The Annals of Applied Probability
1995, Vol. 5, No. 4, 1025–1060

STEPPING-STONE MODELS WITH EXTINCTION AND RECOLONIZATION

By Hyun-Chung Kang, Stephen M. Krone[1]
and Claudia Neuhauser[2]

USDA Forest Service and University of Wisconsin, University of Wisconsin and University of Wisconsin

The stepping-stone model is widely used in population genetics to describe the evolution of a population with mating and geographical structure. It is typically formulated on a countable set, where each element of the set corresponds to a colony. Each colony consists of a population of a fixed number of haploid individuals. Individuals undergo random mating within each colony and migrate to neighboring colonies. In the model considered here, we are interested in the changes that may occur at a single locus on a single chromosome at which two alleles are possible. Changes in the gene frequencies are caused by migration between colonies, in addition to mutation, selection and random mating within each colony. It is frequently assumed that local populations persist indefinitely. This, however, is a reasonable assumption only if the average time until extinction of a local population is much larger than the time interval over which the whole population is studied. In this paper, using the framework of interacting particle systems, we address the question of how extinction and recolonization affect the spatial distribution of gene frequencies in both homogeneous and inhomogeneous environments. It turns out that extinction does not change the qualitative behavior of the model in a homogeneous environment; that is, in one and two dimensions, the population clusters in the absence of mutation (i.e., it consolidates into larger and larger blocks), whereas in higher dimensions, coexistence is possible. However, there are quantitative differences. For example, we find that extinction speeds up clustering in one dimension. In inhomogeneous environments, extinction can drastically change the behavior and new phenomena, such as coexistence in $d = 2$, appear.

1. Introduction. The stepping-stone model was introduced by Malécot (1948) and, independently, by Kimura and Weiss (1964) to describe the evolution of the genetic composition of a population with mating and geo-

---

Received November 1994; revised June 1995.

[1] Partially supported by NSF Grant 93-53149 (Presidential Faculty Fellowship held by R. Pemantle) and a grant from the USDA Forest Service, St. Paul, Minnesota.

[2] Alfred P. Sloan Research Fellow, partially supported by the NSF Grant DMS-94-03644 and a grant from the USDA Forest Service, St. Paul, Minnesota. The author also wishes to acknowledge support from the Institute for Mathematics and its Applications during the winter and spring quarters of 1994.

AMS 1991 *subject classifications*. 60K35, 92D10, 92D40.

Key words and phrases. Interacting particle systems, stepping-stone model, extinction, tence, periodic environment, random environment, multitype contact process, voter Wright–Fisher model, Moran model.

graphical structure. It is typically formulated on a countable set, where each element of the set corresponds to a colony. Each colony consists of a population of N haploid individuals. For each such individual, we are interested in a single locus on a single chromosome at which two alleles (or genes), A_1 and A_2, can occur. Genetical evolution is caused by migration between colonies, in addition to mutation, selection and random mating within each colony. The spatial component permits nontrivial behavior which depends on the spatial dimension.

Models for the evolution of the genetic composition of populations in subdivided habitats, such as Wright's (1931) island model, Levins' (1970) metapopulation model or the stepping-stone model, frequently assume that local populations persist indefinitely. This is only a reasonable assumption as long as the average time until extinction of a local population is much larger than the time interval over which the whole population is studied. Several studies [see, e.g., Simberloff and Wilson (1969) or Crowell (1973)] found that local extinctions may in fact be quite frequent. Since it is not obvious how extinction and recolonization may affect the population structure, it is important to incorporate these features into theoretical models, as was pointed out by McCauley (1991). In fact, there are contradicting statements in the biological literature on whether extinction enhances or diminishes genetic differentiation. We refer the reader to Neuhauser, Krone and Kang (1994) for further discussion of the biological issues, as well as references regarding biological and mathematical work which has been done in this area.

We only mention here some of the mathematical results on the stepping-stone model. Sawyer (1976, 1977, 1979) investigated the discrete-time stepping-stone model on $\mathbb{Z}^d$. He was particularly concerned with the rate at which clusters of individuals of common origin grow in the selectively neutral case with or without mutation. Clustering rates were also the focus of two articles by Cox and Griffeath (1987, 1990), who studied a simplified version of the continuous-time stepping-stone model (the so-called multitype voter model). Itatsu (1985, 1989) investigated the equilibrium measures of the discrete-time stepping-stone model with mutation and selection. There are, however, no rigorous results on the stepping-stone model with extinction and recolonization.

We now carefully define the model we wish to investigate. Let the d-dimensional integer lattice $\mathbb{Z}^d$ be the set of possible locations of colonies. Each existing colony has a population of N genes evolving according to a continuous-time Moran model. (The Moran model is the continuous-time analog of the usual Wright–Fisher model and will be defined below.) In addition, we allow for extinction of colonies and migration from neighboring colonies.

More explicitly, we consider a continuous-time Markov process (η_t, ζ_t), $t \geq 0$, where $\eta_t\colon \mathbb{Z}^d \to \{0, 1\}$ is the occupancy process and $\zeta_t\colon \mathbb{Z}^d \to \{1, 2\}^N \cup \Upsilon$ gives the allele compositions of the colonies. The colony size N is the same for all colonies. If $\eta_t(x) = 0$ (resp., 1), then we say that the colony at $x \in \mathbb{Z}^d$ is vacant (resp., occupied) at time t. If $\eta_t(x) = 1$, then $\zeta_t(x) =$

$(\zeta_t(x,1),\dots,\zeta_t(x,N)) \in \{1,2\}^N$ describes the alleles of the N individuals in colony x at time t, with $\zeta_t(x,k) = 1$ (resp., 2) if the kth individual in colony x has allele A_1 (resp., A_2) at time t; if $\eta_t(x) = 0$, then set $\zeta_t(x) = \Upsilon$. (Υ is just a symbol to define the state of a vacant site.) The A_1-allele frequency at time t in an occupied colony x is denoted $\rho_t(x)$; that is, $\rho_t(x) = (1/N)\Sigma_{k=1}^N(2 - \zeta_t(x,k))$. If $x = (x_1, x_2, \dots, x_d) \in \mathbb{Z}^d$, we set $\|x\|_1 = \Sigma_{i=1}^d|x_i|$. The dynamics of the process are as follows:

(i) For each pair (x,y) with $\|x-y\|_1 = 1$, if x is occupied at time t, then at rate 1 the colony at x sends an exact copy of its gene composition, that is, $\zeta_t(x)$, to y. If y is vacant when this happens, then y becomes occupied and $\zeta_t(y) = \zeta_{t-}(x)$. If y is already occupied, then for each k, $k = 1,2,\dots,N$, an independent coin is tossed with probability $\theta \in [0,1]$ of heads. If heads comes up, set $\zeta_t(y,k) = \zeta_{t-}(x,k)$; if tails comes up, $\zeta_t(y,k) = \zeta_{t-}(y,k)$.

(ii) If x is occupied, then it becomes vacant at rate $\gamma(x)$. The extinction rate $\gamma(x)$ may depend on location.

(iii) If x is occupied, then a reproduction event at that location occurs at rate $N\lambda$ (corresponding to an individual reproduction rate λ in a colony of size N). The relative fitnesses of A_1 and A_2 in colony x may depend on x in general, say $1 + s_1(x)$ and $1 + s_2(x)$, respectively, with $s_i(x) \in [0,\infty)$. If the A_1-allele frequency at x is $\rho(x)$, then, after taking selection into account, its weighted frequency is

$$\sigma(\rho(x)) = \frac{(1+s_1(x))\rho(x)}{(1+s_1(x))\rho(x) + (1+s_2(x))(1-\rho(x))}. \tag{1.1}$$

Reproduction is defined as follows: at the time t of a reproduction event, a single individual $k \in \{1,\dots,N\}$ is randomly chosen and a coin is tossed, with probability $\sigma(\rho_{t-}(x))$ of heads. If heads comes up, the value of $\zeta_t(x,k)$ is replaced by 1 at time t; if tails comes up, it is replaced by 2 at time t. All other individuals are unaffected.

Step (i) describes migration to neighboring colonies. If a colony is vacant, it can be recolonized from a neighboring "parent" colony and this new colony has the same gene composition as the parent colony. Already occupied colonies can receive genes from neighboring colonies. The parameter θ is a measure of how successfully the invading population can replace members of the original population. Note that the parent colony at x remains unchanged. One can think of the migration here as a cloning mechanism. Step (ii) describes the extinction mechanism. Typically, we will have $\gamma(x) = \delta$ or Δ, two constants, with $0 \le \delta \le \Delta$. Colonies will be called *δ-colonies* of *Δ-colonies*, depending on this extinction rate. We will treat several different cases. In the case of a homogeneous environment, we will set $\gamma(x) = \Delta$ for all x. We will treat the case of a periodic environment in $\mathbb{Z}$ by placing δ-colonies D units apart and filling in the remaining lattice sites with Δ-colonies. The case of a random environment in $\mathbb{Z}^d$ will be obtained by independently assigning a value δ or Δ to each $\gamma(x)$, with probability p for δ. These values are assigned at the beginning and then held fixed forever. Step (iii) describes the selection and reproduction mechanism of a continuous-time Moran model. Reproduction here is simply replacement: a single individual chosen at the time of reproduction is replaced by another individual whose type is determined according to the weighted frequency distribution at that site. The replaced individual can be thought of as being killed. In this way, the population size of a colony is always N. (In discrete-time models, this could be replaced by the Wright–Fisher model in which all N individuals are replaced at the same time via binomial sampling.) In the case of an inhomogeneous environment, if $\gamma(x) = \delta$, then we set $s_1(x) = s_1$ and $s_2(x) = 0$; if $\gamma(x) = \Delta$, we set $s_1(x) = 0$ and $s_2(x) = s_2$, where s_1 and s_2 are nonnegative constants. In other words, A_1 has a selective advantage (or, at least no disadvantage) on δ-colonies and A_2 has a selective advantage on Δ-colonies. Note that when $\rho_t(x) = 1$ (resp., 0), fixation of allele A_1 (resp., A_2) will have occurred. Since there is no mutation in our model, this fixation can only be interrupted by gene flow (migration) from neighboring sites. We call the above process a *stepping-stone model with extinction and recolonization.*

Note that (i) and (ii) imply that η_t is a basic contact process (in a homogeneous, periodic or random environment). Later, we will need the notion of the *critical death rate* for a basic contact process in a homogeneous environment. Consider a contact process in which particles die at rate δ and give birth on neighboring vacant sites at rate 1. Since this process is attractive [cf. Liggett (1985)], starting from a configuration of all 1's, there is a unique stationary measure ν_1. The all 0 configuration is a trap. It is known that there is a unique number, $\delta_c > 0$, depending on the dimension, for which ν_1 is the point mass on the configuration of all 0's when $\delta \geq \delta_c$, and is a nontrivial invariant measure when $\delta < \delta_c$. It is clear that, to get any interesting behavior in our model, we will need at least one of the two extinction rates Δ and δ to be less than δ_c.

Let $\mathscr{P}(\mathscr{X})$ (resp., $\mathscr{S}$) denote the set of probability measures (resp., translation-invariant probability measures) on the space of configurations,

$$\mathscr{X} \equiv \left(\{1,2\}^N \cup \Upsilon\right)^{\mathbb{Z}^d},$$

and let μ_1 (resp., μ_2) $\in \mathscr{P}(\mathscr{X})$ be the limit law corresponding to initial distribution δ_1 (resp., δ_2) $\in \mathscr{P}(\mathscr{X})$, which puts 1's (resp., 2's) at each $(x, k) \in \mathbb{Z}^d \times \{1, \dots, N\}$. These invariant measures exist because the particle system starting with all 1's (resp., all 2's) is attractive. Note that, due to the absence of mutation, once the process is void of 1's or 2's, it will remain so. In addition, let $\pi_\beta \in \mathscr{P}(\mathscr{X})$ denote the product measure which assigns values $\{1, 2\}$ in an i.i.d. manner to the coordinates (x, k), with probability β for 1 and $1 - \beta$ for 2. We write ζ_t^μ for the process ζ_t with initial distribution $\mu \in \mathscr{P}(\mathscr{X})$; when $\mu = \pi_\beta$, we simply write ζ_t^β. If $\mu, \nu \in \mathscr{P}(\mathscr{X})$, then we use the notation $\zeta_t^\mu \Rightarrow \nu$ to denote the weak convergence $\mu S_t \Rightarrow \nu$, where μS_t is the law of ζ_t^μ.

The first theorem considers the case of a homogeneous environment in which one of the alleles has a selective advantage. Not surprisingly, the allele with the selective advantage takes over.

THEOREM 1. *Suppose* $d \geq 1$ *and* $\gamma(x) = \Delta$, $0 \leq \Delta < \delta_c$, *for all* $x \in \mathbb{Z}^d$. *Further assume that, for each* x, $s_1(x) \equiv s_1$ *and* $s_2(x) \equiv s_2$, *with* $s_1 < s_2$. *Then for any* $\theta \in (0,1]$, $\lambda > 0$ *and* $N > 0$, *the "1's die out." That is, if* ζ_0 *is translation invariant and* $P(\zeta_0(x,k) = 2) > 0$, *then* $\zeta_t \Rightarrow \mu_2$.

The key to proving Theorem 1 is duality. We will give more details after the next theorem. Our next result deals with the case of selectivity neutral alleles in a homogeneous environment.

THEOREM 2. *Suppose* $\gamma(x) = \Delta$, $0 \leq \Delta < \delta_c$, *and* $s_1(x) = s_2(x)$, *for all* $x \in \mathbb{Z}^d$.

(a) *If* $d \leq 2$, *then clustering occurs: for any* $\theta \in (0,1]$, $\lambda > 0$, $N > 0$ *and any initial configuration* ζ_0,

$$P(\zeta_t(x,j) = 1, \zeta_t(y,k) = 2) \to 0 \quad \text{as } t \to \infty,$$

holds for all $(x,j) \neq (y,k) \in \mathbb{Z}^d \times \{1,\dots,N\}$. *Furthermore, if* $\mu \in \mathcal{S}$, *then there is a constant* $q \in [0,1]$ *such that*

$$\zeta_t^{\mu} \Rightarrow q\mu_1 + (1-q)\mu_2$$

as $t \to \infty$.

(b) *If* $d \geq 3$, *then coexistence is possible: for any* $\theta \in (0,1]$, $\lambda > 0$, $N > 0$ *and any* $\beta \in (0,1)$, *there is a unique translation-invariant stationary distribution* ν_β, *such that*

$$\zeta_t^{\beta} \Rightarrow \nu_\beta$$

as $t \to \infty$; ν_β *gives zero probability to the set of configurations with no* 1*'s or no* 2*'s.*

REMARKS. (a) As in Neuhauser (1992), page 502, one can show that the set of extreme points of the set of translation-invariant stationary measures is $\{\nu_\beta \colon \beta \in [0,1]\}$. Moreover, starting from any translation-invariant distribution, the process converges weakly to a mixture of the ν_β's.

(b) Comparing this theorem with Sawyer (1976), we see that the dichotomy between clustering and coexistence is the same, with and without extinction. Without extinction, the model closely resembles the voter model [see, e.g., Liggett (1985) or Durrett (1988)]; with extinction, it can be viewed as a mixture of the voter model and the multitype contact process [see Neuhauser (1992) for results on the multitype contact process]. In both models, this dichotomy between clustering in $d \leq 2$ and coexistence in $d \geq 3$ can be observed. Duality (which will be explained in the next section) is the key to the proof. Roughly speaking, a dual process traces the ancestral line of a particle; here, the duals behave similarly to coalescing random walks. Two individuals are of the same type if their duals coalesce and may be of different types if their duals do not coalesce. The dichotomy can thus be explained by the recurrence of random walks in $d \leq 2$ and transience in $d \geq 3$. The proofs of both theorems are similar to what was done in Neuhauser

(1992) for the multitype contact process. Therefore, we will only sketch the proofs and point out the differences.

Even though the model with extinction does not exhibit qualitatively different behavior compared to the one without extinction, the next result will show that extinction causes a change in the rate of clustering. We show that extinction speeds up clustering when $d = 1$. To do this, we set

$$\begin{aligned} &g_t(x, y; \Delta) \\ (1.2)\quad &= \max_{1 \le j,\, k \le N} P_\Delta\big(\zeta_t(\lfloor x\sqrt{t} \rfloor, j) = 1, \\ &\qquad \zeta_t(\lfloor y\sqrt{t} \rfloor, k) = 2 \,\big|\, \zeta_t(\lfloor x\sqrt{t} \rfloor, j) > 0,\, \zeta_t(\lfloor y\sqrt{t} \rfloor, k) > 0\big) \end{aligned}$$

for $x, y \in \mathbb{Z}$ and $\Delta \ge 0$, where the subscript Δ refers to the extinction probability in each colony and $\lfloor \cdot \rfloor$ denotes the integer part. Note that when $\Delta = 0$ and the process starts with all sites occupied, this simplifies to

$$(1.3)\quad g_t(x, y; 0) = \max_{1 \le j,\, k \le N} P_0\big(\zeta_t(\lfloor x\sqrt{t} \rfloor, j) = 1,\, \zeta_t(\lfloor y\sqrt{t} \rfloor, k) = 2\big).$$

In the definition of $g_t(x, y; \Delta)$, we condition on the survival set because, without conditioning, the probability would be made smaller simply because some of the sites are not occupied. Our interest is in showing that the occupied sites are less likely to be different by a certain time due to extinction.

THEOREM 3. *Assume the neutral setting of Theorem* 2, *with* $d = 1$ *and initial distribution* π_β, $\beta \in (0, 1)$. *Then, for any* $N > 0$, $\lambda > 0$, $\Delta \in (0, \delta_c)$ *and pair* $x \ne y$ *in* $\mathbb{Z}$, *there exists* $\theta_0 > 0$ *such that* $\theta \in (0, \theta_0)$ *implies*

$$g_t(x, y; 0) > g_t(x, y; \Delta)$$

for all sufficiently large t.

To prove this result, we will again use duality. This time, we will utilize the fact that increasing the death rate Δ makes it easier for colonies to invade neighboring sites since the density of occupied sites is decreasing in Δ and it is much easier to colonize a vacant site than an occupied site for small θ. (Recall that θ is the probability that an allele can successfully invade a neighboring site that is already occupied.) The scaling in (1.2) is just the natural Brownian motion scaling used to approximate the one-dimensional random walk that arises in the dual process. It seems that Theorem 3 should also be true for the case in which the initial distribution is an ergodic translation-invariant measure with positive densities of 1's and 2's, but we will not pursue this.

The main point of this work is to illustrate some of the differences between stepping-stone models with and without extinction. In a homogeneous environment, the differences appear only on a quantitative level; qualitatively, the systems behave quite similarly. In an inhomogeneous environment, the behavior can change drastically as illustrated in the following theorems. We begin by making a change in the *extinction rule*.

H1. On Δ-sites, all colonies die at rate Δ as before. However, on δ-sites, colonies with $\rho(x) > 0$ become extinct at rate $\delta \leq \Delta$ and colonies with $\rho(x) = 0$ become extinct at rate Δ.

In other words, 2's cannot use a δ-site as a refuge unless there are 1's present.

The first of these theorems concerns the behavior of the above stepping-stone model in a one-dimensional *periodic environment* with δ-colonies at $x = 0, \pm D, \pm 2D, \dots$, where D is some positive integer, and Δ-colonies at the remaining lattice points. The A_1 alleles will have a strong selective advantage on δ-colonies and the A_2 alleles will have a strong selective advantage on Δ-colonies. That is, we will assume the following.

H2. On δ-colonies, $s_1(x) = s_1$ and $s_2(x) = 0$; on Δ-colonies, $s_1(x) = 0$ and $s_2(x) = s_2$.

The values $s_2(x) = 0$ on δ-colonies and $s_1(x) = 0$ on Δ-colonies are not important; the key is that we have $s_1(x) \gg s_2(x)$ on δ-colonies, and the reverse on Δ-colonies.

THEOREM 4. *Consider the one-dimensional stepping-stone model on the periodic environment defined above. We assume* H1 *with* $\delta = 0$ *and* H2.

(i) *Let* $\Delta = 0$ *and fix* $N \geq 1$. *Assume initially that there are infinitely many* Δ-*colonies with* $\rho(x) < 1$. *Then, for any* $D \geq 3$, $s_1 \geq 0$ *and* $\theta \in (0,1)$, *the* A_1 *alleles die out with probability* 1 *for sufficiently large* λ *and* s_2.

(ii) *Let* $\Delta > 0$, $D > 0$ *and* $\theta \in (0,1)$ *be fixed. Assume initially that there are infinitely many* δ-*colonies with* $\rho(x) > 0$. *There exists* $N_0 = N_0(\theta, D, \Delta)$ *so that, for* $N \geq N_0$ *fixed and for any* $s_2 \geq 0$, *the* A_2 *alleles die out with probability* 1 *for sufficiently large* λ *and* s_1.

The statement "allele A_i dies out" means that any finite interval will contain A_i alleles for only a finite amount of time. The first part of the theorem is not surprising. Here, there is no extinction on any site, and the heterogeneous environment gives strong selective advantage to different alleles on different sites. The A_2 alleles, however, have a "spatial advantage" since there are more sites where A_2 alleles are favored over A_1 alleles. The second part says that this "spatial advantage" can be outweighed by a "survival disadvantage" due to extinction.

Our last theorem states that, in $\mathbb{Z}^2$, one can obtain coexistence if the environment is random. Let p_c^{site} denote the critical probability for ordinary two-dimensional site percolation [cf. Durrett (1988)]. The *random environment* is obtained by choosing $\gamma(x)$ independently at each location with $P(\gamma(x) = \delta) = p$ and $P(\gamma(x) = \Delta) = 1 - p$, where $1 - p > p_c^{\text{site}}$. The environment, denoted by e, is chosen at the beginning and held fixed for the life of the process. The law of the process, given the environment e, is denoted by P_e. By *coexistence* here we mean that, for any $\eta > 0$ and suitably large L (depending on η), any $L \times L$ square will have both types of alleles with P_e-probability at least $1 - \eta$ at all sufficiently large times. This implies that there exists a nontrivial stationary distribution with positive proportions of each type.

THEOREM 5. *Suppose $d = 2$ and $\theta \in (0,1)$, and let the random environment be defined as above. Assume the extinction rates and selection coefficients are as in* H1, *with $\delta = 0$, and* H2. *Then, if N is fixed and sufficiently large, we have the following. For almost every environment, there exists $\Delta_1 \in (0, \delta_c)$ such that, if $\Delta \in (0, \Delta_1)$ and λ, s_1 and s_2 are sufficiently large, we get coexistence starting from any initial configuration with infinitely many* 1's *and* 2's.

REMARK. The condition $1 - p > p_c^{\text{site}}$ implies that there is an infinite connected cluster of Δ-colonies and the δ-colonies occur only in finite clusters. The random environment is chosen at the beginning and held fixed for the life of the process. Also, Δ_1 is much smaller than the critical value for the two-dimensional contact process since we need to ensure the survival of one-dimensional embedded contact processes in some steps of the proof.

2. Construction and dual process. Proofs of Theorems 1 and 2. We will now use the standard tool of *graphical representation* to construct our process in terms of a collection of independent Poisson processes. This will also allow us to construct a dual process which will be useful in some of the proofs. We will give the construction for the homogeneous case in which $s_2 \geq s_1$ only. (It will be obvious how to do it for $s_2 \leq s_1$ or for the periodic or random environment cases.) We can think of our particle system as being constructed on $\mathbb{Z}^d \times \{1, \dots, N\}$. A typical element $(x, k) \in \mathbb{Z}^d \times \{1, \dots, N\}$ represents the kth position in colony x. Write $\|z\|_1 = |z_1| + \cdots + |z_d|$ for $z = (z_1, \dots, z_d) \in \mathbb{Z}^d$. For $x, y \in \mathbb{Z}^d$ with $\|x - y\|_1 = 1$ and $k \in \{1, \dots, N\}$, let $\{U_n^{(x,y)}: n \geq 1\}$, $\{V_n^x: n \geq 1\}$ and $\{W_n^{(x,k)}: n \geq 1\}$ be the successive arrival times of independent Poisson processes with rates $1/(2d)$, $\gamma(x)$ and λ, respectively. These correspond to times of migration, extinction and reproduction events. Let $\{\Theta_n^{(x,y,k)}: x, y \in \mathbb{Z}^d, \|x - y\|_1 = 1, k \in \{1, \dots, N\}, n \geq 1\}$ be i.i.d. $\{0,1\}$-valued random variables with $P(\Theta_n^{(x,y,k)} = 1) = \theta$, let $\{\Phi_{n,i}^{(x,k)}: x \in \mathbb{Z}^d, k \in \{1, \dots, N\}, n \geq 1, i \geq 1\}$ be an independent set of i.i.d. $\{1, \dots, N\}$-valued random variables with $P(\Phi_{n,i}^{(x,k)} = j) = 1/N$ for each $j \in \{1, \dots, N\}$ and let $\{\Lambda_{n,i}^{(x,k)}: (x,k) \in \mathbb{Z}^d \times \{1, \dots, N\}, n \geq 1, i \geq 1\}$ be an independent set of i.i.d. $\{0,2\}$-valued random variables with $P(\Lambda_{n,i}^{(x,k)} = 2) = (s_2(x) - s_1(x))/(1 + s_2(x))$. At migration times $U_n^{(x,y)}$, we draw an arrow from (x,k) to (y,k), for each $k \in \{1, \dots, N\}$. This indicates that, if (x,k) is occupied, then (y,k) will become occupied with the same value if either $\eta_{U_n^{(x,y)}}(y) = 0$ or $\Theta_n^{(x,y,k)} = 1$. We refer to these arrows as *migration arrows*. If $\Theta_n^{(x,y,k)} = 1$ (resp., 0), we label the arrow with a "θ" (resp., "$1 - \theta$") and call the arrow a "θ-arrow" [resp., "$(1-\theta)$-arrow"]. At extinction times V_n^x we put a † at (x,k) for each $k \in \{1,\dots,N\}$. The effect of a † is to kill the "particle" at each (x,k) in colony x, if it is occupied. At reproduction times $W_n^{(x,k)}$ we choose a sequence of members $l_i \in \{1,\dots,N\}$ of colony x at random according to $\Phi_{n,i}^{(x,k)}$, for $i = 1,2,\dots,$ and draw for each i an arrow from (x,l_i) to (x,k). We call these arrows *replacement arrows.* If $\Lambda_{n,i}^{(x,k)} = 2$, we put a "2" on the arrow to indicate that only 2's can pass through. If the particle type at (x,l_1) is either 1 or 2 and the arrow from (x,l_1) to (x,k) does not have a 2 on it, or if the arrow from (x,l_1) to (x,k) does have a 2 on it and the individual at (x,l_1) is a 2, then the individual at (x,k) is replaced by one of the same type as the individual at (x,l_1) and we say that a successful reproduction event has taken place (and no further l_i's are needed). If the arrow from (x,l_1) to (x,k) has a 2 on it and the individual at (x,l_1) is a 1, the replacement is suppressed and we try the same procedure with $i = 2$. We continue until the first time a successful replacement occurs. Note that if $s_1 = s_2$, the first trial will always be successful. The effect of this is to replace the particle type at (x,k) with that at (x,l_I), where l_I is the first member of the colony to reproduce successfully. [Note that we allow for an arrow from (x,k) to (x,k). In the Moran reproduction scheme, this covers the case where the individual at (x,k) gives birth and then is killed.] If $s_2 < \infty$, the number of trials until the first successful mating is geometrically distributed and hence, I will be finite almost surely. (To include the case where $s_2 = \infty$, we make the convention that if a colony consists of alleles which are all of the same type, then the reproduction is suppressed.) It is easy to see that this representation describes the reproduction scheme defined in the Introduction: let ρ (resp., ρ') be the A_1-allele frequency before (resp., after) reproduction, and let $a = (s_2 - s_1)/(1 + s_2)$ be the probability that an arrow is labeled with a 2. Then a simple calculation shows

$$\rho' = \sum_{k=0}^{\infty} (1-a)\rho(a\rho)^k = \frac{(1-a)\rho}{1-a\rho} = \frac{(1+s_1)\rho}{1+s_2-(s_2-s_1)\rho}.$$

Comparing this with (1.1) shows that $\rho' = \sigma(\rho)$.

With these basic ingredients, it is now standard fare [cf. Harris (1972)] to construct the process described above, starting from any initial configuration (ζ_0, η_0), using the *percolation diagram* resulting from the graphical representation.

To determine the state of some site (x,j) at time t, it is convenient to work backwards in time using the graphical representation. This backward process is known as the *dual process.* To construct it, consider first the case in which $s_1(x) = s_2(x)$. For each fixed $t > 0$, define $\tilde{\zeta}_t$ by reversing the arrows in the graphical representation of ζ and reversing time through the map $\tilde{s} = t - s$. Say there is a *path* from $(x,j;\tilde{0})$ to $(y,k;\tilde{t})$ if there is a sequence of times $\tilde{0} = \tilde{s}_0 < \tilde{s}_1 < \tilde{s}_2 < \cdots < \tilde{s}_n < \tilde{s}_{n+1} = \tilde{t}$ and spatial locations $(x,j) = (x_0,k_0),(x_1,k_1),\dots,(x_n,k_n) = (y,k)$ such that (i) for $i = 1,\dots,n$, there is an arrow from (x_{i-1},k_{i-1}) to (x_i,k_i) at time $\tilde{s}_i$, and (ii) the vertical segments

$\{(x_i, k_i)\} \times (\tilde{s}_i, \tilde{s}_{i+1})$, $i = 0, 1, \dots, n$, do not contain any †'s. More specifically, the dual process is a set-valued process whose state at time t, starting at the singleton $\{(x, j)\}$, is given by

$$\tilde{\zeta}_t^{(x,j)} = \big\{(y, k): \text{there is a path from } \big(x, j; \tilde{0}\big) \text{ to } \big(y, k; \tilde{t}\big)\big\}. \tag{2.1}$$

Thus, $\tilde{\zeta}_t^{(x,j)}$ is the (random) set of sites at time 0 in $\mathbb{Z}^d \times \{1, \dots, N\}$ that are capable of determining the state of site (x, j) at time t. We refer to the sites in $\tilde{\zeta}_t^{(x,j)}$ as *ancestors* of (x, j). To see how the dual determines the type of a given site at time t, we require the notion of an *ancestral hierarchy*. If we imagine for a moment that each site at time 0 is occupied by a particle of a unique "color," then one of these sites will paint (x, j) its color, if it is occupied at time t. We denote this site by $\tilde{\zeta}_i^{(x,j)}(1)$ and call it the *first ancestor* or *distinguished particle*. It is the first member of the hierarchy. If we make $\tilde{\zeta}_i^{(x,j)}(1)$ empty, then the color will change to that of some other site $\tilde{\zeta}_t^{(x,j)}(2)$, the second member of the hierarchy (or *second ancestor*) and so on. (We will explain below how to obtain the ancestral hierarchy from the graphical representation.) Now we go back to our process in which there are only two "colors." To see which of the ancestors (if any) gives (x, j) its type, it is necessary to trace the paths from $(x, j; t)$ backward to the various ancestors at time 0. We want to know whether $(x, j; t)$ is occupied and, if so, with what type of particle. Follow the path from $(x, j; \tilde{0})$ to $(\tilde{\zeta}_t^{(x,j)}(1); \tilde{t})$ in the reversed percolation diagram; there is exactly one such path. If the initial configuration, ζ_0, has a 1 (resp., 2) at that site, then we know that $(x, j; t)$ is occupied by a 1 (resp., 2), as well. If the initial configuration has no particle at that site, we try the second ancestor and so on. Since, for any fixed time t, there are only a finite number of ancestors, we will be done after a finite number of steps. Either there is an ancestor that can successfully paint $(x, j; t)$ with its color or there is not. In the latter case, $(x, j; t)$ is empty.

We remark that the above construction with the reversed percolation diagram could be done just as easily by using the original percolation diagram and traversing the arrows in the reverse direction. In this case, we would not allow backward paths to go through vertical segments with †'s. This is how we prefer to look at $\tilde{\zeta}_t$ below. It also saves us from having to draw a separate picture for the dual.

Now let us be more specific about how to read off the ordered set of ancestors from the percolation diagram. To determine the ancestors of $(x, j; t)$ at time 0, begin at $(x, j; t)$ and travel backward on the percolation diagram until time 0 or until the first time a death occurs in the path, due to either a † or the tip of a θ-arrow or replacement arrow touching the time line at (x, j). In case of a †, we reverse our direction and go forward in time (i.e., retrace our steps) either until we reach the starting point $(x, j: t)$ or until the first time a tip of an arrow touches our path. In the first case, $(x, j; t)$ does not have any ancestors at time 0 and thus the set of ancestors is empty. In the second case, the arrow is necessarily a $(1 - \theta)$-arrow. We follow this arrow in the reverse direction and then continue down from the new location using the same algorithm. In case of a θ-arrow, follow the arrow in the reverse direction and continue down from the new location. In case of a replacement arrow, follow the arrow from (x, l_I) to (x, j) in the reverse direction and continue down from the new location. Following this recipe, we either land at time 0 at some location, which is then the location of the first ancestor, or the set of ancestors is empty. To find the second ancestor, erase the part of the path in the diagram that ends at $\tilde{\zeta}_t^{(x,j)}(1)$ and is not shared by any of the other ancestors. Using the same procedure as above, we can find the second ancestor. We continue in this way until there are no more paths between $(x, j; t)$ and time 0. The resulting ordered set is then the hierarchy of ancestors.

When $s_2(x) > s_1(x)$, the set of ancestors is the same. This time, however, the paths may contain arrows labeled with 2's. To determine whether (x, j) is occupied at time t, follow the path from $(x, j; t)$ to $(\tilde{\zeta}_t^{(x,j)}(1); 0)$ as before. If the initial configuration has a 2 at that site, then we know that $(x, j; t)$ is occupied by a 2. If it has a 1 at that site, then $(x, j; t)$ is occupied by a 1 only if there are no 2-arrows in the path. If there are 2-arrows in the path and the first ancestor lands on a 1, then none of the other ancestors which grow (down) out of the tail of the lowest such 2-arrow (i.e., the one that the first ancestor runs into first when going in the direction of increasing t) can succeed. They are blocked by the 1 coming from the first ancestor. So "remove" these paths from the diagram and check the next remaining ancestor, if there is one, as above. Finally, if the first ancestor lands on a vacant site, move to the second ancestor and so on. Proceeding in this way, we will either find an ancestor that succeeds in giving $(x, j; t)$ its type or we will run out of ancestors and $(x, j; t)$ will be vacant.

The above dual, over the interval $[\tilde{0}, \tilde{t}]$, is convenient for defining the ancestral hierarchy. As a function of t, however, it is rather badly behaved. So, as is usually done in particle systems, we will work with a slightly different dual which has the same one-dimensional distributions, but nicer sample paths. This dual, denoted by $\hat{\zeta}_t$, is constructed for all times directly from its own (dual) percolation diagram. More specifically, we construct $\hat{\zeta}_\cdot$ from a graphical representation that has, for each x, a set of arrows from (x, k) to (y, k), $k \in \{1, \dots, N\}$, at rate 1 when $\|x - y\|_1 = 1$, each such arrow being independently assigned a label θ or $1 - \theta$ with probabilities θ and $1 - \theta$, respectively; replacement arrows between a randomly selected pair of sites (possibly the same) in colony x at rate $N\lambda$; and †'s at all sites in colony x at rate $\gamma(x)$. The construction of this dual process from the dual percolation diagram is as in the construction of ζ_t, with paths blocked by †'s. In particular,

$$(2.2)\qquad \hat{\zeta}_t^{(x,j)} = \bigl\{(y, k): \text{there is a dual path from } \bigl(x, j; \hat{0}\bigr) \text{ to } \bigl(y, k; \hat{t}\bigr)\bigr\}.$$

Note that, even though this dual process is constructed "forward in time," we want to think of its time scale as running in the opposite direction as that of ζ_t, so we label space–time points in the dual percolation diagram with carets on the times. It is easy to see that $\hat{\zeta}_t$ has the same distribution as $\tilde{\zeta}_t$, for each fixed t. For example, in the neutral case, we have $P(\zeta_t(x,j) = i) = P^*(\zeta_0(\hat{\zeta}_t^{(x,j)}(m)) = i)$, $i = 1, 2$, where $\hat{\zeta}_t^{(x,j)}(m)$ is the first ancestor to land on an occupied site, if there is one, and $P^* = P \times \hat{P}$ is the joint law of ζ and $\hat{\zeta}$ on a suitably enlarged probability space. In what follows, we will not distinguish between the various probability measures. We simply write P for a generic probability measure; the meaning will be clear from the context.

To prove Theorems 1 and 2, we will need more information on the structure of the dual $\hat{\zeta}_t$. For the classical voter model on $\mathbb{Z}^d$, the dual process, starting at a single site, consists of a single particle (ancestor) performing a random walk in $\mathbb{Z}^d$. Two such duals (corresponding to different starting points) move independently until they collide, at which point they coalesce into a single particle. For our stepping-stone model, the dual is much more complicated. Starting at a single site, it typically consists of many particles (the hierarchy of ancestors) and thus the sample paths have a tree structure (cf. Figure 1). The key observation is that embedded in the path of the first ancestor is a sequence of "renewal points" whose space–time coordinates perform a random walk. We have enough control over the behavior of the dual process between such renewal points to show that the distinguished particles corresponding to two distinct sites may eventually coalesce, similar to the voter model.

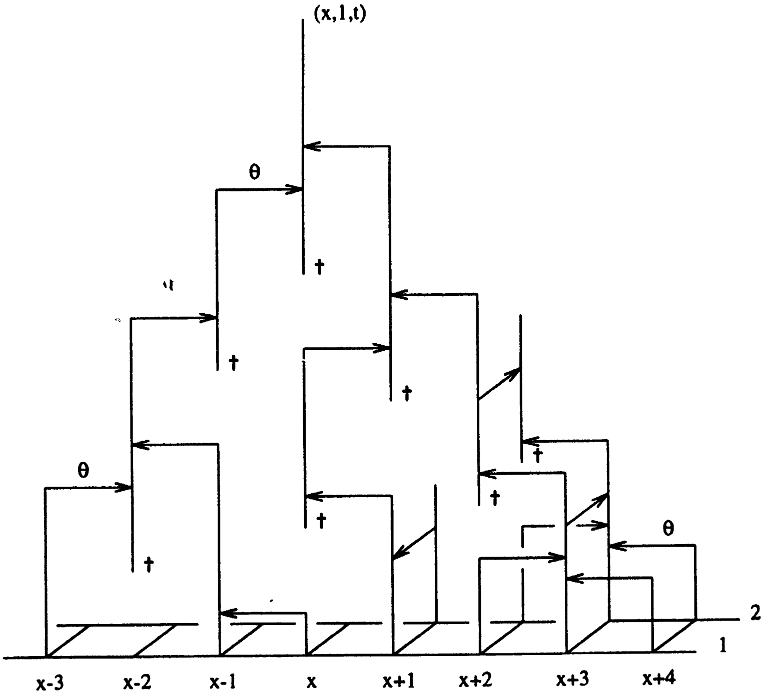

FIG. 1.

For any starting point $(x, j; \hat{0})$ in the dual percolation diagram, either the tree growing out of this point will die in finite time (due to †'s) or it will survive forever. Since we are dealing with the supercritical case, $\Delta < \delta_c$, after a finite (geometric) number of trials we can find a starting point that has an infinite tree growing out of it. These are the only starting points that will affect the asymptotic behavior of the dual and hence the process. Let $\Omega_\infty = \Omega_\infty^{(x,j;\hat{0})}$ be the event that the dual process starting at $(x, j; \hat{0})$ lives forever. In the translation-invariant case, the probability of this event does not depend on $(x, j; \hat{0})$.

Let us consider the tree growing out of some initial point $(x, j; \hat{0})$ in the dual percolation diagram. Suppose for the moment that we can see the whole tree and that we color all infinite paths in the tree with red; there will be at least one such path if we are viewing a sample from Ω_∞. Growing out of the red paths will typically be many "dead ends," branches of the tree that die due to †'s. Now follow the path of the first ancestor, $\hat{\zeta}_t^{(x,j)}(1)$. It will spend some time in "dead end" branches and, in such a case, it goes all the way to the end of this branch before jumping to a new branch according to the algorithm described above. Eventually, the first ancestor will jump back to a red path and move along this path for a while before moving off to other dead end branches and so on. We say that the dual process is *good* while the first ancestor is on this red path. The random space–time points at which the first ancestor rejoins a red path (and hence is back on an infinite branch of the tree) are called *renewal points*. Conditioned on Ω_∞, these renewal points break up the path of the distinguished particle into i.i.d. pieces. When jumping within a given colony (due to replacement arrows), the distinguished particle simply selects a new level according to a uniform distribution on $\{1, \dots, N\}$. To describe jumping between colonies, write $X_i \in \mathbb{Z}^d$ for the spatial displacement between the x-coordinates (colonies) of the ith and $(i-1)$th renewal points, and τ_i for the corresponding temporal displacement. [The 0th renewal point is $(x, j; \hat{0})$.] Then

$$S_n = x + \sum_{i=1}^{n} X_i \quad \text{and} \quad T_n = \sum_{i=1}^{n} \tau_i$$

will be the spatial (colony) and temporal coordinates of the renewal point after its nth jump. The next result says that this defines a random walk, embedded in the "colony path" of the distinguished particle, that jumps to site S_n at the random time T_n.

PROPOSITION 2.1 [Neuhauser (1992)]. *Conditioned on the event* Ω_∞, $\{(X_i, \tau_i)\}_{i \geq 1}$ *form an i.i.d. family of random vectors in* $\mathbb{Z}^d \times \mathbb{R}_+$. *Furthermore, the tail distributions of* X_i *and* τ_i *have exponential bounds; there are constants* $C, \gamma \in (0, \infty)$ *such that*

$$P(\|X_i\|_1 > t) \leq Ce^{-\gamma t} \quad \text{and} \quad P(\tau_i > t) \leq Ce^{-\gamma t}$$

for all $i \geq 1$.

PROOF. Proposition 2.1 appears in Neuhauser (1992) in the case of a multitype contact process. The proof given there is based on an analysis of the amount of time it takes a dead end to die out due to †'s. In our process we have θ-arrows and replacement arrows in addition to the "contact-type" recolonization arrows. These new arrows affect the path that the first ancestor follows, but do not affect the lengths of the dead ends. For example, the θ's are assigned to arrows after they are put in the tree. In other words, the distribution of the (X_i, τ_i) does not depend on θ. With this in mind, it is easy to see that the proof of Proposition 1 in Neuhauser (1992) goes through in this setting. We omit the details. □

We are now ready to prove Theorems 1 and 2.

PROOF OF THEOREM 1. We assume that $s_1 < s_2$ and that ζ_0 has a translation-invariant distribution. Since weak convergence in this setting is just convergence of finite-dimensional distributions, it clearly suffices to show

$$P(\zeta_t(x,j) = 1) \to 0 \tag{2.3}$$

as $t \to \infty$.

This was basically done in Neuhauser (1992), so we just sketch the idea. We select a subsequence of ancestors that are candidates for painting (x,j) with the color 2. The first member of the subsequence is $\hat{\zeta}_t^{(x,j)}(1)$. Follow the path this ancestor takes (going up the dual percolation diagram) in order to paint (x,j) its color. If it encounters any 2-arrows on this path, discard all but the first ancestor that are "offspring" from the tip of the last 2-arrow encountered when going up the percolation diagram for the dual process. (These will be the next few members of the ancestral hierarchy.) The first ancestor that is left after discarding these ancestors is the second member of the subsequence. Repeat the above steps for this ancestor, and continue until there are no more ancestors. Now extract a further subsequence so that all the candidates are different: start with $\hat{\zeta}_t^{(x,j)}(1)$ and discard all members that occupy the same site. Then take the next ancestor that is left and so on. We are left with a subsequence $(\alpha_t(1), \alpha_t(2), \dots)$ of the ordered ancestor vector. The idea now is that, for k and t suitably large, with probability close to 1 each of the first k of these ancestors will contain a 2-arrow in its path and one of these k ancestors will determine the type of (x,j). □

PROOF OF THEOREM 2. Given Proposition 2.1, the proof follows very closely the proofs of Theorems 2 and 3 in Neuhauser (1992). For example, all the random walk estimates are the same. For this reason, we only sketch the proof.

Proposition 2.1 tells us that, with high probability, the x-coordinate of the distinguished particle will stay inside a set which grows linearly with time. We refer to such a set as a *triangle*. When a renewal occurs, the current triangle stops growing and a new triangle starts growing from the renewal point. The new triangle will begin at a point in the base of the previous triangle. We thus get a chain of connected triangles, inside of which the x-coordinate of the distinguished particle lies.

To prove clustering, one essentially uses the fact that symmetric random walks are recurrent in dimensions 1 and 2. This is complicated by the fact that two embedded random walks are independent only as long as their triangles do not collide (i.e., overlap). To overcome this, one shows that there is some $\delta > 0$, independent of the starting points, such that with probability at least δ one can bring any two distinguished particles within a fixed finite colony distance K without a collision of their triangles. As soon as they are within distance K, it is easy to see that, with positive probability, one can glue the two duals together (i.e., the x-coordinates of their distinguished particles will eventually coincide). After such a "colony coalescence," there is positive probability that the two distinguished particles will coalesce within the colony before being separated. Standard arguments, such as those for clustering in the voter model, finish the proof of (a).

Two main ingredients are needed to prove coexistence as stated in part (b). The first is to show convergence of one-dimensional distributions. For this we use the fact that, going back k renewals, the "tree" growing out of this renewal point has a limiting distribution. This, together with the continuous mapping theorem, establishes the convergence of the one-dimensional distributions. For the convergence of higher-dimensional distributions, we exploit the fact that two dual processes either coalesce or get separated due to the transience of symmetric random walks with $d \geq 3$. When they get separated, they are asymptotically independent. □

3. Proof of Theorem 3.

We must prove $g_t(x, y; \Delta) < g_t(x, y; 0)$ [defined in (1.2)] for t large, when starting our process with distribution π_β, $\beta \in (0, 1)$. We do this by showing, for sufficiently large t,

$$g_t(x, y; 0) \geq \beta(1-\beta) f(\theta), \tag{3.1}$$

where $f(\theta) \to 1$ as $\theta \to 0$, and

$$g_t(x, y; \Delta) \leq c\beta(1-\beta), \tag{3.2}$$

where $c = c(x, y) < 1$ is a constant which does not depend on θ.

PROOF OF (3.1). We start our process with distribution π_β, $\beta \in (0, 1)$, and assume $\Delta = 0$. This is just a voter model on $\mathbb{Z} \times \{1, \dots, N\}$ in which the voter at (x, k) adopts the opinion of the voter at (y, k), $\|y - x\|_1 = 1$, at rate θ and adopts the opinion of a voter in its own colony, chosen at random, at rate $N\lambda$. Since there is no death or selection, the state of a site is always determined by the first ancestor of that site. We have

$$\begin{aligned} &P_0\big(\zeta_t(\lfloor x\sqrt{t}\rfloor, j) = 1, \zeta_t(\lfloor y\sqrt{t}\rfloor, k) = 2\big) \\ &\quad = \beta(1-\beta) P\big(\hat{\zeta}_t^{(\lfloor x\sqrt{t}\rfloor, j)}(1) \neq \hat{\zeta}_t^{(\lfloor y\sqrt{t}\rfloor, k)}(1)\big) \\ &\quad \geq \beta(1-\beta) P(\tau_1 > t), \end{aligned} \tag{3.3}$$

where τ_1 is the time of the first *colony coalescence* of the two distinguished particles (i.e., the first time that they occupy the same colony). Since the difference in the colony variables of the two distinguished particles performs a simple random walk which jumps at rate 2θ, this last probability can be estimated, for large t, with the help of Donsker's theorem and the reflection principle. Indeed, letting B_t be a standard Brownian motion in $\mathbb{R}^1$, we have

$$\begin{aligned}\lim_{t\to\infty} P(\tau_1 > t) &= P_{x-y}(B_s \neq 0,\ s \leq 2\theta)\\ &= 2\int_0^{|x-y|} \frac{\exp(-z^2/4\theta)}{\sqrt{4\pi\theta}}\,dz = 2\int_0^{|x-y|/\sqrt{\theta}} \frac{\exp(-z^2/4)}{\sqrt{4\pi}}\,dz \qquad (3.4)\\ &\to 1 \quad \text{as } \theta \to 0. \qquad \square\end{aligned}$$

PROOF OF (3.2). We assume $\Delta \in (0, \delta_c)$. The dual particles now jump not only when they encounter θ-arrows and replacements arrows, but also when they encounter †'s. The presence of death events means that the state of a site will not always be determined by its first ancestor. However, the behavior of the embedded random walk will be enough for our purposes:

$$\begin{aligned}&P_\Delta\big(\zeta_t(\lfloor x\sqrt{t}\rfloor, j) = 1,\ \zeta_t(\lfloor y\sqrt{t}\rfloor, k) = 2 \,\big|\, \zeta_t(\lfloor x\sqrt{t}\rfloor, j) > 0,\\ &\qquad\qquad\qquad\qquad \zeta_t(\lfloor y\sqrt{t}\rfloor, k) > 0\big) \qquad (3.5)\\ &\quad = \beta(1-\beta)P\Big(\hat{\zeta}_t^{(\lfloor x\sqrt{t}\rfloor, j)} \neq \hat{\zeta}_t^{(\lfloor y\sqrt{t}\rfloor, k)} \,\Big|\, \hat{\zeta}_t^{(\lfloor x\sqrt{t}\rfloor, j)} \neq \varnothing,\ \hat{\zeta}_t^{(\lfloor y\sqrt{t}\rfloor, k)} \neq \varnothing\Big),\end{aligned}$$

where $\hat{\zeta}_t^{(\lfloor x\sqrt{t}\rfloor, j)} \neq \hat{\zeta}_t^{(\lfloor y\sqrt{t}\rfloor, k)}$ means that the duals have not coalesced yet. Note that, once two duals have coalesced, the corresponding ancestor vectors are identical thereafter.

In view of (3.4), the theorem will be proved if we can show that the conditional probability on the right-hand side of (3.5) is less than 1. In other words, we must show that the conditional probability that the two duals coalesce by time t is bounded below by a strictly positive constant (independent of θ) for sufficiently large t. Now, setting

$$\Omega_s^{(x,y)\sqrt{t}} \equiv \Big\{\hat{\zeta}_s^{(\lfloor x\sqrt{t}\rfloor, j)} \neq \varnothing,\ \hat{\zeta}_s^{(\lfloor y\sqrt{t}\rfloor, k)} \neq \varnothing\Big\}, \qquad 0 < s \leq \infty, \tag{3.6}$$

we have that $\Omega_t^{(x,y)\sqrt{t}} \supseteq \Omega_\infty^{(x,y)\sqrt{t}}$ and hence the probability of interest is

$$\begin{aligned}&P\Big(\hat{\zeta}_t^{(\lfloor x\sqrt{t}\rfloor, j)} = \hat{\zeta}_t^{(\lfloor y\sqrt{t}\rfloor, k)} \,\Big|\, \Omega_t^{(x,y)\sqrt{t}}\Big)\\ &\qquad \geq P\Big(\hat{\zeta}_t^{(\lfloor x\sqrt{t}\rfloor, j)} = \hat{\zeta}_t^{(\lfloor y\sqrt{t}\rfloor, k)} \,\Big|\, \Omega_\infty^{(x,y)\sqrt{t}}\Big)\frac{P\big(\Omega_\infty^{(x,y)\sqrt{t}}\big)}{P\big(\Omega_t^{(x,y)\sqrt{t}}\big)}.\end{aligned}$$

However, $P(\Omega_\infty^{(x,y)\sqrt{t}})/P(\Omega_t^{(x,y)\sqrt{t}}) \geq P(\Omega_\infty^{(x,y)\sqrt{t}})$, which can be bounded away from 0 as follows. Recall that η_t, the occupancy process defined in the Introduction, is a one-dimensional supercritical contact process starting with all sites occupied (since π_β is the initial distribution). Using a correlation inequality [cf. Griffeath (1979)], we have

$$\begin{aligned} P\left(\Omega_s^{(x,y)\sqrt{t}}\right) &= P\left(\eta_s(\lfloor x\sqrt{t}\rfloor) = 1,\, \eta_s(\lfloor y\sqrt{t}\rfloor) = 1\right) \\ &\geq P\left(\eta_s(\lfloor x\sqrt{t}\rfloor) = 1\right)P\left(\eta_s(\lfloor y\sqrt{t}\rfloor) = 1\right) \\ &= P\left(\eta_s(0) = 1\right)^2 \geq c_1(\Delta)^2/2 \end{aligned}$$

for sufficiently large s, where $c_1(\Delta) \equiv \liminf_{t\to\infty} P(\eta_t^0(0) = 1) > 0$ [cf. Griffeath (1979), page 42]. (Here, η_t^0 is the contact process starting with a single particle at the origin.) Now let $s \to \infty$ to get $P(\Omega_\infty^{(x,y)\sqrt{t}}) \geq c_1(\Delta)^2/2$. Thus, it suffices to show

$$P\left(\hat{\zeta}_t^{(\lfloor x\sqrt{t}\rfloor,j)} = \hat{\zeta}_t^{(\lfloor y\sqrt{t}\rfloor,k)} \,\middle|\, \Omega_\infty^{(x,y)\sqrt{t}}\right) > \text{const.} \tag{3.7}$$

holds for sufficiently large t, where const. > 0 does not depend on θ. Conditioning on $\Omega_\infty^{(x,y)\sqrt{t}}$ allows us to use the embedded random walk structure provided by Proposition 2.1.

Starting with two duals in colonies a distance $L(t) = \|\lfloor x\sqrt{t}\rfloor - \lfloor y\sqrt{t}\rfloor\|$ apart, we will use the first $t/2$ units of time to bring their embedded random walks within distance $(12/\gamma)\log L(t)$, requiring all jumps to be less than $(6/\gamma)\log L(t)$ in magnitude (with γ as in Proposition 2.1). The latter condition ensures that the corresponding triangles do not intersect, and hence the difference of the two random walks is a random walk. The next step is to bring the random walks within some fixed distance K (independent of t) in less than $[(12/\gamma)\log L(t)]^3$ units of time. Since $[(12/\gamma)\log L(t)]^3 \ll t/2$ for large t, it is then easy to see that the duals will coalesce in the remaining time with positive probability.

To handle the first step, let $S_t^{(1)}, S_t^{(2)}$ be the locations of the embedded random walks for the two duals at time t and set

$$\tilde{\tau}_L \equiv \inf\left\{t\colon |S_t^{(1)} - S_t^{(2)}| \leq \frac{12}{\gamma}\log L(t)\right\}.$$

We will show that there is some constant $\eta > 0$ (depending on $|x - y|$ but not θ) such that

$$P\left(\tilde{\tau}_L \leq \frac{t}{2},\ \max_{1\leq i\leq N(t/2)} |X_i| \leq \frac{6}{\gamma}\log L(t) \text{ in both processes} \,\middle|\, \Omega_\infty^{(x,y)\sqrt{t}}\right) \geq \eta \tag{3.8}$$

holds for sufficiently large t. Recall that $N(t/2)$ is the number of jumps of the embedded random walk by time $t/2$ and $|X_i|$ is the size of the ith jump. Now

$$\begin{aligned} &P\left(\tilde{\tau}_L \le \frac{t}{2}, \max_{1\le i\le N(t/2)} |X_i| \le \frac{6}{\gamma}\log L(t) \text{ in both processes} \,\middle|\, \Omega_\infty^{(x,y)\sqrt{t}}\right) \\ &\quad\ge P\left(\tilde{\tau}_L \le \frac{t}{2} \,\middle|\, \Omega_\infty^{(x,y)\sqrt{t}}, \max_{1\le i\le N(t/2)} |X_i| \le \frac{6}{\gamma}\log L(t) \text{ in both processes}\right) \\ &\quad\quad \times P\left(\max_{1\le i\le N(t/2)} |X_i| \le \frac{6}{\gamma}\log L(t) \text{ in both processes} \,\middle|\, \Omega_\infty^{(x,y)\sqrt{t}}\right) \qquad (3.9) \\ &\quad\ge P\left(\tau_0^{L(t)} \le \frac{t}{2}\right) P\left(\max_{1\le i\le N(t/2)} |X_i| \le \frac{6}{\gamma}\log L(t) \text{ in} \right. \\ &\qquad\qquad\qquad\qquad \left. \text{both processes} \,\middle|\, \Omega_\infty^{(x,y)\sqrt{t}}\right), \end{aligned}$$

where τ_0^a is the first hitting time of 0 for a symmetric random walk starting at a and jumping at rate 2. Using Donsker's theorem and the reflection principle again, it is easy to see that the first probability converges to $2P(B_1 \le |x-y|) \equiv g(|x-y|) > 0$, where B_t is standard Brownian motion starting at 0. The second probability is bounded below by $1 - C/[E(\tau_1)L(t)^3]$, where C is a positive constant [cf. Neuhauser (1992)]. This proves (3.8).

For the second part, we need the following result from Neuhauser (1992), which extends easily to our process for the same reason that Proposition 2.1 works. We say that a dual process is *good* at time t if it is still alive at time t and its distinguished particle has not had any offspring since the most recent renewal.

PROPOSITION 3.1 [Neuhauser (1992)]. *Consider the one-dimensional stepping-stone model with no selection. Let* $\tilde{\tau}_K^L$ *be the time it takes two embedded random walks starting distance* L *apart to come within distance* K, *and let* $A(L,K)$ *be the event* $\{\tilde{\tau}_K^L < 2L^3$, *the triangles of the corresponding duals do not collide and both duals are good within one time unit after* $\tilde{\tau}_K^L\}$. *Then there are constants* $K < \infty$ *and* $\alpha > 0$ (*independent of* θ) *such that* $P^*(A(L,K)) \ge \alpha$ *for all* $L > 0$.

From the first step, we have that after $t/2$ units of time the embedded random walks are within distance $(12/\gamma)\log L(t)$. Applying Proposition 3.1, we see that with probability at least α the embedded random walks will move within distance K in less than $[(12/\gamma)\log L(t)]^3 \ll t/2$ units of time, and one unit of time later both duals will be good. It is easy to see that in the remaining $t - t/2 - [(12/\gamma)\log L(t)]^3 - 1$ units of time, there is positive probability that the two duals will coalesce, and hence (3.7) is proved. □

4. Proof of Theorem 4. Recall that in Theorem 4 we have $d = 1$ and a periodic environment in which δ-sites are located at $x = \pm kD$, $k = 0, 1, \dots,$

where D is a fixed positive integer. At the remaining locations in $\mathbb{Z}$, we have Δ-sites. We have $s_1(x) = s_1$, $s_2(x) = 0$ if x is a δ-site, and $s_1(x) = 0$, $s_2(x) = s_2$ if x is a Δ-site. Colonies become extinct at rate Δ on Δ-sites. On δ-sites the extinction rate is 0 if there are any 1's present, and is Δ if the site contains only 2's. We set

$$c(\theta) \equiv 1 - (1 - \theta)^N. \tag{4.1}$$

We begin the proof by considering a simpler process. Here we set the reproduction rate $\lambda = \infty$, as well as $s_1 = s_2 = \infty$. That is, on Δ-colonies $s_1(x) = 0$ and $s_2(x) = \infty$; on δ-colonies $s_1(x) = \infty$ and $s_2(x) = 0$. In this case, there will be no "mixed" colonies, only pure colonies of all 1's or all 2's. A δ-colony of all 1's (resp., a Δ-colony of all 2's) will be replaced by a colony of all 2's (resp., all 1's), due to invasion by a neighboring colony of 2's (resp., 1's), at rate θ^N; that is, only a "total invasion" will be successful. For example, if there are any 1's in a δ-colony after the invasion, the infinite reproduction rate and infinite selective advantage for 1's will cause the 1's to immediately take over that colony. A δ-colony of all 2's (resp., a Δ-colony of all 1's) will be replaced by a colony of all 1's (resp., all 2's), due to invasion by a neighboring colony of 1's (resp., 2's), at rate $c(\theta)$; that is, it just takes one successful invader going into a favorable environment to convert the colony into one containing only the type with the selective advantage. This suggests that we can reduce this to a model whose state space is $\{0, 1, 2\}^{\mathbb{Z}}$, with 0 = vacant, 1 = pure 1-colony, 2 = pure 2-colony. We call this the *extreme model*. Its dynamics are as follows: 2's die at rate Δ regardless of the location; 1's die at rate Δ on Δ-sites and do not die ($\delta = 0$) on δ-sites. A 1 or a 2 will give birth onto a vacant neighboring site at rate 1. A δ-site occupied by a 1 (resp., a Δ-site occupied by a 2) will be converted to a 2 (resp., a 1) by a neighbor of that type at rate θ^N, and a δ-site occupied by a 2 (resp., a Δ-site occupied by a 1) will be converted to a 1 (resp., a 2) by a neighbor of that type at rate $c(\theta)$.

REMARK. To prove (i) [resp., (ii)], we do not really need $s_1 = \infty$ (resp., $s_2 = \infty$) in the corresponding extreme model. However, this added assumption will simplify the bookkeeping and is easy to remove at the end of the proof.

To prove (i), we begin by showing that the 1's die out a.s. in the extreme model for sufficiently large N. More precisely, we use a rescaling argument to show that a "block" of 2's spreads out linearly in time. A perturbation argument will then carry this result over to the stepping-stone model. Clearly, by monotonicity, it is enough to treat the case $D = 3$, that is, where there are two Δ-sites between the δ-sites. Start with 2's at 0 and to the left, and 1's to the right of 0. Since $\delta = \Delta = 0$ in (i), that is, there are no deaths, and since migration is only between nearest neighbors, there is a well-defined edge between the 1's and 2's for all t. Let r_t be the location of the rightmost 2 at time t. We will need the following lemma.

LEMMA 4.1. *There is a positive constant* $v = v(\theta, N, D)$ *such that* $\lim_{t\to\infty} r_t/t = v$ *a.s.*

PROOF. Since the environment has period 3, it is natural to keep track of when the right edge shifts to the left or right by three sites. When it does, everything looks the same again; that is, the process has a renewal. To see which way the right edge moves when it jumps, we consider a Markov chain (X_n) with state space $\{-3, -2, -1, 0, 1, 2, 3\}$. We start with $X_0 = 0$, and interpret this to mean that the right edge is currently on a δ-site. States -3 and 3 correspond to the right edge moving to the next δ-site to the left or right. For X_n to mimic the behavior of the right edge on its way to the next δ-site, we need the following transition probabilities: $p(i, j) = 1/2$ for all adjacent pairs, except for $p(-2, -1) = p(1, 2) = c(\theta)/(c(\theta) + \theta^N)$ and $p(-2, -3) = p(1, 0) = \theta^N/(c(\theta) + \theta^N)$. For $\theta \in (0, 1)$, $c(\theta) > \theta^N$ holds, so

$$P_0((X_n) \text{ hits } 3 \text{ before } -3) \equiv p > 1/2. \tag{4.2}$$

Moreover, a gambler's ruin recursion shows that the mean number of jumps it takes X_n to go from 0 to ± 3 is some number $m \in (0, \infty)$. Inequality (4.2) implies that the right edge has positive drift, and hence $r_t \to \infty$ as $t \to \infty$.

Next we show that r_t has positive asymptotic speed. Set $\sigma_0 = 0$, let $\sigma_n \equiv \inf\{t > \sigma_{n-1}: r_t - r_{\sigma_{n-1}} = \pm D\}$ be the time of the nth renewal, and let N_t be the number of renewals in $(0, t]$. Further, let $Z(n) \equiv r_{\sigma_n}$ define the random walk with values in $D\mathbb{Z}$ which gives the position of r_t on the δ-sites at renewal times. From (4.2), $Z(n)$ jumps to the right by D with probability p and to the left by D with probability $q \equiv 1 - p$. Now $r_t = Z(N_t) + e(t)$, where $|e(t)| < D$, so

$$\frac{r_t}{t} = \frac{Z(N_t)}{N_t}\frac{N_t}{t} + \frac{e(t)}{t}. \tag{4.3}$$

We have $Z(N_t)/N_t \to (p - q)D$ a.s., by the strong law of large numbers, and $N_t/t \to 1/E\sigma_1$ a.s., by the renewal theorem. Now r_t jumps at rate θ^N or $c(\theta)$, depending on its current state, so it is easy to bound $E\sigma_1$ by the expected inter-renewal times when r_t is run at constant rate θ^N or $c(\theta)$. Since the mean number of jumps made by r_t between renewals is m, it follows that

$$\frac{m}{c(\theta)} \le E\sigma_1 \le \frac{m}{\theta^N}. \tag{4.4}$$

Now we can put all this together in (4.3) to get

$$\lim_{t\to\infty} \frac{r_t}{t} = a(\theta, N)(p - q)D \quad \text{a.s.}, \tag{4.5}$$

where $a(\theta, N)$ is a constant satisfying

$$\frac{\theta^N}{m} \le a(\theta, N) \le \frac{c(\theta)}{m}. \qquad \square$$

Next we show that, in the extreme model, there is a space–time cone of 2's; that is, there is an interval of 2's that grows linearly in time. To prove this, we will use a rescaling argument together with Lemma 4.1.

We begin by explaining the rescaling technique. This method was developed by Bramson and Durrett and is reviewed in Durrett (1991). It is by now a standard technique and has been applied frequently [see, e.g., Bramson (1989), Bramson and Durrett (1988) and Bramson and Neuhauser (1994)]. The basic idea is to show that, for appropriate $\eta > 0$, the process under consideration dominates an oriented site percolation process in which sites are open with probability $1 - \eta$. (Sites may have some finite range of dependence.) One then shows that η can be chosen arbitrarily small. Since percolation occurs when η is close enough to 0, this will then imply that the process has the desired property.

We define K-dependent oriented site percolation with density $1 - \eta$ as follows. Let $\mathscr{L} = \{(z, k) \in \mathbb{Z}^2\colon z + k \text{ is even}\}$ denote the "even–odd lattice," where we think of the first coordinate as representing space and the second as representing time. The oriented site percolation process is a collection of random variables $\{\omega(z, k)\colon (z, k) \in \mathscr{L}\}$, with values in $\{0, 1\}$, that indicate whether the sites in $\mathscr{L}$ are open (1) or closed (0). We say the process is *K-dependent with density* $1 - \eta$ if, for any collection $(z_1, k_1), \dots, (z_n, k_n)$ of points in $\mathscr{L}$, (i) the random variables $\omega(z_1, k_1), \dots, \omega(z_n, k_n)$ are independent when, for each $i \neq j$, either $k_i \neq k_j$ or $k_i = k_j$ and $|z_i - z_j| > K$, and (ii) $P(\omega(z_i, k_i) = 1) = 1 - \eta$ for each i. By an *open path* from (y, j) to (z, k) we mean a sequence of points $(y, j) = (z_0, k_0), (z_1, k_1), \dots, (z_n, k_n) = (z, k)$ in $\mathscr{L}$ such that, for $0 \le i \le n - 1$, $(z_{i+1}, k_{i+1}) = (z_i, k_i) + (1, 1)$ or $(z_{i+1}, k_{i+1}) = (z_i, k_i) + (-1, 1)$, and all the (z_i, k_i)'s are open. Let

$$W_k^A = \{z\colon \text{there is an open path from } (y, 0) \text{ to } (z, k) \text{ for some } y \in A\}$$

and

$$\Omega^x = \bigcap_{k=0}^{\infty} \{W_k^{\{x\}} \neq \varnothing\}.$$

Think of W_k^A as the set of *wet sites* at time k connected to a *source* A at time 0. We say that *percolation occurs* starting from x if there is an infinite open path starting at x, that is, if Ω^x occurs. Define the right and left edges by

$$R_n^x = \sup W_n^{\{x\}} \quad \text{and} \quad L_n^x = \inf W_n^{\{x\}}.$$

[A detailed exposition of independent oriented site percolation is given in Durrett (1984).] The following result for K-dependent oriented site percolation appeared in Durrett and Neuhauser (1994), page 327.

LEMMA 4.2 [Durrett and Neuhauser (1994)]. *Let* $\{\omega(z, k)\colon (z, k) \in \mathscr{L}\}$ *be a K-dependent oriented site percolation with density* $1 - \eta$, $\eta \le 6^{-8(2K+1)^2}$. *Then* $P(\Omega^x) > 0$ *and*

$$(4.6) \qquad \liminf_{n \to \infty} \frac{R_{2n}^x}{n} \ge \frac{1}{2} \quad \textit{and} \quad \limsup_{n \to \infty} \frac{L_{2n}^x}{n} \le -\frac{1}{2} \quad \textit{a.s. on } \Omega^x.$$

To compare the extreme model with oriented site percolation, let $v \equiv a(\theta, N)(p-q)D$ denote the asymptotic speed in Lemma 4.1 and set $T = 3L/v$, where L is a large constant to be determined later. Also set

$$\begin{aligned} B &= [-L, L) \times (0, T], \\ \phi(z,k) &= (zL, kT) \quad \text{for } (z,k) \in \mathscr{L}, \\ B(z,k) &= \phi(z,k) + B \quad \text{for } (z,k) \in \mathscr{L}, \\ I &= [-L, L), \qquad I_z = zL + I. \end{aligned} \tag{4.7}$$

The sets $B(z,k)$, $(z,k) \in \mathscr{L}$, are $2L \times T$ squares in space–time "centered" at $\phi(z,k)$ which partition $\mathbb{Z} \times \mathbb{Z}_+$, and I_z is an interval of length $2L$ centered at zL.

The idea is to show that, in T units of time, an interval of 2's of length $2L$ will spread at least L units to the left and right. Moreover, during this interval of time, it will not spread too far, say no more than KL units in either direction. To make this more precise, for $(z,k) \in \mathscr{L}$, let $\xi_t^{z,k}$ denote the set of sites occupied by 2's at time t, in the extreme model starting at time kT with 2's at all sites in I_z and 1's elsewhere. We say that the space–time box $B(z,k)$ is *good* if

$$I_{z-1} \cup I_{z+1} \subseteq \xi_{(k+1)T}^{z,k} \quad \text{and} \quad \xi_t^{z,k} \subseteq [(z-1)L - KL, (z+1)L + KL]$$

for all $t \in [kT, (k+1)T]$. We relate good space–time boxes for the extreme model to "open sites" in the oriented percolation process by saying that $(z,k) \in \mathscr{L}$ is *open*, and setting $\omega(z,k) = 1$, if $B(z,k)$ is good. We will show that the extreme model yields good sites which dominate K-dependent oriented site percolation with density at least $1 - \eta$, where η is as in Lemma 4.2. Then we show that the stepping-stone model, with λ and s_2 sufficiently large, behaves like the extreme model in enough of the space–time boxes that we get 2's taking over.

Let r_t^A (resp., l_t^A) be the position of the rightmost (resp., leftmost) 2 in the extreme model starting with 2's at all sites in A, and 1's at all other sites. Fix $0 < \varepsilon < \frac{1}{8} 6^{-8(2K+1)^2}$. By the weak law of large numbers, we can choose L, and hence T, so large that $P(|T^{-1} r_T^{(-\infty,0]} - v| < v/2) \geq 1 - \varepsilon$. In particular, using the fact that $T = 3L/v$, we have $P(r_T^{(-\infty,(z+1)L]} > (z+1)L + \frac{3}{2}L) \geq 1 - \varepsilon$. [Note that $(z+1)L$ is the right edge of I_z.] Keeping the left and right edges from hitting will ensure that $\xi_t^{z,0} \neq \emptyset$. In fact, by taking L larger if necessary, we can force

$$P\big(r_t^{I_z} \geq (z+1)L - \tfrac{1}{2}L,\, l_t^{I_z} \leq (z-1)L + \tfrac{1}{2}L \text{ for all } t \in [0,T]\big) > 1 - \varepsilon. \tag{4.8}$$

This follows easily, as in the proof of Lemma 4.1, since the probability that a random walk with positive drift ever moves $L/2$ units to the left of its starting position can be made arbitrarily small by taking L sufficiently large (depending on the drift). Finally, note that $r_t^{I_z}$ grows like $r_t^{(-\infty,(z+1)L]}$ as long as $\xi_t^{z,0} \neq \varnothing$. Similar statements hold for $l_T^{I_z}$. Putting all this together, we have, for sufficiently large L,

$$
\begin{aligned}
&P\big(r_T^{I_z} > (z+1)L + \tfrac{3}{2}L\big) \geq 1 - \varepsilon, \\
&P\big(l_T^{I_z} < (z-1)L - \tfrac{3}{2}L\big) \geq 1 - \varepsilon.
\end{aligned} \tag{4.9}
$$

Next, we show

$$
\begin{aligned}
&P\big(r_t^{I_z} \leq (z+1)L + KL, \forall\, t \in [0,T]\big) \geq 1 - \varepsilon, \\
&P\big(l_t^{I_z} \geq (z-1)L - KL, \forall\, t \in [0,T]\big) \geq 1 - \varepsilon.
\end{aligned} \tag{4.10}
$$

By symmetry, it is enough to consider the right edge. Comparing the right edge to a Poisson process R_t which jumps at rate $c(\theta)$, we have, for all $\alpha > 0$,

$$
\begin{aligned}
&P\big(r_t^{I_z} > (z+1)L + KL \text{ for some } t \in [0,T]\big) \\
&\quad \leq P_0(R_T > KL) \leq \exp(-\alpha KL) E \exp(\alpha R_T) \\
&\quad \leq \exp\{\alpha(2c(\theta)T - KL)\}.
\end{aligned} \tag{4.11}
$$

The last inequality holds because $c(\theta) \in (0,1)$. Taking K large enough, we can force this to be less than ε, and (4.10) is proved. In summary, (4.9) and (4.10) show that, by choosing L large enough, each box $B(z,k)$ is good with probability greater than or equal to $1 - 4\varepsilon$.

Since $4\varepsilon < 6^{-8(2K+1)^2}$, the open sites in $\mathscr{L}$ dominate K-dependent oriented site percolation with density greater than $1 - 6^{-8(2K+1)^2}$. Lemma 4.2 implies that the open sites percolate (equivalently, the good boxes percolate), and, moreover, if we start with an interval I_z occupied by 2's, then with positive probability the location of the rightmost (leftmost) 2 will increase (decrease) linearly. Next, note that all sites between the left and right edges must be occupied by 2's as well; to get a 1 between the left and right edges, it would have to cross one of the edges, which is impossible due to the fact that we have nearest-neighbor interactions. Thus, we have that an interval I_z occupied by 2's will spread linearly with positive probability. Note that any Δ-colony with 2's has a positive probability of expanding to cover an interval of length $2L$ within T units of time; this interval will contain only 2's. Thus, because of our choice of ζ_0, there will be infinitely many 2-intervals of length at least $2L$ by time T. Thus, we actually get, with probability 1, a linearly growing interval of 2's which eventually covers all points in $\mathbb{Z}$. This gives an expanding cone of 2's in the space–time setting. Note that this corresponds to an expanding cone of pure 2-colonies in the stepping-stone model with $\lambda = \infty$, $s_1 = s_2 = \infty$.

We now return to the stepping-stone model with finite selection coefficients and reproduction rate. To complete the proof of (i), we employ a perturbation argument. Say that a site is *pure* if it contains a colony of all 1's or all 2's; it is *mixed* if it contains both types. A site is said to be *essentially pure* during a time interval $[t_1, t_2]$ if, whenever the site becomes mixed during the interval, it purifies to a colony containing only the type with a selective advantage before trying to invade another site and being invaded by a neighboring site.

A site is said to become *contaminated* if it moves from a pure to a mixed state, due to an invasion from a neighboring colony.

For $(z,k)\in\mathscr{L}$, let $\bar{\xi}_t^{z,k}$ denote the set of sites occupied by pure 2-colonies at time t, in the stepping-stone model starting at time kT with pure 2-colonies at all sites in I_z and pure 1-colonies elsewhere. We say that the space–time box $B(z,k)$ is $\overline{good}$ if

$$I_{z-1}\cup I_{z+1}\subseteq \bar{\xi}^{z,k}_{(k+1)T} \quad \text{and} \quad \bar{\xi}^{z,k}_t\subseteq [(z-1)L-KL,(z+1)L+KL]$$

for all $t\in[kT,(k+1)T]$. We relate $\overline{good}$ space–time boxes for the stepping-stone model to "open sites" in the oriented percolation process by saying $(z,k)\in\mathscr{L}$ is *open*, and setting $\bar{\omega}(z,k)=1$ if $B(z,k)$ is $\overline{good}$. To show that the $\overline{good}$ boxes percolate, we need to show that the following hold for sufficiently large λ, s_1, s_2:

(a) With probability at least $1-\varepsilon$, there will be no more than cL^2 contamination events in $[(z-1)L-KL,(z+1)L+KL]\times[kT,(k+1)T]$.

(b) With probability at least $1-\varepsilon/cL^2$, when a mixed colony purifies, it will purify to the type with a selective advantage.

(c) With probability at least $1-\varepsilon/cL^2$, a mixed colony will purify before attempting to invade another colony and before being invaded.

To prove (a), note that there are $(2K+2)L+1$ sites in $[(z-1)L-KL,(z+1)L+KL]$ and, since the migration rate is 1, we can choose c' so large that, with probability at least $1-\varepsilon$, there will be no more than $c'T\equiv cL$ contamination events per site during a time interval of length T. (Use the usual exponential bound on the number of jumps made by a rate-1 Poisson process). Criterion (b) follows from a simple gambler's ruin calculation. (Recall that N is fixed.) If a Δ-site is mixed with ρN 1's and $(1-\rho)N$ 2's, then the number of 2's increases by 1 at rate $\lambda(1+s_2)(1-\rho)\times[\rho+(1+s_2)(1-\rho)]^{-1}$ and decreases by 1 at rate $\lambda\rho[\rho+(1+s_2)(1-\rho)]^{-1}$. Taking s_2 sufficiently large gives the desired result. An analogous argument applies to δ-sites and s_1. To prove (c), use a gambler's ruin recursion to show that [with s_2 as in the proof of (b)] the number of reproduction events it takes to purify the colony is less than some sufficiently large constant M with probability at least $1-\varepsilon/2cL^2$. Since the migration rate is 1 and the reproduction rate is λ, we can then take λ sufficiently large (depending on N, M and s_2) so that the colony will purify before invading or being invaded with probability at least $1-\varepsilon/2cL^2$.

Putting (a), (b) and (c) together, we have that, with probability at least $1-3\varepsilon$, all the sites in $[(z-1)L-KL,(z+1)L+KL]$ are essentially pure during $[kT,(k+1)T]$. The stepping-stone model behaves like the extreme model in space–time boxes having all sites essentially pure. In particular, there will still be a cone of pure 2-colonies; only the right and left edges of the space–time cone in the extreme model might be missing from the cone of pure 2-colonies in the stepping-stone model, since these colonies could be mixed. The important thing is that, when all sites in the space–time box are essentially pure, only the edges of the cone are affected since they will purify before invading another colony. So

$$P(B(z,k) \text{ is } \overline{good}) \geq 1 - 3\varepsilon - 4\varepsilon > 1 - 6^{-8(2K+1)^2},$$

when λ, s_1, s_2 are sufficiently large. Thus, the $\overline{good}$ boxes percolate and Theorem 4(i) is proved, except for the assumption about large s_1. However, smaller values of s_1 only help the 2's, so this assumption can be dropped.

The proof of (ii) is very similar to that of (i), so we only give a sketch. Here we assume $\Delta > 0$ and show that the 1's will take over with probability 1. Start with a modified extreme model with 1's at all δ-sites at 0 and to the left, and 2's at all sites to the right of 0. Recall that 1's do not die on δ-sites and die at rate Δ on Δ-sites; 2's always die at rate Δ. Write R_t for the position of the rightmost 1. Whenever R_t reaches the next δ-site to the left or right, make all the Δ-sites to the left of it vacant and put 2's in all the vacant sites to the right of it. The 1's in this process will be contained in the set of 1's in the extreme model, so showing the right edge has a positive asymptotic speed will prove the same for the extreme model. The advantage of this process is that it has renewals whenever the right edge reaches the next δ-site to the left or right.

Given that the above system has just had a renewal, we say that a *κD-gap clearing* occurs if all the 2's in the κD sites to the right of the right edge are killed. If this happens and the rightmost 1 (on a δ-site) is not replaced by an invading 2 before the gap is formed, then we ask if the right edge of 1's can reach the next δ-site to the right before the left edge of 2's reaches it. This is how we will spread the 1's. Since the 1's have to travel (at most) distance D and the 2's must travel (at least) distance $(\kappa - 1)D$, we can choose κ so large that with probability at least $1 - \varepsilon$ the 1's will get there first. (For example, it is enough to consider the right edge $r_t^{\{0\}}$ of a contact process starting with a single particle at 0 which never dies, and the left edge $l_t^{\{\kappa D\}}$ of a contact process starting with a single particle at κD which never dies, and to ask which one hits D first, where each contact process has birth rate 1 and death rate Δ.)

For this κ and fixed $\theta \in (0, 1)$, choose N so large that the probability of getting a κD-gap clearing before a rate-θ^N invasion of the δ-site with the rightmost 1 by a neighboring 2 (if there is one) is at least $1 - \varepsilon$. With these parameter values, it is clear that the Markov chain with state given by the position of the right edge at renewal times has positive drift and hence goes to ∞. It can be shown as in the proof of (i) that the right edge has positive asymptotic speed. Interchanging the roles of 1's and 2's, we can use the same rescaling argument and perturbation argument as in (i) to complete the proof of (ii). □

5. Proof of Theorem 5.

Recall that we put an i.i.d. environment on $\mathbb{Z}^2$ with δ-sites having probability p and Δ-sites having probability $1 - p$, where $1 - p > p_c^{\text{site}}$. With probability 1, there will be an infinite connected cluster of Δ-sites. We assume we have such an environment and it is fixed forever. We will start our proof by showing the coexistence of 1's and 2's in the extreme model (i.e., in a sufficiently large box, we will find both 1's and 2's with high probability) and then finish with a rescaling argument. The extreme model, for a given environment, is defined in the obvious way.

We will first provide the intuition that goes into the proof. To demonstrate the survival of 2's, we will show that the 2's dominate a supercritical contact process in a random environment (CPRE) and that a shape theorem holds for this CPRE (Proposition 5.1). This shape theorem will imply that the CPRE can spread from one rescaling box to a neighboring rescaling box in the two-dimensional environment, thus showing that the 2's can spread in finite time to neighboring boxes. Actually, we will only be using boxes with a sufficiently high density of Δ-colonies contained in the infinite cluster. Thus, in the rescaling argument, we will have to deal with another random environment that comes from choosing these boxes. We will therefore need to modify the standard rescaling argument. This is done in Proposition 5.2. To demonstrate the survival of 1's, we will again use a rescaling argument. This time, the size of the boxes is chosen so that boxes contain δ-colonies with high probability. We then try to spread 1's from a δ-colony in one box to a δ-colony in a neighboring box along a path consisting of vacant sites. The rescaling argument then uses Proposition 5.2 in a similar way as above.

Survival of 2*'s.* Let $\xi_t \subseteq \mathbb{Z}^2$ denote the set of sites occupied by 2's in the extreme model starting with a distribution having a positive density of 2's. In fact, we will show that the 2's can survive on the Δ-sites with sufficiently high density. If a Δ-site is occupied by a 2, then the 2 is killed at a rate less than or equal to $\Delta + 4\theta^N$ since a 2 on a Δ-site can die at rate Δ and can be changed to a 1 at rate θ^N by any of its nearest neighbors which are 1's. Write $\mathscr{N}_x = \{y \in \mathbb{Z}^2\colon \|y - x\|_1 = 1\}$. A Δ-site that does not contain a 2 becomes occupied by a 2 at a rate greater than or equal to $c(\theta)\,|\mathscr{N}_x \cap \xi_t|$. This process dominates a CPRE, $\bar{\xi}_t$, defined on the same environment as follows. Particles die on δ-sites at rate ∞ (i.e., δ-sites cannot be occupied) and on Δ-sites at rate $\Delta + 4\theta^N$. A particle is born on an empty Δ-site x at rate $c(\theta)|\mathscr{N}_x \cap \bar{\xi}_t|$. These processes can be coupled so that $\bar{\xi}_0 \subseteq \xi_0$ implies $\bar{\xi}_t \subseteq \xi_t$, and hence it is enough to show the corresponding result for $\bar{\xi}_t$. For this, we need a shape theorem for a CPRE.

Let $\gamma_t \subseteq \mathbb{Z}^2$ denote the set of occupied sites at time t for a CPRE, with the random environment defined by independently designating each site of $\mathbb{Z}^2$ as *nice* with probability $1 - \alpha$ and *not nice* with probability α. Suppose the dynamics of γ_t are as follows:

(i) Particles are born on vacant nice sites x at rate $|\gamma_t \cap \mathscr{N}_x|$.

(ii) A particle at a nice site x dies at rate δ. The death rate at a site which is not nice is ∞; that is, sites which are not nice can never be occupied (except maybe at time 0).

Let γ_t^A denote the process starting from $A \subseteq \mathbb{Z}^2$, and write γ_t^x if $A = \{x\}$. Let

$$t(x,z) = \inf\{t: z \in \gamma_t^x\}$$

be the first infection time of z starting at x. Define the "hitting set" and "coupled region" as follows ($\|\cdot\|_\infty$ denotes the L^∞-norm in $\mathbb{R}^2$):

$$H_t^x = \{y \in \mathbb{R}^2: \exists\, z \in \mathbb{Z}^2 \text{ with } \|z - y\|_\infty \le 1/2 \text{ and } t(x,z) \le t\} \tag{5.1}$$

and

$$K_t^x = \bigl\{y \in \mathbb{R}^2: \exists\, z \in \mathbb{Z}^2 \text{ with } \|z - y\|_\infty \le 1/2 \text{ and } \gamma_t^x(z) = \gamma_t^{\mathbb{Z}^2}(z)\bigr\}, \tag{5.2}$$

where we have used the coordinate notation $\gamma_t^A(z) = 1$ (resp., 0) if $z \in \gamma_t^A$ (resp., $z \notin \gamma_t^A$). Identifying each $z \in \mathbb{Z}^2$ with the unit cube in $\mathbb{R}^2$ having center z makes the statement of the shape theorem (below) simpler.

A well-known result from percolation theory states that if $1 - \alpha > p_c^{\text{site}}$, the critical value for site percolation on $\mathbb{Z}^2$, then there exists a unique infinite cluster, C_∞, of nice sites with probability 1, and the origin is contained in this infinite cluster with positive probability. Let

$$C_x = \{z \in \mathbb{Z}^2: z \text{ can be reached from } x \text{ by a path of nice sites}\}$$

and

$$\overline{C}_x = \bigl\{y \in \mathbb{R}^2: \exists\, z \in \mathbb{Z}^2 \text{ with } \|z - y\|_\infty \le 1/2 \text{ and } z \in C_x\bigr\}.$$

Here is the shape theorem we need. We write $\delta_c(\mathbb{Z})$ for the critical death rate of a one-dimensional contact process (in a nonrandom environment).

PROPOSITION 5.1. *Let γ_t be the CPRE defined above and suppose $1 - \alpha > p_c^{\text{site}}$. If $\delta < \delta_c(\mathbb{Z})$, then for almost every realization e of the environment there exists a (nonrandom) convex subset U of $\mathbb{R}^2$ such that, for any $\varepsilon > 0$,*

$$t(1-\varepsilon)U \cap \overline{C}_x \subseteq H_t^x \subseteq t(1+\varepsilon)U \quad \textit{eventually}, \tag{5.3}$$

P_e-a.s. on $\{\gamma_t^x \ne \varnothing$ for all $t\}$. Furthermore,

$$t(1-\varepsilon)U \cap \overline{C}_x \subseteq H_t^x \cap K_t^x \subseteq t(1+\varepsilon)U \quad \textit{eventually}, \tag{5.4}$$

P_e-a.s. on $\{\gamma_t^x \ne \varnothing$ for all $t\}$.

This proposition is the "easy" version of Conjecture 1 in Bramson, Durrett and Schonmann (1991). Namely, we assume $\delta < \delta_c(\mathbb{Z})$ rather than the harder case $\delta < \delta_c(\text{CPRE})$, the critical death rate for the contact process in this two-dimensional random environment. The set $\overline{C}_x$ is needed for the lower bound since $\gamma_t^x \subseteq C_x$ for $t > 0$. The most important part of this proposition is the lower bound in (5.4). It says, loosely, that the process is in "contact equilibrium" inside a convex region that grows linearly in time. We prove Proposition 5.1 in the Appendix.

Applying Proposition 5.1, we see that if $(\Delta + 4\theta^N)/c(\theta) < \delta_c(\mathbb{Z})$, then $\bar{\xi}_t^x$ satisfies the shape theorem. That is, for a.e. environment e,

$$(5.5) \qquad t(1-\varepsilon)U \cap \bar{C}_x \subseteq \bar{H}_t^x \cap \bar{K}_t^x \quad \text{eventually,}$$

P_e-a.s. on $\{\bar{\xi}_t^x \neq \varnothing$ for all $t\}$. Here, $\bar{H}_t^x$ and $\bar{K}_t^x$ are the hitting set and coupled region for $\bar{\xi}_t^x$. If $\bar{\xi}_0$ has a positive density of 2's, then P_e-a.s. we can find some $x \in C_\infty$ (depending on the environment e) such that $\bar{\xi}_t^x \subseteq \bar{\xi}_t$ satisfies $\bar{\xi}_t^x \neq \varnothing$ for all t.

To prove Theorem 5 with the above result, we must do several rescalings. For this, we need to consider three-dimensional oriented site percolation (two dimensions for space and one for time) in a random environment. We begin by describing this discrete-time process. First, the i.i.d. random environment on $\mathbb{Z}^2$ independently designates each site as *nice* with probability $1 - \alpha$ and *not nice* with probability α. A realization of the environment, denoted by e, is obtained at the beginning and held fixed. It is on this environment that we run the process; the law of the process, given the environment e, will be denoted P_e. Let

$$\mathscr{L}_2 = \{(x, k) \in \mathbb{Z}^2 \times \mathbb{Z}_+ : x_1 + x_2 + k \text{ is even}\}$$

denote the appropriate even–odd lattice. [Here, $x = (x_1, x_2)$ is a typical element of $\mathbb{Z}^2$.] The three-dimensional oriented site percolation process on the environment e is a collection of random variables $\{\omega(z, k): (z, k) \in \mathscr{L}_2\}$, with values in $\{0, 1\}$, that indicate whether the sites in $\mathscr{L}_2$ are open (1) or closed (0). We say the process is *K-dependent with density* $1 - \eta$ if, for any collection $(z_1, k_1), \dots, (z_n, k_n)$ of points in $\mathscr{L}_2$, (i) the random variables $\omega(z_1, k_1), \dots, \omega(z_n, k_n)$ are independent with respect to P_e when, for each $i \neq j$, either $k_i \neq k_j$ or $k_i = k_j$ and $|z_i - z_j| > K$, and (ii) $P_e(\omega(z, k) = 1) = 1 - \eta$ for each k if z is nice in the environment e, and $P_e(\omega(z, k) = 1) = 0$ for each k if z is not nice in e. By an *open path* from (x, j) to (y, k) we mean a sequence of points $(x, j) = (z_0, k_0), (z_1, k_1), \dots, (z_n, k_n) = (y, k)$ in $\mathscr{L}_2$ such that, for $0 \le i \le n - 1$, $(z_{i+1}, k_{i+1}) = (z_i + u_i, k_i + 1)$, where $u_i \in \{(-1, 0), (0, -1), (1, 0), (0, 1)\}$ and all the (z_i, k_i)'s are open. For $A \subseteq \mathbb{Z}^2_{\text{even}} \equiv \{x \in \mathbb{Z}^2: x_1 + x_2 \text{ is even}\}$, set

$$W_k^A = \{z \in \mathbb{Z}^2: \text{there is an open path from } (y, 0) \text{ to } (z, k) \text{ for some } y \in A\}$$

and let

$$\Omega^x = \bigcap_{k=0}^{\infty} \{W_k^{\{x\}} \neq \varnothing\}.$$

Think of W_k^A as the set of *wet sites* at time k connected to a source A at time 0. Note that $W_k^A \subseteq \mathbb{Z}^2_{\text{even}}$ at even times k and $W_k^A \subseteq \mathbb{Z}^2_{\text{odd}} \equiv \{x \in \mathbb{Z}^2: x_1 + x_2$ is odd$\}$ when k is odd. We say that *percolation occurs* starting from x if there is an infinite open path starting at x, that is, if Ω^x occurs.

Define

$$n(x) = \min\{k: x \in W_k^{\{0\}}\}$$

to be the first time site x gets wet from a source at the origin. [Of course, $n(x) = \infty$ if x is not in the connected cluster of nice sites containing the origin.] Further, let

$$H_n = \{y \in \mathbb{R}^2 : \exists\ x \in \mathbb{Z}^2 \text{ with } \|x - y\|_1 \leq 1 \text{ and } n(x) \leq n\} \tag{5.6}$$

and

$$K_n = \Big\{y \in \mathbb{R}^2 : \exists\ x \in \mathbb{Z}^2 \text{ with } \|x - y\|_1 \leq 1 \text{ and } W_n^{\{0\}}(x) = W_n^{\mathbb{Z}^2_{\text{even}}}(x)\Big\}, \tag{5.7}$$

where we have used the coordinate notation $W_n^A(x) = 1$ (resp., 0) if $x \in W_n^A$ (resp., $x \notin W_n^A$). Note that we use "diamonds" of side $\sqrt{2}$, rather than unit cubes, to fatten up the points here. This is useful because of the even–odd structure of $\mathscr{L}_2$.

Let C_0 and $\overline{C}_0$ be defined as before. Here is a shape theorem for oriented site percolation in a random environment. It will be used several times below.

PROPOSITION 5.2. *Consider the K-dependent oriented site percolation process in a random environment defined above, in which* $1 - \alpha > p_c^{\text{site}}$. *Then, for almost every environment e, the following statements hold. If the fraction,* $1 - \eta$, *of nice sites that are open is sufficiently large, then there exists a (nonrandom) convex subset U of* $\mathbb{R}^2$ *such that, for any* $\varepsilon > 0$,

$$n(1 - \varepsilon)U \cap \overline{C}_0 \subseteq H_n \subseteq n(1 + \varepsilon)U \quad \textit{eventually}, \tag{5.8}$$

P_e-*a.s. on* $\Omega^0 \equiv \{W_n^{\{0\}} \neq \varnothing \text{ for all } n\}$. *Furthermore,*

$$n(1 - \varepsilon)U \cap \overline{C}_0 \subseteq H_n \cap K_n \subseteq n(1 + \varepsilon)U \quad \textit{eventually}, \tag{5.9}$$

P_e-*a.s. on* Ω^0.

For the proof of this proposition, see the Appendix.

We are now ready to begin the rescaling argument, which will consist of several stages. We start by rescaling the random environment. To do this, tile $\mathbb{Z}^2$ into $L \times L$ squares such that the square containing the origin is centered at the origin. Let $B_L = (-L/2, L/2]^2$. We say that B_L (or any of its translates) is G_1 if it contains at least $\beta_1 L^2$ sites that are in the infinite cluster C_∞. (Here and below, G_i refers to the ith level of "goodness" in the multiple rescaling.) The following lemma shows that, for any $\varepsilon > 0$, if we choose $\beta_1 > 0$ small enough and L large enough, an $L \times L$ square will be G_1 with probability at least $1 - \varepsilon$.

LEMMA 5.3. *If* $1 - \alpha > p_c^{\text{site}}$; *then, as* $L \to \infty$,

$$\frac{|B_L \cap C_\infty|}{|B_L|} \to P(x \in C_\infty) > 0 \quad a.s. \tag{5.10}$$

PROOF. Our argument follows a similar argument in Grimmett (1989). First note that $\{\mathbf{1}_{\{x \in C_\infty\}}\}$ is stationary under translations of the lattice. Hence, by the ergodic theorem [see Dunford and Schwartz (1958), Theorem 8.6.9], we have that, as $L \to \infty$,

$$\frac{1}{|B_L|} \sum_{x \in B_L} \mathbf{1}_{\{x \in C_\infty\}} \to E\mathbf{1}_{\{x \in C_\infty\}} \quad \text{a.s.} \tag{5.11}$$

That is,

$$\frac{|B_L \cap C_\infty|}{|B_L|} \to P(x \in C_\infty) \quad \text{a.s.} \tag{5.12}$$

and $P(x \in C_\infty) > 0$ since $1 - \alpha > p_c^{\text{site}}$. □

The G_1-squares form the random environment in the following rescaling argument for $\bar{\xi}_t$. We say that a G_1-square B_L is G_2 at time t if $|\bar{\xi}_t \cap B_L \cap C_\infty| \ge \beta_2 |B_L \cap C_\infty|$.

We want to show that the G_2-squares satisfy the shape theorem of Proposition 5.2. The idea is to try to spread the 2's from a G_2-square B_L to its four neighboring squares, $B_L + (L, 0)$, $B_L + (-L, 0)$, $B_L + (0, L)$, $B_L + (0, -L)$, in such a way that any of these squares which are G_1 at time 0 will be G_2 at time T. It is enough to treat one neighbor, and this is done in the following lemma.

LEMMA 5.4. *Let B'_L be any one of the four neighbors of B_L. For any $\varepsilon > 0$ we can choose L and T large enough and β_2 small enough that, if B_L is G_2 at time* 0 *and B'_L is G_1 at time* 0, *then with probability at least* $1 - \varepsilon$, *B'_L will be G_2 at time T.*

PROOF. We denote by P_e the law of the process $\bar{\xi}_t$ in the fixed environment e. Write $H_t^{B_L}$ and $K_t^{B_L}$ for the hitting set and coupled region corresponding to the process $\bar{\xi}_t$ starting with a single G_2-square B_L at time 0 (i.e., there are at least $\beta_1 \beta_2 L^2$ sites in B_L which are initially occupied by 2's). Write H_t^x and K_t^x for the hitting set and coupled region of $\bar{\xi}_t^x$. Let $\tau^{B_L} = \inf\{t\colon \bar{\xi}_t^{B_L} = \varnothing\}$ and $\tau^x = \inf\{t\colon \bar{\xi}_t^x = \varnothing\}$, where, abusing notation slightly, $\bar{\xi}_t^{B_L}$ denotes the process starting with a single G_2-square B_L. We will continue this abuse of notation below when dealing with B_L, but it should be clear what we mean. Note that $\{\tau^{B_L} = \infty\} = \bigcup_{x \in B_L} \{\tau^x = \infty\}$ and $H_t^x \subseteq H_t^{B_L}$ if $x \in B_L$. Furthermore, using the fact that $\bar{\xi}_t$ is attractive, we have $K_t^x \subseteq K_t^{B_L}$; that is, if $\bar{\xi}_t^x(y) = \bar{\xi}_t^{\mathbb{Z}^2}(y)$, then $\bar{\xi}_t^{B_L}(y) = \bar{\xi}_t^{\mathbb{Z}^2}(y)$. Now let $B_{3L} = (-3L/2, 3L/2]^2$. This is the smallest square that contains B_L and its four neighbors. In the following estimate, we use the fact that $P_e(\tau^{B_L} = \infty)$ can be made arbitrarily close to 1 by starting with sufficiently many sites in $B_L \cap C_\infty$ occupied by 2's, and this can be done by taking L sufficiently large [cf. Durrett (1988), page 285]:

$$
\begin{aligned}
&P_e\big(B_{3L} \cap C_\infty \not\subseteq H_T^{B_L} \cap K_T^{B_L}\big)\\
&\quad = P_e\big(B_{3L} \cap C_\infty \not\subseteq H_T^{B_L} \cap K_T^{B_L}, \tau^{B_L} = \infty\big)\\
&\qquad + P_e\big(B_{3L} \cap C_\infty \not\subseteq H_T^{B_L} \cap K_T^{B_L}, \tau^{B_L} < \infty\big)\\
&\quad \le P_e\bigg(B_{3L} \cap C_\infty \not\subseteq H_T^{B_L} \cap K_T^{B_L}, \bigcup_{x \in B_L} \{\tau^x = \infty\}\bigg) + P_e(\tau^{B_L} < \infty)\\
&\quad \le \sum_{x \in B_L} P_e\big(B_{3L} \cap C_\infty \not\subseteq H_T^{B_L} \cap K_T^{B_L}, \tau^x = \infty\big) + \varepsilon/2\\
&\quad \le \sum_{x \in B_L} P_e\big(B_{3L} \cap C_\infty \not\subseteq H_T^{x} \cap K_T^{x}, \tau^x = \infty\big) + \varepsilon/2\\
&\quad \le L^2 C e^{-\gamma\sqrt{L}} + \varepsilon/2,
\end{aligned}
\tag{5.13}
$$

where the last inequality follows from (A.7) in the Appendix (proof of Proposition 5.1). [Set $l = \sqrt{t}$ in (A.7) and $T = CL$ for C sufficiently large.]

To see that B_L' is G_2 at time T, we use the fact that the distribution of 2's in B_L' is close to the equilibrium measure. Since B_L' is G_1, it follows that it will be G_2 at time T with probability at least $1 - \varepsilon$ for β_2 sufficiently small. □

Lemma 5.4 and Proposition 5.2 now show that the G_2 squares satisfy the shape theorem. A comparison with Richardson's model shows that we can find a number K such that, with probability at least $1 - \varepsilon$, the squares are K-dependent. Combining this with the perturbation argument in Section 4 completes the proof.

Survival of 1*'s*. The argument resembles somewhat the one we used for the survival of 2's. We will again use a rescaling argument to show that the 1's percolate in the extreme model and then use a perturbation argument to extend the result to the stepping-stone model.

We begin by describing the idea behind the proof. Since we are assuming $1 - p > p_c^{\text{site}}$, the sites on which 1's have an advantage, that is, δ-sites, do not percolate. Therefore, in order for the 1's to move from one δ-site to another δ-site, they have to go through hostile territory. By choosing N large enough, we can make the probability that a 1 on a δ-site gets replaced by a 2 arbitrarily small. A 1 on a δ-site can thus send out enough offspring so that eventually one of its descendants will be able to colonize a nearby δ-site successfully.

The setup for the rescaling argument is as follows. As above, we tile $\mathbb{Z}^2$ into $L \times L$ squares. (The square that contains the origin is centered at the origin.) This time, we say a square is G_1 if it contains at least one δ-site.

Since the types of sites are assigned in an i.i.d. way, it follows that for any $\varepsilon > 0$ we can choose L large enough so that $B_L = (-L/2, L/2]^2$ satisfies

$$P(B_L \text{ is } G_1) \geq 1 - (1-p)^{L^2} \geq 1 - \varepsilon. \tag{5.14}$$

The G_1-squares determine the random environment in which the 1's will percolate. If ε is small enough, the good squares percolate.

To show that the 1's percolate in the extreme model, we make the following definition. We call a square G_2 if it is G_1 and at least one of the δ-sites in the extreme model is occupied by a 1. To get enough independence between neighboring squares, we define the following process: let $\bar{\xi}_t$ be the extreme process in which all sites outside of $B_L \cup B'_L$ remain occupied by 2's for all $t \geq 0$. (Note that $\bar{\xi}_t$ has a different meaning here than in the proof of survival of 2's.) The main step is contained in the following lemma.

Lemma 5.5. *Suppose $\bar{\xi}_0$ is such that B_L is G_2 and its neighbor B'_L is G_1. Then, for any $\varepsilon > 0$, we can find T large enough so that B'_L is G_2 at time T with probability at least* $1 - \varepsilon$.

Proof. We will first show that if B_L contains a δ-site occupied by a 1 at time 0, then there exists a time σ such that, by time σ, at least one of the δ-sites in B'_L will be occupied by a 1 with positive probability. To do this, choose one δ-site in B_L which is occupied by a 1 at time 0 and one δ-site in B'_L. Fix a self-avoiding path between these two δ-sites consisting of nearest-neighbor sites contained in $B_L \cup B'_L$. Since the length of the rescaling squares is L, we can choose the path so that its length is at most $3L$. We will spread the 1's from one site to a neighboring site in the following way. Suppose the 1 is at x at time 0 and we wish to spread it to y, a nearest-neighbor site on the path. We say that a *successful propagation* from x to a nearest neighbor y occurs if, in the percolation diagram for $\bar{\xi}_t$, we have the following:

(i) No deaths or arrows pointing into x occur for 1 unit of time.
(ii) A death occurs at y by time $1/2$.
(iii) No deaths occur at y during $[1/2, 1]$.
(iv) A birth occurs from x to y during $[1/2, 1]$.
(v) No other births onto y occur during $[1/2, 1]$.

Since every event in (i)–(v) has positive probability, there exists $q > 0$ so that, with probability at least q, (i)–(v) occur. We have to repeat this at most $3L$ times to get the 1 to the next δ-site. If after $3L$ units of time, the 1 has not successfully propagated along the path connecting our two designated δ-sites, we try again in the next $3L$ units of time. In this way, we perform a sequence of independent trials, each one attempting to spread the 1 to a δ-site in the neighboring square. The probability of *success* on a given trial is greater than or equal to q^{3L}. Thus, taking N sufficiently large, the 1 at the original δ-site will remain long enough to allow enough trials to yield a probability of success greater than or equal to $1 - \varepsilon$. More specifically, the number of trials necessary to get a success is geometric with parameter greater than or equal to q^{3L}. Therefore, we can choose K (depending on q and L) so large that P(success by the Kth trial) $\geq 1 - \varepsilon/3$. Now we just need to make sure that the 1 at the original δ-site remains there long enough to allow K trials (i.e., for at least $3LK \equiv T$ units of time) with probability at least $1 - \varepsilon/3$. However, this can be done by making N sufficiently large, since the rate at which 2's invade this δ-site is less than or equal to $4\theta^N$. Of course, once the 1 has spread from the δ-site in B_L to the δ-site in B'_L, this same value of N will keep the 1 at the new δ-site until time T with probability at least $1 - \varepsilon/3$. Hence the result follows. □

Lemma 5.5, together with Proposition 5.2, now shows that the G_2 boxes satisfy a shape theorem. Fixing the configuration outside of $B_L \cup B'_L$ to be all 2's for all $t \leq T$ in the above argument allows us to use independence rather than K-dependence in the rescaling argument of Section 4. Combining this with the perturbation argument from Section 4 completes the proof. We omit the details. □

APPENDIX

PROOF OF PROPOSITIONS 5.1 AND 5.2. Propositions 5.1 and 5.2 are continuous-time and discrete-time versions of the same result. The proofs are very similar, so we only give the proof of Proposition 5.1.

The upper bound in (5.3), and hence (5.4), is trivial. (Compare the process with a contact process for which $\alpha = 0$.) The proofs of the lower bounds follow closely those of the shape theorem for the d-dimensional contact process in Durrett and Griffeath (1982) [see also Durrett (1988), Chapter 11, for some improvements], so we only sketch the proof.

We must show that radial limits exist and that the process can cover small balls [cf. (A.6) and (A.7)]. There are two modifications of the proof in Durrett and Griffeath (1982) that are needed since the process cannot reach sites which are not nice. First, we cannot show that the process spreads from x to y within a time proportional to the distance $\|x - y\|_1$. Rather, we show that the process spreads from "close to x" to "close to y" within a time proportional to $\|x - y\|_1$. This argument is borrowed from first-passage percolation [see, e.g., Durrett (1988), Chapter 8a]. Second, the proof in Durrett and Griffeath (1982) uses one-dimensional embedded contact processes which live on straight lines. Since sites along straight lines might not be nice and might therefore prevent the embedded one-dimensional contact process from spreading, we will instead use self-avoiding paths through nice sites whose length will be proportional to the distance it covers.

We begin by showing how to obtain the radial limits. Since $1 - \alpha > p_c^{\text{site}}$, for each $x \in \mathbb{Z}^2$, either x is in the infinite cluster or there is an open circuit that is part of the infinite cluster and contains x in its interior. Let $\Gamma(x) = \{x\}$ if x is in the infinite open cluster, and let $\Gamma(x)$ be a minimal open circuit with the above properties otherwise. Note that a minimal open circuit is self-avoiding by definition. Let $\Gamma^0(x)$ denote the bounded component inside $\Gamma(x)$ and set

$$\bar{\Gamma}(x) = \Gamma(x) \cup \Gamma^0(x). \tag{A.1}$$

Let $t(x, y)$ denote the first time that y gets infected if the infection starts at x; that is,

$$t(x, y) = \inf\{t\colon y \in \gamma_t^x\}. \tag{A.2}$$

The quantity $t(x, y)$ is subadditive but does not have finite expectation since y may be closed with positive probability. We will replace $t(x, y)$ with an almost subadditive quantity which has finite expectation, namely,

$$\hat{t}(x, y) = \inf\{t(x', y')\colon x' \in \bar{\Gamma}(x),\ y' \in \bar{\Gamma}(y)\}. \tag{A.3}$$

That is, instead of going from x to y, we will go from "close to x" to "close to y." This idea is borrowed from first-passage percolation. We write $\hat{t}(x)$ for $\hat{t}(0,x)$ and show that $\hat{t}(x)$ is almost subadditive. This means

$$\hat{t}(x + y) \le \hat{t}(x) + \hat{s}(y) + v(x, y), \tag{A.4}$$

where $\hat{s}(y)$ is an appropriately chosen copy of $\hat{t}(y)$ which is independent of $\hat{t}(x)$ and $v(x, y)$ is an error term with

$$E\big(v(x, y)^2 \mid \gamma_t^0 \ne \varnothing \text{ for all } t\big) = O(\|x - y\|_1). \tag{A.5}$$

A standard argument then shows that radial limits exist. That is, for each x, there is a number $\mu(x)$ such that $\mu(x) > 0$ if $x \ne 0$ and

$$\hat{t}(nx)/n \to \mu(x) \tag{A.6}$$

a.s. on $\{\gamma_t^0 \ne \varnothing \text{ for all } t\}$. [See, e.g., Durrett (1988), Chapter 11.]

To prove (5.3), we must also be able to cover small balls in the following sense. Below, $B_{x,r} = \{y \in \mathbb{R}^2\colon \|x - y\|_\infty < r\}$ is the open box in $\mathbb{R}^2$ with sides of length $2r$ and center at x.

Lemma A.1. *There are constants $c, C, \beta \in (0,\infty)$ such that*

$$\begin{aligned} &P\big(B_{x,ct} \cap C_0 \not\subseteq H_{\hat{t}(0,x)+l^2+t} \text{ for some } t \ge 0 \,\big|\, \gamma_t^0 \ne \varnothing \text{ for all } t\big) \\ &\quad \le Ce^{-\beta l}. \end{aligned} \tag{A.7}$$

This was proved in Durrett and Griffeath (1982) for processes in nonrandom environments. The proof in the case of random environments is essentially the same. The only differences occur when we need to find embedded one-dimensional processes which live forever and move along certain paths. Instead of using straight lines to connect two sites x and y, we use the shortest open path connecting the open circuit around x to the open circuit around y. It follows from results in first-passage percolation that, for $\|x - y\|_1$ large enough, the length of the shortest open path is of order $O(\|x - y\|_1)$, and this is enough to complete the proof of the lemma.

Statements (A.4) and (A.5) follow from exponential estimates similar to (4) and (5) in Durrett and Griffeath (1982), and the reader is referred to that paper (cf. pages 543–544) for the details. The only change is that, whenever $B_{x,r}$ appears in Durrett and Griffeath (1982), we need to write in its place $(B_{x,r} \cap C_0) \cup \Gamma(x)$. In addition, we need to show that there are constants $C, \gamma \in (0,\infty)$ such that

$$P(\Gamma(x) \not\subseteq B_{x,r}) \leq Ce^{-\gamma r}. \tag{A.8}$$

However, this follows from well-known estimates on the existence of open circuits in percolation theory.

We can now prove (5.3) by combining (A.6) and (A.7) as in Durrett (1988), Chapter 1. The idea is to cover $t(1-\varepsilon)U$ with open balls of radius δt and then note that there is a finite subcovering $\{B_{x_i,\delta t};\ i = 1,\dots,I\}$ of balls of radius δt centered at x_i, $i = 1,\dots,I$. Using a covering lemma of Besicovitch (1927) [cf. Rudin (1974)], we can assume that the smaller balls, $B_{x_i,\delta t/3}$, $i = 1,\dots,I$, are pairwise disjoint. Using (A.8) again, we can show that a.s. the minimal circuits around the points x_i, $i = 1,\dots,I$, will be contained in their respective balls $B_{\delta t/3}(x_i)$, $i = 1,\dots,I$, eventually.

The proof of (5.4) is the same as in the case when $\alpha = 0$ (nonrandom environment) and details can be found in Durrett and Griffeath (1982). □

Acknowledgment. The authors would like to thank the referees for their helpful comments.

REFERENCES

BESICOVITCH, A. S. (1927). On the fundamental geometrical properties of linearly measurable sets of points. *Math. Ann.* **98** 442–464.

BRAMSON, M. (1989). Survival of nearest particle systems with low birth rate. *Ann. Probab.* **17** 433–443.

BRAMSON, M. and DURRETT, R. (1988). A simple proof of the stability criterion of Gray and Griffeath. *Probab. Theory Related Fields* **80** 293–298.

BRAMSON, M., DURRETT, R. and SCHONMANN, R. (1991). The contact process in a random environment. *Ann. Probab.* **19** 960–983.

BRAMSON, M. and NEUHAUSER, C. (1994). Survival of one-dimensional cellular automata under random perturbations. *Ann. Probab.* **22** 244–263.

COX, J. T. and GRIFFEATH, D. (1987). Recent results for the stepping stone model. In *Percolation Theory and Ergodic Theory of Infinite Particle Systems* (H. Kesten, ed.) 73–83. Springer, New York.

COX, J. T. and GRIFFEATH, D. (1990). Mean field asymptotics for the planar stepping stone model. *Proc. London Math. Soc.* **61** 189–208.

CROWELL, K. L. (1973). Experimental zoogeography: introductions of mice onto small islands. *American Naturalist* **107** 535–558.

DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear Operators* **1**. Wiley, New York.

DURRETT, R. (1984). Oriented percolation in two dimensions. *Ann. Probab.* **12** 999–1040.

DURRETT, R. (1988). *Lecture Notes on Particle Systems and Percolation.* Wadsworth, Belmont, CA.

DURRETT, R. (1991). A new method for proving the existence of phase transition. In *Spatial Stochastic Processes* (K. Alexander and J. Watkins, eds.). Birkhäuser, Boston.

DURRETT, R. and GRIFFEATH, D. (1982). Contact processes in several dimensions. *Z. Wahrsch. Verw. Gebiete* **59** 535–552.

DURRETT, R. and NEUHAUSER, C. (1994). Particle systems and reaction–diffusion equations. *Ann. Probab.* **22** 289–333.

GRIFFEATH, D. (1979). *Additive and Cancellative Interacting Particle Systems. Lecture Notes in Math.* **724**. Springer, New York.

GRIMMETT, G. (1989). *Percolation*. Springer, New York.

HARRIS, T. (1972). Nearest neighbor Markov interaction processes on multidimensional lattices. *Adv. in Math.* **9** 66–89.

ITATSU, S. (1985). Equilibrium measures of the stepping stone model with selection in population genetics. In *Population Genetics and Molecular Evolution* (T. Ohta and K. Aoki, eds.) 257–266. Springer, Berlin.

ITATSU, S. (1989). Ergodic properties of the stepping stone model. *Nagoya Math. J.* **114** 143–163.

KIMURA, M. and WEISS, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49** 561–576.

LEVINS, R. (1970). Extinction. In *Some Mathematical Problems in Biology* (M. Gerstenraber, ed.) **2** 75–108. Amer. Math. Soc., Providence, RI.

LIGGETT, T. (1985). *Interacting Particle Systems*. Springer, New York.

MALÉCOT, G. (1948). *The Mathematics of Heredity*. (English translation, W. H. Freeman, San Francisco, 1969.)

MCCAULEY, D. E. (1991). Genetic consequences of local population extinction and recolonization. *TREE* **6** 5–8.

NEUHAUSER, C. (1992). Ergodic theorems for the multitype contact process. *Probab. Theory Related Fields* **91** 467–506.

NEUHAUSER, C., KRONE, S. and KANG, H. (1994). A note on the stepping stone model with extinction and recolonization. In *IMA Workshop Proceedings in Mathematical Population Genetics* (P. Donnelly and S. Tavare, eds.). To appear.

RUDIN, W. (1974). *Real and Complex Analysis*, 2nd ed. McGraw-Hill, New York.

SAWYER, S. (1976). Results for the stepping stone model for migration in population genetics. *Ann. Probab.* **4** 699–728.

SAWYER, S. (1977). Rates of consolidation in a selectively neutral migration model. *Ann. Probab.* **5** 486–493.

SAWYER, S. (1979). A limit theorem for patch sizes in a selectively-neutral migration model. *J. Appl. Probab.* **16** 482–495.

SIMBERLOFF, D. S. and WILSON, E. O. (1969). Experimental zoogeography of islands: the colonization of empty islands. *Ecology* **50** 278–296.

WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics* **16** 97–159.

H.-C. KANG
DEPARTMENT OF FORESTRY
RUSSELL LABORATORIES
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

S. M. KRONE
DEPARTMENT OF MATHEMATICS
480 LINCOLN DRIVE
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

C. NEUHAUSER
DEPARTMENT OF MATHEMATICS
480 LINCOLN DRIVE
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706