, first observe that, in the limit $t \to \infty$, ϱ_N from Eq. (33) becomes a projector on the PF eigenvector of $\mathcal{H}_N$. Now, application of $\mathcal{H}_N$ to the projector onto its PF eigenvector yields the largest eigenvalue of $\mathcal{H}_N$ (times the projector). Note further that, due to permutation invariance and up to a term of order $O(N^{-1})\mathbf{1}$, $N^{-1}\mathcal{H}_N$ may be replaced by $\mathcal{O} = \mu(\sigma_1^x - \mathbf{1}) + \alpha\sigma_1^z + (\gamma/2)\sigma_1^z\sigma_2^z$ for all $N \geq 2$. Therefore:

$$\bar{r} = \lim_{t\to\infty}\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}\big(\mathcal{H}_N\varrho_N(t)\big) = \lim_{t\to\infty}\lim_{N\to\infty}\mathrm{tr}\big(\mathcal{O}\varrho_N(t)\big), \tag{38}$$

where the limiting processes $t \to \infty$ and $N \to \infty$ may be interchanged due to the uniform convergence established in (25).

Due to the structure of $\mathcal{O}$ and ϱ_∞, consideration of two sites is sufficient. One calculates

$$\begin{aligned}\lim_{N\to\infty}&\frac{1}{N}\mathrm{tr}\big(\mathcal{O}\varrho_N(t)\big)\\ &= \frac{\mathrm{tr}\Big[\big(\mu(\sigma^x\otimes\mathbf{1}-\mathbf{1}) + \alpha\sigma^z\otimes\mathbf{1} + \frac{\gamma}{2}(\sigma^z\otimes\sigma^z)\big)\big(\exp(t\mathfrak{h})\big)\otimes\big(\exp(t\mathfrak{h})\big)\Big]}{\Big(\mathrm{tr}\big(\exp(t\mathfrak{h})\big)\Big)^2}\end{aligned} \tag{39}$$

with $\mathfrak{h}$ as in (35). The key to the evaluation of this expression now lies in the algebraic properties of Pauli's matrices. It is easily verified that

$$\begin{aligned}(\sigma^x)^2 &= (\sigma^z)^2 = \mathbf{1}\\ \sigma^x\sigma^z &= -\sigma^z\sigma^x,\end{aligned} \tag{40}$$

i.e., these operators are involutions, and they anticommute. We further note that

$$\begin{aligned}\mathrm{tr}(\sigma^x) &= \mathrm{tr}(\sigma^z) = 0, \quad \text{and}\\ \mathrm{tr}(\sigma^x\sigma^z) &= \mathrm{tr}(\sigma^z\sigma^x) = 0.\end{aligned} \tag{41}$$

With the help of (41), one first calculates $\mathfrak{h}^2 = \big(\mu^2 + (\alpha+\gamma m)^2\big)\mathbf{1}$. Then, with the abbreviation $\nu := \sqrt{\mu^2 + (\alpha+\gamma m)^2}$ (ν is the larger eigenvalue of $\mathfrak{h}$), evaluation of the exponential series yields

$$\exp(t\mathfrak{h}) = \cosh(\nu t)\mathbf{1} + \frac{\sinh(\nu t)}{\nu}\big(\mu\sigma^x + (\alpha+\gamma m)\sigma^z\big). \tag{42}$$

With the help of (41), this leads to

$$\operatorname{tr}\big(\exp(t\mathfrak{h})\big) = 2\cosh(\nu t). \tag{43}$$

Plugging (42) and (43) into (39), and observing (41) and (43), one arrives at

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr}\big(\mathcal{H}_N \varrho_N(t)\big) = \frac{\mu^2 + \alpha(\alpha+\gamma m)}{\nu} \tanh(\nu t) + \frac{\gamma(\alpha+\gamma m)^2}{2\nu^2}\big(\tanh(\nu t)\big)^2 - \mu. \tag{44}$$

In the limit $t \to \infty$, this becomes

$$\bar{r} = \nu - \frac{\gamma m(\alpha+\gamma m)}{\nu} + \frac{\gamma(\alpha+\gamma m)^2}{2\nu^2} - \mu, \tag{45}$$

where we have also used the definition of ν in the form $\mu^2/\nu^2 = \nu - (\alpha+\gamma m)^2/\nu$. With similar arguments, the self-consistency equation (36) becomes, for $t \to \infty$,

$$m = \frac{\alpha+\gamma m}{\nu} = \frac{\alpha+\gamma m}{\sqrt{(\alpha+\gamma m)^2 + \mu^2}}. \tag{46}$$

With the above, the mean fitness reads

$$\bar{r} = \nu - \frac{\gamma}{2} m^2 - \mu, \tag{47}$$

where m must be determined from the self-consistency equation (46).

However, (46) may have multiple solutions. As we have discussed already in Section 2, we choose the one derived from the relevant maximum (i.e. located at the wildtype). Since the wildtype is characterized by $m = 1$, the relevant solution is that solution of (46) which is closest to 1 (since no solution is larger than one, this is simply the largest solution).

Recapitulating our derivation, we note that the one-site Hamiltonian only appeared in the expression $\exp(t\mathfrak{h})/\operatorname{tr}\big(\exp(t\mathfrak{h})\big)$. But

$$\exp(t\mathfrak{h})/\operatorname{tr}\big(\exp(t\mathfrak{h})\big) = \exp(t\tilde{\mathfrak{h}})/\operatorname{tr}\big(\exp(t\tilde{\mathfrak{h}})\big) \tag{48}$$

for any $\tilde{\mathfrak{h}}$ which differs from $\mathfrak{h}$ only by a constant term, i.e., $\tilde{\mathfrak{h}} = \mathfrak{h} + c\mathbf{1}$. Let us now observe that $\bar{r}$ differs from ν, the leading eigenvalue of $\mathfrak{h}$, only by the constant term $-(\gamma/2)m^2 - \mu$. We may, therefore, alternatively, work with

$$\tilde{\mathfrak{h}} := \mathfrak{h} - \big(\frac{\gamma}{2} m^2 + \mu\big)\mathbf{1}, \tag{49}$$

of which $\bar{r}$ in (47) is the leading eigenvalue. In physical terms, this corresponds to the choice of the energy scale so that the free energy equals the internal energy of the system at zero temperature (note that $t \to \infty$ corresponds to $T \to 0$), i.e., the entropy vanishes at zero temperature.

Therefore, cutting a long derivation short, the determination of the mean fitness boils down to determining the leading eigenvalue of (49), with m the largest solution of (46). Put differently, the largest eigenvalue of the mutation-reproduction matrix, $\lambda_{\max}(\mathcal{H}_N)$, is approximated by

$$\frac{1}{N}\lambda_{\max}(\mathcal{H}_N) = \frac{1}{N}\lambda_{\max}\Big(\sum_i H_i\Big) + o(1) = \lambda_{\max}(\tilde{\mathfrak{h}}) + o(1). \tag{50}$$

Here, $H_i := \mathbf{1} \otimes \ldots \otimes \mathbf{1} \otimes \tilde{\mathfrak{h}} \otimes \mathbf{1} \otimes \ldots \otimes \mathbf{1}$ with $\tilde{\mathfrak{h}}$ in the i'th place, in analogy with the additive case. In contrast to the latter, however, the relationship (50) is strictly limited to the *largest* eigenvalue (since it relies on ϱ_N projecting onto the PF eigenvector of $\mathcal{H}_N$ for $t \to \infty$) – it does not hold for the remainder of the spectrum. Therefore, we may obtain the mean fitness *at equilibrium* ($t \to \infty$) this way, but not for finite times.

We have, so far, drawn on the physical formalism because it can be made mathematically rigorous, and finally provides an easy-to-use recipe to calculate the largest eigenvalue, i.e., the equilibrium mean fitness, to leading order. For those who find the derivation difficult, we shall use another route to arrive at the same result in a less rigorous, more heuristic manner, but with the help of elementary tools only.

b) *Using the Rayleigh coefficient*

If the largest (in modulus) eigenvalue of a matrix A is real (this is the case for our $\mathcal{H}$, since the Perron-Frobenius theorem holds), it may be determined via the Rayleigh coefficient:

$$\lambda_{\max}(A) = \max \frac{(\boldsymbol{z}, A\boldsymbol{z})}{(\boldsymbol{z}, \boldsymbol{z})}. \tag{51}$$

Here, $(.,.)$ denotes the scalar product, and the maximum is taken over all vectors in the space. Equivalently, $\lambda_{\max}$ is obtained by maximizing $(\boldsymbol{z}, A\boldsymbol{z})$ over all vectors $\boldsymbol{z}$ with $\|\boldsymbol{z}\|_2 = \sqrt{(\boldsymbol{z}, \boldsymbol{z})} = 1$, which comes down to the eigenvalue equation, and normalization:

$$A\boldsymbol{z} = \lambda_{\max}\boldsymbol{z} \quad \text{subject to} \quad (\boldsymbol{z}, \boldsymbol{z}) = 1. \tag{52}$$

Of course, (52) is a large system of 2^N equations; but, due to the permutation invariance of the PF eigenvector, it may be reduced to $N + 1$ equations: it is sufficient to specify the number of + sites, regardless of their position. This is a common approach in population genetics, see Barton & Shpak (2000) for a recent application. Actually, one often starts out with this reduced representation right away. We shall see, however, that reference to the larger space is essential for finding the leading eigenvalue.

Since special emphasis must be put on normalization, we shall perform all steps explicitly, starting out from the original variables $x_{\boldsymbol{s}}$. We shall first lump together all configurations with k '+' sites, irrespective of their position, i.e. $y_k := \sum_{\boldsymbol{s}:N_{\boldsymbol{s}}^+=k} x_{\boldsymbol{s}}$, where $N_{\boldsymbol{s}}^+$ again is the number of '+' sites in configuration $\boldsymbol{s}$. This corresponds to the change in representation

$$\boldsymbol{x} = \sum_{\boldsymbol{s}} x_{\boldsymbol{s}} \boldsymbol{e}_{\boldsymbol{s}} = \sum_{k=0}^{N} y_k \boldsymbol{b}_k, \tag{53}$$

where

$$\boldsymbol{b}_k := \binom{N}{k}^{-1} \sum_{\boldsymbol{s}:N_{\boldsymbol{s}}^+=k} \boldsymbol{e}_{\boldsymbol{s}}. \tag{54}$$

$\boldsymbol{b}_k$ is the average over all basis vectors with k '+' sites, of which there are $\binom{N}{k}$. Like the original $\boldsymbol{e}_{\boldsymbol{s}}$, the $\boldsymbol{b}_k$ are unit vectors in the one-norm. Likewise, since $\sum_k y_k = \sum_{\boldsymbol{s}} x_{\boldsymbol{s}} = 1$, it also follows that $\|\boldsymbol{y}\|_1 = \|\boldsymbol{x}\|_1 = 1$. In this representation, the eigenvalue equation (52) reads

$$(m_{k,k} + R_k) y_k + m_{k,k-1} y_{k-1} + m_{k,k+1} y_{k+1} = \lambda_{\max} y_k, \tag{55}$$

where $m_{k,k} = -N\mu$, $m_{k,k-1} = \mu(N+1-k)$, $m_{k,k+1} = \mu(k+1)$, and $R_k = \alpha(N-2k) + \frac{\gamma}{2N}(N-2k)^2$ is the fitness of a type with $N_{\boldsymbol{s}}^+ = k$. These coefficients may be considered as the elements of a tridiagonal matrix (where $m_{k,\ell} = 0$ for $\ell \neq k-1, k, k+1$ is implied).

In order to keep track of the normalization condition in (52), we now change the normalization of the $\boldsymbol{b}_k$ so that they become unit vectors in the two-norm instead of in the one-norm. To this end, we observe that $(\boldsymbol{b}_k, \boldsymbol{b}_k) = \binom{N}{k}^{-1}$. We therefore choose $\boldsymbol{c}_k := \binom{N}{k}^{1/2} \boldsymbol{b}_k$, $k = 0, \dots, N$, as our orthonormal basis (i.e. $(\boldsymbol{c}_k, \boldsymbol{c}_j) = \delta_{k,j}$). With $z_k := \binom{N}{k}^{-1/2} y_k$, one then has $\boldsymbol{x} = \sum_{k=0}^N y_k \boldsymbol{b}_k = \sum_{k=0}^N z_k \boldsymbol{c}_k$. Performing the corresponding similarity transform for the elements of the tridiagonal matrix in (55) (i.e. $R_k \to R_k$, and $m_{kj} \to \binom{N}{k}^{1/2} m_{kj} \binom{N}{j}^{-1/2}$) leads to the modified eigenvalue equation in the new basis

$$\begin{aligned} \mu\sqrt{k(N+1-k)}z_{k-1} + \mu\sqrt{(k+1)(N-k)}z_{k+1} + \big(\alpha(N-2k) + \frac{\gamma}{2N}(N-2k)^2 - N\mu\big)z_k \\ = \lambda_{\max} z_k, \end{aligned} \tag{56}$$

where $\lambda_{\max}$ is extensive in N. Replacing k/N by ζ, z_k by $f(\zeta)$, expanding to leading order in N, and passing to intensive scaling, one obtains

$$\big(-\mu + 2\mu\sqrt{\zeta(1-\zeta)} + \alpha(1-2\zeta) + \frac{\gamma}{2}(1-2\zeta)^2\big)f(\zeta) + o(1) = \frac{\lambda_{\max}}{N} f(\zeta). \tag{57}$$

Replacing $\lambda_{\max}/N$ by $\bar{r}$ and taking the limit $N \to \infty$, we see that $f(\zeta)$ can only differ from zero if

$$\bar{r} = -\mu + 2\mu\sqrt{\zeta(1-\zeta)} + \alpha(1-2\zeta) + \frac{\gamma}{2}(1-2\zeta)^2. \tag{58}$$

Defining $m := 1-2\zeta$ and demanding that the right-hand side be independent of m (and, hence, a stationary point of $(\boldsymbol{z}, A\boldsymbol{z})$ in (51)), one finally arrives at

$$\frac{\partial}{\partial m}(-\mu + \mu\sqrt{1-m^2} + \alpha m + \frac{\gamma}{2}m^2) = -\frac{\mu\, m}{\sqrt{1-m^2}} + \gamma m + \alpha = 0, \tag{59}$$

which an elementary calculation shows to be equivalent to the self-consistency equation (46).

Eq. (46) is quartic and can, in principle, be solved for all parameter values. But even if one resorts to approximate or numerical solutions, it is a vast improvement over the original problem, since the eigenvalue problem of a very large matrix boils down to solving this one and only equation. In special cases, (46) even reduces to a quadratic equation. To be specific, let us consider $\alpha = 0$, $\gamma > 0$ (for this choice, the fitness function has its minimum at the mutation equilibrium). Then, (46) may be rewritten as $m^2(m^2+h^2) = m^2$ (where we have used $h := \mu/\gamma$), with the solutions $m_0 = 0$, and $m_\pm = \pm\sqrt{1-h^2}$. For $h < 1$, m_+ is the largest (and, therefore, relevant) solution; at $h = 1$, all solution branches coalesce, and for $h > 1$, $m_0 = 0$ is the only one left. Thus,

$$m = \begin{cases} \sqrt{1-h^2}, & h \le 1 \\ 0, & h > 1 \end{cases} \tag{60}$$

in agreement with (46). This leads to

$$\bar{r} = \begin{cases} \frac{\gamma}{2}(1-h)^2, & h \le 1 \\ 0, & h > 1, \end{cases} \tag{61}$$

in accordance with (47). Since the landscape is homogeneous with $k = 2$ (up to terms of order $1/N$ which do not contribute in the limit), we can also write down the variance of fitness immediately (cf. 21):

$$v_R = \begin{cases} 2h\gamma^2(1-h)^2, & h \le 1 \\ 0, & h > 1. \end{cases} \tag{62}$$

So we have finally managed to calculate our observables.

So far, we have seen that the quantum chain formalism provides a powerful method to determine the leading eigenvalue of the mutation-reproduction matrix, and can be made plausible by more elementary, but less rigorous methods. Finding that eigenvalue is the primary problem, and once the answer is known, it may be used to determine the distribution of genotypes from Eq. (55), i.e., in terms of our original variables. We therefore put $u := (2k - N)/N$ and $\omega_N(u) := y_k$ in Eq. (55), pass to intensive scaling and let $N \to \infty$ to obtain

$$\frac{\gamma}{2} u^2 \omega(u) = \bar{r}\, \omega(u). \tag{63}$$

With $\bar{r}$ as obtained in Eq. (61), it follows that

$$\omega(u) = \begin{cases} \delta\big(u - (1-h)\big), & h \le 1 \\ \delta(u), & h > 1. \end{cases} \tag{64}$$

This is the limiting probability measure discussed in Section 5. As was noted there, the fact that a delta peak appears here (instead of a distribution with positive variance) is a consequence of the 'crude' $1/N$ scaling and weak convergence (actually, the calculation employed here is heuristic, but weak convergence was shown explicitly in a somewhat more involved derivation (Wagner *et al.*, 1998)). This scaling omits some details, but suffices to determine the expectation of u. Clearly, an inspection of Eq. (64) gives the mean surplus at equilibrium:

$$\bar{u} = \begin{cases} 1 - h, & h \le 1 \\ 0, & h > 1. \end{cases} \tag{65}$$

The quantities $\bar{r}$, $\bar{u}$ and v_R are displayed in Fig. 6 as a function of h. Both fitness and surplus decrease with increasing mutation rate, whereas the variance in fitness first increases and then *decreases* again. The decline indicates that selection becomes less and less efficient and finally ceases to operate when v_R vanishes; at the same time, $\bar{r}$ and $\bar{u}$ vanish. A population at *mutation* equilibrium also has $\bar{r} = v_R = \bar{u} = 0$. (Note that, although there are fit genotypes and fitness differences present in a population of random types, they have vanishing impact on $\bar{r}$ and v_R in the limit $N \to \infty$, since the distribution is clustered very narrowly around $\bar{u} = 0$ even for finite N – hence, $\bar{R} = o(N)$ and $V_R = o(N)$). One may therefore speak of an *error threshold* here: Beyond a critical mutation rate ($\mu > \gamma$), mutation can no longer be counteracted by selection, and the population loses its genetic structure in the sense that it cannot be distinguished from a random population, at least not on the basis of the observables considered here.

However, we must be more precise here. According to the usual definition, an error threshold is identified with the mutation rate at which the fittest type is lost from the population. However, this criterion is taylored to the assumption of a sharply-peaked landscape and need not give meaningful answers in other cases. In the current example, with a large number of closely-spaced fitness values, the fittest type may be exceedingly rare even at small mutation rates, although the population is very near to its fitness optimum. It is seen from Fig. 1 that, in contrast to the sharply-peaked landscape, the distribution of types moves steadily in the direction of more mutations, until a conspicuous change happens at $\mu = \gamma$. We therefore require a more generally applicable definition of error thresholds, and tentatively use the existence of a nonanalytic point of the mean fitness as a function of μ; this is also used in physics to define the analogous phenomenon of a phase transition. In the case just treated, the discontinuous second derivative of $\bar{r}$ at $h = 1$ is obvious, but we shall see a more subtle example below. In any case, a nonanalyticity (i.e., a point at which the function cannot be expanded into a Taylor series) means that extrapolation is impossible, and 'something qualitatively new happens' beyond this point. Some reference to the fittest type remains: Our surplus (which plays a similar role as an *order parameter* in physics) measures the mean agreement with the fittest genotype, averaged over all sites, and all members of the population, and vanishes at the error threshold.

We have, so far, dealt with the completely symmetric fitness landscape with positive epistasis, i.e., $\alpha = 0$, $\gamma > 0$, for which everything may be written down analytically. For other parameters, approximate solutions are more instructive than the exact ones. Let us calculate the mean fitness to second order in the mutation rate, for the general quadratic fitness function. With $m = m(\mu)$, one obtains through implicit differentiation of (46) (and with a little help from MATHEMATICA (Wolfram, 1996)):

$$m'(\mu) = 0 \quad \text{and} \quad m''(\mu) = -\frac{1}{(\alpha+\gamma)^2}. \tag{66}$$

Using this to expand $\bar{r}$ from Eq. (47) to second order in μ yields for the mean fitness

$$\bar{r} = \alpha + \frac{\gamma}{2} - \mu + \frac{\mu^2}{2(\alpha+\gamma)} + O(\mu^3). \tag{67}$$

Since the reference genotype has fitness $r_{\max} = \alpha + \gamma/2$, one obtains for the mutation load

$$\ell = \mu - \frac{\mu^2}{2(\alpha+\gamma)} + O(\mu^3). \tag{68}$$

Clearly, positive epistasis ($\gamma > 0$) reduces the mutation load, whereas negative epistasis ($\gamma < 0$) increases it relative to additive fitness ($\gamma = 0$). To make sure that this is not an artifact of variation in the range of fitness values, let $\alpha + \gamma/2 = c$. Then, $r_1 \equiv c$ and $r_0 \equiv 0$, where r_0 (r_1) refer to the fitness per site of a type with $u = 1$ ($u = 0$). Then, $\ell = \mu - \mu^2/(2(c+\gamma/2)) + O(\mu^3)$, and the interpretation remains unchanged.

But we can also extract the qualitative picture for the general quadratic fitness function from a simple graphical argument based on the self-consistency equation, even without approximation. Let $\alpha, \gamma \neq 0$, and write the self-consistency equation (46) in the form

$$\frac{m}{\sqrt{1-m^2}} = \frac{\alpha + \gamma m}{\mu}, \tag{69}$$

or, with $m := \sin\varphi$,

$$\tan\varphi = \frac{\alpha + \gamma \sin\varphi}{\mu}. \tag{70}$$

The 'rightmost' intersection of the graphs of the left-hand and right-hand sides of (70) defines the relevant equilibrium of the model. This is depicted in Fig. 7 for various α, γ, and μ (but only scenarios of directional selection are considered). An error threshold occurs if this intersection vanishes, i.e., if the intersection on the positive branch of $\tan\varphi$ ceases to exist at some large enough μ (the intersection on the negative branch may be safely ignored, since it belongs to the second, artificial peak of the fitness function). The Figure demonstrates that this requires positive (i.e. diminishing returns) epistasis ($\gamma > 0$), but additionally, $-\gamma \leq \alpha < 0$. In contrast, for unidirectional mutation, error thresholds exist whenever $\gamma > 0$, independently of α (Charlesworth, 1990). Geometrically, in both cases the condition means that the fitness landscape attains its minimum (i.e., the fitness gradient vanishes) at positive surplus values, that is, *before* the mutation equilibrium, which is located at $\bar{u} = 0$; cf. Fig. 7. In the third panel ($0 < \alpha < \gamma$), one has positive epistasis with the fitness minimum situated at $\bar{u} < 0$, and no error threshold. For completeness, let us mention that there is no error threshold in the case of stabilizing selection either, where the fitness maximum is at a surplus value between zero and one (this is the case $\gamma < -\alpha < 0$, which we excluded from the Figure since we have been concentrating on directional selection throughout the article).

We have, and will continue to, restrict ourselves to the equilibrium behaviour here, although we have chosen the setup (in particular the scaling) carefully to also allow the analysis of the time evolution of the model. But this is a separate issue and will be deferred to a future publication.

6.3 Onsager's landscape

As suggested by the verbal description in Section 2.1, the fitness of configuration $\boldsymbol{s}$ under Onsager's landscape is (again, up to an irrelevant constant term) $R_{\boldsymbol{s}} = \gamma(N - 2w) = \gamma \sum_k s_k s_{k+1}$, where $\gamma > 0$, and w is the number of domain walls in the configuration. We assume cyclic boundary conditions for convenience, i.e. $s_{N+1} = s_1$ (other boundary conditions would not change the qualitative behaviour of the solution – so we may go for the simplest case). This corresponds to a spin system with nearest-neighbour interactions as first solved by Onsager, hence the name of this setting. However, it should be noted that the neighbouring relation need not be limited to spatial neighbours; we might, as well, perform any renumbering of sites, thus aiming at interactions of 'logical' neighbours rather than spatial ones.

In Section 2, we have listed a few qualitative properties of this fitness landscape. In particular, it was intuitively clear that there is positive epistasis, since the fraction of compensatory mutations increases with the distance from the peak. We shall make this more precise now by calculating $\hat{r}(u)$, the average fitness of a configuration with surplus u, or $N^+ = N(u+1)/2$ '+' sites. In a random configuration with surplus u, any site reads '+' with probability $p = (1+u)/2$, and '−' with probability $q = (1-u)/2$. Thus, between any pair of neighbouring sites, there is a domain wall with probability $2pq = (1+u)(1-u)/2$, and the expected number of domain walls in the configuration is $w = Npq = N(1-u^2)/2$ (of course, overlapping pairs of sites are not independent, but recall that expectations are additive for sums of random variables, even if these random variables fail to be independent, cf. Feller 1968, Chap. IX). Therefore,

$$\hat{r}(u) = \gamma u^2. \tag{71}$$

That is, as far as the *average* fitness of types with surplus $\boldsymbol{s}$ is concerned, Onsager's landscape coincides with the quadratic fitness function with $\alpha = 0$ and $\gamma > 0$ (up to a factor of 2). But we shall see in a moment that the behaviour is *not* governed by $\hat{r}(u)$ alone. Let us further remark that the whole *distribution* of the number of domain walls is easily calculated, since the problem translates directly into the statistics of *runs* in sequences of Bernoulli trials, which is very well studied (cf. Feller, 1968, Chap. II), but we will not pursue this here.

With $R_{\boldsymbol{s}} = \gamma \sum_k s_k s_{k+1}$, the mutation-reproduction operator for Onsager's landscape reads

$$\mathcal{H} = \mathcal{M} + \mathcal{R} = \mu \sum_i (\sigma_i^x - \mathbf{1}) + \gamma \sum_i \sigma_i^z \sigma_{i+1}^z \,. \tag{72}$$

To solve this model, we now change our strategy. For the previous landscapes, we have performed a line-by-line derivation, since no literature on the method is available that is accessible to the typical theoretical biologist. For Onsager's landscape, however, the solution may be followed in the famous paper by Lieb *et al.* (1961), and there is also a very helpful text available (Thompson, 1972). Adaptations of this method to the biological situation can be found in Wagner *et al.* (1998). To avoid duplication, we shall therefore not dwell on the derivation here, but only summarize the results.

Starting from explicit expressions for eigenvalues and eigenvectors of $\mathcal{H}$ for finite N, the limit $N \to \infty$ is performed and yields two symmetric branches of equilibria, just as with the corresponding quadratic landscape. At $\mu = \gamma$, the branches coalesce, after which there is only one branch left to survive. As a function of $h := \mu/\gamma$, the ground state energy (i.e. mean fitness per site) is derived to read

$$\bar{r} = \frac{2\gamma}{\pi}(1+h)E\big(\frac{\pi}{2}, \theta\big) - \mu, \tag{73}$$

where

$$\theta^2 := \frac{4h}{(1+h)^2}, \quad \text{and} \quad E\big(\frac{\pi}{2}, \theta\big) := \int_0^{\pi/2} \sqrt{1 - \theta^2 \sin^2(\psi)}\, d\psi \tag{74}$$

is the complete elliptic integral of the second kind, cf. Abramowitz & Stegun (1970). Since the landscape is, again, homogeneous (two-site interactions only), one also has $v_R = 4\mu\bar{r}$. Although this is not so obvious, $\bar{r}$ (and, consequently, v_R) both have a nonanalytic point at $h = 1$ (the second derivative is discontinuous, i.e., the phase transition is second order in the fitness variable).

As in the case of the quadratic landscape, the surplus does not correspond to any physical property and must be calculated separately. This was done in Wagner *et al.* (1998) and results in

$$\bar{u} = \begin{cases} \sqrt{1-h}, & h \le 1 \\ 0, & h > 1, \end{cases} \tag{75}$$

for the positive branch of the solution. Clearly, at $h = 1$, the first derivative is discontinuous (i.e., the phase transition is first order with respect to the surplus); hence, the nonanalyticity is more readily diagnosed from this variable than from the mean fitness alone.

The results are depicted in Fig. 6. For small mutation rates, the behaviour of mean and variance in fitness is very similar to the corresponding permutation-invariant model. For larger mutation rates, however, there are obvious differences. The surplus precipitates to zero, and a phase transition occurs in the sense that the mean fitness displays a nonanalyticity. But there is still appreciable fitness left, and the variance remains very large. This reflects the fact that fitness is not a function of the number of deleterious mutations when compensatory mutations are present: Here, we have very fit configurations $(+ + + \cdots + - - - - \cdots -$ as an extreme example) whose distance from the fittest sequence is close to that of a random sequence, and, at high mutation rates, they will be present in appreciable amounts.

Recall that we have, in the 'scaling and limits' section, used this system to illustrate the time evolution in the limit $N \to \infty$. Let us now come back to this point and refer again to Fig. 5, where the time evolution of the mean fitness is depicted for two different (extreme) initial conditions, namely a purely wildtype and a fully randomized population. Clearly, convergence to equilibrium occurs within an evolutionarily reasonable time span. Let us, however, mention that other observables, which we have not discussed here, may show a qualitatively different behaviour. The time evolution of the *variance of the surplus*, for example, shows what is called a 'critical slowing down': In the vicinity of the critical point, convergence towards the equilibrium becomes infinitely slow (in genotype space), and the equilibrium *distribution* of *genotypes* will never be reached, although the mean fitness has long become stationary. We only mention this here to hint at the complicated rearrangements that may go on as a consequence of compensatory mutations; for more details, see Wagner *et al.* (1998).

7 Discussion

There are obvious parallels between the topics of population genetics and statistical physics: Both deal with a huge number of possible states, and there is no way to observe them all. One is therefore interested in a few statistical properties rather than all the details. Since the mathematical tools of statistical physics are highly developed, it is an obvious idea to try and apply such methods to solve problems from population genetics.

However, a one-to-one correspondence between biological and physical models is rare, the differences ranging from subtle to fundamental. In the realm of mutation-selection models, the equivalence with a classical two-dimensional Ising model had been known for fifteen years (Leuthäusser, 1986), but the interactions arising in the biological model (long-range within rows, nearest neighbour between rows) proved so different from those in typical physical models that progress along these lines was limited. The quantum chain picture overcomes this problem, but now care must be exercised to distinguish carefully between quantum-mechanical and classical quantities; in particular, the surplus differs from the quantum-mechanical magnetization in a fundamental way. Apart from this, the translation is one-to-one, and the dictionary reads

fitness	$\leftrightarrow$	energy
additive part of fitness	$\leftrightarrow$	interaction with longitudinal field
epistatic part of fitness	$\leftrightarrow$	within-chain interaction
equilibrium mean fitness	$\leftrightarrow$	ground state energy
mutation	$\leftrightarrow$	interaction with transversal field
mutation rate	$\leftrightarrow$	temperature-like parameter
time	$\leftrightarrow$	inverse temperature
error threshold	$\leftrightarrow$	phase transition

The main benefit of the equivalence is that it makes available the toolbox of quantum statistical mechanics to calculate the leading eigenvalue of the mutation-reproduction matrix, i.e. the mean fitness. This, in turn, gives access to the corresponding eigenvector and the surplus. Much more directly, one obtains the mutation load. This is nontrivial as soon as one leaves the territory where Haldane's Principle holds (i.e., the mutation load in *haploids* (or diploids *without dominance*) equals the mutation rate if mutation is *unidirectional – independently of the fitness function*, in particular, independently of epistasis). Deviations from Haldane's Principle have been investigated in situations with dominance and various schemes of sexual reproduction but still with unidirectional mutation; the overall picture is that positive epistasis enhances the mutation load, whereas negative epistasis alleviates it (Kimura & Maruyama, 1966; Charlesworth, 1990; Higgs, 1994). We have explored the other direction here, with symmetric mutation, but still with haploid, asexual populations. Not surprisingly, symmetric mutation by itself diminishes the mutation load. For quadratic fitness functions with epistasis, we again find that positive (negative) epistasis enhances (decreases) the mutation load relative to this zero-epistasis reference case. The same holds true for Onsager's landscape, which exhibits compensatory mutations but positive epistasis, too – cf. Fig. 6.

A phenomenon closely related to mutation loads is the error threshold, which, so far, has no clear-cut definition, but which we have tentatively identified with a nonanalytic point of the mean fitness. This subsumes various types of behaviour. For example, in the quadratic landscape with $\alpha = 0$, selection ceases to operate beyond a critical mutation rate; in contrast, Onsager's landscape displays a phase transition which is not obvious at the fitness level, but all the more conspicuous at the genetic level. In both cases, there is a nonanalytic point in the limit $N \to \infty$, which marks a qualitative change in the behaviour of the system. The existence of error thresholds has been highly controversial, and has been discussed from various points of views. The first question is whether such a phenomenon is observed in experiments. Answers require fitness measurements at various (artificially elevated) mutation rates, which are practically unfeasible for organisms larger than bacteria. In viruses with small RNA genomes, extinction of the population is observed at increased levels of mutagenesis (Holland *et al.*, 1990; Domingo & Holland, 1997), which is taken as indicative of an error threshold. However, the demonstration of a sharp transition would require much more precise measurements at very closely spaced mutation rates. Inclusion of the genetic level, maybe through DNA hybridization studies with the wildtype, might help to clarify the picture, since this would yield a measurement corresponding to our surplus. From the theoretical point of view, the main question is which fitness landscapes have error thresholds. The examples investigated in Charlesworth (1990), Wiehe (1997), and the present study demonstrate that error thresholds are not restricted to the sharply-peaked landscape. They further allow the speculation that positive epistasis is a necessary, but not sufficient condition for the existence of an error threshold (we have seen in Fig. 7 that a further requirement is that the (local) fitness minimum be located at a positive surplus value). Although the models are admittedly rather special, there is no reason to believe that the result is atypical; this is due to what is known as universality in physics (Kadanoff, 1976; Cardy, 1987).

Of course, the last but hardest question is: What are biologically realistic fitness landscapes? The classical mutation-selection experiments (Mukai *et al.*, 1972) yield data for $\hat{R}_k$, the chromosome-wide average of the fitness of a genotype with k mutations. For the small range of k observed, the data are compatible with a quadratic, permutation-invariant fitness function R_k with positive epistasis. But, as pointed out in Phillips *et al.* (2000), there is no reason to believe that this function models the local interactions correctly. On the contrary, detailed mutation experiments have revealed all kinds of epistatic interactions *locally*, without a general picture emerging so far (Whitlock *et al.*, 1995), and many of the interactions tend to cancel each other when it comes to averages (Elena & Lenski, 1997). Further, it may be argued on theoretical grounds that all kinds of interactions may exist; for example, metabolic control theory is used by Phillips *et al.* (2000) to demonstrate that, with a single gene coding for an enzyme, positive and negative epistasis are equally plausible to occur, depending on the state of the remainder of the metabolic network. The importance of individual interactions is corroborated by our results on Onsager's landscape, which has the same *average* fitness function $\hat{r}(u)$ as a quadratic one, but displays very different behaviour.

But let us return to the *Drosophila* data. Of course, one crucial question concerns $\hat{R}_k$ for larger k than those measured. It is hard to conceive that the fitness function is concave throughout. In a forthcoming paper, it will be shown that even convex *regions* of the fitness function may lead to threshold behaviour. So the case is not yet closed — not even for *Drosophila*.

The limit $N \to \infty$ was a crucial tool to simplify the solution of the models, and to extract the characteristic behaviour in an explicit manner. For finite N, there cannot be a nonanalytic point, which we used to define error thresholds. In a large but finite system with a quadratic fitness function, for example, the second derivative of the mean fitness performs a steep but smooth transition from γ to 0, which turns into a jump for $N \to \infty$, cf. (61). To be conclusive, the limit must be performed in a way which guarantees maximum similarity between the behaviour of the finite (biological) system and the infinitely large (mathematical) one, as defined in (25). We have seen that this requires extensive scaling of both mutation and fitness, and gives access to observables like the mean and variance in fitness, or the fraction of mutated sites, which involve simple averages over all sites, without reference to the spatial structure within the configuration. If 'spatial' aspects (like the size of domains) were to be investigated, a very different approach would be required.

Readers with a physical background may have wondered why we have not even mentioned critical exponents. Critical exponents give the power laws of the decay of a quantity in the vicinity of the critical point; for example, the surplus has exponent 1 in the quadratic landscape with $\alpha = 0$ and $\gamma = 1$, but exponent $1/2$ in Onsager's landscape. These quantities are measureable with high precision in physics and convey crucial information about the nature of the phase transition. As we have discussed above, however, the location of the error threshold (if any) is only measureable with an enormous effort (if at all), and precise characterization of the equilibrium for mutation rates in the vicinity of the threshold are virtually impossible. Also, finite population size would render such observations meaningless, since it shifts error thresholds to smaller mutation rates, and changes the nature of the transition. For small mutation rates, on the other hand, the behaviour is often very similar to that of the infinite population, see Wiehe *et al.* (1995). For these reasons, critical exponents seem to be less meaningful from the biological point of view, and the parameter regions far from the critical point are at least as interesting.

To conclude, this paper has been mainly concerned with the development of novel methods for the analysis of mutation-selection balance. We have illustrated these methods by means of selected examples, which included compensatory mutations in one case. It is worth mentioning that these techniques can also be applied to mutation-selection models with more than two states per site (Hermisson *et al.*, 2000). In a forthcoming paper, we will apply our methods to the general class of permutation-invariant landscapes and draw more general biological conclusions.

8 Appendix: Ising's model

(Ferro-)magnetic materials, like iron, may be in either of two states (or *phases*), an ordered one (at low temperature) and a disordered one (at high temperature)[1]. Which of them is present depends on the interplay of two 'forces'. The interaction between the particles tends to align the magnetic moments, whereas the thermal movement has a randomizing effect. At a critical temperature, the latter becomes so strong that the interaction between the particles is effectively overcome, and a phase transition occurs from the ordered into the disordered phase.

The classical model for the description of such phenomena is the *Ising model*. For readers unfamiliar with the matter, we recommend Thompson (1972) as a really lucid text. According to Ising's model, particles are considered as localized at the vertices of a rectangular lattice patch, and their magnetic moments (or *spins*) may only assume two directions, which may be

[1] More precisely, the phases in question have or do not have spontaneous magnetization.

identified with 'north' and 'south', or $+1$ and -1, respectively. For a patch with N columns and M rows, there are thus 2^{MN} possible configurations $c \in \{+1, -1\}^{MN}$. Every configuration has an energy $E(c)$; E is a mapping from configuration space to the real numbers, and depends on the details of the interaction between the particles. It is well-known that, in thermodynamic equilibrium, the probabilities $p(c)$ of the configurations follow the Boltzmann distribution, i.e.

$$p(c) = \frac{\exp\big(- \beta E(c)\big)}{Z}\,, \tag{76}$$

where $\beta := 1/(kT)$ is the *inverse temperature* (with k being Boltzmann's constant, and T the absolute temperature), and Z is the *partition function*

$$Z := \sum_c \exp\big(- \beta E(c)\big)\,. \tag{77}$$

The partition function appears just as a normalization factor here, but is, in fact, a generating function and, as such, a quantity of fundamental importance. After all, all macroscopic properties of the system may be derived from it (in the literal sense indeed, since the operation involves differentiation). One therefore seeks to calculate Z once the model (i.e. $E(c)$) is specified, but this tends to be a formidable task due to the large number of configurations involved.

For two-dimensional systems on a rectangular patch, the *transfer matrix method* is appropriate, which solves the problem at least formally, as long as the interactions do not span more than two neighbouring rows. It will also provide the link to the evolution model. The (row-to-row) transfer matrix is a $2^N \times 2^N$ matrix which 'transports' the distribution of 'row configurations' from one row of the lattice patch to the next. The calculation of Z for the patch is then achieved by starting from the equidistribution at the bottom row, successive application of the transfer matrix, and summing up the result at the last row.

To be more specific, let a configuration be defined through its row configurations $\boldsymbol{s}^i := (s_1^i, s_2^i, \dots, s_N^i)$, $i = 1, \dots, M$ (i.e. the superscripts label rows, the subscripts label columns). Assume that the interaction energy *within* a row with configuration $\boldsymbol{s}$ is given by $E(\boldsymbol{s})$, and interaction *between* the rows is *nearest neighbour*, i.e. every pair of neighbouring spins in the vertical direction yields an energy contribution of $-J s_j^{i+1} s_j^i$, where J determines the strength of the interactions; so there are negative energy contributions from all neighbouring pairs which are alike, and positive ones from those which are different. The energy of the whole configuration then reads

$$E(c) = \sum_{j=1}^{M} \Big(- J \sum_{i=1}^{N} s_i^{j+1} s_i^j + E(\boldsymbol{s}^j)\Big)\,. \tag{78}$$

The corresponding row-to-row transfer matrix (cf. Thompson, 1972) has elements

$$\mathcal{T}_{\boldsymbol{s}'\boldsymbol{s}} = \exp\big(\beta J \sum_i s_i' s_i\big) \exp\big(- \beta E(\boldsymbol{s})\big)\,. \tag{79}$$

Let us now link the transfer matrix with our mutation-selection model. To this end, we consider a discrete-time version of Eq. (3),

$$\boldsymbol{x}(t+\tau) = \mathcal{V}\mathcal{W}\boldsymbol{x}(t) \tag{80}$$

where τ denotes the duration of one generation. Eq. (80) is obtained through the identifications

$$\mathcal{V} = \exp(\tau \mathcal{M}) \quad \text{and} \quad \mathcal{W} = \exp(\tau \mathcal{R}). \tag{81}$$

More explicitly, the elements of the diagonal matrix $\mathcal{W}$ are $w_{\boldsymbol{s}} = \exp(\tau R_{\boldsymbol{s}})$ (the well-known relationship between Wrightian and Malthusian fitness, cf. Crow & Kimura 1970), and the elements of $\mathcal{V}$ are the mutation probabilities

$$v_{\boldsymbol{s}'\boldsymbol{s}} = p^{D(\boldsymbol{s}',\boldsymbol{s})}(1-p)^{N-D(\boldsymbol{s}',\boldsymbol{s})}\,, \tag{82}$$

with $p = e^{-\mu\tau}\sinh(\mu\tau)$, which belong to the quasispecies model (5), to which the transfer matrix approach was applied in the first place (Leuthäusser, 1986). Actually, Leuthäusser (1986, 1987) showed that the mutation-reproduction matrix $\mathcal{VW}$ is exactly equivalent to the transfer matrix of a two-dimensional Ising model of the type just described. This is because its elements may be rewritten as

$$(\mathcal{VW})_{\boldsymbol{s}'\boldsymbol{s}} = \big(p(1-p)\big)^{N/2}\cdot\exp\big(\frac{\beta}{2}\sum_j s'_j s_j\big)\exp(R_{\boldsymbol{s}}\tau)\,, \tag{83}$$

where $\beta = \ln(p/(1-p))$. The analogy with (79) is then apparent (note that $\big(p(1-p)\big)^{N/2}$ is a constant factor independent of the spin configuration).

We have reproduced this result here because there is a plausible biological picture behind this equivalence. This is the line of descent depicted in Fig. 3, which is obtained when sequences are piled on top of each other generation-wise (i.e. grandmother-mother-daughter-granddaughter sequence); note that the result is a genealogy in the sense of a single lineage, not in the sense of the coalescent (it does not bifurcate). A line of descent may now be identified with a configuration of a two-dimensional Ising lattice patch, where the columns correspond to positions in the sequence, and the rows to generations. The present time corresponds to the (upper) patch surface. The interactions defining this Ising model are anisotropic. They are nearest-neighbour between rows; this corresponds to mutation, which proceeds independently at every site, and has no memory of previous generations. Within rows, on the other hand, interactions may be very complicated and long-ranged—just imagine a protein folding back upon itself. These interactions represent the fitness landscape, i.e. they define the mapping from sequences (i.e. row configurations) into fitness values. The fitness of a sequence, in turn, corresponds to the interaction energy within the row. Since the landscape is stagnant, the interactions (but not the fitness!) are the same for every row. More precisely, the fitness landscape may be decomposed into interaction coefficients η_I in the way described in the main text. The interaction pattern stored in these coefficients may be visualized as in Fig. 3. It may be understood as a recipe of how to determine the fitness given the sequence. This recipe is the same for every sequence; hence the interactions are identical in all rows of the lattice.

The translation between biological and physical quantities may be continued. In particular, an infinite biological population (together with its history) corresponds to an ensemble of Ising configurations, the mutation rate corresponds to a temperature-like parameter (cf. (79) and (83)), and the error threshold may be identified with a phase transition of the system. Apart from its intuitive appeal, this equivalence makes available the highly developed tools of statistical physics for use in population genetics. However, over the past ten years, surprisingly few applications have made use of it, and progress along these lines has been surprisingly sparse. Apart from the case of multiplicative fitness, there are still no landscapes for which exact solutions are available. After all, a few approximate and numerical studies have exploited the transfer matrix method, e.g. Tarazona (1992). The reason for the general sparseness of results seems to be that, in $\exp(\tau\mathcal{M})\exp(\tau\mathcal{R})$, horizontal and vertical interactions of all orders mix. This may be seen by evaluating the matrix exponentials and writing their product in terms of sums of products of Pauli matrices, which results in 'mutation' (σ^x) terms which are influenced by 'reproduction' terms (σ^z), and vice versa. In contrast, mutation and reproduction interactions remain separate in $\mathcal{M} + \mathcal{R}$.

But let now τ in Eqs. (80) and (81) shrink to zero, as in Fig. 4, so that the number of generations per unit time goes to infinity (which corresponds to overlapping generations); biologically, this is the *short-generation limit*. In the corresponding physical picture, this process is known as the *anisotropic limit*, and it leads from the classical Ising model to the quantum chain described in the main text. Mathematically, the corresponding limit reads $n \to \infty$ under $\tau \to 0$ with $n\tau = t = \text{const}$, and it follows from the Trotter formula (Reed & Simon, 1980, Thm. VIII.29) that

$$(\mathcal{V}\mathcal{W})^n = \Big(\exp\Big(\frac{t}{n}\mathcal{M}\Big)\exp\Big(\frac{t}{n}\mathcal{R}\Big)\Big)^n \quad \overset{n\to\infty}{\longrightarrow} \quad \exp\big(t(\mathcal{M}+\mathcal{R})\big)\,; \tag{84}$$

convergence is with respect to any matrix norm, cf. Wagner *et al.* (1998). Hence, the solution of the discrete time equation converges to that of the parallel one in the short generation limit.

To recapitulate, the transfer matrix picture involves a one-to-one correspondence between an *individual* (and its history) and a spin system. In contrast, in the quantum-mechanical picture developed in the main text, the collection of spins corresponds to the whole *population*.

Acknowledgements

It is our pleasure to thank M. Baake, J. Hermisson, and O. Redner for helpful discussions, and for critically reading the manuscript. We are grateful to to Nick Barton, Kevin Dawson, and Paul Higgs for their constructive criticism.

References

Abramowitz, M. & Stegun, I. (1970). *Handbook of Mathematical Functions*. New York: Dover.

Akin, E. (1979). *The Geometry of Population Genetics*, *Lect. Notes Biomath.*, vol. 31, Berlin: Springer.

Amann, H. (1990). *Ordinary Differential Equations*. Berlin: de Gruyter.

Baake, E. (1995). Diploid models on sequence space. *Journal of Biological Systems* **3**, 343–349.

Baake, E., Baake, M. & Wagner, H. (1997). The Ising quantum chain is equivalent to a model of biological evolution. *Physical Review Letters* **78**, 559–562. Erratum, *Physical Review Letters* **79** (1997), 1782.

Baake, E., Baake, M. & Wagner, H. (1998). Quantum mechanics versus classical probability in biological evolution. *Physical Review E* **57**, 1191–1192.

Balian, R. (1991). *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics*, vol. 1. Berlin: Springer.

Barton, N. & Shpak, M. (2000). The stability of symmetric solutions to polygenic models. *Theoretical Population Biology* **57**, 249–263.

Barton, N. & Turelli, M. (1991). Natural and sexual selection on many loci. *Genetics* **127**, 229–255.

Bauer, H. (1996). *Probability Theory.* Berlin: de Gruyter.

Bürger, R. (1998). Mathematical properties of mutation-selection models. *Genetica* **102/103**, 279–298.

Bürger, R. (2000). *The Mathematical Theory of Selection, Recombination, and Mutation.* Chichester: Wiley.

Cardy, J. L. (1987). Conformal invariance. In: C. Domb & J. Lebowitz (eds.), *Phase Transitions and Critical Phenomena*, vol. 11, pp. 55–126. London: Academic Press.

Charlesworth, B. (1990). Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genetical Research Cambridge* **55**, 199–221.

Crow, J. F. (1993). Mutation, mean fitness, and genetic load. *Oxford Surveys in Evolutionary Biology* **9**, 3–42.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory.* New York: Harper & Row.

Crow, J. F. & Simmons, M. J. (1983). The mutation load in *Drosophila.* In: M. Ashburner, H. L. Carson & J. N. Thompson (eds.), *The Genetics and Biology of Drosophila*, vol. 3c, pp. 1–35. London: Academic Press.

Domingo, E. & Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annual Review of Microbiology* **51**, 151–178.

Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* **148**, 1667–1686.

Eigen, M., McCaskill, J. & Schuster, P. (1989). The molecular quasi-species. *Advances in Chemical Physics* **75**, 149–263.

Elena, S. F. & Lenski, R. E. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398.

Ewens, W. (1979). *Mathematical Population Genetics.* Berlin: Springer.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, vol. I, 3rd ed. New York: Wiley.

Gavrilets, S. (1997). Evolution and speciation on holey adaptive landscapes. *Trends in Ecology and Evolution* **12**, 307–312.

Gerisch, T. (1993). Internal symmetries and limiting Gibbs states in quantum lattice mean field theories. *Physica* **A 197**, 284–300.

Hadeler, K. P. (1981). Stable polymorphisms in a selection model with mutation. *SIAM Journal of Applied Mathematics* **41**, 1–7.

Hermisson, J., Wagner, H. & Baake, M. (2000). Four-state quantum chain as a model of sequence evolution. *Journal of Statistical Physics*, in press.

Higgs, P. G. (1994). Error thresholds and stationary mutant distributions in multi–locus diploid genetics models. *Genetical Research Cambridge* **63**, 63–78.

Hofbauer, J. (1985). The selection mutation equation. *Journal of Mathematical Biology* **23**, 41–53.

Holland, J. J., de la Torre, J. C. & Steinhauer, D. A. (1990). Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly be chemical mutagenesis. *Journal of Virology* **64**, 3960–3962.

Horn, R. & Johnson, C. (1991). *Topics in Matrix Analysis*. Cambridge: Cambridge University Press.

Johnson, T. (1999). The approach to mutation-selection balance in an infinite asexual population, and the evolution of mutation rates. *Proceedings of the Royal Society of London B* **266**, 2398–2397.

Kadanoff, L. P. (1976). Scaling, universality and operator algebras. In: C. Domb & M. S. Green (eds.), *Phase Transitions and Critical Phenomena*, vol. 5a, pp. 1–34. London: Academic Press.

Kauffman, S. & Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology* **128**, 11–45.

Kimura, M. & Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440.

Kondrashov, A. S. (1998). Measuring spontaneous deleterious mutation process. *Genetica* **102/103**, 183–197.

Leuthäusser, I. (1986). An exact correspondence between Eigen's evolution model and a two-dimensional Ising system. *Journal of Chemical Physics* **84**, 1884–1885.

Leuthäusser, I. (1987). Statistical mechanics on Eigen's evolution model. *Journal of Statistical Physics* **48**, 343–360.

Li, W. H. (1987). Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *Journal of Molecular Evolution* **24**, 337–345.

Li, W.-H. (1997). *Molecular Evolution*. Sunderland: Sinauer.

Lieb, E., Schultz, T. & Mattis, D. (1961). Two Soluble Models of an Antiferromagnetic Chain. *Annals of Physics* **16**, 407–466.

Maynard Smith, J. & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford: Freeman.

McVean, G.A.T. & Charlesworth, B. (2000). The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929–944.

Moran, P. A. P. (1976). Global stability of genetic systems governed by mutation and selection. *Mathematical Proceedings of the Cambridge Philosophical Society* **80**, 331–336.

Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. (1972). Mutation rate and dominance genes affecting viability. *Genetics* **72**, 335–355.

O'Brien, P. (1985). A genetic model with mutation and selection. *Mathematical Biosciences* **73**, 239–251.

Phillips, P.C., Otto, S.P. & Whitlock, M.C. (2000). Beyond the average: The evolutionary importance of gene interactions and variability of epistatic effects. In: J. D. Wolf, E.D. Brodie & M.J. Wade (eds.), *Epistasis and the Evolutionary Process*, pp. 20–38. Oxford: Oxford University Press.

Raggio, G. & Werner, R. (1989). Quantum statistical mechanics of general mean field systems. *Helvetica Physica Acta* **62**, 980–1002.

Raggio, G. & Werner, R. (1991). The Gibbs variational principle for inhomogeneous mean-field systems. *Helvetica Physica Acta* **64**, 633–667.

Reed, M. & Simon, B. (1980). *Methods of Modern Mathematical Physics*, vol. I: Functional Analysis, 2nd ed. San Diego: Academic Press.

Rozanov, Y. A. (1977). *Probability Theory: A Concise Course*, revised ed. New York: Dover.

Rumschitzky, D. S. (1987). Spectral properties of Eigen evolution matrices. *Journal of Mathematical Biology* **24**, 667–680.

Stadler, P., Seitz, R. & Wagner, G. P. (2000). Population-dependent Fourier decomposition of fitness landscapes over recombination spaces: Evolvability of complex characters. *Bulletin of Mathematical Biology* **62**, 399–428.

Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. (1995). Phylogenetic inference. In: D. M. Hillis, C. Moritz & B. K. Mable (eds.), *Molecular Systematics*, pp. 407–514. Sunderland: Sinauer.

Tarazona, P. (1992). Error threshold for molecular quasispecies as phase transition: From simple landscapes to spin glass models. *Physical Review* **A45**, 6038–6050.

Taylor, C. F. & Higgs, P. G. (2000). A population genetics model for multiple quantitative traits exhibiting pleiotropy and epistasis. *Journal of Theoretical Biology* **203**, 419–437.

Thompson, C. J. (1972). *Mathematical Statistical Mechanics*. New York: Macmillan.

Thompson, C. J. & McBride, J. L. (1974). On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. *Mathematical Biosciences* **21**, 127–142.

Turelli, M. & Barton, N. (1994). Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? *Genetics* **138**, 913–941.

Wagner, H. (1998). *Biologische Sequenzraummodelle und Statistische Mechanik.* Naturwissenschaftl. Reihe No. 18. Darmstadt: Dissertationsdruck Darmstadt.

Wagner, H., Baake, E. & Gerisch, T. (1998). Ising quantum chain and sequence evolution. *Journal of Statistical Physics* **92**, 1017–1052.

Whitlock, M. C., Philips, P. C., Moore, F. B.-G. & Tonsor, S. J. (1995). Multiple fitness peaks and epistasis. *Annual Review in Ecology and Systematics* **26**, 601–629.

Wiehe, T. (1997). Model dependency of error thresholds: the role of fitness functions and contrasts between the finite and infinite sites models. *Genetical Research Cambridge* **69**, 127–136.

Wiehe, T., Baake, E. & Schuster, P. (1995). Error Propagation in Reproduction of Diploid Organisms. *Journal of Theoretical Biology* **177**, 1–15.

Wolfram, S. (1996). *The Mathematica Book*, 3rd ed. Champaign: Wolfram Media.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: *Proceedings of the Sixth International Congress of Genetics*, vol. 1, pp. 356–366.

Yang, C. (1952). The Spontaneous Magnetization of a Two-Dimensional Ising-Model. *Physical Review* **83**, 808–816.

Figure captions

Figure 1:
Mutation-selection balance in a sharply-peaked landscape (left), and under a quadratic fitness function with positive epistasis (right). In the sharply-peaked landscape, the reference type has a selective advantage α over all others, which are equally unfit, i.e., $R_0 = 1$, $R_k = 1 - \alpha$ for $k \geq 1$, where k is the number of mutated sites. Shown is the numerical solution of Eq. (1) for $N = 30$, and selective advantage $\alpha = 0.03$ of the favourable type. Relative frequencies p_k of configurations with k mutant sites are represented as grey scales in the plane with k ($0 \leq k \leq 30$) in the horizontal direction, and the mutation rate per site, μ ($0 \leq \mu \leq 0.002$) in the vertical direction (type freqencies increase with darkness of shading). For mutation rates above the error threshold ($\mu_c \simeq \alpha/N = 0.001$), the population is evenly distributed over type space, i.e. $p_k = \binom{N}{k}/2^N$. For the quadratic fitness function, $R_k = -(ak + ck^2/2N)$ with $N = 30$, $c = -0.004$, and $a = -0.002 + 2 \cdot 10^{-6}$, which corresponds to the Hamiltonian (26) with $\gamma = 0.001$ and $\alpha = 10^{-6}$ (this breaks the symmetry). Here, the type distribution broadens and moves away from the optimum gradually, until the agreement with the fittest type (i.e. the surplus) vanishes at $\mu_c \simeq \gamma = 0.001$. For both fitness functions, this transition become sharp in the limit $N \to \infty$.

Figure 2:
A biological population and a collection of quantum-mechanical spins may both be identified with a vector in $\mathbb{R}^{2^N}$, and with each other. Left: $N = 1$. A vector in the unit simplex under the competing actions of mutation (at rate μ) and selection (at rate $\eta_1 > 0$). The dashed line indicates the direction of mutation (the transversal field), whereas the horizontal direction corresponds to selection (the longitudinal field). Right: Chain of length $N = 5$. The marginal frequencies at each of the sites may be represented by a vector in the unit simplex (the vectors are drawn to have unit one-norm each). Fitness is decomposed into additive ($\eta_1, \dots, \eta_5$) and epistatic (η_{12}, η_{345}) components. The additive components act separately at the sites (longitudinal field), as does mutation (transversal field). The epistatic components introduce interactions within the chain.

Figure 3:
Line of descent as a two-dimensional, anisotropic Ising model. Sequences descending from each other (i.e. grandmother-mother-daughter) form the rows of the lattice. This way, columns correspond to sequence positions, rows to generations, and the transition from one row to the next is governed by the mutation-reproduction matrix $\mathcal{VW}$. If the letters of the sequence are identified with spins, every line of descent corresponds to one possible configuration of a two-dimensional Ising model, and the transfer matrix takes the role of the mutation-reproduction matrix. Interactions are anisotropic: Interactions *between* the rows (dashed) are nearest neighbour and correspond to mutation. Interactions *within* the rows (solid lines) may be arbitrary and long-range, but are identical for every row. Their presence or absence, as well as their strengths, define the fitness landscape in the sense of a mapping from sequence space into the real numbers.

Figure 4:
Short-generation or anisotropic limit. The lattice is as in Fig. 3, with columns corresponding to sequence positions and rows to generations. The distance between rows tends to zero, whereas the distance between columns remains constant. This way, the line of descent with discrete

generations turns into the parallel mutation-selection model in continuous time, and the classical Ising model is replaced by an Ising quantum chain.

Figure 5:
Time evolution of the mean fitness on Onsager's landscape with $\gamma = 0.001$ and $\mu = 0.00025$ for $N = 10$ (dotted), $N = 100$ (dashed) and $N = 1000$ (solid line). The initial conditions are a pure wildtype distribution ($\bar{u} = 1$), and equidistribution of sequences ($\bar{u} = 0$), respectively. The latter is not meant to be a realistic initial condition, but illustrates the properties of the scaling. Furthermore, any intermediate value on the curve may be taken as initial condition due to the semigroup property of the flow. Top: Extensive scaling of mutation and fitness (as in (6) and (8)) yields uniform convergence. If reproduction and mutation rates have the unit 1/year, then convergence to equilibrium occurs roughly within 2000 years, even from the most extreme initial condition. Bottom: Intensive scaling (obtained through $R_s \to R_s/N$, $\mu \to \mu/N$) slows evolution down. Then, one has only pointwise convergence to zero for every value of t.

Figure 6:
Mean fitness, surplus, and variance in fitness per site at equilibrium, in the limit $N \to \infty$, as a function of the mutation rate per site. Dotted: additive fitness (with $\alpha = 0.001$); solid line: quadratic fitness function (with $\alpha = 0, \gamma = 0.002$); dashed: Onsager's landscape (with $\gamma = 0.001$). The parameters are chosen such that, for $N = 1000$ sites, the intensive quantities translate into the extensive ones $R_0 = 1$ (fittest type), $R_1 \simeq 0.998$ or $R_1 \simeq 0.996$ (second-best type in the case of additive fitness, or in the case of quadratic and Onsager landscape, respectively), $R_{N/2} = 0$ (randomized genotype), and V_R is in the range of 10^{-3}.

Figure 7:
Equilibria for the quadratic fitness function. Shown are those cases with a fitness maximum at the wildtype (directional selection). From left to right: $-\alpha < \gamma < 0$; $0 < \gamma < \alpha$; $0 < \alpha < \gamma$; $-\gamma < \alpha < 0$. Upper panels: (intensively-scaled) fitness function r (vertical axes) as a function of u (horizontal axes); lower panels: left-hand side of Eq. (70) (solid line), and right-hand side of Eq. (70), for small μ (dotted) as well as large μ (dashed), as a function of φ. A bullet indicates the solution of (70). An error threshold occurs if the intersection disappears for some large enough μ. This only happens in the regime $-\gamma < \alpha < 0$.

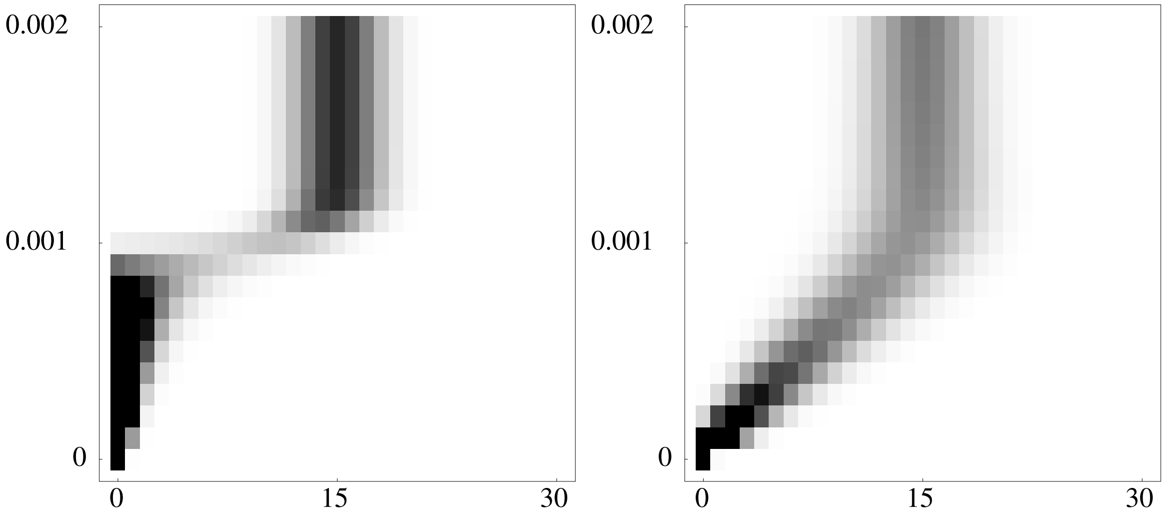

Figure 1:

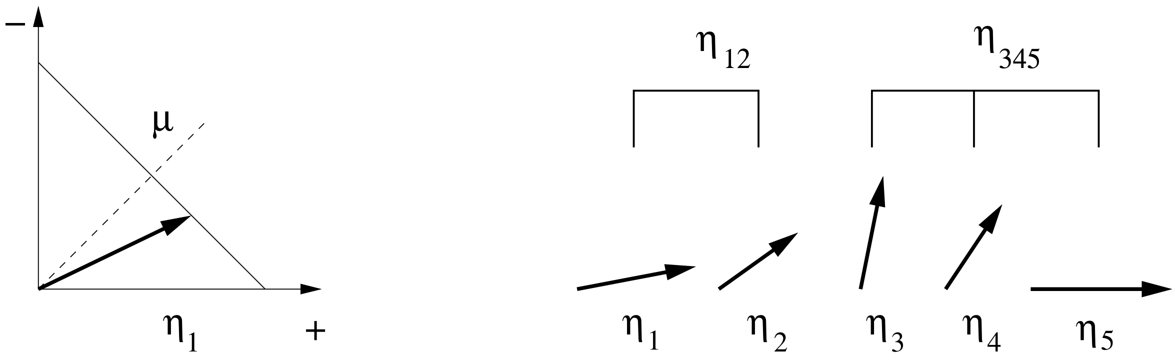

Figure 2:

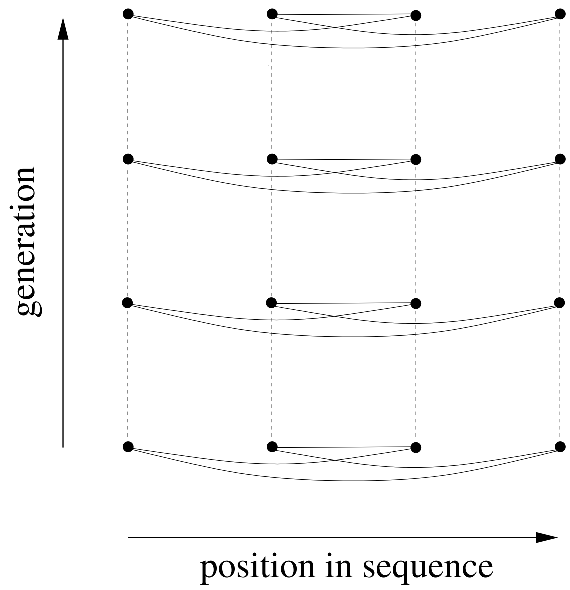

Figure 3:

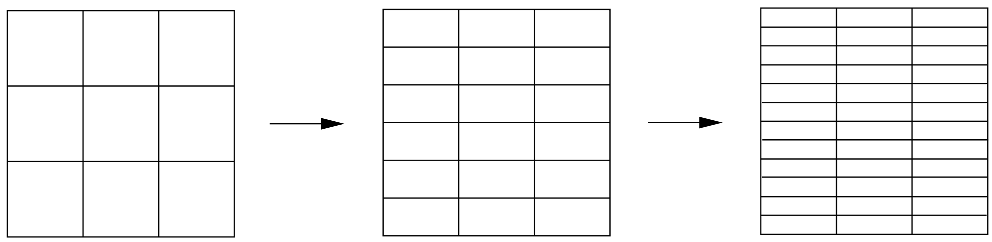

Figure 4:

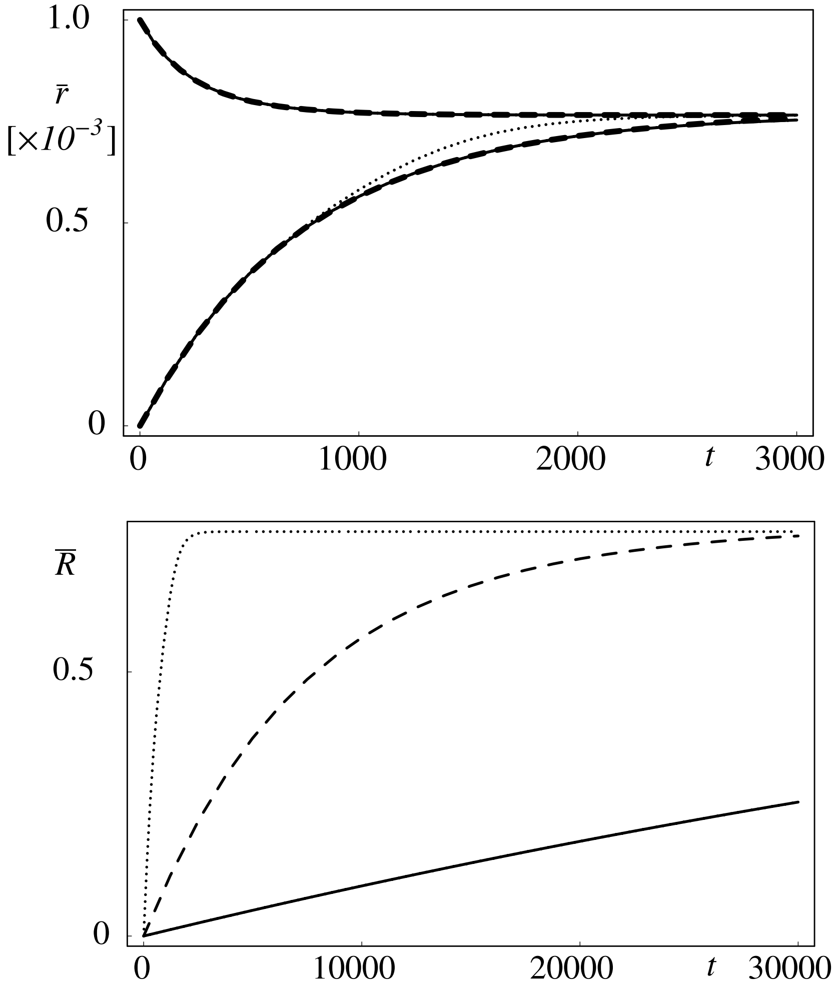
1.0
$\bar{r}$
$[\times 10^{-3}]$
0.5
0
0
1000
2000
t
3000
$\overline{R}$
0.5
0
0
10000
20000
t
30000

Figure 5:

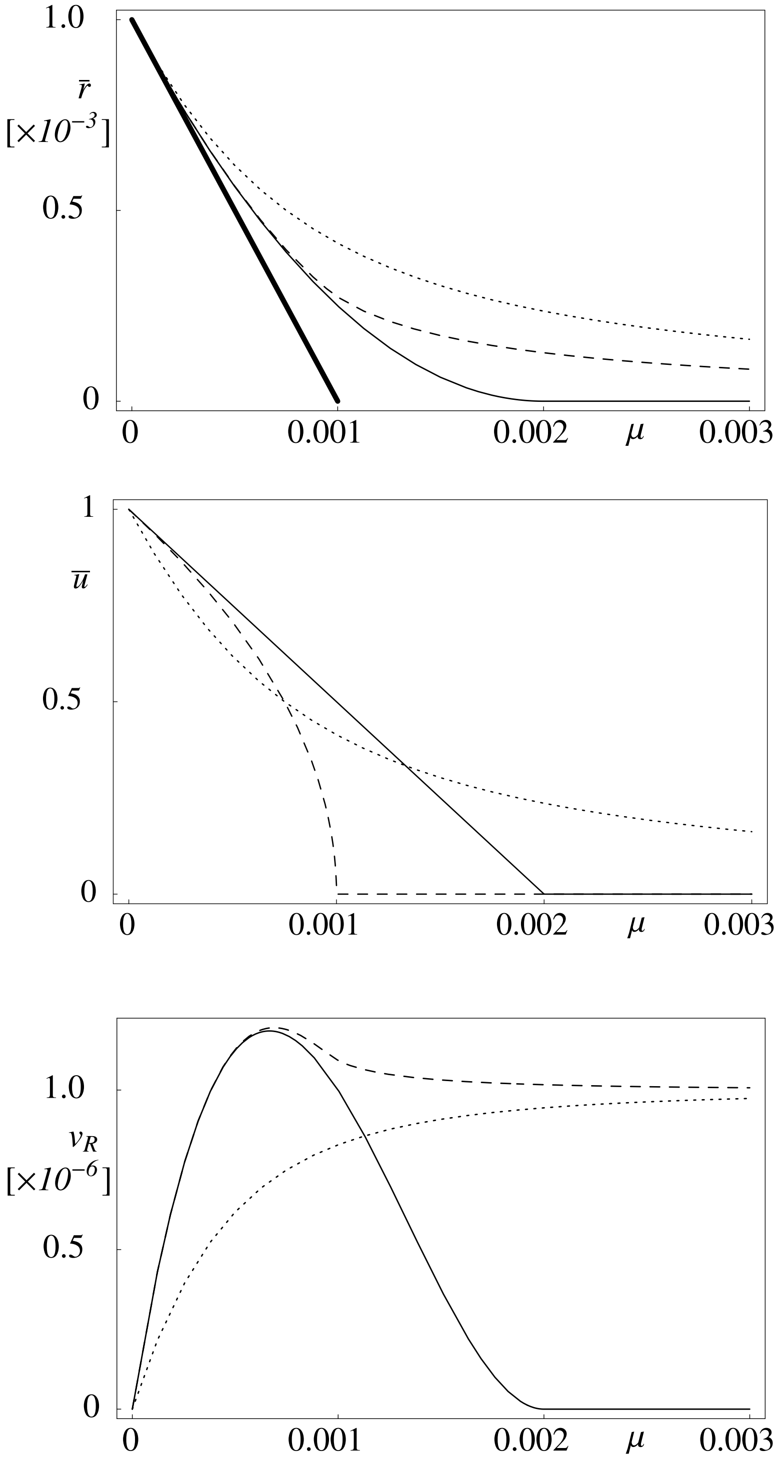
1.0
$\bar{r}$
$[\times 10^{-3}]$
0.5
0
0
0.001
0.002
μ
0.003
1
$\overline{u}$
0.5
0
0
0.001
0.002
μ
0.003
1.0
v_R
$[\times 10^{-6}]$
0.5
0
0
0.001
0.002
μ
0.003

Figure 6:

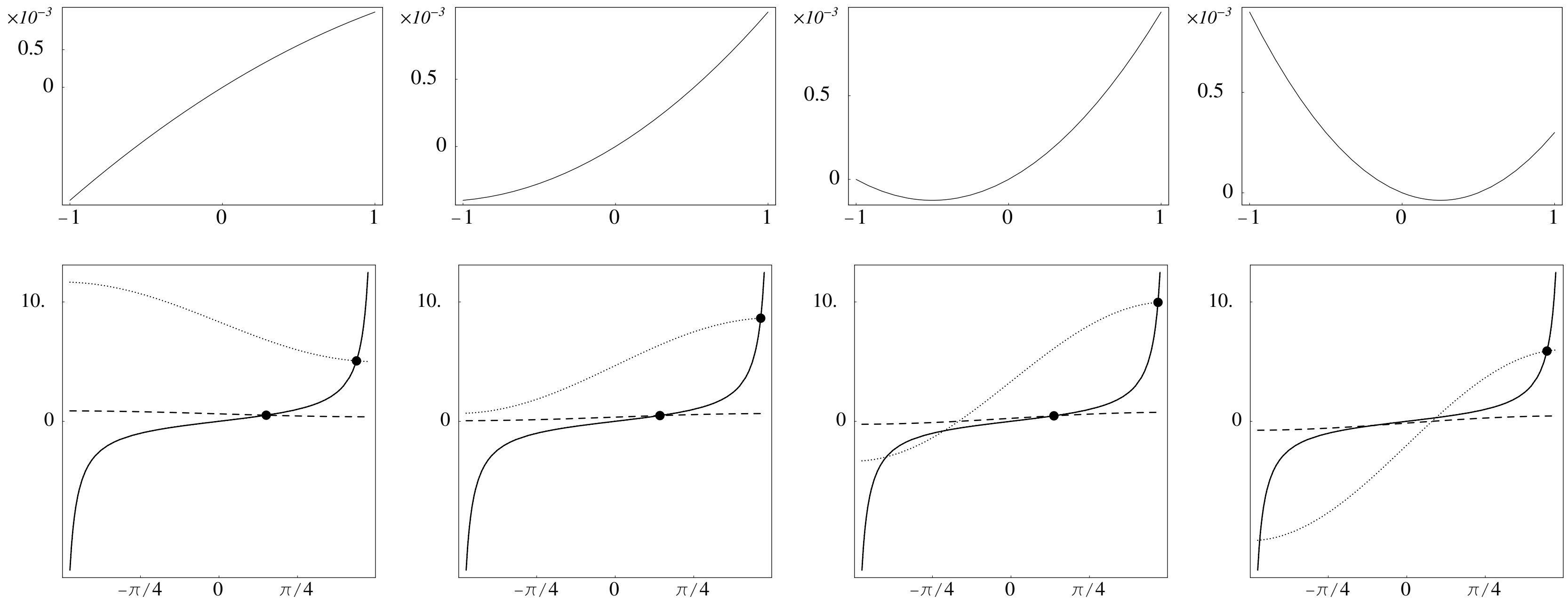

Figure 7: