RICE UNIVERSITY

Statistical Mechanics of Quasispecies theories of molecular evolution

by

Enrique Muñoz Tavera

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE:

Dr. Michael W. Deem, Chair
John W. Cox Professor of Biochemical and Genetic Engineering and Professor of Physics and Astronomy

Dr. Qimiao Si
Harry C. and Olga K. Weiss Professor of Physics and Astronomy

Dr. Oleg Igoshin
Assistant Professor of Bioengineering

Houston, Texas
October, 2008
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction. In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
ABSTRACT

Statistical Mechanics of Quasispecies theories of molecular evolution

by

Enrique Muñoz Tavera

This thesis presents a statistical mechanical analysis of different formulations of quasispecies theory of molecular evolution. These theories, characterized by two different families of models, the Crow-Kimura and the Eigen model, constitute a microscopic description of evolution. These models are most often used for RNA viruses, where a phase transition is predicted, in agreement with experiments, between an organized or quasispecies phase, and a disordered non-selective phase when the mutation rate exceeds a critical value.

The methods of statistical mechanics, in particular field-theoretic methods, are employed to obtain analytic solutions to four problems relevant to biological interest. The first chapter presents the study of evolution under a multiple-peak fitness landscape, with biological applications in the study of the proliferation of viruses or cancer under the control of drugs or the immune system. The second chapter studies
the effect of incorporating different forms of horizontal gene transfer and two-parent recombination to the classical formulation of quasispecies models. As an example, we study the effect of the sign of epistasis of the fitness landscape on the advantage or disadvantage of recombination for the mean fitness. The third chapter considers the relaxation of the purine/pyrimidine assumption in the classical formulation of the models, by formulating and solving the parallel and Eigen models in the context of a four-letter alphabet. The fourth and final chapter studies finite population effects, both in the presence and in the absence of horizontal gene transfer.
Acknowledgments

In the first place, I would like to acknowledge my thesis adviser, Dr. Michael W. Deem, for his stimulating and inspirational mentoring over these years, and for the opportunity of conducting research in the exciting field of theoretical biophysics, particularly in molecular evolution. I also thank Dr. Qimiao Si for his stimulating lectures and discussions on Quantum Many Body theory and Quantum Phase Transitions. I would also like to thank the participation of Dr. Oleg Igoshin in the revision and evaluation of my work. I also thank Dr. Jeong-Man Park and Dr. David Saakian for enlightening discussions over the realization of this work.

I would like to dedicate this thesis to my family. My wife Irene for her love and support, and my daughter Sofia. I thank my parents Enrique and Beatriz, my aunt Maria Angelica and my grand mother Sofia for the invaluable gift of education. A special dedication goes to the memory of my grand fathers, Carlos, who enjoyed physics, and Juan, a paleontologist who dedicated his life to the study of Darwinian evolution.

Finally, I express my gratitude to my former professors Abelardo, Luciano and Zdenka, from whom I got part of the inspiration to get involved in the mathematical and physical sciences.
# Contents

Abstract 

Acknowledgments 

List of Figures 

List of Tables 

1 Introduction 

1.1 Quasispecies theory of evolution ........................................ 1  
1.2 Coherent states and Path Integrals ..................................... 11  
1.2.1 Coherent states ..................................................... 11  
1.2.2 Path Integrals in the coherent states representation ............ 21  

2 Quasispecies theory for multiple-peak fitness landscapes ....... 24  

2.1 Introduction ............................................................ 24  
2.2 Crow-Kimura model with multiple peaks ............................. 27  
2.2.1 The parallel model with two peaks ............................... 29  
2.2.2 Exact solution of the K peak case by a path integral representation 31  
2.2.3 Explicit results for the two peak case ............................ 33  
2.3 Eigen model with multiple peaks ...................................... 37  
2.3.1 Exact solution by a path integral representation .............. 37
3 Quasispecies theory for Horizontal Gene Transfer and Recombination

3.1 Introduction ............................................. 57
3.2 The parallel model ...................................... 65
   3.2.1 Horizontal gene transfer of non-overlapping blocks ....... 66
   3.2.2 Horizontal gene transfer for multiple-size blocks .......... 82
   3.2.3 The parallel model with two-parent recombination .......... 88
3.3 The Eigen model ........................................ 112
   3.3.1 Horizontal gene transfer of non-overlapping blocks ....... 112
   3.3.2 Horizontal gene transfer for multiple-size blocks .......... 118
   3.3.3 The Eigen model with two-parent recombination .......... 123
3.4 Conclusion ............................................ 131
4 Solution of the four-letter alphabet Crow-Kimura and Eigen models

by Schwinger spin coherent states 175

4.1 Introduction ........................................ 176

4.2 The parallel model for an alphabet of size $h = 4$ .................. 178

4.2.1 Fitness landscape .................................. 181

4.2.2 Schwinger spin coherent states representation of the parallel
model ......................................................... 182

4.2.3 Functional integral representation of the parallel model .... 190

4.2.4 The large N limit of the parallel model is a saddle point .. 196

4.2.5 Analytic results for the symmetric mutational scheme .... 197

4.2.6 Quartic fitness landscape .............................. 203

4.2.7 Generalization to arbitrary mutational scheme ............. 205

4.3 The four-state Eigen model .............................. 212

4.3.1 The four-state Eigen model in operator form ............. 213

4.3.2 The large N limit of the four-state Eigen model is a saddle point 218

4.3.3 Analytical results for the symmetric mutation scheme .... 220

4.3.4 The quadratic fitness landscape ........................ 224

4.3.5 The quartic fitness landscape .......................... 225

4.3.6 Generalization to an arbitrary mutation scheme .......... 226

4.4 Conclusion ........................................... 229
5 Quasispecies theory for finite populations

5.1 Introduction ......................................................... 254

5.2 The parallel model in a finite population ......................... 256

5.2.1 Results for the parallel model ............................... 261

5.3 The Eigen model in a finite population .......................... 269

5.3.1 Results for the Eigen model ............................... 273

5.4 Conclusions ......................................................... 275

5.5 Appendix 4 ......................................................... 291

5.6 Appendix 5 ......................................................... 293

6 Conclusions .......................................................... 296

References .............................................................. 298
### List of Figures

1.1 The molecular 'letters' of the genetic alphabet ........................................ 2
1.2 The polymeric structure of DNA ................................................................. 3
1.3 Codons encode for the protein amino acid sequence ....................................... 5
1.4 Codons codify for the protein amino acid sequence ......................................... 8

2.1 A phase diagram for the interaction of a virus and a drug. .......................... 41
2.2 A phase diagram corresponding to the original antigenic sin model. ............. 44
2.3 A phase diagram corresponding to the immune control of cancer .................. 45

3.1 Convention for the sign of epistasis ............................................................. 62
3.2 Pictorial representation of the horizontal gene transfer process considered. .... 67
3.3 Phase diagram of the parallel model for the quadratic fitness, with horizontal gene transfer ........................................................................................................ 76
3.4 Pictorial representation of the two-parent genetic recombination process considered in the theory. ............................................................ 90
3.5 Convergence of the numerical results towards the theoretical value for two-parent recombination in the parallel (Kimura) model for the sharp peak fitness .................. 101
3.6 Probability distributions for two-parent recombination in the parallel model for the quadratic fitness ................................................. 103

3.7 Convergence of the numerical results towards the theoretical value for two-parent recombination in the parallel model for the selective phase S1 ......................................................................................... 104

3.8 Convergence of the numerical results towards the theoretical value for two-parent recombination in the parallel model for the selective phase S2 ......................................................................................... 105

4.1 The Kimura 3 ST mutation scheme ................................................................. 180

4.2 Average composition \( u \), magnetization \( \xi_c \) and fraction of the population at the wild-type sequence \( p_w \), as a function of the parameter \( A/\mu \), for the sharp peak fitness. ................................................................. 200

4.3 Average composition \( u \) and magnetization \( \xi_c \) as a function of the parameter \( \beta/\mu \), for the quadratic fitness when \( \alpha = 0 \). ..................................................... 203

4.4 Average composition \( u \) and magnetization \( \xi_c \) as a function of the parameter \( \beta/\mu \), for the quartic fitness landscape. ................................................................. 204

4.5 The generalized mutation scheme ................................................................. 205

4.6 The average composition \( u \) and magnetization \( \xi_c \) as a function of the parameter \( A/A_0 \), for the sharp peak landscape ................................................................. 223
4.7 The average composition $u$ and magnetization $\xi_c$ are represented as a function of the parameter $\beta/\mu$ for the quadratic fitness, when $\alpha = 0$. 225

4.8 The average composition $u$ and magnetization $\xi_c$ are represented as a function of the parameter $\beta/\mu$ for the quartic fitness landscape. 226

5.1 Theoretical versus results for the probability distribution of the parallel model and quadratic fitness. 263

5.2 Theoretical versus stochastic fluctuations in the probability distribution for the parallel model and quadratic fitness, in the absence of horizontal gene transfer. 264

5.3 Theoretical versus stochastic fluctuations in the probability distribution for the parallel model and quadratic fitness, with horizontal gene transfer. 265

5.4 Fluctuations in the probability distribution as predicted from our theory for the parallel model and quadratic fitness. 266

5.5 Average composition as a function of time from stochastic simulation. 267

5.6 Histograms representing the steady-state composition among 50 independent numerical experiments. 268

5.7 Probability distributions predicted from our theory for the Eigen model. 274

5.8 Fluctuations in the number of individuals with a given sequence as predicted from our theory for the Eigen model. 275
List of Tables

1.1 Standard amino acid codons ........................................... 4

2.1 Analytical versus numerical results for the two peak landscape in the parallel model .............................................. 35

2.2 Analytical versus numerical results for the quadratic fitness landscape in the Eigen model ....................................... 40

3.1 Analytical versus numerical results for horizontal gene transfer in the parallel model for the quadratic fitness ......................... 74

3.2 Analytical versus numerical results for horizontal gene transfer in the parallel model, for the quadratic fitness ..................... 77

3.3 Analytical versus numerical results for horizontal gene transfer in the parallel model for the square-root fitness .................... 79

3.4 Analytical results for horizontal gene transfer with multiple block size in the parallel model for the quadratic fitness ................... 85

3.5 Analytical results for horizontal gene transfer of variable block size in the parallel model for the square-root fitness ................... 87

3.6 Stochastic process versus analytical theory for two-parent recombination in the parallel model for the quadratic fitness ............ 102
3.7 Stochastic process versus differential equation for two-parent recombination in the parallel model for the sharp peak fitness ............... 107

3.8 Analytical theory versus numerical solution for two-parent recombination in the parallel model for the quadratic fitness ................. 109

3.9 Analytical theory versus numerical solution for two-parent recombination in the parallel model for the square-root fitness ............ 110

3.10 Analytical results for horizontal gene transfer in the Eigen model for the square-root fitness .............................................. 117

3.11 Analytical results for horizontal gene transfer in the Eigen model for the square-root fitness .............................................. 122

3.12 Analytical results for two-parent recombination in the Eigen model for the square-root fitness .............................................. 130
Chapter 1
Introduction

1.1 Quasispecies theory of evolution

Molecular evolution constitutes, without a doubts, a cornerstone in modern biophysics. It provides a central dogma [1] that connects diverse fields such as protein folding [2, 3], protein-ligand interactions, and structural molecular biology [1].

Evolutionary dynamics, as understood today, consists of four basic processes: replication, selection, mutation, and different forms of recombination [1, 4]. This last process has been just relatively recently recognized as an important component of evolutionary dynamics [1, 4]. Evolution is believed to be the fundamental source of diversity in biological systems, as it allows living entities to survive environmental changes through adaptation, on a time scale that spans across several generations. Hence, the typical time scales for evolution can be as short as a few months for influenza viruses, as compared with thousands of years for complex mammals such as human beings. Regardless of the time scale and size of the living entity, evolution necessarily involves structural changes at the molecular level, i.e., amino acid substitutions in the protein sequence.

Proteins, the fundamental structural and functional blocks of life, from a physical-chemical perspective, are heterogeneous polymers [5, 6]. Each monomer, called an amino acid, may be one out of 20 different chemical species [7]. Therefore, protein
Figure 1.1 The molecular 'letters' of the genetic alphabet [7]. Left: the different nitrogenated bases are clustered into two chemical groups, purines and pyrimidines. Right: Nucleotides, the monomers which conform the polymeric structure of nucleic acids RNA/DNA. life is written with an alphabet of 20 letters, the amino acids code. However, there exists a more fundamental level at which nature preserves structural information. To synthesize a protein in-vivo, a nucleic acid template is required [7]. The nucleic acids DNA and RNA are also heterogeneous polymers, but the alphabet of monomers is composed just of four different chemical species [7], Fig. 1.1. The alphabet (A,C,G,T) encodes DNA, while (A,C,G,U) encodes RNA. These nitrogenated bases are attached to a pentose sugar structure (ribose or deoxiribose) to constitute nucleotides, the monomers that conform the polymeric structure of nucleic acids, Fig. 1.2.

In analogy with the situation encountered in elementary particle physics, where a combination of three quarks is required to form a composite fermion [8] such as a proton or neutron, a specific triplet of nucleotides in the DNA/RNA, a codon, encodes for a particular amino acid, Table 1.1. Hence, there is a mapping from the
Figure 1.2  The polymeric structure of DNA. Left: DNA is a heterogeneous polymer, constituted by two complementary nucleotide chains with the topology of an α-helix. Right: Nucleotides in the DNA are chemically complementary, and form hydrogen bonds between pairs of purines and pyrimidines [7].

nucleic acids sequence onto the amino acids sequence that constitutes proteins. This mapping, however, is not bijective, since there is redundancy in the number of triplets encoding for a particular amino acid [7], Table 1.1 and Fig. 1.3.

Molecular evolution occurs through the emergence and fixation of substitutions in the nucleic acids sequence of a given species, which translates into the fixation of substitutions in the amino acids sequence of the corresponding encoded proteins.

Different approaches have been proposed for the mathematical modeling of molecular evolution. In particular, quasispecies theory provides an interesting scenario for theoretical physics. Originally conceived to represent molecular evolution in viral populations, quasispecies theory is characterized by two classical models: The parallel,
Table 1.1 Standard amino acid codons [7].

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Single letter code</th>
<th>DNA codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>ATT, ATC, ATA</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>CTT, CTC, CTA, CTG, TTA, TTG</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>GTT, GTC, GTA, GTG</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
<td>TTT, TTC</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>ATG</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
<td>TGT, TGC</td>
</tr>
<tr>
<td>Alanine</td>
<td>A</td>
<td>GCT, GCC, GCA, GCG</td>
</tr>
<tr>
<td>Glycine</td>
<td>G</td>
<td>GGT, GGC, GGA, GGG</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>CCT, CCC, CCA, CCG</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>ACT, ACC, ACA, ACG</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>TCT, TCC, TCA, TCG, AGT, AGC</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>TAT, TAC</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
<td>TGG</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>CAA, CAG</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>AAT, AAC</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>CAT, CAC</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>E</td>
<td>GAA, GAG</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>D</td>
<td>GAT, GAC</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>AAA, AAG</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>CGT, CGC, CGA, CGG, AGA, AGG</td>
</tr>
<tr>
<td>Stop codons</td>
<td>Stop</td>
<td>TAA, TAG, TGA</td>
</tr>
</tbody>
</table>
5

1. Transcription

Polypeptide: Val → His → Leu → ... 

Triplet for leucine.

GUG 
CAC 
GTG 

GUCAUGACU

Triplet for threonine.

DNA: CACOTAAGA
GTGACTCTGACT

Figure 1.3  Codons encode for the protein amino acid sequence. Left: Codons, constituted by triplets of bases along the DNA/RNA, encode for a particular amino acid corresponding to the protein primary structure [7]. Right: Cartoon representation of the protein synthesis mechanism [9].
or Crow-Kimura [10], model and the Eigen model [11, 12]. Both models are expressed in the mathematical language of chemical kinetics, with a system of differential equations representing the time evolution of an infinite population of information carrying macromolecules, equipped with self-reproductive capabilities [11, 12]. These macromolecules, corresponding to the hetero-polymers encoding for genetic information (DNA/RNA) in living cells and viruses, are represented in an abstract picture by binary sequences. The choice of a binary alphabet, which simplifies the mathematical formulation, constitutes a natural coarse-graining criterion according to the two basic chemical structures of the nucleotide bases: Purines (A,G) and pyrimidines (C, T/U) [7], Fig.1.1.

In their original formulation, these classical models include the basic processes of replication, mutation, and selection. Quasispecies models introduce the concept of fitness as the replication rate of a given sequence type. With this operational definition, fitness can be estimated by experimental methods [13, 14]. This direct mapping is a limitation of quasispecies theory, since in reality it is the phenotype and not the genotype characteristics that determine the chances of survival of a living entity in a given environment. Since a molecule or organism interacts with its physical-chemical environment, as well as with other living entities, and that all these conditions fluctuate in time, an exclusive dependence between replication rate and genetic sequence is unlikely to be accurate. Despite its inherent limitations, the
concept of fitness provides a useful approximation to reality, if interpreted in a 'mean-field' sense, by assuming that all the complex environmental effects are 'averaged out' to yield a phenomenological effective replication rate depending only on the genetic sequence. This is in analogy with the concept of effective interactions in many-body systems [15]. For the numerical and analytical study of quasispecies models, it is customary to model the fitness as a deterministic function of the Hamming distance between any mutant in the population and the "wild type." The Hamming distance is defined as the number of digits by which a pair of binary sequences differ [16]. This assumption simplifies considerably the mathematical and numerical analysis, since it provides a projection from the multidimensional sequence space into a single, one-dimensional coordinate.

In the parallel, or Crow-Kimura model, replication and mutation are considered as independent events [10]. Hence, single point mutations are assumed to occur with a fixed rate at each generation. In the Eigen model, mutations are assumed to occur as errors in the replication process [11, 12], Fig. 1.4. Therefore, in the Eigen model multiple mutations may occur along each sequence at each generation.

The most remarkable feature of quasispecies models, both Crow-Kimura and Eigen, is the existence of a phase transition, denoted as the "error threshold," when the mutation rate is below a critical value. This transition separates an organized phase, denoted the "quasispecies", from a disordered phase [11, 12, 17]. The quasis-
Figure 1.4 The replication process of DNA requires the action of several enzymes. Helicases and girases act to open the double strand and modify its topological structure. Polymerases move along the original strand, and insert complementary nucleotides at each site along their path. RNA primers are required to initiate the replication process, which is error prone [7].

Species phase is characterized by a distribution of closely related mutants, rather than by identical sequences. More precisely, the closeness between a pair of sequences is defined in terms of their Hamming distance. The emergence of the quasispecies is a consequence of the auto-catalytic nature of the replication process, which exponentially enriches the population with the fittest mutants [11, 12, 17]. In mathematical terms, as first noticed by Eigen [11] and according to the Perron-Frobenius theorem [18], the matrix that determines the time evolution of the dynamical system of differential equations possesses a positive, non-degenerate maximum eigenvalue that determines the long term evolution of the system towards its corresponding eigenvector, whose components define the composition of the quasispecies.

Quasispecies models have captured the attention of the theoretical physics com-
munity since it was shown that the Eigen model could be exactly mapped into a type of 2D Ising model [19, 20], with sequence and time being the two dimensions. In this picture, it was shown that the average composition of the population in steady-state is equivalent to a surface magnetization [21]. Alternative formulations of quasispecies theory have been reported in terms of a one dimensional quantum spin chain [22, 23, 24, 25, 26, 27, 28] and, more recently, in terms of a quantum field theory formalism [29].

Besides their traditional applications in high energy and condensed matter physics, quantum field theoretical methods provide a powerful technique for the study of classical many-body systems [30, 31], such as reaction diffusion [32, 33], turbulence [30] or birth-death processes [34, 31, 35]. For dynamical systems, these methods are closely related to the Keldysh formalism in non-equilibrium quantum many-body systems [36, 37]. In particular, an exact representation of the classical molecular evolution models in terms of a Schwinger spin Hamiltonian has been obtained [29], and a Path integral representation of the Schrodinger equation in imaginary time leads to a field theoretical expression for the action. The mean fitness or replication rate is then obtained by maximizing the action under suitable constraints [29], because there is a large parameter that enforces the saddle point limit.

More recent studies in quasispecies models consider the possibility of introducing different forms of exchange of genetic material between sequences in the population,
such as horizontal gene transfer and two-parent recombination [38, 39, 40]. In this last direction, as discussed in Chapter 3, one may provide quantitative answers to fundamental biological questions, such as the evolutionary advantage or disadvantage of sex [41, 42, 43, 44, 45, 46, 47, 48]. Another important direction is the study of finite populations evolving according to the quasispecies models dynamics. In this case, the theory must describe the fundamentally stochastic nature of the evolutionary process, and the theoretical task is to solve the corresponding master equation.

In the present thesis, some of these theoretical problems are formulated and solved analytically. Numerical simulations and calculations are presented to test the analytical results. The document is constituted by four chapters, each representing a separate project. Chapter 2 presents the formulation and solution of quasispecies models when the fitness landscape depends on the Hamming distance from multiple peaks, thus representing different biological scenarios where competition exists between an evolving viral population and a drug or the immune system. In this case, the method of one-dimensional quantum spin chains was applied to map the system of differential equations into a path integral. Chapter 3 studies the effect of introducing different forms of horizontal gene transfer and recombination in quasispecies models. In this case, a mapping into a spin boson Hamiltonian, which is treated with a coherent states path integral formalism, was applied to obtain analytical solutions for the steady state properties of the population. Chapter 4 extends the standard
formulation of quasispecies models, by relaxing the assumption of binary sequences
to consider a more realistic alphabet of four letters, to represent evolution of nucleic
acids RNA/DNA. The spin boson coherent states technique is also applied in this
case.

Since coherent states constitute an ubiquitous concept in most of the analytical
methods presented in this thesis, a brief introduction of the terminology, methodology
and properties of these mathematical objects and their connection with path integrals
and partition functions in many-body systems is presented in the next section.

1.2 Coherent states and Path Integrals

1.2.1 Coherent states

The first notion of coherent states was introduced by Schrödinger [49] in his study
of the harmonic oscillator, as a set of propagating wave-packets which coherently
evolve under the action of the harmonic Hamiltonian. The name “coherent states,”
however, was coined much later by Glauber [50] in the different context of photon
statistics in quantum optics. In parallel with their popularity in quantum optics, coherent states have been applied in a variety of different physical scenarios [51],
from condensed matter and statistical mechanics [15, 52, 53] to elementary particle
physics [54] and quantum field theory [55]. In particular, Schwinger [56] developed a
description of angular momentum operators by employing boson spin operators.

There are several formulations of coherent states [51], including more abstract
constructions in terms of generalized Lie algebras [51]. In this chapter, I will present the so-called "canonical formulation," which is the standard in quantum optics [57], and its generalization to many-body quantum mechanics [15].

We first define creation $\hat{a}^\dagger$ and annihilation operators $\hat{a}$, satisfying boson commutation relations

$$[\hat{a}, \hat{a}^\dagger] = 1, \quad [\hat{a}, \hat{a}] = [\hat{a}^\dagger, \hat{a}^\dagger] = 0$$

In the orthogonal basis of occupation states, the action of creation and destruction operators is defined by [58, 34, 33, 32]

$$\hat{a}^\dagger |n\rangle = |n + 1\rangle, \quad \hat{a} |n\rangle = n |n - 1\rangle$$

The choice of coefficients in Eq. (1.2) is standard in dynamics [58, 34, 33, 32], but differs from the usual one of many-body quantum theory. Considering the "vacuum" state $|0\rangle$, and the properties Eq. (1.2), it is straightforward to prove by induction that

$$|n\rangle = (\hat{a}^\dagger)^n |0\rangle, \quad \hat{a} |0\rangle = 0$$

We notice that the definitions Eqs. (1.2) and (1.3) differ from the customary choice of normalization in quantum mechanics, but they are more convenient to represent diffusion-limited reactions [58, 34, 33, 32]. It also implies that the inner product between occupation number states is [58, 34, 33, 32]

$$\langle n | n' \rangle = \langle 0 | (\hat{a})^n (\hat{a}^\dagger)^{n'} | 0 \rangle = n! \delta_{n,n'}$$
and hence the completeness relation for occupation number states becomes

\[ \sum_n \frac{1}{n!} |n\rangle \langle n| = I \]  

(1.5)

By combining the creation and destruction operators, we obtain the "number" operator \( \hat{n} = \hat{a}^\dagger \hat{a} \). It is straightforward to show that the occupation states \( |n\rangle \) constitute eigenstates of the occupation number operator \( \hat{n} \)

\[ \hat{n} |n\rangle = \hat{a}^\dagger \hat{a} |n\rangle = \hat{a}^\dagger |n - 1\rangle = n |n\rangle \]  

(1.6)

Let us define the "displacement" operator

\[ \hat{D}(z) = \exp(z \hat{a}^\dagger - z^* \hat{a}) \]  

(1.7)

By using the Baker–Campbell–Hausdorff formula [59]

\[ e^{A+B} = e^{-[A,\hat{B}]/2} e^A e^B \quad \text{if} \quad [\hat{A},[\hat{A},\hat{B}]] = [\hat{B},[\hat{A},\hat{B}]] = 0 \]  

(1.8)

we have that the displacement operator may be written as

\[ \hat{D}(z) = e^{-|z|^2/2} e^{z \hat{a}^\dagger} e^{-z^* \hat{a}} = e^{|z|^2/2} e^{-z^* \hat{a}} e^{z \hat{a}^\dagger} \]  

(1.9)

From this last expression, it is straightforward to show that the displacement operator is unitary

\[ \hat{D}^\dagger(z) \hat{D}(z) = \hat{D}(z) \hat{D}^\dagger(z) = I \]  

(1.10)

and to prove the semi-group property

\[ \hat{D}(z) \hat{D}(z') = e^{iIm(zz')} \hat{D}(z + z') \]  

(1.11)
The coherent states \(|z\rangle\) are defined by the action of the displacement operator over the vacuum

\[ |z\rangle = \hat{D}(z)|0\rangle \]  

(1.12)

Notice that the property Eq. (1.11) implies

\[ \hat{D}(z)\hat{D}(z')|0\rangle = \hat{D}(z)|z'\rangle = e^{i\mu m(zz')}|z + z'\rangle \]  

(1.13)

An explicit expression for a coherent state \(|z\rangle\) in terms of the orthogonal basis of occupation numbers is obtained by combining Eq. (1.9) and Eq. (1.12)

\[ |z\rangle = e^{-|z|^2/2}e^{z\hat{a}^\dagger}e^{-z^*\hat{a}}|0\rangle = e^{-|z|^2/2}\sum_{n=0}^{\infty} \frac{1}{n!}(z\hat{a}^\dagger)^n|0\rangle \]

\[ = e^{-|z|^2/2}\sum_{n=0}^{\infty} \frac{z^n}{n!}|n\rangle \]

(1.14)

where we have used the properties Eq. (1.3). Therefore, from the orthogonality of the occupation number basis \(\langle n|m\rangle = n!\delta_{n,m}\), we obtain

\[ \langle n|z\rangle = e^{-|z|^2/2}z^n \quad \langle z|n\rangle = e^{-|z|^2/2}(z^*)^n \]

(1.15)

Coherent states are not orthogonal, as can be proved by considering the inner product

\[ \langle z|z'\rangle = \sum_{n=0}^{\infty} \frac{1}{n!}\langle z|n\rangle\langle n|z'\rangle = e^{-|z|^2/2-|z'|^2/2} \sum_{n=0}^{\infty} \frac{(z^*z')^n}{n!} \]

\[ = e^{-|z|^2/2+z^*z'-|z'|^2/2} \]

(1.16)

Here, we introduced the completeness relation Eq. (1.5), and made use of the identities Eq. (1.15). From this result, we notice however that they are normalized to
unity $||z||^2 = \langle z|z \rangle = 1$. Despite not being orthogonal, it is possible to obtain a
resolution of the identity in the representation of coherent states,

$$\int |z\rangle\langle z| \frac{d^2z}{\pi} = I$$

(1.17)

Here, $d^2z \equiv d(Rez)d(Imz) = |z|d|z|d\theta$, the last equivalence in polar coordinates
$z = |z|e^{i\theta}$. To prove that Eq. (1.17) is indeed satisfied, we substitute the expansion
Eq. (1.14) in terms of occupation numbers

$$\int |z\rangle\langle z| \frac{d^2z}{\pi} = \pi^{-1} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{1}{n!m!} \int e^{-|z|^2}(z^*)^n z^m d(Rez)d(Imz) |n\rangle \langle m|$$

$$= \frac{1}{2\pi} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{1}{n!m!} \int_0^{2\pi} d|z|^2 e^{-|z|^2} |z|^{n+m} \int_0^{2\pi} d\theta e^{i(m-n)\theta} |n\rangle \langle m|$$

(1.18)

By substituting the identities

$$\int_0^{2\pi} d\theta e^{i(m-n)\theta} = 2\pi \delta_{n,m} \quad \int_0^{\infty} d|z|^2 e^{-|z|^2} |z|^{2n} = \Gamma(n+1) = n!$$

(1.19)

we finally prove Eq. (1.17),

$$\int |z\rangle\langle z| \frac{d^2z}{\pi} = \sum_{n=0}^{\infty} \frac{1}{n!} |n\rangle \langle n| = I$$

(1.20)

The name coherent states is due to the unique property that, under a harmonic
oscillator Hamiltonian (or in general any Hamiltonian proportional to the number
operator), a coherent state evolves in time as a sequence of coherent states. For a
Hamiltonian $\hat{H} = \hbar \omega \hat{a}^\dagger \hat{a}$, we have that when $|z(t=0)\rangle = |z\rangle$, from Eq. (1.15)

$$|z(t)\rangle = e^{-it\omega \hat{a}^\dagger \hat{a}} |z\rangle = e^{-|z|^2/2} \sum_{n=0}^{\infty} \frac{1}{n!} z^n e^{-i\omega t} |n\rangle = |e^{-i\omega t} z\rangle$$

(1.21)
Another important property of coherent states is the fact that they are eigenstates of the annihilation operator $\hat{a}$. For this purpose, consider again the expansion Eq. (1.15)

$$\hat{a}|z\rangle = e^{-|z|^2/2} \sum_{n=0}^{\infty} \frac{z^n}{n!} \hat{a}|n\rangle = e^{-|z|^2/2} \sum_{n=1}^{\infty} \frac{z^n}{n!} |n-1\rangle$$

$$= ze^{-|z|^2/2} \sum_{n=1}^{\infty} \frac{z^{n-1}}{(n-1)!} = z|z\rangle$$

(1.22)

From the Hermitian adjoint of this identity, also follows

$$\langle z|\hat{a}^\dagger = z^\star \langle z|$$

(1.23)

Thus, the expectation value of the number operator in a coherent state is

$$\langle z|\hat{a}^\dagger \hat{a}|z\rangle = z^\star z \langle z|z\rangle = |z|^2$$

(1.24)

The action of the creation operator on a coherent state is

$$\hat{a}^\dagger|z\rangle = e^{-|z|^2/2} \sum_{n=0}^{\infty} \frac{z^n}{n!} \hat{a}^\dagger|n\rangle = e^{-|z|^2/2} \sum_{n=0}^{\infty} \frac{z^n}{n!} |n+1\rangle$$

$$= e^{-|z|^2/2} \frac{\partial}{\partial z} \sum_{n=0}^{\infty} \frac{z^n}{n!} |n\rangle = \left( \frac{\partial}{\partial z} + \frac{z^\star}{2} \right) |z\rangle$$

(1.25)

Similarly, one can prove

$$\langle z|\hat{a} = \langle z| \left( \frac{\partial}{\partial z^\star} + \frac{z}{2} \right)$$

(1.26)

An operator is defined in normal order : $\hat{O}(\hat{a}^\dagger, \hat{a})$ : when all creation operators are on the left, and destruction operators on the right. For example : $\hat{n} := \hat{a}^\dagger \hat{a}$ is in
normal order. The matrix element of a normal ordered operator between two coherent states, from Eqs. (1.22), (1.23) is

$$\langle z \mid \hat{O}(\hat{a}_1^\dagger, \hat{a}_2) : \mid z' \rangle = O(z^*, z') : \langle z \mid z' \rangle = O(z^*, z') e^{-|z|^2/2 + z^* z' - |z'|^2/2} \quad (1.27)$$

A final result, and following the formalism introduced by Schwinger [56] to describe angular momentum operators through boson operators, is that it is possible to define creation and destruction operators to span a subspace of the occupation number basis with a fixed value of the number operator. In particular, this formulation allows us to treat a spin \( s = 1/2 \) system by employing a pair of boson operators. We define \( \hat{a}_\alpha, \hat{a}^\dagger_\alpha \), for \( \alpha = 1, 2 \)

$$[\hat{a}^\dagger_\alpha, \hat{a}_\beta] = \delta_{\alpha, \beta} \quad [\hat{a}_\alpha, \hat{a}_\beta] = [\hat{a}^\dagger_\alpha, \hat{a}^\dagger_\beta] = 0 \quad (1.28)$$

We define the states \( |n_1, n_2\rangle \) where

$$\hat{a}^\dagger_\alpha \hat{a}_\alpha |n_1, n_2\rangle = n_\alpha |n_1, n_2\rangle \quad \alpha = 1, 2 \quad (1.29)$$

In this formulation, coherent states are defined by a complex vector \( \vec{z} = (z_1, z_2) \).

Using the notation \( \vec{\alpha} = (\hat{a}_1, \hat{a}_2) \), and generalizing Eq. (1.7)–(1.12), we have

$$|\vec{z}\rangle = \hat{D}(\vec{z}) |0,0\rangle = \exp(\vec{z} \cdot \vec{\alpha}^\dagger - \vec{z}^* \cdot \vec{\alpha}) |0,0\rangle$$

$$= e^{-\vec{z} \cdot \vec{\alpha}^\dagger/2} e^{\vec{z} \cdot \vec{\alpha}^\dagger} e^{-\vec{z}^* \cdot \vec{\alpha}} |0,0\rangle = e^{\vec{z} \cdot \vec{\alpha}^\dagger/2} e^{-\vec{z} \cdot \vec{\alpha}} e^{\vec{z}^* \cdot \vec{\alpha}^\dagger} |0,0\rangle \quad (1.30)$$

Equations (1.22), (1.23) and Eq. (1.14) are now expressed in vector form

$$\vec{\alpha} |\vec{z}\rangle = |\vec{z}\rangle \quad \langle \vec{z} | \vec{\alpha}^\dagger = \langle \vec{z} | \vec{z}^* \quad (1.31)$$
All the remaining properties described before for the single-component coherent states are trivially generalized to the two-component case, and moreover to complex vectors of dimension larger than two. In particular, the overlap between two coherent state vectors is, from Eq. (1.16)

$$|\langle \vec{z}|\vec{z}' \rangle| = e^{-\vec{r} \cdot \vec{r}'/2 + \vec{s} \cdot \vec{s}' - \vec{s} \cdot \vec{s}'/2}$$

(1.33)

and the resolution of the identity, generalizing Eq. (1.17) is

$$\int |\vec{z}\rangle \langle \vec{z}||d\vec{z} = I$$

(1.34)

We want to restrict ourselves to the subspace $|1, 0\rangle = |+\rangle$, $|0, 1\rangle = |-\rangle$, that is to enforce the condition $n_1 + n_2 = 1$. This is equivalent to impose a constraint in the number operator, which can be done by defining the projector

$$\hat{P} = \frac{1}{2\pi} \int_0^{2\pi} d\lambda e^{i\lambda(\hat{n}_1 \cdot \hat{n}_1 - 1)}$$

(1.35)

To illustrate the method, let us consider the action of the projector Eq. (1.35) over a coherent state as defined in Eq. (1.32)

$$\hat{P}|\vec{z}\rangle = e^{-\vec{r} \cdot \vec{r}'/2} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \frac{z_1^{n_1} z_2^{n_2}}{n_1! n_2!} \int_0^{2\pi} \frac{d\lambda}{2\pi} e^{i\lambda(n_1 + n_2 - 1)} |n_1, n_2\rangle$$

$$= e^{-\vec{r} \cdot \vec{r}'/2} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \frac{z_1^{n_1} z_2^{n_2}}{n_1! n_2!} \int_0^{2\pi} \frac{d\lambda}{2\pi} e^{i\lambda(n_1 + n_2 - 1)} |n_1, n_2\rangle$$

(1.36)
Notice that the integral over the parameter $\lambda$ becomes a delta function

$$
\int_0^{2\pi} \frac{d\lambda}{2\pi} e^{i\lambda (n_1 + n_2 - 1)} = \delta_{n_2,1-n_1}
$$

(1.37)

Thus, the expression Eq. (1.36) reduces to

$$
\hat{P} |z\rangle = e^{-z^* z/2} \sum_{n=0,1} \frac{z_1^n z_2^{1-n}}{n! (1-n)!} |n, 1-n\rangle = e^{-z^* z/2} (z_1 |+\rangle + z_2 |\rangle)
$$

(1.38)

From Eq. (1.36), it is also straightforward to obtain the equivalent representation

$$
\hat{P} |z\rangle = \int_0^{2\pi} \frac{d\lambda}{2\pi} e^{-i\lambda} |e^{i\lambda} z\rangle
$$

(1.39)

In further applications presented in this thesis, we will consider the representation of spin chains of length $N$ in terms of this formalism. The generalization of the previous results to this case relays in the property that the Hilbert space $\mathcal{H}_N$ of the many-body system is constructed as the tensor product of the single-particle Hilbert space $\mathcal{H}$, i.e. $\mathcal{H}_N = \mathcal{H} \otimes \mathcal{H} \ldots \otimes \mathcal{H}$ [15]. Thus, a many-body vector in the occupation number representation is given by

$$
|\{n\}\rangle = |n_1\rangle \otimes |n_2\rangle \ldots \otimes |n_N\rangle
$$

$$
\equiv |\hat{n}_1, \hat{n}_2, \ldots, \hat{n}_N\rangle
$$

(1.40)

where for each site index $1 \leq j \leq N$, we define the local occupation number vector as in Eq. (1.29), $|\hat{n}_j\rangle = |n_1^j, n_2^j\rangle$. Correspondingly, we add a site index to the two-component creation and annihilation operators in Eq. (1.28), $\hat{a}_\alpha^j(j)$, $\hat{a}_\alpha(j)$, with
\( \alpha = 1, 2 \) as before. The commutation relations for this set of operators generalize the ones stated in Eq. (1.28), where operators defined at different sites commute,

\[
[\hat{a}_\alpha(j), \hat{a}_\beta(j')] = \delta_{\alpha,\beta} \delta_{j,j'} \quad [\hat{a}_\alpha(j), \hat{a}_\beta(j')] = [\hat{a}_\alpha^\dagger(j), \hat{a}_\beta^\dagger(j')] = 0 \quad (1.41)
\]

The properties stated in Eq. (1.2) now become

\[
\hat{a}_\alpha^\dagger(j) |\{\hat{n}\}\rangle = |(n_1^1, n_1^2), \ldots, (n_\alpha^i, n_\alpha^j), \ldots, (n_\beta^N, n_\beta^N)\rangle \\
\hat{a}_\alpha(j) |\{\hat{n}\}\rangle = n_\alpha^j |(n_1^1, n_1^2), \ldots, (n_\alpha^i - 1, n_\alpha^j), \ldots, (n_\beta^N, n_\beta^N)\rangle \quad (1.42)
\]

An equivalent construction is obtained from the tensor product of single-particle coherent states

\[
|\{\tilde{z}\}\rangle = |z_1\rangle \otimes |z_2\rangle \ldots \otimes |z_N\rangle \equiv |\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_N\rangle \quad (1.43)
\]

Here, we defined \(|\tilde{z}_j\rangle = |z_1(j), z_2(j)\rangle\) as the single-site coherent states, for \(1 \leq j \leq N\).

A many-body coherent state as in Eq. (1.43) is created from the vacuum \(|\{0\}\rangle\) by the action of a product of single-particle displacement operators as defined in Eq. (1.30)

\[
|\{\tilde{z}\}\rangle = \prod_{j=1}^{N} \hat{D}(\tilde{z}_j)|\{0\}\rangle = e^{-\sum_{j=1}^{N} z_j^2/2} e^{\sum_{j=1}^{N} z_j \hat{a}_j} e^{-\sum_{j=1}^{N} \hat{a}_j^\dagger z_j}|\{0\}\rangle \quad (1.44)
\]

The inner product between two many-body coherent states is

\[
\langle \{\tilde{z}\} | \{\tilde{z}'\} \rangle = \prod_{j=1}^{N} \langle \tilde{z}_j | \tilde{z}_j' \rangle = e^{\sum_{j=1}^{N} (-z_j^2 \hat{z}_j^2/2 + \hat{z}_j \hat{z}_j^* z_j^2 \hat{z}_j^* /2)} \quad (1.45)
\]
The resolution of the identity in terms of the many-body coherent states is

\[
I = \prod_{j=1}^{N} \int |z_j\rangle \langle z_j| \frac{d^2 z_j}{\pi^2} = \prod_{j=1}^{N} \prod_{\alpha=1,2} \int |z_{j,\alpha}\rangle \langle z_{j,\alpha}| \frac{d^2 z_{j,\alpha}}{\pi}
\]  

(1.46)

1.2.2 Path Integrals in the coherent states representation

We will introduce the path integral representation of the time evolution operator, and then for the partition function, following the approach in [15]. For a system evolving under the quantum Hamiltonian \( \hat{H}(\{\hat{a}\},\{\hat{a}\}) \), let us calculate the matrix element of the time-evolution operator

\[
\mathcal{U}(\{z_f\}, t_f; \{z_i\}, t_i) = \langle \{z_f\}| e^{-\frac{i}{\hbar} \hat{H}(t_f-t_i)} |\{z_i\}\rangle
\]  

(1.47)

We split the finite time interval into a large number \( M \) of time slices \( \epsilon = (t_f - t_i)/M \), and take the limit \( M \to \infty, \epsilon \to 0 \),

\[
e^{-\frac{i}{\hbar} \hat{H}(t_f-t_i)} = \lim_{M \to \infty} \left( e^{-\frac{i}{\hbar} \epsilon \hat{H}} \right)^M
\]  

(1.48)

Between each time-slice factor in Eq. (1.48), we insert a coherent-state resolution of the identity, as defined in Eq. (1.48), to obtain

\[
\mathcal{U}(\{z_f\}, t_f; \{z_i\}, t_i) = \lim_{M \to \infty} \prod_{k=1}^{M-1} \prod_{j=1}^{N} \int \frac{d^2 z_{j,k}}{\pi^2} \prod_{k=1}^{M} \langle \{z_k\}| e^{-\frac{i}{\hbar} \epsilon \hat{H}} |\{z_{k-1}\}\rangle
\]  

(1.49)
Because \( \epsilon \) is a small parameter, we have

\[
\langle \{ \tilde{z}_k \} | e^{-\frac{1}{\hbar} \hat{H}} | \{ \tilde{z}_{k-1} \} \rangle \simeq \langle \{ \tilde{z}_k \} | 1 - \frac{i}{\hbar} \epsilon \hat{H}(\{ \tilde{a}^\dagger \}, \{ \tilde{a} \}) | \{ \tilde{z}_{k-1} \} \rangle \\
= \langle \{ \tilde{z}_k \} | \{ \tilde{z}_{k-1} \} \rangle \left( 1 - \frac{i}{\hbar} \epsilon : H(\{ \tilde{z}_k \}, \{ \tilde{z}_{k-1} \}) : \right) + O(\epsilon^2) \\
\simeq e^{-\sum_{j=1}^{N} \left( \frac{1}{2\epsilon} \tilde{x}_{j,k} \tilde{x}_{j,k-1} + \frac{1}{2\epsilon} \tilde{y}_{j,k-1} \tilde{y}_{j,k} \right)} e^{-\frac{1}{\hbar} \epsilon : H(\{ \tilde{z}_k \}, \{ \tilde{z}_{k-1} \}) :}.
\]

(1.50)

where : \( H(\{ \tilde{z}_k \}, \{ \tilde{z}_k \}) \) : has been normal-ordered. Here, we used

\[
\frac{\langle \{ \tilde{z}_k \} | \hat{H}(\{ \tilde{a}^\dagger \}, \{ \tilde{a} \}) | \{ \tilde{z}_{k-1} \} \rangle}{\langle \{ \tilde{z}_k \} | \{ \tilde{z}_{k-1} \} \rangle} = : H(\{ \tilde{z}_k \}, \{ \tilde{z}_{k-1} \}) :.
\]

(1.51)

Substituting this result in Eq. (1.49), we have

\[
\mathcal{U}(\{ \tilde{z}_j \}, t_j; \{ \tilde{z}_i \}, t_i) = \lim_{M \to \infty} \prod_{k=1}^{M-1} \prod_{j=1}^{N} \int \frac{d\tilde{x}_{j,k}}{\pi^2} e^{\frac{\sum_{j=1}^{N} \left[ -\frac{1}{2} \tilde{x}_{j,k}^2 - \frac{1}{2} \tilde{y}_{j,k-1}^2 \right]}{\epsilon}} e^{-\frac{1}{\hbar} \epsilon : H(\{ \tilde{z}_k \}, \{ \tilde{z}_{k-1} \}) :}.
\]

(1.52)

This expression, in the limit \( M \to \infty, \epsilon \to 0 \) may be written symbolically as

\[
\mathcal{U}(\{ \tilde{z}_j \}, t_j; \{ \tilde{z}_i \}, t_i) = \int_{(\tilde{z}(t_f))=(\tilde{z}_f)} \mathcal{D} \tilde{z} \mathcal{D} \tilde{z}^* e^{\frac{1}{\hbar} \int_{t_i}^{t_f} dt \left( \frac{\partial}{\partial t} \sum_{j=1}^{N} \left[ \tilde{x}_j \frac{\partial \tilde{x}_j}{\partial t} - \frac{\partial \tilde{y}_j}{\partial t} \frac{\partial \tilde{y}_j}{\partial t} \right] - H(\{ \tilde{x} \}, \{ \tilde{z} \}) \right)}.
\]

(1.53)

A similar procedure is employed to obtain the partition function in terms of a path integral. By definition

\[
Z = \text{Tr} e^{-\beta \hat{H}} = \int \frac{d\tilde{z}}{\pi^2} \langle \{ \tilde{z} \} | e^{-\beta \hat{H}} | \{ \tilde{z} \} \rangle
\]

(1.54)
As in the case of the real time integral, we split the inverse temperature interval into \( M \) slices, \( \epsilon = \beta / M \), with \( M \to \infty \). By inserting a coherent states resolution of the identity between every slice, we obtain an expression analogous to Eq. (1.52)

\[
Z = \lim_{M \to \infty} \prod_{k=1}^{M} \prod_{j=1}^{N} \int \frac{d\bar{z}_{j,k}}{\pi^2} \\
\times e^{\sum_{k=1}^{M} \left( \sum_{j=1}^{N} \left[ -\frac{1}{2} \frac{\bar{z}_{j,k} - \bar{z}_{j,k-1}}{\epsilon} + \frac{1}{6} \frac{\bar{z}_{j,k} - \bar{z}_{j,k-1}}{\epsilon} \cdot \bar{z}_{j,k-1} \right] - \epsilon \cdot H(\{\bar{z}_{k}\}, \{\bar{z}_{k-1}\}) \right)}.
\]

In the continuum limit \( M \to \infty, \epsilon \to 0 \), we obtain the formal expression

\[
Z = \int_{\{\bar{z}(0)\} = \{\bar{z}(\beta)\}} D\bar{z} D\bar{z} e^{-\int_{0}^{\beta} d\tau \left( \frac{1}{2} \sum_{j=1}^{N} \left[ \bar{z}_{j} \cdot \frac{\partial H}{\partial \bar{z}_{j}} - \frac{\partial \bar{z}_{j}}{\partial \tau} \cdot \bar{z}_{j} \right] + H(\{\bar{z}\}, \{\bar{z}\}) \right)}.
\]
Chapter 2
Quasispecies theory for multiple-peak fitness landscapes

This chapter was communicated in the article D. B. Saakian, E. Muñoz, C.-K. Hu and M. W. Deem, Phys. Rev. E 73, 041913 (2006).

Abstract

We use a path integral representation to solve the Eigen and Crow-Kimura molecular evolution models for the case of multiple fitness peaks with arbitrary fitness and degradation functions. In the general case, we find that the solution to these molecular evolution models can be written as the optimum of a fitness function, with constraints enforced by Lagrange multipliers and with a term accounting for the entropy of the spreading population in sequence space. The results for the Eigen model are applied to consider virus or cancer proliferation under the control of drugs or the immune system.

2.1 Introduction

Methods of statistical physics have been applied successfully to understand phase transitions of various physical systems in the past few decades [60, 61, 62, 63, 64, 65, 66]. Molecular models of biological evolution also exhibit phase transition behaviors and such models have received much attention in recent decades [11, 12, 10, 67,
In particular, the notion of adaptive evolution on a fitness (replication rate) landscape has proved very fruitful [11, 12, 10, 67]. In the last decade, several exact results [23, 25, 26, 24, 28] have been derived for the Eigen [11, 12] and Crow-Kimura [10, 67, 23] quasispecies models of biological evolution and their generalizations [28] for a single peak fitness landscape.

However, it is widely accepted that biological evolution proceeds on a rugged fitness landscape [69, 70]. In this paper, we consider a multiple peak replication rate landscape as a means to model a rugged fitness landscape. To date, there are few rigorous results for multiple peak fitness landscapes. Such results begin to make the connection with the biologically-relevant case of a rugged fitness landscape. We derive here the exact error thresholds by means of a path integral representation for both the Eigen and Crow-Kimura mutation-selection schemes with an arbitrary number of replication rate peaks.

We first generalize the Crow-Kimura model to the multiple peak case. The solution of the one-peak version of this model, where the replication rate is a function of Hamming distance from one configuration, was provided by a path integral representation in [25, 26]. We provide here the solution to this model for K peaks, where the replication rate is a function of the Hamming distance from K configurations, again by means of a path integral. We find that the mean distances from the peaks maximize the replication rate, with constraints provided by Lagrange multipliers, and with
an additional term that represents the entropy of the population in sequence space. Explicit solutions to this maximization task are given for the two-peak case.

We then generalize and solve the continuous-time Eigen model for $K$ peaks, where the replication and degradation rates are functions of Hamming distances from $K$ configurations. A solution of the discrete-time, single-peak Eigen model, which in a sense interpolates between the Crow-Kimura and continuous time Eigen model [71], was provided in [72]. We solve here the continuous-time Eigen model for $K$ peaks, again by means of a path integral representation. The mean distances from the peaks maximize an excess replication rate with an effective mutation rate, with constraints provided by Lagrange multipliers, and with an additional term that represents the entropy of the population in sequence space under the effective mutation rate.

The Eigen model was first developed to study viral evolution [11], and we use our solution of the two-peak Eigen model to consider viral propagation in the presence of either immune system suppression or an antiviral drug. The preferred viral genome exists at one point in genome space. Conversely, the drug or immune system suppresses the virus most strongly at some other point in genome space. These two points in genome space are the two peaks of the model. The viral growth rate and the suppression rate both decrease with the Hamming distance away from these two unique points.

The rest of the paper is organized as follows: In Sec. 2.3 we describe the gener-
alization of the Crow-Kimura, or parallel, model [10] to multiple peaks and provide a solution of this model for an arbitrary replication rate function that depends on distances from K peaks. In Sec. 2.4, we describe the Eigen model and provide a solution for arbitrary replication and degradation rate functions that depend on distances from K peaks. In Sec. 2.5, we use the Eigen model with two peaks to address the interaction of the immune system with a drug. We consider both adaptable viruses and the original antigenic sin phenomena [73]. We also consider tumor suppression by the immune system. We discuss these results and conclude in Sec. 2.6. We provide a derivation of the path integral representation of the continuous-time Eigen model in the Appendix.

2.2 Crow-Kimura model with multiple peaks

Here we first briefly introduce the Crow-Kimura model [10] and its quantum spin version [22] so that it is easier to understand its generalizations to be studied in the present paper. In the Crow-Kimura model, any genotype configuration $i$ is specified a sequence of $N$ two-valued spins $s_n = \pm 1$, $1 \leq n \leq N$. We denote such configuration $i$ by $S_i \equiv (s^i_1, \ldots, s^i_N)$. That is, as in [12], we consider $s_n = +1$ to represent the purine ($A, G$) and $s_n = -1$ to represent the pyrimidine ($C, T$). Two-values spin models have also been used to study long-range correlations in DNA sequences [74] and DNA unzipping [75, 76] and valuable results have been obtained. The difference between two configurations $S_i$ and $S_j \equiv (s^j_1, \ldots, s^j_N)$ is described by the Hamming distance
\[ d_{ij} = (N - \sum_n s_i^n s_j^n)/2, \] which is the number of different spins between \( S_i \) and \( S_j \).

The relative frequency \( p_i \) of the configuration \( S_i, 1 \leq i \leq 2^N \), satisfies

\[
\frac{dp_i}{dt} = p_i \left( r_i - \sum_{j=1}^{2^N} r_j p_j \right) + \sum_{j=1}^{2^N} \mu_{ij} p_j \tag{2.1}
\]

Here \( r_i \) is the replication rate or the number of offspring per unit period of time (the fitness) of the sequence \( S_i \), and \( \mu_{ij} \) is the mutation rate to move from sequence \( S_i \) to sequence \( S_j \) per unit period of time. In the Crow-Kimura model, only single base mutations are allowed: \( \mu_{ij} = \gamma \Delta(d_{ij} - 1) - N \gamma \Delta(d_{ij}) \). Here \( \Delta(n) \) is the Kronecker delta function.

The fitness of an organism with a given genotype is specified in the Crow-Kimura model by the choice of the replication rate function \( r_i \), which is a function of the genotype: \( r_i = f(S_i) \). It has been observed \([22, 23]\) that the system Eq. (2.1), with \( r_i \equiv f(s_1^i, \ldots, s_N^i) \) evolves according to a Schrödinger equation in imaginary time with the Hamiltonian

\[
-H = \gamma \sum_{n=1}^{N} (\sigma_n^x - 1) + f(\sigma_1^x, \ldots, \sigma_N^x) \tag{2.2}
\]

Here \( \sigma^x \) and \( \sigma^z \) are the Pauli matrices. The mean replication rate, or fitness, of the equilibrium population of genotypes is calculated as (see Reference \[12]\));

\[
\lim_{t \to \infty} \sum_i p_i(t) r_i = \lim_{\beta \to \infty} \frac{1}{\beta} \ln Z \equiv \lim_{\beta \to \infty} \ln \text{Tr} \exp(-\beta \hat{H}) \tag{2.3}
\]

In this way, it is possible to find the phase structure and error threshold of the
equilibrium population. In the generalized setting, the Crow-Kimura model is often called the parallel model.

2.2.1 The parallel model with two peaks

We consider two peaks to be located at two configurations $v^1_n, v^2_n$, $1 \leq n \leq N$, where $v^1_n = \pm 1$, $v^2_n = \pm 1$, and the two configurations have $l$ common spins: $\sum_{n=1}^{N} v^1_nv^2_n = 2l - N$. The value of $l$ determines how close the two peaks are in genome space. Now the replication rate $r_i$ of configuration $S_i$ is a function of the Hamming distances to each peak,

$$r_i = f(2L_1/N - 1, 2L_2/N - 1), \tag{2.4}$$

where $\sum_{n=1}^{N} v^1_ns_n = 2L_1 - N$ and $\sum_{n=1}^{N} v^2_ns_n = 2L_2 - N$.

Due to the symmetry of the Hamiltonian, the equilibrium frequencies are a function only of the distances from the two peaks: $p_i \equiv p(L_1, L_2)$. We define the factors $x_{\alpha_1, \alpha_2}$ that describe the fraction of spins a configuration $S_i$ has in common with the spins of configurations $v^1, v^2$. In particular, we define the fraction of spins that are equal to $\alpha_k$ times the value in peak configuration $v^k$. For $K$ peaks, the general definition is $x_{\alpha_1, \ldots, \alpha_K} = \frac{1}{N} \sum_{n=1}^{N} \delta[s_n, \alpha_1v^1_n] \cdots \delta[s_n, \alpha_Kv^K_n]$. For the two peak case, $x_{\alpha_1, \alpha_2}$ satisfy the relations $x_++x_-+x_+-x_- = 1$. $x_++x_- = L_1/N$, $x_++x_- = l/N$. Thus these factors are related to the distances from the configuration to each peak.
and to the distance between the peaks;

\[
x_{+-}(L_1, L_2) = \frac{(L_1 - L_2 + N - l)}{(2N)},
\]

\[
x_{++}(L_1, L_2) = \frac{(L_1 + L_2 - N + l)}{(2N)},
\]

\[
x_{-+}(L_1, L_2) = \frac{(-L_1 - L_2 + N + l)}{(2N)},
\]

\[
x_{-+}(L_1, L_2) = \frac{(-L_1 + L_2 + N - l)}{(2N)}
\]

(2.5)

With these factors, we find the following equation for the total probability at a given value of \(L_1\) and \(L_2\), \(P(L_1, L_2)\):

\[
\frac{d}{dt}P(L_1, L_2) = f \left( \frac{2L_1}{N} - 1, \frac{2L_2}{N} - 1 \right) P(L_1, L_2) - \gamma NP(L_1, L_2)
\]

\[
+ \gamma \sum_{\alpha_1 = \pm, \alpha_2 = \pm} N x_{\alpha_1, \alpha_2}(L_1 + \alpha_1, L_2 + \alpha_2) P(L_1 + \alpha_1, L_2 + \alpha_2)
\]

\[
- P(L_1, L_2) \sum_{L_1', L_2'} f \left( \frac{2L_1'}{N} - 1, \frac{2L_2'}{N} - 1 \right) P(L_1', L_2')
\]

(2.6)

Only the values of \(L_1\) and \(L_2\) satisfying the conditions \(0 \leq L_1 \leq N, |L_1 + L_2 - N| \leq l, |L_1 - L_2| \leq N - l\) are associated with nonzero probabilities. Equation (2.6) can be solved numerically to find the error threshold and the average Hamming distance of the population to the two peaks. In the next section we solve this equation, and its generalization to \(K\) peaks, analytically.
2.2.2 Exact solution of the K peak case by a path integral representation

We consider the case of $K$ peaks. We consider the replication rate to depend only on the distances from each peak

$$r_i = f \left( \frac{2L_i}{N} - 1, \ldots, \frac{2L_K}{N} - 1 \right) \equiv N f_0(u_1, \ldots, u_K) \tag{2.7}$$

where $N u_k = \sum_{n=1}^{N} v_n^k s_n = 2L_k - N, 1 \leq k \leq K$. The observable value $\langle u_k \rangle$ is called the surface magnetization [21], or surplus [22], for peak $k$.

Characterization of the fitness function that depends on $K$ peaks through the $K$ values of $u_k$ requires more than the $K(K-1)/2$ Hamming distances between the peaks. It proves convenient to define the $2^K$ parameters $y_{\alpha_1, \ldots, \alpha_K} = y_i, 1 \leq i \leq 2^K$. These are defined by $y_i = (1/N) \sum_{n=1}^{N} \prod_{k=1}^{K} \delta(\alpha_{ik}, v_n^k)$. Here $\alpha_{ik}$ is the set of indices $\alpha_1, \ldots, \alpha_K$, and $\alpha_{ik} = \alpha_k$ in the $i$th set of indices $\alpha_1, \ldots, \alpha_K$. The introduction of the $2^K$ parameters $y_i$ is one principle point of this paper.

The Suzuki-Trotter method has been applied in [26, 24] to convert the quantum partition function for a single peak model into a classical functional integral. While calculating $Z = \text{Tr} \exp[-\beta \hat{H}]$, intermediate spin configurations are introduced. We find $Z$ is a functional integral, with the integrand involving a partition function of a spin system in the 2D lattice. In the spin system, there is a nearest-neighbor interaction in horizontal direction and a mean-field-like interaction in the vertical direction. This spin system partition function was evaluated in [26, 24] under the
assumption that the field values are constant. A path integral representation of the
discrete time Eigen model, which is quite similar to the parallel model, was introduced
by Peliti [72].

Here, we generalize this procedure to \( K \) peaks and calculate the time-dependent
path integral and Ising partition function. Since the replication rate is a function of \( K \)
distances, the functional integral is over \( K \) fields that represent the \( K \) magnetizations.
The path integral form of the partition function is

\[
Z = \int \mathcal{D}M_k \mathcal{D}H_k \exp \left\{ N \int_0^\beta d\beta' \left[ f_0[M_1(\beta'), \ldots, M_K(\beta')] \right. \\
\left. - \sum_{k=1}^K H_k(\beta') M_k(\beta') - \gamma \right] + N \sum_{i=1}^{2^K} y_i \ln Q_i \right\}
\]  

(2.8)

where

\[
Q_i = \text{Tr} \hat{T} e^{\int_0^\beta d\beta' \left[ \sigma^2 \gamma + \gamma^2 \sum_{k=1}^K \alpha_{ik} H_k(\beta') \right]}
\]  

(2.9)

Here \( \beta = t \) is the large time to which Eq. (2.1) is solved, and the operator \( \hat{T} \) denotes
time ordering [24], discussed in the Appendix in the context of the Eigen model.

Using that \( N \) is large, we take the saddle point. Considering \( \delta \ln Z / \delta M_k(\beta') = 0 \) and
\( \delta \ln Z / \delta H_k(\beta') = 0 \), we find \( M_k(\beta') \) and \( H_k(\beta') \) independent of \( \beta' \) is a solution. At
long time, therefore, the mean replication rate, or fitness, per site becomes

\[
\ln \frac{Z}{\beta N} = \max_{H_k, M_k} \left\{ f_0(M_1, \ldots, M_K) - \sum_{k=1}^K H_k M_k - \gamma + \sum_{i=1}^{2^K} y_i \left[ \gamma^2 + \left( \sum_{k=1}^K \alpha_{ik} H_k \right)^2 \right]^{1/2} \right\}
\]  

(2.10)
We take the saddle point limit in $H_k$ to find

$$
M_k = \sum_{i=1}^{2^k} y_i \alpha_{ik} \frac{\sum_{k'=1}^{K} \alpha_{ik'} H_{k'}}{\sqrt{\gamma^2 + \left(\sum_{k'=1}^{K} \alpha_{ik'} H_{k'}\right)}}
$$

(2.11)

We note that the observable, surface magnetization given by $\langle u_k \rangle$, is not directly accessible in the saddle point limit, but is calculable from the mean replication rate [22].

In the one peak case one defines the observable surface magnetization for a monotonic fitness function as follows [23]: one solves the equation $f_0(\langle u \rangle) = (\ln Z)/(\beta N)$. For multiple peaks, we use this same trick, considering a symmetric fitness function and assuming $\langle u_1 \rangle = \langle u_2 \rangle \ldots = \langle u_K \rangle$.

### 2.2.3 Explicit results for the two peak case

For clarity, we write the expression for the case of two peaks. In this case, $y_{++} + y_{--} = (1 + m)/2$ and $y_{+-} + y_{-+} = (1 - m)/2$, where $m = (2l - N)/N$. We solve Eq. (2.11) for the fields $H_k$ and put the result into Eq. (2.10). We find that for a pure phase, the bulk magnetizations maximize the function

$$
\frac{\ln Z}{N \beta} = f_0(M_1, M_2) + \frac{\gamma}{2} \sqrt{(1 + m)^2 - (M_1 + M_2)^2} - \frac{\gamma}{2} \sqrt{(1 - m)^2 - (M_1 - M_2)^2} - \gamma
$$

(2.12)
with the constraints

\[-1 \leq M_1 \leq 1, -1 \leq M_2 \leq 1,\]
\[-(1 + m) \leq M_1 + M_2 \leq 1 + m,\]
\[-(1 - m) \leq M_1 - M_2 \leq 1 - m,\]  
(2.13)

In the case of a quadratic replication rate, \( f_0 = k_1 M_1^2 + k_2 M_2^2 + k_3 M_1 M_2 \), Eq. (2.11) becomes

\[
\frac{\ln Z}{N \beta} = k_1 M_1^2 + k_2 M_2^2 + k_3 M_1 M_2 + \frac{\gamma}{2} \sqrt{(1 + m)^2 - (M_1 + M_2)^2}
\]
\[+ \frac{\gamma}{2} \sqrt{(1 - m)^2 - (M_1 - M_2)^2} - \gamma\]  
(2.14)

with the constraints of Eq. (2.12).

As an example, we consider the replication rate function \( f_0 = k(M_1^2 + M_2^2 + M_1 M_2) \).

When \( m > 0 \), and the two peaks are within a Hamming distance of \( N/2 \) of each other, there is a solution with \( M_1 = M_2 = M \) for which

\[
\frac{3kM^2}{2\gamma} + \left[ \left( \frac{1 + m}{2} \right)^2 - M^2 \right]^{1/2} - \frac{1 + m}{2} = \frac{k}{2\gamma} (\langle u_1 \rangle^2 + \langle u_2 \rangle^2 + \langle u_1 \rangle \langle u_2 \rangle)
\]  
(2.15)

where the observable, surface magnetization, is given by \( \langle u_i \rangle = \langle 2L_i/N - 1 \rangle \). We find

\[M_1 = M_2 = M = \sqrt{(1 + m)^2/4 - \gamma^2/(9k^2)}\]  
(2.16)

We have for the mean replication rate, or fitness, per site

\[
\frac{\ln Z}{N \beta} = \frac{3k}{2} \left( \frac{1 + m}{2} - \frac{\gamma}{3k} \right)^2
\]  
(2.17)
Table 2.1 Comparison between the analytical formulas Eqs. (2.18), (2.22) for the two peak landscape in the parallel model and results from a direct numerical solution of the system of differential equations, Eq. (2.6), for sequences of length \( N = 1000 \), with \( p(L_1, L_2, t = 0) = \delta(L_1, N)\delta(L_2, t) \).

<table>
<thead>
<tr>
<th>m</th>
<th>k</th>
<th>\langle u_1 \rangle</th>
<th>\langle u_2 \rangle</th>
<th>\langle u_1 \rangle_{\text{analytic}}</th>
<th>\langle u_2 \rangle_{\text{analytic}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.93</td>
<td>3.0</td>
<td>0.85</td>
<td>0.85</td>
<td>0.853</td>
<td>0.853</td>
</tr>
<tr>
<td>0.93</td>
<td>2.0</td>
<td>0.80</td>
<td>0.80</td>
<td>0.798</td>
<td>0.798</td>
</tr>
<tr>
<td>0.7</td>
<td>3.0</td>
<td>0.74</td>
<td>0.74</td>
<td>0.738</td>
<td>0.738</td>
</tr>
<tr>
<td>0.7</td>
<td>2.0</td>
<td>0.68</td>
<td>0.68</td>
<td>0.683</td>
<td>0.683</td>
</tr>
<tr>
<td>-0.7</td>
<td>3.0</td>
<td>0.52</td>
<td>-0.52</td>
<td>0.517</td>
<td>-0.517</td>
</tr>
<tr>
<td>-0.7</td>
<td>3.0</td>
<td>0.35</td>
<td>-0.35</td>
<td>0.35</td>
<td>-0.35</td>
</tr>
<tr>
<td>-0.93</td>
<td>3.0</td>
<td>0.63</td>
<td>-0.63</td>
<td>0.631</td>
<td>-0.631</td>
</tr>
<tr>
<td>-0.93</td>
<td>2.0</td>
<td>0.46</td>
<td>-0.46</td>
<td>0.465</td>
<td>-0.465</td>
</tr>
</tbody>
</table>
so that [22]

$$\langle u_1 \rangle = \langle u_2 \rangle = \frac{1 + m}{2} - \frac{\gamma}{3k} \quad (2.18)$$

When \( m < 0 \), and the two peaks are greater than a Hamming distance of \( N/2 \) of each other, there is a solution with \( M_1 = -M_2 = M \) for which

$$\frac{kM^2}{2\gamma} + \left[ \left( \frac{1 - m}{2} \right)^2 - M^2 \right]^{1/2} = \frac{1 - m}{2} = \frac{k}{2\gamma} (\langle u_1 \rangle^2 + \langle u_2 \rangle^2 + \langle u_1 \rangle \langle u_2 \rangle) \quad (2.19)$$

One solution is

$$M_1 = -M_2 = \sqrt{(1 - m)^2/4 - \gamma^2/k^2}, \quad (2.20)$$

which gives for a mean replication rate, or fitness, per site

$$\frac{\ln Z}{N\beta} = \frac{k}{2} \left( \frac{1 - m}{2} - \frac{\gamma}{k} \right)^2 \quad (2.21)$$

so that [22]

$$\langle u_1 \rangle = -\langle u_2 \rangle = \frac{1 - m}{2} - \frac{\gamma}{k} \quad (2.22)$$

Numerical solution is in agreement with our analytical formulas, as shown in Table 2.2.3.
2.3 Eigen model with multiple peaks

2.3.1 Exact solution by a path integral representation

In the case of the Eigen model, the system is defined by means of replication rate functions, \( r_j \), as well as degradation rates, \( D_j \),

\[
\frac{dp_i}{dt} = \sum_{j=1}^{2^N} [Q_{ij}r_j - \delta_{ij}D_j]p_j - p_i \left[ \sum_{j=1}^{2^N} (r_j - D_j)p_j \right] \tag{2.23}
\]

Here, the frequencies of a given genome, \( p_i \), satisfy \( \sum_{i=1}^{2^N} p_i = 1 \). The transition rates are given by \( Q_{ij} = q^N - d_{ij}(1-q)^{d_{ij}} \), with \( d_{ij} \) being the Hamming distance between two genomes \( S_i \) and \( S_j \). The parameter \( \gamma = N(1-q) \) describes the efficiency of mutations. We take \( \gamma = O(1) \). As in Eq. (2.6), we take the replication and degradation rate to depend only on the spin state, in particular on the Hamming distances from each peak: \( r_i = f(S_i) \) and \( D_i = D(S_i) \) where

\[
f(S) = Nf_0(u_1, \ldots, u_K), \quad D(S) = Nd_0(u_1, \ldots, u_K) \tag{2.24}
\]

We find a path integral representation of the partition function for the Eigen model for the \( K \) peak case in the limit of long time as

\[
Z = \int DM_kDH_kDm_0Dh_0 \exp \left\{ N \int_0^\beta d\beta' \left[ f_0(M_1, \ldots, M_K)e^{-\gamma(1-m_0)} - h_0m_0 - \sum_{k=1}^K H_kM_k - d_0(M_1, \ldots, M_K) \right] + N \sum_{i=1}^{2^K} y_i \ln Q_i \right\} \tag{2.25}
\]

where

\[
Q_i = \text{Tr} \hat{T} e^{f_0(\beta') + \gamma \sum_{k=1}^K \sigma a_k H_k(\beta')} \tag{2.26}
\]
The $M_k$ are the values of the magnetization, and $\gamma m_0$ is an effective mutation rate. This form is derived in the Appendix. Using that $N$ is large, we take the saddle point. As before, we find the mean excess replication rate per site, $f_m = \lim_{t \to \infty} \sum_i p_i(t) \langle r_i - D_i \rangle / N$, from the maximum of the expression for $Z = \text{Tr} \exp(-\beta \hat{H})$. We find $Z \sim \exp(\beta N f_m)$, where

$$f_m = f_0(M_1, \ldots, M_K)e^{-\gamma(1-m_0)} - d_0(M_1, \ldots, M_K). \tag{2.27}$$

Here $m_0, M_k$ are defined through the fields $H_k$,

$$M_k = \sum_{i=1}^{2^K} y_i \frac{\sum_{k'=1}^K \alpha_{ik'} H_{k'}}{\sqrt{h_0^2 + \left(\sum_{k'=1}^K \alpha_{ik'} H_{k'}\right)^2}},$$

$$m_0 = \sum_{i=1}^{2^K} y_i \frac{h_0}{\sqrt{h_0^2 + \left(\sum_{k'=1}^K \alpha_{ik'} H_{k'}\right)^2}} \tag{2.28}$$

We define

$$m_i = \frac{\sum_{k=1}^K \alpha_{ik} H_k}{\sqrt{h_0^2 + \left(\sum_{k=1}^K \alpha_{ik} H_k\right)^2}} \tag{2.29}$$

We thus find $m_0 = \sum_{i=1}^{2^K} y_i \sqrt{1 - m_i^2}$, giving Eq. (2.28).

### 2.3.2 Simple formulas for the two peak case

In the two peak, $K = 2$, case we can define the $m_i$ from Eq. (2.29) from the system

$$M_1 = \frac{1 + m}{2} (m_1 + m_2) + \frac{1 - m}{2} (m_1 - m_2),$$

$$M_2 = \frac{1 + m}{2} (m_1 + m_2) - \frac{1 - m}{2} (m_1 - m_2) \tag{2.30}$$
where \( m \) is the overlap between two peaks and we have defined \( m_1, m_2 \) in terms of the \( m_i \) from Eq. (2.29) by \( m_{++} = -m_{--} = m_1 + m_2 \) and \( m_{+-} = -m_{-+} = m_1 - m_2 \).

We have for the mean excess replication rate per site

\[
 f_m = f_0(M_1, M_2) \exp \left[ -\gamma \left( 1 - \frac{1 + m}{2} \sqrt{1 - (M_1 + M_2)^2/(1 + m)^2} \right) \right] - d_0(M_1, M_2) \tag{2.31}
\]

### 2.3.3 Eigen model with quadratic replication rate without degradation

We apply our results to model qualitatively the interaction of a virus with a drug. In some situations, one can describe the action of a drug against the virus simply as a one peak Eigen model; that is, the replication rate is a function of the Hamming distance from one peak. The virus may increase its mutation rate, and at some mutation rate there is an error catastrophe [77]. Let us define the critical \( \gamma \) for the replication rate function

\[
 f_0(M) = \frac{kM^2}{2} + 1 \tag{2.32}
\]

According to our analytical solution, Eq. (2.27), we consider the maximum of the mean excess replication rate per site,

\[
 f_m = f_0(M) \exp[-\gamma(1 - \sqrt{1 - M^2})] \tag{2.33}
\]

which can also be obtained from Eq. (2.31) by taking \( m = 1 \) and \( M_1 = M_2 \). The error catastrophe occurs and leads to a phase with \( M = 0 \) when \( k < \gamma \). The error
Table 2.2 Comparison between the analytical formula Eq. (2.33) for the quadratic landscape Eq. (2.31) in the Eigen model and results from a direct numerical solution of the system of differential equations Eq. (2.22), for sequences of length $N = 4000$, with $p(u, t = 0) = \delta(u, l)$. We set $\gamma = 5$.

<table>
<thead>
<tr>
<th>$k/\gamma$</th>
<th>$M$</th>
<th>$\langle u \rangle$</th>
<th>$\langle u \rangle_{\text{analytic}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>0.24</td>
<td>0.065</td>
<td>0.068</td>
</tr>
<tr>
<td>1.4</td>
<td>0.31</td>
<td>0.112</td>
<td>0.113</td>
</tr>
<tr>
<td>1.6</td>
<td>0.35</td>
<td>0.146</td>
<td>0.147</td>
</tr>
<tr>
<td>1.8</td>
<td>0.38</td>
<td>0.172</td>
<td>0.172</td>
</tr>
<tr>
<td>2.0</td>
<td>0.41</td>
<td>0.192</td>
<td>0.193</td>
</tr>
</tbody>
</table>

Table 2.2 illustrates the comparison between the analytical formula and numerical solutions for the quadratic landscape in the Eigen model. The threshold for this quadratic case is the same as in the case of the Crow-Kimura model Eq. (2.1). The average of $u$, $\langle u \rangle$, satisfies the equation

$$f_0(M) \exp[-\gamma(1 - \sqrt{1 - M^2})] = f_0(\langle u \rangle) = \frac{k\langle u \rangle^2}{2} + 1 \quad (2.34)$$

This equation gives $\langle u \rangle_{\text{analytic}}$ shown in Table 2.3.3, which are in agreement with numerical solutions.

### 2.4 Biological applications

The Eigen model is commonly used to consider virus or cancer evolution. We here consider an evolving virus or cancer and its control by a drug or the immune system, using the $K = 2$, two-peak version of the Eigen model. To model this situation, we consider there to be an optimal genome for virus replication, and we consider
Figure 2.1  A phase diagram for the interaction of virus and drug, according to the narrow replication advantage model, Eqs. (2.34) and (2.33). We set $b = 3.5$ in the exponential degradation function Eq. (2.34), and $\gamma = 1$. As the drug overlaps more with the virus, a higher viral replication advantage is required for the virus to survive. In the NS phase, the drug eliminates the virus. In the inset is shown the phase diagram for the interaction of an adaptable virus and a drug, according to the flat peak replication advantage model, Eqs. (2.34) and (2.36). We use $M_0 = 0.9$ to represent a broad peak for the virus replication rate.

the replication rate function $f_0(M_1, M_2)$ to depend only on the Hamming distance of the virus or cancer from this preferred genome, $N(1 - M_1)/2$. Conversely, there is another point in genome space that the drug or immune system suppresses most strongly, and we consider the degradation rate function $d_0(M_1, M_2)$ to depend only on the Hamming distance from this point, $N(1 - M_2)/2$. While each of the functions $f_0$ and $d_0$ depends only on one of the two distances, this is a multiple-peak problem, because both distances are needed to describe the evolution of the system.
2.4.1 Interaction of virus with a drug

We first consider a virus interacting with a drug. We model this situation by the Eigen model with one peak in the replication rate function and one peak in the degradation rate function. The virus replicates most quickly at one point in genome space, with the rate at all other points given by a function that depends on the Hamming distance from this one point. That is, in Eq. (2.31) we have

\[
 f_0(M_1, M_2) = \begin{cases} 
 A, & M_1 = 1 \\
 1, & M_1 < 1 
\end{cases} 
\] (2.35)

At another point in genome space, a drug suppresses the virus most strongly. We consider the case of exponential degradation, a generic and prototypical example of recognition [73],

\[
 d_0(M_1, M_2) = e^{-b(1-M_2)} 
\] (2.36)

Applying the multiple-peak formalism, we find two phases. There is a selected, ferromagnetic phase (FM) with \( M_1 = 1, M_2 = m \) and a mean excess replication rate per site

\[
 f_m = Ae^{-\gamma} - e^{-b(1-m)} 
\] (2.37)

There is also a non-selective (NS) phase, with \( M_1 < 1 \). The values of \( M_1 \) and \( M_2 \) in the NS phase are those which maximize Eq. (2.31) given the constraints of Eq. (2.12).
The error threshold corresponds to the situation when the mean excess replication rate of the FM and NS phases are equal. The phase diagram as a function of the optimal replication rate of the virus and the distance between the points of optimal virus growth and optimal virus suppression is shown in Fig. 2.1. The optimal replication rate is \( A \), and the distance between the points of optimal virus growth and optimal virus suppression is \( N - l \), where the parameter \( m \) is defined as \( m = (2l - N)/N \).

As the point in genome space at which the drug is most effective moves toward the point in genome space at which the virus grows most rapidly, the virus is more readily eradicated. Alternatively, one can say that as the point in genome space at which the virus grows most rapidly, a higher replication rate of the virus is required for its survival.

### 2.4.2 Interaction of an adaptable virus with a drug

We now consider a virus that replicates with rate \( A \) when the genome is within a given Hamming distance form the optimal genome and with rate 1 otherwise. That is, in Eq. (2.31) we have

\[
f_0(M_1, M_2) = \begin{cases} 
A, & M_0 \leq M_1 \leq 1 \\
1, & -1 \leq M_1 < M_0 
\end{cases} \tag{2.38}
\]

where \( M_0 > 0 \) and is close to 1. We consider the suppression of the virus by the drug as expressed in Eq. (2.34). There is again a ferromagnetic (FM) phase with a successful selection. In the FM phase, one has \( M_0 \leq M_1 < 1 \). The evolved values of
Figure 2.2 A phase diagram corresponding to the original antigenic sin model is shown. The degradation function (shown in the inset) is chose to closely reproduce the binding constant behavior in [73], with a limiting degradation rate of $d_0(M_1, M_2 = -1) = 1$. We use $\gamma = 1$. The virus replication advantage required to escape immune system control is a non-monotonic function of the overlap of the vaccine with the virus.

$M_1$ and $M_2$ maximize

$$ f_m = A \exp \left[ -\gamma \left( 1 - \frac{1 + m}{2} \sqrt{1 - (M_1 + M_2)^2/(1 + m)^2} \right) - d_0(M_1, M_2) \right] (2.39) $$

There is also a NS phase where the virus has been driven off its advantaged peak, $M_1 < M_0$. In this case, one seeks a maximum of Eq. (2.30) with $f_0 = 1$ via $M_1$ and $M_2$ in the range $-1 \leq M_1 \leq M_0$, $-1 \leq M_2 \leq 1$, subject to the constraints of Eq. (2.12).

A phase diagram for this case is shown in the inset of Fig. 2.2. The broader range of virus fitness landscape allows the virus to survive under a greater drug pressure
Figure 2.3  A phase diagram corresponding to the immune control of cancer is shown. We use $\gamma = 1$, $B = 1$, $M_0 = 0.9$, and $M_b = 0.54$. The cancer replication advantage required to escape immune system control decreases with the overlap of the cancer with the self. Three of the four selective and one of the two non-selective phases are present for the chosen parameters.

in model Eq. (2.37) versus Eq. (2.33). That is, as the drug overlaps more with the favored virus genotypes, the adaptable virus is still able to persist due to the greater range of genotype space available in the FM phase. For such an adaptable virus, a more specific, multidrug cocktail might be required for eradication. A multidrug cocktail provides more suppression in a broader range of genome space, so that the adaptable virus may be eradicated under a broader range of conditions.

2.4.3 Original antigenic sin

The immune system, in the context of this theoretical analysis, acts much like a drug, as a natural protection against death by infection. Prior exposure, such as
vaccination, typically increases the immune control of a virus. In some cases, the immune control of a virus is non-monotonic in the overlap between the vaccine and the virus [73]. This phenomenon is termed the original antigenic sin. To model original antigenic sin, we consider a non-monotonic degradation function, centered around the second peak, which represents the non-monotonic behavior of the binding constant, as in our previous model [73]. We fit the binding constant data [73] to a sixth order polynomial in \( p \), where \( p = (1 - M) / 2 \) is the relative distance between the recognition of the antibody and the virus. The degradation function is shown in the inset in Fig. 2.2. We consider a single peak virus replication rate, Eq. (2.33).

There is an interesting phase structure as a function of \( m \). From Eq. (2.30), we have a FM phase with \( M_1 = 1, M_2 = m \). We also have a non-selective NS phase, with \( M_1 < 1 \), where \( M_1, M_2 \) are determined by maximization of Eq. (2.30) with \( f_0 = 1 \) under the constraints of Eq. (2.12). The phase diagram for typical parameters [73] is shown in Fig. 2.2. A continuous phase boundary is observed between the FM and NS phases. The virus replication rate required to escape eradication by the adaptive immune system depends on how similar the virus and the vaccine are. When the vaccine is similar to the virus, \( m \) near 1, a large virus replication rate is required to escape eradication. This result indicates the typical usefulness of vaccines in protection against and eradication of viruses. When the vaccine is not similar to the virus, \( m < 0 \), the vaccine is not effective, and only a typical virus replication rate
is required.

When the vaccine is somewhat similar to, but not identical to, the virus, the replication rate required for virus survival is non monotonic. This result is due to the non monotonic degradation rate around the vaccine degradation peak. The minimum in the required virus replication rate, \( m \approx 0.30 \), corresponds to the minimum in the degradation rate, \( M_2 \approx 0.30 \). The competition between the immune system, vaccine, and virus results in a nontrivial phase transition for the eradication of the virus.

2.4.4 Tumor control and proliferation

We consider cancer to be a mutating, replicating object, with a flat replication rate around the first peak, Eqs. (2.38) and (2.30). We consider the immune system to be able to eradicate the cancer when cancer is sufficiently different from self. Thus, the T cells have a constant degradation rate everywhere except near the self, represented by the second peak,

\[
d_0(M_1, M_2) = \begin{cases} 
B, & -1 \leq M_2 < M_b \\
0, & M_b \leq M_2 \leq 1
\end{cases}
\] (2.40)

To be consistent with the biology, we assume \( M_b > 0 \). We also assume \( M_0 > 1/2 \). Typically, also, the Hamming distance between the cancer and the self will be small, \( m \) will be positive and near unity, although we do not assume this.

There are four possible selective, ferromagnetic phases. We find the phase boundaries analytically, as a function of \( m = (2l - N)/N \). For \( mM_0 < M_b \), there is a FM4
phase with $M_1 = M_0$ and $M_2 = mM_0$. The mean excess replication rate per site is $f_m = Ae^{\gamma(\sqrt{1-M_0^2} - 1)} - B$. There is a FM3 phase with $M_1 = M_0$ and $M_2 = M_b$. The mean excess replication rate per site is $f_m = Ae^{\gamma(\sqrt{(1+m)^2-(M_0+M_b)^2} + \sqrt{(1-m)^2-(M_0-M_b)^2})/2}$. This phase is chosen over the FM4 phase when the mean excess replication rate is greater. For $mM_0 \geq M_b$ there is a FM2 phase with $M_1 = M_0$ and $M_2 = mM_0$. The mean excess replication rate per site is $f_m = Ae^{\gamma(\sqrt{1-M_0^2} - 1)}$. For $mM_b \geq M_0$ there is a FM1 phase with $M_1 = mM_b$ and $M_2 = M_b$. The mean excess replication rate per site is $f_m = Ae^{\gamma(\sqrt{1-M_0^2} - 1)}$.

There are two non-selective phases. There is a NS1 phase with $M_1 = mM_b$ and $M_2 = M_b$. The mean excess replication rate per site is $f_m = e^{\gamma(\sqrt{1-M_b^2} - 1)}$. There is a NS2 phase with $M_1 = M_2 = 0$. The mean excess replication rate per site is $f_m = 1 - B$.

In Fig. 2.3 is shown the phase diagram for cancer proliferation. According to our previous model [73, 78], we choose $(1 - M_b)/2 = 0.23$. We choose $M_0 = 0.9$ for the width of the advantaged cancer phase. We choose the immune suppression rate as $B = 1$. As the cancer becomes more similar to the self, the immune control becomes less effective, and the replication rate required for the cancer to proliferate becomes less. Three of the four selective and one of the two non-selective phases are present for this set of parameters.
2.5 Discussion and conclusions

We have used the Eigen model to consider the interaction of a virus or cancer with a drug or the immune system. One can also use the parallel model to represent the replication dynamics of the virus or the cancer. This would be an interesting application of our formalism.

Another application of the formalism would be to consider explicitly the degradation induced by multidrug cocktails. That is, one would consider one peak to represent the preferred virus genome and $K - 1$ degradation peaks to represent the $K - 1$ drugs. We note that in the general case, the $y_i$ parameters depend on the explicit location of the drug degradation peaks, not simply the distance between them. Results from this application of the formalism could be quite illuminating as regards the evolution of multidrug resistance.

In conclusion, we have solved two common evolution models with general fitness, or replication and degradation rate, landscapes that depend on the Hamming distances from several fitness peaks. Why is this important? First, we have solved the microscopic models rather than assuming a phenomenological macroscopic model. As is known in statistical mechanics, a phenomenological model may not always detect the fine structure of critical phenomena. Second, approximate or numerical solutions, while useful, do not always explicitly demonstrate the essence of the phenomenon. With analytical solutions, the essence of the phenomenon is transparent. Third, we
have derived the first path integral formulation of the Eigen model. This formulation may prove useful in other studies of this model of molecular evolution.

Our results for cancer are a case in point. There are four stable selective phases and two stable non-selective phases. These results may help to shed light on the, at present, poorly understood phenomena of interaction with the immune system, and on why the immune response to cancer and to viruses differs in important ways. These phases could well also be related to the different stages, or grades, through which tumors typically progress. Our results are a first step toward making the connection with evolution on rugged fitness landscapes, landscapes widely accepted to be accurate depictions of nature. We have applied our solution of these microscopic complex adaptive systems to model four situations in biology: how a virus interacts with a drug, how an adaptable virus interacts with a drug, the problem of original antigenic sin [73], and immune system control of a proliferating cancer.
Appendix

In this appendix we derive the path integral representation for the solution to the Eigen model. For simplicity, we show the derivation of the \( K = 1 \) case. To our knowledge, this is the first path integral expression representation of the solution to the Eigen model. This path integral representation allows us to make strong analytical progress. We start from the quantum representation of the Eigen model [25]. The Hamiltonian is given by

\[
-H = \sum_{l=0}^{N} Ne^{-\gamma \left( \frac{\gamma}{N} \right)^l} \sum_{1 \leq i_1 < i_2 < \ldots < i_l \leq N} \sigma_{i_1}^x \sigma_{i_2}^x \ldots \sigma_{i_l}^x \times f_0(\sigma^z) - Nd_0(\sigma^z)
\]

where we have used the fact that with \( \gamma/N \) small, we need to consider only \( l \ll N \) spin flips. The partition function is decomposed by a Trotter factorization,

\[
Z = \text{Tr} e^{-\beta \hat{H}} = \text{Tr} \langle S_1 | e^{-\beta \hat{H}/L} | S_2 \rangle \langle S_2 | e^{-\beta \hat{H}/L} | S_3 \rangle \ldots \langle S_L | e^{-\beta \hat{H}/L} | S_1 \rangle
\]

Here

\[
\langle S_{l-1} | e^{-\beta \hat{H}/L} | S_l \rangle = \langle S_{l-1} | e^{\beta N/L} [e^{-\gamma e^{\gamma/N} N \sum \sigma_i^z f_0(\sigma^z) - d_0(\sigma^z)}] | S_l \rangle
\]

\[
= \langle S_{l-1} | I + \frac{\beta N}{L} \left[ e^{-\gamma e^{\gamma/N} (\sum \sigma_i^z f_0 \left( \sum_n s_n^l / N \right) - d_0 \left( \sum_n s_n^l / N \right) \right]} | S_l \rangle
\]

(2.43)
We use the notation $M_l = \sum_n s_n^l/N$. We find

$$\alpha_l = \langle S_{l-1}|I + \frac{\beta N}{L} e^{-\gamma} f_0 \left( \sum_n s_n^l/N \right) e^{\gamma/N} \sum_i \sigma_i^z - \frac{\beta N}{L} d_0 \left( \sum_n s_n^l/N \right) |S_l \rangle$$

$$= \langle S_{l-1}|S_l \rangle \left[ 1 - \frac{\beta N}{L} d_0(M_l) \right] + \frac{\beta N e^{-\gamma}}{L} f_0(M_l)e^{B \sum_n (s_n^l s_{n+1}^l - 1)}$$

(2.44)

where $e^{-2B} = \gamma/N$. We thus find

$$\alpha_l = \Delta(d_l) \left[ 1 - \frac{\beta N}{L} d_0(M_l) \right] + \frac{\beta N}{L} e^{-\gamma} f_0(M_l)e^{B d_l}$$

(2.45)

where $d_l = \sum_n (s_n^l s_{n+1}^l - 1)$. To represent this in path integral form, we consider

$$\frac{1}{(2\pi)^2} \int dh dm \psi e^{\Delta t N e^{-\gamma} f_0(M_l)e^{B m} e^{-\psi_m - \Delta t N d_0(M_l)}} \times e^{\psi_d - h(m-d)}$$

$$= \frac{1}{2\pi} \int dh dm [\delta(d) e^{-\Delta t N d_0(M_l)} + \Delta t N e^{-\gamma} f_0(M_l)e^{B m} \delta(m-d) + O[(\Delta t)^2]]$$

$$= \int dm [\delta(d) \delta(m-d) e^{-\Delta t N d_0(M_l)} + \Delta t N e^{-\gamma} f_0(M_l)e^{B m} \delta(m-d) \delta(m-d)]$$

$$= \delta(d) e^{-\Delta t N d_0(M_l)} + \Delta t N e^{-\gamma} f_0(M_l)e^{B d} \delta(0)$$

$$= \delta(0)[\Delta(d) e^{-\Delta t N d_0(M_l)} + \Delta t N e^{-\gamma} f_0(M_l)e^{B d}]$$

(2.46)

where $\Delta t = \beta/L$. We note that had we used a Fourier representation of the $\delta$ function on the finite domain $[-A/2,A/2]$ instead of the infinite domain $(-\infty,\infty)$, the expression $2\pi \delta(0)$ simply becomes $A$; moreover, such a finite representation of the $\delta$ function is a sufficiently accurate representation of the $\Delta(d_l)$ constraint when $A >> N$. Ignoring the constant prefactor $\delta(0)$ terms, we can write the full partition function as

$$Z = \text{Tr} \int D\psi D\mathbf{h} D\mathbf{m} e^{\sum_l \psi(d_l) + \beta N/L \sum_l [-h_l m_l + h_l d_l]}$$

(2.47)
We now introduce the integral representation of the constraint \( \delta[(\beta/L)(NM_i - \sum_n s_n^i)] \).

After rescaling \( Bm_i \to m_i, h_l \to Bh_l \) we find

\[
Z = \text{Tr} \int \mathcal{D}\psi \mathcal{D}h \mathcal{D}H \mathcal{D}M e^{\beta N/L} \sum_i \left[ e^{-\gamma f_0(M_i)e^{m_i}} e^{-\psi m_i/B - h_m m_i - H_i M_i} \right] \\
\times e^{\beta/L} \sum_i \sum_n s_n^i + \sum_i (\psi + \beta N B h_l/L) \sum_n (s_n^i - 1) \left( \psi \right) \]

We note by an expansion of the

\[
\exp[(\beta N/L)e^{-\gamma f_0(M_i)e^{m_i}e^{-\psi m_i/B}]} \\
= \sum_{k_l=0}^{\infty} [(\beta N/L)e^{-\gamma f_0(M_i)e^{m_i}]^{k_l} \exp(-k_l \psi m_i/B) / k_l! \quad (2.49)
\]

term in Eq. (2.48) to first order in \( \beta N/L \) that the integral over \( \psi_l \) gives nothing more than \( \delta(-k_l m_l/B + d_l) \) for \( k_l = 0, 1 \). This condition, however, is already enforced by the \( h_l \) field when \( k_l = 1 \) and by the \( m_l \) field when \( k_l = 0 \) if we take as a rule to disallow mutations when \( h_l = 0 \). We can, thus, remove the integral over \( \psi \), removing the \( \delta(0) \) that we anticipated, to find

\[
Z = \int \mathcal{D}h \mathcal{D}m \mathcal{D}H \mathcal{D}M e^{\beta N/L} \sum_i \left[ e^{-\gamma f_0(M_i)e^{m_i} - d_0(M_i) - h_m m_i - H_i M_i} \right] Q \quad (2.50)
\]

where

\[
Q = \text{Tr} e^{\beta/L} \sum_i H_i \sum_n s_n^i + \beta N B / L \sum_n (s_n^i - 1) F \quad (2.51)
\]

Here \( Q \) is the partition function of \( N \) 1D Ising models of length \( L \). Here \( F \) enforces the constraint of disallowing mutations when \( h_l = 0 \): \( F = \prod_{l=1}^{L} \Delta(\Delta(h_l) \sum_n (s_n^l - 1) - \)}
We note that \( Q = Q_1^N \), where \( Q_1 \) is the partition function of one of these models. We are not, at this point, allowed to assume that the \( H_i \) or \( h_i \) fields are constant over \( l \). Indeed, by Taylor series expanding the first term in Eq. (2.50) and integrating over \( m_i \), we find that \( h_i = 0 \) or \( h_i = L/(\beta N) \). We evaluate the partition function \( Q_1 \) with an ordered product of transfer matrices. To first order in \( \beta/L \) the matrix at position \( l \) is given by \( T_l = I + \epsilon_l \) where

\[
\epsilon_l = \begin{pmatrix}
\frac{\beta H_i}{L} & \frac{\tau_i}{N} \\
\frac{\tau_i}{N} & -\frac{\beta H_i}{L}
\end{pmatrix} = \begin{pmatrix}
\frac{\beta H_i}{L} & \frac{\beta h_i}{L} \\
\frac{\beta h_i}{L} & -\frac{\beta H_i}{L}
\end{pmatrix}
\] (2.52)

We find

\[
Q_1 = \text{Tr} \prod_l T_l \sim \text{Tr} \prod_l e^{\epsilon_l}
\] (2.53)

We rescale \( h \rightarrow h/\gamma \) and \( m \rightarrow m\gamma \) and take the continuous limit to find

\[
Q_1 = \text{Tr} \hat{T} e^{\int_0^\beta d\beta' [\sigma^x H(\beta') + \sigma^z h(\beta')]} \] (2.54)

where the operator \( \hat{T} \) indicates (reverse) time ordering, and \( \beta' = \beta(L-l)/L \). We find the form of the partition function to be

\[
Z = \int \mathcal{D}h\mathcal{D}m\mathcal{D}H\mathcal{D}Me^N \int_0^\beta d\beta' [e^{-\gamma f_0(M)}e^{\gamma m - d_0(M) - hm - HM} + N \ln Q_1]
\] (2.55)

Noting the \( N \) prefacing the entire term in the exponential, we take the saddle point.

We note that

\[
\frac{\delta Q_1}{\delta H(\beta')} \bigg|_{H(\beta') = H, h(\beta') = h} = \frac{\beta H}{\sqrt{H^2 + h^2}} \times 2 \sinh(\beta \sqrt{H^2 + h^2})
\] (2.56)
\[
\delta Q_1 \left|_{H(\beta')=H,h(\beta')=h} \frac{\delta h(\beta')}{\delta h(\beta')} \right. = \frac{\beta h}{\sqrt{H^2 + h^2}} \times 2 \sinh(\beta \sqrt{H^2 + h^2}) \quad (2.57)
\]

We, thus, find a solution of the saddle point condition to be fields \( H, M, h, m \) independent of \( \beta \) that maximize

\[
\frac{\ln Z}{N} = \beta \left[ f_0(M)e^{-\gamma e^m} - d_0(M) - hm - HM + \ln[2 \cosh(\beta \sqrt{H^2 + h^2})] \right] \quad (2.58)
\]

when the fields are averaged over a range \( \Delta \beta = O(1/N) \) by the saddle point limit.

In the limit of large \( \beta \), we find

\[
\frac{\ln Z}{N\beta} = \max_{M,H,m,h} \left[ f_0(M)e^{-\gamma e^m} - d_0(M) - hm - HM + \sqrt{H^2 + h^2} \right] \quad (2.59)
\]

One can also derive Eq. (2.56) by means of a series expansion in \( \beta \), a "high temperature" expansion.

The generalization of the path integral representation to the multiple peak Eigen case proceeds as in the parallel case. One introduces \( K \) fields for the magnetizations, \( M_k \), and \( K \) fields enforcing the constraint, \( H_k \). One also finds in the linear field part of the Ising model the sum \( \sum_{k=1}^K H_k \sum_n v_n^k s_n^l \) instead of simply \( H \sum_n s_n^l \). The definition of the \( y_i \) and the \( \alpha_{ik} \) allows one to rewrite this in the form that leads to Eqs. (2.7), (2.24).
Chapter 3
Quasispecies theory for Horizontal Gene Transfer and Recombination

This chapter was communicated in the article E. Muñoz, J.-M. Park, and M. W. Deem, Phys. Rev. E (to appear, 2008).

Abstract

We introduce a generalization of the parallel, or Crow-Kimura, and Eigen models of molecular evolution to represent the exchange of genetic information between individuals in a population. We study the effect of different schemes of genetic recombination on the steady-state mean fitness and distribution of individuals in the population, through an analytic field theoretic mapping. We investigate both horizontal gene transfer from a population and recombination between pairs of individuals. Somewhat surprisingly, these nonlinear generalizations of quasispecies theory to modern biology are analytically solvable. For two-parent recombination, we find two selected phases, one of which is spectrally rigid. We present exact analytical formulas for the equilibrium mean fitness of the population, in terms of a maximum principle, which are generally applicable to any permutation invariant replication rate function. For smooth fitness landscapes, we show that when positive epistatic interactions are present, recombination or horizontal gene transfer introduces a mild load
against selection. Conversely, if the fitness landscape exhibits negative epistasis, horizontal gene transfer or recombination introduce an advantage by enhancing selection towards the fittest genotypes. These results prove that the mutational deterministic hypothesis holds for quasispecies models. For the discontinuous single sharp peak fitness landscape, we show that horizontal gene transfer has no effect on the fitness, while recombination decreases the fitness, for both the parallel and the Eigen models. We present numerical and analytical results as well as phase diagrams for the different cases.

3.1 Introduction

It has been argued that genetic recombination provides a mechanism to speed up evolution, at least in finite populations [39]. Moreover, it has been suggested that recombination may provide a way to escape from the phenomenon of “Muller’s ratchet” [79], or suboptimal fitness characteristic of finite populations with asexual reproduction. In bacteria, it has been proposed [80] that horizontal gene transfer allows for the gradual emergence of modularity, through the formation of gene clusters and their eventual organization into operons. In in-vitro systems, protein engineering protocols by directed evolution incorporate genetic recombination in the form of DNA shuffling [81, 82] to speed up the search for desired features such as high binding constants among combinatorial libraries of mutants.

Besides these inherently dynamical effects, it remains a matter of debate if the
exchange of genetic-encoding elements provides a long-term advantage to an infinite population in a nearly static environment. Indeed, it is argued that [83] when advantageous genetic associations have been generated as a result of selection in a given environment, further random recombination is likely to disrupt these associations, thus decreasing the overall fitness. This argument is less cogent if we consider that recombination and horizontal gene transfer preserve the modular structure of the genetic material [80]. That is, entire operational and functional units are recombined, rather than random pieces. It has also been proposed that for recombination to introduce an advantage in infinite populations, negative linkage disequilibrium is required [45, 84, 42, 44]. This situation means that particular allele combinations are present in the population at a lower frequency than predicted by chance. Negative linkage disequilibrium can result as a consequence of negative epistasis: alleles with negative contributions to the fitness interact synergistically, increasing their deleterious effect when combined, and alleles with positive contributions to the fitness interact antagonistically [46, 45, 85], see Fig. 3.1. Under negative epistasis, the mutational deterministic hypothesis [41, 42, 43, 44, 45, 46, 47] postulates that recombination promotes a more efficient removal of deleterious mutations, by bringing them together into single genomes, and hence facilitating selection [41, 48] to discard those genotypes with low fitness. It has been argued that the negative linkage disequilibrium generated by negative epistatic interactions is a factor to promote the evolution of
recombination in nature [45, 47, 86], and conversely that recombination may act as a
mechanism to evolve epistasis [87, 88, 89]. This later statement is controversial, since
it is intuitive that recombination should contribute to weaken correlations between
different genes [90]. Despite these theoretical arguments, experimental studies seem
to indicate that negative epistasis is not so common in nature [91, 92] as recombi-
nation and, moreover, both negative and positive epistasis may coexist as different
fitness components [45] within the same genome in natural organisms.

To address some of these questions, we study the effect of transferring genetic
information between different organisms in an infinite population. We choose the
conceptual framework of “quasispecies” theory, represented by two classical mod-
els of molecular evolution: the Eigen [11, 93, 12, 94] model and the parallel, or
Crow-Kimura, model [95, 96]. These classical models include the basic processes of
mutation, selection, and replication that occur in biological evolution. Our goal is
to solve these two standard models of quasispecies theory, Crow-Kimura and Eigen,
when horizontal gene transfer or recombination are included. Since horizontal gene
transfer and recombination are essential features of evolutionary biology, our solu-
tions bring quasispecies theory closer to modern biology. An operational definition
of fitness is provided in these models by the replication rate, which is considered
to be a function of the genotype. In their simplest formulation quasispecies mod-
els consider a static environment, with a deterministic mapping between individual
genetic sequences and replication rate. Both the Eigen [11, 93] and the parallel, or Crow-Kimura model [95], are formulated in terms of a large system of differential equations, describing the time evolution of the relative frequencies of the different sequence types in an infinite population, a mathematical language that is common in the field of chemical kinetics [11, 93]. Sequences, representing information carrying molecules such as RNA or DNA, are assumed to be drawn from a binary alphabet (e.g. purines/pyrimidines). The most remarkable property of these classical models is that when the mutation rate is below a critical value they exhibit a phase transition in the infinite genome limit [11, 93, 12, 21, 20, 94, 97, 29, 28], with the emergence of a self-organized phase: the quasispecies [11, 93, 12]. This organized phase, characterized by a collection of nearly neutral mutants rather than by a single homogeneous sequence type, is mainly a consequence of the auto-catalytic character of the evolution dynamics, which tends to enrich exponentially the proportion of fittest individuals in the population [11, 93, 12, 94]. The quasispecies concept, with its corresponding "error threshold" transition, has been applied in the interpretation of experimental studies in RNA viruses [98, 99, 100, 101]. In particular, the error-threshold transition has been proposed as a theoretical motivation for an antiviral strategy [77], termed "lethal mutagenesis", which drives an infecting population of viruses towards extinction by enhancing their mutation rate [102, 103, 104]. It has been argued, however, that the mechanism for lethal mutagenesis possesses a strong ecological component.
[105], and that perhaps the mean population fitness is simply driven negative, and so the total number of viral particles in an infecting population decreases in time towards extinction, in contrast with error-threshold theories that describe a shift of the composition of the quasispecies in genotype space.

The existence of the error threshold transition has motivated the attention of theoretical physicists, especially since it was proved that the quasispecies theory can be exactly mapped into an 2D Ising spin system [20, 21], with a phase transition that is first order for a sharp peak fitness, and second or higher order for smooth fitness functions. More recently, exact mappings into a quantum spin chain [22, 23, 27, 24, 25] or field theoretic representations [29] have been developed. Analytical and numerical studies of these systems, in the large genome limit, are possible when the fitness function is considered to be permutation invariant [22, 23, 29, 40, 97], or depending on the overlap with several peaks in sequence space [28]. The mapping of the quasispecies models into a physical system allows for the application of the powerful mathematical techniques of statistical mechanics, thus obtaining exact analytical solutions which provide significant insight over numerical studies [29, 28, 27]. Most of the existing analytical solutions correspond to the case when recombination is absent. Recombination and horizontal gene transfer have been studied by computer simulations of artificial gene networks [46] and digital organisms [84], but relatively few analytical approaches have been reported in the context of quasispecies theory [38, 39, 40, 106].
Figure 3.1 Convention for the sign of epistasis, $\epsilon$. In the figure are represented two smooth fitness landscapes, as a function of $u = 2l/N - 1$, with $N$ the total length of the (binary) genetic sequences and $0 \leq l \leq N$ the number of beneficial mutations (number of ' +' spins) along the sequence. In this representation, positive (synergistic) epistasis $\epsilon > 0$ corresponds to a positive curvature $f''(u) > 0$, while negative (antagonistic) epistasis $\epsilon < 0$ corresponds to a negative curvature $f''(u) < 0$ [85, 46, 45]. The examples shown are a quadratic fitness landscape $f(u) = ku^2/2$ (dashed line), with positive curvature and $\epsilon > 0$, and a square-root fitness landscape $f(u) = k\sqrt{u}$ (solid line), with negative curvature and $\epsilon < 0$. We set $k = 4.0$ in both examples.
A numerical study of a mathematical model for viral super-infection termed uniform crossover, and intermediate between horizontal gene transfer and recombination, has been reported [38], with numerical solutions based on relatively short viral sequences (N=15). More recently, the effect of incorporating horizontal gene transfer in quasispecies theory has been studied in terms of the dynamics [39], reporting numerical studies and approximate analytical expressions. Exact analytical expressions for the equilibrium properties of the population in the presence of horizontal gene transfer have been derived using the methods of quantum field theory [40].

In this article, we study the effect of introducing different schemes of genetic recombination in quasispecies theory. Extending the results in [40], we present an exact field theoretical mapping of the parallel and Eigen models. We remark that field theoretical methods provide a unique and powerful set of tools for the analytical study of dynamical systems, such as reaction-diffusion [32, 58] or birth-death processes [34]. In this paper, we employ these theoretical tools to obtain exact analytical expressions for the equilibrium mean fitness and average composition of the population, for permutation invariant but otherwise arbitrary replication rate functions.

In Section 2 we consider the parallel model. We consider horizontal gene transfer of non-overlapping blocks, as well as of blocks of random size. We also consider a recombination process producing a daughter sequence symmetrically from two parents, as might occur in viral super- or co-infection. In Section 3, we study the effect
of these different genetic recombination schemes in the context of the Eigen model. In both models, recombination leads to two selected phases. Interestingly, beyond a critical recombination rate, the distribution of the population becomes independent of the recombination rate. Also interesting is that the steady-state distribution is independent of the crossover probability.

To study the effect of epistasis, whose sign is determined by the curvature of the fitness landscape (second derivative) when represented as a function of the Hamming distance with respect to the wild-type, we considered two different examples of smooth fitness functions: a quadratic function, representing positive epistasis, and a square-root function representing negative epistasis. We find that, for the quadratic fitness function, horizontal gene transfer and recombination introduce a mild load against selection. The opposite effect is observed for the square-root fitness, that is, horizontal gene transfer and recombination introduce an advantage by enhancing selection towards fittest genotypes. This results provide support for the mutational deterministic hypothesis, which postulates that recombination should be beneficial for negative epistasis fitness functions, and deleterious for positive epistasis fitness functions. Moreover, we prove analytically in Appendix 3.12 that the mutational deterministic hypothesis applies for the parallel model in the presence of horizontal gene transfer. A similar proof is provided in Appendix 3.13 for the Eigen model. We also show analytically that the mutational deterministic hypothesis applies for the case of
two-parent recombination, as presented in Appendix 3.14 for the parallel model, and in Appendix 3.15 for the Eigen model.

The effect of recombination becomes negligible for discontinuous fitness landscapes, such as a single sharp peak. For all these cases, we present exact analytical expressions that determine the phase structure of the population at steady state. Results are explicit for any microscopic fitness function: Eqs. (3.14), (3.31), and (3.62–3.63) for the parallel model and Eqs. (3.82), (3.93), and (3.106–3.107) for the Eigen model. We evaluate these expressions for three permutation invariant fitness functions: sharp peak, quadratic, and square root for the two common forms of quasispecies theory, parallel and Eigen: Eqs. (3.22), (3.23), (3.33), (3.34), (3.68), (3.71), (3.85–3.87), (3.96–3.98), (3.112), and (3.113). We also present numerical tests supporting our analytical equations.

3.2 The parallel model

We consider a generalization [40] of the parallel, or Crow-Kimura [95], model to take into account the transfer of genetic material between pairs of individuals in an infinite population.

\[
\frac{dq_i}{dt} = r_i q_i + \sum_{k=1}^{2^N} \mu_{ik} q_k + \nu N \frac{\sum_{k,l} R_{kl}^i q_k q_l}{\sum_k q_k} - \nu N q_i
\]  

(3.1)

Here, \( q_i \) represents the (unnormalized) frequency of the sequence type \( S_i = (s_1^i, s_2^i, \ldots, s_N^i) \), with \( s_j^i = \pm 1 \), for \( 1 \leq i \leq 2^N \) and \( 1 \leq j \leq N \). The normalized frequencies are ob-
tained from \( p_i = q_i / \sum_{j=1}^{2^N} q_j \). In Eq. (3.1), \( r_i \) is the replication rate of sequence \( S_i \). It is given that \( r_i = N f \left( \frac{1}{N} \sum_{j=1}^{N} s_j^i \right) \). The mutation rate from sequence \( S_j \) into \( S_i \) is \( \mu_{ij} = \mu \delta_{d_{ji},1} - N \mu \delta_{d_{ji},0} \). The Kronecker delta in this expression ensures that mutations involve a single base substitution per unit time (generation). Genetic recombination processes between pairs of sequences in the population are represented by the nonlinear term. They are considered to occur with an overall rate \( \nu \), while the coefficient \( R_{kl}^i \) represents the probability that a pair of parental sequences \( S_k, S_l \) produces an offspring \( S_i \). Depending on the particular recombination mechanism, some of these coefficients will be identically zero. Also, these coefficients must satisfy the condition \( \sum_{i=1}^{2^N} R_{kl}^i = 1, \forall 1 \leq k, l \leq 2^N \).

For this generic process, we will present the analytical solutions for the steady-state mean fitness by considering different schemes of genetic recombination.

### 3.2.1 Horizontal gene transfer of non-overlapping blocks

In this recombination scheme, we consider the exchange of blocks of genetic material between pairs of individuals. We consider these blocks to be non-overlapping in the parental sequences, and of a fixed size \( \bar{M} \). Thus, each sequence is made of \( N/\bar{M} \) blocks. The recombination coefficients in the differential Eq. (3.1) are given for this horizontal gene transfer process by

\[
R_{kl}^i = \sum_{b=0}^{N/\bar{M}-1} \prod_{j_b=\bar{M}b+1}^{\bar{M}(b+1)} \left( \frac{1 + s_j^i s_j^i}{2} \right) \prod_{j \notin \{j_b\}}^{N} \left( \frac{1 + s_j^i s_j^i}{2} \right).
\]
Figure 3.2  Pictorial representation of the horizontal gene transfer process considered.

Here, $0 \leq b \leq N/\bar{M} - 1$ represents the block index, while $\bar{M}b + 1 \leq j_b \leq \bar{M}(b + 1)$ represents the site index within block $b$.

Generalizing the method presented in [40], we write the non-linear term as

$$
\frac{\sum_l q_l \bar{R}_{kl}}{\sum_m q_m} = \sum_{b=0}^{N/\bar{M}-1} \left\langle \prod_{j_b=\bar{M}b+1}^{\bar{M}(b+1)} \left( \frac{1 + s_j^l s_j^i}{2} \right) \prod_{j \neq \{j_b\}}^{N} \left( \frac{1 + s_j^k s_j^i}{2} \right) \right\rangle.
$$

Here, $\langle A_l \rangle = \sum_l q_l A_l / \sum_m q_m$ is a population average. At steady state, this average is independent of the value of $b$, due to the symmetry of the fitness function.

The variance of the composition $u_l = \frac{1}{N} \sum_{j=1}^{N} s_j^l$ is given by $\frac{1}{N^2} \sum_{j,j'=1}^{N} \langle \delta s_j^l \delta s_{j'}^l \rangle$. 

In the absence of recombination or horizontal gene transfer this variance is $O(N^{-1})$, which implies correlations along the sequence are $O(N^{-1})$ [29]. We expect the same scaling of the variance in the presence of recombination or horizontal gene transfer. Therefore, we introduce the factorization

$$\prod_{j_b=M+1}^{M+1} \frac{1 + s_{j_b}^i s_{j_b}^i}{2} \sim \prod_{j_b=M+1}^{M+1} \left( \frac{1 + s_{j_b}^i s_{j_b}^i}{2} + O(M/N) \right)$$

which becomes exact in the $N \rightarrow \infty$ limit. Here, $u(j_b) = \sum_i q_i s_{j_b}^i / \sum_m q_m$ is the average base composition at site $j_b$.

We are interested in the long time behavior of the system, when the average base composition becomes independent of time and position $u(j) \sim u$. Thus, in the formalism of spin Boson operators [40] $\tilde{a}(j) = (\tilde{a}_1(j), \tilde{a}_2(j))$, we define the recombination operator describing this recombination term by

$$\hat{R} = \frac{1}{N} \sum_{b=0}^{N/M-1} \left[ \prod_{j_b=M+1}^{M+1} [\rho_+ \tilde{a}_1(j_b) + \rho_- \tilde{a}_2(j_b)][\tilde{a}_1(j_b) + \tilde{a}_2(j_b)] - \hat{I} \right]$$

Here, $\hat{I}$ is the identity operator. The coefficients $\rho_\pm = (1 \pm u)/2$ represent [40] the steady-state probability (per site) of having a “+1” or a “-1”. Defining the matrix

$$D = \begin{pmatrix} \rho_+ & \rho_- \\ \rho_- & \rho_+ \end{pmatrix},$$

(3.6)
the recombination operator in Eq. (3.5) can be expressed as

\[ \hat{R} = \frac{1}{N} \sum_{b=0}^{N/M-1} \left[ \prod_{j_b=Mb+1}^{Mb+1} \tilde{a}^\dagger(j_b)D\tilde{a}(j_b) - \hat{I} \right]. \quad (3.7) \]

The Hamiltonian

Considering the recombination operator in Eq. (3.7), we formulate the Hamiltonian describing the system

\[ -\hat{H} = Nf \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{a}^\dagger(j)\sigma_3\tilde{a}(j) \right] + \mu \sum_{j=1}^{N} \tilde{a}^\dagger(j)\sigma_1\tilde{a}(j) - \hat{I} \]

\[ + \nu \sum_{b=0}^{N/M-1} \left[ \prod_{j_b=Mb+1}^{Mb+1} \tilde{a}^\dagger(j_b)D\tilde{a}(j_b) - \hat{I} \right]. \quad (3.8) \]

Here, \( \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \) and \( \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \) are the Pauli matrices. We introduce a Trotter factorization

\[ e^{-\hat{H}t} = \lim_{M \to \infty} \int [D\bar{z}^M D\bar{z}] \bar{z}_M \left( \prod_{k=1}^{M} \langle \bar{z}_k | e^{-t\hat{H}} | \bar{z}_{k-1} \rangle \right) \langle \bar{z}_0 \rangle. \quad (3.9) \]

As shown in Appendix 3.1, the partition function that gives the mean population fitness is

\[ Z = \int [D\xi D\bar{\xi} D\phi D\bar{\phi}] e^{-S[\xi,\bar{\xi},\phi,\bar{\phi}]} \sim e^{Nf_{\text{mt}}} \cdot (3.10) \]

Here, the action in the continuous time limit is

\[ S[\xi, \bar{\xi}, \phi, \bar{\phi}] = -N \int_0^t dt \left[ -\bar{\xi} \dot{\xi} - \bar{\phi} \dot{\phi} - \mu - \frac{\nu}{M} + f(\xi) + \frac{\nu}{M} \phi^M \right] - N \ln Q \]

\[ \quad (3.11) \]
The saddle point limit

In the $N \to \infty$ limit, the saddle point is exact and we obtain an analytical expression for the partition function Eq. (3.10). We look for the steady-state solution, when the fields become independent of time, $\xi_c, \bar{\xi}_c, \phi_c, \bar{\phi}_c$. The trace defined by Eq. (3.126) in the long time saddle-point limit becomes

$$
\lim_{t \to -\infty} \frac{\ln Q_c}{t} = \frac{\bar{\phi}_c}{2} + \left[ \bar{\xi}_c(\xi_c + u\bar{\phi}_c) + (\mu + \bar{\phi}_c/2)^2 \right]^{1/2}
$$

(3.12)

Hence, the saddle-point action is

$$
\lim_{N,t\to\infty} \frac{\ln Z}{Nt} = \lim_{t\to\infty} \frac{-S_c}{Nt} = f_m = \max_{\xi_c,\bar{\xi}_c,\phi_c,\bar{\phi}_c} \left\{ f(\xi_c) - \frac{\bar{\xi}_c}{2} - \phi_c \bar{\phi}_c - \mu - \frac{\nu}{M} + \frac{\nu}{M} \phi_c \bar{\phi}_c \frac{\nu}{2} \left[ 1 + \frac{\nu}{2\mu} (1\phi_c(\xi_c)] [\bar{M}^{-1}] \left[ 1 + \frac{\nu}{2\mu} (1\phi_c(\xi_c)] [\bar{M}^{-1}] - u^2 \right]^{1/2} \right. \}
$$

(3.13)

As shown in Appendix 3.2, the mean fitness of the population is

$$
f_m = \max_{-1 \leq \xi_c \leq 1} \left\{ f(\xi_c) - \mu - \frac{\nu}{M} + \frac{\nu}{M} \phi_c(\xi_c)] [\bar{M}^{-1}] \left[ 1 + \frac{\nu}{2\mu} (1\phi_c(\xi_c)] [\bar{M}^{-1}] - u^2 \right]^{1/2} \right. \}
$$

(3.14)

Here, $\phi_c$ is given by Eq. (3.133), and the surplus $u$ is obtained through the self-consistency condition $f_m = f(u)$. Equation (3.14) represents an exact analytical expression for the mean fitness of an infinite population experiencing horizontal gene
transfer. This expression is valid for an arbitrary, permutation invariant replication rate \( f(u) \).

It is worth to notice that Eq. (3.14) is a natural generalization of the single-site horizontal gene transfer process described in [40]. Indeed, specializing the Eqs. (3.133) and (3.14) to the particular case \( M = 1 \), after some algebra, we obtain

\[
 f_m(\bar{M} = 1) = \max_{-1 \leq \xi_c \leq 1} \left\{ f(\xi_c) - \mu - \frac{\nu}{2} + \frac{\nu u}{2} \xi_c + \sqrt{1 - \xi_c^2} \left[ \left( \mu + \frac{\nu}{2} \right)^2 - \left( \frac{w u}{2} \right)^2 \right]^{1/2} \right\},
\]

which reproduces the analytical result in [40].

**Numerical tests and examples**

For numerical calculations, it is convenient to reformulate Eq. (3.1) in terms of the fraction of the population at a distance \( l \) from the wild type, \( P_l = \sum_{j \in C_l} p_j \). Here, \( C_l \) is the class of sequences with \( l \) number of "-1" sites. The number of sequences within this class is \( \binom{N}{l} \).

As an example, for the case \( \bar{M} = 3 \), the differential equation representing the time evolution of the probability distribution of classes within an infinite population
of binary sequences is

\[
\frac{dP_l}{dt} = N \left[ f(2l/N - 1) - \sum_{l' = 0}^{N} P_{l'} f(2l'/N - 1) - \mu \right] P_l + \mu N \left[ \frac{N - l + 1}{N} P_{l-1} + \frac{l + 1}{N} P_{l+1} \right] \\
+ \nu N \left\{ \rho^3_- g_3(N - l + 3)P_{l-3} + [\rho^3_+ h(N - l + 2) + 3\rho^2_- \rho_+ g_3(N - l + 2)]P_{l-2} \\
+ [\rho^3_- h(l - 1) + 3\rho_- \rho^2_+ g_3(N - l + 1) + 3\rho^2_- \rho_+ h(N - l + 1)]P_{l-1} \\
+ [\rho^3_+ h(N - l - 1) + 3\rho_- \rho^2_+ g_3(l + 1) + 3\rho_- \rho_+ h(l + 1)]P_{l+1} \\
+ [\rho^3_+ h(l + 2) + 3\rho_- \rho^2_+ g_3(l + 2)]P_{l+2} + \rho^3_+ g_3(l + 3)P_{l+3} \right\} \\
- \frac{\nu}{3} N \left\{ (\rho^3_- + 3\rho^2_- \rho_+ + 3\rho_- \rho^2_+) g_3(N - l) + (\rho^3_+ + 3\rho^2_- \rho_+ + \rho^3_+) h(N - l) \\
+ (\rho^3_- + 3\rho_- \rho^2_+ + \rho^3_+) h(l) + (\rho^3_+ + 3\rho_+ \rho^2_- + 3\rho_- \rho^2_+) g_3(l) \right\} P_l \\
\right. 
\]

(3.16)

In writing this equation we have made use of the only $O(N^{-1})$ correlations between sites, which holds at long time as well as for short time with suitable initial conditions.

Here, we defined

\[
\rho_\pm = \frac{1 \pm u}{2} 
\]

(3.17)

where the average composition is calculated as

\[
u = \sum_{l=0}^{N} \frac{N - 2l}{N} P_l 
\]

(3.18)

and the functions

\[
g_3(l) = \frac{l(l - 1)(l - 2)}{N(N - 1)(N - 2)} \\
\]

(3.19)

\[
h(l) = 3\frac{l(l - 1)(N - l)}{N(N - 1)(N - 2)} 
\]
A comparison between the analytical expression Eq. (3.14) and the direct numerical solution of the differential Eq. (3.16) for \( N = 1002 \) is presented in Table 3.1, where the quadratic fitness \( f(u) = ku^2/2 \) was considered. We notice that the analytical method and the numerical solution provide the same results within \( \mathcal{O}(N^{-1}) \), as expected from the saddle point limit.

The differential equation representing the horizontal gene transfer of blocks of size \( \tilde{M} = 4 \) within an infinite population of binary sequences is given by

\[
\frac{d}{dt} P_l = N \left[ f(2l/N - 1) - \sum_{l'=0}^{N} P_{l'} f(2l'/N - 1) - \mu \right] P_l + \mu N \left[ \frac{N - l + 1}{N} P_{l-1} + \frac{l + 1}{N} P_{l+1} \right] \\
+ \frac{\nu}{4} N \left\{ g_4(N - l + 4) \rho_-^4 P_{l-4} + [\rho_-^4 h_3(N - l + 3) + 4 \rho_-^3 \rho_+ g_4(N - l + 3)] P_{l-3} \right. \\
+ [\rho_-^4 h_2(l - 2) + 4 \rho_-^3 \rho_+ h_3(N - l + 2) \\
+ 6 \rho_-^2 \rho_+^2 g_4(N - l + 2)] P_{l-2} + [\rho_-^4 h_3(l - 1) + 4 \rho_-^3 \rho_+ h_2(l - 1) \\
+ 6 \rho_-^2 \rho_+^2 h_3(N - l + 1) + 4 \rho_- \rho_+^3 g_4(N - l + 1)] P_{l-1} \right. \\
+ [\rho_+^4 h_3(N - l - 1) + 4 \rho_- \rho_+^3 h_2(l + 1) + 6 \rho_-^2 \rho_+^2 h_3(l + 1) + 4 \rho_- \rho_+ g_4(l + 1)] P_{l+1} \right. \\
+ [\rho_+^4 h_2(l + 2) + 4 \rho_- \rho_+^3 h_3(l + 2) + 6 \rho_-^2 \rho_+^2 g_4(l + 2)] P_{l+2} \right. \\
+ [\rho_+^4 h_3(l + 3) + 4 \rho_- \rho_+^3 g_4(l + 3)] P_{l+3} + \rho_+^4 g_4(l + 4) P_{l+4} \right. \\
\left. \left\} \left( \frac{\nu}{4} N \left\{ [\rho_+^4 + 6 \rho_-^2 \rho_+^2 + 4 \rho_- \rho_+ + \rho_+^4] h_3(N - l) + [\rho_+^4 + 6 \rho_-^2 \rho_+^2 + 4 \rho_- \rho_+^3 + \rho_+^4] h_3(l) \right. \\
+ [4 \rho_-^2 \rho_+ + 6 \rho_-^2 \rho_+^2 + 4 \rho_- \rho_+^3 + \rho_+^4] g_4(N - l) + [4 \rho_-^2 \rho_+ + 6 \rho_-^2 \rho_+^2 + 4 \rho_- \rho_+^3 + \rho_+^4] g_4(l) \right. \\
+ [\rho_+^4 + 4 \rho_-^3 \rho_+ + 4 \rho_- \rho_+^3 + \rho_+^4] h_2(l) \right\} P_l \right. \\
\right. \\
\left. \right. \\
\left. \right. \right. \\
\left. \right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \\
\right. \right. \
\right. 
(3.20)
\]

Here, the parameters \( \rho_{\pm} \) and \( u \) are defined, as before, by Eq. (3.17) and Eq. (3.18),
Table 3.1  Analytical versus numerical results for horizontal gene transfer in the parallel (Kimura) model for the quadratic fitness \( f(u) = ku^2/2 \), with \( \bar{M} = 3 \).

<table>
<thead>
<tr>
<th>( k/\mu )</th>
<th>( \nu/\mu )</th>
<th>( u^{\text{numeric}} )</th>
<th>( u^{\text{analytic}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>0.0</td>
<td>0.4993</td>
<td>0.5000</td>
</tr>
<tr>
<td>2.0</td>
<td>0.5</td>
<td>0.4830</td>
<td>0.4838</td>
</tr>
<tr>
<td>2.0</td>
<td>1.0</td>
<td>0.4668</td>
<td>0.4677</td>
</tr>
<tr>
<td>2.0</td>
<td>1.5</td>
<td>0.4510</td>
<td>0.4519</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0</td>
<td>0.5995</td>
<td>0.6000</td>
</tr>
<tr>
<td>2.5</td>
<td>0.5</td>
<td>0.5915</td>
<td>0.5920</td>
</tr>
<tr>
<td>2.5</td>
<td>1.0</td>
<td>0.5838</td>
<td>0.5844</td>
</tr>
<tr>
<td>2.5</td>
<td>1.5</td>
<td>0.5766</td>
<td>0.5772</td>
</tr>
<tr>
<td>5.0</td>
<td>0.0</td>
<td>0.7998</td>
<td>0.8000</td>
</tr>
<tr>
<td>5.0</td>
<td>0.5</td>
<td>0.7988</td>
<td>0.7990</td>
</tr>
<tr>
<td>5.0</td>
<td>1.0</td>
<td>0.7979</td>
<td>0.7981</td>
</tr>
<tr>
<td>5.0</td>
<td>1.5</td>
<td>0.7970</td>
<td>0.7972</td>
</tr>
</tbody>
</table>
respectively. We also define the functions

\[
\begin{align*}
g_4(l) &= \frac{l(l-1)(l-2)(l-3)}{N(N-1)(N-2)(N-3)} \\
h_3(l) &= 4 \frac{l(l-1)(l-2)(N-l)}{N(N-1)(N-2)(N-3)} \\
h_2(l) &= 6 \frac{l(l-1)(N-l)(N-l-1)}{N(N-1)(N-2)(N-3)}
\end{align*}
\]  

A comparison between the analytical expression Eq. (3.14) and the direct numerical solution of the differential Eq. (3.20) for \( N = 1002 \) is presented in Table 3.2, for the quadratic fitness \( f(u) = ku^2/2 \). As in the former case, the numerical and analytical results agree to within \( \mathcal{O}(N^{-1}) \), as expected.

For the quadratic fitness case in the absence of recombination \((\nu = 0)\), the exact analytical result predicts the existence of a "selected" organized phase, or quasispecies, when \( k > \mu \). In this phase, the average composition is given by \( u = 1 - \mu/k \). For \( k < \mu \), a phase transition occurs and the quasispecies disappears in favor of a disordered or "unselected" phase with \( u = 0 \). In Figure 3.3, we display the phase structure in the presence of horizontal gene transfer. In agreement with the numerical results presented in Table 3.1 and Table 3.2, the recombination scheme considered in this model introduces a mild mutational load. However, near the critical region \( k/\mu \sim 1 \), one observes that horizontal gene transfer distorts the phase boundary which defines the error threshold, from the horizontal line \( k/\mu = 1 \), to a monotonically increasing curve that saturates for large values of \( \nu/\mu \). We obtain an analytical
Figure 3.3  Phase diagram of the parallel (Kimura) model for the quadratic fitness $f(u) = ku^2/2$, with horizontal gene transfer of non-overlapping blocks of size $\bar{M}$. The phase boundary of the error threshold phase transition is given by the curve, and its shape is independent of the block size $\bar{M}$. In the absence of horizontal gene transfer, the phase transition occurs at $k/\mu = 1$. 
Table 3.2 Analytical versus numerical results for horizontal gene transfer in the parallel model for the quadratic fitness \( f(u) = ku^2/2 \), with \( \bar{M} = 4 \).

<table>
<thead>
<tr>
<th>( \frac{k}{\mu} )</th>
<th>( \frac{\nu}{\mu} )</th>
<th>( u^{\text{numeric}} )</th>
<th>( u^{\text{analytic}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>0.0</td>
<td>0.4993</td>
<td>0.5000</td>
</tr>
<tr>
<td>2.0</td>
<td>0.5</td>
<td>0.4832</td>
<td>0.4840</td>
</tr>
<tr>
<td>2.0</td>
<td>1.0</td>
<td>0.4672</td>
<td>0.4680</td>
</tr>
<tr>
<td>2.0</td>
<td>1.5</td>
<td>0.4510</td>
<td>0.4519</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0</td>
<td>0.5995</td>
<td>0.6000</td>
</tr>
<tr>
<td>2.5</td>
<td>0.5</td>
<td>0.5916</td>
<td>0.5921</td>
</tr>
<tr>
<td>2.5</td>
<td>1.0</td>
<td>0.5839</td>
<td>0.5845</td>
</tr>
<tr>
<td>2.5</td>
<td>1.5</td>
<td>0.5766</td>
<td>0.5773</td>
</tr>
<tr>
<td>5.0</td>
<td>0.0</td>
<td>0.7998</td>
<td>0.8000</td>
</tr>
<tr>
<td>5.0</td>
<td>0.5</td>
<td>0.7988</td>
<td>0.7990</td>
</tr>
<tr>
<td>5.0</td>
<td>1.0</td>
<td>0.7979</td>
<td>0.7981</td>
</tr>
<tr>
<td>5.0</td>
<td>1.5</td>
<td>0.7970</td>
<td>0.7973</td>
</tr>
</tbody>
</table>
expression for the phase boundary, by expanding Eqs. (3.133) and (3.14) near the critical region $\xi_c \sim 0$, $u \sim 0$. We find that the boundary is defined by

$$k_{\text{crit}} = \mu \frac{1 + \nu/\mu}{1 + \nu/2\mu}. \quad (3.22)$$

We notice from this expression that for small $\nu$, $k_{\text{crit}} \sim \mu + \nu/2$, whereas for large $\nu$ the phase boundary becomes asymptotically independent of $\nu$, $k_{\text{crit}} \sim 2\mu$. We also notice from this formula that the phase boundary is independent of the block size $M$.

As a second example, we consider a square-root fitness function

$$f(u) = k\sqrt{|u|} \quad (3.23)$$

In Table 3.3, we present a comparison of our analytical result, obtained from Eq. (3.14), with the direct numerical solution of the differential Eq. (3.16), for $\tilde{M} = 3$. As in the quadratic fitness example, the analytical and numerical results agree to order $O(N^{-1})$, as expected.

From the results presented in Table 3.3, it is remarkable that the average composition $u$, and correspondingly the mean fitness of the population $f_m = k\sqrt{|u|}$, increase when increasing the horizontal gene transfer rate $\nu$.

The mutational deterministic hypothesis states that recombination is beneficial for negative epistasis fitness functions (see Fig. 3.1) $f''(u) < 0$, and deleterious for positive epistasis fitness functions, $f''(u) > 0$ [41, 45, 46, 42, 43, 44]. Our results for the quadratic and square-root fitness functions, Eqs. (3.14)–(3.22) and Tables 3.1,
Table 3.3 Analytical versus numerical results for horizontal gene transfer in the parallel (Kimura) model for the square-root fitness $f(u) = k\sqrt{|u|}$, with $M = 3$, $N = 801$.

<table>
<thead>
<tr>
<th>$k/\mu$</th>
<th>$\nu/\mu$</th>
<th>$u^{\text{numeric}}$</th>
<th>$u^{\text{analytic}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>0.0</td>
<td>0.4858</td>
<td>0.4855</td>
</tr>
<tr>
<td>2.0</td>
<td>0.5</td>
<td>0.4892</td>
<td>0.4889</td>
</tr>
<tr>
<td>2.0</td>
<td>1.0</td>
<td>0.4918</td>
<td>0.4915</td>
</tr>
<tr>
<td>2.0</td>
<td>1.5</td>
<td>0.4939</td>
<td>0.4936</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0</td>
<td>0.5399</td>
<td>0.5396</td>
</tr>
<tr>
<td>2.5</td>
<td>0.5</td>
<td>0.5428</td>
<td>0.5425</td>
</tr>
<tr>
<td>2.5</td>
<td>1.0</td>
<td>0.5450</td>
<td>0.5448</td>
</tr>
<tr>
<td>2.5</td>
<td>1.5</td>
<td>0.5469</td>
<td>0.5466</td>
</tr>
<tr>
<td>4.0</td>
<td>0.0</td>
<td>0.6525</td>
<td>0.6523</td>
</tr>
<tr>
<td>4.0</td>
<td>0.5</td>
<td>0.6542</td>
<td>0.6540</td>
</tr>
<tr>
<td>4.0</td>
<td>1.0</td>
<td>0.6556</td>
<td>0.6554</td>
</tr>
<tr>
<td>4.0</td>
<td>1.5</td>
<td>0.6568</td>
<td>0.6565</td>
</tr>
</tbody>
</table>
3.2, and 3.3 provide support for this hypothesis. In fact, we can prove the mutational
deterministic hypothesis holds for the parallel model in the presence of horizontal
gene transfer, Appendix 3.12.

Horizontal gene transfer has less of an effect for the sharp peak fitness, \( f(u) = A\delta_{u,1} \). For general \( \bar{M} \), the maximum in Eq. (3.14) is achieved for \( \xi_c = 1 \), with
\[
\phi_c(1) = (1 + u)/2 \text{ from Eq. (3.133)}.
\]
Thus, one obtains
\[
f_m = A - \mu + \frac{\nu}{\bar{M}} \left[ 1 - \left( \frac{1 + u}{2} \right)^\bar{M} \right]. \tag{3.24}
\]
The error threshold is given for \( u = 0 \) by the condition \( A > \mu + \frac{\nu}{\bar{M}} (1 - 2^{-\bar{M}}) \). However, we notice from Eq. (3.24) that \( f_m(u = 1) = A - \mu > f_m(u = 0) \). Therefore, we have
\( u = 1 - \mathcal{O}(N^{-1}) \) in the selected phase, with the effect of horizontal gene transfer being negligible for finite \( \bar{M} \). We obtain the fraction of the population located at the peak \( P_0 \), from the self-consistency condition \( P_0 A = f_m \), which yields \( P_0 = 1 - \mu/A \). Thus, the true error threshold is at \( A_{\text{crit}} = \mu \), with the condition \( A > \mu + \frac{\nu}{\bar{M}} (1 - 2^{-\bar{M}}) \) defining the limit of metastability for initial conditions with \( u \sim 0 \). These results are similar to the ones obtained in the absence of horizontal gene transfer \([29, 40, 107]\). Thus, we conclude that for the sharp peak fitness, horizontal gene transfer does not spread out the population in sequence space. This result differs from the numerical studies presented in \([38]\), where a mathematical model for 'uniform crossover' recombination between viral strains super-infecting a population of cells was described. We remark that this model studied sequences of finite length \( (N = 15) \), where the error threshold
transition is not really sharp. Our results correspond to the more realistic limit $N \to \infty$ (typical viral genomes are $10^3 - 10^4$).

In summary, from our exact analytical formula for the mean fitness Eq. (3.14), which is valid for any permutation invariant replication rate, we developed the explicit solution of three different examples: a quadratic fitness, a square-root fitness and a single sharp peak. For the case of smooth fitness functions, from our exact analytical formulas for the mean fitness $f_m$ and average composition $u$, we conclude that in agreement with the mutational deterministic hypothesis [41, 45, 42, 43, 44], a population whose fitness represents positive epistasis (i.e. quadratic), will experience an additional load against selection due to horizontal gene transfer. On the contrary, when negative epistasis is present (e.g. square-root), horizontal gene transfer is beneficial by enhancing selection. We provided a mathematical proof for this effect, Appendix 3.12. When the fitness is defined by a single sharp peak, the population steady-state distribution behaves more rigidly in response to horizontal gene transfer. This fundamental difference can be attributed to the structure of the quasispecies distribution, which in the smooth fitness case is a Gaussian centered at the mean fitness, while in the sharp peak it is a fast decaying exponential, sharply peaked at the master sequence [29]. While the Gaussian distribution spreads its tails over a wide region of sequence space, thus allowing for horizontal gene transfer effects to propagate over a large diversity of mutants, the sharp exponential distribution con-
centrates in a narrow neighborhood of the master sequence, acting as a barrier to the propagation of such effects.

3.2.2 Horizontal gene transfer for multiple-size blocks

A natural extension to the model of horizontal gene transfer involving blocks of genes of a given size is to consider a process where each site along the sequence may be transferred with probability $\gamma$, or left intact with probability $1 - \gamma$. The operator describing this process is

$$\hat{R} = \frac{1}{\langle M \rangle} \prod_{j=1}^{N} \left[ (1 - \gamma) \hat{I}_j + \gamma \hat{R}_j \right] - \frac{1}{\langle M \rangle} \hat{I}. \quad (3.25)$$

Here, $\hat{R}_j = \hat{a}^\dagger(j) \hat{D} \hat{a}(j)$ is the single-site recombination operator defined in Eq. (3.5), with the matrix $D$ defined as in Eq. (3.6). Notice that this operator represents a binomial process, where an average number of sites $\langle M \rangle = \gamma N$ is transferred. If we consider, as in the former finite block size case, that $N/\langle M \rangle = \mathcal{O}(N)$, then we have $\gamma = \langle M \rangle/N$, and for very large $N$ Eq. (3.25) reduces to

$$\hat{R} = \frac{1}{\langle M \rangle} \prod_{j=1}^{N} \left[ (1 - \gamma) \hat{I}_j + \gamma \hat{R}_j \right] - \frac{1}{\langle M \rangle} \hat{I} \sim \frac{1}{\langle M \rangle} e^{-\langle M \rangle + \langle M \rangle \sum_{j=1}^{N} \hat{a}^\dagger(j) \hat{D} \hat{a}(j)} - \frac{1}{\langle M \rangle} \hat{I}. \quad (3.26)$$

Considering the recombination operator defined in Eq. (3.26), the spin Boson
Hamiltonian for the Kimura model becomes

$$-\hat{H} = Nf \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{a}^\dagger(j) \sigma_3 \tilde{a}(j) \right] + \mu \sum_{j=1}^{N} [\tilde{a}(j)^\dagger \sigma_1 \tilde{a}(j) - \hat{N}] + \frac{\nu}{\langle M \rangle} Ne^{-(\langle M \rangle \phi)} \sum_{j=1}^{N} \tilde{a}(j) \tilde{D}(j)$$

$$- \frac{\nu}{\langle M \rangle} N \hat{N}. \quad (3.27)$$

We introduce a Trotter factorization

$$e^{-\hat{H}t} = \lim_{M \to \infty} \int \mathcal{D} \tilde{z}^* \mathcal{D} \tilde{z} | \tilde{z}_M \rangle \left( \prod_{k=1}^{M} (\tilde{z}_k | e^{-\hat{H}t} | \tilde{z}_{k-1} \rangle) \right) \langle \tilde{z}_0 |. \quad (3.28)$$

As shown in Appendix 3.3, the partition function becomes

$$Z = \int \mathcal{D} \tilde{z} | e^{-S[\tilde{z}, \phi]} \rangle \sim e^{N f_m t}. \quad (3.29)$$

Here, the action in the continuous time limit is

$$S[\xi, \xi, \phi, \phi] = -N \int_0^t dt' \left[ -\tilde{\xi} \dot{\xi} - \tilde{\phi} \dot{\phi} - \mu - \frac{\nu}{\langle M \rangle} + f(\xi) + \frac{\nu}{\langle M \rangle} e^{-\langle M \rangle (1-\phi)} \right] - N \ln Q \quad (3.30)$$

**The saddle point limit**

As in the previous model, the saddle point limit is exact as $N \to \infty$ in Eq. (3.30).

After a similar procedure as in section II.A.2, we find the saddle-point equation for the mean fitness

$$f_m = \max_{-1 \leq \xi \leq 1} \left\{ f(\xi) - \mu - \frac{\nu}{\langle M \rangle} e^{-(\langle M \rangle (1-\phi(\xi)))} + \mu \sqrt{1-\xi^2} \frac{-\sqrt{1-\frac{\nu^2}{2}} e^{-(\langle M \rangle (1-\phi(\xi)))}}{\left[1 + \frac{\nu}{2} (1-u^2) e^{-(\langle M \rangle (1-\phi(\xi)))} \right]^2 - u^2} \right\}^{1/2} \quad (3.31)$$
Here, $\phi_c(\xi_c)$ is obtained from the equation

$$
\phi_c(\xi_c) = \frac{1 + u\xi_c}{2} + \frac{\sqrt{1 - \xi_c^2}}{2} \left[ \frac{\sqrt{1 - u^2}}{1 - \left(\frac{u}{1 + \frac{u}{\sqrt{2}} (1-u^2)e^{-(\langle M \rangle(1-\phi_c))}}\right)^2} \right]^{1/2}
$$

(3.32)

Eq. (3.31) represents an exact analytical expression for the mean fitness $f_m$ of an infinite population experiencing horizontal gene transfer of multiple size sequences. The formula is valid for an arbitrary, permutation invariant replication rate function $f(u)$.

We notice that recombination introduces an additional mutational load against selection. This load is mild at low values of the fitness constant $k$, and becomes negligibly small at larger values. Numerical evaluation of Eqs. (3.31) and (3.32) is presented in Table 3.4 for the quadratic fitness $f(u) = ku^2/2$, and average block size $\langle \langle M \rangle \rangle = 3$.

An analytical expression for the phase boundary is obtained from Eqs. (3.31) and (3.32), near the error threshold $u \sim 0$, $\xi_c \sim 0$. We find

$$
k_{\text{crit}} = \frac{\mu}{1 + \frac{\nu}{2\mu}}
$$

(3.33)

We notice that for small $\nu$, the critical value is $k_{\text{crit}} \sim \mu + \nu/2$, whereas for large values of $\nu$ it becomes independent of recombination $k_{\text{crit}} \sim 2\mu$. This behavior is similar to the one previously observed in Fig. 3.3 for the case of horizontal gene transfer with blocks of fixed size. The shape of the phase boundary is independent of the block size in the horizontal gene transfer process, assuming that the size of the blocks is finite.
Table 3.4 Analytical results for horizontal gene transfer in the parallel model for the quadratic fitness $f(u) = \frac{k}{2}u^2$, with $\langle M \rangle = 3$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\nu$</th>
<th>$u^{\text{analytic}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>0.0</td>
<td>0.50</td>
</tr>
<tr>
<td>2.0</td>
<td>0.5</td>
<td>0.4840</td>
</tr>
<tr>
<td>2.0</td>
<td>1.0</td>
<td>0.4680</td>
</tr>
<tr>
<td>2.0</td>
<td>1.5</td>
<td>0.4522</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0</td>
<td>0.6000</td>
</tr>
<tr>
<td>2.5</td>
<td>0.5</td>
<td>0.5921</td>
</tr>
<tr>
<td>2.5</td>
<td>1.0</td>
<td>0.5845</td>
</tr>
<tr>
<td>2.5</td>
<td>1.5</td>
<td>0.5773</td>
</tr>
<tr>
<td>4.0</td>
<td>0.0</td>
<td>0.8000</td>
</tr>
<tr>
<td>4.0</td>
<td>0.5</td>
<td>0.7990</td>
</tr>
<tr>
<td>4.0</td>
<td>1.0</td>
<td>0.7981</td>
</tr>
<tr>
<td>4.0</td>
<td>1.5</td>
<td>0.7973</td>
</tr>
</tbody>
</table>
As a second example, we consider the square root fitness \( f(u) = k\sqrt{|u|} \). Analytical results for the average composition, obtained after Eq. (3.14), are represented in Table 3.5 for blocks of average size \((\langle M \rangle = 3)\). From the values displayed in Table 3.5, we notice that horizontal gene transfer introduces a mild increase in the average composition and, correspondingly, in the mean fitness of the population \( f_m = k\sqrt{|u|} \). This trend, which is opposite to the quadratic fitness case, can be attributed to the negative epistasis represented by the square root fitness, by similar arguments as in the case of fixed block size.

Horizontal gene transfer does not affect the phase boundary for the sharp peak fitness, \( f(u) = A \delta_{u,1} \). In this case, Eq. (3.31) is maximized at \( \xi_c = 1 \), with \( \phi_c = (1 + u)/2 \) from Eq. (3.32). Thus, the mean fitness becomes

\[
f_m = A - \mu - \frac{\nu}{\langle M \rangle} [1 - e^{-\langle M \rangle(1-u)/2}]
\]

The error threshold is given, for \( u = 0 \) in Eq. (3.34), by the condition \( A > \mu - \frac{\nu}{\langle M \rangle} [1 - e^{-\langle M \rangle/2}] \). However, we notice that \( f_m(u = 1) = A - \mu > f_m(u = 0) \). Hence, in the selected phase \( u = 1 - \mathcal{O}(N^{-1}) \), and the recombination effect becomes negligible for infinite \( N \). From the self-consistency condition \( f_m = P_0 A \), we obtain the fraction of the population located at the peak \( P_0 = 1 - \mu/A \). Therefore, the true error threshold is given by \( A_{\text{crit}} > \mu \), with \( A > \mu - \frac{\nu}{\langle M \rangle} [1 - e^{-\langle M \rangle/2}] \) the limit of metastability for initial conditions with \( u \sim 0 \).

Therefore, we conclude that horizontal gene transfer for multiple size blocks dis-
Table 3.5  Analytical results for horizontal gene transfer in the parallel model for the square-root fitness \( f(u) = k \sqrt{|u|} \), with \( \langle \hat{M} \rangle = 3 \).

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \nu )</th>
<th>( u^{\text{analytic}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>0.0</td>
<td>0.4855</td>
</tr>
<tr>
<td>2.0</td>
<td>0.5</td>
<td>0.4889</td>
</tr>
<tr>
<td>2.0</td>
<td>1.0</td>
<td>0.4915</td>
</tr>
<tr>
<td>2.0</td>
<td>1.5</td>
<td>0.4936</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0</td>
<td>0.5396</td>
</tr>
<tr>
<td>2.5</td>
<td>0.5</td>
<td>0.5425</td>
</tr>
<tr>
<td>2.5</td>
<td>1.0</td>
<td>0.5448</td>
</tr>
<tr>
<td>2.5</td>
<td>1.5</td>
<td>0.5466</td>
</tr>
<tr>
<td>5.0</td>
<td>0.0</td>
<td>0.6523</td>
</tr>
<tr>
<td>5.0</td>
<td>0.5</td>
<td>0.6540</td>
</tr>
<tr>
<td>5.0</td>
<td>1.0</td>
<td>0.6554</td>
</tr>
<tr>
<td>5.0</td>
<td>1.5</td>
<td>0.6566</td>
</tr>
</tbody>
</table>
plays a qualitatively similar behavior to the corresponding process for fixed block size. A population evolving under a smooth fitness function with positive epistasis (e.g. quadratic, see Fig. 3.1) experiences an additional mutational load due to horizontal gene transfer, which modifies the quasispecies structure, reducing the mean fitness, and hence shifting the error threshold. On the contrary, when epistasis is negative (e.g. square-root, see Fig. 3.1) a beneficial effect is induced by horizontal gene transfer, in agreement with the mutational deterministic hypothesis, as we demonstrate in Appendix 3.12.

A discontinuous sharp peak fitness function does not change the quasispecies distribution or the mean fitness, although it does introduce metastability.

3.2.3 The parallel model with two-parent recombination

Biological recombination, as occurs for example in viral super- or co-infection or in sexual reproduction, involves the crossing over of parental strands at random points along the sequence. The copying process is carried out by the action of polymerase enzymes, which move alternatively along one or the other parental strand. An approximate representation of this process is to consider that the polymerase enzyme starts, with probability 1/2 on either parental strand, copying one base at a time. We consider the crossovers to occur because there exists a probability $p_c$ per site that the polymerase "jumps" from its current position towards the other parental strand. Alternatively, the enzyme progresses along the current strand with probability $1 - p_c$. 
A pictorial representation is shown in Fig. 3.4.

For this particular process representing the wandering path followed by the polymerase enzyme, the recombination coefficients $R^i_{kl}$ in Eq. (3.1) are given by the exact analytical expression

$$R^i_{kl} = \frac{1}{2} \sum_{\{\alpha_j = \pm 1\}} \left( \frac{1 + s^k_i s^i_j}{2} \right)^{\frac{1+\alpha_j}{2} \frac{1-\alpha_j}{2}} \left( \frac{1 + s^l_i s^i_j}{2} \right)^{\frac{1+\alpha_l}{2} \frac{1-\alpha_l}{2}}$$

$$\times [(1 - p_c)^{\frac{1+\alpha_l}{2} \frac{1-\alpha_l}{2}} p_c^\frac{1-\alpha_l}{2} \frac{1+\alpha_l}{2}] \left( \frac{1 + s^k_i s^i_j}{2} \right)^{\frac{1+\alpha_k}{2} \frac{1-\alpha_k}{2}} \left( \frac{1 + s^l_i s^i_j}{2} \right)^{\frac{1+\alpha_l}{2} \frac{1-\alpha_l}{2}}$$

$$\times [(1 - p_c)^{\frac{1+\alpha_k}{2} \frac{1-\alpha_k}{2}} p_c^\frac{1-\alpha_k}{2} \frac{1+\alpha_k}{2}] \left( \frac{1 + s^k_i s^i_j}{2} \right)^{\frac{1+\alpha_k}{2} \frac{1-\alpha_k}{2}} \left( \frac{1 + s^l_i s^i_j}{2} \right)^{\frac{1+\alpha_l}{2} \frac{1-\alpha_l}{2}}$$

$$\times \ldots \times [(1 - p_c)^{\frac{1+\alpha_{N-1}}{2} \frac{1-\alpha_{N-1}}{2}} p_c^\frac{1-\alpha_{N-1}}{2} \frac{1+\alpha_{N-1}}{2}] \left( \frac{1 + s^k_i s^i_j}{2} \right)^{\frac{1+\alpha_k}{2} \frac{1-\alpha_k}{2}} \left( \frac{1 + s^l_i s^i_j}{2} \right)^{\frac{1+\alpha_l}{2} \frac{1-\alpha_l}{2}}$$

(3.35)

Here, the recombining parental sequences are $S_k = (s^k_1, s^k_2, \ldots, s^k_N)$, $S_l = (s^l_1, \ldots, s^l_N)$ and the offspring sequence is $S_i = (s^i_1, s^i_2, \ldots, s^i_N)$, with $s_j = \pm 1$. Using Eq. (3.35), Eq. (3.1) representing the time evolution of an infinite population of binary sequences experiencing replication, point mutations and two-parent recombination, exactly becomes

$$\frac{dq_i}{dt} = r_i q_i + \sum_{k=1}^{2^N} \mu_k q_k + \nu N \sum_{k=1}^{2^N} \frac{1}{2} \sum_{\{\alpha_j = \pm 1\}} \left\{ \prod_{j=2}^{N} p_c^{\frac{1-\alpha_j}{2}} (1 - p_c)^{\frac{1+\alpha_j}{2}} \right\}$$

$$\times \prod_{j=1}^{N} \left( \frac{1 + s^k_j s^j_k}{2} \right)^{\frac{1+\alpha_k}{2} \frac{1-\alpha_k}{2}} \sum_{l=1}^{2^N} p_l \left( \frac{1 + s^l_i s^i_l}{2} \delta_{s^i_l,+1} + \frac{1 - s^l_i s^i_l}{2} \delta_{s^i_l,-1} \right)^{\frac{1-\alpha_l}{2}} q_k - \nu N q_i$$

(3.36)
Figure 3.4 Pictorial representation of the two-parent genetic recombination process considered in the theory.
where, again, \( p_l = q_l / \sum_{l=1}^{2^N} q_l \) is the normalized probability for sequence \( 1 \leq l \leq 2^N \).

From Eq. (3.36), the recombination operator corresponding to this recombination process in the spin Boson representation is

\[
\hat{R} = \frac{1}{2} \sum_{l=1}^{2^N} p_l \sum_{\{ \alpha_i = \pm 1 \}} \left[ I_1^{1+q_1} \hat{R}_l(1)^{1-q_1} \right] \times \left[ (1 - p_c)^{1+q_2} p_c^{1-q_2} \right] \\
\times \left[ I_2^{1+q_3} \hat{R}_l(2)^{1-q_3} \right] \times \left[ (1 - p_c)^{1+q_4} p_c^{1-q_4} \right] \times \left[ I_3^{1+q_5} \hat{R}_l(3)^{1-q_5} \right] \\
\times \ldots \times \left[ (1 - p_c)^{1+q_{N-1}} p_c^{1-q_{N-1}} \right] \times \left[ I_N^{1+q_N} \hat{R}_l(N)^{1-q_N} \right] - \hat{I}
\]

(3.37)

Here, the local recombination operator is \( \hat{R}_l(j) = \hat{a}(j)^\dagger D_j^l \hat{a}(j) \), with

\[
D_j^l = \begin{pmatrix}
\frac{1 + s_j^l}{2} & \frac{1 + s_j^l}{2} \\
\frac{1 - s_j^l}{2} & \frac{1 - s_j^l}{2}
\end{pmatrix}
\]

(3.38)

The \( \hat{I}_j \) are the identity operators acting on site \( 1 \leq j \leq N \), whereas \( \hat{I} = \prod_{j=1}^N \hat{I}_j \) is the identity operator for the entire sequence vector.

**The Hamiltonian**

The Hamiltonian describing the evolution of this system in the spin Boson representation is given by

\[
-\hat{H} = N \frac{1}{N} \sum_{j=1}^{N} \hat{a}(j)^\dagger \sigma_3 \hat{a}(j) + \mu \sum_{j=1}^{N} \left[ \hat{a}(j)^\dagger \sigma_1 \hat{a}(j) - \hat{I} \right] + \nu N \left( g(\{ \hat{R}_l(j) \}) - \hat{I} \right)
\]

(3.39)
We introduce a Trotter factorization
\[
e^{-\hat{H} t} = \lim_{M \to \infty} \int [D\hat{z}^\dagger D\hat{z}] |\hat{z}_M\rangle \left( \prod_{k=1}^M \langle \hat{z}_k | e^{-\epsilon \hat{H} t} | \hat{z}_{k-1} \rangle \right) \langle \hat{z}_0 | \tag{3.40}
\]
As shown in Appendix 3.4 the partition function is
\[
Z = \int [D\hat{\xi} D\hat{\phi}] e^{-S[\hat{\xi}, \hat{\phi}, \phi]} \tag{3.41}
\]
Here, the action in the continuous time limit is given by
\[
S[\hat{\xi}, \hat{\phi}, \phi] = -N \int_0^t dt \left[ -\hat{\xi} \dot{\hat{\xi}} - \dot{\hat{\phi}} \phi - \mu - \nu + f(\xi) + \nu g(\phi) \right] - N \ln Q \tag{3.42}
\]
As shown in Appendix 3.5, the recombination term can be represented, for $0 \leq p_c \leq 1/2$, by the exact finite series
\[
g(\{\psi^l_j\}) = \sum_{l=1}^{2^N} p_l \left\{ \prod_{j=1}^{N} \left( \frac{1 + \psi^l_j}{2} \right) \right. \\
+ \sum_{1 \leq i < j}^N (1 - 2p_c)^{j-i-1} \frac{1 - \psi^l_j}{2} \frac{1 - \psi^l_i}{2} \prod_{k \neq i, j}^N \left( \frac{1 + \psi^l_k}{2} \right) \\
+ \sum_{1 \leq i < j < k < n}^N (1 - 2p_c)^{j-i+n-k} \frac{1 - \psi^l_j}{2} \frac{1 - \psi^l_i}{2} \frac{1 - \psi^l_n}{2} \frac{1 - \psi^l_k}{2} \\
\times \prod_{m \neq i, j, k, n}^N \left( \frac{1 + \psi^l_m}{2} \right) + \ldots + (1 - 2p_c)^{\frac{1}{2}^N} \prod_{j=1}^N \left( \frac{1 - \psi^l_j}{2} \right) \right\} \tag{3.43}
\]
were we used the notation $\psi^l_j = \hat{z}^*_{k}(j) D^l_j \hat{z}_{k-1}(j)$, and $D^l_j$ is defined in Eq. (3.38).

We consider first the case when $p_c = 1/2$ in the above expression. Then, we have
\[
g(\{\psi^l_j\}, p_c = 1/2) = \sum_{l=1}^{2^N} p_l \prod_{j=1}^{N} (1 + \psi^l_j)/2 \tag{3.44}
\]
We notice that the recombination term in the differential Eq. (3.36) satisfies \( \sum_{i=1}^{2N} p_l R_{kl}^i \leq 1, \forall k, i, \) because \( R_{kl}^i \geq 0 \) and \( \sum_{i=1}^{2N} R_{kl}^i = 1. \) In our field theoretic representation of the model, this condition is equivalent to \( g(\{\psi_j^i\}) \leq 1. \) Hence, for any physical state, the product \( \prod_{j=1}^{N} \left( \frac{1+\psi_j^i}{2} \right) \leq 1. \) This condition imposes a boundary for the values of \( \psi_j^i = z_k^i(j) D_j^i z_{k-1}(j) \) in any physical state,

\[
|z^i D_j^i z| \leq 1. \tag{3.45}
\]

For \( N \) very large, we notice that the product in the expression for \( g \) Eq. (3.44) will be \( \sim 0, \) unless \( \mathcal{O}(N) \) of the \( \psi_j^i = 1. \)

For the general case of \( 0 < p_c < 1/2, \) we notice that \( 0 < 1 - 2p_c < 1. \) When \( \psi_j^i \sim 1, \) the first term dominates the series, and the others become arbitrarily small, thus recovering the same expression as for \( p_c = 1/2. \) On the other hand, when \( \psi_j^i \sim -1, \) we notice that the dominant terms are the last ones. However, those terms are proportional to powers of \( 1 - 2p_c \) of order \( N, \) whereas the numerator of these terms is of just polynomial order in \( N. \) Therefore, for \( N \) very large these terms become arbitrary small for \( \psi_j^i < 0. \) Then, we conclude that in the limit \( N \to \infty, \) regardless of the value of \( p_c, \) the function \( g \) is represented by Eq. (3.44).

In the particular case of uniform crossover \( p_c = 1/2, \) when the fitness function is permutation invariant, i.e., it depends only on the average composition of the sequence through the average base composition \( u, \) it is possible to reformulate the differential equation Eq. (3.1) for the evolutionary dynamics of an infinite population of binary
sequences in terms of the distribution of classes:

\[ P_i = \sum_{j \in C_i} p_j \]  

(3.46)

where \( C_i \) represents the class of sequences with \( l \), "—1" spins. Although all the sequences in a given class do not have the same dynamics, we can nonetheless calculate the class dynamics exactly:

\[
\frac{dP_i}{dt} = N \left[ f(2l/N - 1) - \sum_{l'=0}^{N} P_{l'} f(2l'/N - 1) \right] P_i + \mu(N - l + 1)P_{i-1} + \mu(l + 1)P_{i+1} \\
- N\mu P_i + \nu \sum_{l_1, l_2} R(l|l_1, l_2)P_{l_1}P_{l_2} - N\nu P_i,
\]  

(3.47)

The coefficients \( R(l|l_1, l_2) \) represent the probability that a pair of parental sequences in the classes \( C_{l_1}, C_{l_2} \), due to uniform crossover recombination, generate a child sequence in the class \( C_l \). The number of sequences in these classes is \( \binom{N}{l_1}, \binom{N}{l_2} \) and \( \binom{N}{l} \), respectively. For a given pair of parental sequences, let us consider the variables \( n_{++}, n_{+-}, n_{-+} \) and \( n_{--} \), representing the number of pairs of \((+1,+1)\), \((+1,-1)\), \((-1,+1)\) and \((-1,-1)\) spins respectively. These variables satisfy the equation \( N = n_{++} + n_{+-} + n_{-+} + n_{--} \). We further notice that these variables also satisfy \( n_{-+} = l_1 - n_{--} \) and \( n_{+-} = l_2 - n_{--} \). Considering that from each pair of \((+1,-1)\) or \((-1,+1)\) spins in the parental sequences, the child sequence will inherit a "—1" spin with probability \(1/2\), while from a pair of the kind \((-1,-1)\) it will inherit a "—1" spin.
with probability 1, we have the explicit analytical expression for these coefficients

\[
R(l|l_1, l_2) = \sum_{n=\max\{0, l_1 + l_2 - N\}}^{\min\{l_1 + l_2 - l, l_1, l_2\}} \binom{N}{n_1-n_2-n} \binom{N}{l_1} \binom{N}{l_2} \left(\frac{l_1 + l_2 - 2n}{l - n}\right)^2 2^{-(l_1 + l_2 - 2n)}
\]

(3.48)

The first factor is the probability for a configuration with \( n \equiv n_{--} \), given \( l_1, l_2 \) and \( l \). The second factor is the number of ways of picking \( l - n_{--} \) "1" spins among \( n_{++} + n_{--} \). The third factor is just \((1/2)^{n_{++}}(1/2)^{n_{--}}(1)^{n_{--}}\). These coefficients are different from zero only if

\[
\max\{0, l_1 + l_2 - N\} \leq l \leq \min\{N, l_1 + l_2\}
\]

(3.49)

They also satisfy the following properties:

\[
R(l|l_1, l_2) = R(l|l_2, l_1)
\]

(3.50)

\[
\sum_{l=0}^{N} R(l|l_1, l_2) = 1 \quad \forall l_1, l_2
\]

(3.51)

\[
R(N|N, N) = R(0|0, 0) = 1
\]

(3.52)

In the limit of large \( N \), we find that the recombination coefficients satisfy a Gaussian distribution in the variables \( u_1 = 1 - 2l_1/N \), \( u_2 = 1 - 2l_2/N \), and \( u = 1 - 2l/N \) (see Appendix 3.6):

\[
R_{u_1, u_2}^u \sim \frac{e^{-N[(u_1 + u_2)/2 - u]^2/(1-u_2^2)}}{\sqrt{\pi(1-u_2^2)/N}}
\]

(3.53)
where \( f_m = f(u_*) \).

This form of the recombination operator, Eq. (3.53), is equivalent to Eq. (3.44) with \( s_j^l \) replaced by \( u \) in the \( D \) matrix. Alternatively, we notice that when the singular behavior of the function \( g \) can be described as a delta function, we have

\[
g = \sum_{l=1}^{2N} p_l \delta_1^N \sum_{j=1}^{2N} \mathbb{E}(j) D_j^l \mathbb{E}_{k-1}(j,1)
\]

\[
= \sum_{l=1}^{2N} p_l \int_0^{2\pi} \frac{d\lambda}{2\pi} e^{i\lambda} \left[ \frac{1}{N} \sum_{j=1}^{2N} \mathbb{E}(j) D_j^l \mathbb{E}_{k-1}(j,1) \right]
\]

\[
= \sum_{l=1}^{2N} p_l \int_0^{2\pi} \frac{d\lambda}{2\pi} e^{-i\lambda} \left\{ 1 + \frac{i\lambda}{N} \sum_{j=1}^{2N} \mathbb{E}(j) D_j^l \mathbb{E}_{k-1}(j) \right. \\
+ \frac{1}{2!} \left( \frac{i\lambda}{N} \right)^2 \sum_{j,m=1}^{2N} \mathbb{E}(j) D_j^l \mathbb{E}_{k-1}(j) \mathbb{E}(m) D_m^l \mathbb{E}_{k-1}(m) + \ldots \right\} 
\]  

(3.54)

By noticing that correlations between compositions at different sites along the sequence are of order \( O(N^{-1}) \), we have that for the second order correlation

\[
\langle D_j^l D_m^l \rangle - \langle D_j^l \rangle^2 \sim O(N^{-1})
\]

(3.55)

where \( \langle D_j^l \rangle = \sum_{l=1}^{2N} p_l D_j^l \equiv D_j \) is the population average. A similar analysis for the higher order correlations allows us to factorize order by order the terms in the series Eq. (3.54), to obtain

\[
g \sim \delta_1^N \sum_{j=1}^{2N} \mathbb{E}(j) D_j \mathbb{E}_{k-1}(j,1) + O(N^{-1})
\]

(3.56)

We are interested in the long term, steady state distribution, when the average base composition \( u(j) = \langle s_j^l \rangle \sim u \) becomes independent of time. In this limit, the trace
defined by Eq. (3.152) becomes

\[
\lim_{t \to \infty} \frac{\ln Q_c}{t} = \frac{\phi_c}{2} + [\xi_c(\xi_c + u\phi_c) + (\mu + \phi_c/2)^2]^{1/2}
\] (3.57)

Hence, from Eq. (3.42), the saddle point action is

\[
\lim_{N,t \to \infty} \frac{\ln Z}{Nt} = \lim_{t \to \infty} -\frac{S_c}{Nt} = f_m
\]

\[
= \max_{\xi_c,\xi_c,\phi_c,\phi_c} \left\{ -\xi_c\xi_c - \phi_c\phi_c - \mu - \nu + \nu g(\phi_c)
\right.
\]

\[
+ f(\xi_c) + \frac{\phi_c}{2} + \left[ \xi_c(\xi_c + u\phi_c) + \left(\mu + \frac{\phi_c}{2}\right)^2 \right]^{1/2}\}
\] (3.58)

As shown in Appendix 3.7, we find

\[
-\frac{S_c}{Nt} = \max_{\phi_c,\xi_c} \left\{ f(\xi_c) - \mu - \nu + \nu g(\phi_c) + \frac{2\phi_c - 1 - \mu}{1 - u^2}(2\phi_c - 1 - u\xi_c)
\right.
\]

\[
- \frac{\mu|u|}{1 - u^2} \left[ (2\phi_c - 1 - u\xi_c)^2 - (1 - u^2)(1 - \xi_c^2) \right]^{1/2}\}
\] (3.59)

Because of the singular behavior of the function \(g(\phi_c)\), to find the saddle point we need to consider three separate cases: \(\phi_c < 1\), \(\phi_c = 1\), and \(\phi_c = 1 - O(1/N)\).

The existence of different expressions for the mean fitness suggests the possibility of different selected phases in certain conditions. We also notice that the saddle point analysis may not apply exactly, unless \(g(\phi_c) = \delta_{\phi_c,1}\).

Case 1: \(\phi_c < 1\). For this case, we look for a saddle point in the field \(\phi_c\), in the interior of the domain, \(\phi_c < 1\) where \(g(\phi_c) = 0\)

\[
\frac{\delta}{\delta\phi_c} \left( -\frac{S_c}{Nt} \right) = \frac{2\mu}{1 - u^2} - \frac{\mu|u|}{1 - u^2} \left( 2(2\phi_c - 1 - u\xi_c) - (1 - u^2)(1 - \xi_c^2) \right]^{1/2} = 0
\] (3.60)
From Eq. (3.60), we solve for $\phi_c$ as a function of $\xi_c$

$$
\phi_c(\xi_c) = \frac{1 + u\xi_c}{2} + \frac{1}{2}\sqrt{1 - \xi_c^2} 
$$  

Substituting Eq. (3.61) in the saddle-point action Eq. (3.59), we obtain

$$
f_m^{(1)} = \max_{-1 \leq \xi_c \leq 1} \left\{ f(\xi_c) - \mu - \nu + \mu\sqrt{1 - \xi_c^2} \right\} 
$$  

Case 2: $\phi_c = 1$. The mean fitness is obtained from Eq. (3.59) as

$$
f_m^{(2)} = \max_{-1 \leq \xi_c \leq 1} \left\{ f(\xi_c) - \mu + \frac{\mu}{1 - u}\left(1 - u\xi_c - |u\xi_c - u^2|\right) \right\} 
$$  

Case 3: $\phi_c = 1 - \mathcal{O}(1/N)$. In this case, additional analysis is necessary to calculate the mean fitness due to the singular behavior of the $g(\phi_c)$ function. For a smooth fitness function, we can argue this case does not exist. We first consider the Hamiltonian (3.39) for the case $g = 0$. The largest eigenvalue, $f_m$, is shifted by $-\nu$ relative to the $\nu = 0$ case. This allows us to calculate the average composition, $u_*$, from the implicit relation $f_m(\nu) = f_m(\nu = 0) - \nu = f(u_*)$. Alternatively, if we consider the differential equation for the unnormalized class probabilities, $dQ/dt = LQ$, we see that the differential operator $L$ looks like that in the absence of recombination, save for a shift of $-\nu$ in the fitness function. Thus, the variance of the population is given by [29] $\sigma_u^2/N = 2\mu u_*/[Nf'(u_*)]$. Considering more carefully the $g$ function, we find

$$
\int du_1 du_2 R_{u_1 u_2}^u P(u_1) P(u_2) = \exp[-N(u - u_*)^2/(2\sigma^2)]/\sqrt{2\pi\sigma^2 N},
$$

with $\sigma^2 = \sigma_u^2/2 + (1 - u_*)^2/2$. This term is exponentially negligible compared to the
\[ -\nu P(u) \text{ term when } \sigma^2 < \sigma_u^2, \text{ since } P(u) = \exp[-N(u-u_*)^2/(2\sigma_u^2)]/\sqrt{2\pi\sigma_u^2N}. \]

In other words, we must strictly be in case 1 when

\[ 1 - u_*^2 < 2\mu u_*/f'(u_*). \]  

(3.64)

We denote the value of \( \nu \) at which

\[ 1 - u_*^2 = 2\mu u_/f'(u_*) \text{ at } \nu = \nu_* \]  

as \( \nu_* \). Now, at this value of \( \nu_* \) we have \( \int du_1 du_2 R_{u_1u_2} P(u_1)P(u_2) = P(u) \). Thus, the term proportional to \( \nu \) in Hamiltonian (3.39), or differential equation (3.47), exactly vanishes. Thus, we have \( df_m/d\nu = 0 \) and \( dP(u)/d\nu = 0 \) at this value of \( \nu \). There is spectral rigidity. This implies that for \( \nu > \nu_* \), the distribution \( P(u) \) is independent of \( \nu \), and that the value of \( u_* \) is constant. In other words, the value of \( f_m \) in case 2 must be constant with \( \nu \). Assuming \( f_m \) varies continuously with \( \nu \) in case 1, and that the fitness values for case 1 and case 2 are equal at a single value of \( \nu \), therefore, case 2 is simply case 1 with the value \( \nu = \nu_* \).

\[ f_m(\nu > \nu_*) = f_m(\nu = \nu_*) \]  

(3.66)

Eqs. (3.62), (3.63) provide an exact analytical solution for the mean fitness of an infinite population, for a general permutation invariant replication rate represented by a continuous, smooth function \( f(u) \).

For a non-smooth fitness function, additional analysis is necessary, since \( f'(u_*) \) is undefined, and \( P(u) \) may no longer be Gaussian.
Examples and numerical tests

We investigate the phase diagrams, as predicted from our theoretical equations Eqs. (3.62), (3.63) for three different fitness functions: A sharp peak, a quadratic fitness landscape and a square-root fitness landscape.

For the sharp peak landscape \( f(u) = A\delta_{u,1} \), we notice that the maximum is achieved at \( \xi_c = 1 \), with \( u = 1 - O(N^{-1}) \). From Eqs. (3.63) and (3.62), we obtain

\[
\left[ f_m^{(2)} = A - \mu > f_m^{(1)} = A - \mu - \nu \right]
\]

Therefore, for the sharp peak only a single selected phase is observed. In this case, the function \( g(\phi_c) \) is not exactly a Kronecker delta \( \delta_{\phi_c,1} \), we are in case 3, and thus we find a small correction, approximately linear in \( \nu \), to the saddle-point prediction.

In the selected phase, where the population is exponentially localized near \( u = 1 \) for large \( N \), Eq. (3.48) becomes \( R(l|l_1, l_2) \sim (l_1 + l_2)!2^{-l_1-l_2}/[l!(l_1 + l_2 - l)!] \). By analyzing the differential equation at zeroth-order in \( \nu \) for large \( N \), we find that the class distribution is given by \( P_l^{(0)} = P_0^{(0)}(1 - P_0^{(0)})^l \). Hence, we find that at first order in \( \nu \), the fraction of the population \( P_0 \) located at the peak is given by

\[
P_0 = 1 - \mu/A - \nu/A \left[ 1 - 4 \frac{1 - \frac{\mu}{A}}{(2 - \frac{\mu}{A})^2} \right] + O(\nu^2)
\]

We note that this value of \( f_m = AP_0 \) interpolates between \( f_m^{(1)} \) for \( A/\mu = 1 \) and \( f_m^{(2)} \) for \( A/\mu = \infty \).
Figure 3.5 Convergence of the numerical results towards the theoretical value for two-parent recombination in the parallel (Kimura) model for the sharp peak fitness. In this example, $A/\mu = 4.0$.

As a second example, we consider the quadratic fitness landscape, $f(u) = ku^2/2$. This smooth, continuous fitness function allows for the use of the exact analytical formulas Eq. (3.62), (3.63). By maximizing Eq. (3.62) with respect to $\xi_c$, when $\phi_c < 1$ and hence $g(\phi_c) = 0$, we find

$$f_m^{(1)} = \frac{k}{2} \left[ \left( 1 - \frac{\mu}{k} \right)^2 - \frac{2\nu}{k} \right]$$

(3.69)

This mean fitness defines a selective phase S1.

According to our previous analysis, when $\phi_c = 1$ and $g(\phi_c) = 1$, we maximize Eq. (3.63) in $\xi_c$. Here, we consider that the order parameters $\xi_c$ and $u$ have the same sign, $u\xi_c \geq 0$. We then have $u\xi_c \geq u^2$ in Eq. (3.63) [108]. Hence, we find

$$f_m^{(2)} = \frac{k}{2} \left( 1 - \frac{2\mu}{k} \right)$$

(3.70)
Table 3.6  Stochastic process versus analytical theory for two-parent recombination in the parallel model for the quadratic fitness \( f(u) = ku^2/2 \), with \( k/\mu = 4.0 \), \( \nu/\mu = 3.0 \), and \( N = 100 \).

<table>
<thead>
<tr>
<th>( p_c )</th>
<th>( u^{\text{stochastic}} )</th>
<th>( u^{\text{analytic}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.7065</td>
<td>0.7071</td>
</tr>
<tr>
<td>0.3</td>
<td>0.7052</td>
<td>0.7071</td>
</tr>
<tr>
<td>0.5</td>
<td>0.7058</td>
<td>0.7071</td>
</tr>
</tbody>
</table>

which defines a second selective phase S2.

By applying the self-consistency condition \( f_m^{(1,2)} = ku^2/2 \), we find the following phases

\[
S1: \quad u = \left[ \left( 1 - \frac{\mu}{k} \right)^2 - \frac{2\nu}{k} \right]^{1/2}, \quad \frac{2\nu}{\mu} < \frac{\mu}{k} < 1 - \left[ \frac{2\nu}{k} \right]^{1/2}
\]
\[
S2: \quad u = \sqrt{1 - \frac{2\mu}{k}}, \quad \frac{2\nu}{\mu} > \frac{\mu}{k} < \frac{1}{2}
\]
\[
NS: \quad u = 0, \quad \text{otherwise}
\]  

We note that the phase transition between case 1 and case 2 is exactly as predicted by Eq. (3.65). We further note that the mean fitness is independent of \( \nu \) for \( \nu > \nu^* = \mu^2/(2k) \), exactly as predicted by Eq. (3.66).

The system of differential equations (3.47) provides an exact representation of the evolution dynamics for an infinite population, when uniform crossover probability \( p_c = 1/2 \) is assumed. On the other hand, our analytical equations Eq. (3.62), Eq.
Figure 3.6  Probability distributions for two-parent recombination in the parallel model for the quadratic fitness $f(u) = ku^2/2$, with $k/\mu = 4.0$ and $\nu/\mu = 3.0$, obtained from stochastic simulations with $M = 10000$ sequences of $N = 100$ bases and different values of $p_c$.

(3.63) for smooth fitness, or Eq. (3.68) for the discontinuous sharp peak, predict that the equilibrium results should be independent of the crossover probability $p_c$. To test this theory, we performed exact stochastic simulations based on a Lebowitz/Gillespie algorithm [109, 110]. We generate a population of $M = 10000$ sequences initially in the wild-type. The size of the finite population represented in the simulation was chosen large enough such that the results become independent of size $M$. Then, the population is evolved in time by point mutation, recombination and replication with rates proportional to $\mu$, $\nu$, and $f(u^l)$ respectively, with $u^l = \frac{1}{N} \sum_{j=1}^{N} s_j^l$ the average composition of sequence $S_l$, $1 \leq l \leq M$. For that purpose, a list is generated by defining: $\tau_i = \mu + \nu + f(u^l)$, $\tau = \sum_{i=1}^{M} \tau_i$. With probability $\tau_i/\tau$, a sequence $1 \leq l \leq M$ is chosen from the population to undergo either a single point mutation
with probability $\mu/\tau_1$, replication with probability $f(u^i)/\tau_1$, or recombination with another sequence with probability $\nu/\tau_1$ according to the process described in Fig. 3.4.

![Figure 3.7](image)

**Figure 3.7** Convergence of the numerical results towards the theoretical value for two-parent recombination in the parallel model for the selective phase S1 in Eq. (3.71). In this example, $k/\mu = 4.0$ and $\nu/\mu < 1/8$.

To preserve the size $M$ of the population, when replication or recombination is performed, a sequence chosen at random from the population is substituted with the offspring. The time increment after any of these events is performed is calculated as $dt = -\log(w)/(N\tau)$, with $w \in (0,1]$ a uniformly distributed random number. The results obtained from this stochastic simulation are compared with the theoretical prediction in Table 3.7 for the sharp peak fitness landscape and uniform crossover $p_c = 1/2$.

In agreement with our theoretical prediction, as shown in Table 3.6 from stochastic simulations in the quadratic fitness landscape, the effect of recombination is in-
Figure 3.8 Convergence of the numerical results towards the theoretical value for two-parent recombination in the parallel model for the selective phase S2 in Eq. (3.71). In this example, $k/\mu = 4.0$ and $\nu/\mu > 1/8$.

dependent of the polymerase crossover probability $p_c$. The probability distributions obtained for the systems considered in Table 3.6 are displayed in Fig. 3.6. Clearly, the distributions are independent of $p_c$, in agreement with the theory.

We obtain a direct numerical solution of the deterministic system of differential equations Eq. (3.47), which provides an exact representation of the evolution dynamics for an infinite population experiencing uniform crossover recombination $p_c = 1/2$. A comparison between these numerical solutions, and results obtained from the stochastic simulation for a system large enough to eliminate finite size effects, is displayed in Table 3.7 for the sharp peak fitness. The theoretical prediction from the analytical formula Eq. (3.68) is also shown for comparison. It is evident from this table that the effect of recombination is independent of the polymerase crossover
probability $p_c$, in agreement with our theoretical predictions.
Table 3.7  Stochastic process versus differential equation for two-parent recombination in the parallel model for the sharp peak fitness, $A/\mu = 4.0$, $N = 400$.

<table>
<thead>
<tr>
<th>$\nu/\mu$</th>
<th>$u^{\text{stochastic}}$</th>
<th>$u^{\text{diffeq}}$</th>
<th>$P_0^{\text{stochastic}}, p_c = 0.1$</th>
<th>$P_0^{\text{stochastic}}, p_c = 0.3$</th>
<th>$P_0^{\text{stochastic}}, p_c = 0.5$</th>
<th>$P_0^{\text{diffeq}}$</th>
<th>$P_0^{\text{analytic}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.998337</td>
<td>0.998336</td>
<td>0.75017</td>
<td>0.75017</td>
<td>0.75017</td>
<td>0.75016</td>
<td>0.75</td>
</tr>
<tr>
<td>1.0</td>
<td>0.998329</td>
<td>0.998326</td>
<td>0.7455</td>
<td>0.7454</td>
<td>0.74591</td>
<td>0.74544</td>
<td>0.7449</td>
</tr>
<tr>
<td>2.0</td>
<td>0.998312</td>
<td>0.998317</td>
<td>0.7415</td>
<td>0.7414</td>
<td>0.74085</td>
<td>0.74140</td>
<td>0.7398</td>
</tr>
</tbody>
</table>
From the data presented in Table 3.7, we notice that the deterministic system of differential equations provides an accurate representation of the underlying stochastic dynamics for the case of uniform crossover, \( p_c = 1/2 \). Thus, the results obtained from the numerical solution of the deterministic system of differential equations can be fairly compared with the analytical theory.

It is remarkable that the small, but finite, effect introduced by recombination in the structure of the quasispecies distribution for the sharp peak case, is not a consequence of the Muller's ratchet phenomenon [79] characteristic of finite populations. Indeed, the shift in the wild-type probability \( P_0 \) due to recombination, as predicted from our analytical equation Eq. (3.68), was derived from the system of differential equations Eq. (3.47), which describes the time evolution of an infinite population. Moreover, this closed analytical result is in excellent agreement with the numerical solution of the system of differential equations Eq. (3.47), as displayed in Fig. 3.8 and Table 3.8. A good agreement between our analytical and differential equation results, which correspond to the infinite population case, and the stochastic simulation is expected when the later is performed in a large enough population. We determined that for the parameters we consider, \( M = 10000 \) sequences provides simulation results that are independent of the population size for the sharp peak fitness function, thus allowing for a comparison with the infinite population theory expressed by the differential equations Eq. (3.47) and with our analytical solution Eq. (3.68).
Table 3.8 Analytical theory versus numerical solution for two-parent recombination in the parallel model for the quadratic fitness \( f(u) = ku^2/2 \) with \( N = 800 \) and \( k/\mu=4.0 \).

<table>
<thead>
<tr>
<th>( \nu/\mu )</th>
<th>( u_{\text{diffeq}} )</th>
<th>( u_{\text{analytic}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.7499</td>
<td>0.7500</td>
</tr>
<tr>
<td>0.025</td>
<td>0.7417</td>
<td>0.7416</td>
</tr>
<tr>
<td>0.05</td>
<td>0.7329</td>
<td>0.7331</td>
</tr>
<tr>
<td>0.1</td>
<td>0.7202</td>
<td>0.7159</td>
</tr>
<tr>
<td>0.5</td>
<td>0.7091</td>
<td>0.7071</td>
</tr>
<tr>
<td>1.0</td>
<td>0.7083</td>
<td>0.7071</td>
</tr>
<tr>
<td>2.0</td>
<td>0.7075</td>
<td>0.7071</td>
</tr>
<tr>
<td>3.0</td>
<td>0.7073</td>
<td>0.7071</td>
</tr>
</tbody>
</table>
Table 3.9 Analytical theory versus numerical solution for two-parent recombination in the parallel model for the square-root fitness $f(u) = k\sqrt{|u|}$, with $N = 400, 800, 1000$ and $k/\mu = 4.0$.

<table>
<thead>
<tr>
<th>$\nu/\mu$</th>
<th>$u_{\text{diffeq}}, N = 400$</th>
<th>$u_{\text{diffeq}}, N = 800$</th>
<th>$u_{\text{diffeq}}, N = 1000$</th>
<th>$u_{\text{analytic}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.6527</td>
<td>0.6525</td>
<td>0.65249</td>
<td>0.6523</td>
</tr>
<tr>
<td>0.1</td>
<td>0.6650</td>
<td>0.6672</td>
<td>0.6678</td>
<td>0.6710</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6686</td>
<td>0.6697</td>
<td>0.66993</td>
<td>0.6710</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6696</td>
<td>0.6703</td>
<td>0.67043</td>
<td>0.6710</td>
</tr>
<tr>
<td>0.8</td>
<td>0.6703</td>
<td>0.6707</td>
<td>0.67073</td>
<td>0.6710</td>
</tr>
<tr>
<td>1.0</td>
<td>0.6705</td>
<td>0.6708</td>
<td>0.67083</td>
<td>0.6710</td>
</tr>
</tbody>
</table>

Notice that for the quadratic fitness, the analytical theory reproduces the differential equation results within $O(N^{-1})$. The convergence towards the theoretical value as a function of the system size $1/N$, for parameters within the $S1$ phase defined in Eq. (3.71), is displayed in Fig. 3.7, and for the $S2$ phase in Fig. 3.8.

As a final example, we apply our analytical solution Eq. (3.62) and Eq. (3.63) to study the square-root fitness, $f(u) = k\sqrt{|u|}$, as displayed in Table 3.9, where analytical theory and direct numerical solution of the differential equation agree to $O(N^{-1})$.

As shown in Table 3.9, two-parent recombination in the square-root fitness landscape enhances selection towards sequences which are on average more fit, as observed
by a slight increase of the average composition \( u \), with respect to the case when recombination is absent. This effect, which was already observed for the square-root landscape in the presence of horizontal gene transfer, can be attributed to the negative (see Fig. 3.1) epistatic interactions introduced by the square-root fitness, in agreement with the mutational deterministic hypothesis, Appendix 3.14.

An additional interesting effect in two-parent recombination, which was observed in the quadratic as well as in the square-root fitness landscapes, is the presence of spectral rigidity: the effect of recombination becomes independent of the recombination rate for \( \nu > 0 \).

In summary, from our generalization of the parallel or Crow-Kimura model for an infinite population of evolving sequences Eq. (3.36), we conclude that two-parent recombination introduces a mild mutational load over discontinuous fitness functions, such as a single sharp peak, and thus it can shift the error-threshold transition. For smooth fitness functions, the effect of recombination depends on the sign of epistasis (see Fig. 3.1), in agreement with the mutational deterministic hypothesis [41, 43, 44, 42]. We show this analytically in Appendix 3.14.

In contrast with horizontal gene transfer, recombination affects the structure of the quasispecies (nor the error threshold transition) for a sharp peak fitness. We believe that this fundamental difference between horizontal gene transfer and recombination is because of the fact that the later can generate a much larger diversity in the
offspring per recombination event. Hence, the diversity barrier that, as previously discussed in section II, is imposed by the sharp exponential distribution in the sharp peak case can be tunneled through due to the more radical mixing effects of two-parent recombination. Our analytical theory, which provides explicit expressions for the mean fitness $f_m$ and average composition $u$, is developed in the realistic regime ($N \rightarrow \infty$), considering that typical viral genomes are $N \sim 10^3 - 10^4$.

3.3 The Eigen model

In this section, we present a generalization of the classical Eigen model [11, 93, 12], including the exchange of genetic material between pairs of individuals in an infinite population [40],

$$\frac{dq_k}{dt} = \sum_{j,k=1}^{2N} [B_{ij}C_{jk}r_k - \delta_{ij}\delta_{ik}D_i]q_k$$

(3.72)

Here, recombination as well as mutation are considered to be coupled to the replication process. Recombination is represented by the coefficients $C_{jk}$, which in general will be functions of the frequencies $q_k$, $C_{jk} \sim \delta_{jk} + \sum_l q_l C_{kl}^j/\sum_k q_k$.

3.3.1 Horizontal gene transfer of non-overlapping blocks

In this recombination scheme, we consider the exchange of blocks of genetic material between pairs of individuals in the population. We consider the blocks to be non-overlapping, such that we have $N/\bar{M}$ of them. We define a block index $0 \leq b \leq N/\bar{M} - 1$, and a site index within each block to be $\bar{M}b + 1 \leq j_b \leq \bar{M}(b + 1)$. 
For this process, we have that the nonlinear recombination term in the differential Eq. (3.72) is

$$C_{jk} \sim \left(1 - \frac{\nu}{\tilde{M}}\right) \delta_{j,k} + \frac{\nu}{\tilde{M}} \frac{N}{M} \prod_{b=0}^{N/\tilde{M}-1} \prod_{j_b=Mb+1}^{\tilde{M}(b+1)} \left(\delta_{j_b+1, j_b}^{j_b} \cdot \frac{1 + u(j_b)}{2} + \delta_{j_b, j_b}^{j_b} \cdot \frac{1 - u(j_b)}{2}\right) \prod_{m \neq (j_b)} \delta_{s_m, s_m^b}$$

(3.73)

The recombination operator representing this process, assuming the recombination rate per block to be $\nu/\tilde{M}$, becomes

$$\hat{R} = \prod_{b=0}^{N/\tilde{M}-1} \left(1 - \frac{\nu}{\tilde{M}}\right) \frac{M}{N} \prod_{j_b=Mb+1}^{M(b+1)} \hat{I}_{j_b} + \frac{\nu}{\tilde{M}} \frac{M}{N} \prod_{j_b=Mb+1}^{M(b+1)} \hat{R}_{j_b}$$

(3.74)

Here, we defined the single-site recombination operator as $\hat{R}_j = \hat{a}^\dagger(j)D\hat{a}(j)$, with the matrix $D$ defined in Eq. (3.6). We consider the large $N$ limit, while keeping $N/\tilde{M} \simeq \mathcal{O}(N)$. Then, the recombination operator defined in Eq. (3.74) becomes, to order $\mathcal{O}(N^{-1})$

$$\hat{R} = e^{-\frac{\nu}{\tilde{M}} e^{\frac{\nu}{\tilde{M}} \sum_{b=0}^{N/\tilde{M}-1} \prod_{j_b=Mb+1}^{M(b+1)} \hat{a}^\dagger(j)D\hat{a}(j)}}$$

(3.75)

The Hamiltonian

The Hamiltonian operator for the Eigen model, including the horizontal gene transfer process described by the operator Eq. (3.74) is given by

$$-\hat{H} = Ne^{-\mu + \frac{\nu}{\tilde{M}} \sum_{j=1}^N \hat{a}^\dagger(j)\sigma_1\hat{a}(j)} e^{-\frac{\nu}{\tilde{M}} \sum_{b=0}^{N/\tilde{M}-1} \prod_{j_b=Mb+1}^{M(b+1)} \hat{a}^\dagger(j)D\hat{a}(j)}$$

$$\times f \left[\frac{1}{N} \sum_{j=1}^N \hat{a}^\dagger(j)\sigma_3\hat{a}(j)\right] - Nd \left[\frac{1}{N} \sum_{j=1}^N \hat{a}^\dagger(j)\sigma_3\hat{a}(j)\right]$$

(3.76)
The microscopic fitness function is \( f(u) \) and degradation function is \( d(u) \). Here, the matrix \( D \) is defined as in Eq. (3.6). We introduce a Trotter factorization of the evolution operator, in the basis of coherent states

\[
e^{-Ht} = \lim_{M \to \infty} \int \left[ \prod_{k=1}^{M} Dz_k^* Dz_k \right] |\tilde{z}_M\rangle \left( \prod_{k=1}^{M} \langle \tilde{z}_k | e^{-cH} | \tilde{z}_{k-1} \rangle \right) \langle \tilde{z}_0 | (3.77)
\]

As shown in Appendix 3.8, the partition function is

\[
Z = \int [D\xi D\eta D\bar{\eta} D\phi D\bar{\phi}] e^{-S[\xi, \eta, \bar{\eta}, \phi, \bar{\phi}]} (3.78)
\]

Here, the action is defined by

\[
S[\xi, \eta, \bar{\eta}, \phi, \bar{\phi}] = -N \int_0^t dt \left[ -\dot{\xi}^2 - \eta \phi + e^{-\mu(1-\eta)} - \frac{M}{\bar{\eta}} + \frac{\phi^2}{2} \right] f(\xi) - d(\xi) - N \ln Q (3.79)
\]

### The saddle point limit

We consider the saddle point limit of the action defined by Eq. (3.79). In the saddle point limit, for long times, the trace defined by Eq. (3.191) becomes

\[
\lim_{t \to \infty} \frac{\ln Q_c}{t} = \frac{\bar{\phi}_c}{2} + \left[ \xi_c(\bar{\xi}_c + u\bar{\phi}_c) + (\bar{\eta}_c + \bar{\phi}_c/2)^2 \right]^{1/2} (3.80)
\]

In this saddle-point limit, the action is given by

\[
\lim_{N, t \to \infty} \frac{\ln Z}{Nt} = \lim_{t \to \infty} \frac{-S_c}{Nt} = \max_{\xi_c, \bar{\xi}_c, \phi_c, \bar{\phi}_c, \eta_c, \bar{\eta}_c} \left\{ f(\xi_c) e^{-\mu(1-\eta_c)} - \frac{M}{\bar{\eta}_c} + \frac{\phi_c^2}{2} - d(\xi_c) - \xi_c \bar{\xi}_c - \bar{\eta}_c \eta_c - \bar{\phi}_c \phi_c + \frac{\bar{\phi}_c}{2} + \left[ \xi_c(\bar{\xi}_c + u\bar{\phi}_c) + (\bar{\eta}_c + \bar{\phi}_c/2)^2 \right]^{1/2} \right\} (3.81)
\]
As shown in Appendix 3.9 the mean fitness, defined from the saddle point action

\[ f_m = \lim_{N,t \to \infty} \ln \frac{Z}{Nt} = -S_c/N_t, \]

is

\[ f_m = \max_{-1 \leq \xi_c \leq 1} \left\{ e^{-\mu(1-\eta_c(\xi_c))-
\frac{\nu}{2} \{1-\phi_c(\xi_c)\}^M} f(\xi_c) - d(\xi_c) \right\} \quad (3.82) \]

Here, the expressions \( \phi_c(\xi_c) \) and \( \eta_c(\xi_c) \) are given by

\[ \phi_c(\xi_c) = \frac{1 + u \xi_c}{2} + \frac{\sqrt{1 - \xi_c^2}}{2} \frac{\mu + \frac{\nu}{2}(1 - u^2)\phi_c^{M-1}}{\left( (\mu + \frac{\nu}{2}\phi_c^{M-1})^2 - \frac{\nu^2 u^2}{4}[\phi_c^{M-1}]^2 \right)^{1/2}} \quad (3.83) \]

\[ \eta_c(\xi_c) = \sqrt{1 - \xi_c^2} \frac{\mu + \frac{\nu}{2}\phi_c^{M-1}}{\left( (\mu + \frac{\nu}{2}\phi_c^{M-1})^2 - \frac{\nu^2 u^2}{4}[\phi_c^{M-1}]^2 \right)^{1/2}} \quad (3.84) \]

The average composition, \( u \), is obtained from the self-consistency condition

\[ f_m = f(u) - d(u). \]

Eq. (3.82) is an exact analytical expression for the equilibrium mean fitness of an infinite population of evolving sequences. This analytical expression is valid for arbitrary permutation invariant replication rate \( f(u) \) and degradation rate \( d(u) \).

**Examples**

We consider first the quadratic fitness case, \( f(u) = ku^2/2 + k_0 \). By expanding the formulas Eqs. (3.82), (3.83) and (3.84) near the error threshold \( \xi_c \sim 0, u \sim 0 \), we obtain the phase boundary from the critical condition

\[ k_{\text{crit}} = \mu k_0 \frac{1 + \nu/\mu}{1 + \nu/2\mu} \quad (3.85) \]
We notice that the phase boundary is qualitatively similar to the horizontal gene transfer process analyzed in section I.A, Eq. (3.12) for the parallel model. As in this former case, we notice that horizontal gene transfer introduces a mild mutational load against selection for a smooth fitness (i.e. quadratic).

As a second example, we consider the square-root fitness landscape \( f(u) = k\sqrt{u} + 1 \). In Table 3.10, we evaluate our analytical Eqs. (3.82-3.84) for this particular case.

From the results displayed in Table 3.10, we notice that horizontal gene transfer increases the average composition \( u \) and therefore the mean fitness of the population. This effect, which is attributed to the negative epistasis introduced by the square-root fitness (see Fig. 3.1), is in agreement with the previous examples studied in the case of the parallel model, and with the mutational deterministic hypothesis [45, 46, 85, 44], as we prove in Appendix 3.13.

As a third example, we consider the sharp peak fitness \( f(u) = (A - A_0)\delta_{u,1} + A_0 \). In this case, the maximum in Eq. (3.82) corresponds to \( \xi_c = 1 \). From Eqs. (3.83) and (3.84), we have \( \xi_c = (1 + u)/2, \eta_c = 0 \), and hence after Eq. (3.82)

\[
\hat{f}_m = A e^{-\mu - \gamma \mu [1 - (1+\mu)^{\mu}]} 
\]

The error threshold is given, for \( u = 0 \) in Eq. (3.86), by the condition \( A e^{-\mu - \gamma [1 - 1/2^{\mu}]} > A_0 \). However, we notice that \( f_m(u = 1) = A e^{-\mu} > f_m(u = 0) \). Hence, in the selected phase we have \( u = 1 - \mathcal{O}(N^{-1}) \). The fraction of the population located at the peak
Table 3.10  Analytical results for horizontal gene transfer in the Eigen model for the square-root fitness \( f(u) = k\sqrt{|u|} + 1 \), with \( \bar{M} = 3 \).

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \nu )</th>
<th>( u^{\text{analytic}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>0.0</td>
<td>0.3346</td>
</tr>
<tr>
<td>3.0</td>
<td>0.5</td>
<td>0.3398</td>
</tr>
<tr>
<td>3.0</td>
<td>0.8</td>
<td>0.3422</td>
</tr>
<tr>
<td>3.0</td>
<td>1.5</td>
<td>0.3466</td>
</tr>
<tr>
<td>5.0</td>
<td>0.0</td>
<td>0.3588</td>
</tr>
<tr>
<td>5.0</td>
<td>0.5</td>
<td>0.3642</td>
</tr>
<tr>
<td>5.0</td>
<td>0.8</td>
<td>0.3667</td>
</tr>
<tr>
<td>5.0</td>
<td>1.5</td>
<td>0.3713</td>
</tr>
<tr>
<td>8.0</td>
<td>0.0</td>
<td>0.3741</td>
</tr>
<tr>
<td>8.0</td>
<td>0.5</td>
<td>0.3796</td>
</tr>
<tr>
<td>8.0</td>
<td>0.8</td>
<td>0.3822</td>
</tr>
<tr>
<td>8.0</td>
<td>1.5</td>
<td>0.3869</td>
</tr>
</tbody>
</table>
$P_0$ is obtained from the self-consistency condition $f_m = AP_0 + A_0(1 - P_0)$

$$P_0 = \frac{Ae^{-\mu} - A_0}{A - A_0}$$

(3.87)

After Eq. (3.87), we find the true error threshold at $A_{\text{crit}} = A_0e^\mu$, while the condition $Ae^{-\mu - \frac{\mu}{N}[1 - 2 - \frac{\mu}{N}]} > A_0$ represents the limit of metastability for initial conditions with $u \sim 0$. We notice that this result is similar to the exact solution in the absence of horizontal gene transfer [29]. Hence, as previously discussed in section I.A. for the parallel model, we conclude that horizontal gene transfer does not affect the structure of the quasispecies for a discontinuous, single sharp peak fitness.

### 3.3.2 Horizontal gene transfer for multiple-size blocks

In analogy with the model treated in Section II.B, we consider the natural extension of horizontal gene transfer of blocks with multiple size, with average $\langle \tilde{M} \rangle$ and $\langle \tilde{M} \rangle/N = \mathcal{O}(N^{-1})$. Following a similar analysis as in the derivation of Eq. (3.25), we define the recombination operator for multiple-size blocks as

$$\hat{R} \sim e^{-\langle \tilde{M} \rangle} + e^{-\frac{\mu}{N}} \sum_{j=1}^{N}\tilde{s}(j)D\tilde{s}(j)$$

(3.88)

**The Hamiltonian**

We consider horizontal gene transfer to be coupled to the replication process. Moreover, we will consider that when replication occurs, a horizontal gene transfer event also occurs with a probability $0 \leq \nu/\langle \tilde{M} \rangle \leq 1$. The Hamiltonian operator
for the Eigen model, including the horizontal gene transfer process described by the operator Eq. (3.88) is given by

$$-\hat{H} = Ne^{-\mu+\frac{\mu}{N} \sum_{j=1}^N \tilde{a}(j)D\tilde{a}(j)} \left( 1 - \frac{\nu}{\langle M \rangle} + \frac{\nu}{\langle M \rangle} e^{-\langle M \rangle \sum_{j=1}^N \tilde{a}(j)D\tilde{a}(j)} \right) \times f \left[ \frac{1}{N} \sum_{j=1}^N \tilde{a}(j) \sigma_3 \tilde{a}(j) \right] - Nd \left[ \frac{1}{N} \sum_{j=1}^N \tilde{a}(j) \sigma_3 \tilde{a}(j) \right]$$

(3.89)

We introduce a Trotter factorization

$$e^{-\hat{H}t} = \lim_{M\to\infty} \int [D\tilde{z}^* D\tilde{z}] \left| \tilde{z}_M \right> \prod_{j=1}^M \left( \tilde{z}_k | e^{-t\hat{H}} | \tilde{z}_{k-1} \right) \langle \tilde{z}_0 |$$

(3.90)

As shown in Appendix 3.10, the partition function is

$$Z = \int [D\tilde{z} D\xi D\eta D\phi D\phi] e^{-s[\xi,\eta,\eta,\eta,\phi,\phi]}$$

(3.91)

Here, the action in the continuous time limit is

$$S[\xi,\xi,\eta,\eta,\phi,\phi] = -N \int_0^t dt' \left\{ -\tilde{\xi} \xi - \tilde{\eta} \eta - \tilde{\phi} \phi + e^{-\mu(1-\eta)} \left[ 1 - \frac{\nu}{\langle M \rangle} + \frac{\nu}{\langle M \rangle} e^{-\langle M \rangle (1-\phi)} \right] f(\xi) - d(\xi) \right\} - N \ln Q$$

(3.92)

The saddle point limit

The saddle point limit is exact as $N \to \infty$ in Eq. (3.92). After a similar procedure as in Section 3.A.2, we find the saddle point equation for the mean fitness

$$f_m = \max_{-1 \leq \xi_c \leq 1} \left\{ e^{-\mu(1-\nu)} \left[ 1 - \frac{\nu}{\langle M \rangle} + \frac{\nu}{\langle M \rangle} e^{-\langle M \rangle (1-\phi_c)} \right] f(\xi_c) - d(\xi_c) \right\}$$

(3.93)
Here, the fields $\eta_c$ and $\phi_c$ are expressed as functions of $\xi_c$

$$\eta_c(\xi_c) = \sqrt{1 - \xi_c^2} \frac{\frac{\nu}{\langle M \rangle} + \left[1 - \frac{\nu}{\langle M \rangle}\right] e^{(\langle M \rangle)(1-\phi_c)} + \frac{\nu}{2\mu}}{\left[\left(\frac{\nu}{\langle M \rangle} + \left[1 - \frac{\nu}{\langle M \rangle}\right] e^{(\langle M \rangle)(1-\phi_c)} + \frac{\nu}{2\mu}\right)^2 - \frac{\nu^2\mu^2}{4\mu^2}\right]^{1/2}} (3.94)$$

$$\phi_c(\xi_c) = \frac{1 + u \xi_c}{2} + \frac{\sqrt{1 - \xi_c^2}}{2} \frac{\frac{\nu}{\langle M \rangle} + \left[1 - \frac{\nu}{\langle M \rangle}\right] e^{(\langle M \rangle)(1-\phi_c)} + \frac{\nu(1-u^2)}{2\mu}}{\left[\left(\frac{\nu}{\langle M \rangle} + \left[1 - \frac{\nu}{\langle M \rangle}\right] e^{(\langle M \rangle)(1-\phi_c)} + \frac{\nu}{2\mu}\right)^2 - \frac{\nu^2\mu^2}{4\mu^2}\right]^{1/2}} (3.95)$$

Equations (3.93)–(3.95) represent an exact analytical solution for the equilibrium mean fitness of an infinite population experiencing horizontal gene transfer of variable blocks size. This expression is valid for arbitrary, permutation invariant replication rate $f(u)$ and degradation rate $d(u)$.

**Examples**

We consider first the sharp peak fitness $f(u) = (A - A_0)\delta_{u,1} + A_0$. In this case, the maximum in Eq. (3.93) is at $\xi_c = 1$. From Eqs. (3.94) and (3.95), we obtain $\eta_c = 0$ and $\phi_c = (1 + u)/2$. Substituting these values in Eq. (3.93), we obtain for the mean fitness

$$f_m = e^{-\mu} \left[1 - \frac{\nu}{\langle M \rangle} + \frac{\nu}{\langle M \rangle} e^{-(\langle M \rangle)(1-u)/2}\right] A (3.96)$$

The error threshold for $u = 0$ is obtained from Eq. (3.96) by the condition

$$Ae^{-\mu} \left[1 - \frac{\nu}{\langle M \rangle} + \frac{\nu}{\langle M \rangle} e^{-(\langle M \rangle)/2}\right] > A_0 (3.97)$$
However, we notice that $f_m(u = 1) = Ae^{-\mu} > f_m(u = 0)$. Therefore, in the selected phase the average composition $u = 1 - O(N^{-1})$, and the effect of recombination becomes negligible for the sharp peak fitness. The fraction of the population located at the peak $P_0$ is obtained from the self-consistency condition $f_m = AP_0 + A_0(1 - P_0)$

$$P_0 = \frac{Ae^{-\mu} - A_0}{A - A_0} \tag{3.98}$$

From this expression, we find that the true error threshold for the sharp peak fitness is $A_{\text{crit}} = e^\mu A_0$, with the condition $A e^{-\mu} \left[1 - \frac{\nu}{(M)} + \frac{\nu}{(M)} e^{-\langle M \rangle/2} \right] > A_0$ representing the limit for metastability for initial conditions with $u \sim 0$.

As a second example, we consider the quadratic fitness $f(u) = ku^2/2 + k_0$. An analytical expression for the phase boundary is obtained from Eqs. (3.93), (3.94) and (3.95) near the error threshold $\xi_c \sim 0, u \sim 0$. We find

$$k_{\text{crit}} = \mu k_0 \frac{1 + \frac{\nu}{\mu}}{1 + \frac{\nu}{2\mu}} \tag{3.99}$$

For small $\nu$, the critical value is $k_{\text{crit}} \sim k_0(\mu + \nu/2)$.

As a final example, we consider the square-root fitness $f(u) = k\sqrt{|u|} + 1$. Analytical results, as obtained from Eqs. (3.93)-(3.95) for this case, are presented in Table 3.11.

We notice that the results obtained for the horizontal gene transfer process with variable block size agree with the corresponding ones when the size of the recombination blocks is fixed. We recall that this correspondence was also observed and
Table 3.11  Analytical results for horizontal gene transfer in the Eigen model for the square-root fitness $f(u) = k\sqrt{|u|} + 1$, with $\langle M \rangle = 3$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\nu$</th>
<th>$u^{\text{analytic}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>0.0</td>
<td>0.3346</td>
</tr>
<tr>
<td>3.0</td>
<td>0.5</td>
<td>0.3409</td>
</tr>
<tr>
<td>3.0</td>
<td>0.8</td>
<td>0.3450</td>
</tr>
<tr>
<td>3.0</td>
<td>1.5</td>
<td>0.3546</td>
</tr>
<tr>
<td>5.0</td>
<td>0.0</td>
<td>0.3588</td>
</tr>
<tr>
<td>5.0</td>
<td>0.5</td>
<td>0.3654</td>
</tr>
<tr>
<td>5.0</td>
<td>0.8</td>
<td>0.3695</td>
</tr>
<tr>
<td>5.0</td>
<td>1.5</td>
<td>0.3794</td>
</tr>
<tr>
<td>8.0</td>
<td>0.0</td>
<td>0.3741</td>
</tr>
<tr>
<td>8.0</td>
<td>0.5</td>
<td>0.3809</td>
</tr>
<tr>
<td>8.0</td>
<td>0.8</td>
<td>0.3851</td>
</tr>
<tr>
<td>8.0</td>
<td>1.5</td>
<td>0.3950</td>
</tr>
</tbody>
</table>
discussed in the previous section for the parallel model, so similar arguments apply to the Eigen model as well. An analytical proof is provided in Appendix 3.13.

3.3.3 The Eigen model with two-parent recombination

For the Eigen model, we introduce the recombination process described in Section II.C and illustrated in Fig. 3.4, which considers the exchange of genetic material between pairs of sequences due to crossovers governed by the polymerase switching from one parental chromosome to the other with probability \( p_c \) per site. For the Eigen model, mutation and recombination are considered to be coupled to the recombination process, as stated in the generic differential equation Eq. (3.72). We will consider that during replication, a sequence can recombine with probability \( \nu \leq 1 \), or just replicate without recombining with probability \( 1 - \nu \). This process is represented by the coefficients in Eq. (3.72)

\[
C_{jk} = (1 - \nu)\delta_{j,k} + \frac{\nu}{2} \sum_{\{n=-1\}} \left\{ \prod_{n=2}^{N} \left[ p_c^{\frac{1-\alpha_n}{2} - 1} (1 - p_c) \frac{1+\alpha_n}{2} \right] \right\} 
\times \sum_{l=1}^{2^N} p_l \prod_{n=1}^{N} \left( \frac{1 + s_n^l s_n^j}{2} \right)^{\frac{1+\alpha_n}{2}} \left( \frac{1 + s_n^l s_n^{j+1}}{2} \delta_{s_n^{j+1},1} + \frac{1 - s_n^l s_n^{j-1}}{2} \delta_{s_n^{j-1},-1} \right)^{\frac{1-\alpha_n}{2}}
\]

(3.100)

Here, again, \( p_l = q_l / \sum_{l=1}^{2^N} q_l \) is the normalized probability for the sequence \( 1 \leq l \leq 2^N \).

In the spin Boson representation, we express the Eigen model Hamiltonian by the
operator

\[ -\hat{H} = Ne^{-\mu + \frac{1}{N} \sum_{j=1}^{N} \tilde{a}^{\dagger}(j) \sigma_{3}(j)} \left( 1 - \nu \right) \hat{I} + \nu g\{\{\tilde{a}^{\dagger}(j) D_{j}^{\dagger} \tilde{a}(j)\}\} \]

\times \int \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{a}^{\dagger}(j) \sigma_{3} \tilde{a}(j) \right] - Nd \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{a}^{\dagger}(j) \sigma_{3} \tilde{a}(j) \right] \]

(3.101)

Here, \( g\{\{\tilde{K}_{j}\}\} \) was defined in Eq. (3.37), and the matrices \( D_{j}^{\dagger} \) were defined in Eq. (3.38). We introduce a Trotter factorization

\[ e^{-\hat{H}t} = \lim_{M \to \infty} \int [D\bar{z}^{\dagger}D\bar{z}] \prod_{k=1}^{M} \langle \bar{z}_{k} | e^{-\epsilon^{\dagger}} | \bar{z}_{k-1} \rangle \langle \bar{z}_{0} | \]

(3.102)

As shown in Appendix 3.11, the partition function is

\[ Z = \int D\xi D\eta D\phi D\phi = -S[\xi, \eta, \phi] \]

(3.103)

Here, the action is defined by

\[ S[\xi, \eta, \phi] = -N \int_{0}^{t} dt' \left[ -\dot{\xi} \xi - \dot{\eta} \eta - \dot{\phi} \phi \right. \]

\[ + e^{-\mu(1-\eta)}(1 - \nu + \nu g(\phi)) f(\xi) - d(\xi)) - N \ln Q \]

(3.104)

The saddle point limit

For long times, a steady state condition is achieved. Then, the fields become time-independent, and we have

\[ \lim_{t \to \infty} \frac{\ln Q_{c}}{t} = \frac{\phi_{c}}{2} + \left[ \xi_{c}(\xi_{c} + u \phi_{c}) + \left( \eta_{c} + \frac{\phi_{c}}{2} \right)^{2} \left( \xi_{c} + \frac{\phi_{c}}{2} \right) \right]^{1/2} \]

(3.105)
We look for the saddle point solution from the action

$$\lim_{N,t \to \infty} \frac{\ln Z}{Nt} = \lim_{t \to \infty} \frac{-S_c}{Nt} = \max_{\xi_c, \xi_c, \eta_c, \eta_c} \left\{ -\xi_c \xi_c - \eta_c \eta_c - \phi_c \phi_c + e^{-\mu(1-\eta_c)}(1 - \nu + \nu g(\phi_c)) f(\xi_c) - d(\xi_c) + \frac{\phi_c}{2} + \left[ \xi_c (\xi_c + u \phi_c) + \left( \eta_c + \frac{\phi_c}{2} \right) \right]^{1/2} \right\}$$

(3.106)

Because of the singular behavior of the function $g(\phi_c)$, to find the saddle point we need to consider three separate cases: $\phi_c < 1$, $\phi_c = 1$, and $\phi_c = 1 - O(1/N)$. We notice that the saddle point analysis may not apply exactly, unless $g(\phi_c) = \delta_{\phi_c, 1}$.

Case 1: $\phi_c < 1$. The mean fitness is given by

$$f_m^{(1)} = \max_{-1 \leq \xi_c \leq 1} \{(1 - \nu) e^{-\mu \left[ 1 - \sqrt{1 - \xi_c^2} \right]} f(\xi_c) - d(\xi_c) \}$$

(3.107)

We note $\phi_c$ is still given by Eq. (3.61).

Case 2: $\phi_c = 1$. The mean fitness is given by

$$f_m^{(2)} = \max_{-1 \leq \xi_c \leq 1} \left\{ e^{-\mu \left[ \frac{1 - (1 - u) \xi_c - u \xi_c^2}{1 - u^2} \right]} f(\xi_c) - d(\xi_c) \right\}$$

(3.108)

Case 3: $\phi_c = 1 - O(1/N)$. In this case, additional analysis is necessary to calculate the mean fitness due to the singular behavior of the $g(\phi_c)$ function. For a smooth fitness function, we can argue this case does not exist. We first consider the Hamiltonian (3.101) for the case $g = 0$. In this case, the fitness function is simply multiplied by $(1 - \nu)$. If the degradation function is zero, the largest eigenvalue, $f_m$ is simply multiplied by $(1 - \nu)$ relative to the $\nu = 0$ case. Without degradation, this result allows us to calculate the average composition, $u_*$, from the implicit relation
\[ f_m(\nu) = (1 - \nu)f_m(\nu = 0) = f(u_\star). \]

With a non-zero degradation function, the equation for \( f_m(\nu) \) will be a bit more involved. Alternatively, if we consider the differential equation for the unnormalized class probabilities, \( dQ/dt = LQ \), we see that the differential operator \( L \) looks like that in the absence of recombination, save for a multiplication of \((1 - \nu)\) in the fitness function. Thus, the variance of the population is given by \[ \sigma_u^2 / N = 2\mu u_\star (1 - \nu) f(u_\star) / \left[ N((1 - \nu)f'(u_\star) - d'(u_\star)) \right]. \]

Considering more carefully the \( g \) function, we find as before this term is exponentially negligible compared to the \(-\nu P(u)\) term when \( \sigma^2 < \sigma_u^2 \). In other words, we must strictly be in case 1 when

\[ 1 - u_\star^2 < 2\mu u_\star (1 - \nu) f(u_\star) / \left[ (1 - \nu)f'(u_\star) - d'(u_\star) \right] \quad (3.109) \]

We denote the value of \( \nu \) at which

\[ 1 - u_\star^2 = 2\mu u_\star (1 - \nu) f(u_\star) / \left[ (1 - \nu)f'(u_\star) - d'(u_\star) \right] \] at \( \nu = \nu_\star \quad (3.110) \]

as \( \nu_\star \). Now, at this value of \( \nu_\star \) we have \( \int du_1 du_2 R_{u_1 u_2}^u P(u_1) P(u_2) = P(u) \). Thus, the term proportional to \( \nu \) in Hamiltonian (3.101) exactly vanishes. Thus, we have \( df_m/d\nu = 0 \) and \( dP(u)/d\nu = 0 \) at this value of \( \nu \). There is spectral rigidity. This result implies that for \( \nu > \nu_\star \), the distribution \( P(u) \) is independent of \( \nu \), and that the value of \( u_\star \) is constant. In other words, the value of \( f_m \) in case 2 must be constant with \( \nu \). Assuming \( f_m \) varies continuously with \( \nu \) in case 1, and that the fitness values for case 1 and case 2 are equal at a single value of \( \nu \), which mathematically may be
negative, case 2 is simply case 1 with the value $\nu = \nu_*$.

$$f_m(\nu > \nu_*) = f_m(\nu = \nu_*)$$  \hspace{1cm} (3.111)

Equations (3.107), (3.108) constitute an exact analytical expression for the equilibrium mean fitness of an infinite population of sequences evolving under the dynamics of the Eigen model, and experiencing two-parent recombination. These equations are exact for a smooth, permutation invariant replication rate $f(u)$ and degradation rate $d(u)$.

For a non-smooth fitness function, additional analysis is necessary, since $f'(u_*) - d'(u_*)$ is undefined, and $P(u)$ may no longer be Gaussian.

Examples

We investigate the phase diagrams, as predicted from our theoretical equations, for two different fitness functions: A sharp peak and a quadratic fitness landscape.

As an example, we consider the sharp peak fitness, $f(u) = (A - A_0)\delta_{u,1} + A_0$. The maximum is obtained at $\xi_c = 1$, $u = 1 - O(N^{-1})$. From Eqs. (3.108) and (3.107) we have

$$f_m^{(2)} = Ae^{-\mu} > f_m^{(1)} = (1 - \nu)Ae^{-\mu}$$  \hspace{1cm} (3.112)

Hence, for the sharp peak fitness a single selective phase is observed. In this case, the function $g(\phi_c)$ is not exactly a Kronecker delta $\delta_{\phi_c,1}$, we are in case 3, and then
we expect to observe a small correction, approximately linear in \( \nu \) from the prediction of the saddle point analysis. By considering the differential equations for the sharp peak case at zeroth-order in \( \nu \), we find that the class distributions satisfy

\[
e^{-\nu/2} \sum_k (r_k/N) P_k^{(0)}/2^k = f_m^{(0)} \sum_l P_l^{(0)}/2^l \quad \text{with} \quad P_0^{(0)} = (Ae^{-\mu} - A_0)/(A - A_0)
\]

and

\[
f_m^{(0)} = AP_0^{(0)} + A_0(1 - P_0^{(0)}) = Ae^{-\mu}.
\]

Thus we find \( S = \sum_l P_l^{(0)}/2^l = (A - A_0)P_0^{(0)}e^{-\nu/2}/(f_m^{(0)} - A_0e^{-\mu/2}) = (Ae^{-\mu} - A_0)e^{-\mu/2}/(Ae^{-\mu} - A_0e^{-\mu/2}) \). Thus, we find the recombination term

\[
\sum_k (r_k/N) P_k^{(0)}/2^k \sum_l P_l^{(0)}/2^l = Ae^{-\mu/2}S^2.
\]

Hence, we find that at first order in \( \nu \), the fraction of the population located at the peak is given by

\[
P_0 = \frac{Ae^{-\mu} - A_0}{A - A_0} - \nu e^{-\mu} \left[ \frac{A}{A - A_0} - Ae^{-\mu/2} \frac{Ae^{-\mu} - A_0}{(Ae^{-\mu/2} - A_0)^2} \right] + O(\nu^2) \quad (3.113)
\]

We note that this value of \( f_m = AP_0 + A_0(1 - P_0) \) interpolates between \( f_m^{(1)} \) for \( Ae^{-\mu}/A_0 = 1 \) and a value intermediate to \( f_m^{(1)} \) and \( f_m^{(2)} \) for \( Ae^{-\mu}/A_0 = \infty \).

As a second example, we consider the quadratic fitness \( f(u) = ku^2/2 + k_0 \). By maximizing expressions Eq. (3.108) [111] and Eq. (3.107), we obtain two selective phases \( S_1 \) and \( S_2 \), and a non-selective phase \( NS \), defined by the equations

\[
S1: \quad u = \left[ 2(1 - \nu)e^{-\mu} \left[ 1 - \sqrt{1 - \xi_c^2} \right] (\xi_c^2/2 + k_0/k) - 2k_0/k \right]^{1/2}, \quad \nu < \min(\nu_*, \nu_c)
\]

\[
S2: \quad u = \left[ \frac{1 - 2\mu k_0/k}{1 + \mu} \right]^{1/2}, \quad \nu_c > \nu_0 < \nu
\]

\[
NS: \quad u = 0, \quad \text{otherwise} \quad (3.114)
\]

where in the \( S1 \) phase

\[
\xi_c^2 = 2\left[ \sqrt{1 + \mu^2(1 + 2k_0/k)} - 1 - \mu^2 k_0/k \right]/\mu^2 \quad (3.115)
\]
and we have defined

\[ \nu_c = 1 - \frac{k_0}{k} e^{\mu [1 - \sqrt{1 - \xi_c^2}]/(\xi_c^2/2 + k_0/k)} \]

\[ \nu_* = 1 - \frac{k + 2k_0}{2k(1 + \mu)} e^{\mu [1 - \sqrt{1 - \xi_c^2}]/(\xi_c^2/2 + k_0/k)} \]

\( (3.116) \)

where \( \xi_c^2 \) is given by Eq. (3.115). The phase structure is defined by the conditions:

For \( 2\mu k_0/k \geq 1 \), the system is in SI if \( \nu < \nu_c \), or in NS if \( \nu \geq \nu_c \); for \( 2\mu k_0/k < 1 \), the system is in SI if \( \nu \leq \nu_* \), or in S2 if \( \nu > \nu_* \). From Eq. (3.116), we notice that at \( 2\mu k_0/k = 1 \), \( \nu_c = \nu_* \).

We note that the phase transition between case 1 and case 2 is exactly as predicted by Eq. (3.110). We further note that the mean fitness is independent of \( \nu \) for \( \nu > \nu_* \), exactly as predicted by Eq. (3.111).

As a final example, we consider the square-root fitness \( f(u) = k \sqrt{|u|} + 1 \). By maximizing expressions Eq. (3.108) [111] and Eq. (3.107) for the square-root fitness landscape, we obtain the results presented in Table 3.12. From the results displayed in Table 3.12, we observe a similar qualitative behavior as in the two-parent recombination for the parallel case, Table 3.9. In the square-root fitness, recombination introduces a favorable effect over selection, which can be attributed to negative epistasis (see Fig. 3.1) according to the mutational deterministic hypothesis [45, 46, 85, 44], as shown in Appendix 3.15. Spectral rigidity is also observed in this case when \( \nu > 0 \).
Table 3.12  Analytical results for two-parent recombination in the Eigen model for the square-root fitness \( f(u) = k\sqrt{|u|} + 1 \).

<table>
<thead>
<tr>
<th>( k/\mu )</th>
<th>( \nu/\mu )</th>
<th>( u^{\text{analytic}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>0.0</td>
<td>0.3493</td>
</tr>
<tr>
<td>4.0</td>
<td>0.1</td>
<td>0.3892</td>
</tr>
<tr>
<td>4.0</td>
<td>0.2</td>
<td>0.3892</td>
</tr>
<tr>
<td>4.0</td>
<td>0.5</td>
<td>0.3892</td>
</tr>
<tr>
<td>3.0</td>
<td>0.0</td>
<td>0.3346</td>
</tr>
<tr>
<td>3.0</td>
<td>0.1</td>
<td>0.3892</td>
</tr>
<tr>
<td>3.0</td>
<td>0.2</td>
<td>0.3892</td>
</tr>
<tr>
<td>3.0</td>
<td>0.5</td>
<td>0.3892</td>
</tr>
</tbody>
</table>
3.4 Conclusion

We have generalized two classical models of evolutionary biology, the Crow-Kimura and Eigen models. We have introduced inter-individual transfer of genetic information to these models, bringing them closer to the modern understanding of evolutionary biology. For both models, we showed how to incorporate horizontal gene transfer. We showed that these generalized models may be written in an equivalent field-theoretic formulation. This mapping allows us to apply the powerful mathematical techniques of quantum field theory to obtain exact analytical solutions. For fitness landscapes that depend only on distance from a wild-type genome and for long genome lengths, we are able to solve for the mean population fitness for arbitrary functional forms of the fitness. Horizontal gene transfer of $\tilde{M}$ genetic units was shown to be analogous to horizontal gene transfer of one genetic unit, with a suitably scaled horizontal gene transfer rate.

We also showed how to incorporate recombination to these classical models, as might occur in viral super- or co-infection. This case seems at first glance far more non-linear, since on average half of the genetic material is taken from each parent to make the child, rather than $O(1)$ genes as in horizontal gene transfer. Somewhat surprisingly, we were able to exactly solve the two-parent recombination case for both the Eigen and Crow-Kimura model as well. In the limit of a long genome and for fitness landscapes that depend on the distance from a wild-type genome,
we find that the mean population fitness is independent of the average cross-over length in the recombination process. We also find two selected phases. The phase for large recombination rates is spectrally rigid, with the mean fitness and population distribution independent of the rate of recombination.

We proved the mutational deterministic hypothesis holds for horizontal gene transfer or recombination in both the parallel (Kimura) and Eigen models. That is, horizontal gene transfer and recombination reduce the mean fitness in the presence of positive epistasis and increase the fitness in the presence of negative epistasis (see Fig. 3.1 and Appendices 3.12, 3.13, 3.14, and 3.15.

For a discontinuous, sharp peak fitness landscape, we found that horizontal gene transfer does not affect the structure of the quasispecies distribution or the error threshold transition. For the sharp peak fitness function, the only appreciable effect of horizontal gene transfer is related to the potential emergence of metastability depending on the initial conditions, and we analytically determined the region of parameters space in which this situation may occur. On the other hand, even for the sharp peak fitness function, two-parent recombination induces enough mixing to enhance diversity in systems evolving under a sharp peak replication rate, thus changing the quasispecies distribution and shifting the error threshold transition. We found explicit analytical expressions for this shift.

For smooth fitness landscapes, these genetic transfers affect the steady-state pop-
ulation distribution and mean fitness. Recombination and horizontal gene transfer may, of course, dramatically change the dynamics of the evolution process as well. The most dramatic impact of these exchanges of genetic material is expected for fitness landscapes that have a correlated, biological structure that is conjugate to these exchanges [112]. Analytic investigation of such correlated fitness landscapes is perhaps one of the next steps in the development of modern theories of evolution.

Appendix 3.1

We consider Eq. (3.9) for horizontal gene transfer of blocks of fixed length \( \bar{M} \) in the parallel model. For \( \epsilon = t/M \) and \( M \to \infty \), we have

\[
\langle \tilde{z}_k | e^{-\epsilon \hat{H}} | \tilde{z}_{k-1} \rangle \simeq \langle \tilde{z}_k | \tilde{z}_{k-1} \rangle - \epsilon \langle \tilde{z}_k | \hat{H} | \tilde{z}_{k-1} \rangle \simeq \langle \tilde{z}_k | \tilde{z}_{k-1} \rangle e^{-\epsilon \langle \tilde{z}_k | \hat{H} | \tilde{z}_{k-1} \rangle}. \tag{3.117}
\]

For the Hamiltonian matrix elements in the coherent states basis, we obtain to order \( O(N^0) \)

\[
-\frac{\langle \tilde{z}_k | \hat{H} | \tilde{z}_{k-1} \rangle}{\langle \tilde{z}_k | \tilde{z}_{k-1} \rangle} = N f \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k^\dagger (j) \sigma_3 \tilde{z}_{k-1} (j) \right] + \mu \sum_{j=1}^{N} \tilde{z}_k^\dagger (j) \sigma_1 \tilde{z}_{k-1} (j) - 1 \\
+ \nu \sum_{b=0}^{N/\bar{M} - 1} \left[ \prod_{j_b=\bar{M}b+1}^{\bar{M}(b+1)} \tilde{z}_k^\dagger (j_b) D \tilde{z}_{k-1} (j_b) - 1 \right] \tag{3.118}
\]

We introduce the auxiliary field

\[
\xi_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k^\dagger (j) \sigma_3 \tilde{z}_{k-1} (j) \tag{3.119}
\]
and the conjugate field $\xi_k$ to enforce the constraint via a Laplace representation of the delta function. Substituting into Eq. (3.118) into Eq. (3.9), we obtain

$$e^{-\hat{H}t} = \lim_{M \to \infty} \int [D\bar{z}^* D\bar{z}] \int \left[ \prod_{k=0}^{M} \frac{i\epsilon N d\xi_k d\xi_k}{2\pi} \right] |\bar{z}_M\rangle \langle \bar{z}_0|$$

$$\times e^{\sum_{k=1}^{M} \sum_{j=1}^{N} \left\{ -1/2 [\xi_k(j) \bar{\xi}_k(j) + \bar{\xi}_{k-1}(j) \bar{\xi}_{k-1}(j)] - 2\bar{\xi}_k(j) \bar{\xi}_{k-1}(j) + 4\bar{\xi}_k(j) \xi_k(j) \xi_{k-1}(j) \right\}}$$

$$\times e^{-\epsilon N \sum_{k=1}^{M} \left[ \xi_k \xi_k + \mu + \frac{\epsilon}{M} f(\xi_k) - \frac{\epsilon}{\sqrt{M}} \sum_{b=0}^{N/\sqrt{M} - 1} \prod_{j_b = M_b + 1}^{M_b(b+1)} \xi_k(j_b) D\xi_{k-1}(j_b) \right]}.$$  

(3.120)

The contribution of the interaction term $\frac{\epsilon}{N} \sum_{j=1}^{N} \sum_{b=0}^{N/\sqrt{M} - 1} \prod_{j_b = M_b + 1}^{M_b(b+1)} \xi_k(j_b) D\xi_{k-1}(j_b)$ to the partition function can be treated to arbitrary order in perturbation theory using the formula $Z = Z_0 \langle e^{-\delta S} \rangle_0$, and its contribution shown to be site-independent. Moreover, this reference perturbation theory has $O(N^{-1})$ fluctuations. Thus, it can be shown that with an error $O(M/N)$ at all orders in perturbation theory, we obtain the same partition function when substituting this interaction term by

$$\frac{\epsilon}{\sqrt{M}} \left( 1/N \sum_{j=1}^{N} \xi_k(j) D\xi_{k-1}(j) \right)^M.$$  

Therefore, we define the auxiliary field

$$\phi_k = \frac{1}{N} \sum_{j=1}^{N} \xi_k(j) D\xi_{k-1}(j).$$  

(3.121)

We obtain the partition function from the trace of the evolution operator, Eq. (3.120), projected onto physical states [29]

$$Z = \text{Tr} \left[ e^{-\hat{H}t} \hat{P} \right] = \int_0^{2\pi} \left[ \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \right] \lim_{M \to \infty} \int \left[ \prod_{k=0}^{M} D\bar{z}_k^* D\bar{z}_k \right] e^{-S[\bar{z},\xi]} |_{\bar{z}_0 = e^{i\lambda} \bar{z}_M}.$$  

(3.122)
By inserting Eq. (3.121), we obtain

\[
Z = \lim_{M \to \infty} \int \left[ \mathcal{D} \bar{\xi} \mathcal{D} \xi \mathcal{D} \bar{\phi} \mathcal{D} \phi \right] e^{-N \xi^2 \sum_{k=1}^{M} \left( \xi_k \xi_k + \phi_k \phi_k - f(\xi_k) + \mu + \frac{\lambda}{M} - \xi_k \phi_k^* \right)}
\]

\[
\times \int_0^{2\pi} \left[ \prod_{j=1}^{N} \frac{d \lambda_j}{2\pi} e^{-i\lambda_j} \right] \int \left[ \prod_{k=0}^{M} \mathcal{D} \bar{z}_k \mathcal{D} z_k \right] e^{-\sum_{j=1}^{N} \sum_{k,l=1}^{M} \bar{z}_k(j) S_{kl}(j) z_l(j)} \bigg|_{\bar{z}_k = e^{i\lambda_k}}
\]

(3.123)

The matrix \( S(j) \) in Eq. (3.123) is defined by

\[
S(j) = \begin{pmatrix}
I & 0 & 0 & \ldots & -e^{i\lambda_j} A_1 \\
-A_2 & I & 0 & \ldots & 0 \\
0 & -A_3 & I & \ldots & 0 \\
\vdots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 0 & -A_M & I
\end{pmatrix}
\]

(3.124)

Here \( A_k = I + e(\xi_k \sigma_3 + \mu \sigma_1 + \phi_k D). \)

After calculating the Gaussian integral over the coherent state fields, we obtain

\[
\lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^{N} \frac{d \lambda_j}{2\pi} e^{-i\lambda_j} \int \left[ \prod_{k=0}^{M} \mathcal{D} \bar{z}_k \mathcal{D} z_k \right] e^{-\sum_{j=1}^{N} \sum_{k,l=1}^{M} \bar{z}_k(j) S_{kl}(j) z_l(j)}
\]

\[
= \lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^{N} \frac{d \lambda_j}{2\pi} e^{-i\lambda_j} [\det S(j)]^{-1}
\]

\[
= \lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^{N} \frac{d \lambda_j}{2\pi} e^{-i\lambda_j} e^{-\text{Tr} \ln[I - e^{i\lambda_j} \hat{T} \exp(e \sum_{k=1}^{M} \xi_k \sigma_3 + \mu \sigma_1 + \phi_k D)]]}
\]

\[
= \lim_{M \to \infty} \prod_{j=1}^{N} \text{Tr} \hat{T} e^{e \sum_{k=1}^{M} (\xi_k \sigma_3 + \mu \sigma_1 + \phi_k D)} = Q^N,
\]

(3.125)

where \( \hat{T} \) is the time ordering operator and

\[
Q = \text{Tr} \hat{T} e^{\int_0^t dt' (\xi_3 + \mu \sigma_1 + \phi D)}.
\]

(3.126)
With this result the partition function in Eq. (3.123) becomes Eq. (3.10).

Appendix 3.2

From Eq. (3.13), we obtain the saddle-point equations with respect to the fields $\xi_c, \phi_c$ for horizontal gene transfer of blocks of fixed length $\tilde{M}$ in the parallel model:

$$\frac{\delta}{\delta \xi_c} \left( \frac{-S_c}{Nt} \right) = -\xi_c + \frac{2\xi_c + u\phi_c}{2 \left[ \xi_c(\xi_c + u\phi_c) + \left( \mu + \frac{\phi_c}{2} \right) \right]^{1/2}} = 0$$  \hspace{1cm} (3.127)

$$\frac{\delta}{\delta \phi_c} \left( \frac{-S_c}{Nt} \right) = -\phi_c + \frac{1}{2} + \frac{u\xi_c + \mu + \frac{\phi_c}{2}}{2 \left[ \xi_c(\xi_c + u\phi_c) + \left( \mu + \frac{\phi_c}{2} \right) \right]^{1/2}} = 0.$$  \hspace{1cm} (3.128)

Then, the system of Eqs. (3.127) and (3.128) reduces to

$$\xi_c = \frac{\xi_c + \frac{u}{2}\phi_c}{\left[ \xi_c(\xi_c + u\phi_c) + \left( \mu + \frac{\phi_c}{2} \right) \right]^{1/2}}$$  \hspace{1cm} (3.129)

$$\phi_c - \frac{1}{2} = \frac{u\xi_c + \mu + \frac{\phi_c}{2}}{2 \left[ \xi_c(\xi_c + u\phi_c) + \left( \mu + \frac{\phi_c}{2} \right) \right]^{1/2}}.$$  \hspace{1cm} (3.130)

We eliminate $\xi_c, \phi_c$, to obtain

$$\frac{-S_c}{Nt} = \max_{\xi_c, \phi_c} \left\{ f(\xi_c) - \mu - \frac{\nu}{\tilde{M}} + \frac{\nu}{\tilde{M}}\phi_c^\tilde{M} + \frac{\mu}{1-u^2}(2\phi_c - 1 - u\xi_c) - \frac{\mu|u|}{1-u^2} [(2\phi_c - 1 - u\xi_c)^2 - (1-u^2)(1-\xi_c^2)]^{1/2} \right\}.$$  \hspace{1cm} (3.131)

Finally, we look for an extrema in $\phi_c$,

$$\frac{\delta}{\delta \phi_c} \left( \frac{-S_c}{Nt} \right) = \frac{\nu}{\tilde{M}}\phi_c^{\tilde{M}-1} + \frac{2\mu}{1-u^2} - \frac{\mu|u|}{1-u^2} [2(2\phi_c - 1 - u\xi_c)^2 - (1-u^2)(1-\xi_c^2)]^{1/2} = 0.$$  \hspace{1cm} (3.132)
We solve for $\phi_c$ as a function of $\xi_c$ from this equation

$$
\phi_c(\xi_c) = \frac{1 + u\xi_c}{2} + \frac{\sqrt{1 - \xi_c^2}}{2} \frac{\sqrt{1 - u^2}}{1 - \left(\frac{u}{1 + \xi_c(1-u^2)\phi_c^{-1}}\right)^2}^{1/2}.
$$

(3.133)

Substituting into Eq. (3.131), we obtain for the mean fitness or average replication rate Eq. (3.14).

**Appendix 3.3**

We consider Eq. (3.28) for horizontal gene transfer of blocks of variable length in the parallel model. For $\epsilon = t/M$ and $M \to \infty$, we have

$$
\langle z_k | e^{-\epsilon \tilde{H}} | z_{k-1} \rangle \simeq \langle z_k | \tilde{H} | z_{k-1} \rangle \simeq \langle z_k | z_{k-1} \rangle e^{-\epsilon \langle \tilde{z}_k | \tilde{H} | \tilde{z}_{k-1} \rangle}. 
$$

(3.134)

For the Hamiltonian matrix elements in the coherent states basis, we obtain

$$
-\frac{\langle \tilde{z}_k | \tilde{H} | \tilde{z}_{k-1} \rangle}{\langle \tilde{z}_k | \tilde{z}_{k-1} \rangle} = N f \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k^* (j) \sigma_3 \tilde{z}_{k-1} (j) \right] + \mu N \left[ \sum_{j=1}^{N} \tilde{z}_k^* (j) \sigma_1 \tilde{z}_{k-1} (j) - 1 \right]
$$

$$
+ \frac{\nu}{\langle M \rangle} N e^{-\langle M \rangle} + \langle M \rangle \sum_{j=1}^{N} \tilde{z}_k^* (j) D \tilde{z}_{k-1} (j) - \frac{\nu}{\langle M \rangle} N. 
$$

(3.135)

We introduce the fields

$$
\xi_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k^* (j) \sigma_3 \tilde{z}_{k-1} (j) 
$$

(3.136)

$$
\phi_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k^* (j) D \tilde{z}_{k-1} (j) 
$$

(3.137)
and the conjugate fields $\bar{\phi}_k$ and $\bar{\xi}_k$ to enforce the constraints via Laplace representations of the Dirac delta functions. Substituting into Eq. (3.28), we obtain

$$e^{-\hat{H}t} = \lim_{M \to \infty} \int [D\bar{\xi}D\bar{\phi}] \left[ \prod_{k=1}^{M} \frac{i\epsilon N d\bar{\xi}_k d\bar{\phi}_k}{2\pi} \right] |\bar{Z}_M\rangle \langle \bar{Z}_0|$$

$$\times e^{\sum_{k=1}^{M} \sum_{j=1}^{N} \{-1/2(\bar{\xi}_k(j) \cdot \bar{\xi}_k(j) + \bar{\xi}_{k-1}(j) \cdot \bar{\xi}_{k-1}(j) - 2\bar{\xi}_k(j) \cdot \bar{\xi}_{k-1}(j)) + \epsilon(\bar{\xi}_k(j) \cdot (\bar{\xi}_k + \mu + \bar{\phi}_k D) \bar{\xi}_{k-1}(j))\}}$$

$$\times e^{-\epsilon N \sum_{k=1}^{M} \left[ \bar{\xi}_k \bar{\xi}_k + \bar{\phi}_k \bar{\phi}_k + \mu + \frac{\bar{f}(\bar{\xi})}{(\bar{\xi})^2} \right] e^{-\overline{(\bar{\xi})} \cdot \overline{(\bar{\xi})}} e^{-\overline{(\bar{\xi})} (1 - \bar{\phi}_k)} }$$

(3.139)

We obtain the partition function from the trace of the evolution operator Eq. (3.138)

$$Z = \text{Tr} \left[ e^{-\hat{H}t} \hat{P} \right]$$

$$= \int_{0}^{2\pi} \left[ \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \right] \lim_{M \to \infty} \int \left[ \prod_{k=1}^{M} D\bar{\xi}_k D\bar{\phi}_k \right] e^{-\hat{S}[\bar{\xi}, \bar{\phi}]} |\bar{Z}_0 = e^{i\lambda} \bar{Z}_M\rangle$$

(3.139)

By inserting Eq. (3.138), we obtain

$$Z = \lim_{M \to \infty} \int [D\xi D\phi] e^{-N \epsilon \sum_{j=1}^{N} [\bar{\xi}_j \bar{\xi}_j + \bar{\phi}_j \bar{\phi}_j - \bar{f}(\bar{\xi}) + \mu + \frac{\bar{f}(\bar{\xi})}{(\bar{\xi})^2} e^{-\overline{(\bar{\xi})} \cdot \overline{(\bar{\xi})}} e^{-\overline{(\bar{\xi})} (1 - \bar{\phi}_j)} ]$$

$$\times \int_{0}^{2\pi} \left[ \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \right] \int \left[ \prod_{k=1}^{M} D\bar{\xi}_k D\bar{\phi}_k \right] e^{-\sum_{j=1}^{N} \sum_{k=1}^{M} \bar{\xi}_k(j) \bar{\xi}_k(j) S_{jk}(\bar{\xi}(j))} |\bar{Z}_M = e^{i\lambda} \bar{Z}_0\rangle$$

(3.140)

The matrix $S(j)$ in Eq. (3.140) is defined by

$$S(j) = \begin{pmatrix}
I & 0 & 0 & \ldots & -e^{i\lambda_j} A_1 \\
-A_2 & I & 0 & \ldots & 0 \\
0 & -A_3 & I & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & -A_M & I
\end{pmatrix}$$

(3.141)
where \( A_k = I + \epsilon(\xi_k \sigma_3 + \mu \sigma_1 + \phi_k D) \).

After calculating the Gaussian integral over the coherent state fields, we obtain
\[
\lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^N \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \int \left[ \prod_{k=1}^M Dz_k^* Dz_k \right] e^{-\sum_{j=1}^N \sum_{k,l=1}^M z_k(j) s_{kl}(j) z_l(j)}
\]
\[
= \lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^N \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \left[ \det S(j) \right]^{-1}
\]
\[
= \lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^N \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} e^{-\text{Tr} \ln[I - e^{i\lambda_j} \hat{T} \exp(e \sum_{k=1}^M \xi_k \sigma_3 + \mu \sigma_1 + \phi_k D)]}
\]
\[
= \lim_{M \to \infty} \prod_{j=1}^N \text{Tr} \hat{T} e^{e \sum_{k=1}^M (\xi_k \sigma_3 + \mu \sigma_1 + \phi_k D)} = Q^N
\]
(3.142)

where
\[
Q = \text{Tr} \hat{T} e^{\int_0^t dt' (\xi_3 + \mu \sigma_1 + \phi D)}
\]
(3.143)

With this result, in the limit \( M \to \infty \), the partition function in Eq. (3.140) becomes Eq. (3.29).

**Appendix 3.4**

We consider recombination in the parallel model. For the Hamiltonian matrix elements in the coherent states basis, we obtain to order \( \mathcal{O}(N^0) \)
\[
- \frac{\langle z_k | \hat{H} | z_{k-1} \rangle}{\langle z_k | z_{k-1} \rangle} = N f \left[ \frac{1}{N} \sum_{j=1}^N z_k^*(j) \sigma_3 z_{k-1}(j) \right] + \mu \sum_{j=1}^N [z_k^*(j) \sigma_1 z_{k-1}(j) - 1]
\]
\[
+ \nu N \left[ \{z_k^*(j) D_j(z_k) z_k(z_{k-1}(j)) \} - 1 \right]
\]
(3.144)

where the matrices \( D_j \) are defined by Eq. (3.38). We introduce the auxiliary fields
\[
\xi_k = \frac{1}{N} \sum_{j=1}^N z_k^*(j) \sigma_3 z_{k-1}(j)
\]
(3.145)
and the conjugate fields $\xi_k$ to enforce the constraints via a Laplace representation of the delta functions. Substituting into Eq. (3.40), we obtain

\[
e^{-\hat{H}t} = \lim_{M \to \infty} \int [\mathcal{D}z^* \mathcal{D}z] \int \left[ \prod_{k=1}^{M} \frac{i\epsilon N d\xi_k d\xi_k^*}{2\pi} \right] |z_{M}\rangle \langle z_0|
\]

\[
\times e^{\sum_{k=1}^{M} \sum_{j=1}^{N} \left\{ -(1/2)|z_k^*(j)|^2 z_k(j) + |z_{k-1}(j)|^2 z_{k-1}(j) - 2|z_k(j)|^2 z_{k-1}(j) + \epsilon N |z_k(j)| f(z_k(j), z_{k-1}(j)) \right\}}
\]

\[
\times e^{-\sum_{k=1}^{M} [\xi_k \xi_k + \mu + \nu - f(\xi_k) - \nu g(\{z_k(j) D_j z_{k-1}(j)\})]}
\]

\[
(3.146)
\]

We obtain the partition function from the trace of the evolution operator, Eq. (3.146), for recombination in the parallel model

\[
Z = \text{Tr} \left[ e^{-\hat{H}t} \hat{\mathcal{P}} \right] = \int_{0}^{2\pi} \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \left. \lim_{M \to \infty} \int \left[ \prod_{k=1}^{M} Dz_k^* Dz_k \right] e^{-S[z^*, z]} \right|_{z_{M} = e^{i\lambda} z_0}
\]

\[
(3.147)
\]

It is convenient to define the auxiliary field

\[
\phi_k = \frac{1}{N} \sum_{j=1}^{N} z_k^*(j) D_k z_k(j)
\]

and the corresponding $\tilde{\phi}_k$ to enforce the constraint by a Laplace representation of the Dirac delta function. From Eq. (3.147), we have

\[
Z = \lim_{M \to \infty} \int [\mathcal{D}z \mathcal{D}\xi \mathcal{D}\phi \mathcal{D}\phi] e^{-N \epsilon \sum_{k=1}^{M} [\xi_k \xi_k + \overline{\phi}_k \phi_k - f(\xi_k) + \mu + \nu - \nu g(\phi_k)]}
\]

\[
\times \int_{0}^{2\pi} \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \left. \int \left[ \prod_{k=1}^{M} Dz_k^* Dz_k \right] e^{-\sum_{j=1}^{N} \sum_{k=1}^{M} z_k^*(j) S_k(j) z_k(j)} \right|_{z_{M} = e^{i\lambda} z_0}
\]

\[
(3.149)
\]
Here, for large $N$ the function $g(\phi)$ has the singular behavior $g(\phi) = 0$ unless $\phi = 1 - \mathcal{O}(1/N)$. We also notice $g(1) = 1$. The matrix $S(j)$ in Eq. (3.149) is defined by

$$S(j) = \begin{pmatrix}
I & 0 & 0 & \ldots & -e^{i\lambda_j} A_1 \\
-A_2 & I & 0 & \ldots & 0 \\
0 & -A_3 & I & \ldots & 0 \\
\vdots & \ldots & \ldots & \ldots & 0 \\
0 & \ldots & 0 & -A_M & I
\end{pmatrix}$$

(3.150)

Here, $A_k = I + \epsilon(\xi_k \sigma_3 + \mu \sigma_1 + \bar{\phi}_k D)$. After calculating the Gaussian integral over the coherent states fields, we obtain

$$\lim_{M \to \infty} \int_0^{2\pi} \frac{d\lambda_j}{2\pi} \prod_{j=1}^N e^{-i\lambda_j} \int \prod_{k=1}^M D\hat{z}_k D\hat{\bar{z}}_k e^{-\sum_{j=1}^N \sum_{k=1}^M \frac{1}{2} \hat{z}_k S(j) \hat{z}_k}$$

$$= \lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^N \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} [\det S(j)]^{-1}$$

$$= \lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^N \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} e^{-\text{Tr} \ln[I - e^{i\lambda_j} \hat{T} \exp(\epsilon \sum_{k=1}^M (\xi_k \sigma_3 + \mu \sigma_1 + \bar{\phi}_k D))]}$$

$$= \lim_{M \to \infty} \prod_{j=1}^N \text{Tr} \ e^{\epsilon \sum_{k=1}^M (\xi_k \sigma_3 + \mu \sigma_1 + \bar{\phi}_k D)} = Q^N$$

(3.151)

where in the continuous limit

$$Q = \text{Tr} \ e^{\int_0^t dt' (\xi \sigma_3 + \mu \sigma_1 + \bar{\phi} D)}$$

(3.152)

With this result, the partition function in Eq. (3.149) becomes Eq. (3.59).
Appendix 3.5
The recombination operator

For the recombination process, we consider that in the first step, the polymerase enzyme starts the copying path in either of both parental chains with equal probability 1/2. Then, at each site, it can jump to the other chain with probability \(0 < p_c \leq 1/2\) or continue along the same chain with probability \(1 - p_c\).

As presented in Section II.C, this process is represented in the general differential Eq. (3.1) by the coefficients in Eq. (3.35)

\[
R_{kl}^i = \frac{1}{2} \sum_{\{a_j=\pm 1\}} \left( \frac{1 + s_i^k s_i^l}{2} \right)^{1+a_i} \left( \frac{1 + s_i^l s_i^l}{2} \right)^{1-a_i} \\
\times \left[ (1 - p_c) \frac{1+a_1 a_2}{p_c} \frac{1-a_1 a_2}{2} \right] \left( \frac{1 + s_2^k s_2^l}{2} \right)^{1+a_2} \left( \frac{1 + s_2^l s_2^l}{2} \right)^{1-a_2} \\
\times \left[ (1 - p_c) \frac{1+a_3 a_2}{p_c} \frac{1-a_3 a_2}{2} \right] \left( \frac{1 + s_3^k s_3^l}{2} \right)^{1+a_3} \left( \frac{1 + s_3^l s_3^l}{2} \right)^{1-a_3} \\
\times \ldots \times \left[ (1 - p_c) \frac{1+a_{N-1} a_N}{p_c} \frac{1-a_{N-1} a_N}{2} \right] \left( \frac{1 + s_N^k s_N^l}{2} \right)^{1+a_N} \left( \frac{1 + s_N^l s_N^l}{2} \right)^{1-a_N}
\]

(3.153)

The operator for this process in the Schwinger-boson representation is presented in
Here, we define the single-site recombination operator as
\[ \hat{R}(j) = a^l(j)D_j a^l(j), \]
with
\[ D_j = \begin{pmatrix} \frac{1+s_j}{2} & \frac{1+s_j}{2} \\ \frac{1-s_j}{2} & \frac{1-s_j}{2} \end{pmatrix} \]
and \( p_l = q_l/\sum_{l=1}^{2N} q_l \) is the normalized probability for sequence \( 1 \leq l \leq 2^N \).

It is possible to group the different terms in the form of Ising-like traces, by using the definition
\[ J = -(1/2) \ln[p_c/(1-p_c)], \]
the expression
\[ g(\{\hat{R}(j)\}) = \frac{1}{2} [2 \cosh(J)]^{-(N-1)} \sum_{l=1}^{2N} p_l \sum_{\{\alpha_j=\pm1\}} e^{J\sum_{j=2}^{N} \alpha_j \alpha_{j-1}} \prod_{j=1}^{N} \left[ \frac{1+\alpha_j}{2} \hat{I}_j + \frac{1-\alpha_j}{2} \hat{R}_j \right] \]
(3.156)

After the representation in terms of coherent states fields, we have \( \hat{R}(j) \rightarrow \hat{z}_k(j)D_j \hat{z}_k(j) \equiv \psi_j \), and correspondingly \( g \rightarrow g(\{\psi_j\}) \)

\[ g(\{\psi_j\}) = \frac{1}{2} [2 \cosh(J)]^{-(N-1)} \sum_{l=1}^{2N} p_l \sum_{\{\alpha_j=\pm1\}} e^{J\sum_{j=2}^{N} \alpha_j \alpha_{j-1}} \prod_{j=1}^{N} \left[ \frac{1+\alpha_j}{2} \psi_j + \frac{1-\alpha_j}{2} \psi_j \right] \]
(3.157)
It is convenient to reorganize this expression as

\[
g(\{\psi_j\}) = \frac{1}{2} [2 \cosh(J)]^{-(N-1)} \sum_{l=1}^{2N} p_l \prod_{j=1}^{N} \left( \frac{1 + \psi_j}{2} \right) \sum_{\{\alpha_j = \pm 1\}} e^{J \sum_{j=2}^{N} \alpha_j \alpha_{j-1}} \prod_{j=1}^{N} \left[ 1 + \alpha_j \frac{1 - \psi_j}{1 + \psi_j} \right]
\]

(3.158)

We define the transfer matrix

\[
T = \begin{pmatrix} e^J & e^{-J} \\ e^{-J} & e^J \end{pmatrix}
\]

(3.159)

with eigenvalues \( \lambda_+ = 2 \cosh(J) \) and \( \lambda_- = 2 \sinh(J) \).

The Ising trace in Eq. (3.158) is given by

\[
\sum_{\{\alpha_j = \pm 1\}} e^{J \sum_{j=2}^{N} \alpha_j \alpha_{j-1}} = \sum_{\{\alpha_1 = \pm 1\}} \left( \langle \alpha_1 | T^{N-1} | \alpha_1 \rangle + \langle \alpha_1 | T^{N-1} | - \alpha_1 \rangle \right)
= \text{Tr}[T^{N-1}] + \text{Tr}[T^{N-1} \sigma_1]
= \lambda_+^{N-1} + \lambda_-^{N-1} + \lambda_+^{N-1} - \lambda_-^{N-1}
= 2\lambda_+^{N-1} = 2 [2 \cosh(J)]^{N-1}
\]

(3.160)

By considering this formula, and expanding the product in Eq. (3.158), we obtain

\[
g(\{\psi_j\}) = \sum_{l=1}^{2N} p_l \prod_{j=1}^{N} \left( \frac{1 + \psi_j}{2} \right) \left\{ 1 + \sum_{j=1}^{N} \langle \alpha_j \rangle \frac{1 - \psi_j}{1 + \psi_j} + \sum_{1 \leq k < m} \langle \alpha_k \alpha_m \rangle \frac{1 - \psi_k}{1 + \psi_k} \frac{1 - \psi_m}{1 + \psi_m} + \sum_{1 \leq k < m < n} \langle \alpha_k \alpha_m \alpha_n \rangle \frac{1 - \psi_k}{1 + \psi_k} \frac{1 - \psi_m}{1 + \psi_m} \frac{1 - \psi_n}{1 + \psi_n} + \ldots + \langle \alpha_1 \alpha_2 \ldots \alpha_N \rangle \prod_{j=1}^{N} \frac{1 - \psi_j}{1 + \psi_j} \right\}
\]

(3.161)

In this notation, we defined the averages

\[
\langle \alpha_k \alpha_l \ldots \rangle = \frac{1}{2\lambda_+^{N-1}} \sum_{\{\alpha_j = \pm 1\}} e^{J \sum_{j=2}^{N} \alpha_j \alpha_{j-1}} \alpha_k \alpha_l \ldots
\]

(3.162)
We present the first and second order averages, to illustrate the general technique to obtain the higher orders.

The first order average is

\[
\langle \alpha_k \rangle = \frac{1}{2\lambda^N_+} \sum_{\alpha_j = \pm 1} e^{i \sum_{j=2}^N \alpha_j \alpha_{j-1} \alpha_k}
\]

\[
= \frac{1}{2\lambda^N_+} \text{Tr} \left\{ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} T^{k-1} \sigma_3 T^{N-k} \right\}
\]

\[
= \frac{1}{2\lambda^N_+} \text{Tr} \left\{ P^{-1} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} PP^{-1} T^{k-1} PP^{-1} \sigma_3 PP^{-1} T^{N-k} P \right\}
\]

(3.163)

To evaluate the trace, we introduced the matrix \( P \) which diagonalizes the transfer matrix \( T \)

\[
P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}
\]

(3.164)

We use the identities

\[
P^{-1} TP = \begin{pmatrix} \lambda_- & 0 \\ 0 & \lambda_+ \end{pmatrix}, \quad P^{-1} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} P = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}, \quad P^{-1} \sigma_3 P = \sigma_1
\]

(3.165)
Substituting into Eq. (3.163), we obtain

\[ \langle \alpha_k \rangle = \frac{1}{\lambda_{+}^{N-1}} \text{Tr} \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} T^{k-1} \begin{pmatrix} \lambda_{+}^{k-1} & 0 \\ 0 & \lambda_{+}^{k-1} \end{pmatrix} \begin{pmatrix} \lambda_{-}^{N-k} & 0 \\ 0 & \lambda_{+}^{N-k} \end{pmatrix} \right\} = 0 \tag{3.166} \]

a result we expect due to the symmetry of the Hamiltonian in Eq. (3.163). Following a similar procedure, we can express the second order correlation in the form

\[ \langle\alpha_k \alpha_m\rangle = \frac{1}{2\lambda_{+}^{N-1}} \text{Tr} \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} T^{k-1} T^{m-k} \sigma_3 T^{N-m} \right\} \]

\[ = \frac{1}{\lambda_{+}^{N-1}} \text{Tr} \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_{-}^{k-1} & 0 \\ 0 & \lambda_{+}^{k-1} \end{pmatrix} \begin{pmatrix} \lambda_{-}^{m-k} & 0 \\ 0 & \lambda_{+}^{m-k} \end{pmatrix} \right\} \]

\[ = \frac{\lambda_{+}^{k-1+N-m} \lambda_{-}^{m-k}}{\lambda_{+}^{N-1}} = \left( \frac{\lambda_{-}}{\lambda_{+}} \right)^{m-k} \]

\[ = (\tanh(J))^{m-k} = (1 - 2p_{c})^{m-k} \tag{3.167} \]

From the same analysis, we prove that the correlations for an odd number of \( \alpha \)'s vanish, whereas those for an even number become

\[ \langle\alpha_k \alpha_l \alpha_m \alpha_n \ldots\rangle = \left( \frac{\lambda_{-}}{\lambda_{+}} \right)^{l-k+n-m+\ldots} = (\tanh(J))^{l-k+n-m+\ldots} = (1 - 2p_{c})^{l-k+n-m+\ldots} \tag{3.168} \]
Substituting into Eq. (3.161), we obtain the finite series representation

\[
g(\{\psi_j\}) = 2^N \sum_{l=1}^{2^N} p_l \prod_{j=1}^{N} \left( \frac{1 + \psi_j^l}{2} \right) \left( 1 + \sum_{1\leq k < m}^N (1 - 2p_c)^{m-k-1} \frac{1 - \psi_k^l}{1 + \psi_k^l} \frac{1 - \psi_m^l}{1 + \psi_m^l} \right.
\]

\[
+ \sum_{1\leq k < m < n}^N (1 - 2p_c)^{m-k-q-n} \frac{1 - \psi_k^l}{1 + \psi_k^l} \frac{1 - \psi_m^l}{1 + \psi_m^l} \frac{1 - \psi_n^l}{1 + \psi_n^l} \frac{1 - \psi_q^l}{1 + \psi_q^l} \]

\[
+ \ldots + (1 - 2p_c)^{N-1} \prod_{j=1}^{N} \left( \frac{1 - \psi_j^l}{1 + \psi_j^l} \right) \right) \right)
\]

(3.169)

Finally, we can obtain the alternative representation

\[
g(\{\psi_j^l\}) = \sum_{l=1}^{2^N} p_l \left( \prod_{j=1}^{N} \left( \frac{1 + \psi_j^l}{2} \right) \right)
\]

\[
+ \sum_{1\leq k < m}^N (1 - 2p_c)^{m-k} \frac{1 - \psi_k^l}{2} \frac{1 - \psi_m^l}{2} \prod_{j \neq k, l} \frac{1 + \psi_j^l}{2}
\]

\[
+ \sum_{1\leq k < m < n < q}^N (1 - 2p_c)^{m-k-q-n} \frac{1 - \psi_k^l}{2} \frac{1 - \psi_m^l}{2} \frac{1 - \psi_n^l}{2} \frac{1 - \psi_q^l}{2}
\]

\[
\times \prod_{j \neq k, m, n, q} \frac{1 + \psi_j^l}{2} + \ldots + (1 - 2p_c)^{N-1} \prod_{j=1}^{N} \frac{1 - \psi_j^l}{2} \right) \right)
\]

(3.170)

**Appendix 3.6**

For the case of uniform crossover recombination, \( p_c = 1/2 \), a simplified analysis can be carried out to obtain the large N, or Gaussian limit, of the recombination coefficients \( R_{u_1, u_2}^{u} \). For the child sequence created from parental sequences with number of “+1” sites as \( n_1 \) and \( n_2 \), the number of child sequences, \( n \), with “+1” sites is given
by the expression

\[ n = \sum_{i=1}^{N} \left( \frac{1 + \alpha_i}{2} \left( s_i + \frac{1 - \alpha_i}{2} s_i^2 \right) \right) \]  

(3.171)

Here, the path followed by the polymerase while copying from either parental sequence is parametrized by the random variables \( \alpha_i = \pm 1 \), with \( \langle \alpha_i \rangle = 0 \) and \( \langle \alpha_i \alpha_k \rangle = \delta_{ij} \).

From Eq. (3.171), we obtain the corresponding expression for the average composition of the child sequence, \( u = (N - 2n)/N \)

\[ u = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1 + \alpha_i}{2} s_i + \frac{1 - \alpha_i}{2} s_i^2 \right) \]  

(3.172)

From Eq. (3.172), we obtain the average

\[ \langle u \rangle_\alpha = \frac{1}{N} \sum_{i=1}^{N} \frac{s_i^1 + s_i^2}{2} = \frac{u_1 + u_2}{2} \]  

(3.173)

To obtain the variance, we calculate

\[ \langle u^2 \rangle_\alpha = \frac{1}{N^2} \sum_{i,j=0}^{N} \langle \left( \frac{1 + \alpha_i}{2} s_i + \frac{1 - \alpha_i}{2} s_i^2 \right) \left( \frac{1 + \alpha_j}{2} s_j + \frac{1 - \alpha_j}{2} s_j^2 \right) \rangle_\alpha \]

\[ = \frac{1}{4N} \sum_{i,j=1}^{N} (s_i^1 + s_i^2)(s_j^1 + s_j^2) + \frac{1}{4N^2} \sum_{i=1}^{N} \langle (\alpha_is_i^1 - \alpha_is_i^2)^2 \rangle_\alpha \]

\[ = \langle u \rangle_\alpha^2 + \frac{1}{4N^2} \sum_{i=1}^{N} (s_i^1 - s_i^2)^2 \]  

(3.174)

Therefore, we obtain the variance as

\[ \langle (\delta u)^2 \rangle_\alpha = \frac{1}{4N^2} N \cdot 4 \cdot 2 \cdot \frac{1 + u}{2} \cdot \frac{1 - u}{2} = \frac{1 - u^2}{2N} \]  

(3.175)
Hence, in the large \( N \) Gaussian limit, the recombination coefficients are given by the distribution

\[
R_{u_1,u_2}^u \sim \frac{e^{-N[(u_1+u_2)/2-u^2]/(1-u_2^2)}}{\sqrt{\pi(1-u_2^2)/N}}
\]

where \( f_m = f(u_c) \).

**Appendix 3.7**

We consider the saddle point condition for recombination in the parallel model.

First, we look for the saddle-point condition with respect to the fields \( \xi_c, \phi_c \)

\[
\frac{\delta}{\delta \xi_c} \left( -\frac{S_c}{Nt} \right) = -\xi_c + \frac{2\xi_c + u\phi_c}{2 \left[ \xi_c(\xi_c + u\phi_c) + \left( \mu + \frac{\phi_c}{2} \right)^2 \right]^{1/2}} = 0
\]

\[
\frac{\delta}{\delta \phi_c} \left( -\frac{S_c}{Nt} \right) = -\phi_c + \frac{1}{2} + \frac{u\xi_c + \mu + \frac{\phi_c}{2}}{2 \left[ \xi_c(\xi_c + u\phi_c) + \left( \mu + \frac{\phi_c}{2} \right)^2 \right]^{1/2}} = 0
\]

Eqs. (3.177) and (3.178) become

\[
\xi_c = \frac{2\xi_c + u\phi_c}{2 \left[ \xi_c(\xi_c + u\phi_c) + \left( \mu + \frac{\phi_c}{2} \right)^2 \right]^{1/2}}
\]

\[
\phi_c = \frac{1}{2} + \frac{u\xi_c + \mu + \frac{\phi_c}{2}}{2 \left[ \xi_c(\xi_c + u\phi_c) + \left( \mu + \frac{\phi_c}{2} \right)^2 \right]^{1/2}}
\]

By combining Eqs. (3.179) and (3.180), with the saddle-point action Eq. (3.58), we obtain Eq. (3.59).
Appendix 3.8

We consider horizontal gene transfer of blocks of length $M$ in the Eigen model.

The matrix elements of the Hamiltonian in the basis of coherent states are given by

$$
\langle \tilde{z}_k | \hat{H} | \tilde{z}_{k-1} \rangle = N e^{-\mu + \frac{k}{N} \sum_{j=1}^{N} \xi_k(j) \sigma_1 \tilde{z}_{k-1}(j)}
$$

$$
\times e^{-\frac{k}{M} + \frac{k}{N} \sum_{j=0}^{N/M-1} \prod_{z_{k-b+1}}^{M/b} \tilde{z}(j) D \tilde{z}_{k-1}(j)} f \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) \sigma_3 \tilde{z}_{k-1}(j) \right]
$$

$$
-N d \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) \sigma_3 \tilde{z}_{k-1}(j) \right]
$$

(3.181)

We introduce the auxiliary fields

$$
\xi_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) \sigma_3 \tilde{z}_{k-1}(j)
$$

(3.182)

$$
\eta_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) \sigma_1 \tilde{z}_{k-1}(j)
$$

(3.183)

and the corresponding conjugate fields $\bar{\xi}_k$, $\bar{\eta}_k$ to enforce the constraints via Laplace representations of the Dirac delta functions. Therefore, Eq. (3.77) becomes

$$
e^{-\hat{H}t} = \lim_{M \to \infty} \int [ \mathcal{D} \tilde{z}^* \mathcal{D} \tilde{z} ] | \tilde{z}_M \rangle \langle \tilde{z}_0 | \int \left[ \prod_{k=1}^{M} \frac{i e N d \xi_k d \xi_k \eta_k d \eta_k}{2\pi} \prod_{k=1}^{M} \frac{i e N d \eta_k d \eta_k}{2\pi} \right]
$$

$$\times e^{-t \sum_{k=1}^{M} \sum_{j=1}^{N} \tilde{z}_k(j) \tilde{z}_k(j) + \tilde{z}_{k-1}(j) \tilde{z}_{k-1}(j) - 2 \tilde{z}_k(j) \tilde{z}_{k-1}(j)} e^{-t \sum_{k=1}^{M} \bar{\xi}_k \bar{\xi}_k + \bar{\eta}_k \bar{\eta}_k}
$$

$$\times e^{t \sum_{k=1}^{M} \sum_{j=1}^{N} \tilde{z}_k(j) \tilde{z}_k(j) \bar{\xi}_k(j) \bar{\eta}_k(j)} e^{-t N \sum_{k=1}^{M} \bar{\xi}_k \bar{\xi}_k + \bar{\eta}_k \bar{\eta}_k}
$$

$$\times e^{t N \sum_{k=1}^{M} \left[ e^{-\mu(1-\eta_k) - \nu M + \frac{k}{N} \sum_{j=0}^{N/M-1} \prod_{z_{k-b+1}}^{M/b} \tilde{z}(j) D \tilde{z}_{k-1}(j) f(\xi_k) - d(\xi_k) \right]} (3.184)
$$

At this point, a perturbation theory analysis similar to the case of the horizontal gene transfer of finite blocks in the Kimura model leads us to conclude that to within
error $\mathcal{O}(\bar{M}/N)$ at each order in perturbation theory, it is possible to substitute the recombination term by

$$\frac{\nu}{\bar{M}} \left( \frac{1}{N} \sum_{j=1}^{N} \bar{z}_{k-1}(j) \right)^{\bar{M}}$$

(3.185)

Then, it is convenient to introduce the auxiliary field

$$\phi_k = \frac{1}{N} \sum_{j=1}^{N} \bar{z}_k(j) \partial_{\bar{z}_{k-1}}(j)$$

(3.186)

and the corresponding $\bar{\phi}_k$ field to enforce the constraint through a Laplace representation of the Dirac delta function. The partition function is obtained from the trace of the evolution operator in Eq. (3.184)

$$Z = \text{Tr} \left[ e^{-\tilde{H} t} \hat{\tilde{P}} \right] = \int_0^{2\pi} \left[ \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \right] \lim_{M \to \infty} \int \left[ \prod_{k=0}^{M} D\bar{z}_k \bar{\partial}_{\bar{z}_k} \right] e^{-S[\bar{z},\bar{\phi}]} \bigg|_{\bar{z}_0 = e^{i\lambda} \bar{z}_M} (3.187)$$

Thus, we obtain

$$Z = \lim_{M \to \infty} \int \left[ D\bar{\xi} D\xi D\eta D\eta D\bar{\phi} D\phi \right] e^{-e N \sum_{k=1}^{M} [\bar{\xi}_k \xi_k + \eta_k \eta_k + \phi_k \bar{\phi}_k]} e^{e N \sum_{k=1}^{M} [e^{-\mu(1-\eta_k) - \nu/\bar{M} + \bar{\phi} \mu} f(\xi) - d(\xi)]}$$

$$\times \int_0^{2\pi} \left[ \frac{d\lambda_j}{2\pi} \right] \int \left[ \prod_{k=1}^{M} D\bar{z}_k \bar{\partial}_{\bar{z}_k} \right] e^{-\sum_{j=1}^{N} \sum_{k=1}^{M} \bar{z}_k(j) S_{kl}(j) \bar{z}_k(j)} \bigg|_{\bar{z}_M = e^{i\lambda_j} \bar{z}_0} (3.188)$$
The matrix $S(j)$ in Eq. (3.188) is defined by

$$
S(j) = \begin{pmatrix}
I & 0 & 0 & \ldots & -e^{i\lambda_j} A_1 \\
-A_2 & I & 0 & \ldots & 0 \\
0 & -A_3 & I & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & \ldots & \ldots & 0 \\
0 & \ldots & 0 & -A_M & I
\end{pmatrix}
$$

Here $A_k = I + \epsilon(\tilde{\xi}_k \sigma_3 + \tilde{\eta}_k \sigma_1 + \tilde{\phi}_k D)$.

After calculating the Gaussian integral over the coherent states fields, we obtain

$$
\lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^N \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \int \left[ \prod_{k=1}^M D\bar{z}_k Dz_k \right] e^{-\sum_{j=1}^N \sum_{k=1}^M \bar{z}_k (j) S_{kl}(j) \bar{z}_l (j)}
$$

$$
= \lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^N \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} [\det S(j)]^{-1}
$$

$$
= \lim_{M \to \infty} \int_0^{2\pi} \prod_{j=1}^N \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} e^{-\text{Tr} \ln[I - e^{i\lambda_j} \hat{T} \exp(\epsilon \sum_{k=1}^M \tilde{\xi}_k \sigma_3 + \tilde{\eta}_k \sigma_1 + \tilde{\phi}_k D)]}
$$

$$
= \lim_{M \to \infty} \prod_{j=1}^N \text{Tr} \hat{T} e^{\epsilon \sum_{k=1}^M (\tilde{\xi}_k \sigma_3 + \tilde{\eta}_k \sigma_1 + \tilde{\phi}_k D)} = Q^N
$$

(3.190)

where

$$
Q = \text{Tr} \hat{T} e^{\int_0^t dt' (\tilde{\xi} \sigma_3 + \tilde{\eta} \sigma_1 + \tilde{\phi} D)}
$$

(3.191)

With this result the partition function in Eq. (3.188) becomes Eq. (3.78).
Appendix 3.9

We consider the saddle-point equations for horizontal gene transfer of blocks of length $M$ in the Eigen model:

$$\frac{\delta}{\delta \xi_c} \left( \frac{-S_c}{Nt} \right) = -\xi_c + \frac{\xi_c + \frac{u}{2} \phi_c}{\left[\xi_c(\xi_c + u\phi_c) + (\bar{\eta}_c + \bar{\phi}_c/2)^2\right]^{1/2}} = 0 \quad (3.192)$$

$$\frac{\delta}{\delta \phi_c} \left( \frac{-S_c}{Nt} \right) = -\phi_c + \frac{1}{2} + \frac{u\xi_c + \bar{\eta}_c + \bar{\phi}_c/2}{2 \left[\xi_c(\xi_c + u\phi_c) + (\bar{\eta}_c + \bar{\phi}_c/2)^2\right]^{1/2}} = 0 \quad (3.193)$$

$$\frac{\delta}{\delta \phi_c} \left( \frac{-S_c}{Nt} \right) = -\bar{\phi}_c + \nu \phi_c^{M-1} e^{-\mu(\bar{\eta}_c) - \frac{\xi_c}{M} + \frac{\phi_c}{M}} f(\xi_c) = 0 \quad (3.194)$$

$$\frac{\delta}{\delta \bar{\eta}_c} \left( \frac{-S_c}{Nt} \right) = -\bar{\eta}_c + \frac{\bar{\eta}_c + \bar{\phi}_c/2}{2 \left[\xi_c(\xi_c + u\phi_c) + (\bar{\eta}_c + \bar{\phi}_c/2)^2\right]^{1/2}} = 0 \quad (3.195)$$

$$\frac{\delta}{\delta \bar{\eta}_c} \left( \frac{-S_c}{Nt} \right) = -\bar{\eta}_c + \mu e^{-\mu(\bar{\eta}_c) - \frac{\xi_c}{M} + \frac{\phi_c}{M}} f(\xi_c) = 0 \quad (3.196)$$

We obtain the following identities

$$\xi_c = \frac{\bar{\xi}_c + u\bar{\phi}_c/2}{\left[\xi_c(\xi_c + u\phi_c) + (\bar{\eta}_c + \bar{\phi}_c/2)^2\right]^{1/2}} \quad (3.197)$$

$$\eta_c = \frac{\bar{\eta}_c + \bar{\phi}_c/2}{\left[\xi_c(\xi_c + u\phi_c) + (\bar{\eta}_c + \bar{\phi}_c/2)^2\right]^{1/2}} \quad (3.198)$$
\[ \tilde{\eta}_c = \mu e^{-\mu(1-\eta_c)} - \frac{\nu}{M} \left( 1 - \phi_c \right) f(\xi_c) \] (3.199)

\[ \phi_c = \frac{1}{2} + \frac{1}{2} \frac{u\xi_c + \tilde{\eta}_c + \bar{\phi}_c/2}{\left[ \xi_c(\xi_c + u\bar{\phi}_c) + (\tilde{\eta}_c + \bar{\phi}_c/2)^2 \right]^{1/2}} \] (3.200)

\[ \tilde{\phi}_c = \nu \phi_c^{M-1} e^{-\mu(1-\eta_c)} - \frac{\nu}{M} \left( 1 - \phi_c \right) f(\xi_c) \] (3.201)

Combining Eq. (3.199) and Eq. (3.201), we obtain

\[ \nu \tilde{\eta}_c \phi_c^{M-1} = \mu \tilde{\phi}_c \] (3.202)

From the system of Eqs. (3.197)-(3.202), it can be shown that

\[ -\xi_c \xi_c - \tilde{\eta}_c \eta_c - \bar{\phi}_c \phi_c + \frac{\ln Q_c}{t} = 0 \] (3.203)

**Appendix 3.10**

We consider horizontal gene transfer of blocks of variable length in the Eigen model. The Hamiltonian matrix elements in the coherent states basis are given, to \( O(N^{-1}) \), by

\[ \frac{\langle \tilde{z}_k | \hat{H} | \tilde{z}_{k-1} \rangle}{\langle \tilde{z}_k | \tilde{z}_{k-1} \rangle} = N e^{-\mu + \frac{\nu}{M} \sum_{j=1}^{N} i_{\sigma}^{\varepsilon} D_{\sigma j} - \frac{\nu}{M} \sum_{j=1}^{N} i_{\sigma}^{\varepsilon} D_{\sigma j}^{\varepsilon} - \frac{\nu}{M} \sum_{j=1}^{N} i_{\sigma}^{\varepsilon} D_{\sigma j}^{\varepsilon} - \frac{\nu}{M} \sum_{j=1}^{N} i_{\sigma}^{\varepsilon} D_{\sigma j}^{\varepsilon}} \times \left( 1 - \frac{\nu}{\langle M \rangle} + \frac{\nu}{\langle M \rangle} e^{-\nu (\langle M \rangle) + \frac{(\delta \nu)}{M} \sum_{j=1}^{N} i_{\sigma}^{\varepsilon} D_{\sigma j} - \frac{\nu}{M} \sum_{j=1}^{N} i_{\sigma}^{\varepsilon} D_{\sigma j}^{\varepsilon} - \frac{\nu}{M} \sum_{j=1}^{N} i_{\sigma}^{\varepsilon} D_{\sigma j}^{\varepsilon}} \right) \]

\[ \times \left[ \frac{1}{N} \sum_{j=1}^{N} \varepsilon_{\sigma j}^{\varepsilon} (j) \sigma \tilde{z}_{k-1}^{\varepsilon} (j) \right] - N d \left[ \frac{1}{N} \sum_{j=1}^{N} \varepsilon_{\sigma j}^{\varepsilon} (j) \sigma \tilde{z}_{k-1}^{\varepsilon} (j) \right] \] (3.204)
We introduce the auxiliary fields

\[ \xi_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) \sigma_3 \tilde{z}_{k-1}(j) \]  \hspace{1cm} (3.205)

\[ \eta_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) \sigma_1 \tilde{z}_{k-1}(j) \]  \hspace{1cm} (3.206)

\[ \phi_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) D \tilde{z}_{k-1}(j) \]  \hspace{1cm} (3.207)

and the corresponding \( \tilde{\xi}_k, \tilde{\eta}_k, \tilde{\phi}_k \) to enforce the constraints via Laplace representations of the Dirac delta functions. From Eq. (3.90), we obtain

\[
e^{-\hat{H}t} = \lim_{M \to \infty} \int [\mathcal{D} \bar{z} \mathcal{D} \tilde{z}] \int \left[ \prod_{k=1}^{M} \frac{i e N d \tilde{\xi}_k d \xi_k}{2\pi} \frac{i e N d \tilde{\eta}_k d \eta_k}{2\pi} \frac{i e N d \tilde{\phi}_k d \phi_k}{2\pi} \right] |\bar{z}_M \rangle |\bar{z}_0 \rangle \]

\[
\times e^{\sum_{k=1}^{M} \sum_{j=1}^{N} \left\{ -\frac{1}{2} \left[ (\tilde{\xi}_k(j) - \tilde{\xi}_k(j)) + (\tilde{\eta}_k(j) - \tilde{\eta}_k(j)) - 2 \tilde{z}_k(j) \tilde{z}_{k-1}(j) \right] + e^{-i \tilde{\phi}_k(j) (\tilde{\phi}_k(j) + \tilde{\phi}_k(j))} \left[ \tilde{\xi}_k(j) + \tilde{\eta}_k(j) + \tilde{\phi}_k(j) \right] \right\}}
\]

\[
\times e^{eN \sum_{k=1}^{M} \left\{ -\tilde{\xi}_k(\phi_k - \phi_k \eta_k + e^{-\mu(1 - \eta_k)} [1 - \tilde{\xi}_k(j) + \tilde{\eta}_k(j) + \tilde{\phi}_k(j)] / (\tilde{\xi}_k(j) - d(\tilde{\xi}_k)) \right\}} (3.208)
\]

We obtain the partition function from the trace of the evolution operator Eq. (3.208)

\[
Z = \text{Tr} \left[ e^{-\hat{H}t} \hat{P} \right] = \lim_{M \to \infty} \int_{0}^{2\pi} \left[ \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \right] \int \left[ \prod_{k=1}^{M} \mathcal{D} \bar{z}_k \mathcal{D} \tilde{z}_k \right] e^{-S[\bar{z}, \tilde{z}]} |\bar{z}_0 = e^{i\lambda} \tilde{z}_M \rangle (3.209)
\]
By inserting Eq. (3.208), we obtain

\[
Z = \lim_{M \to \infty} \left[ \int d\xi D\xi D\eta D\eta D\phi D\phi \right] e^{\xi N \sum_{k=1}^{M} (-\xi_k \xi_k - \eta_k \eta_k - \phi_k \phi_k)}
\times e^{\xi N \sum_{k=1}^{M} \left( e^{-\eta_k (1-\eta_k)} - \eta_k (1-\eta_k) \right) f(\xi_k) - d(\xi_k)}
\times \int_{0}^{2\pi} \left[ \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \right] \int \left[ \prod_{k=1}^{M} D\bar{z}_k D\bar{z}_k \right] e^{-\sum_{j=1}^{N} \sum_{k=1}^{M} \bar{z}_k(j) S_{kl}(j) \bar{z}_l(j)} \bigg|_{\bar{z}_k=\lambda_k \bar{z}_0} (3.210)
\]

The matrix \( S(j) \) in Eq. (3.210) is defined by

\[
S(j) = \begin{pmatrix}
I & 0 & 0 & \cdots & -e^{i\lambda_j} A_1 \\
-A_2 & I & 0 & \cdots & 0 \\
0 & -A_3 & I & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & 0 & -A_M & I
\end{pmatrix}
\]

(3.211)

Here \( A_k = I + \epsilon(\xi \sigma_3 + \eta \sigma_1 + \phi D) \).

After calculating the Gaussian integral over the coherent states fields, we obtain

\[
\lim_{M \to \infty} \int_{0}^{2\pi} \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \int \left[ \prod_{k=1}^{M} D\bar{z}_k D\bar{z}_k \right] e^{-\sum_{j=1}^{N} \sum_{k=1}^{M} \bar{z}_k(j) S_{kl}(j) \bar{z}_l(j)}
\]

\[
= \lim_{M \to \infty} \int_{0}^{2\pi} \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} [\det S(j)]^{-1}
\]

\[
= \int_{0}^{2\pi} \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} e^{-\text{Tr} \ln[I - e^{i\lambda_j} \hat{T} \exp(\epsilon \sum_{k=1}^{M} \xi_k \sigma_3 + \eta_k \sigma_1 + \phi_k D)]}
\]

\[
= \lim_{M \to \infty} \prod_{j=1}^{N} \text{Tr} \hat{T} e^{\epsilon \sum_{k=1}^{M} (\xi_k \sigma_3 + \eta_k \sigma_1 + \phi_k D)} = Q^N
\]

(3.212)
where,

\[
Q = \text{Tr} \, \hat{T} e^{\int_0^t dt' (\hat{\xi}_{3+\tilde{n}} + \hat{n}_D)}
\]  

(3.213)

With this result the partition function in Eq. (3.210) becomes Eq. (3.91).

**Appendix 3.11**

We consider recombination in the Eigen model. The matrix elements of the Hamiltonian operator in the coherent states basis are given, to order $\mathcal{O}(N)$, by

\[
-\frac{\langle \tilde{z}_k | \hat{H} | \tilde{z}_{k-1} \rangle}{\langle \tilde{z}_k | \tilde{z}_{k-1} \rangle} = Ne^{-\mu/\hbar} \sum_{j=1}^{N} \tilde{z}_k(j) \sigma_1 \tilde{z}_{k-1}(j) \times \left[ 1 - \nu + \nu g(\{ \tilde{z}_k^*(j) D_j^\dagger \tilde{z}_{k-1}(j) \}) \right] f \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k^*(j) \sigma_3 \tilde{z}_{k-1}(j) \right] - Nd \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k^*(j) \sigma_3 \tilde{z}_{k-1}(j) \right]
\]  

(3.214)

Here we notice that the function $g(\{ \tilde{z}_k^*(j) D_j^\dagger \tilde{z}_{k-1}(j) \})$ is the same as in Eq. (3.43). Therefore, the same analysis presented through Eqs. (3.43) – (3.45) regarding the singular behavior of the function $g$ applies for the Eigen model as well. Hence, in the large $N$ limit, we have $g \left( \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k^*(j) D_j \tilde{z}_{k-1}(j) \right)$, with $D = \langle D_j^\dagger \rangle$ being again the matrix defined in Eq. (3.42).

We introduce the auxiliary fields

\[
\xi_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k^*(j) \sigma_3 \tilde{z}_{k-1}(j)
\]  

(3.215)
\[ \eta_k = \frac{1}{N} \sum_{j=1}^{N} \bar{z}_k(j) \sigma_1 \bar{z}_{k-1}(j) \]  

(3.216)

\[ \phi_k = \frac{1}{N} \sum_{j=1}^{N} \bar{z}_k(j) D \bar{z}_{k-1}(j) \]  

(3.217)

and the corresponding conjugate fields \( \bar{\xi}_k, \bar{\eta}_k \) and \( \bar{\phi}_k \) to enforce the constraints via Laplace representations of the Dirac delta functions. Thus, we have

\[ e^{-iHt} = \lim_{M \to \infty} \int [D\bar{z}^* D\bar{z}] \int \left[ \prod_{k=1}^{M} \frac{i e N d\bar{\xi}_k d\xi_k}{2\pi} \frac{i e N d\bar{\eta}_k d\eta_k}{2\pi} \frac{i e N d\bar{\phi}_k d\phi_k}{2\pi} \right] \]

\[ \times \langle \bar{z}_M \rangle \langle \bar{z}_0 | e^{-\frac{1}{2} \sum_{k=1}^{M} \sum_{j=1}^{N} \left[ \bar{z}_k(j) \bar{z}_k(j) + \bar{z}_{k-1}(j) \bar{z}_{k-1}(j) - 2 \bar{z}_k(j) \bar{z}_{k-1}(j) \right]} \]

\[ \times e^{i \sum_{k=1}^{M} \sum_{j=1}^{N} \bar{z}_k(j) [\bar{\xi}_k \sigma_3 + \bar{\eta}_k \sigma_1 + \bar{\phi}_k D] \bar{z}_{k-1}(j)} e^{-\epsilon N \sum_{k=1}^{M} [\bar{\xi}_k \bar{\eta}_k + \bar{\eta}_k \bar{\phi}_k + \bar{\phi}_k \bar{\phi}_k]} \]

\[ \times e^{\epsilon N \sum_{k=1}^{M} [\bar{\xi}_k(\bar{\eta}_k + \bar{\phi}_k) \bar{\phi}_k(\bar{\eta}_k)]} \]

(3.218)

The partition function is expressed by

\[ Z = \text{Tr}[e^{-\hat{H}t} \hat{P}] = \int_0^{2\pi} \left[ \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \right] \lim_{M \to \infty} \int \left[ \prod_{k=1}^{M} D\bar{z}_k D\bar{z}_k \right] \left. e^{-S[z^*, \bar{z}]} \right|_{z_0 = e^{i\lambda} \bar{z}_M} \]

(3.219)
By inserting Eq. (3.218), we obtain

\[
Z = \lim_{M \to \infty} \int [D\xi D\bar{\xi} D\eta D\bar{\eta} D\phi D\bar{\phi}]
\times e^{-\epsilon N \sum_{k=1}^{M} [\bar{\xi}_k \xi_k + \bar{\eta}_k \eta_k + \bar{\phi}_k \phi_k]} e^{\epsilon N \sum_{k=1}^{M} [e^{-\mu(1-\eta_k)(1-\nu+\nu g(\phi_k))} f(\xi_k) - d(\xi_k)]}
\times \int_0^{2\pi} \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \int \left[ \prod_{k=1}^{M} D\bar{z}_k D\bar{z}_k \right] e^{-\sum_{j=1}^{N} \sum_{k=1}^{M} \bar{z}_j(j) \bar{s}_k(j) \bar{z}_j(j)} \bigg|_{\bar{z}_0 = e^{i\lambda} \bar{z}_M}
\]

(3.220)

The Gaussian integral can be performed over the coherent state fields, to obtain the representation in Eq. (3.103). Here, the one-dimensional Ising trace is defined by

\[
Q = \text{Tr} \hat{T} e^{\int_0^t dt (\xi_0 + \eta_0 + \phi D)}
\]

(3.221)

**Appendix 3.12**

We analyze the effect of introducing different schemes of horizontal gene transfer in the parallel model.

For the parallel model in the presence of horizontal gene transfer with blocks of size \( \tilde{M} = 1 \), we obtain

\[
\left. \frac{d\xi}{d\nu} \right|_{\nu = 0} = \frac{u_0 \xi_0 + \sqrt{1 - \xi_0^2} - 1}{2f'(u_0)}
\]

(3.222)

Here, \((\xi_0, u_0)\) represents the solution for \( \nu = 0 \), i.e., they are obtained from the system

\[
F[\xi] = f(\xi) + \mu \sqrt{1 - \xi^2} - \mu
\]

(3.223)

\[
\left. \frac{\partial F}{\partial \xi} \right|_{\xi = \xi_0} = 0 = f'(\xi_0) - \frac{\mu \xi_0}{\sqrt{1 - \xi_0^2}}
\]

(3.224)
\[ f_m = f(u_0) = \mathcal{F}[\xi_0] = f(\xi_0) + \mu \sqrt{1 - \xi_0^2} - \mu \quad (3.225) \]

From Eq. (3.225), we obtain \( u_0 \) from the inverse function

\[ u_0 = f^{-1}[\mathcal{F}[\xi_0]] = f^{-1}[f(\xi_0) + \mu \sqrt{1 - \xi_0^2} - \mu] \quad (3.226) \]

Let us Taylor-expand Eq. (3.226) near \( x = f(\xi_0) \),

\[ u_0 = f^{-1}[x] + (f^{-1})'[x] \delta x + (f^{-1})''[x] \frac{(\delta x)^2}{2} \quad (3.227) \]

with \( \delta x = \mu(\sqrt{1 - \xi_0^2} - 1) \). Here, we use the inverse function theorem to obtain the derivatives

\[ (f^{-1})'[x] = \frac{1}{f'(f^{-1}[x])} = \frac{1}{f'(\xi_0)} \]
\[ (f^{-1})''[x] = \frac{-f''(f^{-1}[x])}{(f'(f^{-1}[x]))^3} = -\frac{f''(\xi_0)}{(f'(\xi_0))^3} \quad (3.228) \]

Hence, Eq. (3.227) becomes

\[ u_0 = \xi_0 + \frac{\delta x}{f'(\xi_0)} - \frac{f''(\xi_0)}{(f'(\xi_0))^3} \frac{(\delta x)^2}{2} \quad (3.229) \]

From Eq. (3.224), we have

\[ \frac{\delta x}{f'(\xi_0)} = \frac{\mu(\sqrt{1 - \xi_0^2} - 1)}{\frac{\mu \xi_0}{\sqrt{1 - \xi_0^2}}} = \frac{1 - \xi_0^2 - \sqrt{1 - \xi_0^2}}{\xi_0} \quad (3.230) \]

From Eq. (3.229) into Eq. (3.228), after multiplying by \( \xi_0 \), we have

\[ u_0 \xi_0 = \xi_0^2 + \xi_0 \frac{\delta x}{f'(\xi_0)} - \xi_0 \frac{f''(\xi_0)}{(f'(\xi_0))^3} \frac{(\delta x)^2}{2} \]
\[ = \xi_0^2 + \xi_0 \frac{(1 - \xi_0^2 - \sqrt{1 - \xi_0^2})}{\xi_0} - \xi_0 \frac{f''(\xi_0)}{f'(\xi_0)} \frac{(\delta x)^2}{2} \]
\[ = 1 - \sqrt{1 - \xi_0^2} - f''(\xi_0) \frac{(\delta x)^2}{2(f'(\xi_0))^2} \frac{\sqrt{1 - \xi_0^2} \mu}{\mu} \quad (3.231) \]
Therefore, we finally obtain

\[ u_0 \xi_0 + \sqrt{1 - \xi_0^2} - 1 = - \frac{f''(\xi_0)}{2} \frac{(\delta x)^2}{(f'(\xi_0))^2} \frac{\sqrt{1 - \xi_0^2}}{\mu} \]  

(3.232)

The sign of this expression is clearly determined by \(-f''(\xi_0)\), and hence after Eq. (3.222) we obtain the condition

\[
\left. \frac{du}{d\nu} \right|_{\nu = 0} = \begin{cases} 
> 0 & \text{if } f''(\xi_0) < 0 \\
< 0 & \text{if } f''(\xi_0) > 0 
\end{cases}
\]  

(3.233)

From Eq. (3.233), we conclude that horizontal gene transfer will enhance selection towards the fittest individuals when negative epistasis is present \([f''(u) < 0]\), while it will introduce an additional load against selection, with the corresponding deleterious effect on the mean fitness, when positive epistasis is present \([f''(u) > 0]\). This result proves that the mutational deterministic hypothesis holds for horizontal gene transfer of blocks of size \(\bar{M} = 1\) in the parallel model.

For the case of horizontal gene transfer of blocks \(\bar{M} > 1\), we obtain the equation

\[
\left. \frac{du}{d\nu} \right|_{\nu = 0} = \frac{1 + \frac{u_0 \xi_0 - 1 + \sqrt{1 - \xi_0^2}}{2 \bar{M} f'(u_0)}}{\bar{M} f'(u_0)} - 1
\]  

(3.234)

We notice by expanding the binomial up to first order, that the leading term in Eq. (3.234) is

\[
\left. \frac{du}{d\nu} \right|_{\nu = 0} \sim \frac{u_0 \xi_0 - 1 + \sqrt{1 - \xi_0^2}}{2 f'(u_0)}
\]  

(3.235)
which is identical to Eq. (3.222), and hence the analysis presented for the case $\bar{M} = 1$ also applies for $\bar{M} > 1$, in particular Eq. (3.233).

For the process of horizontal gene transfer with multiple-size blocks, with average $\langle \bar{M} \rangle$, we obtain the equation

$$
\frac{du}{dv} \bigg|_{v=0} = \frac{e^{-\frac{\bar{M}}{2}(u_0\xi_0-1+\sqrt{1-\xi_0^2})} - 1}{\langle \bar{M} \rangle f'(u_0)} \tag{3.236}
$$

By expanding the exponential at first order, we obtain that the leading term in this case is also Eq. (3.235), which is identical to Eq. (3.222). Therefore, the analysis presented for $\bar{M} = 1$, and in particular Eq. (3.233) applies in this case as well.

In conclusion, we proved that the mutational deterministic hypothesis, expressed in quantitative form by Eq. (3.233), holds for the different forms of horizontal gene transfer discussed in our work for the parallel model.

**Appendix 3.13**

We analyze the effect of introducing different schemes of horizontal gene transfer in the Eigen model.

For the Eigen model in the presence of horizontal gene transfer, and for zero degradation rate $d(u) = 0$, we obtain the equation

$$
\frac{du}{dv} \bigg|_{v=0} = \frac{u_0\xi_0 + \sqrt{1-\xi_0^2} - 1}{2f'(u)} e^{-\mu[1-\sqrt{1-\xi_0^2}]f(\xi_0)} \tag{3.237}
$$

The sign of this derivative is determined by the combination $u_0\xi_0 + \sqrt{1-\xi_0^2} - 1$, where
\((\xi_0, u_0)\) represents the solution for \(\nu = 0\), i.e. they are obtained from the system

\[
\mathcal{F}[\xi] = f(\xi)e^{-\mu(1-\sqrt{1-\xi^2})}
\]  
(3.238)

\[
\left. \frac{\partial \mathcal{F}}{\partial \xi} \right|_{\xi=\xi_0} = 0 = \left( f'(\xi_0) - \frac{\mu \xi_0}{\sqrt{1-\xi_0^2}} \right) e^{-\mu(1-\sqrt{1-\xi_0^2})}
\]  
(3.239)

\[
f_m = f(u_0) = \mathcal{F}[\xi_0] = e^{-\mu(1-\sqrt{1-\xi_0^2})} f(\xi_0)
\]  
(3.240)

By inverting Eq. (3.240), we obtain \(u_0\)

\[
u_0 = f^{-1}[\mathcal{F}[\xi_0]] = f^{-1}[f(\xi_0)e^{-\mu(1-\sqrt{1-\xi_0^2})}]
\]  
(3.241)

We expand Eq. (3.241) near \(x = f(\xi_0)\), by applying identities Eqs. (3.227–3.230)

\[
u_0 = \xi_0 + \frac{\delta x}{f'(\xi_0)} - \frac{f''(\xi_0)}{[f'(\xi_0)]^3} \frac{(\delta x)^2}{2}
\]  
(3.242)

with \(\delta x = \left[ e^{-\mu(1-\sqrt{1-\xi_0^2})} - 1 \right] f(\xi_0) \sim -\mu [1 - \sqrt{1-\xi_0^2}] f(\xi_0)\). From Eq. (3.239), we have

\[
\frac{\delta x}{f'(\xi_0)} = \frac{\mu [\sqrt{1-\xi_0^2} - 1]}{f'(\xi_0)} = \frac{1 - \xi_0^2 - \sqrt{1-\xi_0^2}}{\xi_0}
\]  
(3.243)

From Eq. (3.243) into Eq. (3.242), after multiplying by \(\xi_0\) we find

\[
u_0\xi_0 = \xi_0^2 + \frac{\xi_0}{\xi_0} \frac{1 - \xi_0^2 - \sqrt{1-\xi_0^2}}{\xi_0} - \xi_0 \frac{(\delta x)^2}{2} \frac{f''(\xi_0)}{[f'(\xi_0)]^3}
\]

\[
= 1 - \sqrt{1-\xi_0^2} - f''(\xi_0) \frac{\xi_0(\delta x)^2}{[f'(\xi_0)]^3}
\]  
(3.244)

Hence, we obtain

\[
u_0 \xi_0 + \sqrt{1-\xi_0^2} - 1 = -f''(\xi_0) \frac{\xi_0(\delta x)^2}{[f'(\xi_0)]^3}
\]  
(3.245)
Clearly, the sign of this expression is determined by the sign of \(-f''(\xi_0)\), and hence after Eq. (3.237) we obtain the condition

\[
\left. \frac{du}{d\nu} \right|_{\nu \to 0} = \begin{cases} 
> 0 \text{ if } f''(\xi_0) < 0 \\
< 0 \text{ if } f''(\xi_0) > 0 
\end{cases} \tag{3.246}
\]

which proves that the mutational deterministic hypothesis holds for horizontal gene transfer of blocks of size \(M = 1\) in the Eigen model.

For the case of horizontal gene transfer of blocks of size \(M > 1\), we obtain the equation

\[
\left. \frac{du}{d\nu} \right|_{\nu \to 0} = \frac{\left[ 1 + \frac{u_0\xi_0 - 1 + \sqrt{1 - \xi_0^2}}{2} \right]^M - 1}{M f'(u_0)} e^{-\mu(1-\sqrt{1-\xi_0^2})} f(\xi_0) \tag{3.247}
\]

By expanding the binomial in the numerator of Eq. (3.247) up to first order, we notice that the leading term is given by

\[
\left. \frac{du}{d\nu} \right|_{\nu \to 0} \sim \frac{u_0\xi_0 - 1 + \sqrt{1 - \xi_0^2}}{2f'(u_0)} e^{-\mu(1-\sqrt{1-\xi_0^2})} f(\xi_0) \tag{3.248}
\]

which is identical to Eq. (3.237). Therefore, the analysis presented for the case \(M = 1\), and in particular Eq. (3.246) applies for \(M > 1\) as well.

When considering the process of horizontal gene transfer of blocks of multiple size with average \(\langle M \rangle\), we obtain the equation

\[
\left. \frac{du}{d\nu} \right|_{\nu \to 0} = \frac{e^{\frac{\langle M \rangle}{2}(u_0\xi_0 - 1 + \sqrt{1 - \xi_0^2})} - 1}{\langle M \rangle f'(u_0)} e^{-\mu(1-\sqrt{1-\xi_0^2})} f(\xi_0) \tag{3.249}
\]
By expanding the exponential in Eq. (3.249) up to first order, we notice that the leading term is given by Eq. (3.248) in this case as well, which is identical to Eq. (3.237). Therefore, the analysis presented for the process with $\bar{M} = 1$, and in particular Eq. (3.246), applies for the process of horizontal gene transfer of multiple size blocks as well.

Summarizing, we proved that the mutational deterministic hypothesis, expressed quantitatively in Eq. (3.246), holds for the different forms of horizontal gene transfer studied in this work for the Eigen model.

**Appendix 3.14**

For the case of two-parent recombination in the parallel model, we find that the phase structure is defined by two fitness functions. A low $\nu$-dependent phase S1, defined as the maximum in $\xi$ of

$$\mathcal{F}_{\nu}^{(1)}[\xi] = f(\xi) + \mu(\sqrt{1 - \xi^2} - 1) - \nu$$  \hspace{1cm} (3.250)

The maximum of this expression, attained at $\xi_0$, is obtained from the equation

$$\frac{\partial}{\partial \xi} \mathcal{F}_{\nu}^{(1)}[\xi_0] = f'(\xi_0) - \frac{\mu \xi_0}{\sqrt{1 - \xi_0^2}}$$  \hspace{1cm} (3.251)

We notice that the value $\xi_0$ is the same as in the absence of recombination, when $\nu = 0$. Therefore, from the self-consistency condition, we obtain for this phase

$$f_{m}^{(1)} = \mathcal{F}_{\nu}^{(1)}[\xi_0] = \mathcal{F}_0[\xi_0] - \nu = f(u_\nu)$$  \hspace{1cm} (3.252)
Here, we have denoted $u_v$ as the value of the average composition in phase SI, when the recombination rate is $\nu$. Correspondingly, we also have from Eq. (3.252) the exact relation

$$f(u_v) = f(u_0) - \nu$$ (3.253)

with $f(u_0) = \mathcal{F}_0[\xi_0]$ and $u_0$ the average composition in the absence of recombination, when $\nu = 0$.

Let us define as $u_*$ the value of the average composition at the S2 phase, which is independent of the recombination rate. The value $u_*$ is obtained as the solution of the non-linear equation

$$f'(u_*) = \frac{2\mu u_*}{1 - u_*^2}$$ (3.254)

We consider in Eq. (3.253) the value $\nu = \nu^*$ at which the average fitness of the SI and S2 phases are identical, as the condition $u_{\nu^*} = u_*$,

$$\nu^* = f(u_0) - f(u_*)$$ (3.255)

In Eq. (3.255), let us consider the Taylor expansion of $f(u_*)$ near $u_0$, up to first order in $\epsilon = u_* - u_0$,

$$\nu^* = -\epsilon f'(u_*) + O(\epsilon^2)$$ (3.256)
We expand Eq. (3.254) near \( u_0 \) at first order in \( \epsilon = u_* - u_0 \),

\[
f'(u_0) + \epsilon f''(u_0) = \frac{2\mu(u_0 + \epsilon)}{1 - (u_0 + \epsilon)^2} \sim \frac{2\mu(u_0 + \epsilon)}{1 - u_0^2} \left[ 1 - \frac{2u_0}{1 - u_0^2} \epsilon \right]^{-1} \]

\[= \frac{2\mu u_0}{1 - u_0^2} + 2\mu \frac{1 + u_0^2}{(1 - u_0^2)^2} \epsilon + \mathcal{O}(\epsilon^2) \quad (3.257)\]

We solve explicitly for \( \epsilon \) in Eq. (3.257), and combine with Eq. (3.256), to obtain an expression for \( \nu^* \)

\[
\nu^* = \frac{f'(u_0) \left[ f'(u_0) - \frac{2\mu u_0}{1 - u_0^2} \right]}{f''(u_0) - 2\mu \frac{1 + u_0^2}{(1 - u_0^2)^2}} \quad (3.258)\]

Let us now analyze the sign of \( \nu^* \) as a function of the sign of the curvature of the fitness function, as defined by \( f'' \). We consider the Laurent series of \( f(u) \) for small \( u \). That is,

\[
f(u) = ku^a \\
f'(u) = k\alpha u^{\alpha-1} \\
f''(u) = k\alpha(\alpha-1)u^{\alpha-2} \quad (3.259)\]

where \( \alpha > 0 \) to satisfy the monotonically increasing condition. This family of polynomials provides a representation of arbitrary, monotonically increasing functions for small \( u_0 \).

The case \( \alpha = 0 \), corresponding to a constant identical fitness for all sequence types in the population, possesses the trivial solution after Eq. (3.251) \( \xi_0 = 0 \), which implies \( u_0 = 0 \), and after Eq. (3.254) \( u_* = 0 \). Thus a single non-selective phase is observed for this case, both in the presence and in the absence of recombination.
From Eq. (3.259), we have $f'' < 0$ for $\alpha < 1$, $f'' > 0$ for $\alpha > 1$ and $f'' = 0$ at $\alpha = 1$. We analyze these possible cases separately. From Eq. (3.259) into Eq. (3.258), we have

$$\nu^* = \frac{k\alpha u_0^\alpha (k\alpha u_0^\alpha - \frac{2\mu u_0^2}{1-u_0^2})}{k\alpha (\alpha - 1)u_0^\alpha - 2\mu u_0^2 \frac{1+u_0^2}{(1-u_0^2)^2}}$$ (3.260)

Case 1: $\alpha < 1$, $f'' < 0$.

The denominator in Eq. (3.260) is clearly negative, since $\alpha - 1 < 0$ in this case.

The numerator, for $u_0 \ll 1$

$$k\alpha u_0^\alpha - \frac{2\mu u_0^2}{1-u_0^2} \sim k\alpha u_0^\alpha - 2\mu u_0^2 > 0$$ (3.261)

Therefore, in this case $\nu^* = \frac{(>0)}{(>0)} > 0$, and hence $u_* - u_0 > 0$.

Case 2: $1 < \alpha < 2$, $f'' > 0$.

The denominator in Eq. (3.260), for $u_0 \ll 1$ and $\alpha - 1 > 0$,

$$k\alpha (\alpha - 1)u_0^\alpha - 2\mu u_0^2 \frac{1+u_0^2}{(1-u_0^2)^2} \sim k\alpha (\alpha - 1)u_0^\alpha - 2\mu u_0^2 > 0$$ (3.262)

The numerator is also positive, by the same argument as in Eq. (3.261). Therefore, in this case $\nu^* = \frac{(>0)}{(>0)} > 0$, and hence $u_* < u_0$.

Case 3: $\alpha > 2$, $f'' > 0$.

The denominator in Eq. (3.260), for $u_0 \ll 1$ and $\alpha - 1 > 0$,

$$k\alpha (\alpha - 1)u_0^\alpha - 2\mu u_0^2 \frac{1+u_0^2}{(1-u_0^2)^2} \sim k\alpha (\alpha - 1)u_0^\alpha - 2\mu u_0^2 < 0$$ (3.263)
The numerator is
\[ k\alpha u_0^2 - \frac{2\mu u_0^2}{1 - u_0^2} \sim k\alpha u_0^2 - 2\mu u_0^2 < 0 \] (3.264)

Therefore, in this case \( \nu^* = \frac{(\leq 0)}{< 0} > 0 \), and hence \( u^* - u_0 < 0 \).

For \( \alpha = 1 \), we obtain an exact solution from Eq. (3.251), \( u_0 = \sqrt{1 + \mu^2/k^2} - \mu/k \).

This result in Eq. (3.260) yields \( \nu^* = 0 \), and thus \( u^* = u_0 \) for this particular case.

For \( \alpha = 2 \), we have the analytical solution presented in Eqs. (3.71),
\[ u^* - u_0 = \sqrt{1 - \frac{\mu}{k}} - \frac{1 - \mu}{k} = \sqrt{(1 - \frac{\mu}{k})^2 - \frac{\mu^2}{k^2}} - (1 - \frac{\mu}{k}) < 0 \] (3.265)

with \( \nu^* = \frac{\mu^2}{2k} > 0 \).

Summarizing, we proved that
\[ u^* - u_0 = \begin{cases} > 0, & f'' < 0 \\ < 0, & f'' > 0 \end{cases} \] (3.266)

This result proves the mutational deterministic hypothesis for two-parent recombination in the parallel model.

**Appendix 3.15**

For the case of two-parent recombination in the Eigen model, we find that the phase structure is defined by two fitness functions. A low \( \nu \)-dependent phase S1, defined as the maximum in \( \xi \) of
\[ \mathcal{F}_i^{(1)}[\xi] = (1 - \nu)e^{-\nu[1 - \sqrt{1 - \xi^2}]} \] (3.267)
The maximum of this expression, attained at $\xi_0$, is obtained from the equation

$$
\frac{\partial}{\partial \xi} F^{(1)}_\nu[\xi_0] = 0
$$

$$
f'(\xi_0) = \frac{\mu \xi_0}{\sqrt{1 - \xi_0^2}} f(\xi_0)
$$

$$
[\ln f(\xi_0)]' = \frac{\mu \xi_0}{\sqrt{1 - \xi_0^2}}
$$

We notice that the value $\xi_0$ is the same as in the absence of recombination, when $\nu = 0$. Therefore, from the self-consistency condition, we obtain for this phase

$$
f^{(1)} = F^{(1)}_\nu[\xi_0] = (1 - \nu) F_0[\xi_0] = f(u_\nu)
$$

Here, we have denoted $u_\nu$ as the value of the average composition in phase SI, when the recombination rate is $\nu$. Correspondingly, we also have from Eq. (3.269) the exact relation

$$
f(u_\nu) = (1 - \nu) f(u_0)
$$

with $f(u_0) = F_0[\xi_0]$ and $u_0$ the average composition in the absence of recombination, when $\nu = 0$.

Let us define as $u_*$ the value of the average composition at the S2 phase, which is independent of the recombination rate. The value $u_*$ is obtained as the solution of the non-linear equation

$$
f'(u_*) = \frac{2\mu u_*}{1 - u_*^2} f(u_*)
$$

$$
[\ln f(u_*')] = \frac{2\mu u_*}{1 - u_*^2}
$$

(3.271)
We consider in Eq. (3.269) the value \( \nu = \nu^* \) at which the average fitness of the two phases are equal, as the condition \( u_\nu = u_* \),

\[
1 - \nu^* = \frac{f(u_*)}{f(u_0)}
\]  

(3.272)

We take the logarithm of this expression, and Taylor expand up to first order in \( \epsilon = u_* - u_0 \),

\[
\ln(1 - \nu^*) = \ln[f(u_0 + \epsilon)] - \ln[f(u_0)]
\]

\[
-\nu^* = \epsilon \ln[f(u_0)]'
\]

(3.273)

We expand Eq. (3.271) near \( u_0 \) at first order in \( \epsilon = u_* - u_0 \),

\[
[\ln f(u_0)]' + \epsilon [\ln f(u_0)]'' = \frac{2\mu(u_0 + \epsilon)}{1 - (u_0 + \epsilon)^2}
\]

\[
= \frac{2\mu(u_0 + \epsilon)}{1 - u_0^2} \left[ 1 - \frac{2u_0 \epsilon}{1 - u_0^2} \right]^{-1} + \mathcal{O}(\epsilon^2)
\]

\[
= \frac{2\mu u_0}{1 - u_0^2} + 2\mu \frac{1 + u_0^2}{(1 - u_0^2)^2} \epsilon + \mathcal{O}(\epsilon^2)
\]

(3.274)

We solve explicitly for \( \epsilon \) in Eq. (3.274), and combine with Eq. (3.273), to obtain an expression for \( \nu^* \)

\[
\nu^* = [\ln f(u_0)]' \frac{[\ln f(u_0)]' - \frac{2\mu u_0}{1 - u_0^2}}{[\ln f(u_0)]'' - \frac{2\mu (1 + u_0^2)}{(1 - u_0^2)^2}}
\]

(3.275)

The analysis follows the same lines as in the parallel model case. That is, we analyze the sign of \( \nu^* \) after Eq. (3.275). We consider a family of polynomials \( f(u) = \)
\[ ku^\alpha + k_0, \text{ which for } u_0 \ll 1 \]

\[
\ln f(u) = \ln \left(1 + \frac{k}{k_0} u^\alpha \right) + \ln(k_0) \sim \frac{k}{k_0} u^\alpha + \ln(k_0)
\]

\[
\left[ \ln f(u) \right]' = \frac{k}{k_0} u^{\alpha - 1}
\]

\[
\left[ \ln f(u) \right]'' = \alpha(\alpha - 1) \frac{k}{k_0} u^{-2}
\]

(3.276)

with \( \alpha > 0 \) to satisfy the monotonically increasing condition. This family of polynomials provides a representation of smooth and monotonically increasing functions for small \( u_0 \).

The case \( \alpha = 0 \) corresponds to a constant identical fitness for all sequence types in the population, and possesses the trivial solution after Eq. (3.268) \( \xi_0 = 0 \), which implies \( u_0 = 0 \), and after Eq. (3.271) \( u_* = 0 \). Therefore, a single non-selective phase is observed for this case, both in the presence and in the absence of recombination.

From Eq. (3.276), we have \( f'' < 0 \) for \( \alpha < 1 \), \( f'' > 0 \) for \( \alpha > 1 \) and \( f'' = 0 \) at \( \alpha = 1 \). We analyze these possible cases separately. From Eq. (3.276) into Eq. (3.275), we have

\[
\nu^* = \frac{k}{k_0} \alpha u_0^\alpha \left( \frac{k}{k_0} \alpha u_0^\alpha - \frac{2\mu u_0^2}{1-u_0^2} \right)
\]

(3.277)

Case 1: \( \alpha < 1, f'' < 0 \).

The denominator in Eq. (3.277) is clearly negative, since \( \alpha - 1 < 0 \) in this case.

The numerator, for \( u_0 \ll 1 \)

\[
\frac{k}{k_0} \alpha u_0^\alpha - \frac{2\mu u_0^2}{1-u_0^2} \sim \frac{k}{k_0} \alpha u_0^\alpha - 2\mu u_0^2 > 0
\]

(3.278)
Therefore, in this case $\nu^* = \frac{\{>0\}}{\{<0\}} < 0$, and hence $u_* - u_0 > 0$.

Case 2: $1 < \alpha < 2$, $f'' > 0$.

The denominator in Eq. (3.277), for $u_0 \ll 1$ and $\alpha - 1 > 0$,

$$\frac{k}{k_0}\alpha(\alpha - 1)u^\alpha_0 - 2\mu u^2_0 \frac{1 + u^2_0}{(1 - u^2_0)^2} \sim \frac{k}{k_0}\alpha(\alpha - 1)u^\alpha_0 - 2\mu u^2_0 > 0 \quad (3.279)$$

The numerator is also positive, by the same argument as in Eq. (3.278). Therefore, in this case $\nu^* = \frac{\{>0\}}{\{<0\}}$, and hence $u_* < u_0$.

Case 3: $\alpha > 2$, $f'' > 0$.

The denominator in Eq. (3.277), for $u_0 \ll 1$ and $\alpha - 1 > 0$,

$$\frac{k}{k_0}\alpha(\alpha - 1)u^\alpha_0 - 2\mu u^2_0 \frac{1 + u^2_0}{(1 - u^2_0)^2} \sim \frac{k}{k_0}\alpha(\alpha - 1)u^\alpha_0 - 2\mu u^2_0 < 0 \quad (3.280)$$

The numerator is

$$\frac{k}{k_0}\alpha u^\alpha_0 - \frac{2\mu u^2_0}{1 - u^2_0} \sim \frac{k}{k_0}\alpha u^\alpha_0 - 2\mu u^2_0 < 0 \quad (3.281)$$

Therefore, in this case $\nu^* = \frac{\{<0\}}{\{<0\}} > 0$, and hence $u_* - u_0 < 0$.

For $\alpha = 1$, we find that for $u_* \ll 1$ and $u_0 \ll 1$, $u_* = \frac{k}{2\mu k_0} + \mathcal{O}\left(\frac{k}{2\mu k_0}\right)^2$, $\xi_0 = \frac{k}{\mu k_0} + \mathcal{O}\left(\frac{k}{2\mu k_0}\right)^2$ and $u_0 = \frac{k}{2\mu k_0} + \mathcal{O}\left(\frac{k}{2\mu k_0}\right)^2$. Therefore, $u_* - u_0 = 0$ and $\nu^* = 0$ in this case.

For $\alpha = 2$, we have the exact solution expressed in Eqs. (3.114), (3.115). The region of parameters space where phases S1 and S2 intersect is $2\frac{\mu k_0}{k} < 1$. We analyze these formulas considering that $u_* < 1$ and $u_0 < 1$. It is convenient to define in this
case the small parameter $\epsilon = 2^{\frac{\mu k_o}{k}} < 1$. From Eq. (3.271), we have

$$u_* = \frac{1}{1+\mu} - \mathcal{O}(\epsilon) \quad (3.282)$$

Expanding Eq. (3.114) up to first order in $\epsilon$, we obtain the result

$$u_0 = \sqrt{2\sqrt{1+\mu^2} - 1} - \mathcal{O}(\epsilon) \quad (3.283)$$

Therefore, for $\epsilon \ll 1$, from Eq. (3.283) and Eq. (3.282), when $\alpha = 2$, $u_* < u_0$, and hence $\nu^* > 0$.

Summarizing, we have shown that

$$u_* - u_0 = \begin{cases} 
> 0, & f'' < 0 \\
< 0, & f'' > 0 
\end{cases} \quad (3.284)$$

This result proves the mutational deterministic hypothesis for two-parent recombination in the Eigen model.
Chapter 4
Solution of the four-letter alphabet Crow-Kimura and Eigen models by Schwinger spin coherent states


Abstract

To represent the evolution of nucleic acids, we express the parallel and Eigen models for molecular evolution in terms of a functional integral representation with a four-letter alphabet, lifting the two-state, purine/pyrimidine assumption often made in quasispecies theory. We consider mutation to occur by both the Kimura 3ST and a general mutation scheme. An error catastrophe phase transition occurs in these models, and the order of the phase transition changes from second to first order for smooth fitness functions when the alphabet size is increased from two to four letters. A maximization principle is derived, which gives the mean fitness of the evolved population in terms of microscopic parameters of the evolutionary dynamics. We analyze the general analytic solution for sharp peak, linear, quadratic, and quartic fitness functions.
4.1 Introduction

Quasispecies theory is represented by two classical models of molecular evolution: the Eigen model [11, 93, 12] and the Parallel or Crow-Kimura model [95]. These models were originally formulated in the language of chemical kinetics [11], by a large system of differential equations representing the time evolution of the relative frequencies of each sequence type. Quasispecies models capture the basic microscopic processes of mutation and replication, for an infinite population of binary sequences. The most remarkable feature of these models is the existence of a phase transition, termed the "error threshold" [11, 94], when the mutation rate is below a critical value, separating a disordered non-selective phase from an organized or "quasispecies" phase. The quasispecies is characterized by a population of closely related mutants, rather than by identical sequences [11, 93, 12], and its emergence is related to the autocatalytic character of the replication process [11, 94], which exponentially enriches the proportion of the fittest mutants in the population. Experimental studies provide support for quasispecies theory in the evolution of RNA viruses [98, 99].

The choice of a binary alphabet, which simplifies the mathematical and numerical analysis of the theory, represents a coarse graining of the four-letters alphabet of nucleic acids DNA/RNA (A,C,G,T/U), by considering the two basic chemical structures of nitrogenated bases, purines (A,G) and pyrimidines (C,T/U).

Most numerical and analytical studies on quasispecies models consider the binary
alphabet simplification [11, 93, 12, 95]. In particular, the assumption of a binary alphabet allows for an exact mapping of the quasispecies models into a 2D Ising model [20, 21], or into a quantum spin chain [22, 23, 96, 25, 28]. An exception is [113], where a four-letter alphabet was studied by a quantum spin chain representation of the parallel model. Other approaches to the nucleic acids evolution problem have been presented in [114, 115].

The fact that nucleic acids in nature are constituted by a four-letter alphabet motivates the exploration of the alphabet size effect in quasispecies models, in particular how the alphabet size affects the error-threshold transition. In this article, we present an exact analytical mapping of the four-letter alphabet Crow-Kimura and Eigen models onto a quantum field theory. Our method generalizes the Schwinger spin coherent field theory in [29] for the binary alphabet to a larger alphabet. This method has also been recently applied in the solution of a model that includes transfer of genetic material between sequences in quasispecies theory [40]. We present exact analytical solutions of this field theory, in terms of a maximum principle, for the steady state mean fitness of the population and average composition. For the parallel model, our results are explicit for the Kimura 3 ST mutation scheme [116], Eq. (4.61). We develop in detail the result for the symmetric mutation rates scheme, Eq. (4.66), for four different examples of microscopic fitness functions: sharp peak Eqs. (4.68) and (4.69), Fujiyama landscape Eqs. (4.71)-(4.74), a quadratic landscape
Eqs. (4.76)-(4.79), and a quartic landscape Eqs. (4.81) and (4.82). In section 2.7, we further generalize this theory to consider an arbitrary mutation scheme, and we present an exact analytical expression for the mean fitness of the population in this general case as well, Eq. (4.101).

For the Eigen model, our analytical solutions are explicit for a mutation scheme analogous to the Kimura 3 ST [116], Eq. (4.132). We analyze in detail the solution for the symmetric mutations rate, Eq. (4.142), for four different examples of microscopic fitness functions: sharp peak Eqs. (4.143) and (4.144), Fujiyama landscape Eqs. (4.146-4.148), quadratic fitness landscape Eqs. (4.150-4.152), and quartic fitness landscape Eqs. (4.154) and (4.155).

These results bring quasispecies theory closer to the real microscopic evolutionary dynamics that occurs in the natural four-letter alphabet of nucleic acids.

4.2 The parallel model for an alphabet of size $h = 4$

The parallel model [95] describes the continuous time evolution of an infinite size population of viral genetic sequences. The evolutionary dynamics is driven by point mutations and selection, with mutations occurring in parallel and independently of viral replication. Each viral genome is represented as a sequence of $N$ 'letters', from an alphabet of size $h$, and therefore the total number of different viral genomes in the population is $h^N$.

The probability $p_i$ for a virus to have a genetic sequence $S_i$, $1 \leq i \leq h^N$, evolves
according to the following system of nonlinear differential equations

\[
\frac{dp_i}{dt} = p_i(r_i - \sum_{j=1}^{h^N} r_j p_j) + \sum_{j=1}^{h^N} \mu_{ij} p_j
\]  

(4.1)

Here \( r_i \) is replication rate of sequence \( S_i \), and \( \mu_{ij} \) is the mutation rate from sequence \( S_j \) into sequence \( S_i \). The nonlinear term in Eq. (4.1) represents the average replication rate in the population, or mean fitness. This non-linear term enforces the conservation of probability, \( \sum_i p_i = 1 \). This term can be removed through a simple exponential transformation, to obtain the linear system of differential equations

\[
\frac{dq_i}{dt} = r_i q_i + \sum_{j=1}^{h^N} \mu_{ij} q_j
\]  

(4.2)

where \( p_i(t) = q_i(t) / \sum_j q_j(t) \).

If we describe a viral genetic sequence in the alphabet of nucleic acids (DNA or RNA), the natural choice would be \( h = 4 \), and explicitly the alphabet corresponds to \( \{A,C,G,T\} \). It is common, to simplify the theoretical analysis, to choose instead a coarse grained alphabet of size \( h = 2 \), by 'lumping' together purines \( \{A,T\} \) and pyrimidines \( \{C,G\} \). In what follows, we will consider the case of \( h = 4 \) and assume the Kimura 3 ST mutation scheme \([116, 113, 114, 115]\), Fig. 4.1, as well as a general mutation scheme in Sec. 2.6.

It is worth noticing that regardless of the starting basis \( X = \{A,C,G,T/U\} \) the Kimura 3 ST mutation scheme considers mutation in three directions, with rates
Accordingly, three components of the Hamming distance between a pair of sequences $S_i$ and $S_j$ are defined as follows:

$$
\begin{align*}
\forall_{i_1, \mu_2, \mu_3} & \\
X & \rightarrow \mu_2
\end{align*}
$$

Here, $\#_X \rightarrow Y(S_i, S_j)$ is the number of sites at which $X$ and $Y$ are exchanged between sequences $S_i$ and $S_j$. The total Hamming distance between sequences $S_i$ and $S_j$ is given by

$$
\begin{align*}
d(S_i, S_j) &= d_1(S_i, S_j) + d_2(S_i, S_j) + d_3(S_i, S_j) 
\end{align*}
$$
The mutation rate is therefore modeled by the function

\[ \mu_{ij} = \begin{cases} 
\mu_\alpha, & d_\alpha(S_i, S_j) = d(S_i, S_j) = 1 \\
-N \sum_{\alpha=1}^{3} \mu_\alpha, & S_i = S_j \\
0, & d(S_i, S_j) > 1 
\end{cases} \] (4.5)

### 4.2.1 Fitness landscape

The replication rate for a sequence \( S_i \) in the parallel model Eq. (4.1) is defined by the fitness function

\[ r_i = N f[u_1(S_i), u_2(S_i), u_3(S_i)] \] (4.6)

where \((u_1, u_2, u_3)\) are defined in terms of the Hamming distance components from an arbitrary (but fixed) sequence \( S_w \), representing the 'wild type' genome.

\[ u_1(S_i) = 1 - \frac{2}{N}[d_1(S_i, S_w) + d_3(S_i, S_w)] \]
\[ u_2(S_i) = 1 - \frac{2}{N}[d_2(S_i, S_w) + d_3(S_i, S_w)] \] (4.7)
\[ u_3(S_i) = 1 - \frac{2}{N}[d_1(S_i, S_w) + d_2(S_i, S_w)] \]

In correspondence with Eq. (4.4), we define the average base composition \( u \) in terms of its components as

\[ u(S_i) = \frac{1}{3}(u_1 + u_2 + u_3) = 1 - \frac{4}{3N}(d_1 + d_2 + d_3) = 1 - \frac{4}{3N}d(S_i, S_w) \] (4.8)
Notice that according to this definition, the maximum value of \( u = 1 \) is reached when \( d_1 = d_2 = d_3 = 0 \), that is the sequence \( S_i \) being identical to the wild type \( S_w \).

The minimum value for the average base composition is obtained when one of the Hamming distance components, say \( d_i = N \), while the others are null \( d_{j \neq i} = 0 \). Then, \( d = d_i = N \) and \( u = -1/3 \).

4.2.2 Schwinger spin coherent states representation of the parallel model

We can express the parallel model in operator form, by generalizing the method presented in [29]. We define \( h = 4 \) kinds of creation and annihilation operators: \( \hat{a}^{\dagger}_\alpha(j), \hat{a}_\alpha(j), 1 \leq \alpha \leq 4 \) and \( 1 \leq j \leq N \). These operators satisfy the commutation relations

\[
\left[ \hat{a}_\alpha(i), \hat{a}^{\dagger}_\beta(j) \right] = \delta_{\alpha\beta} \delta_{ij}, \\
[\hat{a}_\alpha(i), \hat{a}_\beta(j)] = [\hat{a}^{\dagger}_\alpha(i), \hat{a}^{\dagger}_\beta(j)] = 0
\] (4.9)

These operators create/annihilate a nucleotide state \( A (\alpha = 1) \), \( C (\alpha = 2) \), \( G (\alpha = 3) \) or \( T/U (\alpha = 4) \) at position \( 1 \leq j \leq N \) in the genetic sequence. Since at each site the nucleotide corresponds to a single state, we enforce the constraint

\[
\sum_{\alpha=1}^{4} \hat{a}^{\dagger}_\alpha(j) \hat{a}_\alpha(j) = 1
\] (4.10)
for all $1 \leq j \leq N$. Therefore, the state at site $j$ corresponds to one of the four possibilities

$$
|1,0,0,0) = \left[\hat{a}_1^+(j)\right]^1 \left[\hat{a}_2^+(j)\right]^0 \left[\hat{a}_3^+(j)\right]^0 \left[\hat{a}_4^+(j)\right]^0 |0,0,0,0) \\
|0,1,0,0) = \left[\hat{a}_1^+(j)\right]^0 \left[\hat{a}_2^+(j)\right]^1 \left[\hat{a}_3^+(j)\right]^0 \left[\hat{a}_4^+(j)\right]^0 |0,0,0,0) \\
|0,0,1,0) = \left[\hat{a}_1^+(j)\right]^0 \left[\hat{a}_2^+(j)\right]^0 \left[\hat{a}_3^+(j)\right]^1 \left[\hat{a}_4^+(j)\right]^0 |0,0,0,0) \\
|0,0,0,1) = \left[\hat{a}_1^+(j)\right]^0 \left[\hat{a}_2^+(j)\right]^0 \left[\hat{a}_3^+(j)\right]^0 \left[\hat{a}_4^+(j)\right]^1 |0,0,0,0) \\
$$

(4.11)

We define $n_a^i(j)$ as the power on $\hat{a}_a^i(j)$ for the sequence state $S_i$, $1 \leq i \leq 4^N$, defined by the vectors

$$
|S_i\rangle = \prod_{j=1}^{N} |\vec{n}_j^i\rangle \quad (4.12)
$$

where $|\vec{n}_j^i\rangle = |(n_1^i,n_2^i,n_3^i,n_4^i)\rangle$.

We introduce the unnormalized population state vector

$$
|\psi\rangle = \sum_{i=1}^{4^N} q(S_i)|S_i\rangle \quad (4.13)
$$

which evolves in time according to the equation

$$
\frac{d}{dt}|\psi\rangle = -\hat{H}|\psi\rangle \quad (4.14)
$$

Here, the Hamiltonian operator, to highest order in $N$, is given by

$$
\hat{H} = \hat{m} + N f(\hat{u}_1,\hat{u}_2,\hat{u}_3) \quad (4.15)
$$
where $\breve{m}$ represents the mutation operator, and $\hat{u}_k$ represents the components of $u$, as defined by Eq. (4.8), in operator form.

Let us first discuss the mutation operator $\breve{m}$. In the most general case, 6 possible different substitutions can occur at each site in the sequence, i.e. $A \leftrightarrow C$, $A \leftrightarrow G$, $A \leftrightarrow T$, $C \leftrightarrow G$, $C \leftrightarrow T$ and $G \leftrightarrow T$. Each individual process can be written in the operator form

$$\bar{a}^\dagger(j) \tau^{\alpha\beta} \bar{a}(j)$$

(4.16)
where the matrices $\tau^{\alpha\beta}$ are explicitly defined by

$$
\tau^{12} = \begin{pmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{pmatrix},
\tau^{13} = \begin{pmatrix}
0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
\end{pmatrix},
\tau^{14} = \begin{pmatrix}
0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 \\
\end{pmatrix},
$$

$$
\tau^{23} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
\end{pmatrix},
\tau^{24} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
\end{pmatrix},
\tau^{34} = \begin{pmatrix}
0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 \\
\end{pmatrix},
$$

(4.17)

By assuming that $\mu_{\alpha\beta}$ is the mutation rate for the substitution $\alpha \leftrightarrow \beta$ at site $1 \leq j \leq N$, the most general form for the mutation operator is given by

$$
\hat{m} = \sum_{j=1}^{N} \sum_{(\alpha<\beta)}^{6} \mu_{\alpha\beta} \left[ \tilde{a}^{\dagger}(j)\tau^{\alpha\beta}\tilde{a}(j) - \tilde{a}^{\dagger}(j)\tau^{00}\tilde{a}(j) \right]
$$

(4.18)

Let us consider the particular case of the Kimura 3 ST model [95, 117, 118, 113], as
displayed in Fig. 4.1. Then, we have the following identities

\[ \mu_{12} = \mu_{34} = \mu_1 \quad \mu_{13} = \mu_{24} = \mu_2 \quad \mu_{14} = \mu_{23} = \mu_3 \] (4.19)

It is therefore convenient to cluster the 6 different terms in Eq. (4.18) into three terms, as follows

\[
\begin{align*}
\tau^{12} &+ \tau^{34} - \tau^{00} = \begin{pmatrix} \sigma^1 & 0 \\ 0 & \sigma^1 \end{pmatrix} = \sigma^1 \otimes \sigma^0 \\
\tau^{13} &+ \tau^{24} - \tau^{00} = \begin{pmatrix} 0 & \sigma^0 \\ \sigma^0 & 0 \end{pmatrix} = \sigma^0 \otimes \sigma^1 \\
\tau^{14} &+ \tau^{23} - \tau^{00} = \begin{pmatrix} 0 & \sigma^1 \\ \sigma^1 & 0 \end{pmatrix} = \sigma^1 \otimes \sigma^1 \\
\tau^{00} &= \begin{pmatrix} \sigma^0 & 0 \\ 0 & \sigma^0 \end{pmatrix} = \sigma^0 \otimes \sigma^0 
\end{align*}
\] (4.20)

Thus, we end up with elements of the algebra generated by \( SU(2) \otimes SU(2) \), with the identity matrix and the Pauli matrices defined as usual

\[
\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \] (4.21)

Therefore, under the assumptions of the Kimura 3 ST model, the mutation operator Eq. (4.18) becomes

\[
\hat{m} = \sum_{j=1}^{N} \left( \mu_1 \hat{T}^{10}(j) + \mu_2 \hat{T}^{01}(j) + \mu_3 \hat{T}^{11}(j) - (\mu_1 + \mu_2 + \mu_3) \hat{T}^{00}(j) \right) \] (4.22)
where we defined

\[ \hat{T}^{\alpha\beta}(j) = \hat{a}^\dagger(j) (\sigma^\alpha \otimes \sigma^\beta) \hat{a}(j) \]  

(4.23)

Let us now consider the Schwinger spin coherent state representation of the average base composition terms. For this purpose, we need first to express the Hamming distances, as defined by Eq. (4.4). We assume that the wild type sequence is given by \( S_w \), which is characterized by the integers \( S_w = \{ n_A^w(j), n_C^w(j), n_G^w(j), n_T^w(j) \} \), with \( n_A^w(j) + n_C^w(j) + n_G^w(j) + n_T^w(j) = 1 \), at position \( j \). Hence, the Hamming distance
components are given by the operators

\[
\hat{d}_1 = \sum_{j=1}^N \tilde{\alpha}^1(j) \begin{pmatrix}
  n_C(j) & 0 & 0 & 0 \\
  0 & n_A(j) & 0 & 0 \\
  0 & 0 & n_T(j) & 0 \\
  0 & 0 & 0 & n_G(j)
\end{pmatrix} \tilde{\alpha}(j)
\]

\[
\hat{d}_2 = \sum_{j=1}^N \tilde{\alpha}^2(j) \begin{pmatrix}
  n_C(j) & 0 & 0 & 0 \\
  0 & n_T(j) & 0 & 0 \\
  0 & 0 & n_A(j) & 0 \\
  0 & 0 & 0 & n_G(j)
\end{pmatrix} \tilde{\alpha}(j)
\]

\[
\hat{d}_3 = \sum_{j=1}^N \tilde{\alpha}^3(j) \begin{pmatrix}
  n_T(j) & 0 & 0 & 0 \\
  0 & n_C(j) & 0 & 0 \\
  0 & 0 & n_A(j) & 0 \\
  0 & 0 & 0 & n_G(j)
\end{pmatrix} \tilde{\alpha}(j)
\]

It is now straightforward to construct the average base composition operators, according to the definition Eq. (4.8)

\[
\hat{u}_i = \frac{1}{N} \sum_{j=1}^N \tilde{\alpha}^i(j) \Theta^{j(i,w)} \tilde{\alpha}(j) \quad i = 1, 2, 3
\]
where we defined the matrices

$$\Theta^{(1,w)}_j = I - 2\begin{pmatrix}
[n^w_C + n^w_F](j) & 0 & 0 & 0 \\
0 & [n^w_A + n^w_C](j) & 0 & 0 \\
0 & 0 & [n^w_C + n^w_F](j) & 0 \\
0 & 0 & 0 & [n^w_A + n^w_C](j)
\end{pmatrix}$$

$$\Theta^{(2,w)}_j = I - 2\begin{pmatrix}
[n^w_C + n^w_F](j) & 0 & 0 & 0 \\
0 & [n^w_C + n^w_F](j) & 0 & 0 \\
0 & 0 & [n^w_A + n^w_C](j) & 0 \\
0 & 0 & 0 & [n^w_A + n^w_C](j)
\end{pmatrix}$$

$$\Theta^{(3,w)}_j = I - 2\begin{pmatrix}
[n^w_C + n^w_F](j) & 0 & 0 & 0 \\
0 & [n^w_A + n^w_C](j) & 0 & 0 \\
0 & 0 & [n^w_C + n^w_F](j) & 0 \\
0 & 0 & 0 & [n^w_A + n^w_C](j)
\end{pmatrix}$$

In what follows, we shall use the vector notation

$$\vec{u} = (\hat{u}_1, \hat{u}_2, \hat{u}_3)$$

$$\Theta^w_j = (\Theta^{(1,w)}_j, \Theta^{(2,w)}_j, \Theta^{(3,w)}_j)$$

and correspondingly Eq. (4.27) becomes

$$\vec{\ddot{u}} = \frac{1}{N} \sum_{j=1}^{N} \vec{a}^t(j) \Theta^w_j \vec{a}(j)$$
Considering the previous expressions, the Hamiltonian operator becomes

\[
-\hat{H} = N_f \left[ \frac{1}{N} \sum_{j=1}^{N} \hat{a}^\dagger(j) \Theta^u_j \hat{a}(j) \right] + \sum_{j=1}^{N} \left[ \mu_1 \hat{\mathcal{T}}^{10}(j) + \mu_2 \hat{\mathcal{T}}^{01}(j) + \mu_3 \hat{\mathcal{T}}^{11}(j) - (\mu_1 + \mu_2 + \mu_3) \hat{\mathcal{T}}^{00}(j) \right]
\]

(4.30)

### 4.2.3 Functional integral representation of the parallel model

We convert the operator representation of the parallel model into a functional integral form by introducing Schwinger spin coherent states [29]. We define a coherent state by

\[
|z(j)\rangle = e^{\xi^\dagger(j)z(j) - z^\dagger(j)\xi(j)} |0,0,0,0\rangle
\]

\[
= e^{-\frac{1}{2}z^\dagger(j)z(j)} \sum_{k,l,m,n=0}^{\infty} \frac{[z_1(j)]^k [z_2(j)]^l [z_3(j)]^m [z_4(j)]^n}{\sqrt{k!l!m!n!}} |(k,l,m,n)\rangle
\]

(4.31)

Coherent states satisfy the completeness relation

\[
I = \int \prod_{j=1}^{N} \frac{d\vec{z}(j) d\vec{z}(j)}{\pi^4} |\{\vec{z}\}\rangle \langle \{\vec{z}\}|
\]

(4.32)

The overlap between a pair of coherent states is given by

\[
\langle \vec{z}'(j)|\vec{z}(j)\rangle = e^{-\frac{1}{2} \{z'^\dagger(j)z(j) - z'^\dagger(j)z(j) - z'^\dagger(j)z(j) + z'^\dagger(j)z(j)\}}
\]

(4.33)
To enforce the constraint Eq. (4.10), we introduce the projector

\[
\hat{P} = \prod_{j=1}^{N} \hat{P}(j) = \prod_{j=1}^{N} \Delta[\hat{a}^\dagger(j) \cdot \hat{a}(j) - 1] = \int_0^{2\pi} \prod_{j=1}^{N} \frac{d\lambda_j}{2\pi} e^{i\lambda_j [\hat{a}^\dagger(j) \hat{a}(j) - 1]}
\]

(4.34)

At long times, due to the Perrone-Frobenius theorem, we find that the system evolution is dominated by the unique largest eigenvalue, \(f_m\), of \(-\hat{H}\) and its corresponding eigenvector \(|\psi^*\rangle\), such that \(e^{-\hat{H}t}|\tilde{n}_0\rangle \sim e^{f_m t}|\psi^*\rangle\).

To evaluate this eigenvalue, we perform a Trotter factorization, for \(\epsilon = t/M\), with \(M \rightarrow \infty\), and introduce resolutions of the identity as defined by Eq. (4.32) at each time slice [29]

\[
e^{-\hat{H}t} = \lim_{M \rightarrow \infty} \int \left[ \prod_{k=1}^{M} \frac{d\tilde{z}_k^*(j) d\tilde{z}_k(j)}{\pi^4} \right] \langle \tilde{z}_M \rangle \prod_{k=1}^{M} \langle \tilde{z}_k | e^{-\epsilon \hat{H}} | \tilde{z}_{k-1} \rangle \langle \tilde{z}_0 \rangle
\]

(4.35)

We define the partition function

\[
Z = \text{Tr} e^{-\hat{H}t} \hat{P} = \int_0^{2\pi} \frac{d\lambda_j}{2\pi} e^{-i\lambda_j} \lim_{M \rightarrow \infty} \int \left[ \prod_{k=1}^{M} \frac{d\tilde{z}_k^*(j) d\tilde{z}_k(j)}{\pi^4} \right] e^{-S[\tilde{z}^*, \tilde{z}]}
\]

(4.36)

Here, we defined

\[
e^{-S[\tilde{z}^*, \tilde{z}]} = \prod_{k=1}^{M} \langle \tilde{z}_k | e^{-\epsilon \hat{H}} | \tilde{z}_{k-1} \rangle
\]

(4.37)
with the boundary condition $z_0(j) = e^{i\lambda_j} z_M(j)$ [29]. An explicit expression for the matrix element in the coherent states representation is

$$\langle \{z_k\}|e^{-\epsilon \hat{H}}|\{z_{k-1}\}\rangle = \exp \left( - \frac{1}{2} \sum_{j=1}^{N} [\bar{z}_k(j) \cdot \bar{z}_k(j) - 2\bar{z}_k(j) \cdot \bar{z}_{k-1}(j) + \bar{z}_{k-1}(j) \cdot \bar{z}_{k-1}(j)] \right)$$

$$- \epsilon N (\mu_1 + \mu_2 + \mu_3) + \epsilon N f \left[ \frac{1}{N} \sum_{j=1}^{N} \bar{z}_k(j) \Theta_j^{(i)} \bar{z}_{k-1}(j) \right]$$

$$+ \epsilon \sum_{j=1}^{N} \bar{z}_k(j) \{\mu_1(\sigma^1 \otimes \sigma^0) + \mu_2(\sigma^0 \otimes \sigma^1) + \mu_3(\sigma^1 \otimes \sigma^1)\} \bar{z}_{k-1}(j)$$

(4.38)

Let us now introduce a 3-component vector field $\xi_k = (\xi_k^1, \xi_k^2, \xi_k^3)$, with

$$\xi_k^i = \frac{1}{N} \sum_{j=1}^{N} \bar{z}_k(j) \Theta_j^{(i)} \bar{z}_{k-1}(j)$$

(4.39)

We enforce this constraint by introducing an integral representation of the corresponding delta function

$$1 = \int D[\xi] \prod_{k=1}^{M} \delta^{(3)} \left[ \vec{\xi}_k - \frac{1}{N} \sum_{j=1}^{N} \bar{z}_k(j) \Theta_j^{(i)} \bar{z}_{k-1}(j) \right]$$

$$= \int \left[ \prod_{k=1}^{M} \prod_{i=1}^{3} \frac{d\xi_k^i d\xi_{k-1}^i}{2\pi} \right] e^{\frac{1}{2} \sum_{k=1}^{M} \xi_k^i \xi_k^i - \frac{1}{2} \sum_{k=1}^{M} \xi_k^i \xi_{k-1}^i}$$

$$\times e^{\epsilon N \sum_{k=1}^{M} \xi_k^i \xi_k^i} e^{\epsilon \sum_{j=1}^{N} \bar{z}_k(j) \Theta_j^{(i)} \bar{z}_{k-1}(j)}$$

(4.40)
Inserting this into the functional integral Eq. (4.36), we have [29]

\[ Z = \lim_{M \to \infty} \int \mathcal{D}[\xi] \mathcal{D}[\bar{\xi}] e^{iN \sum_{k=1}^{M} \{ f(\bar{\xi}_k) - \bar{\xi}_k \xi_k - (\mu_1 + \mu_2 + \mu_3) \}} \]

\[ \times \int \mathcal{D}[\bar{z}] \mathcal{D}[z] \prod_{j=1}^{N} e^{-i\lambda_j e^{\sum_{k=1}^{M} \bar{z}_k S(j) S(j) \bar{z}_k(j)}} |_{\{z_0\} = \{ e^{i\lambda_j \bar{z}_M} \}} \]  

(4.41)

The matrix \( S(j) \) has the structure

\[
S(j) = \begin{pmatrix}
I & 0 & 0 & \ldots & -e^{-i\lambda_j A_1(j)} \\
-A_2(j) & I & 0 & \ldots & 0 \\
0 & -A_3(j) & I & \ldots & 0 \\
0 & \ldots & -A_M(j) & I \\
\end{pmatrix}
\]  

(4.42)

where \( A_k(j) = I + \epsilon [\mu_1 (\sigma^1 \otimes \sigma^0) + \mu_2 (\sigma^0 \otimes \sigma^1) + \mu_3 (\sigma^1 \otimes \sigma^1) + \bar{\xi}_k \cdot \Theta_j^y] \). After performing the Gaussian integration over the coherent state fields, we obtain

\[ Z = \lim_{M \to \infty} \int \mathcal{D}[\xi] \mathcal{D}[\bar{\xi}] e^{iN \sum_{k=1}^{M} \{ f(\bar{\xi}_k) - \bar{\xi}_k \xi_k - (\mu_1 + \mu_2 + \mu_3) \}} \]

\[ \times \int \mathcal{D}[\lambda] \prod_{j=1}^{N} e^{-i\lambda_j [\det S(j)]^{-1}} \]  

(4.43)

Here,

\[
\det S(j) = \det \left[ I - e^{i\lambda_j} \prod_{k=1}^{M} A_k(j) \right] \\
= \det \left[ I - e^{i\lambda_j} \hat{T} e^{\sum_{k=1}^{M} \mu_1 (\sigma^1 \otimes \sigma^0) + \mu_2 (\sigma^0 \otimes \sigma^1) + \mu_3 (\sigma^1 \otimes \sigma^1) + \bar{\xi}_k \cdot \Theta_j^y} \right] \\
= e^{\text{Tr} \ln \left[ I - e^{i\lambda_j} \hat{T} e^{\sum_{k=1}^{M} \mu_1 (\sigma^1 \otimes \sigma^0) + \mu_2 (\sigma^0 \otimes \sigma^1) + \mu_3 (\sigma^1 \otimes \sigma^1) + \bar{\xi}_k \cdot \Theta_j^y} \right]} 
\]  

(4.44)
where the operator \( \hat{T} \) indicates time ordering. Substituting this result in the partition function, we obtain

\[
Z = \lim_{M \to \infty} \int \mathcal{D}[\xi] \mathcal{D}[\bar{\xi}] e^{-s} e^{N \sum_{k=1}^{M} [f(\xi_k) - \xi_k \cdot \bar{\xi}_k - (\mu_1 + \mu_2 + \mu_3)]} \prod_{j=1}^{N} e^{-i \lambda_j \xi_j} e^{\sum_{k=1}^{M} [\mu_1 (\sigma^1 \otimes \sigma^0) + \mu_2 (\sigma^0 \otimes \sigma^1) + \mu_3 (\sigma^1 \otimes \sigma^1) + \xi_k \cdot \Theta_j]} \]

\[
Z = \lim_{M \to \infty} \int \mathcal{D}[\xi] \mathcal{D}[\bar{\xi}] e^{-s} e^{N \sum_{k=1}^{M} [f(\xi_k) - \xi_k \cdot \bar{\xi}_k - (\mu_1 + \mu_2 + \mu_3)]} \prod_{j=1}^{N} Q(j) \tag{4.45}
\]

with

\[
Q(j) = \lim_{M \to \infty} \text{Tr} \hat{T} \prod_{k=1}^{M} \left[ I + \epsilon (\mu_1 (\sigma^1 \otimes \sigma^0) + \mu_2 (\sigma^0 \otimes \sigma^1) + \mu_3 (\sigma^1 \otimes \sigma^1) + \xi_k \cdot \Theta_j) \right] + \sum_{k=1}^{M} [\mu_1 (\sigma^1 \otimes \sigma^0) + \mu_2 (\sigma^0 \otimes \sigma^1) + \mu_3 (\sigma^1 \otimes \sigma^1) + \xi_k \cdot \Theta_j] \tag{4.46}
\]

After taking the limit \( M \to \infty \), Eq. (4.45) becomes

\[
Z = \int \mathcal{D}[\xi] \mathcal{D}[\bar{\xi}] e^{-s(\xi, \bar{\xi})} \tag{4.47}
\]

where the effective action is given by

\[
S[\xi, \bar{\xi}] = -N \int_{0}^{t} dt' [f(\xi(t')) - \xi(t') \cdot \bar{\xi}(t') - (\mu_1 + \mu_2 + \mu_3)] - \sum_{j=1}^{N} \ln Q(j) \tag{4.48}
\]

As explained in Appendix 1, we just need to consider the four different cases for
\( \Theta_j^w \), which correspond to

\[
\Theta_A^w = [a^3 \otimes a^0, a^0 \otimes a^3, a^3 \otimes a^3]
\]
\[
\Theta_C^w = [-a^3 \otimes a^0, a^0 \otimes a^3, -a^3 \otimes a^3]
\]
\[
\Theta_G^w = [a^3 \otimes a^0, -a^0 \otimes a^3, -a^3 \otimes a^3]
\]
\[
\Theta_T^w = [-a^3 \otimes a^0, -a^0 \otimes a^3, a^3 \otimes a^3]
\]

(4.49)

With these definitions, it is convenient to express the sum over site-dependent traces in Eq. (4.48) as

\[
\sum_{j=1}^{N} \ln Q(j) = N_A^w \ln Q_A + N_C^w \ln Q_C + N_G^w \ln Q_G + N_T^w \ln Q_T
\]

(4.50)

where \( N_X^w \) represents the total number of bases of type \( X = (A,C,G, \text{ or } T/U) \) in the wild-type, regardless of their position along the sequence, and the four traces are defined by

\[
Q_X = T \varepsilon t_0^T \varepsilon t'd[\mu_1 (a^1 \otimes \sigma^0) + \mu_2 (a^0 \otimes \sigma^1) + \mu_3 (a^1 \otimes \sigma^1) + \xi(t') \cdot \Theta_X^w]
\]

(4.51)

Considering this last simplification, the effective action in Eq. (4.48) becomes

\[
S[\tilde{\xi}, \xi] = -N \int_0^T dt'[f(\tilde{\xi}(t')) - \tilde{\xi}(t') \cdot \tilde{\xi}(t') - (\mu_1 + \mu_2 + \mu_3)]
\]

\[
- N \sum_{X=(A,C,G,T)} y_X \ln Q_X
\]

(4.52)

Here \( y_X = N_X^w / N \) is the fractional composition of the wild-type, for \( X = (A,C,G \text{ or } T/U) \).
By analyzing the characteristic equation for the eigenvalues of the exponential
matrix in Eq. (4.51), it is shown in Appendix 1 that $Q_A = Q_C = Q_G = Q_T = Q$, with

$$Q = Tr e^\int_0^t dt' [\mu_1 \sigma^1 \otimes \sigma^0 + \mu_2 \sigma^0 \otimes \sigma^1 + \mu_3 \sigma^1 \otimes \sigma^1 + \xi^{3}(t) \sigma^3 \otimes \sigma^0 + \xi^{2}(t') \sigma^0 \otimes \sigma^3 + \xi^{1}(t') \sigma^3 \otimes \sigma^3]$$ (4.53)

After this analysis, the effective action has the final expression

$$S[\bar{\xi}, \xi] = -N \int_0^t dt' [f(\bar{\xi}(t')) - \bar{\xi}(t') \cdot \xi(t') - (\mu_1 + \mu_2 + \mu_3)]$$
$$-N \ln Q$$ (4.54)

Here, after Eq. (4.53) and the analysis presented in Appendix 1, we have that in the
long time limit

$$\lim_{t \to \infty} \frac{\ln Q}{t} = \lambda_{\text{max}}$$ (4.55)

with $\lambda_{\text{max}}$ the maximum eigenvalue of the matrix

$$M(\bar{\xi}_c, \xi_c, \bar{\xi}_c) = \begin{pmatrix}
(\bar{\xi}_c^1 + \bar{\xi}_c^2 + \bar{\xi}_c^3) & \mu_1 & \mu_2 & \mu_3 \\
\mu_1 & (-\bar{\xi}_c^1 + \bar{\xi}_c^2 - \bar{\xi}_c^3) & \mu_3 & \mu_2 \\
\mu_2 & \mu_3 & (\bar{\xi}_c^1 - \xi_c^2 - \xi_c^3) & \mu_1 \\
\mu_3 & \mu_2 & \mu_1 & (-\bar{\xi}_c^1 - \bar{\xi}_c^2 + \bar{\xi}_c^3)
\end{pmatrix}$$ (4.56)

4.2.4 The large N limit of the parallel model is a saddle point

Considering that the sequence length $N$ is very large, $N \to \infty$, we can evaluate
the functional integral Eq. (4.47) for the partition function by looking for a saddle
point. Considering the action defined in Eq. (4.54), we have

\[
\frac{\delta S}{\delta \xi^i} \bigg|_{\xi^c} = -N \left( \frac{\partial f[\xi^c]}{\partial \xi^i} \bigg|_c - \xi^c_i \right) = 0
\]

\[
\frac{\delta S}{\delta \xi^i} \bigg|_{\xi^c, \xi^c} = -N \left( -\xi^c_i + \frac{1}{Q} \frac{\delta Q}{\delta \xi^i} \bigg|_{\xi^c, \xi^c} \right) = 0
\]  

(4.57)

We have therefore the system of equations

\[
\xi^c_i = \frac{\partial f[\xi^c]}{\partial \xi^i} \bigg|_c
\]  

(4.58)

\[
\xi^1_c = \langle \sigma^3 \otimes \sigma^0 \rangle, \quad \xi^2_c = \langle \sigma^0 \otimes \sigma^3 \rangle, \quad \xi^3_c = \langle \sigma^3 \otimes \sigma^3 \rangle
\]  

(4.59)

where we defined

\[
\langle \cdot \rangle = \frac{\text{Tr} \hat{T}(\cdot) e^{S_c \hat{H}} dt' [\mu_1 \sigma^1 \otimes \sigma^0 + \mu_2 \sigma^0 \otimes \sigma^1 + \mu_3 \sigma^1 \otimes \sigma^1 + \xi^c_1 \sigma^3 \otimes \sigma^0 + \xi^c_2 \sigma^0 \otimes \sigma^3 + \xi^c_3 \sigma^3 \otimes \sigma^3]}{\text{Tr} \hat{T} e^{S_c \hat{H}} dt' [\mu_1 \sigma^1 \otimes \sigma^0 + \mu_2 \sigma^0 \otimes \sigma^1 + \mu_3 \sigma^1 \otimes \sigma^1 + \xi^c_1 \sigma^3 \otimes \sigma^0 + \xi^c_2 \sigma^0 \otimes \sigma^3 + \xi^c_3 \sigma^3 \otimes \sigma^3]}
\]  

(4.60)

After this saddle-point analysis, we obtain a general expression for the mean fitness \(f_m\) of the population, for an arbitrary microscopic fitness function \(f(u)\),

\[
f_m = \lim_{N, t \to \infty} \frac{-S_c}{Nt} = \max \left[ f(\xi^c) - \xi^c \cdot \xi^c - (\mu_1 + \mu_2 + \mu_3) + \lambda_{\text{max}} \right]
\]  

(4.61)

with \(\lambda_{\text{max}}\) defined in Eqs. (4.55) and (4.56).

From this general expression, the average composition \(\bar{u} = (u_1, u_2, u_3)\) is obtained by applying the self-consistent condition \(f(\bar{u}) = f_m\).

4.2.5 Analytic results for the symmetric mutational scheme

When a symmetric mutational scheme, \(\mu_1 = \mu_2 = \mu_3 = \mu\) is assumed, it is possible to obtain an exact, closed, analytical expression for the trace in Eq. (4.53). For this
case, we also have $\xi^1_c = \xi^2_c = \xi^3_c \equiv \xi_c$ and $\bar{\xi}^1_c = \bar{\xi}^2_c = \bar{\xi}^3_c \equiv \bar{\xi}_c$. Then, the three equalities in Eq. (4.59) can be added to obtain

$$3\xi_c = \langle \sigma^3 \otimes \sigma^0 + \sigma^0 \otimes \sigma^3 + \sigma^3 \otimes \sigma^3 \rangle = \frac{\partial}{\partial \xi_c} \frac{\ln Q}{t}$$

(4.62)

Thus, as shown in Appendix 1,

$$Q_c = 2 e^{-t(\xi_c + \mu)} + 2 e^{t(\xi_c + \mu)} \cosh \left[ 2t \sqrt{(\xi_c)^2 - \mu \xi_c + \mu^2} \right]$$

(4.63)

In the infinite time limit, this expression becomes

$$\lim_{t \to \infty} \frac{\ln Q_c}{t} = \lambda_{\text{max}} = \xi_c + \mu + 2 \sqrt{(\xi_c)^2 - \mu \xi_c + \mu^2}$$

(4.64)

From Eq. (4.62), we obtain

$$\xi_c = \frac{1}{3} \left( 1 + \frac{2\bar{\xi}_c - \mu}{\sqrt{(\xi_c)^2 - \xi_c \mu + \mu^2}} \right)$$

(4.65)

Substituting into Eq. (4.61), we obtain an expression for the mean fitness

$$f_m = \lim_{N,t \to \infty} \frac{-S_c}{Nt} = \max_{-\frac{1}{2} \leq \xi_c \leq 1} \left\{ f(\xi_c) - \frac{3}{2} \mu (1 + \xi_c) + \frac{3}{2} \mu \sqrt{(1 - \xi_c)(1 + 3\xi_c)} \right\}$$

(4.66)

We remark that Eq. (4.66) represents an exact analytical expression for the mean fitness of the population, for any arbitrary microscopic fitness $f(u)$, with the assumption of symmetric mutation rates $\mu_1 = \mu_2 = \mu_3 = \mu$. From this exact expression, the average composition $u$ is obtained by applying the self-consistency condition
\( f_m = f(u) \). In the following sections, we apply Eq. (4.66) to analyze in detail some examples of microscopic fitness functions: The sharp peak landscape, a Fujiyama landscape, a quadratic fitness landscape, and a quartic fitness landscape.

**The sharp peak landscape**

We shall first consider the sharp peak landscape, which is described by the function

\[
 f(u) = A\delta_{u,1} 
\]  

That is, only sequences identical to the wild-type replicate with a rate \( A > 0 \). From Eq. (4.66), we notice that this implies: \( \xi_c = 1 \), if \( A > 3\mu \), or \( \xi_c = 0 \) otherwise. Therefore, we obtain for the mean replication rate

\[
 f_m = \begin{cases} 
 A - 3\mu, & A > 3\mu \\
 0, & A < 3\mu 
\end{cases} 
\]  

(4.68)

The fraction of the population at the wild-type \( p_w \) is obtained from the self-consistent condition \( f_m = p_w A \),

\[
 p_w = \begin{cases} 
 1 - \frac{3\mu}{A}, & A > 3\mu \\
 0, & A < 3\mu 
\end{cases} 
\]  

(4.69)

There exists an error threshold in this case, which is given by the critical value \( A_{\text{crit}} = 3\mu \), as shown in Eqs. (4.68), (4.69) and displayed in Fig. 4.2. The phase transition is first order as a function of \( A/\mu \).

One may compare this result with the error threshold observed in the binary alphabet case, which is \( A_{\text{crit}} = \mu \). This result is intuitive, because in the 4 letters
Figure 4.2  Average composition $u$, magnetization $\xi_c$ and fraction of the population at the wild-type sequence $p_w$, as a function of the parameter $A/\mu$, for the sharp peak fitness. alphabet, there exist 3 mutation channels to escape from the wild type instead of just one as in the binary alphabet, and therefore a stronger selection pressure is required to retain the wild-type features.

The Fujiyama fitness landscape

The Fujiyama landscape is obtained as a linear function of the composition

$$f[u] = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3$$  \hspace{1cm} (4.70)

We will present analytical results for the symmetric case $\alpha_i \equiv \alpha$, $\mu_i \equiv \mu$. Thus, $\xi_c^1 = \xi_c^2 = \xi_c^3 = \xi_c$. Substituting in Eq. (4.66), we have

$$f_m = \max_{-\frac{1}{\xi_c} \leq \xi_c \leq 1} \left\{ 3\alpha \xi_c - \frac{3}{2} \mu(1 + \xi_c) + \frac{3}{2} \mu \sqrt{(1 - \xi_c)(1 + 3\xi_c)} \right\}$$  \hspace{1cm} (4.71)
We look for a maximum

\[
\frac{\partial f_m}{\partial \xi_c} = 3\alpha - \frac{3}{2}\mu + \frac{3}{2}\frac{\mu}{2\sqrt{1 + 2\xi_c - 3\xi_c^2}} = 0 \tag{4.72}
\]

From this equation, we obtain

\[
\xi_c = \frac{1}{3} \left(1 + \frac{2\alpha - \mu}{\sqrt{\alpha^2 - \alpha\mu + \mu^2}}\right) \tag{4.73}
\]

To obtain the average base composition \(u\), we apply the self-consistent condition \(f_m = f(u) = 3\alpha u\), to obtain

\[
u = \frac{1}{3} \left(1 - 2\frac{\mu}{\alpha} + \frac{2}{\alpha}\sqrt{\alpha^2 - \alpha\mu + \mu^2}\right) \tag{4.74}
\]

Clearly, no phase transition is observed in this fitness landscape, as \(0 < u < 1\) for \(0 < \alpha < \infty\).

**Quadratic fitness landscape**

The quadratic fitness landscape is given by the general quadratic form

\[
f(\vec{u}) = \sum_{i=1}^{3} \left(\frac{\beta_i}{2} u_i^2 + \alpha_i u_i\right) \tag{4.75}
\]

We will present the analytical solution for the symmetric case \(\alpha_i \equiv \alpha, \beta_i \equiv \beta\), with the symmetric mutation scheme \(\mu_i \equiv \mu\). Under these conditions, we have \(\xi_c^1 = \xi_c^2 = \xi_c^3 \equiv \xi_c\), and from Eq. (4.66) we have for the mean fitness

\[
f_m = \max_{-\frac{1}{2} \leq \xi_c \leq 1} \left\{\frac{3}{2} \beta \xi_c^2 + 3\alpha \xi_c - \frac{3}{2}\mu(1 + \xi_c) + \frac{3}{2}\mu \sqrt{(1 - \xi_c)(1 + 3\xi_c)}\right\} \tag{4.76}
\]
The maximum is obtained from the equation

\[
\frac{\partial f_m}{\partial \xi_c} = 3\beta \xi_c + 3\alpha - \frac{3}{2}\mu + \frac{3}{2}\mu \frac{2 - 6\xi_c}{2\sqrt{1 + 2\xi_c - 3\xi_c^2}} = 0
\]  

(4.77)

From Eq. (4.77), we obtain

\[
\beta \xi_c + \alpha - \frac{\mu}{2} = \frac{\mu}{2} \frac{3\xi_c - 1}{2\sqrt{1 + 2\xi_c - 3\xi_c^2}}
\]  

(4.78)

As shown in Appendix 2, this equation can be cast in the form of a quartic equation, whose roots are the values of $\xi_c$. The average composition $u$ is finally obtained through the self-consistency equation

\[
f_m = f(u) = \frac{3}{2}\beta u^2 + 3\alpha u
\]  

(4.79)

We find that the error threshold transition towards a selective phase for $\alpha = 0$ is defined by $\xi_c > 0$, $u > 0$, at $\beta > 3/2\mu$. The value of $u$ is continuous at the transition, although $\xi_c$ jumps from 0 to $2/3$. By expanding Eq. (4.76) near the critical point, we find that the first derivative $df_m/d\beta$ has a discontinuous jump from 0 to $2/3$. Therefore, the phase transition is first order as a function of $\beta/\mu$.

When $0 \leq \alpha/\beta \leq \frac{1}{3} \left(\sqrt{\frac{4}{3}} - 1\right)$, as shown in Appendix 2, we find a finite jump in the magnetization from $\xi_{c,+}$ to $\xi_{c,-}$, with $\xi_{c,\pm} = 1/3(1 \pm \sqrt{1 - 18\alpha/\beta - 27(\alpha/\beta)^2})$.

This result is in agreement with [113]. A complete analysis of the different possible cases other than this, is presented in Appendix 2.
Figure 4.3  Average composition $u$ and magnetization $\xi_c$ as a function of the parameter $\beta/\mu$, for the quadratic fitness when $\alpha = 0$.

4.2.6 Quartic fitness landscape

As a final example, we consider a quartic fitness landscape

$$f(\bar{u}) = \sum_{i=1}^{3} \frac{\beta_i}{4} u_i^4$$  \hspace{1cm} (4.80)

As in the previous cases, we consider the symmetric mutation rates $\mu_i \equiv \mu$, $\beta_i \equiv \beta$, and hence $\xi_{c,i} \equiv \xi_c$. Considering this fitness function in the general equation (4.66), we have that the mean fitness is given by the analytical expression

$$f_m = \max_{-\frac{1}{3} \leq \xi_c \leq 0} \left\{ \frac{3}{4} \beta \xi_c^4 - \frac{3}{2} \mu (1 + \xi_c) + \frac{3}{2} \mu \sqrt{(1 - \xi_c)(1 + 3\xi_c)} \right\}$$  \hspace{1cm} (4.81)

The average composition $u$ is obtained by applying the self-consistent condition

$$f(u) = \frac{3}{4} \beta u^4 = f_m$$  \hspace{1cm} (4.82)
In Fig. 4.4, we present the values of $u$ and $\xi_c$, as obtained from Eqs. (4.81), (4.82), as a function of the parameter $\beta/\mu$. A discontinuous jump in the bulk magnetization from $\xi_c = 0$ to $\xi_c = 0.971618$ is observed at $\beta/\mu = 3.67653$. By expanding Eq. (4.81) near the critical point, we find a discontinuous jump in the derivative $df_m/d\beta$, from 0 to 0.66841101. Therefore, the phase transition is first order in $\beta/\mu$. The average composition $u$, however, experiences a fast but smooth transition. This behavior is much alike the one observed in the sharp peak fitness landscape, Eq. (4.68) and Fig. 4.2, except for the fact that the average composition $u$ is continuous at the transition. Indeed, from a purely mathematical perspective, a fitness function following a power law $f_n(u) = ku^n$, for $0 < u < 1$, will satisfy the limit

$$\lim_{n \to \infty} f_n(u) = k\delta_{u,1} \quad (4.83)$$
4.2.7 Generalization to arbitrary mutational scheme

In the previous section, we applied the theory and presented explicit examples when the Kimura 3 ST mutation scheme is assumed. It is possible, however, to formulate a field theoretical representation of the four states parallel model for a general mutation scheme, as depicted in Fig. 4.5.

We define Hamming distance components as follows
The total Hamming distance is defined as

\[ d(S_i, S_j) = \sum_{\alpha < \beta} d_{\alpha\beta}(S_i, S_j) \]  

(4.85)

With these definitions, the mutation rate matrix corresponding to the generalized scheme in Fig. 4.5 is given by

\[ \mu_{ij} = \begin{cases} 
\mu_{\alpha\beta}, & d_{\alpha\beta}(S_i, S_j) = 1 \\
-N \sum_{1 \leq \alpha < \beta \leq 4} \mu_{\alpha\beta}, & S_i = S_j \\
0, & d(S_i, S_j) > 1 
\end{cases} \]  

(4.86)

The generalized average composition components in this expression become

\[ u_{\alpha\beta}(S_i) = 1 - \frac{2}{N} d_{\alpha\beta}(S_i, S_w), \quad -1 \leq u_{\alpha\beta} \leq 1 \]  

(4.87)
We define the total average composition as

\[ u(S_i) = \frac{1}{6} \sum_{1 \leq \alpha < \beta \leq 4} u_{\alpha \beta} = 1 - \frac{1}{3N} d(S_i, S_w), \quad 2/3 < u \leq 1 \]  

(4.88)

and the total mutation rate \( \mu = \sum_{\alpha < \beta} \mu_{\alpha \beta} \).

We define the matrices

\[ \Theta_{12}^w(j) = I - 2 \begin{pmatrix} n_{c}^w(j) & 0 & 0 & 0 \\ 0 & n_{A}^w(j) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \]

\[ \Theta_{13}^w(j) = I - 2 \begin{pmatrix} n_{c}^w(j) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & n_A(j) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \]

\[ \Theta_{14}^w(j) = I - 2 \begin{pmatrix} n_T^w(j) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & n_A(j) \end{pmatrix} \]
\[ \Theta_{23}(j) = I - 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & n_T(j) & 0 & 0 \\ 0 & 0 & n_T(j) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \]

\[ \Theta_{24}(j) = I - 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & n_T(j) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & n_T(j) \end{pmatrix} \]

\[ \Theta_{34}(j) = I - 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & n_T(j) & 0 \\ 0 & 0 & 0 & n_T(j) \end{pmatrix} \]  \hspace{1cm} (4.89)

The average composition operators are defined after the Hamming distance components

\[ \hat{u}_{\alpha\beta} = \frac{1}{N} \sum_{j=1}^{N} \tilde{a}^\dagger(j) \Theta_{\alpha\beta}(j) \tilde{a}(j) \]  \hspace{1cm} (4.90)

Considering these definitions, the Hamiltonian corresponding to the generalized mutation scheme is

\[ -\hat{H} = N f[\tilde{u}] + \sum_{j=1}^{N} \tilde{a}^\dagger(j) \left[ \sum_{1 \leq \alpha < \beta \leq 4} \mu_{\alpha\beta} \tau_{\alpha\beta} - \mu I \right] \tilde{a}(j) \]  \hspace{1cm} (4.91)
Here, \( r^{\alpha \beta} \) are defined in Eq. (4.17), and \( I \) is the identity matrix.

The microscopic fitness function should depend only on three independent components, related to the bases composition along the sequences. Therefore, we define the vector \( \vec{u} = (\hat{u}_1, \hat{u}_2, \hat{u}_3) \), with

\[
\begin{align*}
\hat{u}_1 &= \frac{1}{2}(\hat{u}_{12} + \hat{u}_{34}) \\
\hat{u}_2 &= \frac{1}{2}(\hat{u}_{13} + \hat{u}_{24}) \\
\hat{u}_3 &= \frac{1}{2}(\hat{u}_{14} + \hat{u}_{23})
\end{align*}
\]

Introducing auxiliary fields \( \xi^{\alpha \beta} \), \( \bar{\xi}^{\alpha \beta} \), and following a similar method as in the Kimura 3 ST theory, we express the generalized mutation scheme model in terms of the partition function

\[
Z = \int [\mathcal{D}\xi^{\alpha \beta}][\mathcal{D}\bar{\xi}^{\alpha \beta}] e^{-S[\xi^{\alpha \beta}, \bar{\xi}^{\alpha \beta}]}
\]

Here, the action is defined by

\[
S = -N \int_0^t dt' [f(\bar{\xi}(t')) - \sum_{1 \leq \alpha < \beta \leq 4} \xi^{\alpha \beta}(t') \bar{\xi}^{\alpha \beta}(t')] - \sum_{j=1}^{N} \ln Q(j)
\]

with

\[
Q(j) = \text{Tr} \hat{T} e^{\int_0^t dt' \sum_{\alpha < \beta} \mu_{\alpha \beta} r^{\alpha \beta} + \xi^{\alpha \beta}(t') \Theta^{\alpha \beta}(j)}
\]

and \( \bar{\xi} = (\xi^1, \xi^2, \xi^3) \). The components of this vector are defined according to the
components of \( \vec{u} \), as in Eq. (4.92)

\[
\begin{align*}
\xi^1 &= \frac{1}{2}(\xi^{12} + \xi^{24}) \\
\xi^2 &= \frac{1}{2}(\xi^{13} + \xi^{24}) \\
\xi^3 &= \frac{1}{2}(\xi^{14} + \xi^{23})
\end{align*}
\]

After the analysis presented in Appendix 4, since \( n_{A}^{w}(j) + n_{C}^{w}(j) + n_{G}^{w}(j) + n_{T}^{w}(j) = 1 \), the matrices \( \Theta_{\alpha\beta}^{w}(j) \) can be just one out of four different combinations \( \Theta_{X,\alpha\beta}^{w} \). These combinations, as defined in Appendix 4, can be summarized in the following matrix arrays

\[
\begin{align*}
\Theta_{A}^{w} &= \begin{bmatrix}
(\sigma^3, 0), (\sigma^0, 0), (\sigma^0, 0), I, I, I \\
(0, \sigma^0), (0, -\sigma^3), (0, \sigma^3), I, I, I
\end{bmatrix} \\
\Theta_{C}^{w} &= \begin{bmatrix}
(-\sigma^3, 0), I, I, (\sigma^0, 0), (\sigma^0, 0), I, I, I, I
\end{bmatrix} \\
\Theta_{G}^{w} &= \begin{bmatrix}
I, (\sigma^3, 0), I, (\sigma^0, 0), (\sigma^0, 0), I, I, I, I
\end{bmatrix} \\
\Theta_{T}^{w} &= \begin{bmatrix}
I, I, (-\sigma^3, 0), I, (\sigma^3, 0), (\sigma^0, 0), I, I, I, I
\end{bmatrix}
\end{align*}
\]

Therefore, by defining the fractional composition of the wild-type as \( y_{X} = \frac{N_{X}^{w}}{N} \)
for \( X = (A, C, G, T/U) \), we have the explicit extensive form for the action Eq. (4.94)

\[
S = -N \int_0^t dt' \left[ f(\xi(t')) - \sum_{1 \leq \alpha < \beta \leq 4} \xi^{\alpha \beta}(t') \tilde{\xi}^{\alpha \beta}(t') - \sum_{1 \leq \alpha < \beta \leq 4} \mu_{\alpha \beta} \right]
\]

\[
- N \sum_{X=A,C,G,T/U} y_X \ln Q_X
\]

(4.98)

with the four different traces defined by

\[
Q_X = \text{Tr} e^{\int_0^t dt' \sum_{\alpha < \beta} [\mu_{\alpha \beta} \tau^{\alpha \beta} + \xi^{\alpha \beta}(t') \Theta^{\alpha \beta}_{X,\alpha \beta}]}\]

(4.99)

and \( \Theta^{\alpha \beta}_{X,\alpha \beta} \) as the different components of the matrix arrays defined by Eq. (4.97), for \( X = (A, C, G, T/U) \). After the analysis in Appendix 4, we have the long time limit

\[
\lim_{t \to \infty} \frac{\ln Q_X}{t} = \lambda_X^{\max}
\]

(4.100)

where \( \lambda_X^{\max} \) is the largest eigenvalue of the matrix \( M_X(\xi^{\alpha \beta}) \) defined in Eq. (4.219), for \( X = (A, C, G, T/U) \).

Since the action in Eq. (4.98) is extensive in the sequence length \( N \), a saddle point calculation provides and exact expression for the mean fitness, in the limit \( N \to \infty \),

\[
f_m = \max_{\{\xi^{\alpha \beta}, \tilde{\xi}^{\alpha \beta}\}} \left[ f(\tilde{\xi}^{\alpha \beta}) - \sum_{1 \leq \alpha < \beta \leq 4} \xi^{\alpha \beta} \tilde{\xi}^{\alpha \beta} - \sum_{1 \leq \alpha < \beta \leq 4} \mu_{\alpha \beta} + \sum_{X=(A,C,G,T/U)} y_X \lambda_X^{\max} \right]
\]

(4.101)

We remark that Eq. (4.101) represents an exact analytical expression for the steady-state mean fitness of the population, in the limit of very long sequences \( N \to \infty \), for an arbitrary microscopic fitness function \( f(\tilde{u}) \) and arbitrary mutation scheme as represented by Fig. 4.5.
4.3 The four-state Eigen model

The Eigen model conceptually differs from the parallel or Kimura model because it is assumed that mutations arise as a consequence of errors in the replication process. For an alphabet of size \( h \), the system of equations which describes the time evolution of the probabilities \( p_i \), with \( 1 \leq i \leq h^N \), is

\[
\frac{dp_i}{dt} = \sum_{j=1}^{N} [B_{ij} r_j - \delta_{ij} D_j] p_j - p_i \left[ \sum_{j=1}^{N} (r_j - D_j) p_j \right]
\]  
(4.102)

Here, \( r_i \) is the replication rate of sequence \( S_i \), and the components of the matrix

\[
B_{ij} = (1 - q)^{N-d_{ij}} \left( \frac{q}{h - 1} \right)^{d_{ij}}
\]  
(4.103)

represent the transition rates from sequence \( S_j \) into \( S_i \), where \( 1 - q \) is the probability to copy a nucleotide without error, and \( q \) is the probability per site for a base substitution during the replication process.

We consider a mutation scheme in the replication process analogous to the Kimura 3 ST scheme. Let us assume that for each nucleotide \( X = \{A,C,G,T/U\} \), the probability of making an exact copy is given by \( 1 - q \), and the probabilities for the three possible mistakes which increase the Hamming distance along its three different components are given by \( q_1, q_2, q_3 \), where \( q_1 + q_2 + q_3 = q \).

\[
\begin{align*}
\uparrow q_1 \\
X &\rightarrow q_2 \\
\downarrow q_3
\end{align*}
\]
We thus obtain the natural generalization

\[ (1 - q)^{N - d_{ij}} \left( \frac{q}{h - 1} \right)^{d_{ij}} \to (1 - q)^{N - (d_{1} + d_{2} + d_{3})} q_{1}^{d_{1}} q_{2}^{d_{2}} q_{3}^{d_{3}} \]  \hspace{1cm} (4.104)

### 4.3.1 The four-state Eigen model in operator form

By similar arguments as in the parallel model, we formulate a Hamiltonian operator for the Eigen model

\[ -\hat{H} = \prod_{j=1}^{N} \left[ (1 - q)\hat{T}^{00}(j) + q_{1}\hat{T}^{01}(j) + q_{2}\hat{T}^{10}(j) + q_{3}\hat{T}^{11}(j) \right] \]

\[ \times N f \left[ \frac{1}{N} \sum_{l=1}^{N} \tilde{a}^{\dagger}(l) \Theta_{l}^{\nu} \tilde{a}(l) \right] - N d \left[ \frac{1}{N} \sum_{l=1}^{N} \tilde{a}^{\dagger}(l) \Theta_{l}^{\nu} \tilde{a}(l) \right] \]

\hspace{1cm} (4.105)

Let us define the coefficients \( \mu_{1} = Nq_{1}, \mu_{2} = Nq_{2} \) and \( \mu_{3} = Nq_{3} \). Then, we have

\[ -N \ln(1 - q) \approx Nq = N(q_{1} + q_{2} + q_{3}) = \mu_{1} + \mu_{2} + \mu_{3} \]  \hspace{1cm} (4.106)

The Hamiltonian operator Eq. (4.105) is expressed, to \( \mathcal{O}(\infty/N) \), by

\[ -\hat{H} = N e^{-(\mu_{1} + \mu_{2} + \mu_{3})} e^{\sum_{j=1}^{N} \left[ \frac{\mu_{1}}{N} \hat{T}^{10}(j) + \frac{\mu_{2}}{N} \hat{T}^{01}(j) + \frac{\mu_{3}}{N} \hat{T}^{11}(j) \right]}

\[ \times f \left[ \frac{1}{N} \sum_{l=1}^{N} \tilde{a}^{\dagger}(l) \Theta_{l}^{\nu} \tilde{a}(l) \right] - N d \left[ \frac{1}{N} \sum_{l=1}^{N} \tilde{a}^{\dagger}(l) \Theta_{l}^{\nu} \tilde{a}(l) \right] \]

\hspace{1cm} (4.107)

To study the equilibrium properties of the system, as in the case of the parallel model, we calculate the partition function by performing a Trotter factorization

\[ Z = \text{Tr} e^{-\hat{H} \hat{P}} \]

\[ = \int_{0}^{2\pi} \left[ \prod_{j=1}^{N} \frac{d\lambda_{j}}{2\pi} \right] e^{-i\lambda_{j}} \lim_{M \to \infty} \int \left[ \prod_{k=1}^{M} \prod_{j=1}^{N} \frac{d^{4} z_{k}(j) d^{4} z_{k}(j)}{\pi^{4}} \right] e^{-S[\tau, \pi]} \]

\hspace{1cm} (4.108)
Here,

\[ e^{-S[\varphi, \bar{\varphi}]} = \prod_{k=1}^{M} \langle \{ \tilde{z}_k \} | e^{-\epsilon \bar{H}} | \{ \tilde{z}_{k-1} \} \rangle \]

(4.109)

where the matrix elements in the coherent states basis are given by the expression

\[ \langle \{ \tilde{z}_k \} | e^{-\epsilon \bar{H}} | \{ \tilde{z}_{k-1} \} \rangle = e^{-\frac{1}{2} \sum_{j=1}^{N} (\tilde{z}_k(j) \tilde{z}_k(j) - 2\tilde{z}_k(j) \tilde{z}_k_{-1}(j) + \tilde{z}_k_{-1}(j) \tilde{z}_k_{-1}(j))} \times \exp \left( \epsilon N e^{-(\mu_1 + \mu_2 + \mu_3)} e^{\sum_{j=1}^{N} \tilde{z}_k(j) \left( \mu_1 \sigma^1 \otimes \sigma^0 + \mu_2 \sigma^0 \otimes \sigma^1 + \mu_3 \sigma^1 \otimes \sigma^1 \right) \tilde{z}_k_{-1}(j)} \times f \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) \Theta^w_j \tilde{z}_k_{-1}(j) \right] - \epsilon N d \left[ \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) \Theta^w_j \tilde{z}_k_{-1}(j) \right] \right) \]

(4.110)

Let us introduce the 3-component vector field \( \vec{\xi}_k = (\xi_k^1, \xi_k^2, \xi_k^3) \), and an integral representation of the corresponding delta function

\[ 1 = \int D[\vec{\xi}] \prod_{k=1}^{M} \delta^{(3)} \left[ \vec{\xi}_k - \frac{1}{N} \sum_{j=1}^{N} \tilde{z}_k(j) \Theta^w_j \tilde{z}_k_{-1}(j) \right] \]

\[ = \int \left[ \prod_{k=1}^{M} \prod_{i=1}^{3} \frac{i \epsilon N d \xi_k^i d \bar{\xi}_k^i}{2\pi} \right] e^{-\epsilon N \sum_{k=1}^{M} \xi_k \cdot \xi_k + \epsilon \sum_{k=1}^{M} \sum_{j=1}^{N} \tilde{z}_k(j) \tilde{z}_k_{-1}(j) \xi_k \cdot \Theta^w_j \tilde{z}_k_{-1}(j)} \]

(4.111)
Similarly, let us introduce a second, 3-component spin field \( \bar{\eta}_k = (\eta_k^1, \eta_k^2, \eta_k^3) \),

\[
1 = \int D[\bar{\eta}] \prod_{k=1}^{M} \delta \left[ \eta_k^1 - \frac{1}{N} \sum_{j=1}^{N} \bar{z}_k^1(j) \sigma^1 \otimes \sigma^0 \bar{z}_{k-1}(j) \right] \\
\times \delta \left[ \eta_k^2 - \frac{1}{N} \sum_{j=1}^{N} \bar{z}_k^2(j) \sigma^0 \otimes \sigma^1 \bar{z}_{k-1}(j) \right] \times \delta \left[ \eta_k^3 - \frac{1}{N} \sum_{j=1}^{N} \bar{z}_k^3(j) \sigma^1 \otimes \sigma^1 \bar{z}_{k-1}(j) \right] \\
= \int \left[ \prod_{k=1}^{M} \prod_{i=1}^{3} \frac{e^{N d\bar{\eta}_i^k d\bar{\eta}_i^k}}{2\pi} \right] \\
x e^{-e\sum_{k=1}^{M} (\bar{z}_k^i - \eta_k^i) + \epsilon \sum_{k=1}^{M} \sum_{j=1}^{N} \bar{z}_k^i(j) (\bar{z}_k^i \sigma^0 \otimes \sigma^0 + \eta_k^i \sigma^0 \otimes \sigma^1 + \eta_k^i \sigma^1 \otimes \sigma^1) \bar{z}_{k-1}(j)}
\]

(4.112)

Inserting both constraints Eq. (4.111) and Eq. (4.112) in the expression for the trace Eq. (4.108), we obtain

\[
Z = \lim_{M \to \infty} \int D[\xi] D[\bar{\xi}] D[\eta] D[\bar{\eta}] \\
x e^{\epsilon \sum_{k=1}^{M} (-\bar{\xi}_k \cdot \bar{\eta}_k + \epsilon \sum_{m=1}^{(\mu_1 + \mu_2 + \mu_3)} \epsilon \sum_{j=1}^{N} \bar{z}_k(j) \sigma^1 \otimes \sigma^1 \bar{z}_{k-1}(j))} \\
\times \int D[\bar{z}] D[\hat{z}] D[\lambda] \prod_{j=1}^{N} e^{-i\lambda_j} e^{\epsilon \sum_{k=1}^{M} \bar{z}_k(j) S_{kl}(j)} \left| \{\bar{z}_0\} = \{e^{i\lambda_j} \hat{z}_M\} \right|
\]

(4.113)

After performing the Gaussian integral over the fields \( \bar{z}, \hat{z} \), we obtain

\[
Z = \lim_{M \to \infty} \int D[\xi] D[\bar{\xi}] D[\eta] D[\bar{\eta}] \\
x e^{\epsilon \sum_{k=1}^{M} (-\bar{\xi}_k \cdot \bar{\eta}_k + \epsilon \sum_{m=1}^{(\mu_1 + \mu_2 + \mu_3)} \epsilon \sum_{j=1}^{N} \bar{z}_k(j) \sigma^1 \otimes \sigma^1 \bar{z}_{k-1}(j))} \\
\times D[\lambda] \prod_{j=1}^{N} e^{-i\lambda_j} [\text{det} S(j)]^{-1}
\]

(4.114)
The matrix $S(j)$ has the structure

$$
S(j) = \begin{pmatrix}
I & 0 & 0 & \ldots & e^{-\imath \lambda_j A_1(j)} \\
-A_2(j) & I & 0 & \ldots & 0 \\
0 & -A_3(j) & I & \ldots & 0 \\
& & & \ddots & \vdots \\
0 & \ldots & -A_M(j) & I \\
\end{pmatrix}
$$

(4.115)

where $A_k(j) = I + \epsilon[\bar{\eta}_1 \sigma^1 \otimes \sigma^0 + \bar{\eta}_2 \sigma^0 \otimes \sigma^1 + \bar{\eta}_3 \sigma^1 \otimes \sigma^1 + \bar{\xi}_k \cdot \Theta_j^w]$. We obtain

$$
det S(j) = det \left[ I - e^{i \lambda_j} \prod_{k=1}^{M} A_k(j) \right]
$$

$$
= \det \left[ I - e^{i \lambda_j} \mathcal{P} e^{\sum_{k=1}^{M} [\bar{\eta}_1 (\sigma^1 \otimes \sigma^0) + \bar{\eta}_2 (\sigma^0 \otimes \sigma^1) + \bar{\eta}_3 (\sigma^1 \otimes \sigma^1) + \bar{\xi}_k \cdot \Theta_j^w]} \right]
$$

$$
= e^{\text{Tr ln} \left[ I - e^{i \lambda_j} \mathcal{P} e^{\sum_{k=1}^{M} [\bar{\eta}_1 (\sigma^1 \otimes \sigma^0) + \bar{\eta}_2 (\sigma^0 \otimes \sigma^1) + \bar{\eta}_3 (\sigma^1 \otimes \sigma^1) + \bar{\xi}_k \cdot \Theta_j^w]} \right]}
$$

(4.116)

Substituting this last expression into the functional integral Eq. (4.114), and then performing the integrals over the $\lambda$ fields, we obtain

$$
Z = \lim_{M \to \infty} \int \mathcal{D}[\bar{\xi}] \mathcal{D}[\xi] \mathcal{D}[\bar{\eta}] \mathcal{D}[\eta]
$$

$$
\times e^{\mathcal{N} \sum_{k=1}^{M} \left[ -\xi_k \bar{\xi}_k \eta_k + e^{-(\mu_1 + \mu_2 + \mu_3)} \epsilon_1 \eta_1^2 + \mu_2 \eta_2^2 + \mu_3 \eta_3^2 \mathcal{f}[\xi_k] - d[\bar{\xi}_k] \right]} \prod_{j=1}^{N} Q(j)
$$

(4.117)
Here,

\[ Q(j) = \lim_{M \to \infty} \text{Tr} \hat{T} \prod_{k=1}^{M} \left[ (I + e(\bar{\eta}_k^0 \sigma^0 \otimes \sigma^0 + \bar{\eta}_k^1 \sigma^0 \otimes \sigma^1 + \bar{\eta}_k^2 \sigma^1 \otimes \sigma^1) + \bar{\xi}_k \cdot \Theta^j \right] \]

\[ = \lim_{M \to \infty} \text{Tr} \hat{T} \mathcal{E} \sum_{k=1}^{M} [\eta_k^0 \sigma^0 \otimes \sigma^0 + \eta_k^1 \sigma^0 \otimes \sigma^1 + \eta_k^2 \sigma^1 \otimes \sigma^1 + \xi_k \cdot \Theta^j] \]

\[ = \text{Tr} \hat{T} \mathcal{E} \int \mathcal{D} \mathcal{E} \mathcal{D}[\tilde{\eta}] \mathcal{D} \mathcal{D} e^{-S[\tilde{\xi}, \tilde{\eta}, \tilde{\eta}]} \quad (4.118) \]

After taking the limit \( M \to \infty \), we obtain

\[ Z = \int \mathcal{D}[\tilde{\xi}] \mathcal{D}[\tilde{\eta}] \mathcal{D}[\tilde{\eta}] \mathcal{D}[\tilde{\eta}] e^{-S[\tilde{\xi}, \tilde{\eta}, \tilde{\eta}]} \quad (4.119) \]

Here,

\[ S[\tilde{\xi}, \tilde{\eta}, \tilde{\eta}] = -N \int_0^t dt' \left\{ -\tilde{\xi}(t') \cdot \tilde{\xi}(t') - \tilde{\eta}(t') \cdot \tilde{\eta}(t') \right\} \]

\[ + e^{-(\mu_1 + \mu_2 + \mu_3) \mathcal{E} \eta_1(t') + \mu_2 \eta_2(t') + \mu_3 \eta_3(t')} \mathcal{E} \eta(t') - \mathcal{D}[\tilde{\xi}(t')] - \sum_{j=1}^{N} \ln Q(j) \]

\[ (4.120) \]

From the same arguments as in the parallel model, and after Appendix 3, we have

\[ \sum_{j=1}^{N} \ln Q(j) = N \ln Q \quad (4.121) \]

where

\[ Q = \text{Tr} \hat{T} \mathcal{E} \int \mathcal{D} \mathcal{E} \mathcal{D}[\tilde{\eta}] \mathcal{D}[\tilde{\eta}] (t') \sigma^0 \otimes \sigma^0 + \tilde{\eta}_1(t') \sigma^0 \otimes \sigma^1 + \tilde{\eta}_2(t') \sigma^1 \otimes \sigma^1 + \tilde{\xi}(t') \sigma^3 \otimes \sigma^3 + \xi(t') \sigma^3 \otimes \sigma^3) \]

\[ (4.122) \]
With this last simplification, the effective action becomes

\[ S[\bar{\xi}, \bar{\zeta}, \bar{\eta}, \bar{\eta}] = -N \int_0^t dt' (-\bar{\xi}(t') : \bar{\xi}(t') - \bar{\eta}(t') : \bar{\eta}(t') + e^{-(\mu_1 + \mu_2 + \mu_3) \bar{\eta}^3(t') + \mu_3 \bar{\eta}^3(t')} f[\bar{\xi}(t')] - d[\bar{\xi}(t')]) - N \ln Q \]

(4.123)

After the definition Eq. (4.122), and the analysis presented in Appendix 3, we have that for the long time limit,

\[ \lim_{t \to \infty} \frac{\ln Q}{t} = \lambda_{\text{max}} \]

(4.124)

with \( \lambda_{\text{max}} \) the largest eigenvalue of the matrix

\[ M(\bar{\xi}_c, \bar{\eta}_c) = \begin{pmatrix}
(\bar{\xi}_c^1 + \bar{\xi}_c^2 + \bar{\xi}_c^3) & \bar{\eta}_c^1 & \bar{\eta}_c^2 & \bar{\eta}_c^3 \\
\bar{\eta}_c^1 & (\bar{\xi}_c^1 + \bar{\xi}_c^2 - \bar{\xi}_c^3) & \bar{\eta}_c^3 & \bar{\eta}_c^2 \\
\bar{\eta}_c^2 & \bar{\eta}_c^3 & (\bar{\xi}_c^1 - \bar{\xi}_c^2 - \bar{\xi}_c^3) & \bar{\eta}_c^1 \\
\bar{\eta}_c^3 & \bar{\eta}_c^2 & \bar{\eta}_c^1 & (\bar{\xi}_c^1 - \bar{\xi}_c^2 + \bar{\xi}_c^3)
\end{pmatrix} \]

(4.125)

4.3.2 The large N limit of the four-state Eigen model is a saddle point

By assuming that the sequence length N is very large, \( N \to \infty \), we can evaluate the functional integral Eq. (4.119) by a saddle point method. Considering the action
defined in Eq. (4.123), we have

\[
\frac{\delta S}{\delta \xi^i} \bigg|_{\xi_0, \xi_0, \eta_c, \eta_c} = N \left( \xi^i_c - e^{-(\mu_1 + \mu_2 + \mu_3)} e^{\mu_1 \eta_c^1 + \mu_2 \eta_c^2 + \mu_3 \eta_c^3} \frac{\partial f[\xi]}{\partial \xi^i} \right) _c + \frac{\partial d[\xi_c]}{\partial \xi^i} \bigg|_c = 0
\]

\[
\frac{\delta S}{\delta \eta^i} \bigg|_{\xi_0, \xi_0, \eta_c, \eta_c} = N \left( \eta^i_c - \mu_i e^{-(\mu_1 + \mu_2 + \mu_3)} e^{\mu_1 \eta_c^1 + \mu_2 \eta_c^2 + \mu_3 \eta_c^3} f[\xi_c] \right) = 0
\]

\[
\frac{\delta S}{\delta \bar{\xi}^i} \bigg|_{\xi_0, \xi_0, \eta_c, \eta_c} = N \left( \bar{\xi}^i_c - \frac{1}{Q} \frac{\partial Q}{\partial \xi^i} \right) = 0
\]

\[
\frac{\delta S}{\delta \bar{\eta}^i} \bigg|_{\xi_0, \xi_0, \eta_c, \eta_c} = N \left( \bar{\eta}^i_c - \frac{1}{Q} \frac{\partial Q}{\partial \eta^i} \right) = 0
\]

(4.126)

We have therefore the system of equations

\[
\xi^i_c = e^{-(\mu_1 + \mu_2 + \mu_3)} e^{\mu_1 \eta^1_c + \mu_2 \eta^2_c + \mu_3 \eta^3_c} \frac{\partial f[\xi]}{\partial \xi^i} \bigg|_{c} - \frac{\partial d[\xi_c]}{\partial \xi^i} \bigg|_{c} \quad (4.127)
\]

\[
\eta^i_c = \mu_i e^{-(\mu_1 + \mu_2 + \mu_3)} e^{\mu_1 \eta^1_c + \mu_2 \eta^2_c + \mu_3 \eta^3_c} f[\xi_c] \quad (4.128)
\]

\[
\xi^1_c = \langle \sigma^3 \otimes \sigma^0 \rangle, \quad \xi^2_c = \langle \sigma^0 \otimes \sigma^3 \rangle, \quad \xi^3_c = \langle \sigma^3 \otimes \sigma^3 \rangle \quad (4.129)
\]

\[
\eta^1_c = \langle \sigma^1 \otimes \sigma^0 \rangle, \quad \eta^2_c = \langle \sigma^0 \otimes \sigma^1 \rangle, \quad \eta^3_c = \langle \sigma^1 \otimes \sigma^1 \rangle \quad (4.130)
\]

where we defined

\[
\langle (\cdot) \rangle = \frac{\int_{t_0} \mathcal{T} \mathcal{T} \mathcal{T} \mathcal{T} (\cdot) e^{\int_{t_0} dt' [\eta_1 \sigma^1 \otimes \sigma^0 + \eta_2 \sigma^0 \otimes \sigma^1 + \eta_3 \sigma^1 \otimes \sigma^1 + \xi_1 \sigma^3 \otimes \sigma^0 + \xi_2 \sigma^0 \otimes \sigma^3 + \xi_3 \sigma^3 \otimes \sigma^3]}}{\int_{t_0} \mathcal{T} \mathcal{T} \mathcal{T} \mathcal{T} e^{\int_{t_0} dt' [\eta_1 \sigma^1 \otimes \sigma^0 + \eta_2 \sigma^0 \otimes \sigma^1 + \eta_3 \sigma^1 \otimes \sigma^1 + \xi_1 \sigma^3 \otimes \sigma^0 + \xi_2 \sigma^0 \otimes \sigma^3 + \xi_3 \sigma^3 \otimes \sigma^3]}}
\]

(4.131)

After the saddle-point analysis, we obtain an exact analytical expression for the mean fitness $f_m$ of the population, in the limit of very large sequences $N \to \infty$, for
an arbitrary microscopic fitness function \( f(\bar{u}) \) and degradation rate \( d(\bar{u}) \)

\[
f_m = \lim_{N,t \to \infty} \frac{-S_c}{Nt} = \max_{\{\xi_c, \eta_c, \tilde{\eta}_c\}} \left[ e^{-\left(\mu_1 + \mu_2 + \mu_3\right) t} e^{\mu_1 \eta_1^2 + \mu_2 \eta_2^2 + \mu_3 \eta_3^2} f(\xi_c) - d(\xi_c) - \xi_c \cdot \tilde{\xi}_c - \eta_c \cdot \tilde{\eta}_c + \lambda_{\text{max}} \right]
\]

(4.132)

The average composition of the population \( u \) is obtained after Eq. (4.132) by the self-consistent condition \( f_m = f(\bar{u}) \).

### 4.3.3 Analytical results for the symmetric mutation scheme

If the mutation rates are identical \( \mu_1 = \mu_2 = \mu_3 = \mu \), then we have the symmetric case \( \xi_1 = \xi_2 = \xi_3 = \xi_c, \ \xi_1 = \xi_2 = \xi_3 = \xi_c \) and \( \eta_1 = \eta_2 = \eta_3 = \eta_c, \ \tilde{\eta}_1 = \tilde{\eta}_2 = \tilde{\eta}_3 = \tilde{\eta}_c \).

As shown in Appendix 3, the trace in Eq. (96) becomes

\[
Q = 2e^{-(\xi_c + \eta_c) t} + 2e^{(\xi_c + \eta_c) t} \cosh \left( 2t \sqrt{(\xi_c)^2 - (\eta_c)^2} \right)
\]

(4.133)

In the long time limit, this expression becomes

\[
\lim_{t \to \infty} \frac{\ln Q}{t} = \lambda_{\text{max}} = \xi_c + \eta_c + 2\sqrt{(\xi_c)^2 - (\eta_c)^2}
\]

(4.134)

The three identities in Eq. (4.129) can be added to obtain

\[
3\xi_c = \langle \sigma^3 \otimes \sigma^0 + \sigma^0 \otimes \sigma^3 + \sigma^3 \otimes \sigma^3 \rangle = \frac{\partial}{\partial \xi_c} \lambda_{\text{max}}
\]

(4.135)

Similarly, by adding the three identities in Eq. (4.130), we obtain

\[
3\eta_c = \langle \sigma^1 \otimes \sigma^0 + \sigma^0 \otimes \sigma^1 + \sigma^1 \otimes \sigma^1 \rangle = \frac{\partial}{\partial \eta_c} \lambda_{\text{max}}
\]

(4.136)
After Eq. (4.136),
\[ \eta_c = \frac{1}{3} \left( 1 + \frac{2\eta_c - \xi_c}{\sqrt{(\xi_c)^2 - \eta_c \xi_c + (\eta_c)^2}} \right) \] (4.137)

and after Eq. (4.135)
\[ \xi_c = \frac{1}{3} \left( 1 + \frac{2\xi_c - \eta_c}{\sqrt{(\xi_c)^2 - \eta_c \xi_c + (\eta_c)^2}} \right) \] (4.138)

By elementary algebraic manipulations combining equations (4.137), (4.138), we obtain the identities

\[ 3(\eta_c \eta_c + \xi_c \xi_c) = \eta_c + \xi_c + 2\sqrt{(\xi_c)^2 - \eta_c \xi_c + (\eta_c)^2} = \lambda_{\text{max}} \] (4.139)

and

\[ (\eta_c)^2 + \xi_c \eta_c + (\xi_c)^2 - \xi_c - \eta_c = 0 \] (4.140)

This last equation allows us to eliminate \( \eta_c \) in favor of \( \xi_c \)
\[ \eta_c = \frac{1 - \xi_c \pm \sqrt{(1 + 3\xi_c)(1 - \xi_c)}}{2} \] (4.141)

Substituting in Eq. (4.132), we obtain that [119]
\[ \lim_{N,t \to \infty} \frac{-S_c}{Nt} = f_m = \max_{-\frac{1}{3} \leq \xi_c \leq 1} \left\{ e^{-3\mu} e^{\frac{3}{2}\mu(1 - \xi_c + \sqrt{(1 + 3\xi_c)^2})} f[\xi_c] - d[\xi_c] \right\} \] (4.142)

**The sharp peak fitness landscape**

Let us first consider the sharp peak landscape \( f(u) = (A - A_0)\delta_{u,1} + A_0 \), with \( A > A_0 \). That is, the replication rate is \( f(u = 1) = A \) for sequences identical to the
wild type, and \( f(u \neq 1) = A_0 \), for all other sequences. With zero degradation rate, 
\( d = 0 \), we notice that this result implies: \( \xi_c = 1 \) if \( A > A_0e^{3\mu} \), or \( \xi_c = 0 \) otherwise. 
Therefore, we obtain the mean replication rate

\[
    f_m = \begin{cases} 
        e^{-3\mu} A, & A > A_0e^{3\mu} \\
        A_0, & A < A_0e^{3\mu} 
    \end{cases} \tag{4.143}
\]

The system experiences a phase transition which is first order in \( A \). The steady-state probability for the wild-type is obtained from the self-consistent condition: 
\( f_m = Ap_w + A_0(1 - p_w) \),

\[
    p_w = \begin{cases} 
        e^{-3\mu} A - A_0 \over A - A_0, & A > A_0e^{3\mu} \\
        0, & A < A_0e^{3\mu} 
    \end{cases} \tag{4.144}
\]

Notice that the error threshold is reached at the critical value \( A_{\text{crit}} = A_0e^{3\mu} \), as follows from Eqs. (4.143), (4.144) and as displayed in Fig. 4.6. We notice that this result differs from the analytical value obtained for the binary alphabet [29], \( A_{\text{crit}}^{\text{binary}} = A_0e^\mu \).

The additional factor of three which is explicit in the exponent is clearly a consequence of the existence of three mutation channels into which evolving sequences can escape from the wild-type. This effect, which is purely entropic- and not fitness-like, is an explicit consequence of the larger alphabet size.
The Fujiyama fitness landscape

We will consider the Fujiyama fitness landscape, which is a linear function of the composition

\[ f(\bar{\nu}) = \sum_{i=1}^{3} (\alpha_i u_i) + \alpha_0 \]  

(4.145)

For the symmetric case, \( \alpha_i = \alpha \), \( \mu_i = \mu \). Therefore, we have \( \xi_1 = \xi_2 = \xi_3 \equiv \xi_c \). The mean fitness, in the absence of degradation, from Eq. (4.142) becomes

\[ f_m = \max_{-\frac{3}{2} \leq \xi_c \leq 1} \left\{ (3\alpha \xi_c + \alpha_0) e^{-3\mu \xi_c^2 \mu (1-\xi_c) \sqrt{1+3\xi_c} (1-\xi_c)} \right\} \]

(4.146)

By maximizing with respect to \( \xi_c \), \( \frac{\partial f_m}{\partial \xi_c} = 0 \), we obtain the nonlinear equation

\[ \alpha - \frac{\mu}{2} \alpha_0 - \frac{3}{2} \mu \alpha \xi_c = \frac{\mu (3\alpha \xi_c + \alpha_0)(3\xi_c - 1)}{2 \sqrt{(1 + 3\xi_c)(1 - \xi_c)}} \]

(4.147)
No error threshold is observed for this fitness landscape, except for the trivial limit \( \alpha \to 0, \alpha_0 \geq 0 \). The average surplus \( u \) is obtained by the self-consistent equation

\[ f_m = 3\alpha u + \alpha_0 \]  \hspace{1cm} (4.148)

### 4.3.4 The quadratic fitness landscape

Next we consider the quadratic fitness landscape

\[ f(\vec{u}) = \sum_{i=1}^{N} \left( \frac{\beta_i}{2} u_i^2 + \alpha_i u_i \right) + 1 \]  \hspace{1cm} (4.149)

For the symmetric case, \( \beta_i = \beta, \alpha_i = \alpha, \mu_i = \mu \), we have \( \xi_c^f = \xi_c \) and \( \bar{x}_c^f = \bar{x}_c \).

Thus, the mean fitness, for a null degradation rate, after Eq. (4.142) is

\[ f_m = \max_{-1/2 \leq \xi_c \leq 1} \left\{ \left( \frac{3}{2} \beta \xi_c^2 + 3\alpha \xi_c + 1 \right) e^{-3\mu \xi_c^2 \mu (1-\xi_c+\sqrt{(1+3\xi_c)(1-\xi_c)})} \right\} \]  \hspace{1cm} (4.150)

We maximize with respect to \( \xi_c \), \( \frac{\partial f_m}{\partial \xi_c} = 0 \), to obtain

\[ \beta \xi_c + \alpha = \frac{\mu}{2} \left( \frac{3}{2} \beta \xi_c^2 + 3\alpha \xi_c + 1 \right) \left[ 1 + \frac{3\xi_c - 1}{\sqrt{(1+3\xi_c)(1-\xi_c)}} \right] \]  \hspace{1cm} (4.151)

The average base composition \( u \) is obtained from the self-consistent condition

\[ f_m = f(u) = \frac{3}{2} \beta u^2 + 3\alpha u + 1 \]  \hspace{1cm} (4.152)

The selected phase, \( \xi_c > 0, u > 0 \), occurs for \( \beta > 1.8096\mu \) when \( \alpha = 0 \). The value of \( u \) is continuous at the transition, although \( \xi_c \) jumps from 0 to 0.2289\( \mu \). By expanding Eq. (4.150) near the critical point, we find a discontinuous jump in \( df_m/d\beta \) from 0 to 0.066213. Therefore, the phase transition is of first order in \( \beta \). A graphical representation is displayed in Fig. 4.7.
The average composition $u$ and magnetization $\xi_c$ are represented as a function of the parameter $\beta/\mu$ for the quadratic fitness, when $\alpha = 0$.

### 4.3.5 The quartic fitness landscape

As a final example, we consider the quartic fitness landscape,

$$f(\bar{u}) = \sum_{i=1}^{3} \frac{\beta_i}{4} u_i^4 + 1$$

(4.153)

We further consider the symmetric case $\mu_i \equiv \mu$, $\beta_i \equiv \beta$, and hence $\xi_{c,i} \equiv \xi_c$. From the general expression Eq. (4.142), we obtain an analytical expression for the mean fitness

$$f_m = \max_{\{-1/3 \leq \xi_c \leq 1\}} \left\{ \left( \frac{3}{4} \beta \xi_c^4 + 1 \right) e^{-3\mu + \frac{3}{2} \mu (1-\xi_c + \sqrt{(1+3\xi_c)(1-\xi_c)})} \right\}$$

(4.154)

The average composition of the population $u$ is obtained from the self-consistent condition

$$f(u) = \frac{3}{4} \beta u^4 + 1 = f_m$$

(4.155)
Figure 4.8  The average composition $u$ and magnetization $\xi_c$ are represented as a function of the parameter $\beta/\mu$ for the quartic fitness landscape.

In Fig. 4.8, we present the values of $u$ and $\xi_c$, as obtained from Eqs. (4.154), (4.155), as a function of the parameter $\beta/\mu$. We notice that a discontinuous jump in the bulk magnetization from $\xi_c = 0$ to $\xi_c = 0.779856$ is observed at $\beta/\mu = 10.776165$. By expanding Eq. (4.154) near the critical point, we find a discontinuous jump in $df_m/d\beta$, from 0 to 0.066213. Therefore, the phase transition is of first order in $\beta$.

The average composition shows a fast but continuous transition. This behavior is much like the one observed in the sharp peak fitness landscape, Eq. (4.143), and in the corresponding example for the parallel model.

4.3.6 Generalization to an arbitrary mutation scheme

In this section, we consider a generalized mutation scheme, according to Fig. 4.5. We consider the mutation probabilities $q_{\alpha\beta}$, and the mutation matrices defined in Eq.
(4.89). We also define a total mutation probability \( q = \sum_{\alpha<\beta} q_{\alpha\beta} \).

By employing the mutation matrices defined in Eq. (4.17), and the matrices in Eq. (4.89), we have that the Hamiltonian is given by

\[
-\hat{H} = \prod_{j=1}^{N} \left[ (1 - q) \hat{I} + \hat{a}^\dagger(j) \left( \sum_{\alpha<\beta} q_{\alpha\beta} \tau^\alpha \right) \hat{a}(j) \right] N f[\tilde{u}] - Nd[\tilde{u}]
\]

(4.156)

Let us define the coefficients \( \mu_{\alpha\beta} = Nq_{\alpha\beta} \), for \( 1 \leq \alpha < \beta \leq 4 \). Then, we have

\[
-N \ln(1 - q) \simeq Nq = N \sum_{\alpha<\beta} q_{\alpha\beta} = \sum_{\alpha<\beta} \mu_{\alpha\beta} \equiv \mu
\]

(4.157)

With this definition, in the large \( N \) limit, the Hamiltonian can be written as

\[
-\hat{H} = e^{-\mu + \frac{1}{N} \sum_{j=1}^{N} \hat{a}^\dagger(j) \left( \sum_{\alpha<\beta} \mu_{\alpha\beta} \right) \hat{a}(j)} N f[\tilde{u}] - Nd[\tilde{u}]
\]

(4.158)

Following a similar procedure as in the Kimura 3 ST case, we express the partition function for the Hamiltonian Eq. (4.158) as a functional integral

\[
Z = \int [D\xi^{\alpha\beta}] [D\tilde{\xi}^{\alpha\beta}] [D\eta^{\alpha\beta}] [D\tilde{\eta}^{\alpha\beta}] e^{-S[\xi^{\alpha\beta}, \tilde{\xi}^{\alpha\beta}, \eta^{\alpha\beta}, \tilde{\eta}^{\alpha\beta}]}
\]

(4.159)

Here, the action is defined as

\[
S = -N \int_0^t dt' \left( e^{-\mu + \sum_{\alpha<\beta} \mu_{\alpha\beta} \eta^{\alpha\beta} f[\xi]} - d[\xi] - \sum_{\alpha<\beta} \left\{ \eta^{\alpha\beta} \eta^{\alpha\beta} + \xi^{\alpha\beta} \tilde{\xi}^{\alpha\beta} \right\} \right)
\]

\[
- \sum_{j=1}^{N} \ln Q(j)
\]

(4.160)
Here, we have

\[ Q(j) = \text{Tr} e^{\int_0^t dt' \sum_{\alpha<\beta}(\eta^{\alpha\beta}r^{\alpha\beta} + \xi^{\alpha\beta}g^w_{\alpha\beta}(j))} \quad (4.161) \]

As explained in Appendix 4, and as previously discussed for the parallel model, since \( n_\alpha^w(j) + n_\beta^w(j) + n_\gamma^w(j) + n_\delta^w(j) = 1 \) for all \( 1 \leq j \leq N \), the action Eq. (4.160) has the explicit extensive form

\[ S = -N \int_0^t dt' \left( e^{-\mu \sum_{\alpha<\beta} \mu_{\alpha\beta} n^{\alpha\beta}} f[\xi] - d[\xi] - \sum_{\alpha<\beta} \{\bar{\eta}^{\alpha\beta} \eta^{\alpha\beta} + \bar{\xi}^{\alpha\beta} \xi^{\alpha\beta}\} \right) \]

\[ -N \sum_{X=A,C,G,T/U} y_X \ln Q_X \quad (4.162) \]

Here, \( y_X = N_X^w/N \) is the fractional composition of the wild-type sequence, and the four different traces are defined by

\[ Q_X = \text{Tr} e^{\int_0^t dt' \sum_{\alpha<\beta}(\eta^{\alpha\beta}r^{\alpha\beta} + \xi^{\alpha\beta}g^w_{\alpha\beta})} \quad (4.163) \]

with the four matrix arrays \( g^w_X \) defined in Eq. (4.97). After the definition Eq. (4.163), and the analysis presented in Appendix 4, we have that in the long time limit

\[ \lim_{t \to \infty} \frac{\ln Q_X}{t} = \lambda^{\infty}_{\text{max}} \quad (4.164) \]

with \( \lambda^{\infty}_{\text{max}} \) the largest eigenvalue of the matrix \( M_X(\bar{\xi}^{\alpha\beta}, \bar{\eta}^{\alpha\beta}) \) defined by Eq. (4.221) for \( X = \{A, C, G, T/U\} \).

As the action defined by Eq. (4.162) is extensive in the sequence length \( N \), a saddle point limit provides an exact expression for the mean fitness of the population, when
\[ N \to \infty \]

\[ f_m = \max_{\xi^\alpha_\beta, \zeta^\alpha_\beta, \eta^\alpha_\beta, \xi^\alpha_\beta} \left[ e^{-\frac{1}{\theta} \sum_{\alpha<\beta} \mu_{\alpha\beta} (1-\eta^\alpha_\beta)} f(\xi) - d[\xi^\alpha_\beta] \right. \]

\[ - \sum_{1 \leq \alpha < \beta \leq 4} \left\{ \eta^\alpha_\beta \xi^\alpha_\beta + \zeta^\alpha_\beta \xi^\alpha_\beta \right\} + \sum_{1 \leq \alpha < \beta \leq 4} y x \lambda_{\max}^x \]  

(4.165)

### 4.4 Conclusion

Using the quantum spin chain approach, the 2-state, purine/pyrimidine assumption for quasispecies theory has been lifted, and exact results have been derived for linear and quadratic fitness cases [113]. We have here expressed the general result for the fitness of the evolved population as a maximization principle. We have derived the solution for a general fitness function using the Schwinger spin coherent states approach. We have presented analytic results for the sharp peak, as well as linear, quadratic, and quartic fitness functions. For the Kimura 3 ST mutation scheme, we have presented an explicit solution for a general fitness function, expressed as a maximization principle. For the general mutation scheme, a maximization was also presented.

We have also derived the general solution to the Eigen model of mutation and selection. We have presented analytic results for the sharp peak, linear, quadratic, and quartic fitness functions. For the Kimura 3 ST mutation scheme, we have presented an explicit solution for a general fitness function, expressed as a maximization principle. For a general mutation scheme, the maximization principle was also presented.
These results bring quasispecies theory closer to the evolutionary dynamics that occurs at the genetic level. The theory could be generalized to include an alphabet of 20 amino acids, rather than 4 bases. The theory could also be generalized to include finite population size effects.

Appendix 1

By considering the definitions of the wild-type, site dependent matrices $\Theta_{j}^{(i,w)}$ in Eq. (4.27), we see that there are only four possible different cases, since $n_{A}^{w}(j) + n_{C}^{w}(j) + n_{T}^{w}(j) + n_{G}^{w}(j) = 1$ for all $1 \leq j \leq N$.

Case 1: $n_{A}^{w}(j) = 1$ and $n_{C}^{w}(j) = n_{G}^{w}(j) = n_{T}^{w}(j) = 0$. We re-define $\Theta_{j}^{(i,w)} \rightarrow \Theta_{A}^{(i,w)}$, ...
and explicitly we find

\[
\Theta_A^{(1,w)} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix} = \begin{pmatrix}
\sigma^3 & 0 \\
0 & \sigma^3
\end{pmatrix} = \sigma^3 \otimes \sigma^0
\]

\[
\Theta_A^{(2,w)} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix} = \begin{pmatrix}
\sigma^0 & 0 \\
0 & -\sigma^0
\end{pmatrix} = \sigma^0 \otimes \sigma^3
\]

\[
\Theta_A^{(3,w)} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} = \begin{pmatrix}
\sigma^3 & 0 \\
0 & -\sigma^3
\end{pmatrix} = \sigma^3 \otimes \sigma^3 \quad (4.166)
\]

Case 2: \( n_C^{w}(j) = 1 \) and \( n_A^{w}(j) = n_B^{w}(j) = n_T^{w}(j) = 0 \). We re-define \( \Theta_j^{(t,w)} \rightarrow \Theta_C^{(t,w)} \),
and explicitly find

\[
\Theta^{(1,w)}_C = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
= \begin{pmatrix}
-\sigma^3 & 0 \\
0 & -\sigma^3
\end{pmatrix}
= -\sigma^3 \otimes \sigma^0
\]

\[
\Theta^{(2,w)}_C = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
= \begin{pmatrix}
\sigma^0 & 0 \\
0 & -\sigma^0
\end{pmatrix}
= \sigma^0 \otimes \sigma^3
\]

\[
\Theta^{(3,w)}_C = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
= \begin{pmatrix}
-\sigma^3 & 0 \\
0 & -\sigma^3
\end{pmatrix}
= -\sigma^3 \otimes \sigma^3
\]  \hspace{1cm} (4.167)

Case 3: \( n_C^n(j) = 1 \) and \( n_A^n(j) = n_C^n(j) = n_F^n(j) = 0 \). We re-define \( \Theta^{(i,w)}_J \rightarrow \Theta^{(i,w)}_G \),
and explicitly find

\[
\Theta_{G}^{(1,w)} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix} = \begin{pmatrix}
\sigma^3 & 0 \\
0 & \sigma^3
\end{pmatrix} = \sigma^3 \otimes \sigma^0
\]

\[
\Theta_{G}^{(2,w)} = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} = \begin{pmatrix}
-\sigma^0 & 0 \\
0 & \sigma^0
\end{pmatrix} = -\sigma^0 \otimes \sigma^3
\]

\[
\Theta_{G}^{(3,w)} = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix} = \begin{pmatrix}
-\sigma^3 & 0 \\
0 & \sigma^3
\end{pmatrix} = -\sigma^3 \otimes \sigma^3
\]

Case 4: \( n_{T}^{w}(j) = 1 \) and \( n_{A}^{w}(j) = n_{G}^{w}(j) = n_{C}^{w}(j) = 0 \). We re-define \( \Theta_{j}^{(i,w)} \rightarrow \Theta_{T}^{(i,w)} \),
and explicitly find

\[
\Theta_T^{1,w} = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} = \begin{pmatrix}
-\sigma^3 & 0 \\
0 & -\sigma^3
\end{pmatrix} = -\sigma^3 \otimes \sigma^0
\]

\[
\Theta_T^{2,w} = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} = \begin{pmatrix}
-\sigma^0 & 0 \\
0 & \sigma^0
\end{pmatrix} = -\sigma^0 \otimes \sigma^3
\]

\[
\Theta_T^{3,w} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} = \begin{pmatrix}
\sigma^3 & 0 \\
0 & -\sigma^3
\end{pmatrix} = \sigma^3 \otimes \sigma^3
\]

To calculate the traces involved in the different models, we need to obtain the eigenvalues \(\{\lambda_i\}_{1 \leq i \leq 4}\) of the matrix

\[
M(\vec{e}_1, \vec{e}_2, \vec{e}_3) = \mu_1 \sigma^1 \otimes \sigma^0 + \mu_2 \sigma^0 \otimes \sigma^1 + \mu_3 \sigma^1 \otimes \sigma^1
\]

\[
+ \xi_1 \sigma^3 \otimes \sigma^0 + \xi_2 \sigma^0 \otimes \sigma^3 + \xi_3 \sigma^3 \otimes \sigma^3
\]

(4.170)
The characteristic equation

$$\det \left| M(\xi_c^1, \xi_c^2, \xi_c^3, \xi_c^4) - \lambda I_{4 \times 4} \right| = 0$$

(4.171)

adopts the explicit form

$$\begin{vmatrix}
(\xi_c^1 + \xi_c^2 + \xi_c^3) - \lambda & \mu_1 & \mu_2 & \mu_3 \\
\mu_1 & (-\xi_c^1 + \xi_c^2 - \xi_c^3) - \lambda & \mu_3 & \mu_2 \\
\mu_2 & \mu_3 & (\xi_c^1 - \xi_c^2 - \xi_c^3) - \lambda & \mu_1 \\
\mu_3 & \mu_2 & \mu_1 & (-\xi_c^1 - \xi_c^2 + \xi_c^3) - \lambda
\end{vmatrix} = 0$$

(4.172)

By inspection, we notice that, by performing standard row-column operations, the characteristic equation is exactly the same for the four matrices

$$M(\xi_c^1, \xi_c^2, \xi_c^3) = M(-\xi_c^1, \xi_c^2, -\xi_c^3) = M(\xi_c^1, -\xi_c^2, -\xi_c^3) = M(-\xi_c^1, -\xi_c^2, \xi_c^3)$$

(4.173)

Therefore, all four matrices are equivalent in the sense that they possess the same eigenvalues. In particular, this implies that the trace of their exponential, has the same value in all four cases, thus implying the equality

$$Q_A = Q_C = Q_G = Q_T = \text{Tr} e^{Mt} = \sum_{i=1}^{4} e^{\lambda_i t} \equiv Q$$

(4.174)

In particular, for the symmetric case $\mu_i \equiv \mu$, and $\xi_c^i \equiv \xi_c$, it is possible to obtain
simple analytical expressions for the eigenvalues:

\[
\begin{align*}
\lambda_1 &= \lambda_2 = -(\xi_e + \mu) \\
\lambda_3 &= (\xi_e + \mu) + 2\sqrt{(\xi_e)^2 - \mu \xi_e + \mu^2} \\
\lambda_4 &= (\xi_e + \mu) - 2\sqrt{(\xi_e)^2 - \mu \xi_e + \mu^2}
\end{align*}
\] (4.175)

Therefore, the expression for the trace in this symmetric case becomes

\[
Q = 2e^{-(\xi_e+\mu)t} + 2e^{(\xi_e+\mu)t} \cosh \left(2t \sqrt{(\xi_e)^2 - \mu \xi_e + \mu^2}\right)
\] (4.176)

**Appendix 2**

By performing elementary algebraic manipulations Eq. (4.78)

\[
3\beta \xi_e + 3\alpha - \frac{3}{2} \tilde{\mu} + \frac{3}{2} \sqrt{1 + 2 \xi_e - 3 \xi_e^2} = 0
\]

can be cast into the standard form of a quartic equation

\[
A\xi_e^4 + B\xi_e^3 + C\xi_e^2 + D\xi_e + E = 0
\] (4.177)

where, by defining \(\tilde{\mu} \equiv \mu / \beta\) and \(\tilde{\alpha} \equiv \alpha / \beta\), the coefficients correspond to

\[
\begin{align*}
A &= 3 \\
B &= 6\tilde{\alpha} - 3\tilde{\mu} - 2 \\
C &= 3\tilde{\mu}^2 - 3\tilde{\alpha}\tilde{\mu} + 3\tilde{\alpha}^2 + 2\tilde{\mu} - 4\tilde{\alpha} - 1 \\
D &= -2\tilde{\alpha} - 2\tilde{\alpha}^2 + \tilde{\mu} + 2\tilde{\alpha}\tilde{\mu} - 2\tilde{\mu}^2 \\
E &= -\tilde{\alpha}^2 + \tilde{\alpha}\tilde{\mu}
\end{align*}
\] (4.178)
We remark that this quartic equation introduces additional, unphysical solutions to the original Eq. (4.78). However, discarding these unphysical solutions whenever appropriate, the quartic Eq. (4.177) allows us to obtain explicit analytical expressions for $\xi_c$ in the entire region of parameters. Following Ferrari's method [120], we define the parameters

$$a_1 = \frac{3B^2}{8A^2} + \frac{C}{A}$$
$$= \frac{1}{2} - \frac{\bar{\alpha}}{3} - \frac{\bar{\alpha}^2}{2} + \frac{\bar{\mu}}{6} + \frac{\bar{\alpha}\bar{\mu}}{2} + \frac{5\bar{\mu}^2}{8} \quad (4.180)$$

$$a_2 = \frac{B^3}{8A^3} - \frac{BC}{2A^2} + \frac{D}{A}$$
$$= -\frac{4}{27} - \frac{4\bar{\alpha}}{9} + \frac{2\bar{\mu}}{9} - \frac{\bar{\mu}^2}{4} - \frac{3\bar{\alpha}\bar{\mu}^2}{4} + \frac{3\bar{\mu}^3}{8} \quad (4.181)$$

$$a_3 = -\frac{3B^4}{256A^4} + \frac{CB^2}{16A^3} - \frac{BD}{4A^2} + \frac{E}{A}$$
$$= -\frac{5}{432} - \frac{7\bar{\alpha}}{108} + \frac{5\bar{\alpha}^2}{72} + \frac{\bar{\alpha}^3}{12} + \frac{\bar{\alpha}^4}{16} + \frac{7\bar{\mu}}{216} + \frac{5\bar{\alpha}\bar{\mu}}{72} - \frac{\bar{\alpha}^2\bar{\mu}}{8} - \frac{\bar{\alpha}^3\bar{\mu}}{8} + \frac{\bar{\mu}^2}{288} + \frac{3\bar{\alpha}\bar{\mu}^2}{16} + \frac{9\bar{\alpha}^2\bar{\mu}^2}{32} - \frac{7\bar{\mu}^3}{96} - \frac{7\bar{\alpha}\bar{\mu}^3}{32} + \frac{13\bar{\mu}^4}{256} \quad (4.182)$$

and solve the depressed quartic equation in the auxiliary variable $z = \xi_c + B/4A$,

$$z^4 + a_1z^2 + a_2z + a_3 = 0 \quad (4.183)$$

We analyze the different cases in the parameter space that defines the possible solutions of this equation.

Case 1: $a_2 = 0$. This situation arises at the critical value

$$\bar{\mu}_c^{(1)} = \frac{2}{3} + 2\bar{\alpha} \quad (4.184)$$
We obtain four possible roots, according to the general formula

\[
\xi_c = -\frac{B}{4A} \pm \sqrt{-\frac{a_1 \pm \sqrt{a_1^2 - 4a_3}}{2}}
\]

\[
= \frac{1}{6} \left( 2 \pm \sqrt{2} \sqrt{1 - 9\alpha(2 + 3\alpha) \pm |1 - 9\alpha(2 + 3\alpha)|} \right)
\]

(4.185)

Depending on the sign of the term in the square root, we have the following solutions

i) If \( 1 - 18\bar{\alpha} - 27\bar{\alpha}^2 > 0 \). This situation occurs when \(-\frac{1}{3} \leq \bar{\alpha} \leq \frac{1}{3} \left( \sqrt{\frac{4}{3}} - 1 \right) \), and the solution is

\[
\xi_{c,\pm} = \frac{1}{3} (1 \pm \sqrt{1 - 18\bar{\alpha} - 27\bar{\alpha}^2}), \quad \xi_c = \frac{1}{3}
\]

(4.186)

ii) If \( 1 - 18\bar{\alpha} - 27\bar{\alpha}^2 \leq 0 \). This situation occurs when \( \bar{\alpha} > \frac{1}{3} \left( \sqrt{\frac{4}{3}} - 1 \right) \).

\[
\xi_c = \frac{1}{3}
\]

(4.187)

We shall consider \( \bar{\alpha} \geq 0 \) in the region of physically meaningful parameters. When \( \bar{\alpha} = 0 \), a non-selective phase is obtained, from Eq. (4.186), if \( \beta < \frac{3}{2} \mu \). At \( \beta = \frac{3}{2} \mu \), for \( \alpha = 0 \), a finite 'jump' in the value of \( \xi_c \) from 0 to 2/3 defines a phase transition, where the value of \( \mu \) varies continuously from 0 to a positive value.

When \( 0 \leq \bar{\alpha} \leq \frac{1}{3} \left( \sqrt{\frac{4}{3}} - 1 \right) \), a finite jump in the bulk magnetization from \( \xi_{c,-} \) to \( \xi_{c,+} \) is observed. This result is in agreement with [113].

Case 2: \( a_3 = 0, \ a_2 \neq 0 \). This situation occurs at the critical values

\[
\tilde{\mu}_c^{(2)} = \frac{2}{39} \left( 1 + 3\bar{\alpha} + 2\sqrt{49 - 18\bar{\alpha} - 27\bar{\alpha}^2} \right), \quad 0 \leq \bar{\alpha} \leq 1.054444
\]

(4.188)
\[ \tilde{\mu}_c^{(3)} = \frac{2}{39} \left( 1 + 3\tilde{\alpha} - 2\sqrt{49 - 18\tilde{\alpha} - 27\tilde{\alpha}^2} \right), \quad 1 \leq \tilde{\alpha} \leq 1.054444 \quad (4.189) \]

In this case, the quartic equation in \( z \) factorizes,

\[ z(z^3 + a_1z + a_2) = 0 \quad (4.190) \]

There is a solution \( z = 0 \) for Eq. (4.190). This is however not a solution of Eq. (4.78), but an artifact of introducing the algebraic transformation into the fourth order polynomial Eq. (4.177).

The solutions corresponding to the remaining cubic equation in Eq. (4.190) are analyzed as follows. Let us define the parameters,

\[
\begin{align*}
\sigma_1 &= \left[ -\frac{a_2}{2} + \left( \frac{a_1^3}{27} + \frac{a_2^2}{4} \right)^{1/2} \right]^{1/3} \\
\sigma_2 &= \left[ -\frac{a_2}{2} - \left( \frac{a_1^3}{27} + \frac{a_2^2}{4} \right)^{1/2} \right]^{1/3}
\end{align*}
\quad (4.191)\]

Then, we have the following cases,

Case 2.a: Consider \( \mu = \tilde{\mu}_c^{(2)} \), defined by Eq. (4.188). This situation is possible when \( 0 \leq \tilde{\alpha} \leq 1.054444 \). Within this range of values for \( \tilde{\alpha} \), the parameter \( \frac{a_1^3}{27} + \frac{a_2^2}{4} \geq 0 \).

Then, we find a single real solution

\[ \xi_c = \frac{1}{39} \left( 7 - 18\tilde{\alpha} + \sqrt{49 - 18\tilde{\alpha} - 27\tilde{\alpha}^2} \right) + s_1 + s_2 \quad (4.192) \]

Case 2.b: Consider \( \tilde{\mu} = \tilde{\mu}_c^{(3)} \), defined by Eq. (4.189). This situation is possible
when $1 \leq \tilde{\alpha} \leq 1.054444$. Within this range of values for $\tilde{\alpha}$, the parameter $\frac{a_1^3}{27} + \frac{a_2^2}{4} \geq 0$.

Then, we find a single real solution

$$\xi_c = \frac{1}{39} \left(7 - 18\tilde{\alpha} - \sqrt{49 - 18\tilde{\alpha} - 27\tilde{\alpha}^2}\right) + s_1 + s_2 \quad (4.193)$$

Case 3: $a_3 \neq 0, a_2 \neq 0$. In this case, we consider again the general quartic Eq. (4.177). Following Ferrari’s method [120], we find 4 possible roots

$$\xi_{c,(1,2)} = -\frac{B}{4A} + \frac{1}{2}P \pm \frac{1}{2}Q$$

$$\xi_{c,(3,4)} = -\frac{B}{4A} - \frac{1}{2}P \pm \frac{1}{2}U \quad (4.194)$$

Here, we defined

$$P = \sqrt{\frac{B^2}{4A^2} - \frac{C}{A} + y_1} \quad (4.195)$$

$$Q = \begin{cases} 
\sqrt{\frac{3B^2}{4A^2} - P^2 - 2\frac{C}{A} + \frac{1}{4}(4\frac{BC}{A^2} - 8\frac{D}{A} - \frac{B^3}{A^3})P^{-1}}, & P \neq 0 \\
\sqrt{\frac{3B^2}{4A^2} - 2\frac{C}{A} + 2\sqrt{y_1^2 - 4\frac{E}{A}}}, & P = 0 
\end{cases} \quad (4.196)$$

$$U = \begin{cases} 
\sqrt{\frac{3B^2}{4A^2} - P^2 - 2\frac{C}{A} - \frac{1}{4}(4\frac{BC}{A^2} - 8\frac{D}{A} - \frac{B^3}{A^3})P^{-1}}, & P \neq 0 \\
\sqrt{\frac{3B^2}{4A^2} - 2\frac{C}{A} - 2\sqrt{y_1^2 - 4\frac{E}{A}}}, & P = 0 
\end{cases} \quad (4.197)$$

From Eq. (4.194), the largest real root corresponds to the physical solution of Eq. (4.78).
The parameter $y_1$ in Eqs. (4.195–4.197) is obtained as the real root of the auxiliary cubic equation

$$y^3 + \gamma_2 y^2 + \gamma_1 y + \gamma_0 = 0 \quad (4.198)$$

Here, we defined the parameters

$$\gamma_2 = \frac{C}{A}$$
$$\gamma_1 = \frac{BD}{A^2} - \frac{E}{A}$$
$$\gamma_0 = \frac{CE}{A^2} - \frac{D^2}{A^2} - \frac{B^2 E}{A^3} \quad (4.199)$$

Let us define

$$q = \frac{\gamma_1}{3} - \frac{\gamma_2}{9}$$
$$r = \frac{1}{6}(\gamma_1 \gamma_2 - 3\gamma_0) - \frac{\gamma_3}{27}$$
$$\Delta = q^3 + r^2 \quad (4.200)$$

We have three possible cases: $\Delta > 0$, $\Delta = 0$, and $\Delta < 0$.

Case 3.a: $\Delta > 0$. In this case, we have one real root $y_1$ for the auxiliary cubic Eq. (4.198), and two complex roots. The real root to be used in Eqs. (4.194–4.197) is given by

$$y_1 = (r + \Delta^{1/2})^{1/3} + (r - \Delta^{1/2})^{1/3} - \frac{\gamma_2}{3} \quad (4.201)$$

Case 3.b: $\Delta = 0$. In this case, all roots of the auxiliary cubic Eq. (4.198) are real,
with two of them identical, and given by
\[
\begin{align*}
y_1 & = 2r^{1/3} - \frac{\gamma_2}{3} \\ y_2 & = y_3 = -r^{1/3}
\end{align*}
\tag{4.202}
\]
In this case, we take the root \( y_1 \) in Eq. (4.202), to be used in the formulas Eqs. (4.194–4.197).

**Case 3.c: \( \Delta < 0 \).** In this case, all three roots of the auxiliary cubic Eq. (4.198) are real and different.

\[
\begin{align*}
y_1 & = 2(r^2 - \Delta)^{1/6} \cos(\theta/3) - \frac{\gamma_2}{3} \\ y_2 & = -2(r^2 - \Delta)^{1/6} \cos(\theta/3 + \pi/3) - \frac{\gamma_2}{3} \\ y_3 & = -2(r^2 - \Delta)^{1/6} \cos(\theta/3 - \pi/3) - \frac{\gamma_2}{3}
\end{align*}
\tag{4.203}
\]
Here, \( \theta = \tan^{-1} \left( \frac{(-\Delta^{1/2})}{r} \right) \). We take the root \( y_1 \) to be used in Eqs. (4.194–4.197).

**Appendix 3**

To calculate the traces involved in the Eigen model, we need to obtain the eigenvalues \( \{\lambda_i\}_{1 \leq i \leq 4} \) of the matrix
\[
M(\xi_c, \xi_c, \xi_c, \eta_c) = \eta_1^1 \sigma^1 \otimes \sigma^0 + \eta_2^2 \sigma^0 \otimes \sigma^1 + \eta_3^3 \sigma^1 \otimes \sigma^1 \\
+ \xi_1^3 \sigma^3 \otimes \sigma^0 + \xi_2^2 \sigma^0 \otimes \sigma^3 + \xi_3^1 \sigma^3 \otimes \sigma^3
\tag{4.204}
\]
The characteristic equation
\[
\det \left| M(\xi_c, \xi_c, \xi_c, \eta_c) - \lambda I_{4 \times 4} \right| = 0
\tag{4.205}
\]
has the explicit form

\[
\begin{vmatrix}
(\xi_1 + \xi_2 + \xi_3) - \lambda & \eta_1^1 & \eta_1^2 & \eta_1^3 \\
\eta_2^1 & (-\xi_1 + \xi_2 - \xi_3) - \lambda & \eta_2^2 & \eta_2^3 \\
\eta_3^1 & \eta_3^2 & (-\xi_1 - \xi_2 - \xi_3) - \lambda & \eta_3^3 \\
\eta_4^1 & \eta_4^2 & \eta_4^3 & (-\xi_1 - \xi_2 + \xi_3) - \lambda
\end{vmatrix} = 0 \tag{4.206}
\]

By inspection, we notice that, by performing standard row-column operations, the characteristic equation is exactly the same for the four matrices

\[
M(\xi_1, \xi_2, \xi_3, \eta_c) = M(-\xi_1, \xi_2, -\xi_3, \eta_c)
\]

\[
= M(\xi_1, -\xi_2, -\xi_3, \eta_c) = M(-\xi_1, -\xi_2, \xi_3, \eta_c)
\]

\[
= M(\xi_1, \xi_2, -\xi_3, \eta_c) = M(-\xi_1, \xi_2, -\xi_3, \eta_c)
\]

\[
(4.207)
\]

Therefore, all four matrices are equivalent in the sense that they possess the same eigenvalues. In particular, this implies that the trace of their exponential has the same value in all four cases,

\[
Q_A = Q_C = Q_G = Q_T = \text{Tr} e^{Mt} = Q \tag{4.208}
\]

In particular, for the symmetric case \( \eta_c^i \equiv \eta_c \), and \( \xi_c^i \equiv \xi_c \), it is possible to obtain simple analytical expressions for the eigenvalues:

\[
\lambda_1 = \lambda_2 = -(\xi_c + \eta_c)t
\]

\[
\lambda_3 = (\xi_c + \eta_c)t + 2t \sqrt{(\xi_c)^2 - \eta_c \xi_c + (\eta_c)^2}
\]

\[
\lambda_4 = (\xi_c + \eta_c)t - 2t \sqrt{(\xi_c)^2 - \eta_c \xi_c + (\eta_c)^2} \tag{4.209}
\]
Therefore, the expression for the trace in this symmetric case becomes

\[ Q = 2e^{-(\xi_c + \eta_c)t} + 2e^{(\xi_c + \eta_c)t} \cosh \left( 2t \sqrt{(\xi_c)^2 - \eta_c \xi_c + (\eta_c)^2} \right) \]  \hspace{1cm} (4.210)

**Appendix 4**

By considering the matrices defined in the generalized mutational scheme Fig. 4.5, Eq. (4.89), since \( n\alpha(j) + n\beta(j) + n\gamma(j) + n\delta(j) = 1 \) for all \( 1 < j < N \), we find that there is only four possible different cases

Case 1: \( n\alpha(j) = 1 \) and \( n\beta(j) = n\gamma(j) = n\delta(j) = 0 \). We re-define \( \Theta_{\alpha\beta}(j) \rightarrow \Theta_{\alpha\beta}^{\prime} \), and explicitly we find
\[ \Theta^w_{A,12} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^0 \end{pmatrix} \]

\[ \Theta^w_{A,13} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^3 \end{pmatrix} \]

\[ \Theta^w_{A,14} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} \sigma^0 & 0 \\ 0 & \sigma^3 \end{pmatrix} \]

\[ \Theta^w_{A,23} = \Theta^w_{A,24} = \Theta^w_{A,34} = \begin{pmatrix} \sigma^0 & 0 \\ 0 & \sigma^0 \end{pmatrix} = I \quad (4.211) \]

**Case 2:** \( n_C^w(j) = 1 \) and \( n_A^w(j) = n_G^w(j) = n_T^w(j) = 0 \). We re-define \( \Theta^w_{a\beta}(j) \rightarrow \Theta^w_{C,a\beta} \),

and explicitly we find
\[
\begin{align*}
\Theta_{C,12}^w &= \begin{pmatrix} -\sigma^3 & 0 \\ 0 & \sigma^0 \end{pmatrix} \\
\Theta_{C,23}^w &= \begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^3 \end{pmatrix} \\
\Theta_{C,24}^w &= \begin{pmatrix} \sigma^0 & 0 \\ 0 & \sigma^3 \end{pmatrix}
\end{align*}
\]

\[
\Theta_{C,13}^w = \Theta_{C,14}^w = \Theta_{C,34}^w = \begin{pmatrix} \sigma^0 & 0 \\ 0 & \sigma^0 \end{pmatrix} = I \tag{4.212}
\]

Case 3: \( n_{\text{C}}^w(j) = 1 \) and \( n_{\text{A}}^w(j) = n_{\text{T}}^w(j) = 0 \). We re-define \( \Theta_{\alpha\beta}^w(j) \rightarrow \Theta_{C,\alpha\beta}^w \), and explicitly we find
\[ \Theta_{G,13}^w = \begin{pmatrix} -\sigma^3 & 0 \\ 0 & \sigma^0 \end{pmatrix} \]
\[ \Theta_{G,23}^w = \begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^0 \end{pmatrix} \]
\[ \Theta_{G,24}^w = \begin{pmatrix} \sigma^0 & 0 \\ 0 & \sigma^3 \end{pmatrix} \]

\[ \Theta_{G,13}^w = \Theta_{G,14}^w = \Theta_{G,34}^w = \begin{pmatrix} \sigma^0 & 0 \\ 0 & \sigma^0 \end{pmatrix} = I \quad (4.213) \]

Case 4: \( n_C^w(j) = 1 \) and \( n_A^w(j) = n_E^w(j) = n_T^w(j) = 0 \). We re-define \( \Theta_{\alpha\beta}^w(j) \rightarrow \Theta_{G,\alpha\beta}^w \), and explicitly we find
We define the fractional composition of the wild type as \( y_X = \frac{N^w_X}{N} \), for \( X = (A, C, G, T/U) \). Then, taking into account the previous analysis, the trace in Eq. (4.95) has the explicit extensive form

\[
\sum_{j=1}^{N} \ln Q(j) = N \sum_{X=A,C,G,T/U} y_X \ln Q_X
\]  

(4.215)

Here, for the parallel or Crow-Kimura model, we define

\[
\ln Q_X = \text{Tr} \hat{T} e^{\int_0^t dt' \sum_{a<\delta} [\eta_{a\delta} \gamma_{a\delta} + \xi_{a\delta} \Theta_{X,a\delta}]} \quad X = (A, C, G, T/U)
\]  

(4.216)

and \( \Theta_{X,a\delta} \) as defined by Eqs. (4.211–4.213).

The corresponding expression for the Eigen model is

\[
\ln Q_X = \text{Tr} \hat{T} e^{\int_0^t dt' \sum_{a<\delta} [\eta_{a\delta} \gamma_{a\delta} + \xi_{a\delta} \Theta_{X,a\delta}]} \quad X = (A, C, G, T/U)
\]  

(4.217)
Here, again, $\Theta_{X,\alpha\beta}^\psi$ are defined by Eqs. (4.211-4.213).

Explicitly, the traces defined in Eq. (4.216) are given by

$$Q_X = \text{Tr} e^{t M_X(\xi_{\alpha\beta})} = \sum_{i=1}^{4} e^{t \lambda_i^X}$$

(4.218)

Here, $\lambda_i^X$ are the eigenvalues of the matrices $M_X(\xi_{\alpha\beta})$, defined by

$$M_A(\xi_{\alpha\beta}) = \begin{pmatrix}
\mu_{23} + \mu_{24} + \mu_{34} + \xi_{12} + \xi_{13} & \mu_{12} & \mu_{13} & \mu_{14} \\
\mu_{12} & \mu_{13} + \mu_{14} + \mu_{34} - \xi_{12} - \xi_{13} & \mu_{23} & \mu_{24} \\
\mu_{13} & \mu_{23} & \mu_{12} + \mu_{14} + \mu_{24} + \xi_{12} - \xi_{13} & \mu_{34} \\
\mu_{14} & \mu_{24} & \mu_{34} & \mu_{12} + \mu_{13} + \mu_{23} \xi_{12} + \xi_{13} \\
\end{pmatrix}$$
\[ \mathbf{M}_G(\xi^{\alpha\beta}) = \begin{pmatrix} \mu_{23} + \mu_{24} + \mu_{34} - \xi^{12} + \xi^{13} & \mu_{12} & \mu_{13} & \mu_{14} \\ + \xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & & & \\ \mu_{12} & \mu_{13} + \mu_{14} + \mu_{34} + \xi^{12} + \xi^{13} & \mu_{23} & \mu_{24} \\ + \xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & & & \\ \mu_{13} & \mu_{23} & \mu_{12} + \mu_{14} + \mu_{24} + \xi^{12} + \xi^{13} & \mu_{34} \\ + \xi^{14} - \xi^{23} + \xi^{24} + \xi^{34} & & & \\ \mu_{14} & \mu_{24} & \mu_{34} & \mu_{12} + \mu_{13} + \mu_{23} \xi^{12} + \xi^{13} \\ + \xi^{14} + \xi^{23} - \xi^{24} + \xi^{34} & & & \end{pmatrix} \]

\[ \mathbf{M}_C(\xi^{\alpha\beta}) = \begin{pmatrix} \mu_{23} + \mu_{24} + \mu_{34} - \xi^{12} - \xi^{13} & \mu_{12} & \mu_{13} & \mu_{14} \\ + \xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & & & \\ \mu_{12} & \mu_{13} + \mu_{14} + \mu_{34} + \xi^{12} + \xi^{13} & \mu_{23} & \mu_{24} \\ + \xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & & & \\ \mu_{13} & \mu_{23} & \mu_{12} + \mu_{14} + \mu_{24} + \xi^{12} + \xi^{13} & \mu_{34} \\ + \xi^{14} - \xi^{23} + \xi^{24} + \xi^{34} & & & \\ \mu_{14} & \mu_{24} & \mu_{34} & \mu_{12} + \mu_{13} + \mu_{23} \xi^{12} + \xi^{13} \\ + \xi^{14} + \xi^{23} - \xi^{24} + \xi^{34} & & & \end{pmatrix} \]
We are interested in the long time limit of the traces in Eq. (4.218), which is given by

\[ \lim_{t \to \infty} \frac{\ln Q_X}{t} = \lambda^X_{\text{max}} \]  

(4.220)

Here, \( \lambda^X_{\text{max}} \) is the largest eigenvalue of the matrix \( M_X(\xi^\alpha_\beta) \) defined in Eq. (4.219).

A similar analysis applies to the Eigen model, where the matrices \( M_X(\xi^\alpha_\beta, \eta^\alpha_\beta) \)
have the same structure as in Eq. (4.219), with the substitution $\mu_{\alpha\beta} \rightarrow \eta_{\alpha\beta}$.

\[
M_A(\zeta^{\alpha\beta}, \eta^{\alpha\beta}) = \begin{pmatrix}
\eta^{23} + \eta^{24} + \eta^{34} + \xi^{12} + \xi^{13} & \eta^{12} & \eta^{13} & \eta^{14} \\
+\xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & \eta^{12} & \eta^{13} + \eta^{14} + \eta^{34} - \xi^{12} + \xi^{13} & \eta^{23} & \eta^{24} \\
\eta^{13} + \eta^{14} + \eta^{34} - \xi^{12} + \xi^{13} & \eta^{23} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} - \xi^{13} & \eta^{34} \\
+\xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & \eta^{13} + \eta^{14} + \eta^{34} - \xi^{12} + \xi^{13} & \eta^{23} & \eta^{24} \\
\eta^{13} + \eta^{14} + \eta^{34} - \xi^{12} + \xi^{13} & \eta^{23} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} - \xi^{13} & \eta^{34} \\
\eta^{14} + \eta^{24} + \eta^{34} - \xi^{12} + \xi^{13} & \eta^{23} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} - \xi^{13} & \eta^{24} \\
-\xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} - \xi^{13} & \eta^{23} & \eta^{24} \\
-\xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} - \xi^{13} & \eta^{23} & \eta^{24} \\
\end{pmatrix}
\]

\[
M_C(\zeta^{\alpha\beta}, \eta^{\alpha\beta}) = \begin{pmatrix}
\eta^{23} + \eta^{24} + \eta^{34} - \xi^{12} + \xi^{13} & \eta^{12} & \eta^{13} & \eta^{14} \\
+\xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & \eta^{12} & \eta^{13} + \eta^{14} + \eta^{34} + \xi^{12} + \xi^{13} & \eta^{23} & \eta^{24} \\
\eta^{13} + \eta^{14} + \eta^{34} + \xi^{12} + \xi^{13} & \eta^{23} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} + \xi^{13} & \eta^{34} \\
+\xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & \eta^{13} + \eta^{14} + \eta^{34} + \xi^{12} + \xi^{13} & \eta^{23} & \eta^{24} \\
\eta^{13} + \eta^{14} + \eta^{34} + \xi^{12} + \xi^{13} & \eta^{23} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} + \xi^{13} & \eta^{34} \\
\eta^{14} + \eta^{24} + \eta^{34} + \xi^{12} + \xi^{13} & \eta^{23} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} + \xi^{13} & \eta^{24} \\
+\xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} + \xi^{13} & \eta^{23} & \eta^{24} \\
+\xi^{14} + \xi^{23} + \xi^{24} + \xi^{34} & \eta^{13} + \eta^{14} + \eta^{24} + \xi^{12} + \xi^{13} & \eta^{23} & \eta^{24} \\
\end{pmatrix}
\]
Then, after Eq. (4.215), we have that for long times

\[
\lim_{t \to \infty} \sum_{j=1}^{N} \frac{\ln Q(j)}{t} = \sum_{X=A,C,G,T/U} y_X \lambda_X^{\max} 
\]
Chapter 5
Quasispecies theory for finite populations

Abstract

We present a general formulation of quasispecies theories, the Crow-Kimura and Eigen models, generalized to include horizontal gene transfer for a finite population of evolving binary sequences. By means of a field-theoretic representation, we obtain analytical expressions for the average population distribution, as well as for the fluctuation of the population numbers. Our theoretical results show that horizontal gene transfer reduces the fluctuations in the population numbers of a finite population.

5.1 Introduction

Biological populations in nature are finite. In particular, it is clear that the number of individuals in a population is much smaller than the number of possible genetic sequences, even for genomes of modest length. For example, the largest populations observed in biological systems, RNA viruses, are on the order of \( N = 10^{12} \) viral particles within a single infected organism [121]. These viruses possess a relatively short genome of length \( L \sim 10^3 - 10^4 \) bases [121], and hence the theoretical size of the sequence space is \( 4^L \sim 10^{6000} \gg N \). From this example, it is clear that no real biological population will be able to sample the entire sequence space during evolutionary dynamics [122], and therefore finite population size effects may
be important for a realistic description of evolution [123]. Finite populations with asexual reproduction are subject to the "Muller's ratchet" effect [79], which is the tendency to accumulate deleterious mutations in finite populations [79, 45, 43]. It has been suggested that horizontal gene transfer and recombination may provide a way to escape Muller's ratchet in small populations [124, 125, 48, 126], and this mechanism has been proposed as one of the evolutionary advantages of sex, despite the additional mutational load for fitness functions with positive epistasis [79, 83, 45, 43, 42, 86, 47, 48].

Quasispecies models for molecular evolution, represented by the Crow-Kimura model [95] and the Eigen model [11, 93, 12, 94], are traditionally formulated in the language of chemical kinetics. That is, they describe the basic processes of mutation and selection in an infinite population of self-replicating, information encoding molecules such as RNA or DNA, which are assumed to be drawn from a binary alphabet (e.g. purines/pyrimidines). These models exhibit a phase transition in the infinite genome limit [11, 93, 12, 21, 20, 94, 97, 29, 28], separating an organized or quasispecies phase from a disordered phase. The quasispecies is composed by a collection of nearly neutral mutants rather than by a single sequence type. Despite its abstract character, the quasispecies model has been successfully applied to interpret experimental studies in RNA viruses [98, 99, 100, 101].

In this infinite population limit, the mean field approach that is customary in
chemical kinetics is justified and the evolution of the probability distribution of sequence types can be described by a deterministic system of differential equations. This mean field approach cannot capture the fluctuations in the numbers of individuals with different sequences, which are a consequence of the stochastic dynamics of the process. An accurate description of a finite population therefore requires a master equation formulation [123].

In this article, we extend the parallel and Eigen models of quasispecies theory to include finite population effects, by considering a master equation formulation. By means of a field-theoretic method [29, 34, 58] we derive from the master equation a system of coupled differential equations for the probability distribution and the fluctuation of numbers of individuals with given sequences. We furthermore study the effects of horizontal gene transfer on the steady-state probability distribution and fluctuations.

5.2 The parallel model in a finite population

We consider a finite population, composed of $N < \infty$ binary purine/pyrimidine sequences, of length $L$. We assume that the replication rate, or microscopic fitness, is a function of the Hamming distance from the wild-type genome, and hence of the one-dimensional coordinate $0 \leq \xi \leq L$ representing the total number of purines along the sequence. We define $r(\xi) = Lf(\xi)$ as the replication rate, and hence we characterize the finite population in terms of the occupation numbers $\{n_\xi\}_{0 \leq \xi \leq L}$ for
each Hamming distance class $C_\xi$, which must satisfy the conservation equation

$$\sum_{\xi=0}^{L} n_\xi = N. \quad (5.1)$$

We define the probability distribution for the entire population, $P\{\{n_\xi\};t\}$, as a function of the set of occupation numbers $\{n_\xi\}_{0 \leq \xi \leq L}$. We consider a mutation rate $\mu$, and horizontal gene transfer rate $\nu$ [40, 127]. Therefore, the master equation governing the time evolution of the probability distribution is

$$\frac{\partial}{\partial t} P\{\{n_\xi\};t\} = \frac{1}{N} \sum_{\xi \neq \xi'} r(\xi)[(n_\xi - 1)(n_{\xi'} + 1)P(\{n_\xi - 1, n_{\xi'} + 1\};t) - n_\xi n_{\xi'} P\{\{n_\xi\};t\}]$$

$$+ \mu \sum_{\xi=0}^{L} [(L - \xi)(n_\xi + 1)P(\{n_\xi + 1, n_{\xi+1} - 1\};t)$$

$$+ \xi(n_\xi + 1)P(\{n_{\xi-1} - 1, n_\xi + 1\};t) - Ln_\xi P\{\{n_\xi\};t\}]$$

$$+ \nu \sum_{\xi=0}^{L} [\rho_+(L - \xi)(n_\xi + 1)P(\{n_\xi + 1, n_{\xi+1} - 1\};t)$$

$$+ \xi\rho_-(n_\xi + 1)P(\{n_{\xi-1} - 1, n_\xi + 1\};t) - n_\xi\rho_+(L - \xi) + \rho_\xi] P\{\{n_\xi\};t\}] \quad (5.2)$$

with $\rho_\pm = (1 \pm u)/2$ the probability to insert a purine or pyrimidine, and average base composition, $u = \frac{1}{N} \sum_{\xi=0}^{L} (2\xi/L - 1) n_\xi$.

We introduce an exact representation of the classical master equation (5.2) in terms of a many-body quantum theory. For that purpose, we define the population state vector

$$|\Psi(t)\rangle = \sum_{\{n_\xi\}} P\{\{n_\xi\};t\}|\{n_\xi\}\rangle \quad (5.3)$$
with \(|\{n_\xi\}\rangle = |n_0, n_1, \ldots, n_L\rangle = \prod_{\xi=0}^L \otimes |n_\xi\rangle\). This population state vector evolves according to a Schrödinger equation in imaginary time

\[
\frac{d}{dt} |\Psi(t)\rangle = -\hat{H} |\Psi(t)\rangle \tag{5.4}
\]

The Hamiltonian operator \(\hat{H}\) is expressed in terms of boson creation and destruction operators \([\hat{a}_\xi, \hat{a}_\xi^\dagger] = \delta_{\xi, \xi'}\), whose action on the occupation number vectors is

\[
\hat{a}_\xi |n_\xi\rangle = n_\xi |n_\xi - 1\rangle \quad \hat{a}_\xi^\dagger |n_\xi\rangle = |n_\xi + 1\rangle \tag{5.5}
\]

The Hamiltonian operator is given by

\[
-\hat{H} = \frac{1}{N} \sum_{\xi, \xi' = 0}^L r(\xi)\hat{a}_\xi^\dagger(\hat{a}_\xi^\dagger - \hat{a}_\xi^\dagger)\hat{a}_\xi^\dagger\hat{a}_\xi + \mu \sum_{\xi' = 0}^L (L - \xi)(\hat{a}_{\xi+1}^\dagger - \hat{a}_\xi^\dagger)\hat{a}_\xi + \xi(\hat{a}_{\xi-1}^\dagger - \hat{a}_\xi^\dagger)\hat{a}_\xi \\
+ \nu \sum_{\xi = 0}^L |\rho_+(L - \xi)(\hat{a}_{\xi+1}^\dagger - \hat{a}_\xi^\dagger)\hat{a}_\xi + \rho_- \xi(\hat{a}_{\xi-1}^\dagger - \hat{a}_\xi^\dagger)\hat{a}_\xi| \tag{5.6}
\]

Here, the first term represents replication of sequences, while preserving the population size \(N\). The second term represents single point mutations, and the third term represents horizontal gene transfer of single sites from the population \([40, 127]\).

The evolution equation Eq. (5.4) is formally integrated in time

\[
|\Psi(t)\rangle = e^{-\hat{H}t} |\Psi(0)\rangle \tag{5.7}
\]

with \(|\Psi(0)\rangle = |\{n_\xi^0\}\rangle\) representing the initial configuration of the population.

The population average of a normal-ordered classical observable \(F(\{\hat{a}_\xi\})\) is obtained, in this formalism, by taking the product on the left with the 'standard bra'
\[ \langle \psi \rangle = \langle 0 \vert \left( \prod_{\xi=0}^{L} e^{\hat{a}_{\xi}} \right) \vert \Phi \rangle [34, 58], \]

\[ \langle F \rangle = \langle \psi | F(\{\hat{a}_{\xi}\}) | \psi(t) \rangle = \langle \psi | F(\{\hat{a}_{\xi}\}) e^{-\hat{H}t} | \psi \rangle \] (5.8)

We introduce a Trotter factorization in Eq. (5.8), with a basis of coherent states \( |z_{\xi} \rangle \)
defined by \( \hat{a}_{\xi} |z_{\xi} \rangle = z_{\xi} |z_{\xi} \rangle \) at each time-slice \( 0 \leq k \leq t/\epsilon \), and \( \epsilon \to 0 \),

\[ \langle F \rangle = \int [dz^{*} \mathcal{D}z] \langle 0 \vert \left( \prod_{\xi=0}^{L} e^{\hat{a}_{\xi}} \right) F(\{\hat{a}_{\xi}\}) |z_{\xi}(t/\epsilon) \rangle \prod_{k=1}^{t/\epsilon} \langle \{z_{\xi}(k)\} | e^{-\epsilon \hat{H}} | \{z_{\xi}(k-1)\} \rangle \{z_{\xi}(0)\} \{n_{\xi}^{0}\} \right] \]

\[ = \frac{\mu \epsilon}{N} \sum_{\xi, \xi'} \sum_{k=1}^{L} \langle \{z_{\xi}(0)\} | \{z_{\xi}(k)\} | \{z_{\xi}(0)\} \{z_{\xi}(k-1)\} \} \right] \]

The action in the exponent of Eq. (5.9) is defined, after the change of variables \( z^{*} = 1 + \bar{z} \),

\[ S[\{\bar{z}\}, \{z\}] = \sum_{\xi=0}^{L} [z_{\xi}(0) z_{\xi}(0) - n_{\xi}^{0} \ln(1 + z_{\xi}(0))] + \sum_{k=1}^{t/\epsilon} \bar{z}_{\xi}(k) \{z_{\xi}(k) - z_{\xi}(k-1)\} \]

\[ = \mu \epsilon \sum_{\xi=0}^{L} \sum_{k=1}^{t/\epsilon} [(L - \xi) \bar{z}_{\xi+1}(k) + \xi z_{\xi}(k) - L \bar{z}_{\xi}(k)] z_{\xi}(k-1) \]

\[ = \nu \epsilon \sum_{\xi=0}^{L} \sum_{k=1}^{t/\epsilon} [(L - \xi) \rho_{+} \bar{z}_{\xi+1}(k) + \xi \rho_{-} \bar{z}_{\xi}(k) - \{(L - \xi) \rho_{+} + \xi \rho_{-}\} \bar{z}_{\xi}(k)] z_{\xi}(k-1) \]

\[ = \frac{\epsilon}{N} \sum_{\xi, \xi'} \sum_{k=1}^{L} \langle \{z_{\xi}(0)\} | \{z_{\xi}(k)\} | \{z_{\xi}(0)\} \{z_{\xi}(k-1)\} z_{\xi}(k-1) z_{\xi'}(k-1) \} \right] \]

We look for a saddle-point in the action Eq. (5.10), \( \frac{\delta S}{\delta z_{\xi}(k)} \bigg|_{c} = 0 \), which has the solution \( z_{\xi}(k) = 0, \forall \xi, k \).

The second saddle-point condition, \( \frac{\delta S}{\delta z_{\xi}(k)} \bigg|_{c} = 0 \), yields for \( k = 0 \), \( z_{\xi}(0) = n_{\xi}^{0} \). For \( k > 0 \), we define \( z_{\xi}(k) = N P_{\xi}(k) \), and in the limit \( \epsilon \to 0 \), the saddle-point condition
translates into a differential equation for the class probabilities $P_\xi(t)$ (see Appendix 1),

$$
\frac{d}{dt} P_\xi(t) = \mu[(L - \xi + 1)P_{\xi-1}(t) + (\xi + 1)P_{\xi+1}(t) - LP_\xi(t)]
$$

$$
+ \nu[\rho_+(L - \xi + 1)P_{\xi-1}(t) + \rho_-(\xi + 1)P_{\xi+1}(t) - \{(L - \xi)\rho_+ + \xi\rho_-\}P_\xi(t)]
$$

$$
+ \left[ r(\xi) - \sum_{\xi' = 0}^L r(\xi')P_{\xi'}(t) \right] P_\xi(t)
$$

subject to the initial condition $P_\xi(0) = n^0_\xi/N$. This equation is exactly infinite pop­
ulation quasispecies theory generalized to include horizontal gene transfer [40].

We obtain the fluctuations around the saddle-point of this theory, by defining

$$
\delta \tilde{z}_\xi(k) = \tilde{z}_\xi(k) - \bar{\xi}(k), \text{ and } \delta z_\xi(k) = z_\xi(k) - \bar{\xi}(k) = z_\xi(k) - NP_\xi(k).
$$

The action is expanded to quadratic order as (see Appendix 1),

$$
\Delta S = S - S_c = \frac{1}{2}X^T \Pi^{-1}X + O(X^3)
$$

with $X^T = (\{\delta \tilde{z}(0), \delta z(0)\}, \ldots, \{\delta \tilde{z}(t/\epsilon), \delta z(t/\epsilon)\})$. We are interested in the correla­
tor $(\{X(t/\epsilon), \{X(t/\epsilon)\})$. Therefore, we define the generating functional

$$
Z[J] = \int [Dx] e^{-\frac{1}{2}X^T \Pi^{-1}X + J^TX} = (2\pi)^{Lt/\epsilon}(\det \Pi)^{-Lt/\epsilon e^{\frac{1}{2}J^T \Pi J}}
$$

By functional differentiation of Eq. (5.13), we obtain

$$
\langle (X(t/\epsilon), \{X(t/\epsilon)\}) = \frac{1}{Z[J]} \frac{\delta^2 Z[J]}{\delta \{J(t/\epsilon)\} \delta \{J(t/\epsilon)\}} \bigg|_{J=0} = [\Pi]_{t/\epsilon, t/\epsilon}
$$

(5.14)
As shown in Appendix 1, we obtain \([\Pi]_{t/e,t/e} = b_{t/e,t/e}\), with
\[
b_{k,k} = \begin{pmatrix} 0 & I \\ I & C(k) \end{pmatrix}
\]  
(5.15)
The matrices \(C(k)\) are symmetric \(C = C^T\), and in the continuous time limit \(\epsilon \rightarrow 0\) evolve according to the Lyapunov equation
\[
\frac{d}{dt} C = AC + CA^T + B
\]  
(5.16)
subject to the initial condition \(C_{\xi,\xi'} = -n^0_\xi \delta_{\xi,\xi'}\), with matrices \(A, B\) defined in Appendix 1, Eqs. (5.40) and (5.41).

The fluctuations in number of individuals with a given sequence are obtained from the relation \((\delta n_\xi)^2 = \langle n^2_\xi \rangle - \langle n_\xi \rangle^2 = NP_\xi + C_{\xi,\xi}\). Therefore, the fluctuations in the class probability distribution are obtained from the formula (see Appendix 3),
\[
\frac{\langle (\delta n_\xi)^2 \rangle}{N^2} = \frac{1}{N} (P_\xi + \frac{1}{N} C_{\xi,\xi})
\]  
(5.17)

In summary, from the master equation (5.1), we obtained an explicit analytical equation for the class probability distribution, Eq. (5.11), and for the fluctuations of the number of individuals, Eqs. (5.16) and (5.17). This system of coupled differential equations is solved in the next section.

5.2.1 Results for the parallel model

We studied the probability distributions and fluctuations for two different fitness functions, a discontinuous sharp peak \(f(u) = A\delta_{u,+1}\), and a smooth, quadratic fitness
landscape \( f(u) = \frac{k}{2} u^2 + 1 \).

For the sharp peak in the absence of horizontal gene transfer \((\nu = 0)\), we obtain from Eq. (5.11) that the wild-type probability \( P_{\xi=L} = 1 - \frac{\mu}{A} + O(L^{-1}) \). Horizontal gene transfer \((\nu > 0)\) does not affect this distribution (see Appendix 4). From Eq. (5.16) and Eq. (5.17), we obtain that the fluctuation is given by \( \langle (\delta n_{\xi=L})^2 \rangle / N^2 = \frac{1}{N} \frac{\mu}{A} \), a result first given in [128].

For the quadratic fitness, we tested our analytical theory with stochastic simulations based on a Lebowitz/Gillespie algorithm [110, 109] in which we explicitly simulate a population of size \( L \) undergoing the stochastic processes of mutation, horizontal gene transfer, and replication.

In Fig. 5.1, the steady-state probability distribution obtained from the numerical solution of Eq. (5.11) is compared with the distributions obtained from stochastic simulations, for different sizes \( N \) of the population. The stochastic results represent an average over 50 independent numerical experiments. An excellent agreement is observed between the theory and stochastic simulations, which improves systematically as the population size increases.

A similar comparison for the fluctuations in the probability distribution is shown in Fig. 5.2, in the absence of horizontal gene transfer \( \nu = 0 \), at different sizes of the population \( N \). The fluctuations are large. A slow convergence towards the predictions of the theory is observed with increasing the population size, with a close agreement
Figure 5.1  Theoretical versus results from stochastic simulations for the probability distribution of the parallel model and quadratic fitness. The theory is obtained from Eq. (5.11). The stochastic results are obtained by averaging over 50 independent numerical experiments. In the inset is shown the same comparison in the absence of horizontal gene transfer ($\nu = 0$). Note the Muller's ratchet phenomenon, whereby fitness is reduced for finite populations, is greatly suppressed for $\nu > 0$. Here $k = 4$, $L = 200$ for the stochastic simulations, and $\nu = 7$ in the main figure.
Figure 5.2  Theoretical versus stochastic fluctuations in the probability distribution for the parallel model and quadratic fitness, in the absence of horizontal gene transfer, $\nu = 0$. Here, $L = 100$ and $k = 3.0$. The theory is obtained from Eqs. (5.16) and (5.17).

At $N = 10^8$.

In Fig. 5.3, we compare our theory with stochastic simulations, at different rates of horizontal gene transfer. Here, the results obtained from stochastic simulations converge toward the theoretical value calculated from Eqs. (5.16) and (5.17) as the size of the population $N$ increases. We notice that the fluctuations are much smaller and convergence is achieved at smaller values of the population size $N$, as compared to the case when horizontal gene transfer is absent $\nu = 0$, Fig. 5.2.

As depicted in Fig. 5.4, our theoretical results show that horizontal gene transfer
Figure 5.3  Theoretical versus stochastic fluctuations in the probability distribution for the parallel model and quadratic fitness, with horizontal gene transfer rate $\nu = 7.0$. Here, $L = 200$ and $k = 4.0$. The theory is obtained from Eqs. (5.16) and (5.17). The stochastic results are obtained by averaging over 50 independent numerical experiments. For $\nu = 0$, the theoretical curve would rise to values of the order of $10^6$. 
Figure 5.4 Fluctuations in the probability distribution as predicted from our theory for the parallel model and quadratic fitness landscape, at different recombination rates $\nu > 0$. Here, $L = 200$ and $k = 3.0$. For $\nu = 0$, the curve would rise to values of the order of $10^7$.

has a dramatic effect in reducing the magnitude of the fluctuations in the distribution. Indeed, a small rate of horizontal gene transfer is enough to reduce by several orders of magnitude these fluctuations, as compared to the case without horizontal gene transfer $\nu = 0$.

These results are consistent with the trend observed in our stochastic simulations for the variance of the average composition, Fig. 5.5. Here, we see that horizontal gene transfer reduces the variance between different numerical experiments.

This reduction in variance is also observed in the corresponding histograms for
Figure 5.5  Average composition as a function of time, averaged over 50 independent numerical experiments. Also shown are one standard deviation envelopes $\pm \sigma(t)$. The steady-state averages $<u> \pm \sqrt{\langle (\delta u)^2 \rangle}$ are displayed as solid lines for reference. The population size is $N = 10^4$ sequences.
Figure 5.6  Histograms representing the steady-state average composition among 50 independent numerical experiments, at different horizontal gene transfer rates, \( \nu \). The population size is \( N = 10^4 \) sequences.

The 50 independent numerical experiments, as displayed in Fig. 5.6. The histograms become more compact due to a decrease in the standard deviation between independent numerical experiments, as the horizontal gene transfer rate \( \nu \) is increased. These numerical examples, for the small population size of \( N = 10^4 \), illustrate the effect of horizontal gene transfer in counter-balancing “Muller’s ratchet” effect in finite populations [79, 48, 124, 125, 126].
5.3 The Eigen model in a finite population

We consider a finite population, composed of \( N < \infty \) binary sequences, of length \( L \). We assume that the replication rate is \( r(\xi) = Lf(\xi) \) as in the parallel model. In this case, it is assumed that multiple mutations can occur along each sequence as a consequence of errors in the replication process, and consequently \( Q_{\xi,\xi'} \) is the transfer matrix for mutations from class \( \xi' \) into class \( \xi \), with

\[
Q_{\xi,\xi'} = \sum_{\xi_1=0}^{\min\{\xi+\xi',2L-(\xi+\xi')\}} q^{L-(2\xi_1+|\xi'-\xi|)}(1-q)^{2\xi_1+|\xi'-\xi|} \left( \xi_1 + \frac{|\xi'-\xi'|}{2} \right) \left( \xi_1 + \frac{|\xi'-\xi|}{2} \right)
\]

(5.18)

Here, \( q \approx 1 \) characterizes the fidelity in the replication process, when \( 1-q \) is the probability (per site) that an incorrect base is placed by the polymerase enzyme.

In the representation of occupation numbers \( \{n_\xi\}_{0 \leq \xi \leq L} \) satisfying the conservation
Eq. (5.1), we formulate the master equation for the population probability $P({\{n_{\xi}\};t})$

$$\frac{\partial}{\partial t} P({\{n_{\xi}\};t}) = \left(1 - \frac{\nu}{L}\right) \left\{ \sum_{\xi=0}^{L} r(\xi) Q_{\xi,\xi} \left[ (n_{\xi} - 1) \sum_{\xi''\neq\xi} \frac{n_{\xi''} + 1}{N} P({\{n_{\xi} - 1, n_{\xi''} + 1\};t)} \right] 
- n_{\xi} \sum_{\xi''\neq\xi} \frac{n_{\xi''}}{N} P({\{n_{\xi}, n_{\xi''}\};t}) \right\} 
+ \sum_{\xi=0}^{L} r(\xi) \sum_{\xi'\neq\xi} Q_{\xi',\xi} \left[ n_{\xi} \sum_{(\xi'',\xi''\neq\xi')} \frac{n_{\xi''} + 1}{N} P({\{n_{\xi'}, n_{\xi''} + 1\};t}) \right] 
- (n_{\xi} - 1) \frac{n_{\xi}}{N} P({\{n_{\xi}, n_{\xi'}\};t}) \right\} 
+ \sum_{\xi=0}^{L} d(\xi) \left[ (n_{\xi} + 1) \sum_{\xi'\neq\xi} \frac{n_{\xi'} - 1}{N} P({\{n_{\xi} + 1, n_{\xi'} - 1\};t}) 
- n_{\xi} \sum_{\xi'\neq\xi} \frac{n_{\xi'}}{N} P({\{n_{\xi}, n_{\xi'}\};t}) \right] 
+ \sum_{\xi,\xi'=0}^{L} Q_{\xi',\xi+1} \frac{\nu}{L} \rho_{+}(L - \xi) r(\xi) n_{\xi} \sum_{(\xi'',\xi''\neq\xi')} \left[ \frac{n_{\xi''} + 1}{N} P({\{n_{\xi'}, n_{\xi''} + 1\};t}) \right] 
- \frac{n_{\xi}}{N} P({\{n_{\xi'}, n_{\xi''}\};t}) \right\} 
+ \sum_{\xi,\xi'=0}^{L} Q_{\xi',\xi-1} \frac{\nu}{L} \rho_{-}\xi r(\xi) n_{\xi} \sum_{(\xi'',\xi''\neq\xi')} \left[ \frac{n_{\xi''} + 1}{N} P({\{n_{\xi'}, n_{\xi''} + 1\};t}) \right] 
- \frac{n_{\xi}}{N} P({\{n_{\xi'}, n_{\xi''}\};t}) \right\} 
\quad (5.19)$$

This system can be exactly mapped, as in the case of the parallel model, into a
many-body quantum Hamiltonian, defined by

$$-\hat{H} = \left(1 - \frac{\nu}{L}\right) \sum_{\xi, \xi', \xi''=0}^L Q_{\xi', \xi''} r(\xi) \hat{a}_{\xi}^\dagger (\hat{a}_{\xi'} - \hat{a}_{\xi''}) \hat{a}_{\xi'} \hat{a}_{\xi''} + \frac{1}{N} \sum_{\xi, \xi'=0}^L d(\xi') \hat{a}_{\xi'}^\dagger (\hat{a}_{\xi'}^\dagger - \hat{a}_{\xi'}) \hat{a}_{\xi'}$$

$$+ \frac{1}{N} \sum_{\xi, \xi', \xi''=0}^L Q_{\xi', \xi''} \rho_+(L - \xi) r(\xi) \hat{a}_{\xi}^\dagger (\hat{a}_{\xi'}^\dagger - \hat{a}_{\xi''}) \hat{a}_{\xi'}$$

$$+ \frac{1}{N} \sum_{\xi, \xi', \xi''=0}^L Q_{\xi', \xi''} \rho_- \xi r(\xi) \hat{a}_{\xi}^\dagger (\hat{a}_{\xi'}^\dagger - \hat{a}_{\xi''}) \hat{a}_{\xi'}$$

(5.20)

From the formulation expressed in Eq. (5.9), and introducing an analogous Trotter factorization, we obtain in this case the action

$$S[\{\tilde{z}\}, \{z\}] = \left(1 - \frac{\nu}{L}\right) \left\{ \sum_{\xi=0}^L \left[ \tilde{z}_{\xi}(0) z_{\xi}(0) - n_{\xi}^0 \ln[1 + \tilde{z}_{\xi}(0)] + \sum_{k=1}^{t/\epsilon} \tilde{z}_{\xi}(k) \{ z_{\xi}(k) - z_{\xi}(k-1) \} \right] 
- \frac{\epsilon}{N} \sum_{k=1}^{t/\epsilon} \sum_{\xi, \xi', \xi''=0}^L Q_{\xi', \xi''} \rho_+(L - \xi) r(\xi) \tilde{z}_{\xi}(k) \{ \tilde{z}_{\xi}(k) - \tilde{z}_{\xi''}(k) \} z_{\xi}(k-1) z_{\xi''}(k-1) \right\}$$

$$- \frac{\epsilon}{N} \sum_{k=1}^{t/\epsilon} \sum_{\xi, \xi'=0}^L d(\xi') \{ 1 + \tilde{z}_{\xi}(k) \} \{ \tilde{z}_{\xi'}(k) - \tilde{z}_{\xi''}(k) \} z_{\xi}(k-1) z_{\xi'}(k-1)$$

$$- \frac{\nu}{L} \frac{\epsilon}{N} \sum_{k=1}^{t/\epsilon} \sum_{\xi, \xi', \xi''=0}^L \left[ Q_{\xi', \xi''} \rho_+(L - \xi) + Q_{\xi', \xi''} \rho_- \xi \right] r(\xi) \{ 1 + \tilde{z}_{\xi}(k) \}$$

$$\times \{ \tilde{z}_{\xi'}(k) - \tilde{z}_{\xi''}(k) \} z_{\xi}(k-1) z_{\xi'}(k-1)$$

(5.21)

We look for a saddle-point in the action Eq. (5.21). From the condition \( \frac{\partial S}{\partial z_{\xi}(k)} \bigg|_c = 0 \), we obtain \( \tilde{z}_{\xi}(k) = 0 \ \forall k, \xi \) (see Appendix 2). From the second equation \( \frac{\partial S}{\partial \tilde{z}_{\xi}(k)} \bigg|_c = 0 \),
after taking the continuous time limit $\epsilon \to 0$, we obtain (see Appendix 2)

\[
\frac{d}{dt} P_\xi(t) = \left(1 - \frac{\nu}{L}\right) \left[ \sum_{\xi'=0}^{L} Q_{\xi \xi'} r(\xi') P_{\xi'}(t) - P_\xi(t) \sum_{\xi'=0}^{L} r(\xi') P_{\xi'}(t) \right] \\
- P_\xi(t) \left[ d(\xi) - \sum_{\xi'=0}^{L} P_{\xi'}(t) d(\xi') \right] \\
+ \frac{\nu}{L} \left[ \sum_{\xi'=0}^{L} \left\{ Q_{\xi \xi'+1} \rho_+ (L - \xi') + Q_{\xi \xi'-1} \rho_- \xi' \right\} r(\xi') P_{\xi'}(t) \right] \\
- P_\xi(t) \sum_{\xi'=0}^{L} \{ \rho_+ (L - \xi') + \rho_- \xi' \} r(\xi') P_{\xi'}(t) 
\]

(5.22)

Here, $P_\xi = z_\xi^2/N$, and the initial condition corresponds to $P_\xi(0) = n_\xi^0/N$ (see Appendix 2). This is exactly infinite population quasispecies theory generalized to include horizontal gene transfer [40].

We expand the action Eq. (5.21) around the saddle-point, by defining $\delta z_\xi(k) = z_\xi(k) - z_\xi^2(k) = z_\xi(k) - NP_\xi(k)$, and $\delta \tilde{z}_\xi(k) = \tilde{z}_\xi(k) - \bar{z}_\xi^2(k) = \bar{z}_\xi^2(k)$. The action is therefore given by

\[
\Delta S = S - S_c = \frac{1}{2} X^T \Pi^{-1} X + O(X^3) 
\]

(5.23)

Here, $X^T = \{ \delta \tilde{z}_0, \delta z_0 \} \ldots \{ \delta \tilde{z}(t/e), \delta z(t/e) \}$ (see Appendix 2).

We are interested in the correlator $\langle \{ X(t/e) \}, \{ X(t/e) \} \rangle$. Therefore, we define the generating functional

\[
Z[J] = \int [DX] e^{-\frac{1}{2} X^T \Pi^{-1} X + J^T X} = (2\pi)^{Lt/\epsilon} (\det \Pi)^{-Lt/\epsilon} e^{\frac{1}{2} J^T \Pi J} 
\]

(5.24)
By functional differentiation of Eq. (5.13), we obtain

\[
\langle \{X(t/\epsilon)\} \{X(t/\epsilon)\} \rangle = \left. \frac{1}{Z[J]} \frac{\delta^2 Z[J]}{\delta J(t/\epsilon) \delta J(t/\epsilon)} \right|_{J=0} = [\Pi]_{t/\epsilon, t/\epsilon} \tag{5.25}
\]

As shown in Appendix 2, we obtain \([\Pi]_{t/\epsilon, t/\epsilon} = b_{t/\epsilon, t/\epsilon}\), with

\[
b_{k,k} = \begin{pmatrix} 0 & I \\ I & C(k) \end{pmatrix}
\] \tag{5.26}

The matrices \(C(k)\) are symmetric \(C = C^T\), and in the continuous time limit \(\epsilon \to 0\) evolve according to the Lyapunov equation

\[
\frac{d}{dt} C = AC + CA^T + B \tag{5.27}
\]

subject to the initial condition \(C_{\xi,\xi'} = -n_{\xi}^0 \delta_{\xi,\xi'}\), with matrices \(A, B\) defined in Appendix 2, Eqs. (5.56) and (5.57).

In summary, we obtained a coupled system of differential equations representing the probability distribution, Eq. (5.22) and the fluctuations in this distribution, Eq. (5.27).

### 5.3.1 Results for the Eigen model

For the Eigen model, in the absence of horizontal gene transfer \((\nu = 0)\), we obtain from Eq. (5.22) that the wild type probability is (see Appendix 5)

\[
P_L = \frac{e^{-\mu A - A_0}}{A - A_0} \tag{5.28}
\]
Figure 5.7  Probability distributions, in an infinite population as predicted from our theory Eq. (5.22), for the Eigen model and quadratic fitness, at different horizontal gene transfer rates $\nu$.

From Eq. (5.27) we obtain that the fluctuation is given by (see Appendix 5).

$$\frac{\langle (\delta n_L)^2 \rangle}{N^2} = \frac{1}{N} \frac{e^{-\mu}(1 - e^{-\mu})A^2}{(A - A_0)^2}$$  \hspace{1cm} (5.29)

In Fig. 5.7, we present the steady-state probability distributions, for different rates of horizontal gene transfer, as obtained from our theory Eq. (5.22). We observe the theory predicts that horizontal gene transfer shifts the population distribution away from the wild type at $\xi = L (u = 1)$, thus decreasing the mean fitness, for this quadratic fitness function with positive epistasis.
Figure 5.8  Fluctuations in the number of individuals with a given sequence, as predicted from our theory Eq. (5.27), for the Eigen model and quadratic fitness, at different horizontal gene transfer rates $v$. For $v = 0$, the curve would rise to values of the order of $10^7$.

In Fig. 5.8 we present the fluctuations in the number of individuals with a given sequence for the quadratic fitness, as predicted from our theory Eq. (5.27). In agreement with the parallel model case, we observe that a moderate horizontal gene transfer rate reduces by orders of magnitude the fluctuations.

5.4 Conclusions

We formulated the master equations for quasispecies theories, Crow-Kimura and Eigen, in a finite population of binary sequences. By using a field theory formalism,
we derived analytical expressions for the Hamming distance class probability distributions and for their fluctuations. The theoretical expressions are valid both in the absence and in the presence of horizontal gene transfer. We compared the predictions of our theory, for the parallel model, with numerical results obtained from stochastic simulations based on a Lebowitz/Gillespie algorithm. Excellent agreement was obtained between theory and stochastic simulations.

For both the parallel and Eigen models, we find that horizontal gene transfer reduces by orders of magnitude the fluctuations in the probability distribution when fitness is represented by smooth functions, such as quadratic. For a discontinuous, sharp peak fitness, horizontal gene transfer does not modify the steady-state distribution of fluctuations.

From our stochastic simulations, we find that horizontal gene transfer also reduces the variability between independent experiments, for a smooth quadratic fitness function.

In conclusion, our theoretical results indicate that for quasispecies models in finite populations, horizontal gene transfer reduces by orders of magnitude the fluctuations, and hence the statistical properties of the finite population become closer to those of an infinite population. This effect can also be interpreted in terms of an improvement of the sampling of sequence space induced by horizontal gene transfer. Alternatively, increased horizontal gene transfer rates can be said to increase the effective popula-
tion size of the system. This reduction of the fluctuations should be observable in experiments.

**Appendix 1**

We consider details of the derivation for the parallel model. The action in the exponent of Eq. (5.9) is given by

\[
S[z^*, \{z\}] = \sum_{\xi=0}^{L} \left[ \sum_{k=0}^{t/\epsilon} z^{*}_\xi(k)z_\xi(k) - \sum_{k=1}^{t/\epsilon} z^{*}_\xi(k)z_\xi(k-1) - z_\xi(t/\epsilon) - n_\xi^0 \ln[z^{*}_\xi(0)] \right] 
- \epsilon \mu \sum_{\xi=0}^{L} \sum_{k=1}^{t/\epsilon} [(L - \xi)[z^{*}_{\xi+1}(k) - z^{*}_\xi(k)]z_\xi(k-1) + \xi[z^{*}_{\xi-1}(k) - z^{*}_\xi(k)]z_\xi(k-1)] 
- \frac{\epsilon}{N} \sum_{\xi, \xi'=0}^{L} \sum_{k=1}^{t/\epsilon} r(\xi)z^{*}_\xi(k)[z^{*}_\xi(k) - z^{*}_{\xi'}(k)]z_\xi(k-1)z_{\xi'}(k-1) 
- \epsilon \nu \sum_{\xi=0}^{L} \sum_{k=1}^{t/\epsilon} \rho_+(L - \xi)[z^{*}_{\xi+1}(k) - z^{*}_\xi(k)]z_\xi(k-1) 
+ \xi \rho_-[z^{*}_{\xi-1}(k) - z^{*}_\xi(k)]z_\xi(k-1) 
\]

(5.30)

After introducing the shift \(z^* \rightarrow 1 + \tilde{z}\), we obtain Eq. (5.10).

We look for a saddle-point in the action,
The solution for this saddle point equation is $\bar{z}_\xi(k) = 0, \forall \xi, k$.

The second saddle-point equation is
\[
\frac{\delta S}{\delta \bar{z}_\xi(k)} \bigg|_{c} = 0 = \bar{z}_\xi(k) - \bar{z}_\xi(k - 1)(1 - \delta_{k,0}) - \frac{n_\xi^0}{1 + \bar{z}_\xi(0)} \delta_{k,0}
- \epsilon \mu \left[ (L - \xi) \bar{z}_{\xi+1}(k - 1) + \xi \bar{z}_{\xi-1}(k - 1) - L \bar{z}_\xi(k - 1) \right] (1 - \delta_{k,0})
- \epsilon \nu \left[ \rho_+ (L - \xi) \bar{z}_{\xi+1}(k - 1) + \rho_- \xi \bar{z}_{\xi-1}(k - 1) \right] (1 - \delta_{k,0})
- \left\{ \left[ (L - \xi) \rho_+ + \xi \rho_- \right] \bar{z}_\xi(k - 1) \right\} (1 - \delta_{k,0})
- \frac{\epsilon}{N} \sum_{\xi_1, \xi_2 = 0}^{L} r(\xi_1) \left[ \delta_{\xi_1, \xi} \left[ \delta_{\xi_1, \xi}^e (k) - \bar{z}_{\xi_2}(k) \right] + [1 + \bar{z}_{\xi_1}(k)] \right]
\times \left( \delta_{\xi_1, \xi} - \bar{z}_{\xi_2}(k) \right) (1 - \delta_{k,0})
\] (5.32)

For $k = 0$, and recalling $\bar{z}_\xi(k) = 0$, Eq. (5.32) becomes $\bar{z}_\xi(0) = n_\xi^0$, thus providing the initial condition. Notice that, by summing this equation over $\xi$, we obtain the
additional identity

\[
\sum_{\xi=0}^{L} z_{\xi}^c(k) = \sum_{\xi=0}^{L} z_{\xi}^c(k-1) = \text{constant} = \sum_{\xi=0}^{L} z_{\xi}^c(0) = \sum_{\xi=0}^{L} n_{\xi}^0 = N
\]  

(5.33)

For \( k > 0 \), Eq. (5.32) becomes

\[
z_{\xi}^c(k) = z_{\xi}^c(k-1) + \epsilon \mu \left[ (L - \xi + 1)z_{\xi-1}^c(k-1) + (\xi + 1)z_{\xi+1}^c(k-1) - Lz_{\xi}^c(k-1) \right] \\
+ \epsilon \nu \left[ \rho_+(L - \xi + 1)z_{\xi-1}^c(k-1) + \rho_-(\xi + 1)z_{\xi+1}^c(k-1) \right] \\
- \{ \rho_+(L - \xi) + \rho_-(\xi) \} z_{\xi}^c(k-1) \\
+ \frac{\epsilon}{N} \left[ r(\xi)z_{\xi}^c(k-1) \sum_{\xi'=0}^{L} z_{\xi'}^c(k-1) - z_{\xi}^c(k-1) \sum_{\xi'=0}^{L} r(\xi')z_{\xi'}^c(k-1) \right]
\]

(5.34)

We define in Eq. (5.34) \( z_{\xi}^c(k) = NP_{\xi}(k) \), to obtain

\[
P_{\xi}(k) = P_{\xi}(k-1) + \epsilon \mu \left[ (L - \xi + 1)P_{\xi-1}(k-1) + (\xi + 1)P_{\xi+1}(k-1) - LP_{\xi}(k-1) \right] \\
+ \epsilon \nu \left[ \rho_+(L - \xi + 1)P_{\xi-1}(k-1) + \rho_-(\xi + 1)P_{\xi+1}(k-1) \right] \\
- \{ (L - \xi)\rho_+ + \xi \rho_- \} P_{\xi}(k) \\
+ \epsilon \left[ r(\xi) - \sum_{\xi'=0}^{L} r(\xi')P_{\xi'}(k-1) \right] P_{\xi}(k-1)
\]

(5.35)

with the initial condition \( P_{\xi}(0) = n_{\xi}^0/N \), and after Eq. (5.33), the conservation of probability \( \sum_{\xi=0}^{L} P_{\xi}(k) = 1, \forall k \). After taking the continuous time limit \( \epsilon \to 0 \), Eq. (5.35) becomes Eq. (5.11).

We next consider the expansion of the action Eq. (5.31) near the saddle-point \( S_c \).
We define \( \delta z_\xi(k) = z_\xi(k) - z_\xi(k) \), and \( \delta \tilde{z}_\xi(k) = \tilde{z}_\xi(k) - \tilde{z}_\xi(k) \). This gives

\[
S - S_c = \sum_{\xi, \xi'=0}^L \left[ \delta \tilde{z}_\xi(0) \delta \tilde{z}^{\xi'}(0) \delta_{\xi,\xi'} + \frac{1}{2} n_\xi^0 \delta \tilde{z}_\xi(0) \delta \tilde{z}^{\xi'}(0) \delta_{\xi,\xi'} \\
+ \sum_{k=1}^{t/\epsilon} \left\{ \delta \tilde{z}_\xi(k) \delta \tilde{z}^{\xi'}(k) \delta_{\xi,\xi'} - \epsilon \delta \tilde{z}_\xi(k) \delta \tilde{z}^{\xi'}(k) [\delta_{\xi,\xi'} r(\xi) N P_\xi(k - 1) \\
- r(\xi) N P_\xi(k - 1) P_{\xi'}(k - 1)] \right\} \\
+ \sum_{k=1}^{t/\epsilon} \delta \tilde{z}_\xi(k) \delta \tilde{z}^{\xi'}(k - 1) \{ - \delta_{\xi,\xi'} - \epsilon [n_\xi - \xi + 1) \delta \tilde{z}_{\xi - 1}\xi' + (\xi + 1) \delta \tilde{z}_{\xi + 1}\xi' - L \delta_{\xi,\xi'} \} \\
- \epsilon [\{ r(\xi) - \sum_{\xi} r(\xi) P_\xi(k - 1) \} \delta_{\xi,\xi'} + (r(\xi) - r(\xi')) P_\xi(k - 1)] \right\} \\
- \epsilon \nu \sum_{\xi=0}^L \sum_{k=1}^{t/\epsilon} \left[ \rho_\xi (L - \xi) \delta \tilde{z}_{\xi + 1}(k) + \rho_{-\xi} \delta \tilde{z}_{\xi - 1}(k) \\
- \{ \rho_\xi (L - \xi) + \rho_{-\xi} \} \delta \tilde{z}_\xi(k) \right] \delta z_\xi(k - 1) + O[(\delta \tilde{z}, \delta z)^3] \\
= \frac{1}{2} X^T \Pi^{-1} X + O(X^3)
\]

(5.36)

Here, we defined \( X = (\{ \delta \tilde{z}(0), \delta z(0) \}, \ldots, \{ \delta \tilde{z}(t/\epsilon), \delta z(t/\epsilon) \}) \)

The matrix \( \Pi^{-1} \) is banded tri-diagonal, with

\[
\Pi^{-1} = \begin{pmatrix}
\Pi_{00}^{-1} & -\Pi_{01}^{-1} & 0 & 0 & \cdots & 0 \\
-\Pi_{10}^{-1} & \Pi_{11}^{-1} & -\Pi_{12}^{-1} & 0 & \cdots & 0 \\
0 & -\Pi_{12}^{-1} & \Pi_{22}^{-1} & -\Pi_{23}^{-1} & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \Pi_{t/\epsilon,t/\epsilon}^{-1}
\end{pmatrix}
\]

(5.37)
where

$$\Pi_{\nu_0}^{-1} = \begin{pmatrix} N^0 & I \\ I & 0 \end{pmatrix}, \quad [N^0]_{\xi,\xi'} = n^0_{\xi} \delta_{\xi,\xi'}$$

$$\Pi_{k,k}^{-1} = \begin{pmatrix} -\epsilon B(k-1) & I \\ I & 0 \end{pmatrix}, \quad k \neq 0$$

$$\Pi_{k,k-1}^{-1} = \begin{pmatrix} 0 & I + \epsilon A(k-1) \\ 0 & 0 \end{pmatrix}$$

$$\Pi_{k-1,k}^{-1} = \begin{pmatrix} 0 & 0 \\ I & \epsilon A^T(k-1) \end{pmatrix}$$

The matrices $A$ and $B$ are defined by

$$[A(k)]_{\xi,\xi'} = \mu[(L - \xi + 1)\delta_{\xi-1,\xi'} + (\xi+1)\delta_{\xi+1,\xi'} - L\delta_{\xi,\xi'}]$$

$$+ \nu\rho_+(L - \xi + 1)\delta_{\xi-1,\xi'} + \rho_-(\xi + 1)\delta_{\xi+1,\xi'} - \{(L - \xi)\rho_+ + \xi\rho_-\}\delta_{\xi,\xi'}$$

$$+ \delta_{\xi,\xi'}[r(\xi) - \sum_{\xi_i} r(\xi_i)P_{\xi_i}(k)] + [r(\xi) - r(\xi')]P_{\xi}(k)$$

with $A$ a symmetric matrix $[A^T(k)]_{\xi,\xi'} = [A(k)]_{\xi',\xi}$.

$$[B(k)]_{\xi,\xi'} = \delta_{\xi,\xi'}2r(\xi)NP_{\xi}(k) - [r(\xi) + r(\xi')]NP_{\xi}(k)P_{\xi}(k)$$
We use the matrix inversion formula

\[
\begin{bmatrix}
A & B \\
C & D
\end{bmatrix}^{-1} = \begin{bmatrix}
A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\
-(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1}
\end{bmatrix}
\]

(5.42)

We notice that

\[
\Pi(t/\epsilon) = \left[ \Pi^{-1}(t/\epsilon) \right]^{-1} = \begin{bmatrix}
[\Pi^{-1}(t/\epsilon - 1)] & \begin{pmatrix}
0 \\
0 \\
\vdots \\
-\Pi^{-1}_{t/\epsilon - 1,t/\epsilon}
\end{pmatrix}
\end{bmatrix}
\]

(5.43)

Calculating the inverse in Eq. (5.43), we obtain

\[
\begin{bmatrix}
\Pi(t/\epsilon)
\end{bmatrix}_{t/\epsilon,t/\epsilon} = b_{t/\epsilon,t/\epsilon}
\]

\[
= \Pi^{-1}_{t/\epsilon,t/\epsilon} - \left( 0 0 \ldots - \Pi^{-1}_{t/\epsilon,t/\epsilon - 1} \right) \begin{bmatrix}
\Pi(t/\epsilon - 1)
\end{bmatrix}^{-1} \begin{pmatrix}
0 \\
0 \\
\vdots \\
-\Pi^{-1}_{t/\epsilon - 1,t/\epsilon}
\end{pmatrix}
\]

\[
= \left[ \Pi^{-1}_{t/\epsilon,t/\epsilon} - \Pi^{-1}_{t/\epsilon,t/\epsilon - 1}b_{t/\epsilon - 1,t/\epsilon - 1}\Pi^{-1}_{t/\epsilon - 1,t/\epsilon} \right]^{-1}
\]

(5.44)
From this recursive equation, we find

\[
\begin{align*}
    b_{00} &= \left[ \Pi_{00}^{-1} \right]^{-1} = \begin{pmatrix} 0 & I \\ I & -N^0 \end{pmatrix} \\
    b_{11} &= \left[ \Pi_{11}^{-1} - \Pi_{10}^{-1} b_{00} \Pi_{01}^{-1} \right]^{-1} \\
          &= \begin{pmatrix} 0 & I \\ I & \{I + \epsilon A(0)\}\{-N^0\}\{I + \epsilon A^T(0)\} + \epsilon B(0) \end{pmatrix}
\end{align*}
\] (5.45)

From Eq. (5.45), proceeding by induction, we prove that the matrices \( b_k \) possess the structure

\[
b_{k,k} = \begin{pmatrix} 0 & I \\ I & C(k) \end{pmatrix}
\] (5.46)

which after the recursion relation

\[
b_{k,k} = \left[ \Pi_{k}^{-1} - \Pi_{k,k-1}^{-1} b_{k-1,k-1} \Pi_{k-1,k}^{-1} \right]^{-1}
\] (5.47)

implies the following equations for \( C(k) \)

\[
\begin{align*}
    C(k) &= [I + \epsilon A(k - 1)] C(k - 1) [I + \epsilon A^T(k - 1)] + \epsilon B(k - 1) \\
    C(0) &= -N^0
\end{align*}
\] (5.48)

In the continuous time limit, Eq. (5.48) becomes a Lyapunov equation

\[
\frac{d}{dt} C = \begin{pmatrix} B + AC + CA^T \end{pmatrix} \\
C(0) = -N^0
\] (5.49)

with \([N^0]_{\xi,\xi'} = \delta_{\xi,\xi'} n^0_\xi\).
Appendix 2

We consider details of the derivation for the Eigen model. In the Eigen model, we obtain the action

\[
S\{\{z^*\}, \{z\}\} = \left(1 - \frac{\nu}{L}\right) \left\{ \sum_{\xi=0}^{L} \left[ \frac{1}{2} z^*_\xi(0) z_\xi(0) - n^0_\xi \ln[z^*_\xi(0)] + \frac{1}{2} z^*_\xi(t/e) z_\xi(t/e) - z_\xi(t/e) \right] + \sum_{\xi=0}^{L} \sum_{k=1}^{t/e} \frac{1}{2} \bigg[ z^*_\xi(k) [z_\xi(k) - z_\xi(k - 1)] - [z^*_\xi(k) - z^*_\xi(k - 1)] z_\xi(k - 1) \bigg] + \frac{\epsilon}{N} \sum_{k=1}^{t/e} \sum_{\xi_\xi', \xi''=0}^{L} Q_{\xi_\xi', \xi''}(\xi) z^*_\xi(k) [z^*_\xi'(k) - z^*_\xi''(k)] z_\xi(k - 1) z_\xi''(k - 1) \right\}.
\]

It is convenient to introduce in Eq. (5.50) the change of variables \(z^* \rightarrow 1 + \bar{z}\), and
then the action becomes

\[
S[z, \bar{z}] = \left(1 - \frac{\nu}{L}\right) \left\{ \sum_{\xi=0}^{L} \left[ \bar{z}_{\xi}(0)z_{\xi}(0) - n_{\xi}^{0}\ln[1 + \bar{z}_{\xi}(0)] + \sum_{k=1}^{t/\epsilon} \bar{z}_{\xi}(k)\{z_{\xi}(k) - z_{\xi}(k - 1)\} \right] \\
- \frac{\epsilon}{N} \sum_{k=1}^{t/\epsilon} \sum_{\xi, \xi', \xi'' = 0}^{L} Q_{\xi', \xi} r(\xi)[1 + \bar{z}_{\xi}(k)][\bar{z}_{\xi'}(k) - \bar{z}_{\xi''}(k)]z_{\xi}(k - 1)z_{\xi''}(k - 1) \right\}
\]

\[
- \frac{\epsilon}{N} \sum_{k=1}^{t/\epsilon} \sum_{\xi, \xi', \xi'' = 0}^{L} d(\xi)[1 + \bar{z}_{\xi}(k)][\bar{z}_{\xi}(k) - \bar{z}_{\xi'}(k)]z_{\xi}(k - 1)z_{\xi'}(k - 1)
\]

\[
- \frac{\nu}{L} \frac{\epsilon}{N} \sum_{k=1}^{t/\epsilon} \sum_{\xi, \xi', \xi'' = 0}^{L} [Q_{\xi', \xi} + 1] r(\xi)[1 + \bar{z}_{\xi}(k)]
\]

\[
\times [\bar{z}_{\xi'}(k) - \bar{z}_{\xi''}(k)]z_{\xi}(k - 1)z_{\xi'}(k - 1)
\]

(5.51)

We look for a saddle-point in the action Eq. (5.51)

\[
\frac{\delta S}{\delta \bar{z}_{\xi}(k)} = 0 = \left(1 - \frac{\nu}{L}\right) \left\{ \bar{z}_{\xi}(k) - \bar{z}_{\xi}(k + 1)(1 - \delta_{k,t/\epsilon}) \\
- \frac{\epsilon}{N} \left[ \sum_{\xi_1, \xi_2, \xi_3 = 0}^{L} Q_{\xi_2, \xi_1} r(\xi_1)[1 + \bar{z}_{\xi_1}(k)][\bar{z}_{\xi_2}(k) - \bar{z}_{\xi_3}(k)] \right. \\
\times \left\{ \delta_{\xi_1, \xi_2} z_{\xi_3}(k - 1) + \bar{z}_{\xi_1}(k - 1)\delta_{\xi_1, \xi_3} \right\} \right\} \\
- \frac{\epsilon}{N} \sum_{\xi_1, \xi_2 = 0}^{L} d(\xi_2)[1 + \bar{z}_{\xi_1}(k)][\bar{z}_{\xi_2}(k) - \bar{z}_{\xi_1}(k)]\delta_{\xi_1, \xi_2} z_{\xi_2}(k - 1)
\]

\[
+ \bar{z}_{\xi_1}(k - 1)\delta_{\xi_1, \xi_2} (1 - \delta_{k,t/\epsilon})
\]

\[
- \frac{\nu}{L} \frac{\epsilon}{N} \sum_{k=1}^{t/\epsilon} \sum_{\xi_1, \xi_2, \xi_3 = 0}^{L} \left[ Q_{\xi_2, \xi_1} + 1 \right] r(\xi_1)[1 + \bar{z}_{\xi_1}(k)]
\]

\[
\times [\bar{z}_{\xi_2}(k) - \bar{z}_{\xi_3}(k)]\delta_{\xi_1, \xi} z_{\xi_3}(k - 1) + \delta_{\xi_1, \xi} z_{\xi_1}(k - 1)
\]

(5.52)

Eq. (5.52) has the saddle point solution \( \bar{z}_{\xi}(k) = 0, \forall \xi, k. \)
The second saddle-point equation is

\[
\frac{\delta S}{\delta z_\xi(k)} c = 0 = \left(1 - \frac{\nu}{L}\right) \left\{ z_\xi(k) - z_\xi(k-1)(1 - \delta_{k,0}) - n_0^0 \delta_{k,0} \right. \\
- \frac{\epsilon}{N} \sum_{\xi_1,\xi_2 = 0}^L [Q_{\xi_2,\xi_1} r(\xi_1) z_{\xi_1}^c (k-1) z_{\xi_2}^c (k-1)](1 - \delta_{k,0}) \\
- Q_{\xi_2,\xi_1} r(\xi_1) z_{\xi_1}^c (k-1) z_{\xi_1}^c (k-1)(1 - \delta_{k,0}) \right\} \\
- \frac{\epsilon}{N} \sum_{\xi_2 = 0}^L d(\xi_2) z_\xi^c (k-1) z_{\xi_2}^c (k-1) - \sum_{\xi_1 = 0}^L d(\xi) z_{\xi_1}^c (k-1) z_{\xi_2}^c (k-1) \right\} (1 - \delta_{k,0}) \\
- \frac{\nu \epsilon}{L N} \sum_{\xi_1,\xi_2,\xi_3 = 0}^L [Q_{\xi_3,\xi_1+1} r_+ (L - \xi_1) + Q_{\xi_2,\xi_1-1} r_- \xi_1] r(\xi_1) \\
\times \{ z_{\xi_2}^c (k) - z_{\xi_3}^c (k) \} [1 + z_{\xi_1}^c (k)] (\delta_{\xi_2,\xi_3} - \delta_{\xi_3,\xi_2}) \} z_{\xi_1}^c (k-1) z_{\xi_3}^c (k-1) \\
\] (5.53)

As in the parallel model, we define \( z_\xi = NP_\xi \), and after taking the continuous time limit \( \epsilon \to \infty \), we obtain the differential equation for the probability distribution

\[
\frac{d}{dt} P_\xi = \left(1 - \frac{\nu}{L}\right) \left[ \sum_{\xi' = 0}^L Q_{\xi,\xi'} r(\xi') P_{\xi'} - P_\xi \sum_{\xi' = 0}^L r(\xi') P_{\xi'} \right] - P_\xi \left[ d(\xi) - \sum_{\xi' = 0}^L P_{\xi'} d(\xi') \right] \\
+ \frac{\nu}{L} \left[ \sum_{\xi' = 0}^L \{Q_{\xi,\xi'+1} r_+ (L - \xi') + Q_{\xi,\xi'-1} r_- \xi' \} r(\xi') P_{\xi'} \right] - \left[ \sum_{\xi' = 0}^L \{ r_+ (L - \xi') + \rho_- \xi' \} r(\xi') P_{\xi'} \right] \\
\] (5.54)
We expand the action Eq. (5.51) near the saddle point, to obtain

\[
S - S_c = \left(1 - \frac{\nu}{L}\right) \sum_{\xi, \xi'} \left[ \sum_{k=1}^{t/e} \delta z(0) \{ \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \}
\]

\[
+ \frac{1}{2} n_\xi \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) 
\]

\[
- \epsilon \sum_{k=1}^{t/e} \left\{ \left( \sum_{\xi_1} r(\xi_1) P_{\xi_1} \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \right) \right\}
\]

\[
+ \left( \sum_{\xi_1} r(\xi_1) P_{\xi_1} \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \right)
\]

\[
- \left( \sum_{\xi_1} r(\xi_1) P_{\xi_1} \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \right)
\]

\[
- \epsilon \sum_{k=1}^{t/e} \left[ \left( \sum_{\xi_1} \delta z(0) P_{\xi_1} \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \right) \right]
\]

\[
+ \left( \sum_{\xi_1} \delta z(0) P_{\xi_1} \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \right)
\]

\[
- \left( \sum_{\xi_1} \delta z(0) P_{\xi_1} \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \right)
\]

\[
- \nu \epsilon \sum_{k=1}^{t/e} \sum_{\xi_1, \xi_2, \xi_3} \left\{ \left( \sum_{\xi_1} \delta z(0) P_{\xi_1} \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \right) \right\}
\]

\[
+ \left( \sum_{\xi_1} \delta z(0) P_{\xi_1} \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \right)
\]

\[
- \left( \sum_{\xi_1} \delta z(0) P_{\xi_1} \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \delta z(0) \right)
\]

\[
+ \frac{1}{2} X^T \Pi^{-1} X + O(X^3)
\]  

(5.55)

Here, we defined \( X^T = \{ (\delta z(0), (\delta z(0)), \ldots, (\delta z(t/e), (\delta z(t/e))) \} \). The matrix \( \Pi^{-1} \) is traditional by blocks, as in the case of the parallel model. A similar analysis holds.
for the Eigen model as well, with matrices $A$ and $B$ defined as

$$
[B(k)]_{\xi,\xi'} = \left(1 - \frac{\nu}{L}\right) \left[Q_{\xi',\xi} r(\xi) N P_{\xi}(k) + Q_{\xi,\xi'} r(\xi') N P_{\xi'}(k) \right]
$$

$$
- \left( r(\xi) + r(\xi') \right) N P_{\xi}(k) P_{\xi'}(k)
$$

$$
+ 2 \left( \sum_{\xi_1=0}^{L} d(\xi_1) P_{\xi_1}(k) \right) N P_{\xi}(k) \delta_{\xi,\xi'} - \left( d(\xi) + d(\xi') \right) N P_{\xi}(k) P_{\xi'}(k)
$$

$$
+ \frac{\nu}{L} \left[ \left( Q_{\xi,\xi+1} + Q_{\xi,\xi-1} \right) r(\xi) N P_{\xi}(k) \right]
$$

$$
+ \left( Q_{\xi,\xi'} + Q_{\xi',\xi} \right) r(\xi') N P_{\xi'}(k)
$$

$$
- \left( r(\xi + \xi') N P_{\xi}(k) P_{\xi'}(k) \right)
$$

$$
- \left( r(\xi' + \xi) N P_{\xi}(k) P_{\xi'}(k) \right)
$$

$$
(5.56)
$$

$$
[A(k)]_{\xi,\xi'} = \left(1 - \frac{\nu}{L}\right) \left[ \sum_{\xi''=0}^{L} Q_{\xi,\xi''} r(\xi'') P_{\xi''}(k) + Q_{\xi,\xi'} r(\xi') \right]
$$

$$
- \delta_{\xi,\xi'} \sum_{\xi''=0}^{L} r(\xi'') P_{\xi''}(k) - r(\xi') P_{\xi}(k) \right]
$$

$$
+ d(\xi') P_{\xi}(k) + \delta_{\xi,\xi'} \sum_{\xi_1=0}^{L} d(\xi_1) P_{\xi_1}(k) - d(\xi) \delta_{\xi,\xi'} - d(\xi) P_{\xi}(k)
$$

$$
+ \frac{\nu}{L} \left[ \sum_{\xi''=0}^{L} \left( Q_{\xi,\xi''+1} + Q_{\xi,\xi''-1} \right) r(\xi'') P_{\xi''}(k) \right]
$$

$$
+ \left( Q_{\xi,\xi'} + Q_{\xi',\xi} \right) r(\xi')
$$

$$
- \delta_{\xi,\xi'} \sum_{\xi''=0}^{L} \left( r(\xi'') \right) P_{\xi''}(k)
$$

$$
- \left( r(\xi' + \xi') P_{\xi}(k) \right)
$$

$$
(5.57)
$$
The recursion relation Eq. (5.47) yields a Lyapunov equation in the Eigen model as well, of the form Eq. (5.49) for the matrix $C$, which in the continuous time limit is

$$\frac{d}{dt} C = B + AC + CA^T$$  \hspace{1cm} (5.58)

with initial condition $C_{\xi,\xi'} = -\delta_{\xi,\xi'} n_{\xi}$. 

**Appendix 3**

We are interested in the following observables from this theory, the average particle number $\langle n_\xi \rangle$ in the Hamming distance class $C_\xi$, and the correlator $\langle n_\xi n_{\xi'} \rangle$.

For the average number of individuals with sequences in class $C_\xi$, we have

$$\langle n_\xi \rangle = \langle \hat{a}_\xi^\dagger \hat{a}_\xi \rangle = \langle z_\xi(t/\epsilon) \rangle = z_\xi^\epsilon(t/\epsilon)$$

$$= NP_\xi$$  \hspace{1cm} (5.59)

Hence, we obtain

$$P_\xi = \frac{1}{N} \langle n_\xi \rangle$$  \hspace{1cm} (5.60)

For the correlator in the number of individuals with sequences in classes $C_\xi$ and
Therefore, from Eq. (5.59) and Eq. (5.61), we have

\[
\langle (\delta n_\xi)^2 \rangle = \langle n_\xi^2 \rangle - \langle n_\xi \rangle^2 = NP_\xi + C_{\xi,\xi}
\]

and from Eq. (5.62) we finally obtain the fluctuations in the probability distribution

\[
\frac{1}{N^2} \langle (\delta n_\xi)^2 \rangle = \frac{1}{N} \left( P_\xi + \frac{1}{N} C_{\xi,\xi} \right)
\]
5.5 Appendix 4

We consider the sharp peak fitness function in the parallel model. The probability equation for the sharp peak fitness \( f(\xi) = A\delta_{\xi,L} \) in the parallel model is

\[
\frac{d}{dt} P_\xi = L A \delta_{\xi,L} P_\xi - P_\xi L A P_L \\
+ \mu \left[ (L - \xi + 1) P_{\xi-1} + (\xi + 1) P_{\xi+1} - L P_\xi \right] \\
+ \nu \left[ \rho_+ (L - \xi + 1) P_{\xi-1} + \rho_- (\xi + 1) P_{\xi+1} - \{\rho_+ (L - \xi) + \rho_- \xi\} P_\xi \right]
\]

(5.64)

For large \( L \),

\[
\frac{d}{dt} P_L = L A P_L (1 - P_L) + \mu [P_{L-1} - LP_L] + \nu [\rho_+ P_{\xi-1} - \rho_- LP_L] \\
\approx L A P_L (1 - P_L) - L (\mu + \nu \rho_-) P_L
\]

(5.65)

The stationary solution is obtained from the equation

\[
P_L [A - (\mu + \nu \rho_-) - A P_L] = 0
\]

(5.66)

and is given by

\[
P_L = \begin{cases} 
0, & \frac{\mu + \nu \rho_-}{A} > 1 \\
1 - \frac{\mu + \nu \rho_-}{A}, & \frac{\mu + \nu \rho_-}{A} < 1 \\
\end{cases}
\]

(5.67)
This result is consistent with [40, 127], and given that \( \rho_- = (1 - u)/2 \) and \( u = 1 - O(L^{-1}) \) indicates that horizontal gene transfer does not affect the steady-state wild-type probability.

For the correlation matrix element \( C_{L,L} \), we have the equation

\[
\frac{d}{dt} C_{L,L} = 2LANP_L - 2LANP_L^2 \\
+ 2 \sum_{\xi_1=0}^{L} \left[ \mu \{ \delta_{L-1,\xi_1} - L\delta_{L,\xi_1} \} C_{\xi_1,L} + \nu \{ \rho_+ \delta_{L-1,\xi_1} - \rho_- L\delta_{L,\xi_1} \} C_{\xi_1,L} \\
+ \delta_{L,\xi_1} (LA - LAP_L) C_{\xi_1,L} + (LA - LA\delta_{\xi_1,L}) P_L C_{\xi_1,L} \right] \\
= 2LANP_L(1 - P_L) + 2\mu(C_{L-1,L} - LC_{L,L}) \\
+ 2\nu(\rho_+ C_{L-1,L} - \rho_- LC_{L,L}) + 2LA(1 - P_L)C_{L,L} \\
+ 2LAP_L \sum_{\xi_1=0}^{L} C_{\xi_1,L} - 2LAP_L C_{L,L}
\]

(5.68)

The terms \( C_{L,L\pm1} \) are \( O(L^{-1}) \). We also notice that \( \sum_{\xi_1=0}^{L} C_{\xi_1,L} = -NP_L \), and find that the stationary solution of Eq. (5.69) is given by

\[
0 = LANP_L(1 - P_L) - \mu LC_{L,L} - \nu \rho_- LC_{L,L} \\
+ LA(1 - P_L)C_{L,L} - LANP_L^2 - LAP_L C_{L,L} \\
= ANP_L(1 - 2P_L) + [(A - \mu + \nu \rho_-) - 2AP_L]C_{L,L}
\]

(5.69)

From Eq. (5.67), we note that \( A - \mu + \nu \rho_- = A_P L \), and substituting into Eq. (5.69)
we obtain

\[ C_{L,L} = N(1 - 2P_L) \quad (5.70) \]

Therefore, for the fluctuations in the number of individuals with sequences in class \( C \), we apply the equations in Appendix 3 to find

\[
\frac{\langle \langle (\delta n_L)^2 \rangle \rangle}{N^2} = \frac{1}{N}[P_L + \frac{1}{N}C_{L,L}]
\]

\[
= \frac{1}{N}[P_L + (1 - 2P_L)]
\]

\[
= \frac{1}{N}(1 - P_L) \quad (5.71)
\]

Then, we finally obtain

\[
\frac{\langle \langle (\delta n_L)^2 \rangle \rangle}{N^2} = \begin{cases} 
0, & \frac{\mu + \nu \rho}{A} > 1 \\
\frac{1}{N} \frac{\mu + \nu \rho}{A}, & \frac{\mu + \nu \rho}{A} < 1 
\end{cases}
\]

\[
\approx \begin{cases} 
0, & \frac{\mu}{A} > 1 \\
\frac{1}{N} \frac{\mu}{A}, & \frac{\mu}{A} < 1 
\end{cases} \quad (5.72)
\]

### 5.6 Appendix 5

For the Eigen model in the sharp peak fitness \( f(\xi) = A_0 + (A - A_0)\delta_{\xi,L} \), we find that in the absence of horizontal gene transfer \( (\nu = 0) \) the steady state probability for the wild type class is given by the equation

\[
\sum_{\xi'=0}^{L} Q_{L,\xi'} f(\xi') P_{\xi'} - P_L \sum_{\xi'=0}^{L} f(\xi') P_{\xi'} = 0 \quad (5.73)
\]
with the transition matrix \( Q_{L,\xi'} = q^{\xi'} (1 - q)^{L-\xi'} \) from Eq. (5.18). By substituting in Eq. (5.73), we obtain

\[
\sum_{\xi' = 0}^{L} q^{\xi'} (1 - q)^{L-\xi'} f(\xi') P_{\xi'} - P_L [AP_L + A_0 \sum_{\xi' \neq L} P_{\xi'}] = 0
\]

(5.74)

Since \( q \approx 1 \), (the fidelity in the replication process is very high), then \( 1 - q \ll 1 \) and Eq. (5.74) becomes.

\[
q^{L} AP_L - P_L [(A - A_0) P_L + A_0] = 0
\]

(5.75)

Noting \( q^{L} \sim e^{-\mu} \), the solution for the wild-type probability, from Eq. (5.75) is therefore

\[
PL = \begin{cases} 
0, & e^{-\mu} < \frac{A_0}{A} \\
\frac{e^{-\mu} A - A_0}{A - A_0}, & e^{-\mu} \geq \frac{A_0}{A}
\end{cases}
\]

(5.76)

For the correlation matrix, we define \( D_{\xi,\xi'} = \frac{1}{N} C_{\xi,\xi'} \), and find that the stationary solution for \( D_{L,L} \) in the absence of degradation \( d(\xi) = 0 \) is given by

\[
0 = \frac{1}{N} B_{L,L} + \sum_{\xi_1 = 0}^{L} [A_{L,\xi_1} D_{\xi_1,L} + A_{L,\xi_1} D_{\xi_1,L}]
\]

(5.77)

From this equation, we find \( \sum_{\xi_1} A_{L,\xi_1} D_{\xi_1,L} = -\frac{1}{2N} B_{L,L} \). Hence, expanding the left hand side explicitly, we find

\[
\sum_{\xi_1 = 0}^{L} \left[ \sum_{\xi'' = 0}^{L} Q_{L,\xi''} f(\xi'') P_{\xi''} + Q_{L,\xi_1} f(\xi_1) - \left( \sum_{\xi' \neq \xi_1} f(\xi'_1) P_{\xi'_1} \right) \delta_{L,\xi_1} - f(\xi_1) P_L \right] D_{\xi_1,L}
= -[Q_{L,L} f(L) P_L - f(L) P_L^2]
\]

(5.78)
Expanding this equation when $L$ is large and $q \simeq 1$, we find

\[ [q^L A - (A - A_0) P_L - A_0 - (A - A_0) P_L] D_{L,L} = A P_L (P_L - q^L) + q^L A P_L^2 - A_0 P_L \]

(5.79)

Substituting the result $P_L = \frac{q^L A - A_0}{A - A_0}$ from Eq. (5.76), we find

\[ D_{L,L} = \frac{1}{(A - A_0)^2} [AA_0 - A_0^2 - (q^L A)^2 + q^L A A_0] \]

(5.80)

The fluctuation in the number of individuals with sequence in the wild-type class is find from the formulas in Appendix 3

\[ \frac{\langle (\delta n_L)^2 \rangle}{N^2} = \frac{1}{N} \left( P_L + D_{L,L} \right) = \frac{1}{N} \frac{q^L (1 - q^L) A^2}{(A - A_0)^2} \]

(5.81)

Finally, we note $q^L \sim e^{-\mu}$, and so

\[ \frac{\langle (\delta n_L)^2 \rangle}{N^2} = \frac{1}{N} \frac{e^{-\mu} (1 - e^{-\mu}) A^2}{(A - A_0)^2} \]

(5.82)
Chapter 6
Conclusions

In this doctoral thesis, several scenarios and generalizations of quasispecies theories, the Eigen and Crow-Kimura models, were studied by using field-theoretical methods. These powerful methods of statistical mechanics allowed us to obtain analytical solutions in many cases, which could be compared with numerical results generated by stochastic simulations.

We developed an extension of quasispecies models to include several mechanisms of horizontal gene transfer and recombination. Of particular interest is the derivation, for the first time, of exact expressions for the average composition and mean fitness of an infinite population of individuals undergoing two-parent recombination. Our analytical expressions also allowed us to prove the applicability of the mutational deterministic hypothesis in the context of quasispecies models, that is that the beneficial or detrimental effects of horizontal gene transfer and recombination depend on the sign of epistasis imposed by the fitness function.

We also presented extensions of quasispecies theory, originally formulated in the simplified purine/pyrimidine alphabet, to include the four letters alphabet characteristic of nucleic acids, DNA and RNA. We obtained analytical solutions and phase diagrams for different fitness functions in this case as well.
Finally, we extended the classical formulation of quasispecies to study finite populations. In this case, we formulated and solved by a field-theoretical method the master equation, obtaining a deterministic system of differential equations for the probability distribution and fluctuations in the average number of individuals with a given sequence type. In this context, in agreement with stochastic simulations, we discovered that horizontal gene transfer dramatically reduces the fluctuations.

Molecular evolution is still a vastly unexplored area in the biophysical sciences, and several questions remain to be answered. I hope that the four research projects presented in this doctoral thesis will constitute a small contribution for a better understanding of some of these problems.
References


85. P. C. Phillips, S. P. Otto, and M. C. Whitlock. The evolutionary importance of
gene interactions and variability of epistatic effects. In J. B. Wolf, E. D. Brodie
The evolutionary importance of gene interactions and variability of epistatic

86. R. D. Kouyos, S. P. Otto, and S. Bonhoeffer. Effect of varying epistasis on the


2007.

89. U. Liberman and M. Feldman. On the evolution of epistasis III: The haploid

90. R. L. Malmberg. The evolution of epistasis and the advantage of recombination


mutations in haploid strains of the yeast *Saccharomyces cerevisiae*. *J. Evol.


2005.


96. E. Baake and H. Wagner. Mutation-selection models solved exactly with meth­


107. There was a typo in [27] for the sharp peak fitness case: the formula reads $P_0 = 1 - \mu/A$ [instead of $u$, which is unity to $O(1/N)$]. Similarly for the Eigen model $P_0 = (Ae^{-\mu} - A_0)/(A - A_0)$ (rather than $u$, which is unity).

108. The alternative choice $u\xi_c < u^2$, combined with the self-consistency condition, leads to the equation $\max_{-1 \leq \xi_c \leq 1} f(\xi_c) = f(u)$, whose unique solution for the quadratic fitness is $|\xi_c| = |u| = 1$, in contradiction with the assumption $u\xi_c < u^2$. Hence, we necessarily have $u\xi_c \geq u^2$. 

305


111. For the case $\phi_c = 1$ there is an alternative solution $\eta_c = 1$. This implies the equation $u\xi_c = u[f'(\xi_c) - d'(\xi_c)] = 0$, whose solution is $u = 0$, unless there is an absolute maxima for $f(\xi_c) - d(\xi_c)$ in the interior of the region $|\xi_c| < 1$, both of which possibilities are contained in Eq. (104).


119. Only the solution with '+' in $\eta_c$ is chosen, since the corresponding exponential term $\exp(+\sqrt{(1+3\xi_c)(1-\xi_c)})$ provides a monotonic increasing contribution in the fitness Eq. (118), which is to be maximized.


