PHYSICAL REVIEW E **90**, 022712 (2014)

Punctuated equilibrium and shock waves in molecular models of biological evolution

David B. Saakian[*]
Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan; Yerevan Physics Institute, Alikhanian Brothers Str. 2, Yerevan 375036, Armenia; and National Center for Theoretical Sciences: Physics Division, National Taiwan University, Taipei 10617, Taiwan

Makar H. Ghazaryan[†]
Yerevan Engineering University, Teryan St., 105, Yerevan 9, Armenia

Chin-Kun Hu[‡]
Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan

(Received 23 September 2013; revised manuscript received 23 January 2014; published 15 August 2014)

We consider the dynamics in infinite population evolution models with a general symmetric fitness landscape. We find shock waves, i.e., discontinuous transitions in the mean fitness, in evolution dynamics even with smooth fitness landscapes, which means that the search for the optimal evolution trajectory is more complicated. These shock waves appear in the case of positive epistasis and can be used to represent punctuated equilibria in biological evolution during long geological time scales. We find exact analytical solutions for discontinuous dynamics at the large-genome-length limit and derive optimal mutation rates for a fixed fitness landscape to send the population from the initial configuration to some final configuration in the fastest way.

DOI: 10.1103/PhysRevE.90.022712 PACS number(s): 87.23.Kg, 64.60.De

I. INTRODUCTION

Applications of the methods of statistical physics to the study of the molecular models of biological evolution [1–17] and the origin of life [1,18] have attracted much attention in recent decades. In this paper, we use the methods of statistical physics to study punctuated equilibria [19–21]: a well-known phenomenon of biological evolution during long geological time scales, related to the the life tree [22]. A molecular model of biological evolution for such phenomena is still lacking.

It has been observed from fossil records that many species can maintain the same original forms for a very long geological time (called stasis), then rapidly change to very different forms or branch to new species with very different forms. In 1972, Eldredge and Gould called such phenomena “punctuated equilibria” [19–21]. In 1986, Raup observed that magnitudes and time separations for extinctions of species appear in many different scales [23]. In 1987, Bak, Tang, and Wiesenfeld observed that many natural systems can self-organize into critical states without tuning a parameter and proposed the concept of self-organized criticality (SOC) [24]. They also proposed a lattice sand-pile model which can show the behavior of SOC [24,25]. In 1989, Bak, Chen, and Creutz proposed that ecology of interacting species can evolve to a self-organized critical state [26]. In 1993, Bak and Sneppen (BS) proposed a simple lattice model for punctuated equilibrium, which can evolve into a self-organized critical state [27]. In the BS model [27], each lattice site represents a species, which is far from a molecular evolution model for punctuated equilibria. In the present paper, we present a molecular biological evolution model for the dynamics of the mean Hamming distance of mutations $\langle n\rangle$ when the fitness of a sequence depends only on the total number n of mutations from the main (reference) sequence, thus it is defined as $g(n)$. We find that the mean overlap $x^*(t) \equiv 1 - 2\langle n\rangle/L$ between sequences and the reference sequence can show the behavior of a shock wave when $g(n)$ has positive epistasis (L is the genome length). We identify shock waves with punctuated equilibria and thus can have a molecular model of biological evolution with punctuated equilibrium.

An important concept in the modern molecular theory of genetics and biological evolution is epistasis, which means that different genes or mutations are not independent. Epistasis is positive (negative) when the second derivative of the fitness with respect to the number of mutations is positive, i.e., $g''(n) > 0$ [negative, i.e., $g''(n) < 0$] [17]. In positive (negative) epistasis, the effects of two mutations is larger (smaller) than the effects of the addition of two separate mutations.

In the usual statistical physical models of biological evolution, the genome of length L is considered as a collection of L alleles of two types: $+1$ and -1 [1,3,13,14,17], and there are 2^L different sequences labeled by S_i with $0 \leqslant i \leqslant 2^L - 1$. $S_i \equiv (s_i^1, \ldots, s_i^L)$ has the probability $p_i(t)$ to appear at time t and the reproduction rate (fitness) r_i which is independent of time. In the Crow–Kimura (CK) model [3,14], $p_i(t)$ satisfy coupled differential equations in which the mutation and the reproduction appear in different terms. In the Eigen model [1,13], the mutation and the reproduction appear in the same term.

The CK model [3,14] and the Eigen model [1,13] were often studied with the single peak fitness function (also called landscape), in which one reference sequence, say S_0, has higher reproduction rate, and other sequences have small reproduction rate. Without the loss of generality, one can choose every component of S_0 to be $+1$.

A more general version is the symmetric fitness landscape. In the symmetric fitness landscape, r_i depends only on the the Hamming distance n between S_i and S_0, i.e., the number of mutations from S_0 to S_i, and one can write r_i as $g(n)$. This (symmetric) fitness landscape is the first approximation of reality.

---

[*] saakian@yerphi.am
[†] makarghazaryan@gmail.com
[‡] huck@phys.sinica.edu.tw

The CK model [3,14] can be written as

$$\begin{aligned}\frac{dp_i}{dt} &= \sum_j m_{ij} p_j + r_i p_i - p_i \sum_j p_j r_j \\ &\equiv \sum_j A_{ij} p_j - p_i \sum_j p_j r_j, \\ A_{ij} &= \delta_{ij} r_j + m_{ij}. \end{aligned} \tag{1}$$

Here, m_{ij} is the mutation rate from state S_j to state S_i, and r_i is the fitness of state S_i. State S_i state S_j have a Hamming distance $d_{ij} = (L - \sum_k s_i^k s_j^k)/2$. As $d_{ij} = 1$, we choose a mutation rate $m_{ij} = \gamma_f$ when $d_{i0} > d_{j0}$, $m_{ij} = \gamma_b$ when $d_{i0} < d_{j0}$, and $m_{ii} = -d_{i0}\gamma_b - (L - d_{i0})\gamma_f$; as $d_{ij} > 1$, $m_{ij} = 0$ [9]. We can write the fitness r_i, the rescaled mean fitness R, and the mean overlap x^*, respectively, as

$$r_i \equiv L f(s_1, \ldots, s_L), \tag{2}$$

$$LR \equiv \sum_i p_i r_i, \tag{3}$$

$$x^* \equiv 1 - 2 \sum_i p_i d_{i0}/L. \tag{4}$$

There are different ways to obtain the dynamic solution for a biological evolution model. One approach uses some methods of spin glasses [13]. One can use this approach in application to the models with single peaks [13,28–31] or the random-energy-like fitness landscape [32].

Another approach uses the Hamilton–Jacobi equation (HJE) [33–37]. This approach gives the exact dynamics in the case of a smooth fitness landscape [36] in Crow–Kimura [3,14] and Eigen models [1,2,13]. The dynamics was solved in terms of the mean number of mutations in a population for a general symmetric fitness landscape case.

For the quadratic fitness function only smooth dynamics has been found for the case $x_0 \neq 0$ in Ref. [36], where $x_0 \equiv 1 - 2n/L$ (n is the Hamming distance between S_i and S_0) is the overlap of S_i with S_0; for the exponential or quarter fitness functions, discontinuous transitions in the dynamics have been found numerically [36]. Some results of Ref. [36] have been confirmed recently by the quantum-field theoretical method in Ref. [38].

Reference [36] fails to give an analytical theory for these discontinuous dynamics. In this article we give the exact analytical description for this discontinuous dynamics and apply our exact solution of dynamics to investigate some optimization aspects in the dynamics of evolution models: what mutation rates give the fastest dynamics to send the evolving population from the initial state to some final state?

We will consider how the characteristics of the dynamics is influenced by the sign of epistasis, looking for shock waves in the dynamics of $x^*(t)$. It is equivalent to shock waves in the mean fitness $R(t)$ because there is a simple relation between $R(t)$ and $x^*(t)$: $R(t) = f(x^*(t))$ where f is the fitness function of Eq. (2).

The investigation of optimization aspects of the evolution processes is a subject of much current interest [39–45]. Such works are especially important, because the evolutionary-dynamics approach to the cancer cell is one of the central directions in cancer biology [46].

Two types of optimization problems are considered: via mutation rate [40] or via fitness landscapes [41]. In particular, the first type of optimization, via mutation rate, is relevant to biology. Experimental results suggest that mutation rates can vary, e.g., increasing during certain adverse conditions [40]. This phenomenon is referred to as "adaptive mutation," indicating that the mutation rate varies in response to selective pressure [40]. By optimization we mean the fastest dynamics to send the population from the given initial configuration to some final configuration.

Traulsen *et al.* [41] considered this in the context of a related mathematical problem—the famous Brachistochrone problem suggested in 1696 by Johann Bernoulli. Given two mutants, A and B, separated by n mutational steps, the problem is to find the evolutionary trajectory which allows a homogeneous infinite population of A to reach B in the shortest time. In Ref. [41] an approximate solution was given for the case of finite populations with the optimization of the fitness landscape: what fitness landscape gives the fastest dynamics to send the population from the fixed original configuration to the final configuration?

The goals of this paper are to solve the dynamics of biological evolution models, including the case with discontinuities, and to apply exact analytical results to find the fastest dynamics via mutation rate or fitness.

In such studies, we consider the infinite-population model. In reality the evolutionary dynamics is described through the finite number of replicators. In general, the number of different types changes randomly and fluctuates. When the total number of population is much larger than the number of all possible types (genomes), we can describe the evolutionary dynamics by using deterministic equations, and such equations represent the infinite-population model. This is a rather theoretical abstraction but allows calculating rather accurately some mean characteristics of evolving populations like the mean fitness. We assume that the solution of the infinite-population model is the first step in the solution of the finite-population case.

This paper is organized as follows: In Sec. II A, we solve the dynamics of the Crow–Kimura model [3,14] for the asymmetric-mutation case, simply generalizing earlier results in Ref. [36] with symmetric mutations. Asymmetric mutations are chosen according to the methods presented in Ref. [41]. In Sec. II B, we present an analytical theory of discontinuities, which is the key result of the article. In Sec. II C, we investigate the relation of shock waves with the sign of epistasis. In Sec. II D optimal mutation rates are calculated for the original distribution with a fixed overlap. In Sec. III, we solve the optimization problem via fitness. All the results of Secs. II B, IIC, and III are for symmetric-mutation schemes. In Sec. IV we calculate the evolution dynamic for the case of unidirectional mutations and confirm our general result for discontinuities. In Sec. V, we summarize and discuss our results.

II. THE FASTEST DYNAMICS

A. Crow–Kimura model with asymmetric mutations

We consider the infinite-population model with different forward and backward mutation rates and the symmetric-fitness landscape, where the fitness depends on a Hamming distance (the number of different alleles at the corresponding positions) from the reference sequence. Thus $r_i \equiv g(i) = Lf(1-2d_{i,0}/L)$, where the Hamming distance $d_{i,0}$ is the number of "-1" alleles in the ith sequence. We define the lth Hamming class as a collection of all sequences at the Hamming distance l from the reference sequence. In the case that all the sequences with the same l number of -1 alleles have the same probability, it is possible to write the following system of equations for relative class probabilities $P_l \equiv \sum_i p_i \delta(l,d_{i,0})$, $0 \leqslant n \leqslant L$, with p_i from Eq. (1):

$$\begin{aligned}\frac{dP_l}{dt} &= P_l[Lf(m_l) - (L-l)\gamma_f - l\gamma_b]\\ &\quad + \gamma_f(L-l+1)P_{l-1} + \gamma_b(l+1)P_{l+1}, \qquad (5)\end{aligned}$$

where $m_l = 1 - 2l/L$, and P_l are relative probabilities at the Hamming distance l (l mutations), $f(x)$ is a fitness function, and γ_f and γ_b are the forward and backward mutation rates, respectively. The forward mutation means the the mutation increases the Hamming distance from the reference sequence, and backward mutation decreases the Hamming distance to the reference sequence. In Eq. (5), for $l = 0$ and $l = L$ we omit P_{-1} and P_{L+1}. To derive Eq. (5), we neglect the last term in the first two lines of Eq. (1); one can use a procedure similar to that for the Eigen model presented from Eq. (1) to (4) in [13] to show that such neglect does not influence the final result. As in Refs. [33,36], at discrete values of overlap $x = 1 - 2l/L$ we use the ansatz

$$P_l(t) \equiv P(x,t) \sim \exp[Lu(x,t)],$$

then Eq. (5) can be written as the Hamilton–Jacobi equation (HJE) for $u \equiv \ln P(x,t)/L$ [33]:

$$\begin{aligned}&\frac{\partial u}{\partial t} + H(x,u') = 0,\\ &-H(x,p) = f(x) - \frac{\gamma_f(1+x)}{2} - \frac{\gamma_b(1-x)}{2}\\ &\qquad + \gamma_f\frac{1+x}{2}e^{2p} + \gamma_b\frac{1-x}{2}e^{-2p}. \qquad (6)\end{aligned}$$

Here $p \equiv u' \equiv \partial u/\partial x$, the domain of x is $-1 \leqslant x \leqslant 1$, and the initial distribution is $u(x,0) = u_0(x)$. Let us denote the location for the maximum of the distribution $P(x,t)$ as $x^*(t)$. Thus at $x^*(t)$, we have $p = u' = 0$. The differentiation of Eq. (6) with respect to x at $p = 0$ gives

$$-\frac{dx^*}{dt} = \frac{[f'(x^*(t))]}{u''(x^*(t),t)} + [(1+x^*)\gamma_f - (1-x^*)\gamma_b], \qquad (7)$$

where $u'' \equiv \partial^2 u/(\partial x)^2$. We see that the dynamics of the maximum depends on the mutation rates, the fitness and the curvature of the distribution.

Minimizing $-H(x,p)$ via p, we obtain the expression of the evolution potential,

$$U(x) = f(x) + \sqrt{\gamma_b\gamma_f}\sqrt{1-x^2} - \gamma_f\frac{1+x}{2} - \gamma_b\frac{1-x}{2}. \qquad (8)$$

The evolution behavior is defined by the evolution potential [36]. The mean fitness R and the surplus are defined as

$$R = \max[U(x)]|_x, \quad f(s) = R. \qquad (9)$$

In Ref. [36], we solve Eq. (6) for the $\gamma_f = \gamma_b = \gamma$ case by a method of characteristics [47,48]. For the characteristics line $x(t)$ we have a Hamilton equation $dx/dt = dH(x,p)/dp$. By using the identity

$$\begin{aligned}k &\equiv \gamma_f\frac{1+x}{2}e^{2p} + \gamma_b\frac{1-x}{2}e^{-2p},\\ \gamma_f\frac{1+x}{2}e^{2p} - \gamma_b\frac{1-x}{2}e^{-2p} &= \pm\sqrt{k^2 - \gamma_f\gamma_b(1-x^2)}, \qquad (10)\end{aligned}$$

and the Hamilton equation with the Hamiltonian H given by Eq. (6), we obtain

$$\begin{aligned}\dot{x} &= \pm 2\sqrt{k^2 - \gamma_f\gamma_b(1-x^2)},\\ k &= q + \gamma_f\frac{1-x}{2} + \gamma_b\frac{1+x}{2} - f(x), \qquad (11)\end{aligned}$$

where $q \equiv \partial u(x,t)/\partial t$ is constant along the characteristics, like the energy of the particle in classical mechanics, because $q = -H(x,p)$ and $H(x,p)$ does not depend t. As we are interested in the dynamics of the maximum of distribution x^*, and at x^* we have $p = 0$, thus Eq. (6) gives $q = f(x^*)$, which will be used to derive Eq. (13) below. At every point we have two characteristics: moving to the right and left.

We consider the dynamics of the population, initially having a fixed overlap x_0 with the reference (master) sequence. Let us look at the manner of change in the mean overlap of the population $x^*(t^*) = \sum_j P_j(1-2d_{i,0}/L)$ at the moment in time t^*. P_j is the fraction of type j in the population, $d_{j,0}$ is the number of mutations in the jth type (compared with the master sequence), and such a mutant has a fitness $Lf(1-2d_{i,0}/L)$. As time progresses, the overlap distribution spreads out and we therefore focus on the time evolution of the overlap x^* which yields the maximum of this distribution.

Integrating $dt = (dt/dx)dx$, we obtain for the large initial x_0

$$t^* = \frac{1}{2}\left|\int_{x_0}^{x^*} d\xi[F(\gamma,x^*,\xi)]^{-1/2}\right|, \qquad (12)$$

where we have the following expression for F:

$$\begin{aligned}F(\gamma,x^*,\xi) &= \left[f(x^*) + \gamma_f\frac{1+\xi}{2} + \gamma_b\frac{1-\xi}{2} - f(\xi)\right]^2\\ &\quad - \gamma_f\gamma_b(1-\xi^2). \qquad (13)\end{aligned}$$

Equation (12) corresponds to the motion along one characteristic. For the small x_0, we should consider the motion along two characteristics: after the point x_1 we should take the characteristics with sign $-$ in Eq. (12). We get

$$t^* = \frac{1}{2}\int_{x_0}^{x_1}\frac{d\xi}{\sqrt{F(\gamma,x^*,\xi)}} + \frac{1}{2}\int_{x^*}^{x_1}\frac{d\xi}{\sqrt{F(\gamma,x^*,\xi)}}, \qquad (14)$$

and x_1 is the solution of

$$F(\gamma,x^*,x_1) = 0. \qquad (15)$$

In Ref. [36] a symmetric-mutation scheme is considered; $\gamma_f = \gamma_b = \gamma$ with F_s instead of F:

$$F_s(\gamma,x^*,\xi) = [f(x^*) + \gamma - f(\xi)]^2 - \gamma^2(1-\xi^2). \qquad (16)$$

For the quadratic fitness function

$$f(x) = \frac{c}{2}x^2, \tag{17}$$

with the parameter $c > 0$, Eqs. (12) and (14) have real solutions provided that $x^* < 1 - \gamma/c$. This upper bound determines the asymptotic value of the overlap with the reference sequence. Of course, in the case $\gamma/c > 1$ the selective phase is lost and the dynamics drifts in the sequence space so that the asymptotic regime is characterized by a zero overlap with the reference sequence.

To decide which equation to use, we should find the time period when Eq. (14) coincides with Eq. (12). It happens when the first term in Eq. (14) is 0, or $x_1 = x_0$. Thus we need to calculate t_h:

$$t_h = \frac{1}{2}\int_{x_h}^{x_0} d\xi [F_s(\gamma,x_h,\xi)]^{-1/2}, \tag{18}$$

where x_h is a root of $F_s(\gamma,x_h,x_0) = 0$. This equation has a solution provided that $f(x_h) \leqslant f(x_0)$ which, in the case of monotonically increasing fitness, implies $x_h \leqslant x_0$. Thus for a given x_0 and t^* we calculate x_h and then t_h. If $t^* < t_h$ we use Eq. (12), otherwise we use Eq. (14), to obtain $x^* = x^*(t^*)$. Having an analytical dynamics, we can now investigate the optimization problem and the shock waves.

B. Discontinuous dynamics in case of quadratic fitness function

In Ref. [36], analytical dynamics for the symmetric-mutation case, Eqs. (12) and (14) with $F = F_s$, had been derived. Discontinuous dynamics occurs in these formulas as the point of the maximum (or the mean overlap) jumps from one point to another. However, the authors of Ref. [36] failed to describe analytically the discontinuities of $x^*(t)$. The mean fitness is defined as the maximum of $U(x)$. When this function has two maxima at $1 \geqslant x > 0$, there could be a discontinuity in the dynamics, while Ref. [36] could not identify the position of this discontinuity. The point is that there can be singularities in the dynamics, even for the fitness with a single maximum at $x > 0$, when the fitness is too steep.

We performed numerical calculations for symmetric mutations ($\gamma_f = \gamma_b = \gamma$) in order to clarify the character of discontinuous dynamics; see Figs. 1–3. In the selective phase $c > \gamma$, the potential $U(x)$ has a single maximum at $1 > x > 0$. Nevertheless, in some cases the function $x^*(t^*)$ displays jumps.

Figure 1 illustrates the time evolution of x^* for $x_0 = 0.01$. If γ/c is not too small, $x^*(t^*)$ is a monotonic function, and the direct numerics of the system of Eq. (5) well supports the theoretical formulas for $x^*(t^*)$.

For small γ/c the S-shaped curves indicate the existence of a discontinuity in the $x^*(t^*)$ dynamics, calculated by Eq. (5). This threshold phenomenon was overlooked in a previous analysis of this problem which considered a single parameter setting, $c = 2$ and $\gamma = 1$ [36].

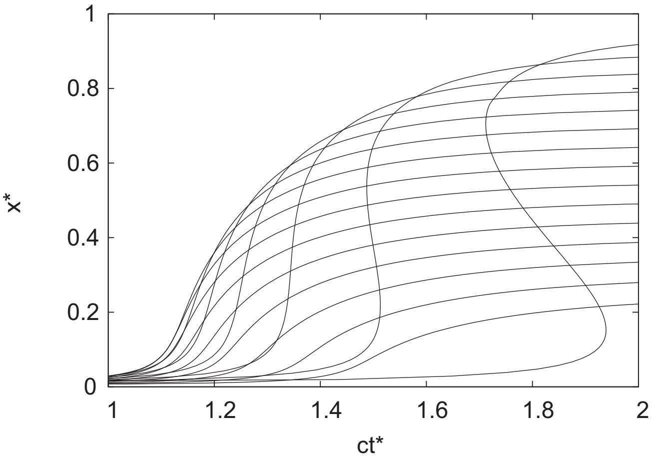

FIG. 1. The dynamics of $x^*(t^*)$ (mean overlap with the reference sequence) as function of time t^*) by Eqs. (12) and (14) for the symmetric case $\gamma_f = \gamma_b = \gamma = 1$ and (top to bottom at $ct^* = 2$) $\gamma/c = 0.05, 0.1, \ldots, 0.7, 0.75$. The initial population has overlap $x_0 = 0.01$ with the reference sequence. For $t^* \to \infty$, we find $x^* = 1 - \gamma/c$.

The unusual time dependence of x^* exhibited in Fig. 1 is rather counterintuitive. We present in Fig. 2 the results of the numerical solution of the ordinary differential equation (ODE) system (5) for different sequence lengths. These results not only confirm the theoretical predictions but complement them. The discontinuous dynamics arises when the curve $x^*(t^*)$, defined by Eq. (14), has an S-like shape. The solution corresponding to the lower branch of the S-like shape is the stable one. This information allows us to obtain the value $t^* = t_d^*$ at which the discontinuity takes place as well as the size of the discontinuity Δx^*. This can be done by locating the lower value of $t^* = t_d$ for which

$$dt^*/dx^* = 0, \tag{19}$$

where t^* as a function of x^* is given by Eq. (14).

Below t_d in same range of t^* there are three different solutions of $x^*(t)$ by Eq. (14), as shown in Fig. 2. We denote them by $x_1^*(t^*)$, $x_2^*(t^*)$, $x_3^*(t^*)$, where $x_1^*(t^*)$ is the low branch

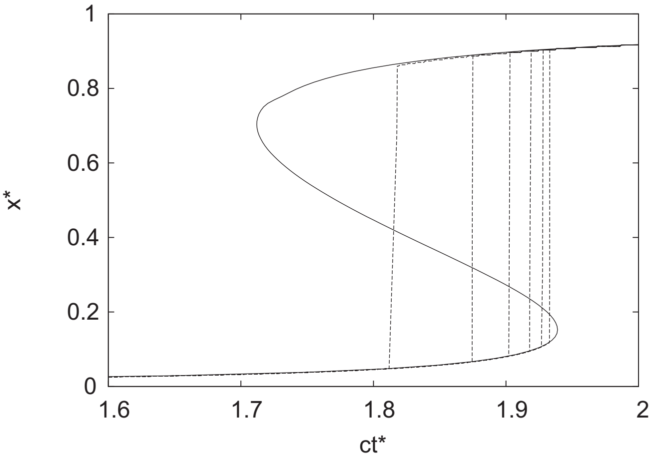

FIG. 2. The dynamics of $x^*(t^*)$ by Eqs. (12) and (14) for $x_0 = 0.01$, $\gamma/c = 0.05$ and by Eq. (1) (dashed vertical lines from left to right) $L = 2000, 4000, \ldots, 12\,000$. For $L \to \infty$ the jump in x^* takes place at $ct^* = ct_d^* = 1.939$ and has the size $\Delta x^* = 0.755$.

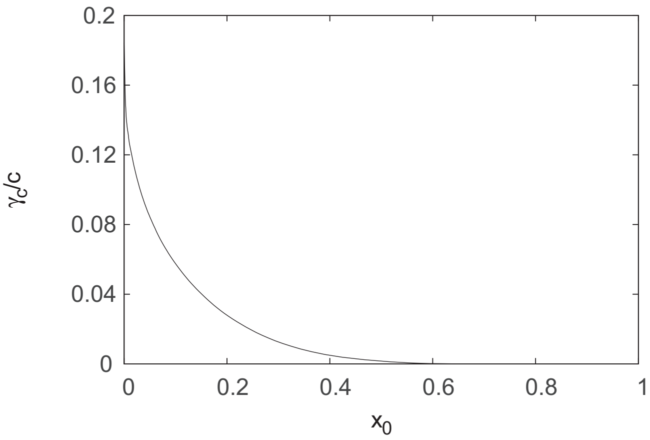

FIG. 3. The critical line γ_c/c vs x_0 at which $\Delta x^* = 0$. The critical c is defined from the system of equations $dx^*/dt^* = 0$, $d^2x^*/d^2t^* = 0$. For the γ values below this curve, the most probable overlap x^* undergoes a discontinuous transition at $t^* = t_d^*$ (see Fig. 2).

and $x_3^*(t^*)$ is the high branch of the S-like shape. At $t^* > t_d$ there is only the solution $x_3^*(t^*)$.

At the $L \to \infty$ limit we have the following solution for the system by Eq. (5):

$$\begin{aligned} x^* &= x_1^*(t^*), \quad t \leqslant t_d, \\ x^*(t^*) &= x_3^*(t^*), \quad t > t_d. \end{aligned} \tag{20}$$

We checked that Eqs. (19) and (20) are also valid for the exponential fitness landscape, as shown in Fig. 4.

C. Sign of epistasis and shock waves

We investigate the existence of shock solutions for the

$$f(m) = cm^n, \tag{21}$$

and

$$f(m) = c(1-m)^n. \tag{22}$$

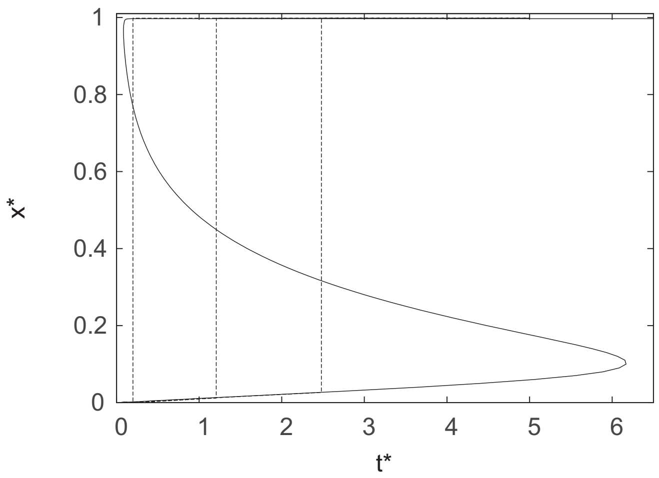

FIG. 4. The relaxation from the original distribution with $x_0 = 0$ for the fitness function $f(m) = 4\exp[(m-1)]$, $\gamma = 0.1$. The dashed lines from left to right correspond to $L = 1000, 5000, 10\,000$. The jump is at the point $dt^*/dx^* = 0$ at $L = \infty$.

Our aim is to investigate the dynamics with small x_0. We construct the curve $x^*(t)$ by Eqs. (12)–(14), and then checked the condition (19).

We use the following expression for the derivatives dt^*/dx^*:

$$\begin{aligned} 2\frac{dt^*}{dx^*} = &-\frac{1}{\gamma x^*} + \frac{\frac{\partial F(x^*,x_0)}{\partial x^*}}{\frac{\partial F(x^*,x_0)}{\partial x_0}} \frac{1}{\sqrt{F(x^*,x_0)}} \\ &- \frac{f'(x^*)}{[-f'(x^*) + \gamma x^*]\,\gamma x^*} \\ &- \int_{x^*}^{x_0} \frac{1}{\sqrt{F(x^*,\xi)}} \frac{d}{d\xi} \frac{\frac{\partial F(x^*,\xi)}{\partial x^*}}{\frac{\partial F(x^*,\xi)}{\partial \xi}} d\xi, \end{aligned} \tag{23}$$

for $t < T_1$, and

$$\begin{aligned} 2\frac{dt^*}{dx^*} = &-\frac{1}{\gamma x^*} - \frac{\frac{\partial F(x^*,x_0)}{\partial x^*}}{\frac{\partial F(x^*,x_0)}{\partial x_0}} \frac{1}{\sqrt{F(x^*,x_0)}} - \frac{f'(x^*)}{[-f'(x^*) + \gamma x^*]\gamma x^*} \\ &- \int_{x_0}^{x_1} \frac{1}{\sqrt{F(x^*,\xi)}} \frac{d}{d\xi} \frac{\frac{\partial F(x^*,\xi)}{\partial x^*}}{\frac{\partial F(x^*,\xi)}{\partial \xi}} d\xi \\ &- \int_{x^*}^{x_1} \frac{1}{\sqrt{F(x^*,\xi)}} \frac{d}{d\xi} \frac{\frac{\partial F(x^*,\xi)}{\partial x^*}}{\frac{\partial F(x^*,\xi)}{\partial \xi}} d\xi, \end{aligned} \tag{24}$$

for $t > T_1$. We analyzed the dynamics by using Eqs. (23) and (24) and *Mathematica*. Within the accuracy of our numerics we could find shocks only for $n > 1.2$

Figure 5 illustrates the loop in $x^*(t)$ dynamics by Eqs. (12)–(14) for the case Eq. (21) with $c = 20$, $x_0 = 0.01$, $n = 1.8$.

For Eq. (22), we considered $n = 1.8$, $x_0 = 0.01$, $c = 10\,000$, as shown in Fig. 6.

According to our numerics, the shock waves exists for the positive epistasis [the sign $f''(m)$ is positive], and absent for the negative epistasis.

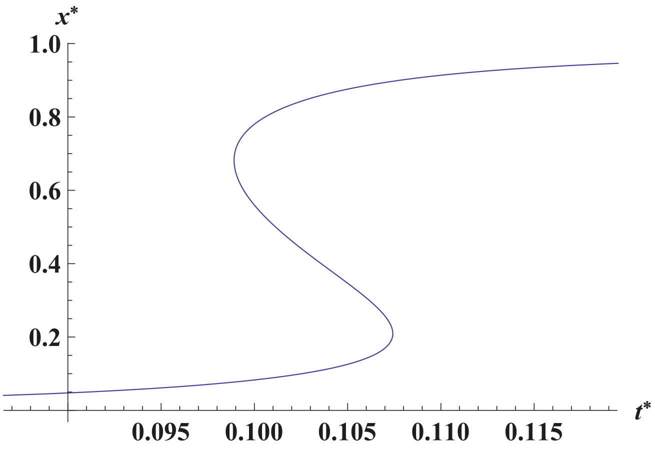

FIG. 5. (Color online) Relaxation from the original distribution with $x_0 = 0.01$, $c = 20$ for the fitness function $f(m) = c|m|^{1.8}$.

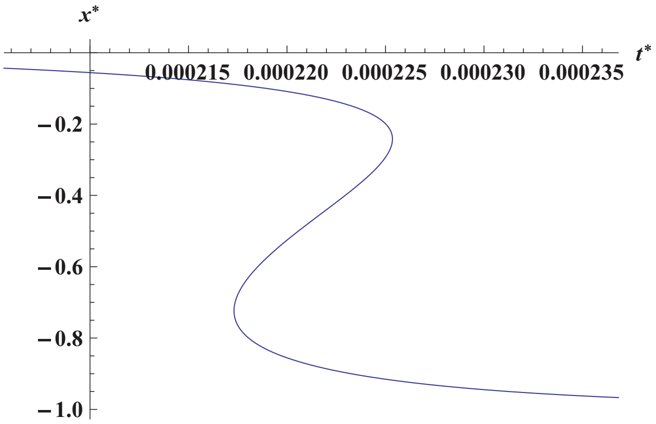

FIG. 6. (Color online) Relaxation from the original distribution with $x_0 = 0.01, c = 10\,000$ for the fitness function $f(m) = c|1-m|^{1.8}$.

D. Mutation rate giving fastest dynamics

Let us now calculate the optimal mutation rates, to get finally $L(1-x^*)$ mutations, when at the start there are more: $L(1-x_0)$ mutations.

We have chosen $x_0 = 0.01$, $x^* = 0.1, 0.2, 0.3, \ldots, 0.8$, and $c = 20$ for the quadratic fitness function of Eq. (17), then used Eqs. (12) and (14) to calculate ct^* as a function of γ/c. We draw the calculated results in Fig. 7.

According to Fig. 7 there is an optimal value of the scaled mutation rate γ/c which minimizes the evolutionary time to go from x_0 to $x^* > x_0$. We note first that this evolutionary trajectory also confirms our theoretical result for the exponential fitness case (possible only for $\gamma/c < 1 - x^*$). We can see that, to reach the endpoint, say, $x^* = 0.2$, it is a bad strategy to choose both small and large values of γ/c. In fact, there is an optimal value of the mutation rate, which for the parameter setting of this example ($x_0 = 0.01$ and $x^* = 0.2$) is $\gamma_{\rm opt}/c = 0.3632$.

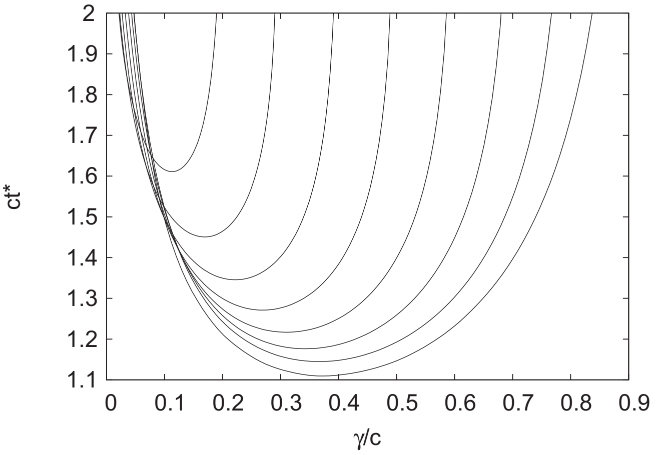

FIG. 7. Time period t^* needed for the maximum of the overlap distribution to reach the values (right to left at $ct^* = 2$) $x^* = 0.1, 0.2, \ldots, 0.8$ as functions of γ/c. The initial population has overlap $x_0 = 0.01$ with the reference sequence. The dynamics can reach x^* provided that $\gamma/c < 1 - x^*$.

To find the optimal mutation rate we just put the condition $dt^*/d\gamma = 0$ to get

$$\frac{1}{\sqrt{\epsilon}}\frac{\frac{dF_s(\gamma,x^*,\xi)}{d\gamma}}{\frac{dF_s(\gamma,x^*,\xi)}{dx_1}} = -\frac{1}{4}\int_{x_0}^{x_1}\frac{d\xi}{F_s^{3/2}}\frac{dF_s(\gamma,x^*,\xi)}{d\gamma} - \frac{1}{4}\int_{x^*}^{x_1}\frac{d\xi}{F_s^{3/2}}\frac{dF_s(\gamma,x^*,\xi)}{d\gamma},$$
$$F_s(\gamma,x^*,x_1) = \epsilon. \tag{25}$$

For the numerical calculations we take $\epsilon = 10^{-6}$. What is surprising is that the optimal mutation rate $\gamma_{\rm opt}$ grows very steeply as x^* departs from x_0 and quickly reaches a maximum value. According to Eq. (25), the optimization depends on the behavior of the fitness function outside the interval $[x_0,x^*]$ when $x_1 > x^*$.

III. THE FASTEST DYNAMICS VIA THE CHOICE OF FITNESS

Although the selection of a fitness function which minimizes the evolution time between any two points x_0 and x^* (the maximum of the overlap distribution in two distinct times) is not as biologically significant as the selection of the optimal mutation rate, it has a considerable aesthetical appeal because the problem is somewhat akin to the Brachistochrone problem in physics [41]. The authors of Ref. [41] assume that the fastest finite population evolution dynamics between two sequences is given by a single peak fitness. However, one should accurately formulate the optimization task. The first possibility is to look at the arrival of finite fraction of population at the peak sequence. The second approach is to look for the arrival of the maximal population at the small distance (the Hamming distance is small compared with L) to the peak sequence, or just arrival of the maximum population at the Hamming class of the peak sequence. The situation is highly nontrivial. If we choose the first approach with some small fraction, then the linear fitness could give better results than the single-peak fitness; see Fig. 8.

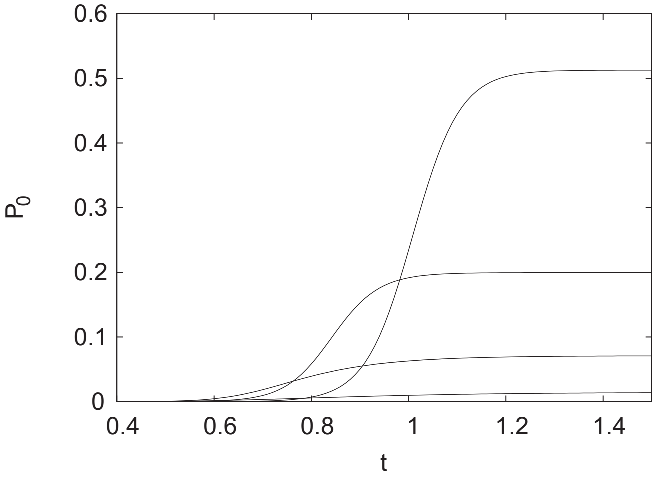

FIG. 8. The dynamics $p_0(t)$ for $L = 20$, symmetric-mutation rate $\gamma = 1$. Fitness $f(x) = 2Lx^a$ for $a = 1, 2, 4$ (from bottom to up at $t = 1.4$). The single peak fitness corresponds to the choice $f(x) = 0$ for $x < 1$ and $f(1) = 2L$.

If we take the second approach to optimization of the dynamics in evolution models, the single-peak fitness appears to be the fastest one. For the case considered (from sequence to sequence), we can provide the expression of the minimal time following the results, following Ref. [13].

For the symmetric-mutation case, the fastest relaxation results in the single-peak fitness landscape ($r_0 = J$ for the peak sequence and $r_i = 0$ for other sequences). In Ref. [13] the relaxation period has been found to send the population from the given sequence [at the Hamming distance $N(1-m)/2$ from the peak sequence] to the peak sequence. To find the minimal time t we add the optimization condition via the choice of γ to the solution of Ref. [13]:

$$\begin{aligned}
t &= \frac{\phi(x,t_1)+Jt_1}{J-\gamma},\\
&\frac{1+x}{2}\tanh(\gamma t_1)+\frac{1-x}{2\tanh(\gamma t_1)}-\frac{J}{\gamma}=0,\\
\phi(x,t) &= \left[\frac{1+x}{2}\ln\cosh(\gamma t)+\frac{1-x}{2}\ln\sinh(\gamma t)\right],\\
&\frac{\partial \frac{\phi(x,t_1)+Jt_1}{J-\gamma}}{\partial\gamma}=0.
\end{aligned} \tag{26}$$

We have performed numerical calculations (see Fig. 8), supporting the choice of single-peak fitness as an optimal fitness for the fastest relaxation in some class of fitness functions, and t by Eq. (26) as a minimal time period.

Consider now the fitness-optimization problem in the case of the distribution with a fixed overlap (to send the population with the initial overlap $m = x_0$ to the eventual distribution with $m = x^*$) and a symmetric fitness landscape. We are looking at the optimization problem for the non-negative fitness function $f(x)$:

$$f(x) \leqslant J, \quad x < x^*; \quad f(x^*) = J. \tag{27}$$

For the fitness function $f(x) = 0,\ x < x^*$, we have

$$t^* = \frac{1}{2}\int_{x_0}^{x^*}\frac{d\xi}{\sqrt{(J+\gamma)^2-\gamma^2(1-\xi^2)}}. \tag{28}$$

It is easy to check that the minimal time is given by the fitness of Eq. (27). Because

$$\begin{aligned}
&\sqrt{(J+\gamma)^2-\gamma^2(1-\xi^2)}\\
&\quad > \sqrt{[(J+\gamma-f(\xi)]^2-\gamma^2(1-\xi^2)},
\end{aligned} \tag{29}$$

the time given by Eq. (28) is less than the time given by any $f(m) > 0$.

IV. UNIDIRECTIONAL-MUTATION CASE

Consider the case of asymmetric mutations [41] with $\gamma_b = 0$. We have an initial distribution at some x_0, and our goal is to send the population to the overlap x^*. Now there is a single characteristic curve at any point. Therefore, contrary to the symmetric-mutation case, all the properties are defined via the behavior of the fitness function in the considered interval $[x_0, x^*]$. Equation (12) gives the following:

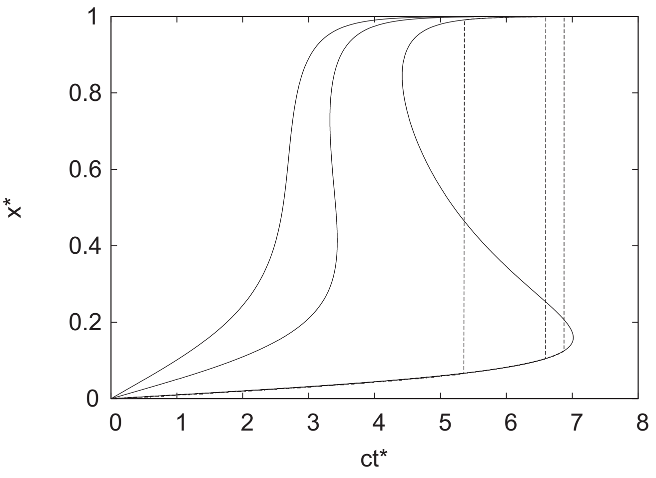

FIG. 9. The dynamics of x^* as a function of time t^* for unidirectional-mutation case with $\gamma_b = 0, x_0 = 0$, (left to right) $\gamma_f/c = 0.1$, 0.05, and 0.01. The numerical solution of the system (5) is given by the dashed vertical lines for (left to right) $L = 1000$, 5000, and 10 000.

$$t^* = \frac{1}{2}\int_{x_0}^{x^*}\frac{dx}{f(x_*)+\gamma_f\frac{1-x}{2}-f(x)}. \tag{30}$$

We see that the optimization via mutation is trivial: by raising the mutation rate we can send the population to the point x^* immediately. The optimization via fitness is also trivial: the fastest trajectory is via the fitness $f(m) = 0,\ m < m^*$, and $f(1) = J_0$.

We have performed numerical calculations for the quadratic fitness case (see Fig. 9). We see that the results again support the conjecture that the jumps occur at the point $dt^*/dx^* = 0$. This is the main result of this section.

V. DISCUSSION

We give conditions for the discontinuous dynamics of evolution models in Eqs. (19) and (20). Such conditions are typical even for the smooth fitness landscape. We have considered the fastest dynamics in evolution models of infinite population, symmetric fitness landscape, and large genome length and found the first exact solution of the problem. Approximate methods for the dynamics of evolution models are too crude [16]. Although good progress has recently been made in the finite-population-optimization problem in evolution, the related problem in infinite population has remained entirely unsolved. In this article, we provide a comprehensive investigation of the problem, including the first exact analytical formulas for the optimization (what mutation rate gives the fastest dynamics for evolution model with fixed fitness landscape).

We found dynamical discontinuities even in the case of quadratic fitness. While the formulas are rather complicated [see Eq. (19)], we found a simple criteria when the shock waves exists: our numerics supports the view that shock waves exist in the case of positive epistasis. Actually, our Eq. (2) describes the evolution in phenotypic space, described via the continuous parameter m. While in genotypic space the evolution is defined via the mutation and selection, in phenotypic space there is a new player, the curvature of the distribution in the phenotypic space [see Eq. (3)]. Another important feature of our main equation (2) is the existence of high degrees of p in the expression of the Hamiltonian. We assume that a similar discontinuous transition exists in other versions of models describing the evolution in phenotype space.

The sharp transitions in evolution are important, because they qualitatively resemble the punctual evolution phenomenon [19–22,27]. The fastest dynamics via the choice of mutation rate is the most intriguing result regarding the adaptive mutation phenomenon. We have calculated the minimal time to send the population from the original sequence with small overlap (with the master sequence) and low fitness to the some final one (with a higher fitness); see Eq. (25). The solution to the mathematical optimization problem is nontrivial; in particular, there exists a certain optimal mutation rate.

The numerical results confirm our analytical predictions. The optimization via fitness landscape (to send the population from the original sequence to the final sequence in the fastest way) should be carefully defined as a mathematical problem. When we are interested in sending some small fraction of the population to the master sequence, the linear fitness can give better results than the single-peak one. When we are looking for the arrival of the maximal population at the small distance (the Hamming distance is small compared with L) to the peak sequence, or just arrival of the maximum of population to the Hamming class of the peak sequence, the optimum is given by a single-peak fitness landscape. Such a hypothesis was assumed first in Ref. [3]. Here, we provide an exact analytical expression for this optimal time period, as well as numerical results illustrating the optimization. If we are looking for a way to send the population with the initial overlap x_0 to the final overlap x^*, then, as we have rigorously proved, the minimal time is given by the single-peak-like fitness (20).

Our infinite-population solutions could be an initial step in consideration of the real biological situation for more involved cases than those considered in Refs. [3,14] with asexual biological evolution models. For example, we can extend the study of this paper to the finite-population problem [49] or the sexual biological evolution model with an approximate neutral fitness function [17]. It will be interesting to apply our findings to cancer, relating them with different stages of tumor progression [46].

ACKNOWLEDGMENTS

D. B. Saakian thanks F. Fontanari for the help in some numerical calculations. This work was supported by Academia Sinica, the Taiwan–Russia collaboration Grant No. NSC 101-2923-M-001-003-MY3, and the National Center for Theoretical Sciences (North).

---

[1] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[2] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[3] J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper Row, New York, 1970).
[4] C. L. Thompson and J. L. McBride, Math. Biosci. **21**, 127 (1974); B. L. Jones, R. H. Enns, and S. S. Rangnekar, Bull. Math. Biophys. **38**, 15 (1976).
[5] I. Leuthausser, J. Stat. Phys. **48**, 343 (1987).
[6] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).
[7] S. Franz and L. Peliti, J. Phys. A: Math. Gen. **30**, 4481 (1997).
[8] L. Peliti, arXiv:cond-mat/9712027.
[9] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559 (1997).
[10] E. Baake and H. Wagner, Genet. Res. **78**, 93 (2001).
[11] B. Drossel, Advances in Physics **50**, 209 (2001).
[12] J. Hermisson, O. Redner, H. Wagner, and E. Baake, Theor. Popul. Biol. **62**, 9 (2002).
[13] D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 021913 (2004).
[14] D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 046121 (2004).
[15] D. B. Saakian and C.-K. Hu, Proc. Natl. Acad. Sci. U. S. A. **103**, 4935 (2006); D. B. Saakian, E. Muñoz, C.-K. Hu, and M. W. Deem, Phys. Rev. E **73**, 041913 (2006).
[16] J. M. Park and M. W. Deem, Phys. Rev. Lett. **98**, 058101 (2007).
[17] D. B. Saakian and C.-K. Hu, Phys. Rev. E **88**, 052717 (2013).
[18] D. B. Saakian, C. K. Biebricher, and C.-K. Hu, PLoS One **6**, e21904 (2011); Z. Kirakosyan, D. B. Saakian, and C.-K. Hu, J. Phys. Soc. Jpn **81**, 114801 (2012).
[19] N. Eldredge and S. J. Gould, in *Models in Paleobiology*, edited by Thomas J. M. Schopf (Freeman, Cooper and Co., San Francisco, 1972), pp. 82–115.
[20] S. J. Gould and N. Eldredge, Paleobiology **3**, 115 (1977).
[21] S. J. Gould and N. Eldredge, Nature (London) **366**, 223 (1993).
[22] S. J. Gould, *The Structure of Evolutionary Theory* (Harvard University Press, Cambridge, 2002).
[23] D. M. Raup, Science **231**, 1528 (1986).
[24] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987); Phys. Rev. A **38**, 364 (1988); P. Bak and K. Chen, Sci. Am. **264**, 46 (1991).
[25] For a recent review, see C.-K. Hu, Chin. J. Phys. **52**, 1 (2014); see also C.-K. Hu, E. V. Ivashkevich, C.-Y. Lin, and V. B. Priezzhev, Phys. Rev. Lett. **85**, 4048 (2000).
[26] P. Bak, K. Chen, and M. Creutz, Nature (London) **342**, 780 (1989).
[27] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993).
[28] M. Ancliff and J. M. Park, Phys. Rev. E **80**, 061910 (2009).
[29] A. G. M. Neves, Phys. Rev. E **82**, 031915 (2010).
[30] M. Ancliff and J. M. Park, Phys. Rev. E **82**, 021904 (2010).
[31] G. Wonpyang, J. Korean Phys. Soc. **57**, 287 (2010).
[32] D. B. Saakian and J. F. Fontanari, Phys. Rev. E **80**, 041903 (2009).
[33] D. B. Saakian, J. Stat. Phys. **128**, 781 (2007).
[34] K. Sato and K. Kaneko, Phys. Rev. E **75**, 061909 (2007).
[35] D. B. Saakian, Z. Kirakosyan, and C.-K. Hu, Phys. Rev. E **77**, 061907 (2008).
[36] D. B. Saakian, O. Rozanova, and A. Akmetzhanov, Phys. Rev. E **78**, 041908 (2008).
[37] N. N. Subbotina and L. G. Shagalova, Proc. Steklov Inst. Math. **277**, 234 (2012).

[38] J. M. Park and M. Antiff, J. Korean Phys. Soc. **61**, 1898 (2012).
[39] P. L. Foster, Annu. Rev. Genet. **33**, 57 (1999).
[40] P. L. Foster, BioEssays **22**, 1067 (2000).
[41] A. Traulsen, Y. Iwasa, and M. A. Nowak, J. Theor. Biol. **249**, 617 (2007).
[42] D. T. Kysela and P. E. Turner, J. Theor. Biol. **249**, 411 (2007).
[43] S. Lenhart and J. T. Workman, *Optimal Control Applied to Biological Models*, in *CRC Mathematical and Computational Biology* (Chapman and Hall, Boca Raton, 2007).
[44] P. Schuster, Complexity **12**, 9 (2007).
[45] R. Chakrabarti, H. Rabitz, S. L. Springs, and G. L. McLendon, Phys. Rev. Lett. **100**, 258103 (2008).
[46] M. Greaves and C. C. Maley, Nature (London) **481**, 361 (2012).
[47] A. Melikyan, *Generalized Characteristics of First Order PDEs* (Birkhäuser, Boston, 1998).
[48] L. C. Evans, *Partial Differential Equations* (AMS, Providence, 2002).
[49] D. B. Saakian, M. W. Deem, and C.-K. Hu, EPL **98**, 18001 (2012).