Phase diagram for the Eigen quasispecies theory with a truncated fitness landscape

David B. Saakian,[1,2,3] Christof K. Biebricher,[4] and Chin-Kun Hu[2,5]

[1]*Yerevan Physics Institute, Alikhanian Brothers St. 2, Yerevan 375036, Armenia*

[2]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*

[3]*National Center for Theoretical Sciences: Physics Division, National Taiwan University, Taipei 10617, Taiwan*

[4]*Max Planck Institute for Biophysical Chemistry, D-37070 Göttingen, Germany*

[5]*Center for Nonlinear and Complex Systems and Department of Physics, Chung-Yuan Christian University,*

Chungli 32023, Taiwan

Using methods of statistical physics, we present rigorous theoretical calculations of Eigen's quasispecies theory with the truncated fitness landscape which dramatically limits the available sequence space of information carriers. As the mutation rate is increased from small values to large values, one can observe three phases: the first (I) selective (also known as ferromagnetic) phase, the second (II) intermediate phase with some residual order, and the third (III) completely randomized (also known as paramagnetic) phase. We calculate the phase diagram for these phases and the concentration of information carriers in the master sequence (also known as peak configuration) x_0 and other classes of information carriers. As the phase point moves across the boundary between phase I and phase II, x_0 changes continuously; as the phase point moves across the boundary between phase II and phase III, x_0 has a large change. Our results are applicable for the general case of a fitness landscape.

I. INTRODUCTION

Developing realistic evolution models poses an important challenge for evolution research [1,2]. After the seminal work of Eigen quasispecies theory [3] and successful experiments with the self-replication of macromolecules [4,5], there have been intensive theoretical studies on molecular models of biological evolution [6–22]. In Eigen's quasispecies theory [3,10], an information carriers (DNA or RNA) of length L is represented by a one-dimensional spin model of length L with +1 representing purines (R) and −1 pyrimidines (Y). The whole space of information carriers of length L contains $M=2^L$ different sequences, which can be represented by S_k, $k=0,1,\ldots,M-1$; the relative frequency and the reproduction rate of the sequence S_k are represented by x_k and r_k, respectively. Such x_k satisfy the normalization condition: $(\Sigma_{k=0}^{M-1} x_k=1)$. S_0 is the master sequence (also called peak configuration) with the highest reproduction rate $r_0=A>1$. The structure of r_k is called the fitness landscape. The number of different bases between S_i and S_k is represented by d_{ik} and is called the Hamming distance between S_i and S_k. In the symmetric fitness landscape, the fitness of S_k is a function of the Hamming distance d_{0k} between S_k and the master sequence S_0, i.e., sequences with the same Hamming distance from the master sequence have the same reproduction rate.

The simplest fitness landscape is the single peak fitness, in which $r_0=A>1$ and $r_k=1$ for $k\neq 0$ [3,10,15]. In earlier papers, we have studied the fitness as a general function of the Hamming distance from the master sequence [16,17] or Hamming distances from several peak configurations [18]. We have also used the Hamilton-Jacobi equation (HJE) method [20–22] to a diploid evolution model with the symmetric fitness landscape [23]. In such studies, all sequences have nonzero reproduction rate.

However, realistic fitness landscapes are not smooth; they include neutral and lethal types, as observed in recent experimental studies with RNA viruses [24–26]. In the neutral type, each mutant on the neutral network has about the same reproduction rate as the master sequence and the fitness of a mutant is not simply a function of the Hamming distance from the master sequence. In the lethal type, some mutants have zero reproduction rate. In a special case of the lethal type, all mutants S_k with Hamming distances d_{0k} larger than a critical value n have zero reproduction rate, where d_{0k} is the Hamming distance between S_k and S_0. Such a lethal type is called the truncated fitness landscape.

In most evolution papers, symmetric fitness landscapes are considered. While solving evolution models, the vast majority of results for the mean fitness have been derived using uncontrolled approximations even for the symmetric fitness landscapes, with too simplified sequence space [27] and ignoring back mutations [10]. A simplified geometry with only two (Hamming) classes for sequences with nonzero fitness was used in studies that investigate the role of lethal mutants in evolution [28]. Furthermore, in most evolution models the whole sequence space is assumed to be available for the evolving genome.

However, the sequence space that a limited population can use is severely restricted to a small part of the sequence space surrounded by an unsurmountable moat of lethal mutations. In this paper, we attempt to rigorously solve the case of a truncated fitness landscape for a symmetric fitness landscape.

II. THE SYSTEM

In Eigen's quasispecies theory [3,10], the sequence S_i produces offspring of the parental type with the probability $Q_{ii}=q^L\equiv Q$ and offspring of another (mutant) type S_k $(k\neq i)$ with the probability $Q_{ki}=q^{L-d_{ik}}(1-q)^{d_{ik}}$, where q is the average incorporation fidelity and d_{ik} is the Hamming distance

(i.e., the number of different bases) between S_i and S_k.

Eigen proposed that x_i satisfy the equations [3,10]

$$\frac{dx_i}{dt} = \left\{ Q_{ii}r_i - \sum_k r_k x_k(t) \right\} x_i(t) + \sum_{k \neq i} Q_{ik}r_k x_k(t). \quad (1)$$

The mean fitness of the system is

$$R = \sum_{k=0}^{M-1} r_k x_k(t).$$

It is convenient to work with the error rate $u \equiv L(1-q)$, leading to $Q=e^{-u}$. In this paper, we consider only the case $L \rightarrow \infty$, u is finite. Eigen maps each genotype precisely into a node on the L-dimensional hypercube and has thus the correct connectivity for each type. The minimal number of steps leading from one position i in sequence space to another one, j, is the Hamming distance d_{ji}.

In the evolution process, the information content of the population can be maintained only when the selection force is higher than the dissipating one (mutation). Otherwise, above the *error threshold*, the information gets lost.

It has been shown that the system of nonlinear differential equations in Eq. (1) can be transformed into an infinite system of linear equations, connecting Eigen's model with statistical mechanics [7,9]. For a single peak fitness landscape ($r_0=A$ and $r_{i\neq 0}=1$), the following condition for conserving the master sequence in the population holds [3,10]:

$$AQ > 1, \quad (2)$$

where $AQ=1$ is the error threshold.

At the selective phase one has [29]

$$x_0 = \frac{QA-1}{A-1} \sim 1, \quad (3)$$

and $x_i \sim 1/L^d$ for $d \ll L$, where d is the Hamming distance from the wild sequence; see Eq. (21) in [17]. We choose the sequences with $1 \leqslant l \leqslant L$ from the corresponding lth Hamming classes. A scaling by Eq. (3) exists also for the rugged (random-energy-model-like) fitness landscapes [11]. Scaling like the one in Eq. (3) has also been applied in models of population genetics with few alleles. In realistic fitness landscapes, however, the wild type is present only in a few percent. Assuming neutrality, we can attain such scaling: a substantial fraction of one-mutation neighbors of the wild sequence has the same high fitness. Neutrality increases the probability of such mutants and suppresses x_0 as low as

$$x_0 \sim 1/\sqrt{L}. \quad (4)$$

This result could be derived easily using Eq. (6) in [30] for the case in which there is a central neutral sequence and large fracture of neutral sequences among the neighbor sequences of the central sequence. In nonselective phase, one has

$$x_0 \sim 1/M. \quad (5)$$

The error threshold phenomenon closely resembles the ferromagnetic-paramagnetic phase transition [10], where the fitness of the system corresponds to the microscopic energy

of the physics system, the mean fitness of the quasispecies to the free energy, and the mutation rate to temperature. To identify the different phases in statistical physics, one uses the free energy and also the order parameters. A phase transition occurs when, during a change of temperature, the analytical expression of the free energy changes. Order parameter changes also: while magnetization is nonzero in the ferromagnetic phase, it is zero at high temperatures and in the absence of a magnetic field. A phase transition in evolution is identified by observing the mean fitness $R=\sum_i x_i r_i$ and choosing proper order parameters, for instance the degree of distribution around the master sequence, and the surplus production, $s=\sum_i x_i(1-2d_{0i}/L)$, where d_{0i} is the Hamming distance from wild type.

Instead of the four-letter alphabet of genotypes, we consider only two symbols in a genome: the spins (alleles) "+" and "−," thus now $M=2^L$ [6]. Base substitutions correspond to sign changes of the spins. It is particularly easy to analyze landscapes where the fitness values are simple functions of the Hamming distance [6]. The L-dimensional sequence space is then transformed into a quasi-one-dimensional linear chain of mutant classes l, where $p_l=N_l x_l$ comprises all types with the Hamming distance number l from the master and the fitness value J_l. $N_l=\frac{L!}{l!(L-l)!}$ is the number of different genotypes in the l class. The parameter l can be identified as a phenotype parameter.

There is a principal difference between the quasi-one-dimensional model, derived rigorously from the initial sequence space with 2^L sequences [6,13,31], and the one-dimensional one considered in [27] and other articles. In contrast to other one-dimensional models used earlier [27] where each class contains only one type, in our case any class l is composed of N_l types and thus retains the connectivity; the Hamming distance between two sequences in the same class can take any value from 0 to $2l$. In the case of the parallel model [12], when evolution equations are formulated for class probabilities, the effective mutation rates to the lower class $\sim (L-l)/L$, and to the higher class l/L are different and change with l [13,31]. In contrast, in the one-dimensional model of [27] these mutation rates are l-independent.

In our quasi-one-dimensional model, J_l can be transformed into the $f(k)$, where $f(k)$ is an appropriate smooth function with the maximum at $k=1$, and $f(0)=1$. The "magnetization" parameter k is defined as $k \equiv (1-2l/L)$.

Consider now the solution of the Eigenmodel with symmetric fitness landscape. The mean fitness R for the fitness function f has been derived as follows [17]:

$$R = \max \{ f(k)Q^{(1-\sqrt{1-k^2})} \}|_{-1 \leqslant k \leqslant 1} = f(k_0)Q^{(1-\sqrt{1-k_0^2})}, \quad (6)$$

where k_0 is the location of the maximum on the right-hand side of the first equation. s can be identified from the mean fitness expression using an equation

$$f(s) = R \quad (7)$$

as has been derived in [13] for the parallel model. Thus in Eq. (6) the maximum is at some k_0, an order parameter of the system quantifying the bulk spin magnetization, while the

surplus s corresponds to the surface magnetization. Equation. (6) is an exact expression (at the infinite genome limit), while in other studies [10] back mutations have been ignored. As shown by Tarazona [9], the Eigenmodel is not equivalent to the simple ferromagnetic system of spins in the lattice, but only to those spins interacting both inside the bulk of the lattice and on the surface of the lattice. In this work, different phases will be characterized by R, the mean fitness; by k_0, the bulk magnetization; by x_0, the fraction of the wild type of the total population; and by s, the surplus. When $k_0 = 0$, resulting in $s = 0$, the population spreads statistically in sequence space, indicating a nonselective phase.

We gave the mean fitness and error threshold [when s in Eq. (7) becomes 0] for the symmetric fitness landscape. The point is that this transition has also an information-theoretical meaning. Eigen actually found the error threshold from an information-theoretical consideration of his model. Eigen's idea (information-theoretical content of a model) resembles the investigation of information-theoretical (optimal coding) aspects of disordered systems, developed in statistical physics two decades later [32,33]. In the random energy model of spin glass [34], the phase transition point was derived using the information theory analogy [32,33], and was found to yield results corresponding to those derived by Eigen. The deep information-theoretical meaning of error threshold transition in evolution models (equivalent to Shannon inequality for optimal coding) is a solid argument that a transition such as the one in Eq. (2) exists for any (irregular, with lethal or neutral mutants) fitness landscapes.

III. WAGNER AND KRALL THEOREM

Wagner and Krall [27] considered a population composed of the master and an infinite linear chain of mutants, where each type mutates only to its next neighbor and the fitness r_i decreases monotonically. When there is no low bound of the fitness, an absence of the error threshold transition was derived. Indeed, when in Eq. (6) $f(0) = 0$, there is no error threshold transition. But in more general symmetric fitness landscapes with a finite $f(0)$, this ceases to be valid. The proof is as follows: the maximum types are located at the Hamming distance class $L/2$ or, equivalently, at $k = 0$. Consider the logarithm of the right-hand side in Eq. (6), and expand near $k = 0$,

$$(1 - \sqrt{1-k^2})\ln Q + \ln f(k) \approx -u\frac{k^2}{2} + \ln[f(0)] + ck^\epsilon, \quad (8)$$

where c and ϵ are parameters describing the function $f(k)$, and $\ln f(k) - \ln f(0) \sim ck^\epsilon$ at $k \to 0$. When $\epsilon < 2$, Eq. (8) has a maximum at $k > 0$, fulfilling the condition for selection. When $\epsilon \geq 2$, it can be demonstrated that there is a maximum at $k = 0$ for a sufficiently low reproduction fidelity Q, therefore a sharp error threshold transition results. In the too simplistic model of Wagner and Krall, the right-hand side of Eq. (8) lacks the quadratic term, resulting in a monotonic function of k and the absence of phase transition. In the Eigenmodel the quadratic term holds, breaking the monotonic character of R in Eq. (8) and invoking the error threshold.
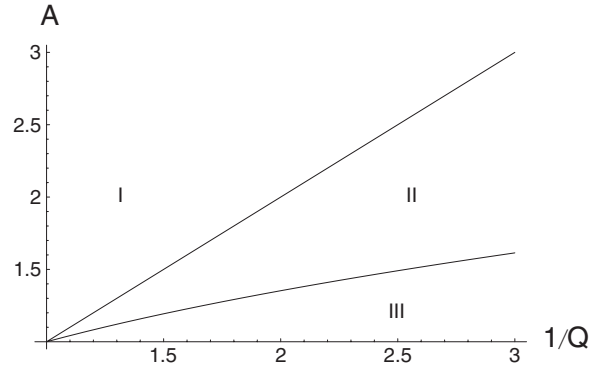

FIG. 1. Phase structure for the model with the overlap parameter for truncation point $K = 0.9$ and mutation rate $u = 1$. I, selective phase at $QA > 1$, where A is the fitness at the peak and $Q = e^{-u}$ is the errorless copying probability of the genome. II, intermediate phase at $QA < 1$ and $QA > Q^{1-\sqrt{1-K^2}}$. III, nonselective phase at $QA < Q^{1-\sqrt{1-K^2}}$.

IV. TRUNCATED SINGLE PEAK FITNESS LANDSCAPE

Let us consider a symmetric fitness landscape, where there is nonzero fitness only to some Hamming distance d from the reference sequence S_0. Here we define the truncated landscape as a single-peak one where all sequences beyond the Hamming distance $d \equiv L(1-K)/2$ are lethal,

$$r_0 = A; \quad r_l = 1, \quad 1 \leq l \leq d; \quad r_l = 0 \quad l > d. \quad (9)$$

Now we have

$$M = \sum_{l=0}^{d} \frac{L!}{l!(L-l)!} \quad (10)$$

nonlethal sequences.

To define the mean fitness, we compare the expression of Eq. (6) inside the region $K < k \leq 1$ and at the border $k = K$.

The investigation of this model is instructive; see Figs. 1–4. When $QA > 1$ the phase is selective (phase I); x_0 is given by Eq. (3), $k_0 = 1$, $R = QA$ [35]. When

$$1 > QA > Q^{1-\sqrt{1-K^2}}, \quad (11)$$

a new phase II prevails with

$$k_0 = 1, \quad R = QA. \quad (12)$$

In phase II, x_0 decreases exponentially with L. The expression of x_0 is calculated in the Appendix [see Eq. (A19)],

$$x_0 \sim \exp\left(L\int_K^1 dm \frac{1}{2} \ln \frac{\frac{\ln A}{u} + \sqrt{\left(\frac{\ln A}{u}\right)^2 - 1 + m^2}}{1+m}\right). \quad (13)$$

At the transition point between phase I and phase II, the expression in the exponent becomes zero, therefore the transition is continuous.
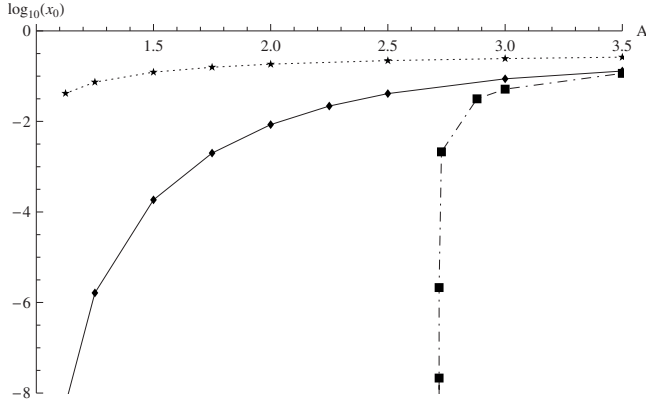
When

FIG. 2. Semilogarithmic plot of x_0 vs A for a single-peak fitness landscape truncated at $d > 8$ with $u=1$, $L=3000$. The upper curve corresponds to the Summers-Litwin model; the middle curve corresponds to the truncated fitness landscape by Eq. (9). The square boxes correspond to a single-peak fitness model. For $\ln(A) < 1$, the Eigenmodel with single-peak fitness has $x_0 = 0$.

$$QA < Q^{1-\sqrt{1-K^2}}, \qquad (14)$$

the k_0 is at the border and we have the nonselective phase III,

$$k_0 = K, \quad R = Q^{1-\sqrt{1-K^2}}. \qquad (15)$$

The expression for x_0 is defined in the Appendix, Eq. (A15). There is some focusing around reference sequence, and x_0 is higher than $1/M$. For $K=0.9$, we have $x_0 \approx 1/M^{0.66}$.

The transition between phases II and III is a discontinuous one; x_0 decreases M_1 times, [see Eq. (A21)],
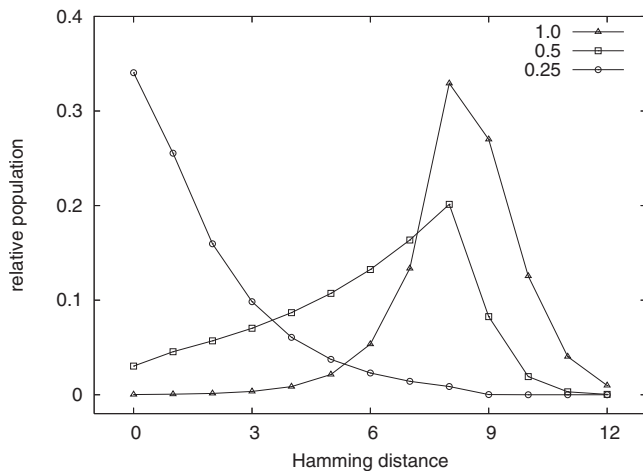

FIG. 3. Truncated Eigenmodel with $d=8$, $A=3$, $u=1$, and $L =3000$. The distribution of probabilities of Hamming classes p_l at different mutation rates (u values). At high accuracy (circles) the population groups around the master type, at intermediate accuracy the population peaks at the truncation border, and at low accuracy the master type has practically disappeared and the majority of the progeny is lethal.
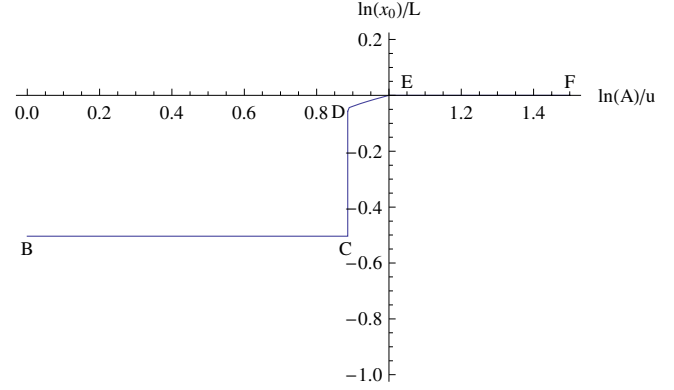

FIG. 4. (Color online) The graphics of $(\ln x_0)/L$ for the truncated Eigenmodel at infinite L with $K=1/2$, $u=1$. The I phase (curve EF) is at $1 < \ln A/u$, the second (curve DE) at $0.88 < \ln A/u < 1$, and the third (curve BC) at $0 < \ln A/u < 0.88$. The curve CD corresponds to the jump at the border between phases II and III. In the first phase, $(\ln x_0)/L \sim O(1/L)$.

$$M_1 \sim \exp\left(\frac{L}{2}\int_K^1 dm \ln \frac{\sqrt{1-K^2}+\sqrt{m^2-K^2}}{\sqrt{1-K^2}-\sqrt{m^2-K^2}}\right). \qquad (16)$$

Figure 4 illustrates different behavior of x_0 in three phases.

While in the ordinary Eigenmodel (without truncation of fitness) there is a sharp phase transition with the jump of the x_0 behavior from Eq. (3) to the $1/M$, in the truncated case this sharp transition is moved to the transition point between phases II and III. Now x_0 is continuous at transition point between phases I and II. For the Summers-Litwin case with $K=1$ we have $M_1 \rightarrow 1$, therefore the transition disappears, as has been obtained in [28], and for the $K=0$ case we get the result of the Eigenmodel, $M_1 \sim 2^L$. Our formulas are derived for the case $K \geqslant 0$.

Is phase II a selective one in the ordinary meaning? Clearly, its mean fitness is higher than in a typical nonselective phase like phase III. This point was clarified by calculating the surplus, replacing the steplike fitness function near the borderline ($k=K$) with a smooth function $f(k)$ that changes its value from 1 to 0 near K. In both phases II and III, with $QA > Q^{1-\sqrt{1-K^2}}$, the majority of the population is near the borderline, both on the viable and the lethal side. Therefore, while there is a kind of phase transition with some population rearrangement, phase II is identified as an intermediate one, with $x_0 \ll 1$, as in the nonselective phase. Summers and Litwin [28] first realized that $k_0 \rightarrow 1$ in a truncated fitness landscape and tried to analyze the phenomenon. Unfortunately, they used too simplistic a model where all mutants except the nearest neighbors of the master type were lethal. Figure 2 compares the relative concentration of the master sequence x_0 at various A in the single-peak fitness landscape described in [3,10], with the truncated fitness landscape with $d=8$, and the case considered by Summers and Litwin. Note the strong dependence of the master concentration on the $(\ln A)/u$ in the more realistic landscape, leading to an error threshold, in contrast to the case in [28]. The population profiles of the truncated fitness landscape at different mutation rates are shown in Fig. 3, showing the tran-

TABLE I. Eigenmodel with truncated single-peak fitness landscape for $L=1000$, $u=1$, $K=1/2$. The transition between phases I and II is at $A \approx 2.778$, and between phases II and III at $A \approx 2.377$. R is the numerical mean fitness.

Phase	III	II	II	II	II	I
A	2.3	2.4	2.5	2.6	2.7	2.8
R	0.859	0.883	0.919	0.957	0.993	1.030
R theory	0.875	0.882	0.919	0.956	0.992	1.029
x_0	7×10^{-54}	2×10^{-24}	4×10^{-16}	10^{-9}	4×10^{-4}	0.0167
x_0 theory						0.0164

sition from a master-dominated population to a widespread mutant distribution in the nonselective case.

In Tables I–III, we give the results of numerics for different phases. Mean fitness is well confirmed numerically, while the accuracy of numerics is poor to get correct values of x_0 in the second phase, $2.37 < A < 2.77$ corresponding to $0.88 < \ln(A)/u < 1$.

How can the phases be identified? The fact is that the parameters k_0 and s have different meanings in a statistical physics approach. This subject has been well analyzed in a series of articles by Baake and her co-authors [13,14]. s and k_0 have been identified with the transverse and longitudinal magnetization of spins in the corresponding quantum model. We just link s with the mean characteristic of the phenotype, and k_0 with the repertoire of genotypes. The consensus sequence should be determined experimentally not only via distribution x_i, but also via distribution $\hat{x}_i \equiv (x_i)^2 / \Sigma_{j=0}^{M}(x_j)^2$. Consider

$$\hat{k}_l = \frac{N_l (x_i)^2}{\sum_j (x_j)^2}, \qquad (17)$$

where the sequence i belongs to the lth phenotype class with N_l sequences. $N_l x_i$ is just the total "probability" measure of the phenotype according to the measure $\hat{x}_i$. Having such data, one can simply identify the phase structure; see Eq. (A29).

In the Appendix, we solve the truncated fitness models for the general monotonic piecewise smooth function $f(x)$. The numerics confirms well our analytical results for the new phase.

V. DISCUSSION

We rigorously solved (at the infinite genome length limit) Eigen's model for the truncated selection using the method of [17] as well as methods of statistical mechanics, including the analogy of the error threshold to the ferromagnetic-paramagnetic transition. This analogy is a complicated critical phenomenon, presented by Leuthäusser and Tarazona [7,9] and well analyzed by Baake and co-authors [12,13]. Instead of using only one order parameter to identify the phase of the model (magnetization), it is necessary to take into account several order parameters describing the order of spins in the bulk lattice and at the surface. Recently the existence of an error threshold was questioned [28] in the case of truncated selection. In this model, the available sequence space has been shrunk to an extremely small size. Figure 2 illustrates that the unrealistic assumption in [28] changed the relative concentration of the master type by more than three orders of magnitude. Nevertheless, this work was certainly useful for clarifying the concept of quasispecies: the authors first realized the intriguing features of a truncated selection landscape. We found a new evolution (intermediate) phase, when there is no successful selection via phenotype trait (the majority does not share the trait), while there is some grouping of population at the genotype level. The intermediate evolution phase differs from the nonselective phase, the frequency of wild type being $\frac{1}{\sqrt{M}}$ or higher in the intermediate phase compared with $\sim \frac{1}{M}$ in the nonselective phase of the Eigenmodel. We proposed a parameter to measure the hidden grouping of population in a genotype level, Eq. (17). Such hidden ordering could be important in the case of a changing environment: it is possible to force the whole population to

TABLE II. Eigenmodel with truncated fitness, $K=1/2$, $L=500$, $u=1$, $f(x)=\exp(cm^2/2)$. R is the mean fitness calculated from numerics. The transition between phases I and II is at $c=2$, and between phases II and III at $c=1/\sqrt{1-K^2} \approx 1.1547$. In phase I we have $s=1-1/c$, $\ln(R)/u=c(1-1/c)^2/2$. In phase II, $s=K$, $\ln(R)/u=c(1-1/c)^2/2$. In phase III, $s=K$, $\ln(R)/u=cK^2/2+\sqrt{1-K^2}-1$.

Phase	III	III	II	II	II	I	I	I
c	0.5	1	1.2	1.3	1.5	1.9	2.1	3
$\ln(R)$	−0.090	−0.016	0.0163	0.035	0.084	0.212	0.287	0.665
$\ln(R)$ theory	−0.071	−0.009	0.0160	0.034	0.083	0.213	0.288	0.666
s	0.507	0.507	0.507	0.508	0.510	0.525	0.541	0.663
s theory	0.5	0.5	0.5	0.5	0.500	0.5	0.523	0.666
$-\ln(x_0)$	223	199	185	177	160	131	120	84

TABLE III. Eigenmodel with truncated fitness, $K=1/2$, $L=500$, $u=1$, $f(x)=\exp(cm^3/3)$. The transition point is at $c\approx 2.31$, $m_0=0.866$ in the model without truncation. The transition point between phases I and II is at $c\approx 2.79$.

Phase	III	III	III	II	II	II	I	I
c	0.5	1	2.2	2.4	2.5	2.6	3	4
$\ln(R)$	−0.135	−0.016	−0.022	0.0202	0.0433	0.0676	0.172	0.459
$\ln(R)$ theory	−0.11	−0.092	−0.042	0.0201	0.0435	0.0678	0.171	0.460
s	0.507	0.507	0.510	0.514	0.518	0.523	0.563	0.698
s theory	0.5	0.5	0.5	0.5	0.5	0.5	0.556	0.701
x_0	230	216	142	128	122	117	98	70

extinction changing the fitness of a small fraction (much smaller than $1/L$, but much higher than $1/M$, where M is the number of different genotypes) of viruses in the population. We recommend virologists to measure the consensus sequence not only using the probabilities x_i, but also the $x_i^2/\Sigma_j x_j^2$. The evolution picture of the virus population is robust when two versions of consensus sequence are close to each other. In experiments [25], an evolution picture has been observed that is qualitatively similar to the intermediate evolution phase.

How the error threshold transition is connected to the virus extinction in virus experiments is another story. Several mechanisms are possible: an error catastrophe, as well as a critical mean fitness in order to maintain viral growth [36]. In this work, we observed the new phase with single peak and symmetric landscapes, but this phase exists probably for any (including irregular) fitness landscape with a lethal wall in sequence space.

APPENDIX A: APPLICATION OF THE HJE FOR A TRUNCATED SYMMETRIC LANDSCAPE

Let us apply the Hamilton-Jacobi equation (HJE)[37,38] method to the Eigenmodel [20–22] with the truncated symmetric fitness landscape defined by a piecewise smooth, monotonic fitness function $f(m)$,

$$f(m) = f_0(m), \quad K < m < 1;$$

$$f(m) = 0, \quad -1 < m < K. \tag{A1}$$

Here $f_0(m)$ is a monotonic analytical function.

We denote by k_0 the maximum point in Eq. (6). As s is defined by Eq. (7), for the monotonic fitness function we obtain

$$s \lesssim k_0, \tag{A2}$$

and there is only one solution of Eq. (7).

Using an ansatz

$$p_l = \exp[LU_0(m,t)], \tag{A3}$$

$m=1-2l/L$, one can derive the following equation [20]:

$$L\frac{\partial U_0(m,t)}{\partial t} = f_0(m)e^{-u}\exp\left\{u\left[\cosh\left(2\frac{\partial U_0(m,t)}{\partial m}\right)\right.\right.$$
$$\left.\left. + m \sinh\left(2\frac{\partial U_0(m,t)}{\partial m}\right)\right]\right\}. \tag{A4}$$

In [20] the HJE has been derived for the sequence probabilities, while Eq. (A4) is for the class probabilities p_l. Equation (A4) transforms into the corresponding HJE for the sequence probabilities [20,22] after the mapping $U_0(m,t)\to U_0(m,t)+\frac{1+m}{2}\ln\frac{1+m}{2}+\frac{1-m}{2}\ln\frac{1-m}{2}$. The HJE approach works for any piecewise smooth fitness functions. Equation (A1) is such a case. To solve our truncated case, we should just use different analytical solutions for the regions $-1<m<K$ and $K<m<1$.

Assuming an asymptotic $U_0(m,t)=\frac{R}{L}t+U(m)$, we derive [20]

$$R = f_0(m)e^{-u}\exp\left\{u\left[\cosh\left(2\frac{dU(m)}{dm}\right) + m \sinh\left(2\frac{dU(m)}{dm}\right)\right]\right\}, \tag{A5}$$

where R is calculated by Eq. (6). The surplus s is defined as the value of m where $U(m)$ has a maximum. When s is inside the region $[K,1]$, $U'(s)=0$. At the extremum point with $U'(m)=0$, Eq. (A5) gives $f(s)=R$. As for a monotonic fitness function there is a single solution for Eq. (6), $U(x)$ has a single maximum point in this case, and we take $U(s)=0$.

We use Eq. (A5) to define p_l with an accuracy $O(1)$ for $\ln p_l$, calculating $U(m)=U(s)+\int_s^m U'(m)dm$ for the corresponding $m=1-2l/L$. Moreover, it is possible to calculate $\ln p_l$ with a higher accuracy $O(1/L)$. In [20], we gave explicit formulas for the case of a parallel model. It is possible to construct similar results for the Eigenmodel as well.

We have two branches of solutions for Eq. (A5)

$$U'(m) = \frac{1}{2}\ln\frac{q \pm \sqrt{q^2-1+m^2}}{1+m},$$

$$q = \frac{1}{u} \ln \frac{R}{f(m)} + 1, \tag{A6}$$

where $U'(m) = \frac{dU(m)}{dm}$.

The choice of different solutions is a principal point. We choose the proper branch assuming the following: (i) $U(m)$ is a continuous function, (ii) $U'(m)$ is a continuous function at $K < m < 1$, and (iii) $U(m)$ is a convex function.

The transition between two branches [$\pm$ solutions in Eq. (A6)] is only at the point where $q^2 - 1 + m^2 = 0$ or

$$R = f(m)e^{-u+u\sqrt{1-m^2}}. \tag{A7}$$

According to Eq. (6), R is the maximum of the right-hand side. Thus we should choose only the branch with a "$-$" sign when k_0 is at the border, $k_0 = K$. When k is inside the interval $[K, 1]$, we choose the "$-$" solution for the interval $[k_0, 1]$ and the "$+$" solution in the interval $[K, k_0]$.

For the $V(m) = \ln(p_l/\sqrt{N_l})/L$ we have another equation [20],

$$R = f_0(m)e^{-u}\exp\left\{u\left[\cosh\left(2\frac{dV(m)}{dm}\right)\sqrt{1-m^2}\right]\right\}. \tag{A8}$$

The minimum of the right-hand side via V' just gives the $f(m)e^{-u+u\sqrt{1-m^2}}$. Thus at the maximum point $m = k_0$ of function $V(m)$ we have $V'(k_0) = 0$. In this paper, we consider the case in which Eq. (6) has a single solution $k = k_0$.

Solutions of Eq. (A5) and (A8) are simply related [20],

$$V(m) = U(m) + \frac{(1+m)\ln\frac{1+m}{2}}{4} + \frac{(1-m)\ln\frac{1-m}{2}}{4}. \tag{A9}$$

Consider now the following different phases of our model: the selective one with $K < k_0 < 1$, $K < s < 1$; the nonselective one with $k_0 = K$, $s = K$; and the intermediate one with $K < k_0 < 1$, $s = K$.

Selective phase

Now R is given by Eq. (6) with $K < k_0 < 1$. We used the "$-$" solution of Eq. (A6) for $k_0 < m < 1$ and the "$+$" solution for $K < m < k_0$. The maximum points of both functions $U(m)$ and $V(m)$ are inside the interval $[K, 1]$. We have $U'(s) = 0$ and $V'(k_0) = 0$. The formulas for the steady-state distributions are the same as in [20]. We have a mean fitness

$$R = f(k_0)e^{-u+u\sqrt{1-k_0^2}}. \tag{A10}$$

We have $U(s) = 0$. For the p_l, $m = (1 - 2l/L)$, $k_0 < m \leq 1$, we derive an expression

$$p_l = \exp\left(\frac{L}{2}\int_{k_0}^m dm \ln\frac{q - \sqrt{q^2 - 1 + m^2}}{1 + m}\right.$$
$$\left. + \frac{L}{2}\int_s^{k_0} dm \ln\frac{q + \sqrt{q^2 - 1 + m^2}}{1 + m}\right). \tag{A11}$$

For $m < k_0$, we have

$$p_l = \exp\left(\frac{L}{2}\int_s^m dm \ln\frac{q + \sqrt{q^2 - 1 + m^2}}{1 + m}\right). \tag{A12}$$

Nonselective phase

Now the maximum of Eq. (6) is at the border $k_0 = K$, and we have

$$R = f(K)e^{-u+u\sqrt{1-K^2}}. \tag{A13}$$

We use the "$-$" solution of Eq. (A6) for the whole interval $K < m < 1$. We take $U(K) = 0$ as the maximum of population is at the border with $m = K$. For the p_l, $m = (1 - 2l/L)$ we have an expression

$$p_l = \exp\left(L\int_K^m dm\frac{1}{2}\ln\frac{q - \sqrt{q^2 - 1 + m^2}}{1 + m}\right) \tag{A14}$$

and the maximum is for p_d with $m \equiv 1 - \frac{2d}{L} = K$.

For the single-peak fitness case [$f(1) = A$ and $f(m) = 1$ for $m < 1$], we have $q = \sqrt{1 - K^2}$ and

$$x_0 \equiv p_0 \sim \exp\left(\frac{L}{2}\int_K^1 dm \ln\frac{\sqrt{1-K^2} - \sqrt{m^2-K^2}}{1 + m}\right). \tag{A15}$$

For the $1/M$ we have an expression

$$\frac{1}{M} \sim \exp\left(\frac{L}{2}\int_K^1 dm \ln\frac{1 - m}{1 + m}\right). \tag{A16}$$

Intermediate phase

Now mean fitness is given by Eq. (A10) with $K < k_0 < 1$ and $s = K$. We used the "$-$" solution of Eq. (A6) for $k_0 < m < 1$ and the "$+$" solution for $K < m < k_0$. We take $U(K) = 0$ as the maximum of population is at the border with $m = K$. When $m > k_0$, we have

$$p_l = \exp\left(\frac{L}{2}\int_K^{k_0} dm \ln\frac{q + \sqrt{q^2 - 1 + m^2}}{1 + m}\right.$$
$$\left. + \frac{L}{2}\int_{k_0}^m dm \ln\frac{q - \sqrt{q^2 - 1 + m^2}}{1 + m}\right). \tag{A17}$$

In the case of $K < m < k_0$, we have

$$p_l = \exp\left(\frac{L}{2}\int_K^m dm \ln\frac{q + \sqrt{q^2 - 1 + m^2}}{1 + m}\right). \tag{A18}$$

For the SP case, we have $q = \frac{\ln A}{u}$ and

$$x_0 = \exp\left(\frac{L}{2}\int_K^1 dm \ln\frac{\frac{\ln A}{u} + \sqrt{\left(\frac{\ln A}{u}\right)^2 - 1 + m^2}}{1 + m}\right). \tag{A19}$$

Above the transition point $(\ln A)/u = \sqrt{1 - K^2}$, Eq. (19) gives

$$x_0 = \exp\left(\frac{L}{2}\int_K^1 dm \ln \frac{\sqrt{1-K^2}+\sqrt{m^2-K^2}}{1+m}\right). \quad (A20)$$

Comparing Eqs. (A15) and (A20), we see that at the transition point there is a jump, x_0 decreases M_1 times,

$$M_1 \sim \exp\left(\frac{L}{2}\int_K^1 dm \ln \frac{\sqrt{1-K^2}+\sqrt{m^2-K^2}}{\sqrt{1-K^2}-\sqrt{m^2-K^2}}\right). \quad (A21)$$

For $K=0.5$, Eq. (A20) gives $\ln(x_0)/L \approx -0.057$, while $\ln(M)/L \approx -0.56$. Thus above the transition point to the third phase

$$\ln(x_0) \sim \frac{1}{M^{0.1}}. \quad (A22)$$

Consider the case $K=0.99$. Now we have $\ln(x_0)/L \approx -0.01$ and $\ln(M)/L \approx -0.031$. Equation (A20) gives

$$\ln(x_0) \sim \frac{1}{M^{0.32}}. \quad (A23)$$

Jumps of x_l in general case

Consider the evolution model with general fitness function $f(x)$. Assume that without truncation the error threshold transition is a discontinuous one, and there is a jump from nonzero $k_0 > 0$ in the selective phase to $k=0$ solution in Eq. (6) for the nonselective phase. Let us introduce the truncation. Choosing $K < k_0$, we have three phases (see Table III) and x_0 decreases M_2 times at the transition point between phases II and III,

$$M_2 \sim \exp\left(\frac{L}{2}\int_{k_0}^1 dm \ln \frac{\sqrt{1-K^2}+\sqrt{m^2-K^2}}{\sqrt{1-K^2}-\sqrt{m^2-K^2}}\right). \quad (A24)$$

If in the original (without truncation) model the error threshold transition is a continuous one with $k_0=0$, after truncation

we have different expressions for x_0 in the II ($k_0 > K$, $s=K$) and III ($k_0=K$, $s=K$) phases while continuous transitions I $\rightarrow$ II and II $\rightarrow$ III; see Table II. We have a similar behavior for the phase transitions in the case of an originally (without truncation) discontinuous error threshold transition, if the truncation parameter K is chosen too large, $K > k_0$.

Characterization of the intermediate phase

Consider again the intermediate phase. Let us derive an important constraint for the population of the class at the Hamming distance $n \equiv L(1-k_0)/2$. For the corresponding V, we have

$$V(k_0) \equiv \frac{1}{L}\ln \frac{p_n}{\sqrt{N_n}}. \quad (A25)$$

As $p_d \sim 1$ [the majority of population is at the border with the overlap parameter $K=(1-2d/L)$], we have

$$V(K) \equiv \frac{1}{L}\ln \frac{1}{\sqrt{M}} \quad (A26)$$

as $M \approx N_K$ [or $(\ln M - \ln N_K) \ll \ln M$]. We proved before that $V(m)$ has a single maximum [in our case with a single solution for the maximum point k_0 in Eq. (6)]. Thus

$$V(k_0) > V(K), \quad (A27)$$

which gives

$$p_n > \frac{\sqrt{N_n}}{\sqrt{M}}. \quad (A28)$$

The last inequality supports the choice of order parameter in Eq. (17).

For the single-peak case $n=0$, and we get from Eq. (A28)

$$x_0 > \frac{1}{\sqrt{M}}. \quad (A29)$$

Equations (A22) and (A23) give even higher values for x_0.

[1] E. Domingo, C. K. Biebricher, M. Eigen, and J. J. Holland, *Quasispecies and RNA Virus Evolution: Principles and Consequences* (Landes Bioscience, Austin, TX, 2001).

[2] M. Eigen, Proc. Natl. Acad. Sci. U.S.A. **99**, 13374 (2002).

[3] M. Eigen, Naturwiss. **58**, 465 (1971).

[4] C. K. Biebricher, Cold Spring Harbor Symp. Quant. Biol. **52**, 299 (1987).

[5] D. R. Mills, R. L. Peterson, and S. Spiegelman, Proc. Natl. Acad. Sci. U.S.A. **58**, 217 (1967).

[6] J. Swetina and P. Schuster, Biophys. Chem. **16**, 329 (1982).

[7] I. Leuthäusser, J. Stat. Phys. **48**, 343 (1987).

[8] P. Schuster and J. Swetina, Bull. Math. Biol. **50**, 635 (1988).

[9] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).

[10] M. Eigen, J. S. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).

[11] S. Franz and L. Peliti, J. Phys. A **30**, 4481 (1997).

[12] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559

(1997).

[13] E. Baake and H. Wagner, Genet. Res. **78**, 93 (2001).

[14] J. Hermisson, O. Redner, H. Wagner, and E. Baake, Theor Popul. Biol. **62**, 9 (2002).

[15] D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 021913 (2004); **69**, 046121 (2004).

[16] D. B. Saakian, C.-K. Hu, and H. Khachatryan, Phys. Rev. E **70**, 041908 (2004).

[17] D. B. Saakian and C.-K. Hu, Proc. Natl. Acad. Sci. U.S.A. **103**, 4935 (2006).

[18] D. B. Saakian, E. Munoz, C.-K. Hu, and M. W. Deem, Phys. Rev. E **73**, 041913 (2006).

[19] J.-M. Park and M. W. Deem, Phys. Rev. Lett. **98**, 058101 (2007).

[20] D. B. Saakian, J. Stat. Phys. **128**, 781 (2007).

[21] K. Sato and K. Kaneko, Phys. Rev. E **75**, 061909 (2007).

[22] D. B. Saakian, O. Rozanova, and Andrei Akmetzhanov, Phys.

Rev. E **78**, 041908 (2008).

[23] D. B. Saakian, Z. Kirakosyan, and C.-K. Hu, Phys. Rev. E **77**, 061907 (2008).

[24] R. Sanjuan, A. Moya, and S. F. Elena, Proc. Natl. Acad. Sci. U.S.A. **101**, 8396 (2004).

[25] E. Lázaro, C. Escarmis, J. Perez-Mercader, S. C. Manrubia, and E. Domingo, Proc. Natl. Acad. Sci. U.S.A. **100**, 10830 (2003).

[26] M. Pariera, G. Fernandez, B. Clotet, and M. A. Martinez, Mol. Biol. Evol. **24**, 382 (2007).

[27] G. P. Wagner and P. Krall, J. Math. Biol. **32**, 33 (1993).

[28] J. Summers and M. Litwin, J. Virol. **80**, 20 (2006).

[29] B. Drossel, Adv. Phys. **50**, 209 (2001).

[30] E. V. Nimwegen, J. P. Crutchfield, and M. Huynen, Proc. Natl. Acad. Sci. U.S.A. **96**, 9716 (1999).

[31] H. Woodcock and P. G. Higgs, J. Theor. Biol. **179**, 61 (1996).

[32] N. Sourlas, Nature **339**, 693 (1989).

[33] D. B. Saakian, JETP Lett. **55**, 198 (1992).

[34] B. Derrida, Phys. Rev. B **24**, 2613 (1981).

[35] M. Kimura and T. Maruyama, Genetics **54**, 1337 (1966).

[36] J. J. Bull, R. Sanjuan, and C. O. Wilke, J. Virol. **81**, 2930 (2007).

[37] L. C. Evans, *Partial Differential Equations* (American Mathematical Society, Providence 2002).

[38] A. Melikyan, *Generalized Characteristics of First Order PDEs* (Birkhäuser, Boston, 1998).