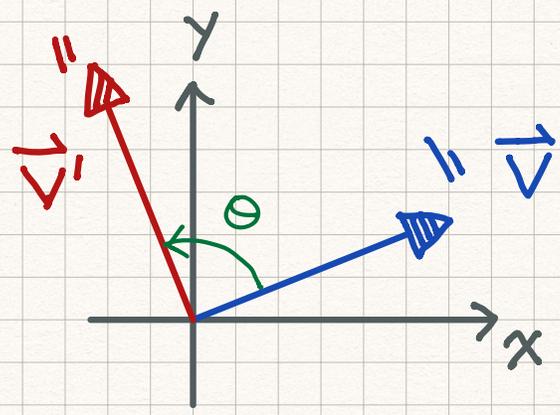


# Linear Algebra an introduction

Vectors and linear operators are commonly used in theoretical analyses. Most of the times, linear operators can be represented by matrices. 🧐

## example 2D rotations



Linear operator  $R(\theta)$ : rotate the vector by the angle  $\theta$  — counterclockwise

$$\vec{V}' = R(\theta) \vec{V}$$

But, this expression can be rather abstract for most newcomers ... We can write down the rotation operator in matrix form ☺

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

— "orthogonal" matrix

Set  $\theta = \frac{\pi}{2}$ . The rotation matrix is simple ~

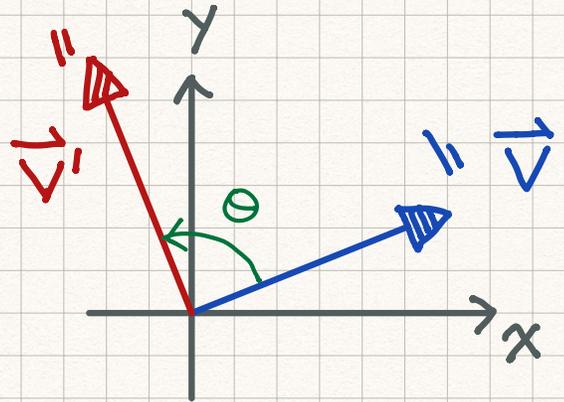
$$R\left(\frac{\pi}{2}\right) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad \text{Write the vectors in column}$$

form and the abstract relation becomes concrete.

$$\begin{pmatrix} v'_x \\ v'_y \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \end{pmatrix}$$

$$v'_x = -v_y$$

$$v'_y = v_x$$

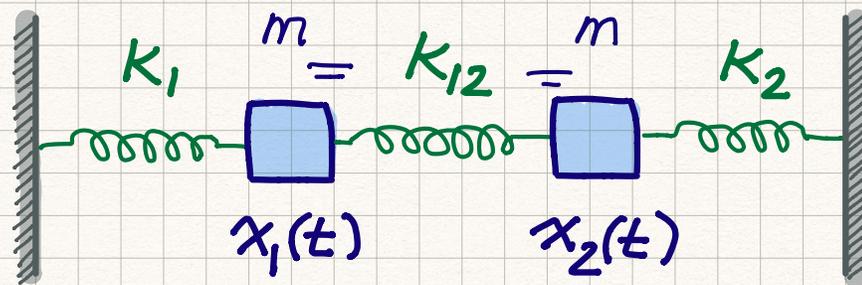


The rotation (a linear operator) can thus be represented by a matrix  $\hat{U}$

Many properties can be verified by the matrix algebra as well.

$$\begin{aligned} R(\theta_1)R(\theta_2) &= \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 & -\cos\theta_1\sin\theta_2 - \sin\theta_1\cos\theta_2 \\ \cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2 & \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{pmatrix} = R(\theta_1+\theta_2) \quad \text{YES } \hat{U} \end{aligned}$$

# Coupled Oscillators



Write down EOM for the oscillators ~

$$m \frac{d^2 x_1}{dt^2} = -K_1 x_1 + K_{12} (x_2 - x_1)$$

$$m \frac{d^2 x_2}{dt^2} = -K_2 x_2 - K_{12} (x_2 - x_1)$$

$$m \begin{pmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{pmatrix} = - \begin{pmatrix} K_1 + K_{12} & -K_{12} \\ -K_{12} & K_2 + K_{12} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

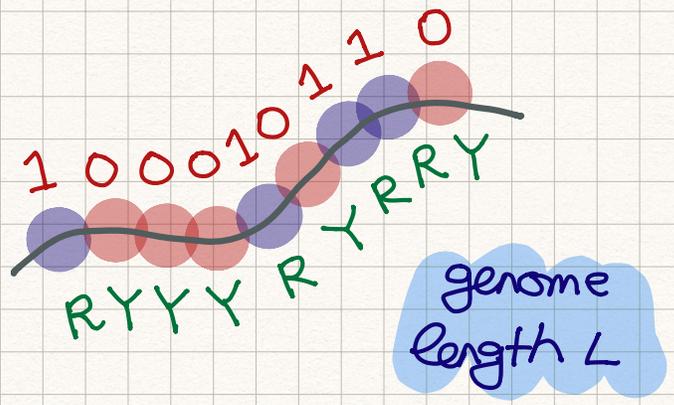
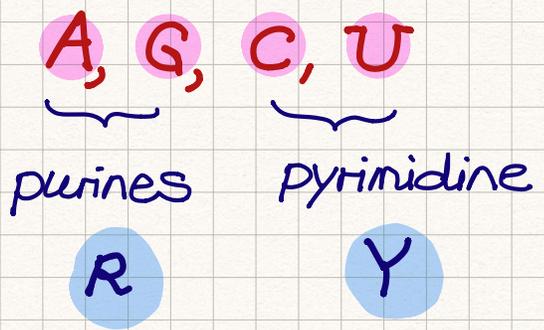
The spring constant turns into a real and symmetric  $2 \times 2$  matrix  $K$ ,

$$K = \begin{pmatrix} K_1 + K_{12} & -K_{12} \\ -K_{12} & K_2 + K_{12} \end{pmatrix}$$

If the coupling  $K_{12} = 0$ , the matrix is diagonal so that  $x_1(t)$  &  $x_2(t)$  are decoupled.

# Origin of Life? — a molecular perspective

nucleotides in RNA



There are  $2^L$  sequences with diff. fitness

$$\frac{dN_i}{dt} = f_i N_i \quad \rightarrow \quad N_i(t) = N_i(0) e^{f_i t}$$

Upon selection pressure, the species with max fitness prevails ☺

relative population —  $x_i(t) = \frac{N_i(t)}{N(t)}$  — total population

$\sum_i N_i(t)$

Suppose  $f_0 = \max \{f_i\}$  in the fitness landscape.

It is easy to see that

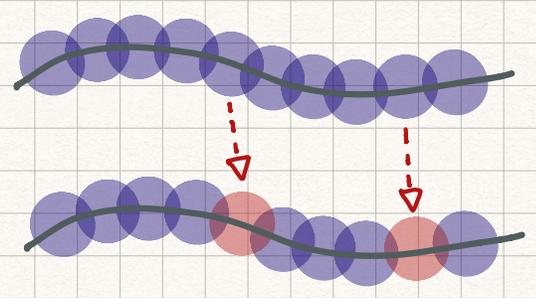
$$x_0(t \rightarrow \infty) = 1 \quad \text{and} \quad x_i(t \rightarrow \infty) = 0$$

↓  
 $i \neq 0$

# Error Threshold

mutations in real life

Assuming point mutations dominate



mutation probability  
 $u^2 (1-u)^{L-2}$

One can introduce a "distance" between two sequences  $d_{ij} = d_{ji}$  by point mutations.

$$m_{ij} = u^{d_{ij}} (1-u)^{L-d_{ij}}$$

mutation matrix

Sum rule of the mutation matrix  $m_{ij} \sim$

$$\sum_i m_{ij} = \sum_i u^{d_{ij}} (1-u)^{L-d_{ij}}$$

$$= [u + (1-u)]^L = 1$$

why?

The EOM of species evolution now becomes

$$\frac{dN_i}{dt} = \sum_j \underbrace{m_{ij} f_j}_{F_{ij}} N_j = \sum_j F_{ij} N_j$$

$F_{ij}$  fitness matrix

# What happens after evolutionary competitions?

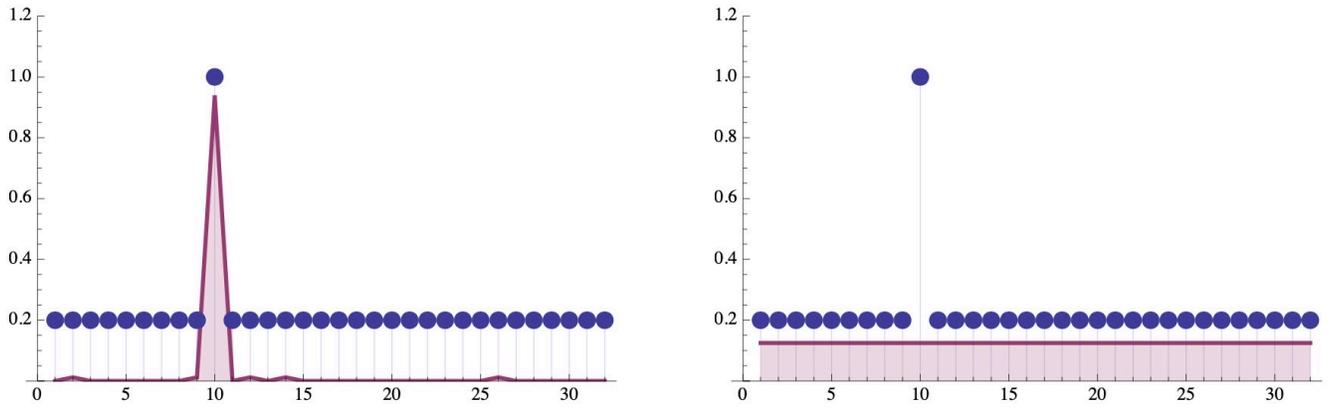


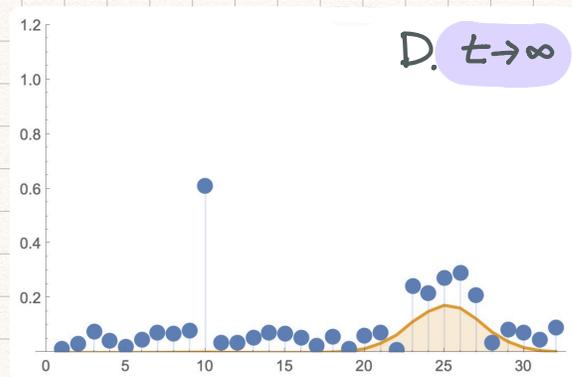
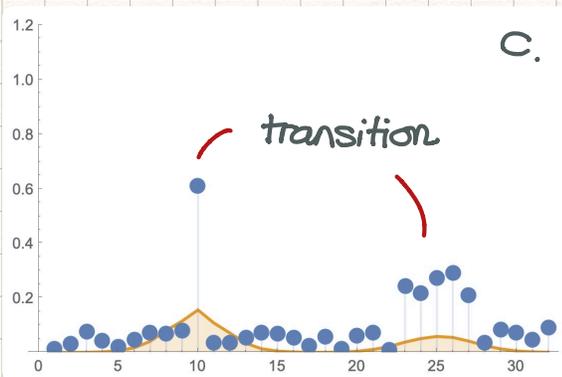
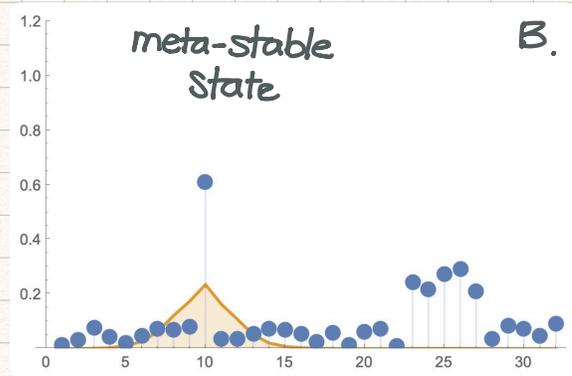
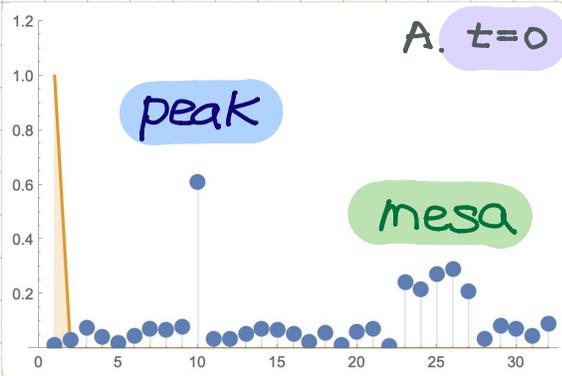
Figure 1: For the single-peak fitness landscape, there exists a mutation threshold  $u_c$ . For  $u < u_c$ , frequency profile is localized near the fitness peak. On the other hand, for  $u > u_c$ , an extended state (not necessarily uniform) emerges and the notion of quasi-species no longer exists.

### Error threshold in Eigen theory

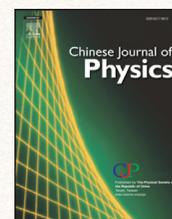
When the mutation rate is small ( $u < u_c$ ),  
 a stable population distribution emerges in  
 the long run  $\rightarrow$  quasi-species!

When the mutation rate is large ( $u > u_c$ ),  
 the quasi-species is unstable and we are  
 left with an ocean of chemical soup!

# Competitions between and quasi-species



When the mutation rate is not-too-small yet not-too-large, the winner is NOT the QS with highest fitness. It is the QS with reasonable tolerance will dominate in the end.



# Variation approach to error threshold in generic fitness landscape



Ching-I Huang<sup>a</sup>, Min-Feng Tu<sup>a</sup>, Hsiu-Hau Lin<sup>a,\*</sup>, Chun-Chung Chen<sup>b,c,\*</sup>

<sup>a</sup> Department of Physics, National Tsing Hua University, Hsinchu 30013, Taiwan

<sup>b</sup> Physics Division, National Center for Theoretical Sciences, Hsinchu 300, Taiwan

<sup>c</sup> Institute of Physics, Academia Sinica, Taipei 115, Taiwan

## ARTICLE INFO

### Article history:

Received 30 November 2016

Revised 16 March 2017

Accepted 13 April 2017

Available online 18 April 2017

### Keywords:

Evolutionary dynamics

Eigen model

Error threshold

Quasispecies theory

Fitness landscape

Peak–mesa competition

## ABSTRACT

Error threshold is the upper bound of the mutation rate such that the genetic information can pass on during reproduction. In Eigen's original proposal, the fitness landscape consists of a single peak along with other background sequences. However, realistic fitness landscapes are far more complicated and thus obstruct deeper understanding from either analytic or numerical approaches. Here we propose a variational approach to derive the error thresholds in generic fitness landscapes. When applying to the Eigen model, it delivers the correct error threshold and reveals the scaling relations between relevant parameters. For the fitness landscape with a peak–mesa profile, the two-dimensional phase diagram in the mesa fitness and the mutation rate can be computed with minimal cost. When the mesa fitness is larger than some critical value, upon increasing the mutation rate, the quasispecies changes from the peak to the mesa and then dissolves into the random background. The phase boundaries can be computed analytically within the effective-fitness approximation. At the end, the universal phase diagram in terms of the rescaled variables is explained in detail.

© 2017 The Physical Society of the Republic of China (Taiwan). Published by Elsevier B.V. All rights reserved.

## 1. Introduction

In 1971, Eigen proposed a simple yet profound model to investigate the processes of molecular replication, leading to the important concept of quasispecies [1]. The model stimulates both experimental and theoretical studies on molecular perspective for biological evolution [2–5] and advances our understanding in evolutionary dynamics. In Eigen's original proposal, selection moves the population towards the better adapted mutants and explores for even better ones in the sequence space by random mutations. A quasispecies, consisting of a wild type with the largest fitness accompanied by nearby mutant types in the sequence space [6], emerges as the robust population profile after some transient period. This simple model is widely applied to general evolutionary processes, particularly suitable for viral evolution [4,5] and population genetics [7–9].

It has been shown that there exists an error threshold [10,11] as an upper bound on the mutation rate, above which the quasispecies becomes unstable and no effective selection can occur. It places limits on how genetic information can be maintained and passed on from generations to generations and restricts possible theories for the origins of life. The expression of error threshold for the Eigen model was first derived by ignoring mutations back to the wild type, known as

\* Corresponding authors.

E-mail addresses: [hsiehau.lin@phys.nthu.edu.tw](mailto:hsiehau.lin@phys.nthu.edu.tw) (H.-H. Lin), [cjj@phys.sinica.edu.tw](mailto:cjj@phys.sinica.edu.tw) (C.-C. Chen).

the assumption of no back mutation [8]. The same expression was recovered by the exact solution of the model [12–15] later. And, through extensive studies by mapping to equivalent statistical models [16–20], the existence of the error threshold can be viewed as a first-order phase transition [21] from the order phase (quasispecies) to the disordered phase (no effective selection). In recent cancer studies, it has been suggested that the tumor cells can be viewed as quasispecies [22,23] with instability triggered by artificially-increased mutation (mutagenesis) [24–26] exceeding the error threshold. However, it is puzzling that some RNA viruses [27,28] seem to have mutation rates close to (or even larger than) the error threshold.

While the exact solution available for the single-peak Eigen model [12–15,29], many realistic features for viral evolution are left out, including multiple peaks in fitness landscape [30,31], survival of the flattest [32], spatial effects [33–35], mutational fitness effects [36] and so on. An exact solution incorporating these features seems out of reach. It is desirable to devise a mean-field approach capturing the essence of the phase transition so that the population profile of the quasispecies for arbitrary fitness landscapes of interest can be computed and analyzed with minimal efforts.

The variational approach is particularly useful to study the first-order phase transitions because the number of the low-lying states is often limited. In this Article, we present a simple variational method to investigate the phase transitions of the quasispecies. We first show how to map the replication-mutation equations into a quantum mechanical system in imaginary time [37]. Finding the stationary population profile in the long-time limit is thus equivalent to searching the ground state of the mapped quantum system. Applying the variational method to the single-peak Eigen model, the quasispecies is well captured in simple analytic form, agreeing with the exact solution exceedingly well. In addition, one can show that the profile of the quasispecies depends solely on a single rescaled variable  $g = uL/\ln(f_p/f)$ , where  $f_p$  and  $f$  denote the fitness of the peak sequence and the random background;  $u$  is the mutation probability during reproduction and  $L$  is the genome length.

In evolutionary dynamics with mutations, a fault-tolerant mesa [30,37] in the fitness landscape sometimes reigns over a peak with higher fitness. Due to the tremendous size of the sequence space, the stability of the quasispecies in the fitness landscape with the peak-mesa structure is hard to analyze. The variational method simplifies the task down to diagonalizing a  $3 \times 3$  matrix. When the mesa fitness is larger than the critical value  $f_m^c = f_p e^{-ul}$ , where  $l$  is the fault-tolerant length of the whole genome, the quasispecies around the peak is first replaced by the mesa and eventually dissolves into the random background when the mutation probability exceeds the error threshold. On the other hand, when the mesa fitness is smaller than  $f_m^c$ , the evolution of the quasispecies remains more or less the same as in the original Eigen model.

Using numerical calculation, we locate the mesa domain and map out the entire phase diagram for various genome lengths. The variational method captures the essence of the phase transition including the error threshold and the evolution of the wild-type population. It is rather remarkable that error thresholds for different mutation probabilities, fitness and genome lengths collapse onto universal curves, giving credits to the power of the variational approach.

This article is organized as follows. In Section 2, we briefly introduce the Eigen model and the gauge transformation mapping it to the equivalent Hamiltonian with imaginary-time dynamics. In Section 3, the equivalent Hamiltonian is solved by the variation method, giving the error threshold and also the evolution of the wild-type population. In Section 4, we calculate the back-mutation rate to the wild type and justify the assumption of no back-mutation in previous literature. In Section 5, the variational method is applied to the peak-mesa fitness landscape. The phase diagram is obtained and explained within the effective fitness picture. Additional technical details about the  $s$ -wave decomposition and the Crow-Kimura model can be found in the Appendices.

## 2. Eigen model

Consider a DNA chain with genome length  $L$ . The primary structure at each site of the chain takes on four different nucleotides ( $G$ ,  $A$ ,  $C$  and  $U$ ). For simplicity, we choose to distinguish only among purines ( $R$ ) and pyrimidines ( $Y$ ). Thus, the total size of sequence space is reduced to  $N_s = 2^L$ , yet still huge for numerical simulations. A viral genome with typical length  $L \sim 1000$  will lead to a (reduced) sequence space of the size  $\sim 10^{300}$ . For more complex forms of life, the size of the sequence space increases tremendously and methods developed in statistical mechanics seem appropriate and powerful to address the dynamics of the genome evolution.

The time-dependent relative population of a particular sequence  $i$  is denoted as  $x_i(t)$ . The selection is on the genotype level described by the fitness  $f_i$  for each sequence. In general, the fitness function can be rather complicated and evolves with time. Here it is assumed to take on its time-averaged value and treated as a constant. Eigen proposed a deterministic dynamics to describe the relative populations of the sequences in the presence of mutations,

$$\frac{dx_i}{dt} = \sum_j (m_{ij}f_j - \phi\delta_{ij})x_j, \quad (1)$$

where  $\phi = \sum_i x_i f_i$  is the average fitness for all sequences. The relative populations are normalized with  $\sum_i x_i = 1$  initially and remain so under the time evolution of the above quasispecies equation. The mutation matrix  $m_{ij}$  describes the probability to mutate from sequence  $j$  to sequence  $i$  in a generation and satisfies  $\sum_i m_{ij} = 1$ . If only point mutations are considered, it is straightforward to work out the mutation matrix

$$m_{ij} = u^{d_{ij}}(1-u)^{L-d_{ij}}, \quad (2)$$

where  $u$  is the probability for a single point mutation to occur within one generation and  $d_{ij}$  is the Hamming distance [38] between the sequences. Since  $d_{ij} = d_{ji}$ , it is clear that the mutation matrix  $m_{ij}$  is also symmetric.

The quasispecies equation can be transformed into an imaginary-time Schrödinger equation by the gauge transformation,

$$\Psi_i(t) = \sqrt{f_i} x_i(t) e^{W(t)}, \quad (3)$$

where  $\dot{W}(t) = \phi(t)$ , which is a form of integration factor transformation [39,40] allowing us to do away with the nonlinearity in Eq. (1). Applying the gauge transformation to the quasispecies equation, the dynamics of  $\Psi_i(t)$  is captured by an effective Hamiltonian  $H_{ij}$ ,

$$\frac{d\Psi_i}{dt} = -\sum_j H_{ij} \Psi_j, \quad \text{with } H_{ij} = -\sqrt{f_i f_j} m_{ij}. \quad (4)$$

One notices that the above equation can be obtained from the usual Schrödinger equation by replacing the time variable  $t \rightarrow it$  (and setting  $\hbar = 1$ ). Following the standard textbooks, the general solution for the “imaginary-time” Schrödinger equation is

$$\Psi_i(t) = \sum_n c_n \Phi_i^n e^{-E_n t}, \quad (5)$$

where  $E_n$  and  $\Phi_i^n$  are the eigenvalues and eigenvectors of the Hamiltonian  $H_{ij}$ . Note that the factor  $\sqrt{f_i}$  in the gauge transformation is crucial to make the Hamiltonian hermitian (real and symmetric here). One can still find the steady state of the dynamical Eq. (1) as the eigenstate with the maximum eigenvalue without the rescaling factor [41] but the variational approach presented below requires a hermitian matrix.

Unlike the intrinsic oscillatory nature of the usual Schrödinger equation, its imaginary-time version only keeps the ground state in the infinite-time limit,

$$\lim_{t \rightarrow \infty} \Psi_i(t) \rightarrow c_0 \Phi_i^0 e^{-E_0 t}, \quad (6)$$

and greatly simplifies the calculations. Making use of the normalization condition,  $\sum_i x_i = 1$ , the exponential factor in the gauge transformation can be expressed as

$$e^{W(t)} = \sum_i \frac{1}{\sqrt{f_i}} \Psi_i(t). \quad (7)$$

The inverse gauge transformation can thus be found,

$$\begin{aligned} x_i(t) &= \frac{1}{\sqrt{f_i}} \Psi_i(t) e^{-W(t)} \\ &= \left( \frac{1}{\sum_j \Psi_j / \sqrt{f_j}} \right) \frac{1}{\sqrt{f_i}} \Psi_i(t). \end{aligned} \quad (8)$$

In the infinite-time limit, only the ground-state wave function survives and the relative populations are

$$x_i^* = \lim_{t \rightarrow \infty} x_i(t) = \left( \frac{1}{\sum_j \Phi_j^0 / \sqrt{f_j}} \right) \frac{1}{\sqrt{f_i}} \Phi_i^0. \quad (9)$$

Note that the constant  $c_0$  cancels out and does not appear in  $x_i^*$ . To find the survival populations  $x_i^*$ , one only needs to find the ground state  $\Phi_i^0$  of the Hamiltonian.

### 3. Variational method

In Eigen's original proposal, the fitness landscape consists of a peak  $f_0 = f_p$  for the wild-type sequence (labeled as “0”) and the random background for other mutants. For simplicity, all other sequences  $i \neq 0$  have a common background fitness  $f_i = f$  smaller than the fitness peak. Even though the single-peak fitness landscape looks simple, the size of the sequence space is still huge  $N_s = 2^L$ . Here we choose two symmetric base functions,

$$\Psi^1 = (1, 0, \dots, 0), \quad \Psi^2 = \frac{1}{\sqrt{N_s - 1}} (0, 1, 1, \dots, 1), \quad (10)$$

and use variational approach to estimate the error threshold for the Eigen model. The wild-type sequence is represented by  $\Psi^1$  and the uniform distribution of the background mutant is represented by  $\Psi^2$ . All other configurations are ignored in the following variational calculations.

Construct the wave function for the ground state with two constrained variational parameters,

$$\Psi_i = c_1 \Psi_i^1 + c_2 \Psi_i^2, \quad (11)$$

where  $c_1^2 + c_2^2 = 1$  arises from the normalization condition. The variational energy takes the form,

$$E(c_1, c_2, \lambda) = \sum_{n,m=1,2} c_n \mathcal{H}_{nm} c_m + \lambda(c_1^2 + c_2^2), \tag{12}$$

where the Lagrange multiplier arises from the normalization condition. The variational Hamiltonian  $\mathcal{H}_{nm}$  is a  $2 \times 2$  real symmetric matrix constructed from the original Hamiltonian in the full sequence space,

$$\mathcal{H}_{nm} = \sum_{i,j=1}^{N_s} \Psi_i^n H_{ij} \Psi_j^m. \tag{13}$$

Because only point mutations are considered here, the matrix elements can be computed after some algebra,

$$\mathcal{H}_{11} = -f_p(1 - u)^L, \tag{14}$$

$$\mathcal{H}_{22} = -f + \frac{f}{N_s - 1} [1 - (1 - u)^L], \tag{15}$$

$$\mathcal{H}_{12} = \mathcal{H}_{21} = -\sqrt{\frac{f_p f}{N_s - 1}} [1 - (1 - u)^L]. \tag{16}$$

Minimizing the variational energy  $E(c_1, c_2, \lambda)$  is equivalent to diagonalizing the  $2 \times 2$  variational Hamiltonian  $\mathcal{H}_{nm}$  rewritten as

$$\mathcal{H}_{nm} = C\mathbf{1} + f \begin{pmatrix} -\epsilon & -\Delta \\ -\Delta & \epsilon \end{pmatrix}. \tag{17}$$

The constant term does not affect the frequencies  $x_i(t)$  after the inverse gauge transformation,

$$C = -\frac{1}{2} \left[ f_p m_{00} + f - \frac{f}{N_s - 1} (1 - m_{00}) \right], \tag{18}$$

with  $m_{00} = (1 - u)^L$  and will be dropped in the following calculations. The dimensionless parameters  $\epsilon$  and  $\Delta$  in the variational Hamiltonian are

$$\epsilon(u, L, r) = \frac{1}{2} [r m_{00} - 1] + \frac{1 - m_{00}}{2(N_s - 1)}, \tag{19}$$

$$\Delta(u, L, r) = \sqrt{\frac{r}{N_s - 1}} [1 - m_{00}], \tag{20}$$

where  $r = f_p/f$  is the ratio of fitness for the wild-type sequence and the background. The eigenvalues of the variational Hamiltonian are

$$E_{\pm} = \pm \sqrt{\epsilon^2 + \Delta^2}. \tag{21}$$

Therefore, the ground-state energy within the variational approximation corresponds to the negative eigenvalue  $E = E_-$  with the eigenvector

$$\frac{c_1}{c_2} = \frac{\Delta}{\sqrt{\epsilon^2 + \Delta^2} - \epsilon} = \frac{\sqrt{\epsilon^2 + \Delta^2} + \epsilon}{\Delta}. \tag{22}$$

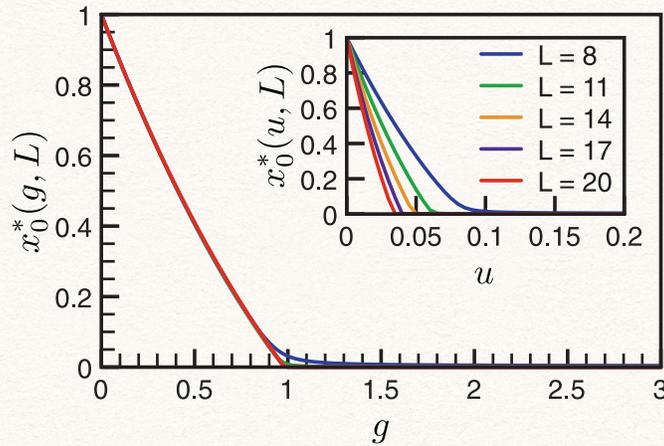
With the coefficients  $c_0$  and  $c_1$  at hand, the ground state within the variational approximation is

$$\Phi = \left( c_1, \frac{c_2}{\sqrt{N_s - 1}}, \frac{c_2}{\sqrt{N_s - 1}}, \dots, \frac{c_2}{\sqrt{N_s - 1}} \right). \tag{23}$$

The relative population of the wild-type sequence in the infinite-time limit can be computed from the variational ground state,

$$\begin{aligned} x_0^*(u, L, r) &= \left( \frac{1}{\sum_j \Phi_j / \sqrt{f_j}} \right) \frac{1}{\sqrt{f_0}} \Phi_0 \\ &= \frac{\sqrt{\epsilon^2 + \Delta^2} + \epsilon}{\sqrt{\epsilon^2 + \Delta^2} + \epsilon + \sqrt{r(N_s - 1)}\Delta}. \end{aligned} \tag{24}$$

As shown in Fig. 1, the stationary population of the wild-type sequence  $x_0^*$  is plotted against the mutation probability  $u$  for different genome lengths ranging from  $L = 8$  to  $L = 20$ . The fitness landscape consists of a single sharp peak with relative



**Fig. 1.** The stationary population of the wild-type sequence  $x_0^*$  plotted versus the point mutation rate  $u$  and the rescaled variable  $g = uL/\ln r$  with fitness ratio  $r = 2$ . The genome length varies from  $L = 8$  (blue) to  $L = 20$  (red). The error threshold  $g_c = 1$  can be spotted clearly when plotted versus the rescaled variable  $g$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fitness  $r = f_p/f = 2$  to a uniform background fitness  $f$  for all other sequences. The error threshold, where the quasispecies disappears, marks the phase transition and, in theory, is only well defined in the thermodynamic limit with infinite genome length. However, it is clear that the transition from the quasispecies to the random background (without any effective selection) is already sharp for finite genome length. Furthermore, the evolution of the quasispecies with different genome lengths collapses onto a universal curve when plotted with the rescaled variable,

$$g = \frac{uL}{\ln(f_p/f)}. \quad (25)$$

As is evident in Fig. 1, the collapse is almost perfect even for such short genome lengths. The error threshold occurs at  $g_c = 1$ , that is,

$$u_c L = \ln(f_p/f), \quad (26)$$

agreeing with the exact solution for the Eigen model.

The variational approach allows one to derive the universal curves for arbitrary fitness ratios. The thermodynamic limit corresponds to taking the genome length  $L$  to infinity while keeping the rescaled variable  $g = uL/\ln r$  finite. The probability to stay in the wild-type sequence during reproduction simplifies,  $m_{00} = (1 - u)^L \rightarrow r^{-g}$ . Because the off-diagonal term of the Hamiltonian  $\Delta$  is negligibly small, the stationary population of the wild-type sequence can be put down in analytic form,

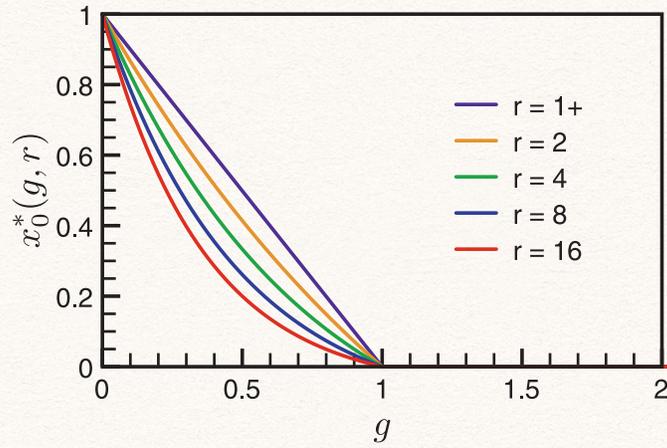
$$\lim_{L \rightarrow \infty} x_0^* = \frac{\frac{1}{2}[|r^{1-g} - 1| + (r^{1-g} - 1)]}{\frac{1}{2}[|r^{1-g} - 1| + (r^{1-g} - 1)] + r(1 - r^{-g})}. \quad (27)$$

The above form can be further simplified: since the relative fitness  $r$  is greater than unity, the numerator is zero for  $g > 1$ . Finally, the stationary population  $x_0^*$  in the thermodynamic limit falls onto the universal curve,

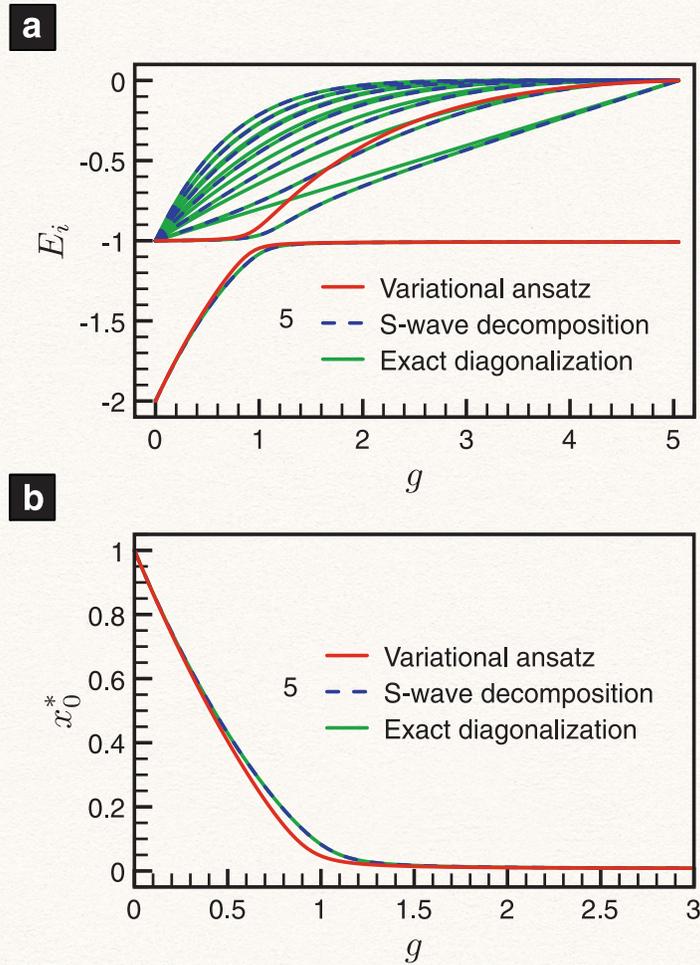
$$x_0^*(g, r) = \Theta(1 - g) \frac{r^{1-g} - 1}{r - 1}. \quad (28)$$

The universal curves for different fitness ratios is shown in Fig. 2. When the fitness peak is barely larger than the background ( $r \rightarrow 1^+$ ), the population of the wild-type sequence,  $x_0^* = 1 - g$ , falls off linearly to zero at the error threshold. It is quite interesting that these universal curves are identical to those derived under the assumption of no back mutations to the wild-type sequence. We shall come back to this coincidence in the next section.

How reliable is the variational approach? The use of the two-dimensional ansatz, Eqs. (10) and (11), here assumes that all the background sequences have the same population. In the original problem, the population of a sequence should depend on its Hamming distance from the peak. For a given chain length, the uniform distribution is approached when the mutation rate is large. Near the transition  $g = 1$ , this is the case when the peak-to-background fitness ratio  $r$  is large. Larger deviation can be expected when  $r \rightarrow 1^+$  but the general trend is not expected to change qualitatively. We check the validity of the variational results by comparing with the exact diagonalization in Fig. 3. The ground-state energy and the stationary population of the wild-type sequence agree extremely well. Since the peak structure is spherical symmetric in the sequence space, it is inspiring to project the full Hamiltonian onto the reduced  $s$ -wave basis, where calculation details can be found in Appendix A. Note that, after the  $s$ -wave decomposition, the dimension of the sequence space  $N_s = 2^L$  is reduced to  $L + 1$ , characterized by the Hamming distance from the fitness peak. This corresponds to the ansatz of having different stationary populations  $x_i$  for different classes of Hamming distance from the peak [42]. For the genome length  $L = 7$ , both the  $2 \times 2$  variational Hamiltonian and the  $(L + 1) \times (L + 1)$   $s$ -wave projected Hamiltonian agree with the exact diagonalization remarkably well. We see an exact match of some eigenvalues for the  $s$ -wave decomposed and the exact Hamiltonian due to the spherical symmetry of the fitness landscape. Some additional eigenvalues of the full Hamiltonian, missed in the  $s$ -wave decomposition, correspond to asymmetrical modes.



**Fig. 2.** The universal scaling function  $x_0^*(g, r)$  in the limit of infinite genome length. The transition is sharp and the universal scaling function varies with relative fitness from  $r = 1^+$  (purple) to  $r = 16$  (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

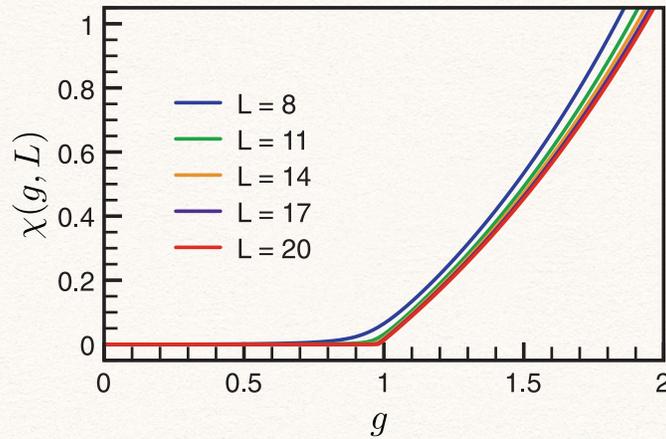


**Fig. 3.** Comparison of (a) the eigenvalues and (b) stationary population of the wild-type sequence among the variational, the s-wave decomposed and the exact Hamiltonians with genome length  $L = 7$  and the fitness ratio  $f_p/f = 2$ .

**4. No back mutation?**

The variational method also provides a helping hand to checking how good the assumption of no back mutations is. A good indicator is the ratio between the back-mutating rate and the rate to stay in the wild-type sequence,

$$\chi(u, L, r) = \frac{\sum_{j \neq 0} m_{0j} f_j x_j^*}{m_{00} f_0 x_0^*}. \tag{29}$$



**Fig. 4.** The back-mutation ratio  $\chi(u, L, r)$  plotted versus the rescaled variable  $g = uL/\ln r$  with fitness ratio  $r = f_p/f = 2$  and genome lengths ranging from  $L = 8$  (blue) to  $L = 20$  (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Within the variational approximation, all  $x_j^* = x_1^*$  for  $j \neq 0$  and the summation in the numerator can be carried out without difficulty,

$$\begin{aligned} \chi(u, L, r) &= \frac{x_1^* f(1 - m_{00})}{x_0^* f_M m_{00}} \\ &= \sqrt{\frac{f}{f_M}} \frac{1 - m_{00}}{m_{00}} \frac{c_1}{\sqrt{(N_s - 1)c_0}}. \end{aligned} \quad (30)$$

For  $r = 2$ , the back-mutation indicator  $\chi$  is shown in Fig. 4. When the quasispecies exists, the back-mutation rate is exponentially small even for finite genome length. Therefore, the assumption of no back mutations is justified in this regime. On the other hand, when the mutation probability  $u$  exceeds the error threshold, the indicator  $\chi$  grows exponentially and shows that the back-mutation rate dominates and cannot be ignored. In fact, the dynamics is dominated by mutual mutations between different sequences and no effective selection is at work. The variational method gives useful estimate for back mutations rates. Again, as shown in Fig. 4, when plotted against the rescaled variable  $g = uL/\ln r$ , all curves collapse onto a universal function in the thermodynamic limit. This is consistent with the exact results accounting for the back-mutation by Saakian and Hu [15] in the  $L \rightarrow \infty$  limit. Finite-size corrections of the results has also been reported by Saakian [29] for the mean fitness of the population as well as the distributions of species. Comparing to these exact results from the large  $L$  limit, our definition of the ratio  $\chi$  and the variational approximation provide an intuitive perspective on the importance of back mutation from a different angle.

Similarly, by taking the genome length to infinity, the universal function for  $\chi(g, r)$  can be derived. The only tricky term in  $\chi(g, r)$  is

$$\lim_{L \rightarrow \infty} \frac{c_1}{\sqrt{(N_s - 1)c_0}} = \Theta(g - 1) \frac{1 - r^{1-g}}{\sqrt{r(1 - r^{-g})}}. \quad (31)$$

After some algebra, the universal function for the back-mutation indicator in the thermodynamic limit is quite simple,

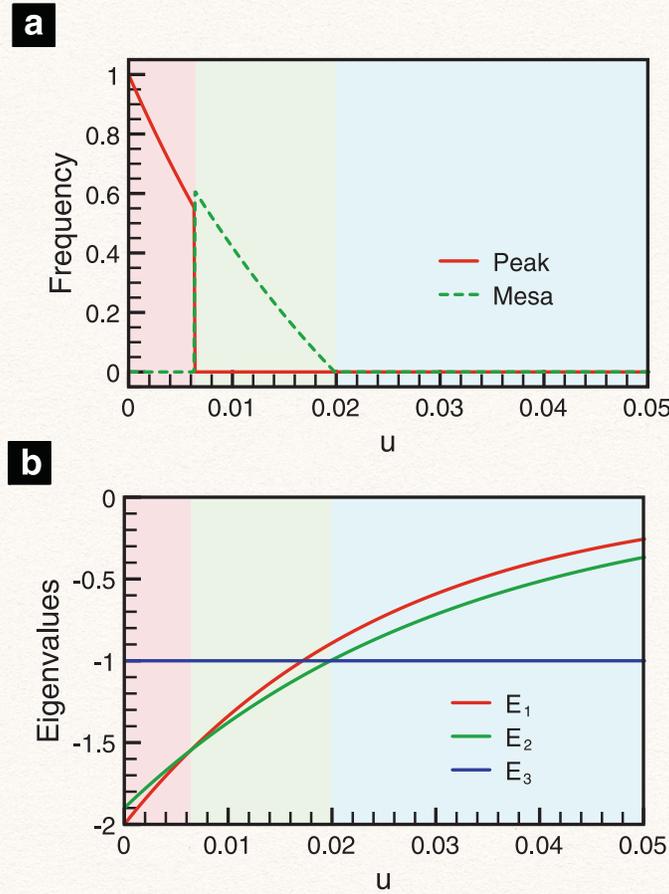
$$\chi(g, r) = \Theta(g - 1) (r^{g-1} - 1). \quad (32)$$

The universal function  $\chi(g, r)$  is zero below the error threshold and grows exponentially above it. Therefore, when considering the stability of the quasispecies, the assumption of no back mutations to the wild-type sequence is well justified in the thermodynamic limit. We note that for finite  $L$ , the Eigen's derivation based on no back mutation also matches surprisingly well to the exact solution [1,29]. This has recently been shown to result from subtleties, such as, the normalization condition of the sequence distribution [43] adopted by Eigen.

## 5. Peak-mesa landscape

Due to inevitable mutations during reproduction, “the survival of the fittest” do not always work. For instance, a mesa in the fitness landscape consists of a collection of sequences with roughly the same fitness as long as mutations occur within the tolerant base pairs. A single peak is fragile upon mutations while the mesa is more robust. Thus, it is possible that a mesa with lower fitness prevails a peak with higher fitness. Here we demonstrate how to apply the variational method to the peak-mesa fitness landscape.

To be precise, let us consider the sequence space for genome length  $L = 8$ . Without loss of generality, we assume the wild-type sequence to be 00000000 with peak fitness  $f_p$ . A mesa with tolerant length  $l = 2$  and distance  $d = 3$  from the wild-type consists of the sequences 111000 $b_1 b_2$ , where  $b_i = 0, 1$  denote the neutral base pairs. The number of sequences in



**Fig. 5.** (a) Stationary populations of the peak and the mesa versus the mutation probability  $u$  with  $f_p = 2$ ,  $f_m = 1.9$  and  $f = 1$ . The genome length  $L = 40$ , the tolerant length  $l = 8$  and the peak-mesa distance is  $d = 20$ . The peak-mesa transition is rather sharp even with finite genome length. (b) The transition points can be located by spotted the crossings between different eigenvalues.

the mesa is clearly  $N_m = 2^l$ . For simplicity, we assume the same fitness  $f_m$  for all sequences in the mesa. All other sequences are considered as background mutants with fitness  $f$  as before. The numbers of the sequences associated with the peak, the mesa and the background are  $N_p = 1$ ,  $N_m = 2^l$  and  $N_b = 2^L - 2^l - 1$  respectively and add up to the dimension of the sequence space  $N_s = 2^L$ . From the above explanations, generalization to arbitrary  $L$ ,  $l$  and  $d$  shall be straightforward.

Because we are interested in the interplay between the peak, mesa and background states (represented by sets of sequences  $P$ ,  $M$ , and  $B$ ), we can choose the three population distribution wave functions  $\Psi^p$ ,  $\Psi^m$ ,  $\Psi^b$  as the basis of our variational wave function:

$$\Psi = c_p \Psi^p + c_m \Psi^m + c_b \Psi^b, \tag{33}$$

where the  $i$ -th component for the base function  $\Psi_i^x = 1/\sqrt{N_x}$  if  $i \in X$  ( $X = P, M, B$ ) and 0 otherwise. The prefactors make sure all base functions  $\Psi^x$  are properly normalized. The variational energy takes the form

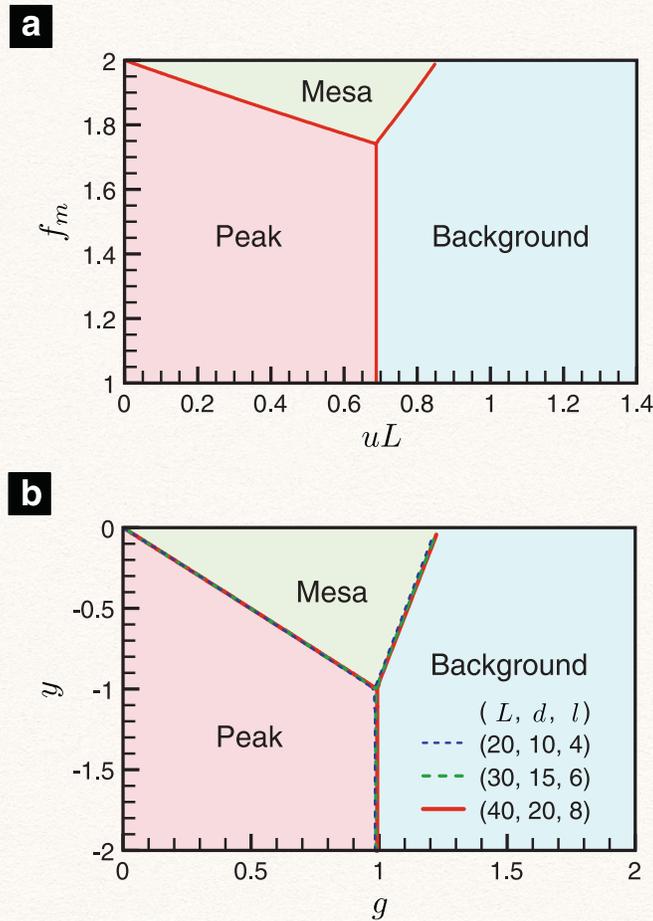
$$E(c_p, c_m, c_b, \lambda) = \sum_{x,y} c_x \mathcal{H}_{xy} c_y + \lambda \sum_x c_x^2, \tag{34}$$

where the Lagrange multiplier arises from the normalization constraint  $\sum_x c_x^2 = 1$ . Within the variational approach, the Hamiltonian is reduced to a  $3 \times 3$  real symmetric matrix with the elements:

$$\begin{aligned} \mathcal{H}_{xy} &= \sum_{ij} \Psi_i^x H_{ij} \Psi_j^y \\ &= -\sqrt{\frac{f_x f_y}{N_x N_y}} \sum_{i \in X, j \in Y} m_{ij}. \end{aligned} \tag{35}$$

Minimizing the variational energy  $E(c_p, c_m, c_b, \lambda)$  is equivalent to diagonalizing the  $3 \times 3$  variational Hamiltonian  $\mathcal{H}$ , which can be performed with minimal cost comparing to the full matrix.

The resulting stationary populations of the peak and the mesa are plotted in Fig. 5. One of the notable features of the peak-mesa transition is the sharpness of the population change: The peak population is almost entirely replaced by the mesa as soon as one enters the mesa-dominated regime. Such drastic transformation has been reported [30,37] for



**Fig. 6.** (a) Phase diagram for the peak–mesa fitness landscape with  $f_p = 2$ ,  $f = 1$  and  $(L, d, l) = (40, 20, 8)$ . (b) Phase diagram plotted with the rescaled variables  $g = uL/\ln(f_p/f)$  and  $y = (L/l)\ln(f_m/f_p)/\ln(f_p/f)$ . It is quite remarkable that the phase boundaries for different parameters all collapse onto the universal lines.

competing species and is in contrast with the peak-to-background and the mesa-to-background transitions where the peak populations vanish continuously as mutation rate increases. While the  $3 \times 3$  variational Hamiltonian produces only 3 low-lying eigenvalues, we have verified through exact diagonalization for genome lengths up to 12 that the ground state of the full Hamiltonian is non-degenerate and generally maintains a finite gap from the rest of the energy levels. As expected, the transition points obtained by the exact diagonalization deviates from the variational approach. However, for realistic values of the mutation probability, the variational method not only provides the correct qualitative picture but also serves as a reasonably good quantitative description.

The major results of this Article is shown in Fig. 6: Varying the mesa fitness  $f_m$  and the mutation probability  $u$ , one can map out the complete phase diagram for the quasispecies associated with the peak and the mesa. Fixing  $f_p = 2$  and  $f = 1$ , it leaves  $1 < f_m < 2$  as the range of interest for the mesa fitness. When the mesa fitness is larger than a certain critical value, there are two phase transitions: peak to mesa first, followed by mesa to background. As the mesa fitness grows larger, the mesa territory in the phase diagram becomes larger as well. These two phase boundaries meet at the *triple point*, below which only direct transition from peak to background occurs. It is quite interesting that the error threshold  $g_c = 1$  remains intact when the quasispecies of the mesa is unstable.

Let us try to construct an intuitive picture to understand the phase diagram. Employing the assumption of no back mutation again, one can introduce the notion of “effective fitness” integrating the fitness and the mutation effects together. The effective fitness for the peak is the actually reproduction rate with finite mutation probability in the quasispecies equation,

$$f_p^{\text{eff}} = f_p(1 - u)^L \approx f_p e^{-uL}, \quad (36)$$

as long as the mutation probability  $u \ll 1$ . Note that the back mutations to the wild-type sequence are ignored in the above expression. The peak fitness acquires a renormalized factor  $e^{-uL}$  due to point mutations. Similarly, the effective fitness for the mesa can be computed,

$$f_m^{\text{eff}} = f_m(1 - u)^{L-l} \approx f_p e^{-u(L-l)}. \quad (37)$$

Because of the tolerant length  $l$ , the renormalization factor for the mesa  $e^{-u(L-l)}$  is larger than that for the peak. It reflects the truth that the mesa is more robust against mutations. Finally, because the background sequences are the majority, a small mutation probability does not change the fitness at all,  $f^{\text{eff}} = f$ .

Because the mutation effects are taken into account, the population evolution follows “the survival of the effectively fittest”! The peak–mesa boundary is determined by equating the corresponding effective fitness,

$$f_p^{\text{eff}} = f_m^{\text{eff}} \rightarrow \ln(f_p/f_m) = ul. \quad (38)$$

Following similar steps, the peak–background and the mesa–background phase boundaries are determined,

$$f_p^{\text{eff}} = f^{\text{eff}} \rightarrow \ln(f_p/f) = uL; \quad (39)$$

$$f_m^{\text{eff}} = f^{\text{eff}} \rightarrow \ln(f_m/f) = u(L-l). \quad (40)$$

The analytic forms of these phase boundaries provide an easy way to understand the phase diagram.

But, how reliable are these estimated phase boundaries? It is convenient to introduce the rescaled variables:  $g$  for the mutation probability and  $y$  for the mesa fitness,

$$g = \frac{uL}{\ln(f_p/f)}, \quad y = \frac{L \ln(f_m/f_p)}{l \ln(f_p/f)}. \quad (41)$$

In terms of the rescaled variables, the peak–mesa, the peak–background and the mesa–background phase boundaries are

$$y = -g, \quad g = 1, \quad (42)$$

$$y = (L/l - 1)g - L/l. \quad (43)$$

For fixed  $L/l$  ratios, the phase boundaries should collapse onto the above universal lines when plotted with the rescaled variables. As shown in Fig. 6, for different parameters  $(L, d, l) = (20, 10, 4), (30, 15, 6), (40, 20, 8)$ , the phase boundaries indeed fall onto the predicted universal lines. The triple point where all three phases join takes on the universal value,  $(g_t, y_t) = (1, -1)$  in terms of the rescaled variables. The effective fitness turns out to be a simple yet power concept to understand the phase diagram for the peak–mesa fitness landscape.

In conclusion, we apply the variational method to the single-peak and the peak–mesa fitness landscapes and obtain the stationary population of the quasispecies (Figs. 1 and 5). A quantitative theory is important in providing testable predictions that can be compared with observations. Unfortunately, solutions for Eigen’s quasispecies equation are not available for generic fitness landscapes except for the simplest case. Numerical approach to the same problem is also challenged by the sheer size of the mutation matrix that grow exponentially with the sequence length. Our mapping of the Eigen model to a Hamiltonian system opens up standard approximation techniques developed for the later to the treatment of the former. The variational method we apply for the peak–mesa fitness landscape allows the computation of the phase diagram (Fig. 6) with minimal numerical cost. The variational approach offers a quantitative estimate for the critical mesa fitness, i.e. the triple point. In addition, without the guidances within the variational approach, it would be extremely hard to find these universal curves in terms of the rescaled variables by brute-force numerical endeavors.

The phase diagram also suggests that the original error threshold for the peak-to-background transition remains unaffected before the mesa gains enough fitness. Even then, within the peak-dominant regime, the peak population stays insensitive to the mesa state until the dramatic takeover by the mesa population at the peak–mesa boundary. While we do not expect the single-point mutation rate for organisms in the wild to vary significantly, the fitness of species generally depends on environmental factors which can vary over time. Based on the calculated results, we can expect to see a rapid and complete displacement of a highly optimized species with specific genetic combination by a less optimized species with a more robust genetic combination.

## Acknowledgment

We acknowledge support from [Ministry of Science and Technology](#) in Taiwan through Grant No. [MOST 103-2112-M-007-011-MY3](#).

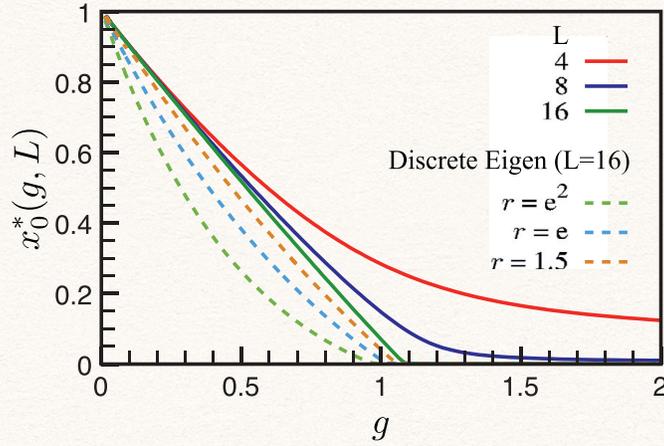
## Appendix A. S-wave decomposition

In the Eigen model, the fitness landscape is spherically symmetric around the wild-type peak. Thus, it is convenient to introduce the  $s$ -wave decomposition for generic configurations,

$$\Psi = (\psi^0, \psi^1, \dots, \psi^L), \quad (A.1)$$

where  $\psi^d$  is a row vector representing states with Hamming distance  $d$  to the peak and thus has the dimension of  $C_d^L = L!/d!(L-d)!$ . Because we are interested in the low-energy sector of the effective Hamiltonian, it is sufficient to keep the so-called “ $s$ -wave” sector [15] only. That is to say, out of the many states with the same Hamming distance to the wild-type sequence, one only needs to keep the totally symmetric state for each Hamming distance,

$$\Psi^{0s} = (\psi^0, 0, \dots, 0) = (1, 0, 0, \dots, 0), \quad (A.2)$$



**Fig. A.7.** Stationary population of the wild-type sequence  $x_0$  plotted against the rescaled variable  $g$  in the Crow–Kimura model ( $L = 4, 8, 6$ ) and the discrete Eigen model ( $L = 16$  and  $r = e^2, e, 1.5$ ). For genome length  $L = 16$ , the stationary populations derived from different models exhibit similar trends with the same error threshold.

$$\Psi^{1s} = (0, \psi^1, 0, \dots, 0) = \sqrt{\frac{1}{L}}(0, 1, 1, \dots, 1, 0, 0, \dots, 0), \quad (\text{A.3})$$

and all other  $\Psi^{ns}$  can be constructed in a similar way. Writing down the effective Hamiltonian explicitly,

$$H_{ij} = -\sqrt{f_i f_j} u^{d_{ji}} (1 - u)^{L - d_{ji}}, \quad (\text{A.4})$$

one can construct the reduced  $(L + 1) \times (L + 1)$  Hamiltonian in the  $s$ -wave sector,

$$\mathcal{H}_{nm} = \sum_{ij} \Psi_i^{ns} H_{ij} \Psi_j^{ms}, \quad (\text{A.5})$$

and numerically diagonalize the reduced Hamiltonian to obtain the ground state. The  $s$ -wave symmetry helps tremendously and reduces the size of the relevant sequence space from  $N_s = 2^L$  to just  $(L + 1)$ . Comparisons among the variational, the  $s$ -wave decomposed and the exact solutions are summaries in Fig. 3.

## Appendix B. Crow–Kimura model

Eigen's original proposal is formulated in terms of the coupled differential equations. To address the discrete reproduction generations, it is helpful to consider the discrete Eigen model for the evolution of populations  $X_i(t)$  (not relative populations) in discrete generations,

$$X_i(t + 1) = \sum_j m_{ij} f_j X_j(t), \quad (\text{B.1})$$

where the mutation matrix  $m_{ij}$  and the fitness function  $f_i$  are defined as in the main text. It is convenient to introduce the growth matrix  $\mathbf{G}$  defined as

$$G_{ij} = m_{ij} f_j - \delta_{ij}. \quad (\text{B.2})$$

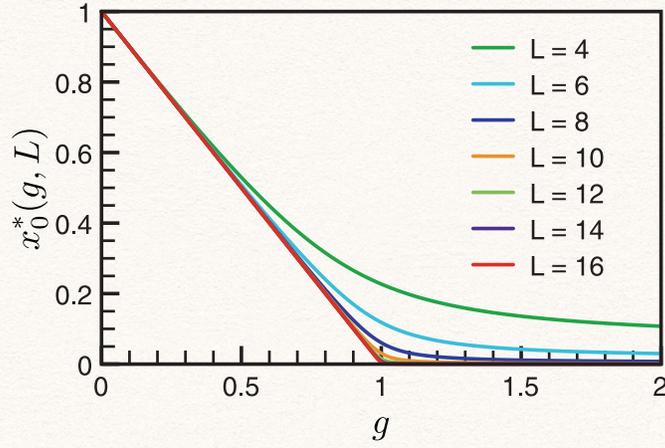
Positive eigenvalues of the growth matrix imply population growth and negative ones mean diminishing populations. The stationary state then corresponds to the eigenstate of  $\mathbf{G}$  with the largest eigenvalue. As demonstrated in Fig. A.7, the numerical results for evolution in discrete time steps agrees with those obtained in the original Eigen model.

Another closely related approach to the error threshold is known as the Crow–Kimura model, described by the set of differential equations,

$$\frac{dX_i}{dt} = G_i X_i + \sum_j \Gamma_{ij} X_j, \quad (\text{B.3})$$

where  $G_i = G + \delta_{i0}(G_p - G)$  is the grow rate for the  $i$ th sequence and  $\Gamma_{ij}$  is the mutation-rate matrix with elements  $\Gamma_{ij} = \Gamma$  for  $d_{ij} = 1$ ,  $\Gamma_{ii} = -\Gamma L$ , and zero otherwise. It is known in the literature, by taking the evolutionary time step to the continuous limit, the discrete Eigen model becomes the Crow–Kimura model. When comparing results from these two models, the rescaled variable  $g$  plays a crucial role. The relations between  $f_i$ ,  $u$  (discrete Eigen) and  $G_i$ ,  $\Gamma$  (Crow–Kimura) are

$$f_i = 1 + G_i \Delta t, \quad u = \Gamma \Delta t, \quad (\text{B.4})$$



**Fig. A.8.** The stationary population of the wild-type sequence  $x_0$  in the continuous limit ( $u \rightarrow 0^+$  and  $r \rightarrow 1^+$ ) for genome lengths varying from  $L = 4$  (green) to  $L = 16$  (red). Note that the scaling function with infinite genome length in the continuous limit is just the linear function  $x_0^*(g) = 1 - g$  below the error threshold  $g_c = 1$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where  $\Delta t$  is the microscopic time step before taking continuous limit. The rescaled variable  $g$  thus takes a slightly different form,

$$\begin{aligned} g &= \frac{uL}{\ln(f_p/f)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Gamma L \Delta t}{\ln(1 + G_p \Delta t) - \ln(1 + G \Delta t)} \\ &= \frac{\Gamma L}{G_p - G}. \end{aligned} \quad (\text{B.5})$$

Plotted versus the rescaled variable  $g$ , the stationary populations of the wild-type sequence for the discrete Eigen model and the Crow–Kimura model are shown in Fig. A.7. For discrete Eigen model with genome length  $L = 16$ , the stationary population  $x_0$  is very close to the universal scaling function in Fig. 2. The numerical results from the discrete Eigen model agree very well with the variational ansatz we proposed for the quasispecies equation. As the relative fitness  $r \equiv f_p/f$  approaches unity, the curve moves closer to the linear function  $x_0^* = 1 - g$  as expected. The Crow–Kimura model delivers similar trend for the stationary population of the wild-type sequence and leads to the same error threshold for the quasispecies. The success of the variational approach relies on the fact that the ground state is mainly composed of the chosen base functions, verified by the exact diagonalization.

The equivalence between the Crow–Kimura model and the Eigen model emerges in the continuous time limit  $\Delta t \rightarrow 0$ . Utilizing the relations in Eq. (B.4), it corresponds to  $u \rightarrow 0^+$  and  $r \rightarrow 1^+$  in the Eigen model. Taking the continuous limit in the Eigen model,

$$\epsilon_c(g, L) \equiv \lim_{r \rightarrow 1} \frac{\epsilon}{r - 1} = \frac{1}{2} \left[ 1 - g + \frac{g}{2^L - 1} \right], \quad (\text{B.6})$$

$$\Delta_c(g, L) \equiv \lim_{r \rightarrow 1} \frac{\Delta}{r - 1} = \frac{1}{\sqrt{2^L - 1}} g. \quad (\text{B.7})$$

After some algebra, the stationary population of the dominant sequence  $x_0^*(g, L)$  in the continuous limit is

$$x_0^*(g, L) = \frac{\sqrt{\epsilon_c^2 + \Delta_c^2} + \epsilon_c}{\sqrt{\epsilon_c^2 + \Delta_c^2} + \epsilon_c + g}. \quad (\text{B.8})$$

These curves for various genome lengths from  $L = 4$  to  $L = 16$  are plotted in Fig. A.8 and agree rather well with those obtained from Crow–Kimura model in Fig. A.7. Furthermore, the above expression becomes extremely simple when the genome length goes to infinity,

$$\lim_{r \rightarrow 1} x_0^*(g, r) = \Theta(1 - g) \times (1 - g). \quad (\text{B.9})$$

The universal scaling function in the thermodynamic (infinite genome length) and the continuous (infinitesimal generation time) limits is just a straight line.

## References

- [1] M. Eigen, Selforganization of matter and the evolution of biological macromolecules, *Naturwissenschaften* 58 (10) (1971) 465–523.
- [2] J.F. Crow, M. Kimura, *An Introduction to Population Genetics Theory*, Blackburn Press, 2009.
- [3] I. Leuthäusser, Statistical mechanics of eigen's evolution model, *J. Stat. Phys.* 48 (1–2) (1987) 343–360.
- [4] E. Domingo, C.R. Parrish, J.J. Holland (Eds.), *Origin and Evolution of Viruses*, Second Edition, second ed., Academic Press, 2008.
- [5] E. Domingo, C.K. Biebricher, M. Eigen, J.J. Holland, *Quasispecies and RNA Virus Evolution: Principles and Consequences*, first ed., Landes Bioscience, 2002.
- [6] M.A. Nowak, What is a quasispecies? *Trends Ecol. Evol. (Amst.)* 7 (4) (1992) 118–121.
- [7] J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, 1998.
- [8] M.A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life*, Belknap Press of Harvard University Press, 2006.
- [9] M.A. Nowak, Five rules for the evolution of cooperation, *Science* 314 (5805) (2006) 1560–1563.
- [10] M. Eigen, P. Schuster, A principle of natural self-organization, *Naturwissenschaften* 64 (11) (1977) 541–565.
- [11] J.M. Smith, E. Szathmáry, *The Major Transitions in Evolution*, Oxford University Press, USA, 1995.
- [12] S. Galluccio, Exact solution of the quasispecies model in a sharply peaked fitness landscape, *Phys. Rev. E* 56 (4) (1997) 4526.
- [13] D. Saakian, C.-K. Hu, Eigen model as a quantum spin chain: exact dynamics, *Phys. Rev. E* 69 (2) (2004) 021913.
- [14] D.B. Saakian, C.-K. Hu, Solvable biological evolution model with a parallel mutation-selection scheme, *Phys. Rev. E* 69 (4) (2004) 046121.
- [15] D.B. Saakian, C.-K. Hu, Exact solution of the eigen model with general fitness functions and degradation rates, *Proc. Natl. Acad. Sci. U.S.A.* 103 (13) (2006) 4935–4939.
- [16] P. Schuster, K. Sigmund, Dynamics of evolutionary optimization, *Berichte der Bunsengesellschaft für physikalische Chemie* 89 (6) (1985) 668–682.
- [17] I. Leuthäusser, An exact correspondence between eigen's evolution model and a two-dimensional ising system, *J. Chem. Phys.* 84 (3) (1986) 1884–1885.
- [18] M. Nowak, P. Schuster, Error thresholds of replication in finite populations mutation frequencies and the onset of muller's ratchet, *J. Theor. Biol.* 137 (4) (1989) 375–395.
- [19] P. Tarazona, Error thresholds for molecular quasispecies as phase transitions: from simple landscapes to spin-glass models, *Phys. Rev. A* 45 (8) (1992) 6038.
- [20] E. Baake, M. Baake, H. Wagner, Ising quantum chain is equivalent to a model of biological evolution, *Phys. Rev. Lett.* 78 (3) (1997) 559–562.
- [21] M. Eigen, Natural selection: a phase transition? *Biophys. Chem.* 85 (2–3) (2000) 101–123.
- [22] R.V. Solé, T.S. Deisboeck, An error catastrophe in cancer? *J. Theor. Biol.* 228 (1) (2004) 47–54.
- [23] M. Greaves, C.C. Maley, Clonal evolution in cancer, *Nature* 481 (7381) (2012) 306–313.
- [24] C. Lengauer, K.W. Kinzler, B. Vogelstein, Genetic instabilities in human cancers, *Nature* 396 (6712) (1998) 643–649.
- [25] E.J. Fox, L.A. Loeb, Lethal mutagenesis: targeting the mutator phenotype in cancer, *Semin. Cancer Biol.* 20 (5) (2010) 353–359.
- [26] L.A. Loeb, Human cancers express mutator phenotypes: origin, consequences and targeting, *Nat. Rev. Cancer* 11 (6) (2011) 450–457.
- [27] J.W. Drake, Rates of spontaneous mutation among RNA viruses, *Proc. Nation. Acad. Sci.* 90 (9) (1993) 4171–4175.
- [28] E. Domingo, C. Escarmís, N. Sevilla, A. Moya, S.F. Elena, J. Quer, I.S. Novella, J.J. Holland, Basic concepts in RNA virus evolution., *FASEB J.* 10 (8) (1996) 859–864.
- [29] D.B. Saakian, A new method for the solution of models of biological evolution: derivation of exact steady-state distributions, *J. Stat. Phys.* 128 (3) (2007) 781–798.
- [30] P. Schuster, J. Swetina, Stationary mutant distributions and evolutionary optimization, *Bull. Math. Biol.* 50 (6) (1988) 635–660.
- [31] D.B. Saakian, E. Muñoz, C.-K. Hu, M.W. Deem, Quasispecies theory for multiple-peak fitness landscapes, *Phys. Rev. E* 73 (4) (2006) 041913.
- [32] J. Sardanyés, S.F. Elena, R.V. Solé, Simple quasispecies models for the survival-of-the-flattest effect: the role of space, *J. Theor. Biol.* 250 (3) (2008) 560–568.
- [33] S. Altmeyer, J.S. McCaskill, Error threshold for spatially resolved evolution in the quasispecies model, *Phys. Rev. Lett.* 86 (25) (2001) 5819–5822.
- [34] S. Toyabe, M. Sano, Spatial suppression of error catastrophe in a growing pattern, *Physica D* 203 (1–2) (2005) 1–8.
- [35] J. Sardanyés, S.F. Elena, Quasispecies spatial models for RNA viruses with different replication modes and infection strategies, *PLoS ONE* 6 (9) (2011) e24884.
- [36] J. Sardanyés, C. Simó, R. Martínez, R.V. Solé, S.F. Elena, Variability in mutational fitness effects prevents full lethal transitions in large quasispecies populations, *Sci. Rep.* 4 (2014) 4625.
- [37] A. Wolff, J. Krug, Robustness and epistasis in mutation-selection models, *Phys Biol* 6 (3) (2009) 036007.
- [38] R.W. Hamming, *Coding and Information Theory*, second ed., Prentice Hall, 1986.
- [39] C.J. Thompson, J.L. McBride, On eigen's theory of the self-organization of matter and the evolution of biological macromolecules, *Math. Biosci.* 21 (1–2) (1974) 127–142.
- [40] B.L. Jones, R.H. Enns, S.S. Rangnekar, On the theory of selection of coupled macromolecular systems, *Bull. Math. Biol.* 38 (1) (1976) 15–28.
- [41] M. Eigen, J. McCaskill, P. Schuster, The molecular quasi-species, in: I. Prigogine, S.A. Rice (Eds.), *Advances in Chemical Physics*, vol. 75, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1989, pp. 149–263.
- [42] J. Swetina, P. Schuster, Self-replication with errors: a model for polynucleotide replication, *Biophys. Chem.* 16 (4) (1982) 329–345.
- [43] P. Schuster, Quasispecies on fitness landscapes, in: E. Domingo, P. Schuster (Eds.), *Quasispecies: From Theory to Experimental Systems*, Current Topics in Microbiology and Immunology, Springer International Publishing, 2015, pp. 61–120.