Optimal mutation rates in dynamic environments: The Eigen model

Mark Ancliff and Jeong-Man Park

Department of Physics, The Catholic University of Korea, Bucheon, Gyeonggi-do 420-743, Korea
(Received 20 April 2010; published 5 August 2010)

We consider the Eigen quasispecies model with a dynamic environment. For an environment with sharp-peak fitness in which the most-fit sequence moves by k spin-flips each period T we find an asymptotic stationary state in which the quasispecies population changes regularly according to the regular environmental change. From this stationary state we estimate the maximum and the minimum mutation rates for a quasispecies to survive under the changing environment and calculate the optimum mutation rate that maximizes the population growth. Interestingly we find that the optimum mutation rate in the Eigen model is lower than that in the Crow-Kimura model, and at their optimum mutation rates the corresponding mean fitness in the eigenmodel is lower than that in the Crow-Kimura model, suggesting that the mutation process which occurs in parallel to the replication process as in the Crow-Kimura model gives an adaptive advantage under changing environment.

I. INTRODUCTION

After the seminal work of Eigen's quasispecies theory [1], the quasispecies models have been popular starting points for the theoretical study of molecular evolution. Quasispecies models describe the evolution of a population of sequences whose dynamics is driven by error-prone replication and selection. Two most widely applied quasispecies models are the Eigen model [1] and the Crow-Kimura model [2]. In the Eigen model, the sequences are subjected to point mutations only during the process of replication, so that the mutation process is connected with the selection process. For this reason the Eigen model is called a connected mutation-selection scheme. On the other hand, in the Crow-Kimura model, mutation and selection are two independent processes and the Crow-Kimura model is called a parallel mutation-selection (ParaMuSe) scheme [3]. Both schemes of mutation-selection are relevant to biology as in DNA-based organisms mutations can occur both during replication and postreplication. Which scheme is more appropriate to describe the evolution in nature depends on whether real evolutionary rates are constant per unit time (as in the Crow-Kimura model) or constant per generation (as in the Eigen model) which is an important topic in current genetics [4]. Moreover, recent theoretical work has shown that the population relaxation rates for two mutation-selection schemes are significantly different [5,6]. Thus it is interesting to compare the dynamic characteristics of two models and make relevant connection to the real evolutionary process.

Studies of the static and dynamic characteristics of quasispecies models until now have focused on static environments [3,7–11]. There have been relatively few studies of quasispecies models with dynamic environments [12–16]. However, real organisms undoubtedly live in changing environments and rate of environmental change has been cited as a key factor in determining the organism's mutation rate [17]. Hence the effort of understanding of the influence of dynamic environments is necessary for more realistic description of evolution.

In our previous work [18], we analyzed the dynamics of the ParaMuSe model with a changing environment. For a dynamic environment with the sharp-peak fitness landscape, we found an asymptotic stationary state and estimated the maximum and the minimum mutation rates for which a quasispecies can survive, and the optimum mutation rate that gives the highest mean fitness. In this paper, we investigate the dynamics of the Eigen model under a changing environment. For a dynamic environment with the sharp-peak fitness function in which the most-fit sequence changes by k spin flips every period T, we find an asymptotic stationary state. By analyzing the stationary state, we obtain the maximum and the minimum mutation rates for a quasispecies to survive and the optimum mutation rate that maximizes the population growth. We also calculate the mean fitness of the Eigen model to compare the result with that of the ParaMuSe model, and show that at their optimum mutation rates the ParaMuSe model outperforms the Eigen model, implying that the parallel mutation-selection scheme gives an adaptive advantage under the changing environment considered in this description.

The paper is organized as follows. In Sec. II we introduce the Eigen quasispecies model and the theoretical methods we use. In Sec. III we describe the regularly changing dynamic environment and find the long-time behavior of the population. In Sec. IV we analyze the long-time behavior of the population and estimate the maximum and the minimum mutation rate thresholds for formation of a quasispecies, as well as the optimum mutation rate. Finally in Sec. V we compare the results for the Eigen model to those of the ParaMuSe model and discuss the significance of these results.

II. EIGEN MODEL

The Eigen model describes the evolution of a population of sequences subject to mutation and selection. A sequence can be represented as a chain of spins $S_i = (s_i^1, \ldots, s_i^N)$ with $s_i^n = \pm 1$ for $1 \le i \le 2^N$ and $1 \le n \le N$. Each sequence in the population represents a different individual and the spins in a sequence represent base pairs in the genome. The number of possible configurations or genotypes is 2^N for a sequence of length N. The probability for a randomly selected sequence

to be in a given configuration i, $1 \leq i \leq 2^N$, at time t is denoted as $x_i(t)$ and satisfies the Eigen model dynamics

$$\frac{dx_i(t)}{dt} = \left[\sum_j W_{ij} f(S_j) x_j(t) \right] - x_i(t) \sum_j f(S_j) x_j(t). \quad (1)$$

Here $f(S_i)$ is the replication rate at which the sequence S_i produces offspring, known as the *fitness*, and the mutation matrix W_{ij} is the probability that an offspring produced by a sequence of type S_j is mutated into a sequence of type S_i. Since we are dealing with asexual reproduction each offspring has only one parent. In a dynamic environment the fitness $f(S_i)$ is time-dependent. The nonlinear term in Eq. (1) serves to enforce the conservation of probability, $\Sigma_i x_i(t) = 1$. Equation (1) can be linearized by a transformation [3]

$$p_i(t) = x_i(t) \exp \left[\int_0^t \sum_j f(S_j) x_j(s) ds \right] \quad (2)$$

leading to a linear system of equations

$$\frac{dp_i(t)}{dt} = \sum_j W_{ij} f(S_j) p_j(t). \quad (3)$$

Introducing the copying fidelity q with which the monomer is copied without error, the probability of any spin being mutated during replication is $(1-q)$ and the mutation matrix W_{ij} can be written as

$$W_{ij} = q^{N-d(i,j)} (1-q)^{d(i,j)} \quad (4)$$

with $d(i,j) = (N - \Sigma_l s_i^l s_j^l)/2$ being the Hamming distance between S_i and S_j, which measures the number of different spins in the two sequences.

Mapping to a quantum system

The linearized Eigen equation can be mapped onto a quantum spin system with introduction of the state vector

$$|\Psi(t)\rangle = \sum_j p_j(t) |S_j\rangle, \quad (5)$$

where $|S_j\rangle$ is the state vector for the spin configuration S_j. Each sequence corresponds to a spin configuration by the mapping $+1 \mapsto \uparrow$, $-1 \mapsto \downarrow$, and the spin configuration $|S_j\rangle$ is given the amplitude p_j in the population state vector $|\psi\rangle$. Equation (3) is equivalent to the evolution of the quantum mechanical spin system described by the imaginary time Schrödinger equation

$$\frac{d}{dt} |\Psi(t)\rangle = -H |\Psi(t)\rangle \quad (6)$$

with the Hamiltonian

$$H = \sum_{l=0}^N q^{N-l} (1-q)^l \sum_{1 \leq i_1 < i_2 \cdots < i_l \leq N} \sigma_{i_1}^x \cdots \sigma_{i_l}^x f(\sigma_1^z, \ldots, \sigma_N^z), \quad (7)$$

where σ_i are the Pauli spin operators acting on the ith spin in the sequence. When they act on the spins, σ^x flips the spins

$\uparrow \mapsto \downarrow$ and $\downarrow \mapsto \uparrow$ describing the mutation process, and σ^z picks up the spin value $+1$ for $\uparrow$ and -1 for $\downarrow$ giving the fitness as a function of the spin configuration. The Hamiltonian describes that the mutation process is connected with the replication process.

We use the sharp-peak fitness function, in which all sequences except one (say S_0) have the same fitness value of 1 and S_0 has higher fitness value of A,

$$f(S_i) = \begin{cases} A & \text{for } S_i = S_0 \\ 1 & \text{else.} \end{cases} \quad (8)$$

The sequence S_0 is referred to as the "most-fit sequence" which is the sequence with the highest replication rate. In a static environment of this kind, there exists a phase transition at the error threshold point $q_c = A^{-1/N}$. If $q > q_c$, a quasispecies is formed around the most-fit sequence, whereas for $q < q_c$ stability of a quasispecies breaks down and no effective selection can occur.

In Ref. [5], Saakian and Hu calculated the transition matrix elements between two configurations in the static environment of the sharp-peak fitness function. To analyze the evolution of the Eigen model under the changing environment, we will use these results, which we summarize below.

If S_i and S_j are a Hamming distance k apart and neither is the most-fit sequence, then the transition matrix element is

$$\langle S_i | e^{-Ht} | S_j \rangle = \exp[N\phi(t, k, \gamma)], \quad (9)$$

with

$$\phi(t, k, \gamma) = \frac{k}{N} \ln \frac{\gamma k}{N} - \frac{k}{N} + \ln \cosh \frac{\gamma e^{-\gamma t}}{N}$$

$$+ \frac{z(1 - \ln z)}{\gamma} \tanh \frac{\gamma e^{-\gamma t}}{N} + \frac{e^{-\gamma t}}{N}$$

$$\times [e^z(1 - z \ln z) - z(1 - \ln z) - 1], \quad (10)$$

where z satisfies the transcendental equation

$$z \left[(e^z - 1) e^{-\gamma t} + \frac{N}{\gamma} \tanh \frac{\gamma e^{-\gamma t}}{N} \right] = k. \quad (11)$$

For S_i a Hamming distance k from S_0, the transition matrix element is found to be

$$\langle S_i | e^{-Ht} | S_0 \rangle = \langle S_0 | e^{-Ht} | S_i \rangle = \exp[N\phi(t_0, k, \gamma)] \langle S_0 | e^{-H(t-t_0)} | S_0 \rangle, \quad (12)$$

where $t_0 \leq t$ is determined by the saddle-point condition

$$-Ae^{-\gamma} + N \frac{d\phi}{dt}(t_0) = 0. \quad (13)$$

Finally, the transition matrix element $\langle S_i | e^{-Ht} | S_j \rangle$ for $S_i = S_j = S_0$ is

$$\langle S_0 | e^{-Ht} | S_0 \rangle = \exp[Ae^{-\gamma t}], \quad (14)$$

where $e^{-\gamma} = q^N$ is the probability of an offspring having no mutations.

III. DYNAMIC ENVIRONMENT

We consider a regularly changing dynamic environment with the sharp-peak fitness function, Eq. (5), in which the most-fit sequence changes by k spin-flips, after each period of time T. The $k=1$ case of the Eigen model in which the most-fit sequence changes by one spin flip at each period T has been studied by Nilsson and Snoad [17]. An equivalent dynamic environment with the ParaMuSe model has been analyzed by exactly up to order of $1/N$ by the authors [18]. The analysis for the Eigen model follows much the same route as that for the ParaMuSe model. We present here the main steps in the derivation, referring the reader to Ref. [18] for more details.

A. Time-scale comparison

In the environment considered, the most-fit sequence changes by k spin flips each period T. Suppose at a certain time all the population is situated at a sequence of a Hamming distance k from the most-fit sequence. The relaxation period, the time taken for a quasispecies to form around the most-fit sequence, is found to be [5]

$$\tau = k \frac{\ln\left(\dfrac{eN\ln(A+1)}{k\gamma}\right)}{Ae^{-\gamma} - 1}. \tag{15}$$

If $T \gg \tau$ then the time scale of the environmental change is much longer than the time taken for the population to relax to the most-fit sequence, and the quasispecies has more than enough time to relax to the new most-fit sequence between each environmental change. Therefore we expect the most-fit sequence population to remain a significant fraction of the total population after many environmental changes. However, if $T \ll \tau$ then the environmental change is too rapid for the population to adapt, in which case we expect no quasispecies can be formed. Thus we are interested in the nontrivial intermediate case, $T \sim \tau$. In this regime, to ensure τ/T tends to a finite nonzero value in the limit of $N \to \infty$, we choose the scaling of the fitness value

$$A = a\ln N, \tag{16}$$

and assume a and γ do not scale with N.

B. Definition of the regularly changing solution

In a regularly changing environment, the asymptotic solution to the Eigen model equation, Eq. (3), is the population distribution that changes regularly with the same period as the environmental change.

In the environment considered here the master sequence changes by k spin flips every period T. We assume these k spins are chosen randomly, and independently of which spins were flipped in the previous environmental changes. Denoting the nth master sequence by S_n, the distance from S_n to S_{n-2} can be anywhere from 0 to $2k$ (less than $2k$ if the same spin is flipped twice in successive environmental changes). For this reason our environment is regularly changing – the distance from the master sequence S_n to $S_{n-1}, S_{n-2}, \ldots$ need

not be the same as the distance from the master sequence S_{n+1} to $S_n, S_{n-1}, \ldots$. However, if $k \ll N$ then the chance of the same spin being flipped twice is very small. More precisely, it is shown in Appendix of Ref. [18] that for $m = N^{1/3}$, the probability that all of the mk spins flipped in m environmental changes are distinct tends to 1 as $N \to \infty$. Thus for large enough N we can assume that the distance from S_n to any S_p with $n - p < N^{1/3}$ is the same as the distance from S_{n+1} to S_{p+1}, so that in the large N limit the environment is approximately regularly changing. Hence we look for a regularly changing asymptotic solution to the Eigen model in the large N limit.

Let $P(t)$ and $p_m(t)$ denote the total population and the population of the most-fit sequence at time t, respectively. In the regularly changing solution we have

$$\frac{p_m(t+T)}{P(t+T)} = \frac{p_m(t)}{P(t)}. \tag{17}$$

In the Eigen model with sharp-peak fitness, the rate of change of total population is given by

$$\frac{dP(t)}{dt} = \sum_j f(S_j)p_j(t) = P(t) + (A-1)p_m(t). \tag{18}$$

From Eqs. (17) and (18) it can be checked that in a regularly changing solution

$$\frac{d}{dt}\left(\frac{P(t+T)}{P(t)}\right) = 0 \tag{19}$$

In other words, both $P(t)$ and $p_m(t)$ increase by the same constant factor over the period T. We denote this factor by the growth constant B,

$$\frac{P(t+T)}{P(t)} = \frac{p_m(t+T)}{p_m(t)} = \text{const} =: B. \tag{20}$$

Hence, in the regularly changing solution the total and the most-fit sequence populations follows a geometric series

$$P(t+nT) = B^n P(t), \quad p_m(t+nT) = B^n p_m(t), \tag{21}$$

and calculating the growth constant B is enough to find the time-averaged mean fitness over the period T. A larger value of B implies the population is better adapted to the environment.

C. Calculation of B

Suppose that at some time t there is a significant fraction of the population at a sequence S_i. To find the fraction of population at another sequence S_j some time t' later, we need to evaluate the transition matrix element $\langle S_j | \mathcal{U}(t', t) | S_i \rangle$, where $\mathcal{U}(t', t)$ is the time-ordered evolution operator $Te^{-\int_t^{t'} H(s)ds}$ with the time dependent Hamiltonian $H(t)$ given by Eq. (7). In principle, to evaluate this matrix element we must take a sum over all possible intermediate configurations $S_i, \ldots, S_k, S_l, \ldots, S_j$, and for each interval (S_k, S_l) evaluate the matrix element $\langle S_k | \mathcal{U} | S_l \rangle$ by using Eqs. (14) and (9), or (12), depending on whether neither, one, or both of (S_k, S_l) are the most-fit sequence for that time period.

For the calculation of $p_m(t)$ in the regularly changing solution only one path gives a significant contribution—
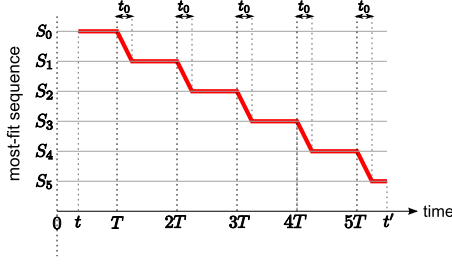
FIG. 1. (Color online) A diagrammatic representation of the significant path for the calculation of $p_m(t')$, representing the matrix element $\langle S_5|\mathcal{U}(t,t')|S_0\rangle$. The horizontal sections of the path correspond to the time spent at the most-fit sequence at a given period, and the slanting sections correspond to the time taken for transitions between most-fit sequences of two neighboring periods.

namely, the path which remains at each successive most-fit sequence for a time $(T-t_0)$, before traveling from one most-fit sequence to the next in time t_0, where t_0 is the saddle-point time defined by Eq. (13). See Fig. 1 for an illustration of this path. The proof that only the contribution from this path is significant is rather involved and can be found in the Appendix.

We use this path to calculate the change in the most-fit sequence population over one period,

$$\frac{p_m((n+1)T-\delta)}{p_m(nT-\delta)} = B, \qquad (22)$$

where δ is a positive infinitesimal. Over the time period from nT to $(n+1)T$ the path travels from the nth most-fit sequence, S_n, to the $(n+1)$-th most fit sequence, S_{n+1} in time t_0, and then remains at S_{n+1} for a time $T-t_0$. Thus,

$$\begin{aligned}
p_m[(n+1)T-\delta] &= p_m(nT-\delta)\langle S_{n+1}|e^{H(T-t_0)}|S_{n+1}\rangle \\
&\quad \times \langle S_{n+1}|e^{Ht_0}|S_n\rangle \\
&= p_m(nT-\delta)\exp[Ae^{-\gamma}(T-t_0)] \\
&\quad \times \exp[N\phi(t_0,k,\gamma)].
\end{aligned} \qquad (23)$$

To obtain the growth constant B, we need to find the saddle-point time t_0. The saddle-point condition Eq. (13) gives

$$-A + \gamma\tanh\left(\frac{\gamma e^{-\gamma}t_0}{N}\right) + z\cosh^{-2}\left(\frac{\gamma e^{-\gamma}t_0}{N}\right) + e^z - z - 1 = 0, \qquad (24)$$

where it should be remembered that z has an implicit time dependence given by Eq. (11). Since we assume $t_0 \ll N$ in the large N limit as we will see later, this equation simplifies to

$$e^z \approx A + 1 \approx A. \qquad (25)$$

Substituting this into Eq. (11), we have

$$\ln A\left[Ae^{-\gamma}t_0 + \frac{N}{\gamma}\tanh\left(\frac{\gamma e^{-\gamma}t_0}{N}\right)\right] = k, \qquad (26)$$
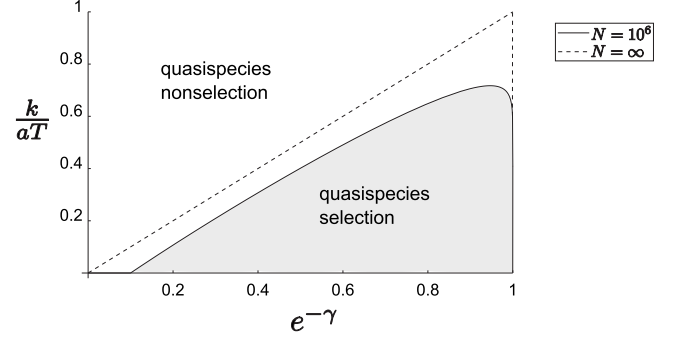
which gives the saddle-point time

FIG. 2. A phase diagram showing the transition between selection and nonselection of the quasispecies ($B>1$ and $B=1$, respectively), as a function of the parameters k/aT and $e^{-\gamma}$. The solid line is the approximate result for $k=10$, $A=10$, and $N=10^6$. The dashed line shows the phase transition in the $N\to\infty$ limit.

$$t_0 \approx \frac{k}{Ae^{-\gamma}\ln A}. \qquad (27)$$

From this we see that $t_0 \sim \frac{1}{\ln N} \ll N$, as assumed.

Putting the saddle-point time t_0 into Eq. (10) and discounting terms of order N^{-2} and smaller gives

$$\begin{aligned}
\phi(t_0,k,\gamma) = \frac{k}{N}\Bigg(&\ln\frac{\gamma k}{N} - 1 + \frac{1}{\ln A} - \ln\ln A - \frac{1-\ln\ln A}{A} \\
&- \frac{1}{A\ln A}\Bigg).
\end{aligned} \qquad (28)$$

The growth constant B is obtained from Eq. (23) with Eqs. (27) and (28),

$$\begin{aligned}
B &= \frac{p_m((n+1)T-\delta)}{p_m(nT-\delta)} \\
&= \exp\Bigg[(ae^{-\gamma}T-k)\ln N - k\Bigg(\ln\ln A - \ln\gamma k + 1 \\
&\quad - \frac{\ln\ln A - 1 - 1/\ln A}{A}\Bigg)\Bigg].
\end{aligned} \qquad (29)$$

IV. ANALYSIS OF THE REGULARLY-CHANGING SOLUTION

When the environmental change is too severe and a stable quasispecies cannot be formed, $p_m=0$ and we have $B=e^T$. If $p_m \neq 0$, then $B>e^T$ and a stable quasispecies is formed. Thus the transition between the two regimes occurs when $Be^{-T}=1$. This gives the criterion for the transition

$$\begin{aligned}
(ae^{-\gamma}T-k)&\ln N + k \\
&\times\left(\ln\ln A - \ln\gamma k + 1 - \frac{\ln\ln A - 1 - 1/\ln A}{A}\right) - T = 0.
\end{aligned} \qquad (30)$$

Figure 2 gives a phase diagram showing this transition between selection and nonselection of the quasispecies in the phase space of the parameters k/aT and $e^{-\gamma}$.
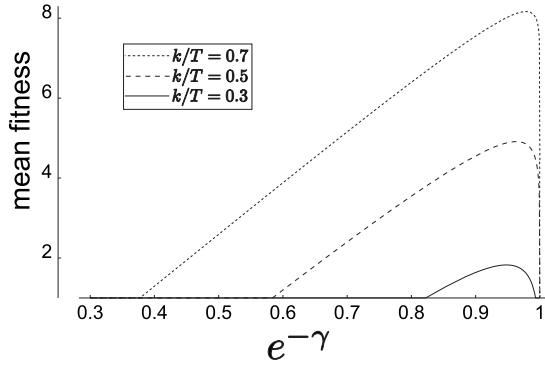
FIG. 3. Time-averaged mean fitness as a function of $e^{-\gamma}$ for three values of the environmental severity parameter k/T. The fixed parameters are $a=1$, $T=20$ and $N=10^6$.

For large enough N, we keep only terms of order $\ln N$ in Eq. (30), giving the upper bound for the mutation rate γ,

$$\gamma_{\max} = \ln \frac{aT}{k}, \tag{31}$$

beyond which the quasispecies cannot be formed. Furthermore, since $\gamma = N \ln q^{-1} \geq 0 (q \leq 1)$, the quasispecies is never selected in the changing environments with $k/aT > 1$.

In the $N \to \infty$ limit, there is no minimal mutation rate threshold - the quasispecies survives for any $0 < \gamma < \gamma_{\max}$. However, for finite N we can give an estimate of the minimum mutation rate threshold by assuming the scaling $\gamma_{\min} \sim N^{-\alpha}$. From Eq. (30), the minimum mutation rate threshold occurs when

$$[k(1 + \alpha) - aT]\ln N = 0, \tag{32}$$

which gives the minimum mutation rate threshold

$$\gamma_{\min} \sim N^{-(aT/k-1)}. \tag{33}$$

We now compute the optimum mutation rate, γ_{opt}, which maximizes the mean fitness, that is, the growth in population over a period. From the relation between the mean fitness and B, the expression for the mean fitness averaged over a period is found to be

$$\langle \bar{f} \rangle = \frac{1}{T} \ln B = A e^{-\gamma} - \frac{k}{T} \bigg(\ln N + \ln \ln A - \ln \gamma k + 1 \\ - \frac{\ln \ln A - 1 - 1/\ln A}{A} \bigg). \tag{34}$$

The first term is the mean fitness of the steady state in the corresponding static environment. Maximizing the mean fitness, we find the optimum mutation rate

$$\gamma_{\mathrm{opt}} = \frac{k}{aT \ln N}. \tag{35}$$

Nilsson and Snoad gave approximate results for the same changing environment with $k=1$. Our findings, Eqs. (31), (32), and (35), are consistent with Nilsson and Snoad's results with $k=1$ in the large N limit. Figure 3 shows the time-

averaged mean fitness as a function of $e^{-\gamma}$ for three values of the environmental severity parameter k/T.

V. DISCUSSION

We investigated the Eigen model under changing environment and found the maximum, minimum, and optimum mutation rates for the given environmental change. In our previous work [18], we examined an equivalent dynamic environment for the Crow-Kimura model (the ParaMuSe model). In our previous findings, the minimum, maximum and optimum mutation rates only depended on the parameters k and T in the combination k/T, suggesting that this parameter was therefore appropriate to describe the "severity" of the environmental changes. For the Eigen model we also find that $\gamma_{\min}$, $\gamma_{\max}$, and γ_{opt} all contain k and T in the combination k/T only.

In order to make a more direct comparison with the ParaMuSe model we construct a mapping between the parameters of the two models. In Ref. [18] we chose the sharp-peak fitness function,

$$f(S_i) = \begin{cases} JN & \text{for } S_i = S_0 \\ 0 & \text{else.} \end{cases} \tag{36}$$

The dynamics of the Crow-Kimura model are invariant under change of the fitness function by a constant $f(S_i) \to f(S_i) + c$. Taking $c=1$ and comparing Eq. (36) to Eq. (8) we find that the Eigen and ParaMuSe fitness functions are equivalent when $A = JN + 1 \approx JN$.

For the mutation rates, Γ in the ParaMuSe model is a mutation rate per unit time, whereas $N(1-q) \approx \gamma$ in the Eigen model is a mutation rate per replication. To calculate the mutation rate per unit time in the Eigen model we multiply γ by the time-averaged mean fitness Eq. (34), which is the average number of replications per unit time in the population. Thus we equate

$$\Gamma \equiv (1 - q)\langle \bar{f} \rangle. \tag{37}$$

At the maximum and minimum mutation rate thresholds we have $\langle \bar{f} \rangle = 1$ since the most-fit sequence population is zero at the threshold. Writing "ParaMuSe" for the ParaMuSe model and using Γ to denote mutation rate per unit time, we find

$$\Gamma_{\max}(\text{ParaMuSe}) = J - \frac{k \ln N}{T},$$

$$\Gamma_{\max}(\text{Eigen}) = \frac{1}{N} - \frac{k \ln N}{NAT} \equiv \frac{1}{A}\Gamma_{\max}(\text{ParaMuSe}). \tag{38}$$

For the minimum mutation rate we find,

$$\Gamma_{\min}(\text{ParaMuSe}) \sim N^{-jT/k} \ln N,$$

$$\Gamma_{\min}(\text{Eigen}) \sim N^{-aT/k} \sim \frac{1}{\ln N}\Gamma_{\min}(\text{ParaMuSe}). \tag{39}$$

Hence the mutation rate per unit time at the thresholds is smaller by a factor $1/\ln N$ in the Eigen model.

For the optimal mutation rate,

$$\Gamma_{\mathrm{opt}}(\mathrm{ParaMuSe}) = \frac{k}{NT},$$

$$\Gamma_{\mathrm{opt}}(\mathrm{Eigen}) = \frac{k}{NT}\left(1 - \frac{k}{aT}\right) = \Gamma_{\mathrm{opt}}(\mathrm{ParaMuSe})\left(1 - \frac{k}{aT}\right).$$

(40)

Thus the optimum mutation rate in the Eigen model is lower, particularly for more severe environments. This result is surprising, since it was shown in Ref. [6] that the relaxation rate for the ParaMuSe model with sharp-peak fitness function is faster than the Eigen model with an equivalent environment. The shorter relaxation time of the ParaMuSe model naively suggest lower optimal mutation rate in comparison with the equivalent Eigen model.

It is also interesting to compare the time-averaged mean fitness at the optimal mutation rate in the two models. Substituting Eq. (35) into Eq. (34) we find the expression for a fitness function in the Eigen model

$$\langle \bar{f} \rangle = Ae^{-k/AT} - \frac{k}{T}\left(\ln N + \ln \ln A - \ln\frac{k^2}{AT} + 1\right.$$

$$\left. - \frac{\ln \ln A - 1 - 1/\ln A}{A}\right)$$

$$\approx A - \frac{k}{T}\left(\ln N + \ln \ln A - \ln\frac{k^2}{AT} + 2\right.$$

$$\left. - \frac{\ln \ln A - 1 - 1/\ln A}{A}\right).$$

(41)

The equivalent expression in the ParaMuSe model is

$$\langle \bar{f} \rangle = A - \frac{k}{T}\left(\ln N - \ln\frac{k^2}{AT} + 2\right).$$

(42)

For any values of the parameters A, k, T for which a quasispecies can be formed, the ParaMuSe model gives a higher time-averaged mean fitness, which implies that the ParaMuSe quasispecies—with mutation occurring in parallel with replication—adapts more effectively to the changing environment at the optimal mutation rate. The difference in mean fitnesses is most significant close to the threshold $k/aT = 1$ above which a quasispecies cannot be formed, and the ratio of mean fitnesses approaches unity as k/aT decreases to zero. See Fig. 4 for an illustration.

In summary, we have studied the Eigen quasispecies model with a dynamic environment. We have derived analytic expressions for the maximum, minimum and optimum mutation rates which are exact in the large N limit. These expressions are similar to equivalent expressions found for the ParaMuSe model in Ref. [18], however, the optimum mutation rate in the Eigen model is lower than that of the ParaMuSe model, especially at high environmental severity k/T. We have shown that, in general, the ParaMuSe model gives a higher time-averaged mean fitness at the optimal mutation rate than the Eigen model. This means that a populat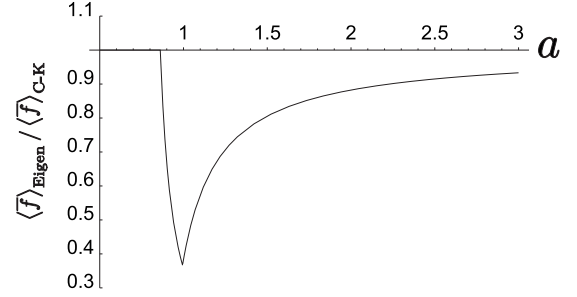ion in which mutation and selection occur in parallel outperforms a population in which mutation occurs only at the moment of reproduction, if the mutation rates are optimal. These findings cast some light on the role of mutations in adaptation to environmental change in asexually reproducing organisms.

FIG. 4. Ratio of time-averaged mean fitness at the optimal mutation rate in the Eigen and ParaMuSe models. Parameters are $N = 10^6$, $k = 15$, $T = 20$. Below $a \approx 0.85$ no quasispecies can be formed and both models have mean fitness 1.

APPENDIX: CALCULATION OF THE SIGNIFICANT PATH

We denote by S_n the most-fit sequence in the time period $nT \leq t < (n+1)T$, and assume that at $t = 0$ all of the population is situated at the sequence S_0. We will find an expression for the population of the sequence S_n just after the nth environment change, i.e., for $p_m(t)$ where $nT < t < (n+1)T$. To do this we consider the sum over all paths from S_0 to S_n in time nT.

Following Ref. [18], for a given path we define the set $V \subseteq \{0, 1, 2, \ldots, n\}$ such that m belongs to V if and only if the path visits the sequence S_l in the time period $lT \leq t < (l+1)T$. We discount paths which visit the sequence S_l more than once in this period, since their contribution is small compared to the path which remains at S_l during the intervening time. Labeling the elements of V by v_i, where $0 = v_1 < v_2 < \cdots < v_m = n$, we define τ_{v_i} as the time spent at most-fit sequence S_{v_i}, and $t_{v_i, v_{i+1}}$ as the time between leaving the sequence S_{v_i} and arriving at sequence $S_{v_{i+1}}$. Due to the regular changes of the environment we have the conditions

$$0 < \tau_{v_i} \leq T,$$

(A1)

$$(v_{i+1} - v_i - 1)T \leq t_{v_i, v_{i+1}} < (v_{i+1} - v_i + 1)T,$$

(A2)

$$\sum_{i=1}^{m} \tau_{v_i} + \sum_{i=1}^{m-1} t_{v_i, v_{i+1}} = t.$$

(A3)

Using the results of Sec. II, the contribution from this path to $p_m(t)$ is

$$\langle S_{v_m}|\exp[-H\tau_{v_m}]|S_{v_m}\rangle\langle S_{v_m}|\exp[-Ht_{v_{m-1},v_m}]|S_{v_{m-1}}\rangle$$

$$\times \langle S_{v_{m-1}}|\exp[-H\tau_{v_{m-1}}]|S_{v_{m-1}}\rangle\langle S_{v_{m-1}}|\exp[-Ht_{v_{m-2},v_{m-1}}]$$

$$\times |S_{v_{m-2}}\rangle \times \cdots \times \langle S_{v_2}|\exp[-H\tau_{v_2}]|S_{v_2}\rangle\langle S_{v_2}|$$

$$\times \exp[-Ht_{v_1,v_2}]|S_{v_1}\rangle \times \langle S_{v_1}|\exp[-H\tau_{v_1}]|S_{v_1}\rangle$$

$$= \exp[Ae^{-\gamma}t]\prod_{i=1}^{m-1}\exp[-Ae^{-\gamma}t_{v_i v_{i+1}} + N\phi(t_{v_i v_{i+1}},d_{v_i v_{i+1}},\gamma)]. \tag{A4}$$

Here $d_{v_i v_{i+1}}$ is the distance between sequences S_{v_i} and $S_{v_{i+1}}$.

We approximate the sum over all paths with the same set V by taking saddle-point values for the times $t_{v_i v_{i+1}}$ in the above expression. If $v_{i+1}=v_i+1$ then the saddle-point time is given by Eq. (27). Otherwise we have from Eq. (24)

$$e^z = A + z + 1 - \gamma\tanh\left(\frac{\gamma e^{-\gamma}t_0}{N}\right) - z\cosh^{-2}\left(\frac{\gamma e^{-\gamma}t_0}{N}\right) > A + 1$$

$$- \gamma. \tag{A5}$$

From the definition of z, Eq. (11), we find

$$t_{v_i v_{i+1}} = \frac{d_{v_i v_{i+1}}}{z(e^z-1)e^{-\gamma}} - \frac{N}{\gamma}\tanh\frac{\gamma e^{-\gamma}t}{N}$$

$$< \frac{k(v_{i+1}-v_i)}{\ln(A+1-\gamma)(A-\gamma)e^{-\gamma}}. \tag{A6}$$

Since $A \sim \ln N$, for large enough N, the saddle-point time in Eq. (A6) is less than the minimum allowable time specified by condition Eq. (A2). Therefore for $v_{i+1}-v_i > 1$ instead of the saddle-point time we take the minimum allowable time

$$t_{v_i v_{i+1}} = (v_{i+1} - v_i - 1)T. \tag{A7}$$

Thus, from Eq. (A4), our approximation for the population of the master sequence at time t is

$$p_m(t) = \exp[Ae^{-\gamma}t]\left(\sum_{V\subseteq\{0,\dots,n\}}\prod_{i=1}^{m-1}\exp[-Ae^{-\gamma}t_{v_i,v_{i+1}}\right.$$

$$\left. + N\phi(t_{v_i,v_{i+1}},d_{v_i,v_{i+1}},\gamma)\right), \tag{A8}$$

where $t_{v_i,v_{i+1}}$ is given by Eqs. (27) and (A7). From Eq. (A8) it follows that if $nT < t_1, t_2 < (n+1)T$ then

$$p_m(t_2) = p_m(t_1)\exp[Ae^{-\gamma}(t_2-t_1)]. \tag{A9}$$

We now consider the population of S_n just after the nth environment change, i.e., $p_m(nT+\delta)$. We split the sum over V in. Equation (A8) as in Ref. [18]. First we consider all paths which visit S_{n-1} (i.e., $n-1 \in V$). These paths must leave S_{n-1} at time $(nT-t_0)$, where t_0 is given by Eq. (27). Second, we consider paths which do not visit S_{n-1} but do visit S_{n-2} (i.e., $n-1 \notin V, n-2 \in V$). We have $t_{n-2,n}=T$, so these paths must remain at S_{n-2} until time $(n-1)T$. Continuing this argument (next considering those paths which visit S_{n-3} but not S_{n-2} or S_{n-1}, and so on), we find

$$p_m(nT+\delta) = \exp[N\phi(t_0,k,\gamma)]p_m(nT-t_0)$$

$$+ \sum_{m=2}^{n}\exp[N\phi((m-1)T,d_{n-m,n},\gamma)]$$

$$\times p_m((n-m+1)T-\delta). \tag{A10}$$

We wish to show that only the first term in Eq. (A10) is significant, since this is exactly the contribution from the path displayed in Fig. 1. Combining Eqs. (27), (A9), and (A10) we have

$$\exp[-Ae^{-\gamma}T] = B^{-1}e^{-k/\ln A}\exp\left[N\phi\left(\frac{k}{Ae^{-\gamma}\ln A},k,\gamma\right)\right]$$

$$+ \sum_{m=2}^{n}B^{-m}\exp[N\phi((m-1)T,d_{n-m,n},\gamma)]. \tag{A11}$$

We can evaluate the first term using Eqs. (10) and (25)

$$e^{-k/\ln A}\exp\left[N\phi\left(\frac{k}{Ae^{-\gamma}\ln A},k,\gamma\right)\right]$$

$$= \exp[k(-\ln N - \ln\ln A + \ln\gamma k - 1)], \tag{A12}$$

where terms in the exponent $\sim \mathcal{O}(\frac{1}{\ln N})$ have been discarded.

We now show that the sum over m in Eq. (A11) can be discarded. As in Ref. [18], we split the sum into two halves, the first from $m=2$ to $m=N^{1/3}$, the second over $m > N^{1/3}$. For $m \le N^{1/3}$ we have $d=mk$ with probability tending to 1 as $N \to \infty$ (see Ref. [18] for a proof), and so

$$\exp\{N\phi[(m-1)T,mk,\gamma]\} = \exp[mk(-\ln N + \ln mk\gamma - 1)$$

$$+ e^{-\gamma}(m-1)T\{A[1-(\ln A)$$

$$\times(\ln\ln A)]-1\}]. \tag{A13}$$

This is a decreasing function of m for $m < N$, and so we can bound each term in the sum of Eq. (A11) by the term with $m=2$. The ratio of the first $N^{1/3}$ terms in the sum to the first term in Eq. (A11) is

$$\frac{\sum_{m=2}^{N^{1/3}}B^{-m+1}\exp[N\phi((m-1)T,mk,\gamma)]}{\exp[k(-\ln N - \ln\ln A + \ln\gamma k - 1)]} \tag{A14}$$

$$< \frac{N^{1/3}\exp[N\phi(T,2k,\gamma)]}{\exp[k(-\ln N - \ln\ln A + \ln\gamma k - 1)]} \tag{A15}$$

$$= \exp\left[\left(-k + \frac{1}{3} - ae^{-\gamma}T((\ln A)(\ln\ln A)-1)\right)\ln N\right.$$

$$\left. + \mathcal{O}(\ln\ln N)\right] \tag{A16}$$

$$\underset{[N\to\infty]}{\to} 0. \tag{A17}$$

Hence the first $N^{1/3}$ terms in the sum can be discarded in the $N \to \infty$ limit.

For the second half of the sum we have

$$\frac{\sum_{m=N^{1/3}+1}^{n} B^{-m+1} \exp\{N\phi[(m-1)T, mk, \gamma]\}}{\exp[k(-\ln N - \ln \ln A + \ln \gamma k - 1)]} \tag{A18}$$

$$< \frac{\sum_{N^{1/3}+1}^{n} B^{-m+1}}{\exp[k(-\ln N - \ln \ln A + \ln \gamma k - 1)]} \tag{A19}$$

$$< \frac{B^{-N^{1/3}}}{1 - B^{-1}} \exp[k(\ln N + \ln \ln A - \ln \gamma k + 1)] \tag{A20}$$

$$\frac{1}{1 - B^{-1}} \exp[-N^{1/3} \ln B + k(\ln N + \ln \ln A - \ln \gamma k + 1)] \tag{A21}$$

$$\underset{[N \to \infty]}{\to} 0, \tag{A22}$$

where we have used the fact that $B > 1$ for a nonzero population p_m. Hence the second half of the sum can also be discarded in the $N \to \infty$ limit, and therefore only the first term in Eq. (A10) is significant, as claimed.

[1] M. Eigen, Naturwiss. **58**, 465 (1971).
[2] J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper and Row, New York, 1970).
[3] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559 (1997).
[4] T. Ohta, Proc. Natl. Acad. Sci. U.S.A. **90**, 10676 (1993).
[5] D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 021913 (2004).
[6] D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 046121 (2004).
[7] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[8] E. Baake, M. Baake, and H. Wagner, Phys. Rev. E **57**, 1191 (1998).
[9] J.-M. Park and M. W. Deem, J. Stat. Phys. **125**, 971 (2006).
[10] D. B. Saakian, E. Munoz, C.-K. Hu, and M. W. Deem, Phys. Rev. E **73**, 041913 (2006).
[11] D. B. Saakian, J. Stat. Phys. **128**, 781 (2007).
[12] M. Nilsson and N. Snoad, Phys. Rev. Lett. **84**, 191 (2000).
[13] C. O. Wilke, C. Ronnewinkel, and T. Martinetz, Phys. Rep. **349**, 395 (2001).
[14] C. Kamp and S. Bornholdt, Phys. Rev. Lett. **88**, 068104 (2002).
[15] Y. Brumer and E. I. Shakhnovich, Phys. Rev. E **71**, 031903 (2005).
[16] M. Ancliff and J.-M. Park, J. Korean Phys. Soc. **53**, 2192 (2008); **54**, 2440 (2009).
[17] M. Nilsson and N. Snoad, Bull. Math. Biol. **64**, 1033 (2002).
[18] M. Ancliff and J.-M. Park, Phys. Rev. E **80**, 061910 (2009).