The Annals of Probability
1994, Vol. 22, No. 2, 833–853

ERGODIC THEOREMS FOR INFINITE SYSTEMS OF LOCALLY INTERACTING DIFFUSIONS

J. T. COX[1] AND ANDREAS GREVEN[2]

Syracuse University and Universität Göttingen

Let $x(t) = \{x_i(t),\ i \in \mathbb{Z}^d\}$ be the solution of the system of stochastic differential equations

$$dx_i(t) = \left(\sum_{j\in\mathbb{Z}^d} a(i,j)x_j(t) - x_i(t)\right) dt + \sqrt{2g\left(x_i(t)\right)}\, dw_i(t), \qquad i \in \mathbb{Z}^d.$$

Here $g\colon [0,1] \to \mathbb{R}^+$ satisfies $g > 0$ on $(0,1)$, $g(0) = g(1) = 0$, g is Lipschitz, $a(i,j)$ is an irreducible random walk kernel on $\mathbb{Z}^d$ and $\{w_i(t),\ i \in \mathbb{Z}^d\}$ is a family of standard, independent Brownian motions on $\mathbb{R}$; $x(t)$ is a Markov process on $X = [0,1]^{\mathbb{Z}^d}$. This class of processes was studied by Notohara and Shiga; the special case $g(v) = v(1-v)$ has been studied extensively by Shiga.

We show that the long term behavior of $x(t)$ depends only on $\hat{a}(i,j) = (a(i,j) + a(j,i))/2$ and is universal for the entire class of g considered. If $\hat{a}(i,j)$ is transient, then there exists a family $\{\nu_\theta, \theta \in [0,1]\}$ of extremal, translation invariant equilibria. Each ν_θ is mixing and has density $\theta = \int x_0\, d\nu_\theta$. If $\hat{a}(i,j)$, is recurrent, then the set of extremal translation invariant equilibria consists of the point masses $\{\delta_0, \delta_1\}$. The process starting in a translation invariant, shift ergodic measure μ on X with $\int x_0\, d\mu = \theta$ converges weakly as $t \to \infty$ to ν_θ if $\hat{a}(i,j)$ is transient, and to $(1-\theta)\delta_0 + \theta\delta_1$ if $\hat{a}(i,j)$ is recurrent. (Our results in the recurrent case remove a mild assumption on g imposed by Notohara and Shiga.) For the case $\hat{a}(i,j)$ transient we use methods developed for infinite particle systems by Liggett and Spitzer. For the case $\hat{a}(i,j)$, recurrent we use a duality comparison argument.

1. Introduction and main results. The objective of this paper is to study the ergodic theory of a class of interacting diffusions with countably many components. This class was first studied by Notohara and Shiga in [15]; related processes have been considered by many others. See [5] and [19] for some recent results and references to the literature. Although the processes we consider are not explicitly "solvable," we are able to determine the set of extremal invariant measures and their domains of attraction among the class of translation invariant measures. Let us begin by defining our model.

The Model. Let $X = [0,1]^{\mathbb{Z}^d}$ be endowed with the product topology. We define a process $x(t) = \{x_i(t),\ i \in \mathbb{Z}^d\}$ taking values in X through a system of (Itô) stochastic differential equations. The basic ingredients of the model are a matrix $a(i,j)$, $i,j \in \mathbb{Z}^d$ and a function g on $I = [0,1]$. The matrix $a(i,j)$ is irreducible

Received August 1991.
[1]Research supported in part by the NSA, the NSF and the SFB 123 (at Heidelberg).
[2]Research supported in part by the SFB 123 (at Heidelberg).
AMS 1991 *subject classification.* Primary 60K35.
Key words and phrases. Interacting diffusions, infinite particle systems.

and satisfies

$$(1.1)\qquad a(i,j)\geq 0,\qquad a(i,j)=a(0,j-i)\quad \forall\, i,j\in\mathbb{Z}^d,\ \sum_j a(0,j)=1.$$

The function g satisfies

$$(1.2)\qquad g>0 \quad \text{on } (0,1),\qquad g(0)=g(1)=0,\qquad \sup_{u\neq v}\frac{|g(u)-g(v)|}{|u-v|}<\infty.$$

Throughout the paper, unrestricted sums are taken over all of $\mathbb{Z}^d$.

For a probability measure μ on X define the process $x(t)$ with initial distribution μ by

$$(1.3)\qquad dx_i(t)=\Big(\sum_j a(i,j)x_j(t)-x_i(t)\Big)dt+\sqrt{2g\big(x_i(t)\big)}\,dw_i(t),\qquad i\in\mathbb{Z}^d,$$

and $\mathcal{L}(x(0))=\mu$, where $\{(w_i(t)),\ i\in\mathbb{Z}^d\}$ is a family of standard, independent Brownian motions on $\mathbb{R}$, and $\mathcal{L}$ denotes law. It was shown by Shiga and Shimizu ([20], Theorem 3.2) that the system (1.1)–(1.3) has a unique, X-valued strong solution $x(t)$.

In the proofs of our results we will have occasion to use the following additional facts (see [20]), especially Remark 3.2): $x(t)$ is a continuous, strong Markov process with Feller semigroup $U(t)$ and generator $\mathfrak{G}$ which acts on C^2 functions f on X which depend on finitely many coordinates according to

$$(1.4)\qquad \mathfrak{G}f(x)=\sum_i g(x_i)\frac{\partial^2}{\partial x_i^2}f(x)+\sum_i\Big(\sum_j\,[a(i,j)-\delta(i,j)]x_j\Big)\frac{\partial}{\partial x_i}f(x),$$

where $\delta(i,j)=1$ if $i=j$ and is zero otherwise. We note that if $\mu=\mathcal{L}(x(0))$ is translation invariant, then so is $\mu U(t)=\mathcal{L}(x(t))$.

Motivation. The process $x(t)$ can be viewed as a diffusion limit type model for the evolution of gene frequencies. In this interpretation individuals have genotype A or B, there is a *colony* of individuals at each site $i\in\mathbb{Z}^d$ and $x_i(t)$ denotes the frequency of genotype A at colony i at time t. Changes occur in gene frequencies for two reasons: fluctuation within a colony (described by g) and migration between colonies [described by $a(i,j)$]. Notohara and Shiga consider a more general model than ours in [15], but impose additional assumptions on g. The special case $g(v)=v(1-v)$ results in a version of the *stepping stone model*, which is of great interest in mathematical genetics. This case has been studied extensively by Shiga and others (see [10] and [17]–[21]). Another example of interest in this context is the Kimura–Ohta model (see [9]), given by $g(v)=v^2(1-v^2)$.

There are several additional reasons for our interest in the class of models defined by (1.1)–(1.3). First, the analysis of the case $g(v)=v(1-v)$ is based on a *duality theory* that does not seem to be available for other choices of g. This duality is similar to the duality theory for the voter model (see [6] and [12]) and is an extremely powerful tool. However, a natural question to consider is whether or not the behavior of $x(t)$ is universal in the class of functions g which satisfy (1.2); that is, is duality *essential* or merely *convenient*? We note that if $\mathbb{Z}^d$ is replaced by a hierarchical structure and a mean field limit is taken, it turns out that the function $g(v) = v(1-v)$ does indeed play a rather special role and is related to a fixed point property of the evolution under a renormalization scheme. This is discussed by Dawson and Greven in [3] and [4].

Second, we have been interested in the long term behavior of systems with a large but finite number of interacting components. A general framework for such systems was developed by Cox and Greven in [1] and [2] and was extended by Dawson and Greven in [3]. We hope eventually to show that the entire class of models described by (1.1)–(1.3) fits into this framework. The first step in doing so is to establish the basic ergodic theory of the infinite systems, which we do here.

Results. We have found that for every g there are two qualitatively different types of behavior possible for $x(t)$, *stability* and *clustering*. This is exactly what happens in the case $g(v) = v(1-v)$. The two types of behavior are determined by the kernel

$$\widehat{a}(i, j) = \frac{a(i, j) + a(j, i)}{2}$$

and *not* the function g. To describe our results we need some additional notation. Let $\mathcal{T}$ be the set of translation invariant measures on X, and let $\mathcal{I}$ be the set of measures which are invariant for $x(t)$ (measures are *probability measures*). For $\theta \in [0,1]$ let $\mathcal{M}_\theta$ consist of all $\mu \in \mathcal{T}$ which are shift ergodic and have density $\theta = \int x_0 \, d\mu$. Let δ_θ be the measure on X which assigns mass 1 to the point $x \in X$, $x_i \equiv \theta$. Let $\Rightarrow$ denote weak convergence, and for a convex set Γ let Γ_e denote the set of extreme point of Γ.

Our first result concerns the *stable* case.

THEOREM 1. *Let $\widehat{a}(i, j)$ be transient.*

(a) *For $\theta \in [0,1]$, if $\mathcal{L}(x(0)) = \delta_\theta$, the weak limit*

$$\nu_\theta = \lim_{t\to\infty} \mathcal{L}\big(x(t)\big) \tag{1.5}$$

exists. Each ν_θ is translation invariant, associated, mixing and has density $\theta = \int x_0 \, d\nu_\theta$.

(b) *For $\theta \in [0,1]$, if $\mathcal{L}(x(0)) \in \mathcal{M}_\theta$, then*

$$\mathcal{L}\big(x(t)\big) \Rightarrow \nu_\theta \qquad \textit{as } t \to \infty. \tag{1.6}$$

(c) $(\mathcal{I} \cap \mathcal{T})_e = \{\nu_\theta,\ 0 \le \theta \le 1\}$.

The situation is quite different if $\widehat{a}(i,j)$ is recurrent, which is the case of *clustering*.

THEOREM 2. *Let $\widehat{a}(i, j)$ be recurrent.*

(a) *For $\theta \in [0,1]$, if $\mathcal{L}(x(0)) \in \mathcal{M}_\theta$, then*

$$\mathcal{L}\big(x(t)\big) \Rightarrow (1-\theta)\delta_0 + \theta\delta_1 \quad \textit{as } t \to \infty. \tag{1.7}$$

(b) $\mathcal{I}_e = \{\delta_0, \delta_1\}$.

REMARKS. The techniques we use to prove Theorems 1 and 2 do not require any special properties of $\mathbb{Z}^d$; *any countable abelian group will do.* Thus our results can be applied to the hierarchical model considered in [3]. We also obtain convergence results which are somewhat stronger than Theorems 1 and 2 (see Theorems 3 and 4, given in Section 4). Notohara and Shiga (see [15]) concentrate primarily on the case $\widehat{a}(i,j)$ recurrent. They obtain the convergence in (1.7) subject to a mild regularity condition on g.

The dichotomy between stability and clustering was first proved by Shiga [17] for the function $g(v) = v(1-v)$ using duality arguments. It does not appear that duality is available for general g. However, in the case of clustering, we are able to make a *duality comparison* argument work which allows us to prove Theorem 2.

As a replacement for duality in the stable ($\widehat{a}(i, j)$ transient) case we are able to make use of a beautiful coupling technique for infinite particle systems introduced by Liggett and Spitzer in [13] and used by Greven in [8] (see [5] for some related ideas). We believe this technique can be applied to other classes of interacting diffusions. For instance, consider the model in which the interval $I = [0,1]$ is replaced by $[0,\infty)$ and the function g satisfies $g(0) = 0$, $g(u)$ is bounded away from 0 as $u \to \infty$ and $g(u) \le cu$, $u \in (0,\infty)$ for some positive finite constant c. This model is similar in spirit to critical branching random walk (which is well understood), and one expects that versions of Theorems 1 and 2 will hold for this model. Another possibility is to replace I with $(-\infty,+\infty)$ and assume that $c_1 \le g \le c_2$ on $(-\infty,+\infty)$ for some finite positive constants c_1 and c_2. In this case one expects a version of Theorem 1 to hold, which is behavior characteristic of the critical Ornstein–Uhlenbeck process (see [5]). The main differences between the analysis of these models and the one we consider here are the complications caused by unbounded coordinates.

The rest of the paper is organized as follows. In Section 2 we study the behavior of moments and mixing properties under the evolution. In Section 3 we construct a coupling modelled on the one in [13], and in Section 4 we complete the proofs of Theorems 1 and 2.

2. Moment equations and density preservation. In order to use the coupling technique of [13] and our duality comparison argument, we need information on the evolution of the correlations of the components of $x(t)$; this will allow us to establish, later on, preservation of density in the limit $t \to \infty$

in the transient case. The first task is to derive some facts concerning $Ex_i(t)$ and $Ex_i(t)x_j(t)$. We need the following ingredients. Using the matrices $a(i,j)$ and $\widehat{a}(i,j)$ define continuous time kernels

$$a_t(i, j) = \sum_{n=0}^{\infty} e^{-t}\frac{t^n}{n!}a^n(i, j), \qquad \widehat{a}_t(i, j) = \sum_{n=0}^{\infty} e^{-t}\frac{t^n}{n!}\widehat{a}^n(i, j).$$

Note that $\widehat{a}_{2t}(i, j) = \sum_k a_t(i, k)a_t(j, k)$. Let us write P^μ and E^μ for probability and expectation if $\mathcal{L}(x(0)) = \mu$, and P^x and E^x in the case $x \in X$ and $\mu(\{x\}) = 1$.

LEMMA 1. *For $x \in X$, $t \geq 0$, and $i, j \in \mathbb{Z}^d$,*

$$E^x x_i(t) = \sum_k a_t(i,k)x_k, \tag{2.1}$$

and

$$\begin{aligned} E^x x_i(t)x_j(t) &= \sum_{k,l} a_t(i,k)a_t(j,l)x_k x_l \\ &\quad + \int_0^t \sum_k a_{t-s}(i,k)a_{t-s}(j,k)E^x g\big(x_k(s)\big)\,ds. \end{aligned} \tag{2.2}$$

PROOF. We start by deriving systems of differential equations for the moments $Ex_i(t)$ and $Ex_i(t)x_j(t)$ which we then solve in terms of random walk systems. By (1.4) applied to $f(x) = x_i$,

$$\frac{d}{dt}E^x x_i(t) = \sum_j \big(a(i, j) - \delta(i, j)\big)E^x x_j(t).$$

Since $a(i, j) - \delta(i, j)$ is the *generator* of $a_t(i, j)$, (2.1) follows.

Next, an application of (1.4) with $f(x) = x_i x_j$, gives

$$\begin{aligned} \frac{d}{dt}E^x x_i(t)x_j(t) &= E^x g\big(x_i(t)\big)\delta(i, j) \\ &\quad + \sum_k \big(a(i,k) - \delta(i,k)\big)E^x x_k(t)x_j(t) \\ &\quad + \sum_l \big(a(j,l) - \delta(j,l)\big)E^x x_i(t)x_l(t). \end{aligned} \tag{2.3}$$

We solve this system of equations by viewing it as a system of random walk equations. Consider the operator $\mathfrak{A}$ acting on bounded functions $h: \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{R}$ defined by

$$\mathfrak{A}h(i, j) = \sum_k \big(a(i,k) - \delta(i,k)\big)h(k, j) + \sum_l \big(a(j,l) - \delta(j,l)\big)h(i,l).$$

If Z_t^1 and Z_t^2 are independent random walks, each with transition function $a_t(i, j)$, then $\mathfrak{A}$ is the generator for $Z_t = (Z_t^1, Z_t^2)$. With functions F_t and H_t defined on $\mathbb{Z}^d \times \mathbb{Z}^d$ by $F_t(i, j) = E^x x_i(t)x_j(t)$ and $H_t(i, j) = E^x g(x_i(t))\delta(i, j)$, (2.3) can be rewritten in the form

$$\frac{d}{dt}F_t = \mathfrak{A}F_t + H_t.$$

By Theorem I.2.15 of [12],

$$F_t = S(t)F_0 + \int_o^t S(t-s)H_s\, ds, \tag{2.4}$$

where $S(t)$ is the semigroup generated by $\mathfrak{A}$. However, (2.4) is the same as (2.2), since $S(t)h(i, j) = \sum_{k,l} a_t(i,k)a_t(j,l)h(k,l)$. The conclusion of Theorem I.2.15 of [12] applies because $\mathfrak{A}$ is a *bounded* operator. □

We shall make use of the following consequences of Lemma 1. If μ is a translation invariant measure on X with $\int x_0\, d\mu = \theta$, then

$$E^\mu x_i(t) = \theta \quad \text{for all } t \geq 0 \tag{2.5}$$

and

$$\begin{aligned} E^\mu x_i(t)x_j(t) = {} & \sum_{k,l} a_t(i,k)a_t(j,l)E^\mu x_k(0)x_l(0) \\ & + \int_0^t \widehat{a}_{2(t-s)}(i, j)E^\mu g\big(x_0(s)\big)\, ds. \end{aligned} \tag{2.6}$$

The next step is to define a set of measures which is preserved under the semigroup $U(t)$. For $\theta \in [0,1]$ let $\mathcal{R}_\theta$ be the collection of all $\mu \in \mathcal{T}$ which satisfy (i) $\int x_0\, d\mu = \theta$ and (ii) $\sum_{k,l} a_t(i,k)a_t(j,l) \int x_k x_l\, d\mu \to \theta^2$ as $t \to \infty$. It is easy to see that if $\mu \in \mathcal{R}_\theta$, then, for all $i \in \mathbb{Z}^d$,

$$\sum_k a_t(i,k)x_k \to \theta \quad \text{in } L_2(\mu) \text{ as } t \to \infty. \tag{2.7}$$

Furthermore, if $\mu \in \mathcal{T}$ and (2.7) holds for some i, then $\mu \in \mathcal{R}_\theta$. This is because bounded harmonic functions of $a(i, j)$ are constant (Choquet–Deny), and a successful coupling exists. See Section II.1 of [12] for more on this. We will omit the proof of the following fact.

REMARK. $\mathcal{M}_\theta \subset \mathcal{R}_\theta$.

Our next result, which parallels Lemma 5.3 of [13], shows that $U(t)$ preserves the class $\mathcal{R}_\theta$, even under weak limits.

LEMMA 2. *Assume* $\mu \in \mathcal{R}_\theta$.

(a) $\mu S(t) \in \mathcal{R}_\theta$ *for all* $t \in [0,\infty)$.
(b) *Assume* $\widehat{a}(i,j)$ *is transient*, $t_n \to \infty$ *and* $\mu S(t_n) \Rightarrow \mu_\infty$. *Then* $\mu_\infty \in R_\theta$.

PROOF. Let $f_t(i,j) = E^{\mu}x_i(t)x_j(t)$. For (a), in view of (2.5), it suffices to check that

$$\sum_{k,l} a_s(i,k)a_s(j,l)f_t(k,l) \to \theta^2 \quad \text{as } s \to \infty. \tag{2.8}$$

By (2.6), the left-hand side above equals

$$\sum_{k,l} a_{s+t}(i,k)a_{s+t}(j,l)f_0(k,l) + \int_0^t \widehat{a}_{2(t-u+s)}(i,j)E^{\mu}g(x_0(u))\,du.$$

Since $\mu \in \mathcal{R}_\theta$, the first term in this expression converges to θ^2 as $s \to \infty$. The second term in the expression is nonnegative and bounded above by

$$\|g\| \int_0^t \widehat{a}_{2(t-u+s)}(i,j)\,du = \frac{\|g\|}{2}\int_{2s}^{2(t+s)} \widehat{a}_v(i,j)\,dv \le t\|g\| \sup_{v\ge 2s} \widehat{a}_v(i,j) \to 0$$

as $s \to \infty$, where $\|g\|$ denotes the sup norm of g. This proves (a).

For (b) let $f_\infty(i,j) = E^{\mu_\infty}x_ix_j$. The problem now is to show that

$$\sum_{k,l} a_s(i,k)a_s(j,l)f_\infty(k,l) \to \theta^2 \quad \text{as } s \to \infty. \tag{2.9}$$

Letting $\widehat{A}_t(i,j) = \int_0^t \widehat{a}_s(i,j)\,ds$, (2.6) implies

$$\begin{aligned}\sum_{k,l} & a_t(i,k)a_t(j,l)E^{\mu}x_k(0)x_l(0) \\ &\le f_t(i,j) \\ &\le \textstyle\sum_{k,l} a_t(i,k)a_t(j,l)E^{\mu}x_k(0)x_l(0) + \frac{1}{2}\|g\|\widehat{A}_{2t}(i,j).\end{aligned}$$

Since $\mu \in \mathcal{R}_\theta$ and $f_\infty(i,j) = \lim_{n\to\infty} f_{t_n}(i,j)$, it follows that

$$\theta^2 \le f_\infty(i,j) \le \theta^2 + \tfrac{1}{2}\|g\|\widehat{A}_\infty(i,j). \tag{2.10}$$

So (2.9) holds provided that $\sum_{k,l} a_s(i,k)a_s(j,l)\widehat{A}_\infty(k,l) \to 0$ as $s \to \infty$. However, this is a consequence of the assumption that $\widehat{a}(i,j)$ is transient:

$$\sum_{k,l} a_s(i,k)a_s(j,l)\widehat{A}_\infty(k,l) = \int_s^\infty \widehat{a}_v(i,j)\,dv \to 0 \quad \text{as } s \to \infty. \qquad \square$$

3. The coupling. In this section we compare two versions of $x(t)$ which start in different initial states. We will do this by constructing a bivariate process $\overline{x}(t) = (x^1(t), x^2(t))$ on $\overline{X} = X \times X$ (i.e., a *coupling*) such that the marginals $x^1(t)$ and $x^2(t)$ are versions of $x(t)$, $x^1(0)$ and $x^2(0)$ have specified distributions and the distribution of the bivariate process is concentrated as closely as possible on the diagonal. The main result of the section is a criterion which guarantees that the coupling is *successful*, which means that $P(|x_i^1(t) - x_i^2(t)| \geq \varepsilon) \to 0$ as $t \to \infty$ for all $i \in \mathbb{Z}^d$ and $\varepsilon > 0$. We will construct such a coupling by using the same collection of Brownian motions for each of the two coordinate processes.

LEMMA 3. *Let $\mu, \nu \in \mathcal{T}$, and consider the system of stochastic differential equations for $\bar{x}(t) = (x^1(t), x^2(t))$:*

$$
\begin{aligned}
dx_i^1(t) &= \left(\sum_j a(i,j)x_j^1(t) - x_i^1(t)\right)dt + \sqrt{2g\left(x_i^1(t)\right)}\,dw_i(t),\\
dx_i^2(t) &= \left(\sum_j a(i,j)x_j^2(t) - x_i^2(t)\right)dt + \sqrt{2g\left(x_i^2(t)\right)}\,dw_i(t),\\
\mathcal{L}\left(\bar{x}(0)\right) &= \nu \times \mu.
\end{aligned}
\tag{3.1}
$$

System (3.1) *has a unique, $\overline{X}$-valued strong solution $\bar{x}(t)$ such that $x^1(t)$ and $x^2(t)$ are both versions of $x(t)$, $\mathcal{L}(x^1(0)) = \mu$, $\mathcal{L}(x^2(0)) = \nu$ and if $\Delta_i(t) = x_i^1(t) - x_i^2(t)$, then*

$$
\frac{d}{dt}E\left|\Delta_i(t)\right| = -2\sum_j a(i,j)E\left(1\left\{\operatorname{sgn}\left(\Delta_i(t)\right) \neq \operatorname{sgn}\left(\Delta_j(t)\right)\right\}\left|\Delta_j(t)\right|\right). \tag{3.2}
$$

Hence $E|x_i^1(t) - x_i^2(t)|$ is decreasing in t for all $i \in \mathbb{Z}^d$.

PROOF. Theorem 3.2 of [20] implies that there is a unique, $\overline{X}$-valued strong solution $\overline{X}(t)$ to (3.1). We want to show that $E|x_i^1(t) - x_i^2(t)|$ is a Liapounov function. This requires calculating $dE|x_i^1(t) - x_i^2(t)|/dt$. In order to do so we would like to apply Itô's formula to the function $|x_i^1 - x_i^2|$. Although $|x_i^1 - x_i^2|$ is not everywhere smooth, we may proceed as in [22], obtaining

$$
\frac{d}{dt}E\left|\Delta_i(t)\right| = E\left(\operatorname{sgn}\left(\Delta_i(t)\right)\sum_j a(i,j)\Delta_j(t)\right) - E\left|\Delta_i(t)\right|, \tag{3.3}
$$

where $\operatorname{sgn}(u) = -1$ if $u \leq 0$ and equals 1 if $u > 0$. Another derivation can be based on the elegant use of local time by Le Gall [11]. Since $\Delta_i(t)$ is a continuous semimartingale, we can write (see Section IV.43 of [16])

$$
\left|\Delta_i(t)\right| - \left|\Delta_i(0)\right| = \int_0^t \operatorname{sgn}\Delta_i(s)\,d\Delta_i(s) + l_t^0, \tag{3.3'}
$$

where l_t^0 is the local time of $\Delta_i(t)$ at 0. Since g is Lipschitz, by Lemma 1.0 of [11] (see also Proposition V.39.3 in [16]), l_t^0 is identically zero. Equation (3.3) is now a simple consequence of this fact and (3.3′).

In order to exploit (3.3), we note that the evolution preserves translation invariance. Since μ and ν are translation invariant,

$$
E\left|\Delta_i(t)\right| = E\left|\Delta_j(t)\right| \quad \forall\, t \geq 0,\ i,j \in \mathbb{Z}^d.
$$

With this equality, some rearrangement of terms in (3.3) yields (3.2). □

The next result is the main step in the analysis of the stable case. It shows that the coupling is successful if $\widehat{a}(i, j)$ is transient.

LEMMA 4. *Assume that $\widehat{a}(i, j)$ is transient, $\mu, \nu \in \mathcal{R}_\theta$ and $\overline{x}(t) = (x^1(t), x^2(t))$ is the bivariate process defined in* (3.1). *Then, for all $i \in \mathbb{Z}^d$,*

$$\lim_{t\to\infty} E\left|x_i^1(t) - x_i^2(t)\right| = 0, \tag{3.4}$$

that is, the coupling is successful.

PROOF. Fix $i \in \mathbb{Z}^d$ and define [recall (3.2)]

$$h(t) = 2\sum_j a(i, j)E\Big\{1\Big(\operatorname{sgn}\big(\Delta_i(t)\big)\Big) \neq \operatorname{sgn}\big(\Delta_j(t)\big)\big|\Delta_j(t)\big|\Big\}.$$

The function $h(t)$ has the following properties:

(i) $h \geq 0$;
(ii) $0 \leq \int_0^\infty h(t)\,dt \leq 1$;
(iii) h is differentiable, and $|h'|$ is bounded.

The first property is obvious, and the second follows from integrating (3.2):

$$E\big|\Delta_i(t)\big| - E\big|\Delta_i(0)\big| = -\int_0^t h(s)\,ds.$$

The proof of the third property is elementary but somewhat tedious. One calculates $h'(t)$ and observes that since $x_i \in [0, 1]$ and $\sum_j a(i, j) = 1$, all terms involved are bounded, and (iii) follows. We omit the details.

From these properties it follows that $h(t) \to 0$ as $t \to \infty$. Thus, for $\varepsilon > 0$ and i,j such that $a(i, j) > 0$,

$$\lim_{t\to\infty} P\Big(\big\{\Delta_i(t) < -\varepsilon,\ \Delta_j(t) > \varepsilon\big\} \cup \big\{\Delta_i(t) > \varepsilon,\ \Delta_j(t) < -\varepsilon\big\}\Big) = 0. \tag{3.5}$$

We must strengthen this to

$$\begin{aligned}\lim_{t\to\infty} P\Big(&\big\{\Delta_i(t) < -\varepsilon,\ \Delta_j(t) > \varepsilon\big\} \\ &\cup\big\{\Delta_i(t) > \varepsilon,\ \Delta_j(t) < -\varepsilon\big\}\Big) = 0 \quad \forall\, i,j \in \mathbb{Z}^d.\end{aligned} \tag{3.6}$$

To do this we rely on the following fact.

LEMMA 5. *Let $\overline{x} \in \overline{X}$ and $i,j,k \in \mathbb{Z}^d$ satisfy $\Delta_i < 0$, $\Delta_j = 0$ and $\Delta_k > 0$. Then*

$$P^{\overline{x}}\big(\exists\, t^* \in [0,1] \text{ such that } \Delta_i(t^*) < 0,\ \Delta_j(t^*) \neq 0,\ \Delta_k(t^*) > 0\big) > 0.$$

The conclusion of Lemma 5 is intuitive, but the proof is somewhat lengthy and will be given in the Appendix. (This is one place where the analogous facts in [13] and [7] seem to be easier to prove than here.)

Using Lemma 5, we can prove (3.6) by contradiction. Suppose that (3.5) holds for the pair i,j, $a(j,k) > 0$, but (3.5) fails for the pair i,k. Then there exist $\varepsilon_0 > 0$, $\delta_0 > 0$ and $t_n \to \infty$ such that, for all n,

$$P\Big(\big\{\Delta_i(t_n) < -\varepsilon_0,\ \Delta_k(t_n) > \varepsilon_0\big\} \cup \big\{\Delta_i(t_n) > \varepsilon_0,\ \Delta_k(t_n) < -\varepsilon_0\big\}\Big) \geq \delta_0. \tag{3.7}$$

Since $\overline{X}$ is compact, by passing to a further subsequence we may assume that $\mathcal{L}(\overline{x}(t_n)) \Rightarrow \overline{\lambda}$ for some measure $\overline{\lambda}$ on $\overline{X}$. The measure $\overline{\lambda}$ must satisfy

$$\begin{aligned}
\overline{\lambda}\Big(\big\{\Delta_i < 0,\ \Delta_j > 0\big\} \cup \big\{\Delta_i > 0,\ \Delta_j < 0\big\}\Big) &= 0,\\
\overline{\lambda}\Big(\big\{\Delta_j < 0,\ \Delta_k > 0\big\} \cup \big\{\Delta_j > 0,\ \Delta_k < 0\big\}\Big) &= 0,\\
\overline{\lambda}\Big(\big\{\Delta_i < 0,\ \Delta_k > 0\big\} \cup \big\{\Delta_i > 0,\ \Delta_k < 0\big\}\Big) &> 0.
\end{aligned}$$

Suppose that $\overline{\lambda}(\Delta_i < 0,\ \Delta_k > 0) > 0$, the case $\overline{\lambda}(\Delta_i > 0,\ \Delta_k < 0) > 0$ being similar. Then $0 < \overline{\lambda}(\Delta_i < 0,\ \Delta_k > 0) = \overline{\lambda}(\Delta_i < 0,\ \Delta_j = 0,\ \Delta_k > 0)$, and thus by Lemma 5,

$$P^{\overline{\lambda}}\big(\exists\, t^* \in [0,1] \text{ such that } \Delta_i(t^*) < 0,\ \Delta_j(t^*) \neq 0,\ \Delta_k(t^*) > 0\big) > 0.$$

By path continuity there must exist $T \in [0,1]$ and $\varepsilon > 0$ such that

$$P^{\overline{\lambda}}\big(\Delta_i(T) < -\varepsilon,\ |\Delta_j(T)| > \varepsilon,\ \Delta_k(T) > \varepsilon\big) > 0.$$

By the Markov property, and the fact that $\overline{x}(t)$ is Feller, we have

$$\liminf_{n\to\infty} P\big(\Delta_i(t_n + T) \leq -\varepsilon,\ |\Delta_j(t_n + T)| \geq \varepsilon,\ \Delta_k(t_n + T) \geq \varepsilon\big) > 0.$$

This contradicts (3.5) for the pair i, j, or the pair j, k, and so (3.7) must fail. Since $a(i, j)$ is irreducible, it follows from an induction argument that (3.5) must hold for all i, j. Thus (3.6) is established.

Let us now complete the proof of Lemma 3. Suppose $t_n \to \infty$ and $\mathcal{L}(\overline{x}(t_n)) \Rightarrow \overline{\lambda}$ as $n \to \infty$. Let

$$\overline{X}_0 = \big\{(x^1,x^2) \in \overline{X}: x_i^1 \leq x_i^2\ \forall\, i \in \mathbb{Z}^d\big\} \cup \big\{(x^1,x^2) \in \overline{X}: x_i^1 \geq x_i^2\ \forall\, i \in \mathbb{Z}^d\big\}.$$

Observe that (3.6) implies $\overline{\lambda}(\overline{X}_0) = 1$. Furthermore, if λ^1 and λ^2 are the marginals of $\overline{\lambda}$, then by Lemma 2, $\lambda^1, \lambda^2 \in \mathcal{R}_\theta$. (It is here that the assumption that $a(i, j)$ is transient is critical.) Finally, note that $\overline{\lambda}$ is translation invariant. Using the last three facts we can compute

$$\begin{aligned}\int_{\overline{X}} |x_i^1 - x_i^2| \, d\overline{\lambda} &= \int_{\overline{X}_0} |x_i^1 - x_i^2| \, d\overline{\lambda} \\ &= \int_{\overline{X}_0} \sum_j a_t(i,j)|x_j^1 - x_j^2| \, d\overline{\lambda} \\ &= \int_{\overline{X}_0} \left| \sum_j a_t(i,j)(x_j^1 - x_j^2) \right| d\overline{\lambda} \\ &\le \int_{\overline{X}_0} \left(\left| \sum_j a_t(i,j)x_j^1 - \theta \right| + \left| \sum_j a_t(i,j)x_j^2 - \theta \right| \right) d\overline{\lambda} \\ &= \int_X \left| \sum_j a_t(i,j)x_j - \theta \right| d\lambda^1 + \int_X \left| \sum_j a_t(i,j)x_j - \theta \right| d\lambda^2.\end{aligned}$$

Both terms on the right-hand side above tend to zero as $t \to \infty$ since $\lambda^1, \lambda^2 \in \mathcal{R}_\theta$. This shows that $\overline{\lambda}$ concentrates on the diagonal of $\overline{X}$.

It is now easy to obtain (3.4). Suppose $t_n \to \infty$ with $E|x_i^1(t_n) - x_i^2(t_n)| \to \delta$ as $n \to \infty$. By passing to a further subsequence if necessary, we can assume that $\mathcal{L}(\overline{x}(t_n)) \Rightarrow \overline{\lambda}$ as $n \to \infty$ for some measure $\overline{\lambda}$ on $\overline{X}$. Thus, by what we have just shown,

$$\delta = \lim_{n\to\infty} E\left|x_i^1(t_n) - x_i^2(t_n)\right| = \int_{\overline{X}} |x_i^1 - x_i^2| \, d\overline{\lambda} = 0. \qquad \square$$

4. Proofs of Theorems 1 and 2.

PROOF OF THEOREM 1. Fix $\theta \in [0,1]$ and let $x(0)$ have distribution δ_θ. Choose $t_n \to \infty$ such that $\mathcal{L}(x(t_n))$ converges weakly as $n \to \infty$. Let ν_θ denote the limit,

$$\nu_\theta = \lim_{n\to\infty} \delta_\theta U(t_n). \tag{4.1}$$

Our first step is to prove that ν_θ is invariant. To do so, fix $s_0 > 0$ and let $\mu = \delta_\theta U(s_0)$ and $\nu = \delta_\theta$. Since $\nu \in \mathcal{R}_\theta$, Lemma 2 implies $\mu \in \mathcal{R}_\theta$. So by (4.1) and Lemma 4,

$$\mu U(t_n) \Rightarrow \nu_\theta.$$

Since $U(t)$ is Feller, it follows that

$$\mu U(t_n) = \delta_\theta U(s_0) U(t_n) = \delta_\theta U(t_n) U(s_0) \Rightarrow \nu_\theta U(s_0).$$

Thus $\nu_\theta = \nu_\theta U(s_0)$, that is, ν_θ is invariant.

The next step is to prove convergence to ν_θ. By Lemma 2, $\nu_\theta \in \mathcal{R}_\theta$, and so if $\mu \in \mathcal{R}_\theta$ and we set $\nu = \nu_\theta$ and apply Lemma 4, the invariance of ν_θ implies

$$\mu U(t) \Rightarrow \nu_\theta \quad \text{as } t \to \infty. \tag{4.2}$$

This proves (1.6) and that ν_θ is indeed given by the limit in (1.5).

We turn now to the question of extreme points. The first step is to show that each $\nu_\theta \in (\mathcal{I} \cap \mathcal{T})_e$. To do this, suppose that $\nu_\theta = c\mu + (1-c)\nu$, where $\mu, \nu \in \mathcal{I} \cap \mathcal{T}$ and $0 < c < 1$. Then

$$\int \Big(\sum_j a_t(i, j)(x_j - \theta) \Big)^2 d\nu_\theta = c \int \Big(\sum_j a_t(i, j)(x_j - \theta) \Big)^2 d\nu$$
$$+ (1-c) \int \Big(\sum_j a_t(i, j)(x_j - \theta) \Big)^2 d\mu.$$

However, since $\nu_\theta \in \mathcal{R}_\theta$, the left-hand side above tends to zero as $t \to \infty$, and hence each term on the right-hand side must also tend to zero. Since μ and ν are translation invariant, Lemma 2 implies that $\mu, \nu \in \mathcal{R}_\theta$. But now (4.2) implies that $\mu U(t) \Rightarrow \nu_\theta$ and $\nu U(t) \Rightarrow \nu_\theta$ as $t \to \infty$. Since μ and ν are invariant, we obtain $\mu = \nu = \nu_\theta$; consequently ν_θ must be an extreme point of $\mathcal{I} \cap \mathcal{T}$.

To show $(\mathcal{I} \cap \mathcal{T})_e \subset \{\nu_\theta,\ 0 \le \theta \le 1\}$, suppose $\mu \in \mathcal{T}$ and D is the $L_2(\mu)$ limit of $\sum_j a_t(0, j) x_j$ as $t \to \infty$. (The existence of D follows from standard L_2 ergodic theory; see [13], page 459, for more on this.) If α is the law of D, and $\mu_\rho(\cdot) = \mu(\cdot \mid D = \rho)$, then $\mu \in \mathcal{R}_\rho$ and

$$\mu = \int_0^1 \mu_\rho \, d\alpha(\rho).$$

It follows easily then from (4.2) and the Feller property that

$$\mu U(t) \Rightarrow \int_0^1 \nu_\rho \, d\alpha(\rho) \quad \text{as } t \to \infty.$$

If μ is invariant, then $\mu = \int_0^1 \nu_\rho \, d\alpha(\rho)$, a mixture of the ν_θ.

It remains to show that each ν_θ is *associated* and *mixing*. Association means that for all bounded functions f_1, f_2 on X which depend on only finitely many coordinates x_i, and which are coordinate-wise increasing,

$$E^{\nu_\theta} f_1(x) f_2(x) \ge E^{\nu_\theta} f_1(x) E^{\nu_\theta} f_2(x).$$

To prove that ν_θ is associated we apply a result of Herbst and Pitt [8]. After a standard approximation argument, Theorem 1.1 (and its extensions) in [8] and the form of the generator in (1.4) imply that $U(t)$ preserves associated measures. Since δ_θ is associated and $\delta_\theta U(t) \Rightarrow \nu_\theta$, ν_θ must be associated.

For mixing we use association and a result of Newman and Wright in [14]. Define

$$Y_0 = \sum_{j \in A} c_j x_j, \qquad Y_n = \sum_{j \in A} d_j x_{j+n},$$

where $A \subset \mathbb{Z}^d$ is finite, and the c_j and d_j, $j \in A$, are positive constants. Then under ν_θ, Y_0 and Y_n are associated, and by (2.2) of [14], for $s, t \in \mathbb{R}$,

$$\left|E^{\nu_\theta} e^{i(sY_0+tY_n)} - E^{\nu_\theta} e^{isY_0} E^{\nu_\theta} e^{itY_n}\right| \le |st| \mathrm{Cov}^{\nu\theta}(Y_0, Y_n).$$

By (2.10),

$$\mathrm{Cov}^{\nu_\theta}(Y_0, Y_n) = \sum_{j,k\in A} c_j d_k \,\mathrm{Cov}^{\nu_\theta}(x_j, x_{k+n}) \le \tfrac{1}{2} \sum_{j,k\in A} c_j d_k \|g\| \widehat{A}(j, k+n),$$

which tends to zero as $n \to \infty$. Thus, Y_0 and Y_n are asymptotically independent under ν_θ as $n \to \infty$, and ν_θ is mixing. □

The previous argument actually proves a somewhat stronger statement than Theorem 1.

THEOREM 3. *Assume $\widehat{a}(i, j)$ is transient, $\mu \in \mathcal{T}$ and α is the distribution of the $L_2(\mu)$ limit as $t \to \infty$ of $\Sigma_j a_t(0, j)x_j$. Then*

$$\mu U(t) \Rightarrow \int_0^1 \nu_\rho \, d\alpha(\rho) \quad \text{as } t \to \infty.$$

We turn now to the case $\widehat{a}(i, j)$ is recurrent. The argument is this case does not depend on the coupling defined in Section 3.

We will now state and prove an ergodic theorem for $x(t)$ starting in fixed initial states, from which we will derive Theorem 2.

THEOREM 4. *Assume $\widehat{a}(i, j)$ is recurrent, and $x \in X$ satisfies*

$$\lim_{t\to\infty} \sum_k a_t(i,k)x_k = \theta \quad \forall\, i \in \mathbb{Z}^d. \tag{4.3}$$

If $x(0) = x$, then

$$\mathcal{L}(x(t)) \Rightarrow (1-\theta)\delta_0 + \theta\delta_1 \quad \text{as } t \to \infty.$$

PROOF. By Lemma 1 and (4.3), $E^x x_i(t) \to \theta$ for all $i \in \mathbb{Z}^d$, and hence it suffices to prove that

$$E^x x_i(t) x_j(t) \to \theta \tag{4.4}$$

for all $i, j \in \mathbb{X}^d$. We do so by using a duality comparison technique.

Consider the Markov chain $W_t = (W_t^0, W_t^1, W_t^2)$ with state space $\{1,2\} \times \mathbb{Z}^d \times \mathbb{Z}^d$ whose evolution is governed by the following transitions:

$$\begin{aligned}
(2,i,j) &\to (2,k,j) && \text{at rate } a(i,k)\\
&\to (2,i,l) && \text{at rate } a(j,l)\\
&\to (1,i,i) && \text{at rate } c\delta(i,j)\\
(1,i,i) &\to (1,k,k) && \text{at rate } a(i,k).
\end{aligned}$$

Here c is a strictly positive parameter which we shall choose later. The process W_t describes the evolution of two random walks which move independently according to $a_t(i, j)$, but *coalesce at rate* c whenever they occupy the same site. The first coordinate $W_t^0 = 1$ if the walks have coalesced by time t; $W_t^0 = 2$ otherwise. Note that walks can occupy the same site and then move apart without coalescing.

We will need the following simple facts concerning W_t. First, W_t^1 and W_t^2 are each random walks with transition kernel $a_t(i, j)$, and hence

$$P^{(2,i,j)}\big(W_t^1 = k\big) = a_t(i,k). \tag{4.5}$$

Second,

$$\lim_{t\to\infty} P^{(2,i,j)}\big(W_t^0 = 2\big) = 0. \tag{4.6}$$

This is a consequence of the *recurrence* of $\widehat{a}_t(i, j)$ and the fact that c is strictly positive.

Now fix $\varepsilon > 0$ and choose $c = c(\varepsilon) > 0$ such that $\widetilde{g}$ defined by

$$\widetilde{g}(v) = c(v - \varepsilon)(1 - \varepsilon - v)$$

satisfies $\widetilde{g} \le g$ on [0, 1]. (The basic assumptions on g in (1.2) make this possible.) Note that $\widetilde{g}$ assumes negative values, so we cannot define a process $\widetilde{x}(t)$ using $\widetilde{g}$ in (1.3). Nevertheless, we will use $\widetilde{g}$, W_t and a duality comparison argument to show that

$$\liminf_{t\to\infty} E^x x_i(t) x_j(t) \ge \theta - \varepsilon(1 - \varepsilon). \tag{4.7}$$

Since $\varepsilon > 0$ is arbitrary and $\limsup E^x x_i(t)x_j(t) \le \theta$, this proves (4.4).

We begin by observing that

$$\frac{d}{dt}E^x\big(x_i(t) - \varepsilon\big) = \sum_k \big(a(i,k) - \delta(i,k)\big)E^x\big(x_k(t) - \varepsilon\big) \tag{4.8}$$

and

$$\begin{aligned}
&\frac{d}{dt}E^x\Big(\big(x_i(t) - \varepsilon\big)\big(x_j(t) + \varepsilon\big)\Big)\\
&\quad= \sum_k \big(a(i,k) - \delta(i,k)\big)E^x\Big(\big(x_k(t) - \varepsilon\big)\big(x_j(t) + \varepsilon\big)\Big)\\
&\qquad+ \sum_l \big(a(j,l) - \delta(j,l)\big)E^x\Big(\big(x_i(t) - \varepsilon\big)\big(x_l(t) + \varepsilon\big)\Big)\\
&\qquad+ E^x\widetilde{g}\big(x_i(t)\big)\delta(i, j) + E^x\Big(g\big(x_i(t)\big) - \widetilde{g}\big(x_i(t)\big)\Big)\delta(i, j).
\end{aligned} \tag{4.9}$$

We wish to view these equations from the point of view of the process W_t and its generator, so we define functions F_t and H_t on $\{1,2\} \times \mathbb{Z}^d \times \mathbb{Z}^d$ by

$$F_t(1,i,i) = E^x(x_i(t) - \varepsilon), \quad F_t(2,i, j) = E^x\big(x_i(t) - \varepsilon\big)\big(x_j(t) + \varepsilon\big)$$

and

$$H_t(1,i,j) \equiv 0, \quad H_t(2,i,j) = E^x\Big(g\big(x_i(t)\big) - \tilde{g}\big(x_i(t)\big)\Big)\delta(i,j).$$

Observe that $H_t \geq 0$ by construction.

We can now rewrite the system of equations in (4.8) and (4.9) in the form

$$\frac{d}{dt}F_t = \mathfrak{B}F_t + H_t, \tag{4.10}$$

where $\mathfrak{B}$ denotes the generator of the process W_t. Since $\mathfrak{B}$ is a bounded operator, we can again apply Theorem I.2.15 of [12] to obtain, with $V(t)$ denoting the semigroup for W_t,

$$F_t = V(t)F_0 + \int_0^t V(t-s)H_s\,ds \geq V(t)F_0.$$

Thus,

$$\begin{aligned}
F_t(2,i,j) &\geq V(t)F_0(2,i,j)\\
&= E^{(2,i,j)}F_0(W_t)\\
&= E^{(2,i,j)}\big(F_0(W_t);\ W_t^0 = 1\big) + E^{(2,i,j)}\big(F_0(W_t);\ W_t^0 = 2\big)\\
&= \sum_k P^{(2,i,j)}(W_t^0 = 1,\ W_t^1 = k)(x_k - \varepsilon) + E^{(2,i,j)}\big(F_0(W_t);\ W_t^0 = 2\big)\\
&= \sum_k P^{(2,i,j)}(W_t^1 = k)(x_k - \varepsilon) - \sum_k P^{(2,i,j)}(W_t^0 = 2,\ W_t^1 = k)(x_k - \varepsilon)\\
&\quad + E^{(2,i,j)}\big(F_0(W_t),\ W_t^0 = 2\big)\\
&\geq \sum_k a_t(i,k)(x_k - \varepsilon) - P^{(2,i,j)}(W_t^0 = 2)\\
&\to \theta - \varepsilon
\end{aligned}$$

as $t \to \infty$, by (4.3). This proves

$$\liminf_{n\to\infty} E^x\big((x_i(t)-\varepsilon)(x_j(t)+\varepsilon)\big) \geq \theta - \varepsilon$$

and (4.7) follows. □

PROOF OF THEOREM 2. Let μ be a measure on X which satisfies

$$\sum_k a_t(i,k)\int x_k\,d\mu \to \theta \quad \text{as } t \to \infty, \tag{4.11}$$

for all i. Then, as above,

$$\begin{aligned}
E^\mu\Big(\big(x_i(t)-\varepsilon\big)\big(x_j(t)+\varepsilon\big)\Big) &= \int E^x\Big(\big(x_i(t)-\varepsilon\big)\big(x_j(t)+\varepsilon\big)\Big)\,d\mu(x)\\
&\geq \int \sum_k a_t(i,k)(x_k-\varepsilon)\,d\mu(x) - P^{2,i,j}(W_t^0 = 2)\\
&\to \theta - \varepsilon.
\end{aligned}$$

Thus, $E^{\mu}x_i(t)x_j(t) \to \theta$, and if $\mathcal{L}(x(0)) = \mu$, then

$$\mathcal{L}\big(x(t)\big) \Rightarrow (1-\theta)\delta_0 + \theta\delta_1. \tag{4.12}$$

Finally, to prove that $(\mathcal{I})_e = \{\delta_0, \delta_1\}$, suppose μ is invariant for $x(t)$. Then, for $t \geq 0$

$$E^{\mu}x_i(0) = E^{\mu}x_i(t) = \sum_j a_t(i, j)E^{\mu}x_j(0).$$

This shows that the function $i \to E^{\mu}x_i(0)$ is a bounded harmonic function for the random walk kernel $a_t(i, j)$, and hence, by the Choquet–Deny lemma, must be constant, say $E^{\mu}x_i(0) \equiv \theta$. In this case (4.11) must hold, and if μ is the initial distribution of $x(t)$, then (4.12) must hold too. However, since μ is *invariant*, $\mu = (1-\theta)\delta_0 + \theta\delta_1$. This proves $\mathcal{I}_e = \{\delta_0, \delta_1\}$. □

APPENDIX

We divide the proof of Lemma 5 into a series of steps. In Steps 1–3 we prove some auxiliary facts [(A.1), (A.3) and (A.6)] concerning the diffusions $x(t)$ and $\overline{x}(t)$. These facts are combined in Step 4 to obtain the proof of Lemma 5. It may be useful to read Step 4 first.

STEP 1. If $x \in X$ with $x_i = 0$ and $x_k > 0$, then

$$P^x\big(\exists\, t^* > 0 \text{ such that } x_i(t) = 0\ \forall\, t \in [0, t^*]\big) = 0. \tag{A.1}$$

PROOF OF (A.1). Writing $(ax)_i(s)$ for $\sum_l a(i,l)x_l(s)$, by (1.3),

$$x_i(t) = \int_0^t \big((ax)_i(s) - x_i(s)\big)\, ds + \int_0^t \sqrt{2g\big(x_i(s)\big)}\, dw_i(s).$$

Fix $T > 0$ and suppose that $x_i(t) = 0$ for all $t \in [0, T]$, which implies $g(x_i(t)) = 0$ for all $t \in [0, T]$. Then

$$\int_0^t (ax)_i(s)\, ds = -\int_0^t \sqrt{2g\big(x_i(s)\big)}\, dw_i(s) = 0, \qquad t \in [0, T]. \tag{A.2}$$

From this it follows that $(ax)_i(t) = 0,\ t \in [0, T]$, and hence if $a(i, l) > 0$, then $x_l(t) = 0,\ t \in [0, T]$. Since the kernel a is irreducible, iteration of this argument shows that $x_k(t) = 0,\ t \in [0, T]$. However, this contradicts the assumption $x_k > 0$. So for fixed $T > 0$, $P^x(x_i(t) = 0\ \forall\, t \in [0, T]) = 0$. This fact and path continuity imply (A.1). □

STEP 2. If $\overline{x} \in \overline{X}$ and $g(x_i^1) \neq g(x_i^2)$, then, for all j,

$$P^{\overline{x}}\big(\exists\, t^* > 0 \text{ such that } \Delta_j(t) = 0\ \forall\, t \in [0, t^*]\big) = 0. \tag{A.3}$$

PROOF OF (A.3). Fix $T > 0$ and suppose that $\Delta_j(t) = 0$ for all $t \in [0, T]$, which implies $g(x_j^1(t)) = g(x_j^2(t))$, $t \in [0, T]$. Writing $(a\Delta)_j(s)$ for $\sum_l a(j,l)\Delta_l(s)$, by (3.1),

$$\text{(A.4)} \qquad \int_0^t (a\Delta)_j(s)\,ds = -\int_0^t \left(\sqrt{2g(x_j^1(s))} - \sqrt{2g(x_j^2(s))}\,\right) dw_j(s) = 0,$$

for all $t \in [0, T]$. From this it follows that $(a\Delta)_j(t) = 0$, $t \in [0, T]$, and writing $(a\Delta)_j(t)$ as a stochastic integral gives

$$\text{(A.5)} \qquad \int_0^t (a^2\Delta)_j(s)\,ds = -\int_0^t \sum_l a(j,l)\left(\sqrt{2g(x_l^1(s))} - \sqrt{2g(x_l^2(s))}\,\right) dw_l(s)$$

for $t \in [0, T]$. Now the right-hand side of (A.5) is a continuous martingale with quadratic variation process

$$\int_0^t \sum_l \left(a(j,l)\left(\sqrt{2g(x_l^1(s))} - \sqrt{2g(x_l^2(s))}\,\right)\right)^2 ds,$$

while the left-hand side is a process of bounded variation. This implies that both integrals in (A.5) must be zero. Therefore $(a^2\Delta)_j(t) = 0$, $t \in [0, T]$, and if $a(j,l) > 0$, then $g(x_l^1(t)) = g(x_l^2(t))$, $t \in [0, T]$. Since the kernel a is irreducible, iteration of this argument shows that $g(x_i^1(t)) = g(x_i^2(t))$, $t \in [0, T]$. This contradicts the assumption that $g(x_i^1) \neq g(x_i^2)$ so $P^{\bar{x}}(\Delta_j(t) = 0\ \forall\, t \in [0, T]) = 0$, and (A.3) must hold. □

STEP 3. If $\bar{x} \in \overline{X}$, $i, k \in \mathbb{Z}^d$ with $\Delta_i < 0$, $\Delta_k > 0$ and $g(x_i^1) = g(x_i^2)$, then

$$\text{(A.6)} \qquad \begin{aligned} &P^{\bar{x}}\big(\exists\, t^* \in \left[0, \tfrac{1}{2}\right] \text{ such that } \Delta_i(t^*) < 0, \\ &\qquad \Delta_k(t^*) > 0,\ g\big(x_i^1(t^*)\big) \neq g\big(x_i^2(t^*)\big) > 0. \end{aligned}$$

PROOF OF (A.6). By assumption, $x_i^1 < 1$ and $x_k^1 > 0$. If $x_i^1 > 0$, set $t_0 = 0$. Otherwise, by (A.1) and path continuity, there exists $t_0 \in [0, \frac{1}{4}]$ such that

$$\text{(A.7)} \qquad P^{\bar{x}}\big(\Delta_i(t_0) < 0,\ \Delta_k(t_0) > 0,\ 0 < x_i^1(t_0) < 1\big) > 0.$$

We will prove that if $\tilde{x} \in \overline{X}$ satisfies $\widetilde{\Delta}_i < 0$, $\widetilde{\Delta}_k > 0$, and $0 < \tilde{x}_i^1 < 1$, then

$$\text{(A.8)} \qquad \begin{aligned} &P^{\tilde{x}}\Big(\exists\, t^* \in [0, \tfrac{1}{4}] \text{ such that } \Delta_i(t^*) < 0, \\ &\qquad \Delta_k(t^*) > 0,\ g\big(x_i^1(t^*)\big) \neq g\big(x_i^2(t^*)\big)\Big) > 0. \end{aligned}$$

By the Markov property, (A.7) and (A.8) imply (A.6).

The idea behind (A.8) is simple, although the details are somewhat tedious. Since $g(\widetilde{x}_i^1) \neq 0$, we can use $w_i(t)$ to drive $x_i^1(t)$ toward zero. Since $g(0) = 0$ and $x_i^1(t)$ and $x_i^2(t)$ move almost in parallel while $g(x_i^1(t)) = g(x_i^2(t))$, at some point $g(x_i^1(t))$ and $g(x_i^2(t))$ must differ. Here are the details.

Define the martingales

$$M_i(t) = \int_0^t \sqrt{2g(x_i^1(s))}\, dw_i(s),$$
$$M_k(t) = \int_0^t \left(\sqrt{2g(x_k^1(s))} - \sqrt{2g(x_k^2(s))}\right) dw_k(s).$$

Using [ ] to denote quadratic variation processes, it is easy to see that

$$\begin{aligned} [M_i](t) &= \int_0^t 2g(x_i^1(s))\, ds, \\ [M_k](t) &= \int_0^t \left(\sqrt{2g(x_k^1(s))} - \sqrt{2g(x_k^2(s))}\right)^2 ds, \end{aligned} \tag{A.9}$$

and $[M_i, M_k](t) \equiv 0$. By Knight's theorem (see [13], Section IV 34), there are standard, *independent* Brownian motions $W_i(t)$ and $W_k(t)$ such that

$$M_i(t) = W_i([M_i](t)), \qquad M_k(t) = W_k([M_k](t)). \tag{A.10}$$

Recall that we are assuming $\widetilde{x}$ satisfies $\widetilde{\Delta}_i < 0$, $\widetilde{\Delta}_k > 0$ and $0 < \widetilde{x}_i^1 < 1$. We may also assume $g(\widetilde{x}_i^1) \neq g(\widetilde{x}_i^2)$, and hence $0 < \widetilde{x}_i^1 < \widetilde{x}_i^2 < 1$; otherwise we may set $t^* = 0$ in (A.8). Now choose $\delta \in (0, \frac{1}{15})$ such that $\widetilde{x}_i^1, \widetilde{x}_i^2 \in [5\delta, 1-5\delta]$, $-\widetilde{\Delta}_i > 5\delta$ and $\widetilde{\Delta}_k > 5\delta$. Let $\xi \in (0, \delta)$ be such that $g(\xi) < \min\{g(u)\colon \delta \le u \le 1-\delta\}$. Let $c_1 = \min\{g(u)\colon \xi \le u \le 1-\xi\}$ and let $c_2 = 2\|g\|$. In view of (A. 9),

$$[M_i](t) \le c_2 t, \qquad [M_k](t) \le c_2 t, \qquad t \ge 0,$$

and

$$[M_i](t) \ge c_1 t \quad \forall\, t \in [0,T] \quad \text{on} \quad \{x_i^1(t) \in [\xi, 1-\xi] \quad \forall\, t \in [0,T]\}.$$

Fix $T \in [0, \xi]$ and define the events

$$\Omega_0 = \left\{ \min_{t\in[0,c_1T]} W_i(t) < -1,\ \max_{t\in[0,c_2T]} W_i(t) \le \delta,\ \max_{t\in[0,c_2T]} |W_k(t)| \le \delta \right\}$$

and

$$\Omega_1 = \{\exists\, t^* \in [0,T] \text{ such that } \Delta_i(t^*) < 0,\ \Delta_k(t^*) > 0,\ g(x_i^1(t^*)) \neq g(x_i^2(t^*))\}.$$

Clearly Ω_0 has positive probability. We will prove $\Omega_0 \subset \Omega_1$, thus showing that $P^{\tilde{x}}(\Omega_1) > 0$, establishing (A.8).

Let us deal with $\Delta_k(t)$ first. Since $|(a\Delta)_k(s) - \Delta_k(s)| \le 2$ and $|M_k(t)| = |W_k([M_k]\,(t))| \le \delta$ on Ω_0, we have

$$\Delta_k(t) = \Delta_k(0) + \int_0^t ((a\Delta)_k(s) - \Delta_k(s))\, ds + M^k(t) \ge 5\delta - 2\delta - \delta = 2\delta,$$

for $t \in [0, T]$. That is, on Ω_0,

$$\Delta_k(t) > 0, \qquad t \in [0, T]. \tag{A.11}$$

Next, since on Ω_0

$$x_i^1(t) = x_i^1(0) + \int_0^t \big((ax^1)_i(s) - x_i^1(s)\big)\, ds + W_i\big([M_i](t)\big), \tag{A.12}$$

we have

$$x_i^1(t) \le 1 - 10\delta + \delta + \delta = 1 - 8\delta, \qquad t \in [0, T]. \tag{A.13}$$

Now define the random times

$$\sigma = \inf\{t \ge 0 \colon x_i^1(t) = \xi\} \quad \text{and} \quad \tau = \inf\big\{t > 0 : \; g\big(x_i^1(t)\big) \ne g\big(x_i^2(t)\big)\big\}.$$

We will prove that, on Ω_0,

$$\tau < \sigma < T \tag{A.14}$$

and

$$x_i^2(\tau) \ge x_i^1(\tau) + 4\delta. \tag{A.15}$$

By combining (A.11), (A.13) and (A.14), path continuity implies there must exist t^* as required in the definition of Ω_1, proving $\Omega_0 \subset \Omega_1$.

To finish we must prove (A.14) and (A.15), and we do this by contradiction. Suppose first that $\sigma \ge T$. In view of (A.13), this certainly implies that

$$x_i^1(t) \in [\xi, 1 - \xi], \qquad t \in [0, T].$$

Then $[M_i](t) \ge c_1 t$, and thus, on Ω_0, there exists $\kappa < T$ such that $M_i(\kappa) = W_i([M_i](\kappa)) = -1$. But then by (A.12), $x_i^1(\kappa) \le 1 - 10\delta + \delta - 1 < 0$, which is impossible. So $\sigma < T$ on Ω_0.

Now suppose that $\tau \ge \sigma$. For $t \le \tau$,

$$\Delta_i(t) = \Delta_i(0) + \int_0^t \big((a\Delta)_i(s) - \Delta_i(s)\big)\, ds$$

implies that

$$x_i^2(t) = x_i^2(0) + x_i^1(t) - x_i^1(0) - \int_0^t \big((a\Delta)_i(s) - \Delta_i(s)\big)\, ds. \tag{A.16}$$

From this and (A.12) it follows that, for $t \le \sigma$,

$$x_i^2(t) \le 1 - 5\delta - 2\delta - 2\delta = 1 - \delta$$

and

$$x_i^2(t) \ge x_i^1(t) + 5\delta - 2\delta \ge 3\delta,$$

or $x_i^2(t) \in [\delta, 1-\delta]$, $t \in [0,\sigma]$. But then the definition of c_1 implies that $g(x_i^1(\sigma)) = c_1 < g(x_i^2(\sigma))$, contradicting $\tau > \sigma$. This establishes (A.14), and (A.15) follows easily from (A.16). □

STEP 4 (Proof of Lemma 5). Assume $\overline{x}$ satisfies $\Delta_i < 0$, $\Delta_j = 0$ and $\Delta_k > 0$. Define

$$\Gamma_0 = \{\overline{x} \in \overline{X}: \Delta_i < 0,\ \Delta_j \neq 0,\ \Delta_k > 0\}$$

and

$$\Gamma_1 = \{\overline{x} \in \overline{X}: \Delta_i < 0,\ g(x_i^1) \neq g(x_i^2),\ \Delta_k > 0\}.$$

Our goal is to prove

$$P^{\overline{x}}\big(\exists\, t^* \in [0,1] \text{ such that } \overline{x}(t^*) \in \Gamma_0\big) > 0.$$

Now by Step 3 there exists $T \in \left[0, \frac{1}{2}\right]$ such that $P^{\overline{x}}(\overline{x}(T) \in \Gamma_1) > 0$. Applying the Markov property, we have

$$\begin{aligned} &P^{\overline{x}}\big(\exists\, t^* \in [0,1] \text{ such that } \overline{x}(t^*) \in \Gamma_0\big) \\ &\quad \geq \int_{\Gamma_1} P^{\overline{x}}\big(\overline{x}(T) \in d\widetilde{x}\big) P^{\widetilde{x}}\big(\exists\, t^* \in \left[0, \tfrac{1}{2}\right] \text{ such that } \overline{x}(t^*) \in \Gamma_0\big). \end{aligned}$$

This integral must be positive since by path continuity and Step 2, for each $\widetilde{x} \in \Gamma_1$,

$$P^{\widetilde{x}}\big(\exists\, t^* \in [0, \tfrac{1}{2}] \text{ such that } \overline{x}(t^*) \in \Gamma_0\big) > 0. \qquad \square$$

Acknowledgments. We thank the Editor for reference [15]. We also thank Rick Durrett for his help with the proof of Lemma 5.

REFERENCES

[1] COX, J. T. and GREVEN, A. (1990). On the long term behavior of some finite particle systems. *Probab. Theory Related Fields* **85** 195–237.

[2] COX, J. T. and GREVEN, A. (1991). On the long term behavior of some finite particle systems: a critical dimension example. In *Random Walks, Brownian Motion, and Interacting Particle Systems* (R. Durrett and H. Kesten, eds.). Birkhäuser, Boston.

[3] DAWSON, D. A. and GREVEN, A. (1993). Multiple time scale analysis of interacting diffusions. *Probab. Theory Related Fields* **95** 467–508.

[4] DAWSON, D. A. and GREVEN, A. (1993). Hierarchical models of interacting diffusions: Multiple time scale phenomena, phase transition and pattern of cluster-formation. *Probab. Theory Related Fields* **96** 435–473.

[5] DEUSCHEL, J.-D. (1994). Algebraic L_2 decay of critical attractive processes on the lattice. *Ann. Probab.* **22** 264–283.

[6] DURRETT, R. (1988). *Lecture Notes on Particle Systems and Percolation*. Wadsworth & Brooks/Cole, Belmont, CA.

[7] GREVEN, A. (1991). A phase transition for the coupled branching process. Part I: Ergodic theory in the range of finite second moments. *Probab. Theory Related Fields* **87** 417–458.

[8] HERBST, I. and PITT, L. (1991). Diffusion equation techniques in stochastic monotonicity and positive correlations. *Probab. Theory Related Fields* **87** 275–312.
[9] KIMURA, M. and OHTA, T. (1971). *Theoretical Aspects of Population Genetics.* Princeton Univ. Press.
[10] KONNO, N. and SHIGA, T. (1988). Stochastic partial differential equations for some measure-valued diffusions. *Probab. Theory Related Fields* **79** 201–225.
[11] LE GALL, J.-F. (1983). Applications du temps local aux équations différentielles stochastiques unidimensionelles. *Séminaire de Probabilités XVII. Lecture Notes in Math.* **986** 15–31. Springer, Berlin.
[12] LIGGETT, T. M. (1985). *Interacting Particle Systems.* Springer, New York.
[13] LIGGETT, T. M. and SPITZER, F. (1981). Ergodic theorems for coupled random walks and other systems with locally interacting components. *Z. Wahrsch. Verw. Gebiete* **56** 443–468.
[14] NEWMAN, C. M. and WRIGHT, A. L. (1981). An invariance principle for certain dependent sequences. *Ann. Probab.* **9** 671–675.
[15] NOTOHARA, M. and SHIGA, T. (1980). Convergence to genetically uniform state in stepping stone models of population genetics. *J. Math. Biol.* **10** 281–294.
[16] ROGERS L. C. G. and WILLIAMS, D. (1987). *Diffusions, Markov Processes, and Martingales* **2.** *Itô Calculus*. Wiley, New York.
[17] SHIGA, T. (1980). An interacting system in population genetics. *J. Math. Kyoto Univ.* **20** 213–243.
[18] SHIGA, T. (1980). An interacting system in population genetics II. *J. Math. Kyoto Univ.* **20** 723–733.
[19] SHIGA, T. (1987). A certain class of infinite-dimensional diffusion processes arising in population genetics. *J. Math. Soc. Japan* **39** 17–25.
[20] SHIGA, T. and SHIMIZU, A. (1980). Infinite-dimensional stochastic differential equations and their applications. *J. Math. Kyoto Univ.* **20** 395–416.
[21] SHIGA, T. and UCHIYAMA, K. (1986). Stationary states and their stability of the stepping stone model involving mutation and selection. *Probab. Theory Related Fields* **73** 87–117.
[22] YAMADA, T. and WATANABE, S. (1971). On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* **11** 155–167.

DEPARTMENT OF MATHEMATICS
SYRACUSE UNIVERSITY
SYRACUSE, NEW YORK 13244

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT GÖTTINGEN
LOTTZESTRASSE 13
W-3400 GÖTTINGEN
GERMANY