PHYSICAL REVIEW E **80**, 061910 (2009)

Maximum, minimum, and optimal mutation rates in dynamic environments

Mark Ancliff and Jeong-Man Park
Department of Physics, The Catholic University of Korea, Bucheon, Gyeonggi-do 420-743, Korea

(Received 11 July 2009; published 15 December 2009)

We analyze the dynamics of the parallel mutation-selection quasispecies model with a changing environment. For an environment with the sharp-peak fitness function in which the most fit sequence changes by k spin flips every period T, we find analytical expressions for the minimum and maximum mutation rates for which a quasispecies can survive, valid in the limit of large sequence size. We find an asymptotic solution in which the quasispecies population changes periodically according to the periodic environmental change. In this state we compute the mutation rate that gives the optimal mean fitness over a period. We find that the optimal mutation rate per genome, k/T, is independent of genome size, a relationship which is observed across broad groups of real organisms.

DOI: 10.1103/PhysRevE.80.061910 PACS number(s): 87.23.Cc, 87.15.Zg, 87.10.Mn

I. INTRODUCTION

Over the last 40 years, quasispecies models have emerged as useful tools in the study of molecular evolution [1–12]. Quasispecies models describe the evolution of a population of sequences driven by error-prone replication and selection, where selection is defined by a fitness function on the sequence space. The quasispecies refers to an equilibrium distribution of sequences localized around a sequence of high fitness.

The quasispecies models have been studied in much detail, and analytical results have been found for equilibrium mean fitness [3,6,13,14] and equilibrium population distributions [4,12,15–17]. More recently, interest has turned to the study of the dynamics of these models [5,18–20] and evolution on dynamic fitness landscapes [7–10,21]. In a static environment the mean fitness of a population in equilibrium is always maximized for zero mutation rate, whereas in dynamic environments it is possible that the mean fitness is maximized by a nonzero optimal mutation rate [8] since when too low a mutation rate can reduce the population's ability to adapt to environmental changes.

The factors influencing the mutation rate of real organisms are an open topic in genetics, and so theoretical study of the effects of environmental change on optimum mutation rates is important [22–25]. However, exact results on dynamic fitness landscapes have so far been limited by the difficulties of the analysis, and analytical results have been found for only the simplest dynamic landscapes. In this paper we use recently developed analytical techniques to obtain results for more general dynamic fitness landscapes than have previously been studied.

The paper is organized as follows. In Sec. II we introduce the parallel mutation-selection (ParaMuSe) quasispecies model. In Sec. III we describe the dynamic fitness landscapes to be studied and define the parameters controlling them. In Sec. IV we analyze the effect of the dynamic environment on a quasispecies population and find that in the long-time limit the quasispecies population fluctuates periodically with the environmental change. We calculate conditions for the survival of the quasispecies and the optimal mutation rate. Finally, in Sec. V we give a discussion of the results of Sec. IV.

II. PARALLEL MUTATION-SELECTION MODEL

Two of the most commonly studied quasispecies models are the Eigen model [1] and the Crow-Kimura or ParaMuSe model [2]. In this paper we study the ParaMuSe model with asexual reproduction. For a space of sequences $\{S_i\}$, where the ith sequence has a population $x_i(t)$ and a fitness $f(S_i)$, the ParaMuSe model defines the rate of change of population of each sequence as

$$\frac{dx_i}{dt}(t) = f(S_i)x_i(t) + \Gamma \sum_{d(i,j)=1} [x_j(t) - x_i(t)] - x_i(t)\sum_k f(S_k)x_k(t). \tag{1}$$

Here, x_i is usually interpreted as frequency, fitness $f(S_i)$ is the growth rate that measures the expected number of surviving offspring by a sequence S_i, and Γ is the mutation rate per base per unit time. The first sum runs over all sequences S_j, which differ from S_i by a single mutation, and the last sum runs over the whole sequence space. The last term ensures that the total population $\Sigma_k x_k(t)$ is time independent. The equation can be linearized by a transformation [13],

$$p_i(t) = x_i(t)\exp\left[\int_0^t \sum_k f(S_k)x_k(s)ds\right], \tag{2}$$

leading to a linear system of equations,

$$\frac{dp_i}{dt}(t) = f(S_i)p_i(t) + \Gamma \sum_{d(i,j)=1} [p_j(t) - p_i(t)]. \tag{3}$$

If the sequence with N number of base pairs is represented by a chain of N spins, $S_i = (s_1, \ldots, s_N)$, then this equation can be mapped onto a quantum spin chain [13]. The quantum dynamics are specified by

$$\frac{d}{dt}|\Psi(t)\rangle = -H|\Psi(t)\rangle, \tag{4}$$

$$|\Psi(t)\rangle = \sum_k p_k(t)|S_k\rangle, \tag{5}$$

$$H = -f(\sigma_1^z, \ldots, \sigma_N^z) - \Gamma \sum_{i=1}^{N} (\sigma_i^x - 1). \tag{6}$$

Here, σ_i are the Pauli spin operators on the ith spin in the chain. Starting from this formulation of the ParaMuSe model, several techniques from statistical physics have been used to solve the dynamics [5,14,19]. In particular, the dynamics in the case of the sharp-peak fitness function in the large-N limit was solved exactly in Ref. [19]. In this paper we use the techniques of Ref. [19] to analyze the ParaMuSe model with a dynamic environment.

III. DYNAMIC ENVIRONMENT

We consider an environment with the sharp-peak fitness function

$$f(S_i) = \begin{cases} JN, & \text{for } S_i = S_0 \\ 0, & \text{else.} \end{cases} \tag{7}$$

The sequence S_0 is referred to as the most fit sequence.

In previous papers, the dynamic environments studied have been restricted to three simple cases, all based on the sharp-peak fitness function: those with a fixed most fit sequence with time-dependent $J(t)$ [18,26], those with a most fit sequence which oscillates between two distant sequences [9,21], and those in which the most fit sequence changes location by one spin flip at a time [7,8,11]. Only the third case displays a nonzero optimum mutation rate. In this paper, we consider an environment in which the most fit sequences change by an arbitrary number of spin flips k, after each period of time T. We study the population of the most fit sequence as a proportion of the total population after many periods T.

Time-scale comparison

Suppose at time $t=0$ all the population is situated at a sequence S_i, a Hamming distance k from the most fit sequence (the Hamming distance between two sequences is defined as the number of spin sites at which the two sequences differ). The relaxation period in this case (the time taken for the most fit sequence population to exceed that of any other sequence) was calculated in Ref. [19] and found to be

$$\tau = \frac{k}{N} \frac{\ln \frac{eJN}{k\Gamma}}{J - \Gamma}, \tag{8}$$

where Γ is the mutation rate per site.

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 therefore we expect the most fit sequence population to remain a significant fraction of the total population after many environmental changes (the quasispecies survives). However, if $T \ll \tau$ then the environment change is too rapid for the population to adapt, in which case we expect no quasispecies to form. We are interested in the intermediate case $T \sim \tau$. Since our method is valid only for large N, we need T/τ to tend to a finite nonzero value as $N \to \infty$. We assume that k and T do not scale with N, and so we choose the scaling

$$J = \frac{\ln N}{N} j, \quad \Gamma = \frac{\ln N}{N} \gamma, \tag{9}$$

where j and γ are N independent [note that the ParaMuSe model (3) is invariant under a transformation $J \to J/\alpha$, $\Gamma \to \Gamma/\alpha$, $t \to \alpha t$]. With this choice of scaling it follows from Eq. (8) that T/τ does not scale with N, as required.

IV. ANALYSIS OF THE DYNAMIC ENVIRONMENT

A. Definition of the periodic solution

In any periodically changing environment, the asymptotic solution to the ParaMuSe model (1) is a population distribution that changes periodically with the same period as the environmental change. The existence and uniqueness of such a solution is guaranteed by the Frobenius-Perron theorem (see Sec. 3 of Ref. [9]).

In the environment considered here the most fit sequence changes by k spin flips every period T. We assume that these k spins are chosen randomly and independently of which spins were flipped in the previous environmental changes. Denoting the nth most fit 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 not strictly periodic—the distance from the most fit sequence S_n to $S_{n-1}, S_{n-2}, \ldots$ does not need to be the same as the distance from the most fit 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 A 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 periodic. Hence, we look for a periodic asymptotic solution to the ParaMuSe model in the large-N limit.

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

$$\frac{p_m(t+T)}{P(t+T)} = \frac{p_m(t)}{P(t)}. \tag{10}$$

Note that the solution is periodic in terms of the normalized populations $x_i(t)$ of Eq. (1), not the populations $p_i(t)$ of Eq. (3).

For the sharp-peak fitness function, summing Eq. (3) over all sequences S_i and integrating over time, we obtain

$$P(t) = P(0) + JN \int_0^t p_m(s) ds. \tag{11}$$

Hence, to calculate the total population at time t it suffices to know the most fit sequence population at all intermediate times. From Eqs. (10) and (11) one can show that, in the periodic solution,

$$\frac{P(T+t)}{P(t)} = \text{const} =: A, \tag{12}$$

where

$$A = 1 + \frac{JN}{P(t)} \int_{nT}^{(n+1)T} p_m(t+s)ds. \tag{13}$$

Hence, in the periodic solution the total and the most fit sequence populations follow a geometric series, i.e.,

$$P(t+nT) = A^n P(t), \tag{14}$$

$$p_m(t+nT) = A^n p_m(t). \tag{15}$$

B. Calculation of the most fit sequence population

We now use the method of Ref. [19] to calculate the most fit sequence population at time t, $p_m(t)$. We assume that at time $t=0$ all the population is situated at the zeroth most fit sequence S_0. The population of the zeroth most fit sequence at time $t<T$ is simply given by

$$\langle S_0|e^{-Ht}|S_0\rangle, \tag{16}$$

where H is the ParaMuSe model Hamiltonian [Eq. (6)].

Following Ref. [19], to calculate a matrix element such as Eq. (16), we consider the sum over all possible intermediate configurations, $S_0,\ldots,S_i,\ldots,S_0$, and for each interval we replace H either with the "transverse" interaction, $H_{\text{diff}} = -\Gamma\sum_{i=1}^{N}(\sigma_i^x - 1)$, or the "longitudinal" one, $H_{\text{int}} = -[f(\sigma_i^z) - N\Gamma]$. In the $N\to\infty$ limit only one of these terms will be significant. For the matrix element (16) it can be shown that the dominant contribution comes from the path in which all intermediate states are S_0, in which we consider only the longitudinal interaction, giving

$$\langle S_0|e^{-Ht}|S_0\rangle = \langle S_0|e^{-H_{\text{int}}t}|S_0\rangle = \exp[N(J-\Gamma)t]. \tag{17}$$

Hence,

$$p_m(t) = \exp[N(J-\Gamma)t] \quad \text{for} \quad t<T. \tag{18}$$

In general, to find the population of the nth most fit sequence we need to calculate the matrix element $\langle S_n|\exp[-Ht]|S_0\rangle$, where $nT<t<(n+1)T$. Again, we consider the sum over all intermediate configurations. For a given path $S_n\cdots S_i\cdots S_0$, we check whether the most fit sequence S_m is or is not an intermediate configuration in the time interval $mT<s<(m+1)T$. If it is not, then we choose the transverse interaction H_{diff} for the period $[mT,(m+1)T]$. If it is, then we take the longitudinal interaction H_{int} for the period of time the path remains at most fit sequence S_m. We can discard paths that reach the most fit sequence S_m at time s_0, migrate to another sequence at time s_1, only to return to the most fit sequence S_m at time s_2 [where $mT \le s_0 < s_1 < s_2 \le (m+1)T$] since it can be shown that the sum of contributions from all such paths is much smaller than the contribution from the path which remains at S_m from time s_0 to time s_2. Hence, we need only consider paths which visit each most fit sequence S_m at most once in the period $[mT,(m+1)T]$. An example of such a path is shown in Fig. 1.

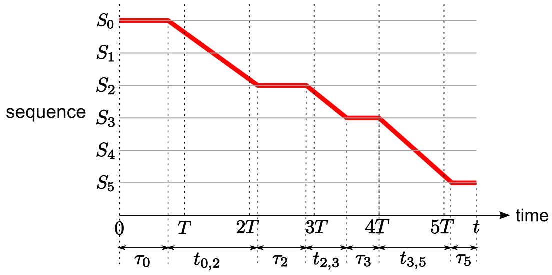

FIG. 1. (Color online) A diagrammatic representation of a path contributing to the matrix element $\langle S_5|\exp[-Ht]|S_0\rangle$. The horizontal sections of the path correspond to periods spent at a most fit sequence [for which we take the longitudinal interaction H_{int} in Eq. (22)], and the diagonal sections correspond to periods between most fit sequences (for which we take the transverse interaction H_{diff}). For this path $V=\{0,2,3,5\}$.

We now compute the contribution from a given path. Let $V \subseteq \{0,1,\ldots,n\}$ be the subset of most fit sequences visited by the path. Denote the elements of V by v_i such that $V = \{v_1,\ldots,v_m\}$ and $0=v_1<v_2<\cdots<v_m=n$. Denote by τ_{v_i} the time spent at most fit sequence S_{v_i} and denote by $t_{v_i,v_{i+1}}$ the time between leaving most fit sequence S_{v_i} and arriving at most fit sequence $S_{v_{i+1}}$ (see Fig. 1). We have the conditions

$$0 < \tau_{v_i} \le T, \tag{19}$$

$$(v_{i+1}-v_i-1)T \le t_{v_i,v_{i+1}} < (v_{i+1}-v_i+1)T, \tag{20}$$

$$\sum_{i=1}^{m} \tau_{v_i} + \sum_{i=1}^{m-1} t_{v_i,v_{i+1}} = t. \tag{21}$$

In the periods τ_{v_i} spent at a most fit sequence we take the longitudinal interaction, and in the periods $t_{v_i,v_{i+1}}$ we take the transverse interaction. The contribution from the path is then

$$\begin{aligned}
&\langle S_{v_m}|\exp[-H_{\text{int}}\tau_{v_m}]|S_{v_m}\rangle\langle S_{v_m}|\exp[-H_{\text{diff}}t_{v_{m-1},v_m}]|S_{v_{m-1}}\rangle\\
&\quad\times\langle S_{v_{m-1}}|\exp[-H_{\text{int}}\tau_{v_{m-1}}]|S_{v_{m-1}}\rangle\\
&\quad\times\langle S_{v_{m-1}}|\exp[-H_{\text{diff}}t_{v_{m-2},v_{m-1}}]|S_{v_{m-2}}\rangle\cdots\langle S_{v_2}|\\
&\quad\times\exp[-H_{\text{int}}\tau_{v_2}]|S_{v_2}\rangle\langle S_{v_2}|\exp[-H_{\text{diff}}t_{v_1,v_2}]|S_{v_1}\rangle\\
&\quad\times\langle S_{v_1}|\exp[-H_{\text{int}}\tau_{v_1}]|S_{v_1}\rangle.
\end{aligned} \tag{22}$$

Reference [19] gives the following expression for a matrix element between two sequences a distance k apart, taking the transverse interaction H_{diff}:

$$\begin{aligned}
\langle S_j|\exp[-H_{\text{diff}}t]|S_i\rangle = \exp[&-N\Gamma t + (N-k)\ln\cosh(\Gamma t)\\
&+ k\ln\sinh(\Gamma t)].
\end{aligned} \tag{23}$$

Equation (22) then becomes

$$
\begin{aligned}
&\exp[N(J-\Gamma)(\tau_{v_1}+\tau_{v_2}+\cdots+\tau_{v_m})]\\
&\quad\times\prod_{i=1}^{m-1}\exp[-N\Gamma t_{v_i,v_{i+1}}+(N-d_{v_i,v_{i+1}})\ln\cosh(\Gamma t_{v_i,v_{i+1}})\\
&\quad+d_{v_i,v_{i+1}}\ln\sinh(\Gamma t_{v_i,v_{i+1}})]\\
&\quad=\exp[N(J-\Gamma)t]\prod_{i=1}^{m-1}\exp[-NJt_{v_i,v_{i+1}}+(N-d_{v_i,v_{i+1}})\\
&\qquad\times\ln\cosh(\Gamma t_{v_i,v_{i+1}})+d_{v_i,v_{i+1}}\ln\sinh(\Gamma t_{v_i,v_{i+1}})], \qquad (24)
\end{aligned}
$$

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

Finally, we approximate the contribution from all paths with the same set V by taking saddle-point values for the times $t_{v_i,v_{i+1}}$. The saddle-point values were found in Ref. [19] [Eq. (17) of Ref. [19]],

$$
\tanh(\Gamma t_{v_i,v_{i+1}})=\frac{2d_{v_i,v_{i+1}}/N}{\dfrac{J}{\Gamma}+\sqrt{\left(\dfrac{J}{\Gamma}\right)^2-\dfrac{4d_{v_i,v_{i+1}}(N-d_{v_i,v_{i+1}})}{N^2}}}. \qquad (25)
$$

If $v_{i+1}=v_i+1$ (i.e., successive most fit sequences are visited), then $d_{v_i,v_{i+1}}=k$, and

$$
t_{v_i,v_{i+1}}=\frac{k}{NJ}+O(N^{-2}) \quad \text{for} \quad v_{i+1}=v_i+1. \qquad (26)
$$

If $v_{i+1}>v_i+1$ then it needs to be checked that the saddle-point time (25) satisfies condition (20). In fact it can be shown that for large enough N the saddle-point time is always less than the minimum time of condition (20), that is, $(v_{i+1}-v_i-1)T>t_{v_i,v_{i+1}}$. The proof of this fact is given in Appendix B. In this case instead of taking the saddle-point time given by Eq. (25), we take the minimum allowable time,

$$
t_{v_i,v_{i+1}}=(v_{i+1}-v_i-1)T \quad \text{for} \quad v_{i+1}>v_i+1. \qquad (27)
$$

Thus, our approximation for the population of the most fit sequence at time t is given by

$$
\begin{aligned}
p_m(t)=\exp[N(J-\Gamma)t]\sum_{V\subseteq\{0,\ldots,n\}}\Bigg(\prod_{i=1}^{m-1}\{\exp[-NJt_{v_i,v_{i+1}}\\
+(N-d_{v_i,v_{i+1}})\ln\cosh(\Gamma t_{v_i,v_{i+1}})\\
+d_{v_i,v_{i+1}}\ln\sinh(\Gamma t_{v_i,v_{i+1}})]\}\Bigg), \qquad (28)
\end{aligned}
$$

where the times $t_{v_i,v_{i+1}}$ are given by Eqs. (26) and (27). From Eq. (28) it follows that, if $nT<t_1,t_2<(n+1)T$ then,

$$
p_m(t_2)=p_m(t_1)e^{N(J-\Gamma)(t_2-t_1)} \qquad (29)
$$

since the sum over V in Eq. (28) is identical for $p_m(t_1)$ and $p_m(t_2)$.

Let us consider the population of S_n at time $t=nT+\delta$, that is, just after the environmental change. We split the sum over all paths arriving at S_n at time t into a sum over paths which visit most fit sequence S_{n-1} $(n-1\in V)$ and a sum over paths which do not $(n-1\notin V)$.

For the paths which visit most fit sequence S_{n-1} we have $t_{n-1,n}=\frac{k}{NJ}$, and thus these paths must leave S_{n-1} at time $(nT-\frac{k}{NJ})$. Hence,

$$
\begin{aligned}
&p_m(nT+\delta)\\
&\quad=\left[\text{sum over paths at } S_{n-1} \text{ at time } \left(nT-\frac{k}{NJ}\right)\right]\\
&\qquad+(\text{sum over paths not visiting } S_{n-1})\\
&\quad=\langle S_n|\exp\left(-H_{\text{diff}}\frac{k}{NJ}\right)|S_{n-1}\rangle\langle S_{n-1}|\exp\left[-H\left(nT-\frac{k}{NJ}\right)\right]\\
&\qquad\times|S_0\rangle+(\text{sum over paths not visiting } S_{n-1})\\
&\quad=\langle S_n|\exp\left(-H_{\text{diff}}\frac{k}{NJ}\right)|S_{n-1}\rangle p_m\left(nT-\frac{k}{NJ}\right)\\
&\qquad+(\text{sum over paths not visiting } S_{n-1}). \qquad (30)
\end{aligned}
$$

In a similar way, we split the sum over paths not visiting S_{n-1} into those which visit S_{n-2} and those which do not. For those which do, we have $t_{n-2,n}=T$, so these paths must remain at S_{n-2} until time $(n-1)T$. Thus,

$$
\begin{aligned}
p_m(nT+\delta)=\langle S_n|\exp\left(-H_{\text{diff}}\frac{k}{NJ}\right)|S_{n-1}\rangle p_m\left(nT-\frac{k}{NJ}\right)\\
+\langle S_n|\exp(-H_{\text{diff}}T)|S_{n-2}\rangle p_m[(n-1)T-\delta]\\
+(\text{sum over paths not visiting } S_{n-1} \text{ or } S_{n-2}). \qquad (31)
\end{aligned}
$$

Continuing this argument, we obtain

$$
\begin{aligned}
p_m(nT+\delta)=\langle S_n|\exp\left(-H_{\text{diff}}\frac{k}{NJ}\right)|S_{n-1}\rangle p_m\left(nT-\frac{k}{NJ}\right)\\
+\sum_{m=2}^{n}\langle S_n|\exp[-H_{\text{diff}}(m-1)T]|S_{n-m}\rangle p_m\\
\times[(n-m+1)T-\delta]. \qquad (32)
\end{aligned}
$$

C. Calculation of A

Using Eq. (32) we now find an expression for the growth constant A in the periodic solution, in terms of the parameters k, T, j, γ, and N. First, we show that in the periodic solution described in Sec. IV A only the first term in Eq. (32) is significant.

In the periodic solution, from Eqs. (15), (29), and (32) we have,

$$\begin{aligned} e^{-N(J-\Gamma)T} &= A^{-1}e^{-(J-\Gamma)k/J}\langle S_n|\exp\left(-H_{\text{diff}}\frac{k}{NJ}\right)|S_{n-1}\rangle \\ &\quad + \sum_{m=2}^{n} A^{-m}\langle S_n|\exp[-H_{\text{diff}}(m-1)T]|S_{n-m}\rangle . \end{aligned} \tag{33}$$

For the first term, using that $\Gamma\frac{k}{NJ} \ll 1$ we have, from Eq. (23),

$$e^{-(J-\Gamma)k/J}\langle S_n|\exp\left[-H_{\text{diff}}\frac{k}{NJ}\right]|S_{n-1}\rangle \approx \exp\left[-k\left(\ln\frac{NJ}{\Gamma k}+1\right)\right]. \tag{34}$$

To calculate the matrix elements for the other terms we need the distance $d_{n,n-m}$ from the nth most fit sequence to the $(n-m)$th most fit sequence. It is shown in Appendix A that, for $m < N^{1/3}$, the probability that this distance is exactly mk tends to 1 as $N \to \infty$. Using this, we can show that all the terms in the sum of Eq. (33) with $m < N^{1/3}$ are negligible for large enough N, i.e.,

$$\lim_{N\to\infty} \frac{\sum_{m=2}^{N^{1/3}} A^{-m}\langle S_n|\exp[-H_{\text{diff}}(m-1)T]|S_{n-m}\rangle}{\exp\left[-k\left(\ln\frac{NJ}{\Gamma k}+1\right)\right]} = 0. \tag{35}$$

From Eq. (23) we have,

$$\begin{aligned} &\langle S_n|\exp[-H_{\text{diff}}(m-1)T]|S_{n-m}\rangle \\ &\quad \approx \exp\{-N(m-1)\Gamma T + mk\ln[(m-1)\Gamma T]\}, \end{aligned} \tag{36}$$

where we have used the fact that $(m-1)\Gamma T = (m-1)\gamma T \ln N/N \ll 1$. The right-hand side of Eq. (36) decreases as m increases (for $m < N$); thus, we can bound each term in the sum of Eq. (35) by the term with $m=2$,

$$\begin{aligned} &\sum_{m=2}^{N^{1/3}} A^{-m}\langle S_n|\exp[-H_{\text{diff}}(m-1)T]|S_{n-m}\rangle \\ &\quad < N^{1/3}\langle S_n|\exp(-H_{\text{diff}}T)|S_{n-2}\rangle \\ &\quad = N^{1/3}\exp[-N\Gamma T + 2k\ln(\Gamma T)] \\ &\quad = \exp[-(2k+\gamma T-1/3)\ln N + 2k\ln(\gamma T\ln N)]. \end{aligned} \tag{37}$$

Hence,

$$\begin{aligned} &\frac{\sum_{m=2}^{N^{1/3}} A^{-m}\langle S_n|\exp[-H_{\text{diff}}(m-1)T]|S_{n-m}\rangle}{\exp\left[-k\left(\ln\frac{NJ}{\Gamma k}+1\right)\right]} \\ &\quad < \frac{\exp[-(2k+\gamma T-1/3)\ln N + 2k\ln(\gamma T\ln N)]}{\exp\left[-k\left(\ln\frac{NJ}{\Gamma k}+1\right)\right]} \\ &\quad = \exp\left[-(k+\gamma T-1/3)\ln N + 2k\ln\left(\sqrt{\frac{\gamma jT}{k}}\ln N\right)+k\right]. \end{aligned} \tag{38}$$

The last line goes to zero as $N \to \infty$, giving the required result.

Now we show that the sum of terms with $m \geq N^{1/3}$ in Eq. (33) can also be discarded. Assuming $A > 1$ (which will always be the case for nonzero most fit sequence population), we have

$$\begin{aligned} &\frac{\sum_{m=N^{1/3}}^{n} A^{-m}\langle S_n|\exp[-H_{\text{diff}}(m-1)T]|S_{n-m}\rangle}{\exp\left[-k\left(\ln\frac{NJ}{\Gamma k}+1\right)\right]} \\ &\quad < \frac{\sum_{m=N^{1/3}}^{n} A^{-m}}{\exp\left[-k\left(\ln\frac{NJ}{\Gamma k}+1\right)\right]} \\ &\quad < \frac{A^{-N^{1/3}}}{(1-A^{-1})\exp\left[-k\left(\ln\frac{Nj}{\gamma k}+1\right)\right]} \\ &\quad = \frac{\exp\left[k\ln\left(\frac{j}{\gamma k}\right)+k\right]}{1-A^{-1}}\exp[-N^{1/3}\ln A + k\ln N]. \end{aligned} \tag{39}$$

Again, the last line goes to zero as $N \to \infty$.

Hence, for large enough N, only the first term in Eq. (33) is significant, and we have

$$\begin{aligned} e^{-N(J-\Gamma)T} &= A^{-1}e^{-(J-\Gamma)k/J}\langle S_n|\exp\left[-H_{\text{diff}}\frac{k}{NJ}\right]|S_{n-1}\rangle \\ &= A^{-1}\exp\left[-k\left(\ln\frac{NJ}{\Gamma k}+1\right)\right]. \end{aligned} \tag{40}$$

Therefore, we obtain the desired expression for A in terms of the parameters k, T, j, γ, and N,

$$A = \exp\left[[-k+(j-\gamma)T]\ln N - k\ln\frac{j}{\gamma k} - k\right]. \tag{41}$$

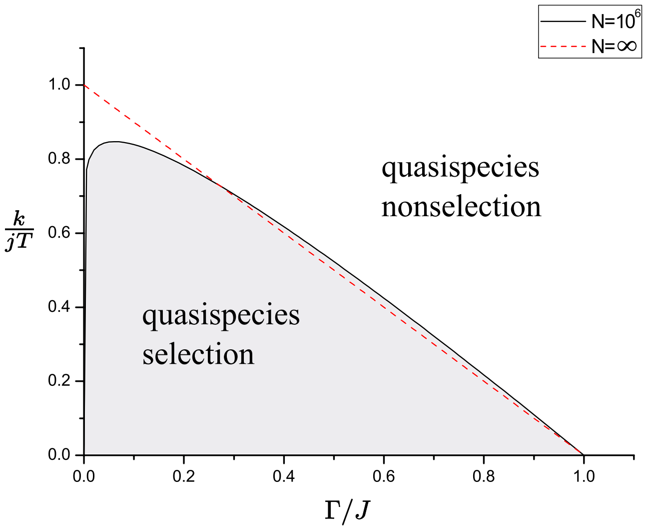

FIG. 2. (Color online) A Phase diagram showing the transition between zero and nonzero average most fit sequence populations (selection and nonselection of the quasispecies) according to the parameters k/jT and Γ/J. The solid boundary is the approximate result for $k=10$, $N=10^6$. The dashed line is the phase transition in the $N\rightarrow\infty$ limit.

D. Analysis of the periodic solution

From Eq. (41) we can calculate the proportion of population at the most fit sequence in the periodic solution as follows. From Eq. (11) we have,

$$P(nT) = P([n-1]T) + JN\int_{(n-1)T}^{nT} p_m(s)ds. \tag{42}$$

We can compute the integral using Eq. (29), leading to

$$P(nT) = P([n-1]T) + \frac{J}{J-\Gamma}p_m(nT-\delta). \tag{43}$$

Using Eqs. (15) and (41) we obtain the proportion of population at the most fit sequence just before an environmental change,

$$\begin{aligned}\frac{p_m(nT-\delta)}{P(nT)} &\\ &= (1-\gamma/j)\left\{1-\exp\left[[k-(j-\gamma)T]\ln N + k\ln\frac{j}{\gamma k}+k\right]\right\}.\end{aligned} \tag{44}$$

The first term is the asymptotic most fit sequence population in the corresponding static environment (sharp-peak fitness function with height J and mutation rate Γ).

The transition between selection and nonselection of the quasispecies occurs when the right-hand side of Eq. (44) is zero, that is,

$$[k-(j-\gamma)T]\ln N + k\ln\frac{j}{\gamma k}+k = 0. \tag{45}$$

Figure 2 gives a diagram showing this transition. For large enough N we can neglect the last two terms, giving the mutation rate threshold

$$\gamma_{\max} = j - \frac{k}{T}. \tag{46}$$

For mutation rates $\gamma > \gamma_{\max}$ the quasispecies is not selected. In terms of the original parameters J and Γ we have,

$$\Gamma_{\max} = J - \frac{k\ln N}{TN}. \tag{47}$$

It is interesting to note that in the $N\rightarrow\infty$ 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 a scaling $\Gamma=\gamma' N^{-\alpha}\ln N$ (it can be checked that the derivations of Secs. IV B and IV C remain valid with this scaling). From Eq. (44) the minimum threshold occurs when

$$(\alpha k - jT)\ln N + k\ln\frac{j}{\gamma' k} + k = 0. \tag{48}$$

Neglecting the last two terms we obtain

$$\alpha = \frac{jT}{k}. \tag{49}$$

So for the minimum mutation rate threshold we have

$$\gamma_{\min} \sim N^{1-jT/k}. \tag{50}$$

In terms of the parameter Γ,

$$\Gamma_{\min} \sim N^{-jT/k}\ln N. \tag{51}$$

We now compute the optimal mutation rate—that is, the mutation rate that optimizes the growth in population over one period, A. From Eq. (41) the optimum is

$$\gamma_{\rm opt} = \frac{k}{T\ln N}. \tag{52}$$

It is interesting to note that the optimal mutation rate is independent of the sharp-peak height J. In terms of the parameter Γ,

$$\Gamma_{\rm opt} = \frac{k}{NT}. \tag{53}$$

Figures 3 and 4 show the most fit sequence populations as a function of mutation rate for a variety of values of the parameters k, T, j, and N.

Finally, we give an expression for the mean fitness averaged over one period,

$$\left\langle JN\frac{p_m(t)}{P(t)}\right\rangle_T = \frac{1}{T}\ln A = (j-\gamma)\ln N - \frac{k}{T}\left(\ln N + \ln\frac{j}{\gamma k} + 1\right). \tag{54}$$

The first term is the mean fitness of the asymptotic solution in the corresponding static environment.

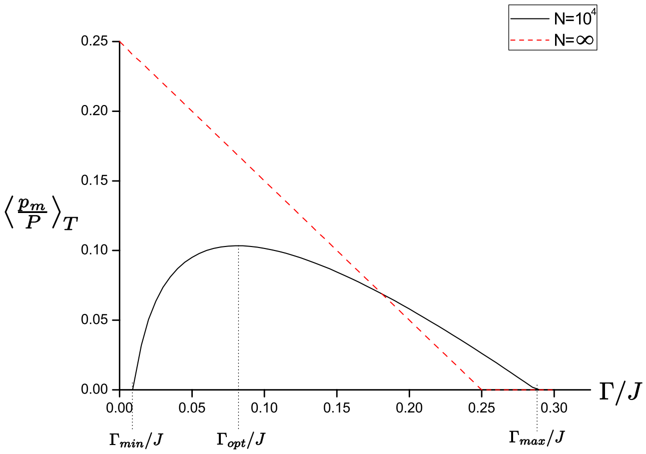

FIG. 3. (Color online) A graph of average most fit sequence population in the periodic solution, against mutation rate. The solid line shows the result for parameters $j=1$, $T=20$, $k=15$, and $N=10^4$. The maximum, minimum, and optimum mutation rates are indicated. The dashed line shows the result in the $N\to\infty$ limit—note that there is no minimum mutation rate threshold in this case.

V. DISCUSSION

In summary we have calculated the asymptotic behavior of a quasispecies population in a periodically changing environment. We have found expressions for the minimal, maximal, and optimal mutation rates, which are exact in the large-N limit.

The minimum, maximum, and optimum mutation rates all involve the parameters k and T in the combination k/T. The parameter k/T can thus be taken as a measure of the severity of the dynamic environment—the higher the value of k/T the more severe the environment. The result for the optimum mutation rate has a particularly simple interpretation: the optimal mutation rate occurs when the average number of mutations per sequence per unit time (ΓN) is the same as the average number of spin flips in the most fit sequence per unit time (k/T).

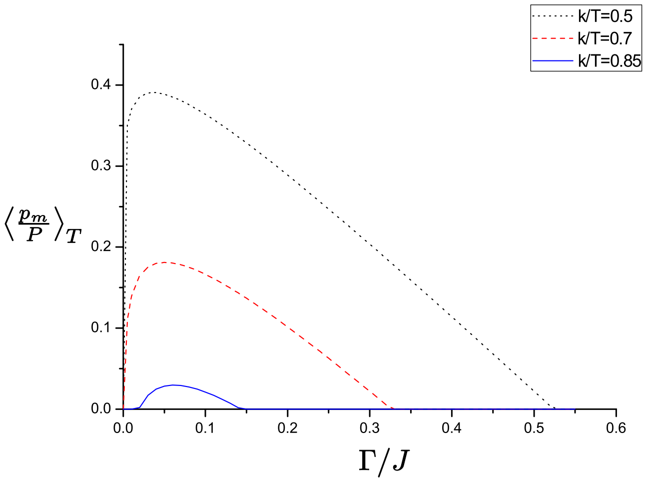

FIG. 4. (Color online) A graph showing the average most fit sequence population against mutation rate for a number of different environmental severities k/T. The fixed parameters are $j=1$, $T=20$, and $N=10^6$.

In future we would like to extend the results presented in this paper to more general fitness functions. A recent analysis of the dynamics of the ParaMuSe and Eigen models for arbitrary symmetric fitness functions [20] may make this tractable. We conjecture that the main findings of this paper—that the dynamic environment can be characterized by a single parameter k/T (average number of base changes in the most fit sequence per unit time), and that the optimal genomic mutation rate is equal to the environmental severity ($\Gamma N=k/T$)—are independent of the exact shape of the most fit sequence peak.

One can also consider the same dynamic environment in the Eigen quasispecies model. The dynamics of the Eigen model with the sharp-peak fitness function have been solved exactly in the large-N limit [18], and an analogous analysis to the one presented here is possible [27]. In Ref. [8] the Eigen model was studied with a changing environment, sharp-peak fitness function, and $k=1$. One interesting conclusion from that paper is that there exist sets of parameters (j,k,T) for which selection does not occur for any value of the mutation rate. We have a similar result here—the maximum and minimum mutation rates meet when $k/jT=1$, and so in environments with $k/jT>1$ selection is impossible.

The fitness landscape considered here is an extremely idealized one, and so it is difficult to make direct comparisons with data from real biological systems. However, one quantitative similarity between the optimal mutation rate found above and mutation rates observed in real organisms is worth highlighting. Drake *et al.* [22] found that the genomic mutation rates (number of mutations per genome per replication) are approximately constant and independent of genome length within broad groups of real organisms, but differ significantly between different groups (the groups studied were lytic RNA viruses, retroviruses, DNA based microbes, and higher eukaryotes). In Ref. [28], Nilsson and Snoad suggested a theoretical explanation of this result in terms of adaptation to a changing environment. In this paper we found that the optimal genomic mutation rate (ΓN) is independent of the genome length N and depends only on the environmental severity. If we assume that organisms within one of the groups listed above inhabit environments with similar severity and that their mutation rates are optimized for the environment, then it follows that their genomic mutation rates should be similar. Thus, our results provide more rigorous theoretical support for the hypothesis given in Ref. [28].

ACKNOWLEDGMENTS

The authors wish to thank David B. Saakian for useful discussions during the preparation of this paper. This research was supported by the Catholic University of Korea research fund 2009 and by a Korea Research Foundation (KRF) grant funded by the Korean Government (Basic Research Promotion Fund Grant No. KRF-2005-015-C00166).

APPENDIX A: PROBABILITY OF *mk* SPIN FLIPS BEING DISTINCT

In this appendix it is shown that for $m\sim N^{1/3}$ the probability of mk successive spin flips being distinct goes to 1 as N

$\to\infty$. First we prove the result for $k=1$. In this case the probability that the first N^{α} $(\alpha<1)$ spin flips are distinct is

$$p_1 = \frac{N!}{(N-N^{\alpha})!\,N^{(N^{\alpha})}}. \tag{A1}$$

For large N we use Stirling's approximation

$$N! = \sqrt{2\pi N}\left(\frac{N}{e}\right)^{N}[1+O(N^{-1})] \tag{A2}$$

leading to

$$p_1 = \frac{e^{-(N^{\alpha})}}{(1-N^{-1+\alpha})^{N-N^{\alpha}+1/2}}[1+O(N^{-\alpha})]. \tag{A3}$$

We approximate

$$\begin{aligned}
&(1-N^{-1+\alpha})^{N-N^{\alpha}+1/2}\\
&\quad=\exp[(N-N^{\alpha}+1/2)\ln(1-N^{-1+\alpha})]\\
&\quad=\exp\{(N-N^{\alpha}+1/2)[-N^{-1+\alpha}-N^{-2+2\alpha}/2+O(N^{-3+3\alpha})]\}\\
&\quad=\exp(-N^{\alpha}+N^{-1+2\alpha}/2)[1+O(N^{-1+\alpha})+O(N^{-2+3\alpha})].
\end{aligned} \tag{A4}$$

Substituting this into Eq. (A3) gives

$$p_1 = \exp(-N^{-1+2\alpha}/2)[1+O(N^{-\alpha})+O(N^{-1+\alpha})+O(N^{-2+3\alpha})]. \tag{A5}$$

We see from Eq. (A5) that, for any $\alpha<\frac{1}{2}$, the probability goes to 1 as $N\to\infty$.

For $k>1$, we need the probability that kN^{α} random spin flips are distinct given that flips $1,\ldots,k$ are distinct; flips $k+1,\ldots,2k$ are distinct; etc. Hence,

$$\begin{aligned}
p_k &= \frac{N!}{(N-kN^{\alpha})!\,N^{(kN^{\alpha})}}\left[\frac{(N-k)!\,N^{k}}{N!}\right]^{(N^{\alpha})}\\
&< \frac{N!}{(N-kN^{\alpha})!\,N^{(kN^{\alpha})}}\\
&= \exp(-N^{-1+2\alpha}k^{2}/2)[1+O(N^{-\alpha})+O(N^{-1+\alpha})+O(N^{-2+3\alpha})].
\end{aligned} \tag{A6}$$

As above, for $\alpha<\frac{1}{2}$ the probability goes to 1 as $N\to\infty$.

APPENDIX B: SADDLE-POINT TIME FOR $v_{i+1}-v_i>1$

In this appendix it is shown that the time t which maximizes the matrix element $\langle S_{v_{i+1}}|\exp[-H_{\text{diff}}t]|S_{v_i}\rangle$ is always less than $(v_{i+1}-v_i-1)T$. If the distance between the two sequences $S_{v_{i+1}}$ and S_{v_i} is d, then the saddle-point time is given by Eq. (25),

$$t = \frac{1}{\Gamma}\tanh^{-1}\left(\frac{2d/N}{\frac{J}{\Gamma}+\sqrt{\left(\frac{J}{\Gamma}\right)^{2}-\frac{4d(N-d)}{N^{2}}}}\right). \tag{B1}$$

From Eq. (B1) it can be shown that

$$\frac{\partial t}{\partial d} = \frac{1}{\Gamma N}\frac{1}{\sqrt{\left(\frac{J}{\Gamma}\right)^{2}-\frac{4d(N-d)}{N^{2}}}} \le \frac{1}{N}\frac{1}{\sqrt{J^{2}-\Gamma^{2}}}. \tag{B2}$$

Hence the saddle-point time t is an increasing function of d, and so the maximum saddle-point time occurs when d takes its maximum value, $k(v_{i+1}-v_i)$. Integrating over d in Eq. (B2) we obtain

$$t \le \frac{1}{N}\frac{k(v_{i+1}-v_i)}{\sqrt{J^{2}-\Gamma^{2}}}, \tag{B3}$$

and so

$$\frac{t}{(v_{i+1}-v_i-1)T} \le \frac{2}{NT}\frac{k}{\sqrt{J^{2}-\Gamma^{2}}} = \frac{2}{T\ln N}\frac{k}{\sqrt{j^{2}-\gamma^{2}}}. \tag{B4}$$

The right-hand side goes to zero as $N\to\infty$, so for large enough N we have the required result

$$t \le (v_{i+1}-v_i)T. \tag{B5}$$

---

[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] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[4] P. Schuster and J. Swetina, Bull. Math. Biol. **50**, 635 (1988).
[5] H. Wagner, E. Baake, and T. Gerisch, J. Stat. Phys. **92**, 1017 (1998).
[6] D. B. Saakian, E. Munoz, C.-K. Hu, and M. W. Deem, Phys. Rev. E **73**, 041913 (2006).
[7] Y. Brumer and E. I. Shakhnovich, Phys. Rev. E **71**, 031903 (2005).
[8] M. Nilsson and N. Snoad, Phys. Rev. Lett. **84**, 191 (2000).
[9] C. O. Wilke, C. Ronnewinkel, and T. Martinetz, Phys. Rep. **349**, 395 (2001).
[10] C. Kamp and S. Bornholdt, Phys. Rev. Lett. **88**, 068104 (2002).
[11] P. Gorodetsky and E. Tannenbaum, Phys. Rev. E **77**, 042901 (2008).
[12] B. Drossel, Adv. Phys. **50**, 209 (2001).
[13] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559 (1997).

[14] J.-M. Park and M. W. Deem, J. Stat. Phys. **125**, 971 (2006).
[15] M. Kimura and T. Maruyama, Genetics **54**, 1337 (1966).
[16] D. B. Saakian and C.-K. Hu, Proc. Natl. Acad. Sci. U.S.A. **103**, 4935 (2006).
[17] D. B. Saakian, J. Stat. Phys. **128**, 781 (2007).
[18] D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 021913 (2004).
[19] D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 046121 (2004).
[20] D. B. Saakian *et al.*, Phys. Rev. E **78**, 041908 (2008).
[21] M. Ancliff and J.-M. Park, J. Korean Phys. Soc. **53**, 2192 (2008).
[22] J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow, Genetics **148**, 1667 (1998).
[23] A. S. Kondrashov, Genet. Res. **66**, 53 (1995).
[24] P. L. Foster, BioEssays **22**, 1067 (2000).
[25] A. Traulsen *et al.*, J. Theor. Biol. **249**, 617 (2007).
[26] M. Nilsson and N. Snoad, Phys. Rev. E **65**, 031901 (2002).
[27] M. Ancliff and J.-M. Park (unpublished).
[28] M. Nilsson and N. Snoad, Bull. Math. Biol. **64**, 1033 (2002).