The Annals of Probability
1998, Vol. 26, No. 2, 533–561

COUPLING AND ERGODIC THEOREMS FOR FLEMING–VIOT PROCESSES[1]

BY S. N. ETHIER AND THOMAS G. KURTZ

University of Utah and University of Wisconsin–Madison

Fleming–Viot processes are probability-measure-valued diffusion processes that can be used as stochastic models in population genetics. Here we use duality methods to prove ergodic theorems for Fleming–Viot processes, including those with recombination. Coupling methods are also used to establish ergodicity of Fleming–Viot processes, first without and then with selection. A special type of selection known as symmetric overdominance is treated by other methods.

1. Introduction. A Fleming–Viot process is a probability-measure-valued Markov process in which the state of the process is interpreted as the frequency distribution of the "types" of the individuals in a large population. The type of an individual is identified with a point in a locally compact, separable metric space (E, r), and hence the state space for the process is $P(E)$, the set of Borel probability measures on E with the topology of weak convergence. In this paper we are concerned with the ergodic properties of Fleming–Viot processes.

Two forms of ergodic theorems are generally proved for Markov processes. The first form typically states that there exists a unique stationary distribution for the process, that is, a unique probability distribution π on the state space S of the process, such that if the Markov process has this distribution as its initial distribution, then the process is stationary. The uniqueness of the stationary distribution ensures that the stationary process X is ergodic in the sense that the σ-field of invariant events I, that is, the collection of events of the form $\{X(\cdot) \in G\}$, $G \in B(S^{[0,\infty)})$, such that $\{X(\cdot) \in G\} = \{X(t+\cdot) \in G\}$ for all $t \geq 0$, includes only events of probability 0 or 1. The ergodic theorem for stationary processes then implies that

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(X(s))\,ds = \int_S f\,d\pi \quad \text{a.s.} \tag{1.1}$$

for all $f \in L^1(\pi)$.

The second form of ergodic theorem is concerned with the asymptotic behavior of the Markov process when the initial distribution is not the stationary distribution. A typical theorem of this form would give conditions

---

Received January 1996; revised September 1997.

[1]Supported by the National Science Foundation.

AMS 1991 *subject classifications.* Primary 60F99, 60J60; secondary 60G10, 60G57, 92D10.

Key words and phrases. Ergodicity, stationary distribution, duality, coupling, measure-valued diffusion, population genetics, recombination, selection.

under which (1.1) holds for at least bounded continuous f. A slightly weaker form would give conditions under which

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t E[f(X(s))]\,ds = \int_S f\,d\pi, \qquad f\in \overline{C}(S). \tag{1.2}$$

[If (1.2) holds, then $\lim_{t\to\infty} E[f(X(t))] = \int_S f\,d\pi$ usually holds as well for all $f \in \overline{C}(S)$.]

If (1.1) holds for all bounded continuous f, then (1.2) holds. If (1.2) holds for all initial distributions, then π is the unique stationary distribution for the process. If S is compact and X is a Feller process [that is, the semigroup corresponding to X maps $C(S)$ into $C(S)$], then uniqueness of stationary distributions implies (1.2) for all initial distributions.

There are a variety of approaches to proving ergodic theorems. Any stationary distribution π for a Markov process with generator A must satisfy

$$\int_S Af\,d\pi = 0, \qquad f\in D(A), \tag{1.3}$$

which, using adjoint operator notation, says

$$A^*\pi = 0. \tag{1.4}$$

One approach to the first type of ergodic theorem is to prove uniqueness of solutions of this *adjoint equation*. The proof of uniqueness is easy, for example, in the case of irreducible finite-state Markov jump processes. Note that if an explicit representation of A^* is used (e.g., as a differential operator in the case of a diffusion process), then one must verify that any solution π of (1.3) is in the domain of the explicit representation.

A second approach is through *renewal arguments*, in particular the generalized renewal arguments of Athreya and Ney (1978) and Nummelin (1978, 1984). When these arguments apply, one can conclude that (1.1) holds for all initial distributions and all $f\in L^1(\pi)$. *Coupling methods* [Doeblin (1940); Griffeath (1976, 1978)] provide a third approach. These methods provide one of the basic approaches for particle systems [Liggett (1985)] and can be applied to diffusions [Lindvall (1983); Lindvall and Rogers (1986)]. Coupling methods also give a simple probabilistic proof of the renewal theorem [Athreya, McDonald and Ney (1978)]. In their strongest form, coupling methods also give (1.1) for all initial distributions and all $f\in L^1(\pi)$. *Duality arguments* can be used to verify (1.2). Duality methods were developed in the context of particle systems [Vasershtein (1969)] and have been applied to prove ergodic theorems for Fleming–Viot processes [Shiga (1982); Dawson and Hochberg (1982); Ethier and Griffiths (1990)].

Our primary concern in this paper is to develop coupling methods for Fleming–Viot processes. This development is carried out in Sections 3 and 5. We also review duality methods in Section 2 and use them to extend the previously known ergodicity results to models with recombination. [For certain two-locus models, this extension already has been carried out by Ethier and Griffiths (1990)]. The case of symmetric overdominance is considered in

Section 4, where it is shown that the process formed by the sequence of descending order statistics of the sizes of the atoms is an infinite-dimensional ergodic diffusion.

A Fleming–Viot process can be characterized as the unique solution of the martingale problem for a generator A defined as follows. For $1 \le i < j \le m$, define the *sampling operators* $\Phi_{ij}^{(m)}$: $B(E^m) \mapsto B(E^{m-1})$ by letting $\Phi_{ij}^{(m)} f$ be the function obtained from f by replacing x_j by x_i and renumbering the variables if necessary [e.g., for $f(x_1, x_2, x_3) \in B(E^3)$, $\Phi_{12}^{(3)} f(x_1, x_2) = f(x_1, x_1, x_2)$, $\Phi_{13}^{(3)} f(x_1, x_2) = f(x_1, x_2, x_1)$ and $\Phi_{23}^{(3)} f(x_1, x_2) = f(x_1, x_2, x_2)$].

Let B be the generator of a Feller semigroup $\{T(t)\}$ on $\hat{C}(E)$. B is called the *mutation operator* for the process and $\{T(t)\}$ is given by a transition function $P(t, x, \Gamma)$, that is,

$$T(t) f(x) = \int_E f(y) P(t, x, dy), \tag{1.5}$$

and in fact (1.5) extends $\{T(t)\}$ to all of $B(E)$. For $m \ge 1$, define $\{T_m(t)\}$ on $B(E^m)$ by

$$\begin{aligned} &T_m(t) f(x_1, \dots, x_m) \\ &\qquad = \int_E \cdots \int_E f(y_1, \dots, y_m) P(t, x_1, dy_1) \cdots P(t, x_m, dy_m) \end{aligned} \tag{1.6}$$

and let $B^{(m)}$ denote the generator of $\{T_m(t)\}$ on $B(E^m)$.

Let $\eta(x_1, x_2, \Gamma)$ be a transition function from $E \times E$ to E. For $i = 1, \dots, m$, define the *recombination operators* $H_i^{(m)}$: $B(E^m) \mapsto B(E^{m+1})$ by

$$\begin{aligned} &H_i^{(m)} f(x_1, \dots, x_{m+1}) \\ &\qquad = \int_E f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_m) \eta(x_i, x_{m+1}, dz). \end{aligned} \tag{1.7}$$

We denote by $\alpha \ge 0$ the *recombination intensity*. Suppose $E = E_1 \times \cdots \times E_N$ and for $I \subset \{1, \dots, N\}$, define h_I: $E \times E \mapsto E$ by letting $h_I(x, y)$ be the element $z \in E$ such that $z_i = x_i$ for $i \in I$ and $z_i = y_i$ for $i \notin I$. If $\eta(x, y, \cdot) = \sum p_I \delta_{h_I(x, y)}$ for some probability distribution $\{p_I\}$ on the set of subsets I of $\{1, \dots, N\}$, then the resulting recombination will be called *physical recombination*.

For σ in $B_{\mathrm{sym}}(E \times E)$, the space of symmetric functions in $B(E \times E)$, set $\bar{\sigma} = \sup_{x, y, z} |\sigma(x, y) - \sigma(y, z)|$ and, for $i = 1, \dots, m$, define the *selection operators* $K_i^{(m)}$: $B(E^m) \mapsto B(E^{m+2})$ by

$$\begin{aligned} &K_i^{(m)} f(x_1, \dots, x_{m+2}) \\ &\qquad = \frac{\bar{\sigma} + \sigma(x_i, x_{m+1}) - \sigma(x_{m+1}, x_{m+2})}{2\bar{\sigma}} f(x_1, \dots, x_m), \end{aligned} \tag{1.8}$$

where $0/0 = 0$. Note that the factor multiplying the function f in (1.8) is nonnegative and bounded by 1. The function σ is called the *selection intensity function*.

For $m \geq 1$ and $f \in B(E^m)$, define $F_f \in B(P(E))$ by $F_f(\mu) = \langle f, \mu^m \rangle \equiv \int_{E^m} f \, d\mu^m$, where μ^m denotes the m-fold product measure of μ, and let $L^m(E) \subset B(P(E))$ be the collection of functions of this form. Note that $L^m(E)$ is a subspace of $B(P(E))$ closed under bounded pointwise convergence. For $f \in D(B^{(m)})$, define AF_f by

$$\begin{aligned} AF_f(\mu) = {} & \sum_{1 \leq i < j \leq m} \left(\langle \Phi_{ij}^{(m)} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle \right) + \langle B^{(m)} f, \mu^m \rangle \\ & + \alpha \sum_{i=1}^{m} \left(\langle H_i^{(m)} f, \mu^{m+1} \rangle - \langle f, \mu^m \rangle \right) \\ & + 2\bar{\sigma} \sum_{i=1}^{m} \left(\langle K_i^{(m)} f, \mu^{m+2} \rangle - \langle f, \mu^m \rangle \right) + \bar{\sigma} m \langle f, \mu^m \rangle. \end{aligned} \tag{1.9}$$

For a derivation and fuller explanation of this generator, see Ethier and Kurtz (1987, 1993).

THEOREM 1.1. *Under the above assumptions on E, B, η, α and σ, there is for each initial distribution in $P(P(E))$ at most one solution of the martingale problem for A. If in addition we assume that*

$$H_1^{(1)}: \bar{C}(E) \mapsto \bar{C}(E^2), \tag{1.10}$$

then there is for each initial distribution exactly one solution. If the martingale problem is well posed, then every solution has a version with continuous sample paths in $P(E)$.

The theorem is a minor modification of a result in Ethier and Kurtz (1987). The assumption that B generates a Feller semigroup on $\hat{C}(E)$ is used in the proof of uniqueness, as well as in the definition of $B^{(m)}$. Though the latter operator can be avoided in the formulation of the Fleming–Viot process, it plays a crucial role in Theorem 5.5 below, for example.

The structure of the generator suggests the existence of a dual process with values in a space of functions. Note that the first, third and fourth terms look like generators of jump processes, where a "jump" would be from $f \in B(E^m)$ to $\Phi_{ij}^{(m)} f$, $H_i^{(m)} f$ or $K_i^{(m)} f$. We review the construction of the function-valued dual process in the next section and apply it to the proof of an ergodic theorem.

2. A function-valued dual. Let E, B, η, α and σ be as above [but we do not assume (1.10)]. Let M be a jump Markov process in $\mathbf{N}$, the set of positive integers, with transition intensities $q_{m,m-1} = m(m-1)/2$, $q_{m,m+1} = \alpha m$, $q_{m,m+2} = 2\bar{\sigma} m$ and $q_{m,l} = 0$ otherwise (unless of course $l = m$). Let $0 = \tau_0 < \tau_1 < \cdots$ be the sequence of jump times of M and let $\Gamma_1, \Gamma_2, \ldots$ be a sequence of random operators that are conditionally independent given M and satisfy

$$P\{\Gamma_k = \Phi_{ij}^{(m)} | M\} = \binom{m}{2}^{-1} I_{\{M(\tau_k -)=m,\ M(\tau_k)=m-1\}}, \qquad 1 \le i < j \le m, \tag{2.1}$$

$$P\{\Gamma_k = H_i^{(m)} | M\} = m^{-1} I_{\{M(\tau_k -)=m,\ M(\tau_k)=m+1\}}, \qquad 1 \le i \le m \tag{2.2}$$

and

$$P\{\Gamma_k = K_i^{(m)} | M\} = m^{-1} I_{\{M(\tau_k -)=m,\ M(\tau_k)=m+2\}}, \qquad 1 \le i \le m. \tag{2.3}$$

Given $Y(0) \in B(E^{M(0)})$, define the $\bigcup_{m\ge 1} B(E^m)$-valued process Y by

$$\begin{aligned} Y(t) = {} & T_{M(\tau_k)}(t-\tau_k)\Gamma_k T_{M(\tau_{k-1})}(\tau_k - \tau_{k-1})\Gamma_{k-1} \\ & \cdots \Gamma_1 T_{M(0)}(\tau_1) Y(0), \qquad \tau_k \le t < \tau_{k+1},\ k \ge 0. \end{aligned} \tag{2.4}$$

If Z is a solution of the martingale problem for A and is independent of Y, then, for each $m \ge 1$ and $f \in B(E^m)$,

$$E\left[\left\langle f, Z(t)^m \right\rangle\right] = E\left[\left\langle Y(t), Z(0)^{M(t)}\right\rangle \exp\left\{\bar\sigma \int_0^t M(s)\, ds\right\}\right], \qquad t \ge 0, \tag{2.5}$$

where $M(0) = m$ and $Y(0) = f$. The validity of this identity is proved in Ethier and Kurtz (1987).

The following lemma provides a useful estimate of the expectation on the right-hand side of (2.5) (showing in particular that it is finite).

LEMMA 2.1. *Let M be as above. Then there exists a function F: $\mathbf{N} \mapsto (0, \infty)$ and a positive constant L such that*

$$E\left[\exp\left\{\bar\sigma \int_0^t M(s)\, ds\right\} \middle| M(0) = m\right] \le F(m) e^{Lt}, \qquad m \ge 1,\ t \ge 0. \tag{2.6}$$

PROOF. Denote by Q the operator associated with (q_{ij}), extended to the space of possibly unbounded functions on $\mathbf{N}$, and define F on $\mathbf{N}$ by $F(m) = (m!)^\beta$, where $0 < \beta < \frac{1}{2}$. Then there exists a positive constant L such that

$$\begin{aligned} & QF(m) + \bar\sigma m F(m) \\ & \quad = \tfrac{1}{2} m(m-1)(F(m-1) - F(m)) + \alpha m (F(m+1) - F(m)) \\ & \qquad + 2\bar\sigma m (F(m+2) - F(m)) + \bar\sigma m F(m) \\ & \quad = \tfrac{1}{2} m(m-1)\big(((m-1)!)^\beta - (m!)^\beta\big) \\ & \qquad + \alpha m \big(((m+1)!)^\beta - (m!)^\beta\big) \\ & \qquad + 2\bar\sigma m ((m+2)!)^\beta - \bar\sigma m (m!)^\beta \\ & \quad = m(m!)^\beta \Big[-\tfrac{1}{2}(m-1)(1 - m^{-\beta}) + \alpha\big((m+1)^\beta - 1\big) \\ & \qquad\qquad + 2\bar\sigma (m+1)^\beta (m+2)^\beta - \bar\sigma \Big] \\ & \quad \le L \end{aligned} \tag{2.7}$$

for all $m \ge 1$, where $F(0) = 1$ and the existence of L follows from the fact that for m sufficiently large the term to the left of the inequality is negative.

Consequently, the optional sampling theorem implies that for $\tau = \inf\{t \geq 0: M(t) \geq z\}$ and $M(0) = m$,

$$
\begin{aligned}
&E\left[\exp\left(\bar{\sigma}\int_0^{t\wedge\tau} M(s)\,ds\right)\right] \\
&\qquad \leq E\left[F(M(t\wedge\tau))\exp\left(\bar{\sigma}\int_0^{t\wedge\tau} M(s)\,ds\right)\right] \\
&\qquad = F(m) + E\left[\int_0^{t\wedge\tau}(QF(M(s)) + \bar{\sigma}M(s)F(M(s)))\right. \\
&\qquad\qquad\qquad\qquad \left.\times\exp\left(\bar{\sigma}\int_0^{s} M(r)\,dr\right)ds\right] \\
&\qquad \leq F(m) + LE\left[\int_0^{t\wedge\tau}\exp\left(\bar{\sigma}\int_0^{s} M(r)\,dr\right)ds\right] \\
&\qquad \leq F(m) + L\int_0^{t}E\left[\exp\left(\bar{\sigma}\int_0^{s\wedge\tau} M(r)\,dr\right)\right]ds, \qquad t \geq 0,
\end{aligned}
\tag{2.8}
$$

and the lemma follows by Gronwall's inequality. □

The duality identity (2.5) determines the one-dimensional distributions of a solution of the martingale problem for A in terms of the initial distribution. Uniqueness of the one-dimensional distributions in turn implies uniqueness of the finite-dimensional distributions and the Markov property of the solution. [See Ethier and Kurtz (1986), Theorem 4.4.2.] Duality also gives a very general ergodic theorem in the neutral case without recombination.

THEOREM 2.2. *Let $\alpha = 0$ and $\sigma = 0$, and let Z be a solution of the martingale problem for A. Suppose that there exists $\pi \in P(E)$ such that $\lim_{t\to\infty} T(t)f(x) = \langle f, \pi\rangle$ for all $f \in \bar{C}(E)$ [resp., $f \in B(E)$] and $x \in E$. Then, for each $m \geq 1$ and $f \in \bar{C}(E^m)$ [resp., $f \in B(E^m)$],*

$$
\lim_{t\to\infty} E\left[\langle f, Z(t)^m\rangle\right] = E\left[\langle Y(\tau_{m-1}), \pi\rangle\right],
\tag{2.9}
$$

where $M(0) = m$ and $Y(0) = f$. If there exists a stationary distribution Π, then it is unique and

$$
\lim_{t\to\infty} E[F(Z(t))] = \int_{P(E)} F(\mu)\Pi(d\mu)
\tag{2.10}
$$

for all $F \in \bar{C}(P(E))$.

PROOF. Note that in the present case, M is a pure death process, so if $M(0) = m \geq 1$, then after $m - 1$ jumps, M absorbs at 1. Therefore

$$
\begin{aligned}
&\lim_{t\to\infty} E\left[\langle f, Z(t)^m\rangle\right] \\
&\qquad = \lim_{t\to\infty} E\left[\langle T(t-\tau_{m-1})Y(\tau_{m-1}), Z(0)\rangle; \tau_{m-1} \leq t\right],
\end{aligned}
\tag{2.11}
$$

which by the ergodicity assumption on $\{T(t)\}$ gives (2.9) in each case. Since the right-hand side of (2.9) does not depend on $Z(0)$ and since the collection of functions of the form $F_f(\mu) = \langle f, \mu^m \rangle$, where $m \geq 1$ and $f \in \overline{C}(E^m)$, is convergence determining, if Π is a stationary distribution, (2.10) follows. $\square$

We next include recombination, but keep $\sigma = 0$. Observe that $\sup_{x \in E^{M(t)}} Y(t, x)$ is nonincreasing in t and $\inf_{x \in E^{M(t)}} Y(t, x)$ is nondecreasing. Ergodicity will follow if there exists a stationary distribution (which is assured if E is compact) and the difference of these quantities goes to zero as $t \to \infty$ for an appropriate class of functions f.

THEOREM 2.3. *Let $\sigma = 0$ and let Z be a solution of the martingale problem for A. Suppose that there exists a Borel function χ: $[0, \infty) \mapsto [0, 1]$ such that*

$$\sup_{x, y \in E} \sup_{\Gamma \in B(E)} |P(t, x, \Gamma) - P(t, y, \Gamma)| \leq \chi(t), \qquad t \geq 0, \tag{2.12}$$

and $\chi(t) \to 0$ as $t \to \infty$. Then, for each $m \geq 1$ and $f \in B(E^m)$, if $M(0) = m$ and $Y(0) = f$, we have that $\lim_{t \to \infty} Y(t, x)$ exists a.s. uniformly in x and is independent of x, and

$$\lim_{t \to \infty} E\Big[\big\langle f, Z(t)^m \big\rangle\Big] = E\Big[\lim_{t \to \infty} Y(t)\Big]. \tag{2.13}$$

If there exists a stationary distribution Π, then it is unique and (2.10) *holds for all $F \in \overline{C}(P(E))$.*

PROOF. Let $\gamma_1 = \inf\{t \geq 0\colon\ M(t) = 1\}$, $\delta_k = \inf\{t > \gamma_k\colon\ M(t) = 2\}$ and $\gamma_{k+1} = \inf\{t > \delta_k\colon M(t) = 1\}$. Then

$$\begin{aligned} &E\Big[\sup_x Y(\gamma_{k+1}, x) - \inf_x Y(\gamma_{k+1}, x) \Big| F_{\gamma_k}\Big] \\ &\quad \leq E\Big[\sup_x Y(\delta_k - , x) - \inf_x Y(\delta_k - , x) \Big| F_{\gamma_k}\Big] \\ &\quad = \int_0^\infty \alpha e^{-\alpha t}\Big(\sup_x T(t) Y(\gamma_k, x) - \inf_x T(t) Y(\gamma_k, x)\Big)\, dt \\ &\quad \leq \overline{\chi}\Big(\sup_x Y(\gamma_k, x) - \inf_x Y(\gamma_k, x)\Big), \end{aligned} \tag{2.14}$$

where

$$\overline{\chi} = \int_0^\infty \alpha e^{-\alpha t} \chi(t)\, dt < 1. \tag{2.15}$$

It follows that $\lim_{t \to \infty}(\sup_x Y(t, x) - \inf_x Y(t, x)) = 0$ a.s. and hence that $Y(t, x)$ converges a.s., uniformly in x, to a limit that is independent of x, which by (2.5) implies (2.13). The second conclusion follows as in Theorem 2.2. $\square$

We can weaken the ergodicity assumption on $\{T(t)\}$ by imposing more structure on the recombination. The fact that we need some additional structure is clear from the observation that if

$$\eta(x, y, \Gamma) = \tfrac{1}{2}(\tilde{\eta}(x, \Gamma) + \tilde{\eta}(y, \Gamma)) \tag{2.16}$$

for some one-step transition function $\tilde{\eta}$, then the inclusion of recombination is equivalent to changing the mutation operator to

$$\tilde{B}f(x) = Bf(x) + \alpha \int_E (f(z) - f(x))\tilde{\eta}(x, dz). \tag{2.17}$$

It is simple to construct examples in which B is ergodic and $\tilde{B}$ is not.

Let β: $E \times E \mapsto [0, \infty)$ be Borel measurable, and for $m \geq 1$, $f \in B(E^m)$ and $s \in [0, \infty]$, define

$$\omega_f(s) = \sup\Big\{|f(x_1, \dots, x_m) - f(y_1, \dots, y_m)|: \max_{1 \leq i \leq m} \beta(x_i, y_i) \leq s\Big\}. \tag{2.18}$$

Note that

$$\omega_{\Phi_{ij}^{(m)} f}(s) \leq \omega_f(s), \qquad 1 \leq i < j \leq m,\ s \in [0, \infty]. \tag{2.19}$$

LEMMA 2.4. *Suppose that for each $m \geq 1$ and $f \in B(E^m)$,*

$$\omega_{H_i^{(m)} f}(s) \leq \omega_f(s), \qquad 1 \leq i \leq m,\ s \in [0, \infty] \tag{2.20}$$

and

$$\omega_{T_m(t) f}(s) \leq \int_{[0,\infty]} \omega_f(u)\gamma_m(t, s, du), \qquad (t, s) \in [0, \infty) \times [0, \infty], \tag{2.21}$$

where $\gamma_m(t, s, du)$ is a transition function from $[0, \infty) \times [0, \infty]$ to $[0, \infty]$. Then, for all $(t, s) \in [0, \infty) \times [0, \infty]$,

$$\begin{aligned}\omega_{Y(t)}(s) \leq \int_{[0,\infty]} \cdots \int_{[0,\infty]} &\omega_{Y(0)}(u_0)\gamma_{M(0)}(\tau_1, u_1, du_0) \\ &\gamma_{M(\tau_1)}(\tau_2 - \tau_1, u_2, du_1) \cdots \gamma_{M(\tau_k)}(t - \tau_k, s, du_k),\end{aligned} \tag{2.22}$$

where $k \geq 0$ is such that $\tau_k \leq t < \tau_{k+1}$.

REMARK 2.5. (a) Condition (2.20) will hold if for each $x, x', y, y' \in E$, one can construct random variables ξ and ζ with distributions $\eta(x, x', \cdot)$ and $\eta(y, y', \cdot)$ such that

$$P\{\beta(\xi, \zeta) \leq \beta(x, y) \vee \beta(x', y')\} = 1. \tag{2.23}$$

For example, if $E = E_1 \times \cdots \times E_N$, $\beta(x, y) = \max_{1 \leq i \leq N} \beta_i(x_i, y_i)$ and η corresponds to physical recombination, then the construction of the desired ξ and ζ is immediate, and (2.20) holds.

(b) If $Bf(x) = \frac{1}{2}\theta \int_E (f(y) - f(x))\nu(dy)$, where $\theta > 0$, then for any β, $\gamma_m(t, s, \cdot)$ can be taken to be $(1 - e^{-\theta t/2})^m \delta_0 + (1 - (1 - e^{-\theta t/2})^m)\delta_s$.

(c) Similarly, under the assumptions of Theorem 2.3, for any β, $\gamma_m(t, s, \cdot)$ can be taken to be $(1 - \chi(t))^m \delta_0 + (1 - (1 - \chi(t))^m)\delta_\infty$ for $s > 0$ and δ_0 for $s = 0$.

(d) If $E = \mathbf{R}$, $Bf(x) = af''(x) - bxf'(x)$, where $a, b > 0$, and $\beta(x, y) = |x - y|$, then $\omega_{T_m(t)f}(s) \le \omega_f(se^{-bt})$ and hence $\gamma_m(t, s, \cdot)$ can be taken to be $\delta_{s\exp(-bt)}$.

(e) If $E = [0,1]^\infty$, $Bf(x) = \frac{1}{2}\theta\int_0^1 (f(z, x_0, x_1, \dots) - f(x_0, x_1, \dots))\,dz$ and $\beta(x, y) = 1/(1 + \min\{k \ge 0\colon x_k \ne y_k\})$, then, for each $m \ge 1$ and $f \in B(E^m)$, we have $\omega_{T(t)f}(s) \le E[\omega_f(1/(Z_m(t) + s^{-1}))]$, where $Z_m(t)$ is the minimum of m independent, mean $\frac{1}{2}\theta t$, Poisson random variables.

(f) If $E = [0,1]^\infty \times \{(a_0, a_1, \dots)\colon 0 \le a_0 \le a_1 \le \cdots\} \subset [0,1]^\infty \times [0,\infty]^\infty$,

$$
\begin{aligned}
Bf(x, a) = \frac{1}{2}\theta\int_0^1 \big(f\big((z, x_0, x_1, \dots), (0, a_0, a_1, \dots)\big) \\
-f\big((x_0, x_1, \dots), (a_0, a_1, \dots)\big)\big)\,dz \\
+ \sum_{k=0}^{\infty} \frac{\partial}{\partial a_k} f\big((x_0, x_1, \dots), (a_0, a_1, \dots)\big)
\end{aligned}
\tag{2.24}
$$

and $\beta((x, a), (y, b)) = 1/(1 + \min\{a_k\colon (x_k, a_k) \ne (y_k, b_k)\})$, then $\omega_{T(t)f}(s) \le \omega_f(1/(t + s^{-1}))$. If the recombinant of (x, a) and (y, b) is obtained by selecting all (x_k, a_k) with $x_k > U$ and all (y_k, b_k) with $y_k \le U$, for a random variable U, then (2.23) and hence (2.20) will be satisfied. Note that in this model, x_k can be interpreted as the location on the chromosome of a mutation and a_k as the time since that mutant first appeared in the population (i.e., its age). Recombination then consists of breaking the chromosomes at U and combining the part to the left of U from one of the chromosomes with the part to the right of U from the other.

PROOF OF LEMMA 2.4. The inequality follows easily from (2.4) by iteratively applying (2.21) followed by (2.19) or (2.20), depending on Γ_k. $\square$

THEOREM 2.6. *Suppose* (2.20) *and* (2.21) *hold with* $\gamma_1, \gamma_2, \dots$ *as in Lemma* 2.4. *Define* $C_\beta(E^m) = \{f \in B(E^m)\colon \lim_{u\to 0} \omega_f(u) = 0\}$ *and suppose that for each* $m \ge 1$, $C_\beta(E^m)$ *is separating for* $P(E^m)$. *For* $x \in E^m$ *and* $s > 0$, *define* $U_m(x, s) = \{y \in E^m\colon \max_{1 \le i \le m} \beta(x_i, y_i) < s\}$ *and suppose for each compact set* $K \subset E^m$, *there exist* $0 < s < \infty$ *and* $x \in E^m$ *such that* $K \subset U_m(x, s)$.

Assume that the martingale problem for A *is well posed. If for each* $\delta > 0$ *and* $s > 0$,

$$
\lim_{t\to\infty} E\left[\int_{[0,\infty]} \cdots \int_{[0,\infty]} I_{[\delta,\infty]}(u_0)\gamma_{M(0)}(\tau_1, u_1, du_0)\cdots \gamma_{M(\tau_k)}(t - \tau_k, s, du_k)\right] = 0,
\tag{2.25}
$$

where $k = \max\{i\colon \tau_i \le t\}$, *then there is at most one stationary distribution for the Fleming–Viot process with generator* A.

REMARK 2.7. (a) The assumption that $C_\beta(E^m)$ is separating for each m implies $\{F_f(\mu) \equiv \langle f, \mu^m \rangle: f \in C_\beta(E^m), m = 1, 2, \dots\}$ is separating for $P(P(E))$.

(b) If for each $m \ge 1$, $\gamma_m(t, s, [0, s]) = 1$ for all $t, s \ge 0$ and

$$\int_{[0,\infty]} \varphi(u) \gamma_m(t, s, du) \le g_m(t, s') \varphi(s'), \qquad t \ge 0, 0 \le s \le s', \tag{2.26}$$

where φ is nondecreasing on $[0, \infty)$ with $\varphi(u) > 0$ for $u > 0$, $0 \le g_m(t, s) \le 1$, and $\lim_{t \to \infty} g_m(t, s) = 0$ for each $s > 0$, then the expectation on the left-hand side of (2.25) is bounded by

$$\frac{\varphi(s)}{\varphi(\delta)} E\left[\min_{\tau_i \le t} g_{M(\tau_{i-1})}(\tau_i - \tau_{i-1}, s) \right] \tag{2.27}$$

and the limit in (2.25) holds. If we take $\varphi(u) \equiv u$, then the examples in Remarks 2.5(b), (d), (e) and (f) satisfy this condition.

(c) Assume that there exist a nondecreasing function φ on $[0, \infty)$ with $\varphi(u) > 0$ for $u > 0$ and a sequence of Borel functions $h_m(t)$ such that

$$\int_{[0,\infty]} \varphi(u) \gamma_m(t, s, du) \le \exp(h_m(t)) \varphi(s), \qquad t, s \ge 0 \tag{2.28}$$

and

$$\sum_{m=1}^{\infty} \pi_m \int_0^\infty h_m(t) \lambda_m \exp(-\lambda_m t)\, dt < 0, \tag{2.29}$$

where $\lambda_m = \frac{1}{2} m(m - 1 + 2\alpha)$ (the jump intensity for M) and $\{\pi_m\}$ is the stationary distribution for M. [Implicitly, we are assuming that the integral in (2.29) exists and the series converges.] Then, for $\tau_k \le t < \tau_{k+1}$,

$$\begin{aligned} &\int_{[0,\infty]} \cdots \int_{[0,\infty]} \varphi(u_0) \gamma_{M(0)}(\tau_1, u_1, du_0) \cdots \gamma_{M(\tau_k)}(t - \tau_k, s, du_k) \\ &\qquad \le \exp\left\{ h_{M(\tau_k)}(t - \tau_k) + \sum_{i=1}^k h_{M(\tau_{i-1})}(\tau_i - \tau_{i-1}) \right\} \varphi(s) \end{aligned} \tag{2.30}$$

and the ergodic theorem for M implies that

$$\begin{aligned} &\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^k h_{M(\tau_{i-1})}(\tau_i - \tau_{i-1}) \\ &\qquad = \sum_{m=1}^\infty \pi_m \int_0^\infty h_m(t) \lambda_m \exp(-\lambda_m t)\, dt < 0, \end{aligned} \tag{2.31}$$

so the right-hand side of (2.30) converges to zero, implying (2.25).

PROOF OF THEOREM 2.6. By (2.22), for each $\delta > 0$,

$$\begin{aligned} \omega_{Y(t)}(s) \le \omega_{Y(0)}(\delta) + \omega_{Y(0)}(\infty) &\int_{[0,\infty]} \cdots \int_{[0,\infty]} I_{[\delta,\infty]}(u_0) \\ &\gamma_{M(0)}(\tau_1, u_1, du_0) \cdots \gamma_{M(\tau_k)}(t - \tau_k, s, du_k), \end{aligned} \tag{2.32}$$

so by (2.25), if $Y(0) \in C_\beta(E^m)$ for some $m \geq 1$, then

$$\lim_{t\to\infty} E\left[\omega_{Y(t)}(s)\right] = 0. \tag{2.33}$$

Suppose Z_1 and Z_2 are stationary Fleming–Viot processes. Let $K \subset E$ be compact, select $x^{(m)} \in E^m$ and $s_m > 0$ for each $m \geq 1$ such that $K^m \subset U_m(x^{(m)}, s_m)$ and define $H_m = E^m - K^m$. (Without loss of generality we can assume that s_m is increasing in m.) Now fix $m \geq 1$ and $f \in C_\beta(E^m)$. Then

$$\begin{aligned}
&\left| E\left[\left\langle f, Z_1(0)^m\right\rangle\right] - E\left[\left\langle f, Z_2(0)^m\right\rangle\right]\right| \\
&\quad = \left| E\left[\left\langle f, Z_1(t)^m\right\rangle\right] - E\left[\left\langle f, Z_2(t)^m\right\rangle\right]\right| \\
&\quad = \left| E\left[\left\langle Y(t), Z_1(0)^{M(t)}\right\rangle\right] - E\left[\left\langle Y(t), Z_2(0)^{M(t)}\right\rangle\right]\right| \\
&\quad \leq 2E\left[\omega_{Y(t)}(s_{M(t)})\right] + \|f\|\Big(E\left[\left\langle I_{H_{M(t)}}, Z_1(0)^{M(t)}\right\rangle\right] \\
&\qquad\qquad + E\left[\left\langle I_{H_{M(t)}}, Z_2(0)^{M(t)}\right\rangle\right]\Big) \\
&\quad \leq 2E\left[\omega_{Y(t)}(s_{m_0})\right] + \|f\|\Big(E\left[\left\langle I_{H_{m_0}}, Z_1(0)^{m_0}\right\rangle\right] \\
&\qquad\qquad + E\left[\left\langle I_{H_{m_0}}, Z_2(0)^{m_0}\right\rangle\right]\Big) \\
&\qquad + 6\|f\|P\{M(t) > m_0\},
\end{aligned} \tag{2.34}$$

and the right-hand side of (2.34) can be made arbitrarily small by taking m_0 large, K large and t large, which proves the uniqueness of stationary distributions. □

3. Coupling neutral Fleming–Viot processes. Let B be the generator for a Markov process in a locally compact, separable metric space (E, r). Let $(\tilde{E}, \tilde{r})$ be another locally compact, separable metric space and let ρ_1: $\tilde{E} \mapsto E$, ρ_2: $\tilde{E} \mapsto E$ and ρ: $E \times E \mapsto \tilde{E}$ be Borel measurable mappings such that $(\rho_i \circ \rho)(x_1, x_2) = x_i$ for $i = 1, 2$. An operator $\tilde{B}$ on $B(\tilde{E})$ and the mappings ρ_1, ρ_2 and ρ determine a *Markov coupling* for B if the martingale problem for $\tilde{B}$ is well posed and, for each $f \in D(B)$, $f \circ \rho_i \in D(\tilde{B})$ and $\tilde{B}(f \circ \rho_i) = (Bf) \circ \rho_i$ for $i = 1, 2$. In particular, if X is a solution of the martingale problem for $\tilde{B}$, then $X_1 \equiv \rho_1 \circ X$ and $X_2 \equiv \rho_2 \circ X$ are solutions of the martingale problem for B. We say that $\tilde{B}$, ρ_1, ρ_2 and ρ determine a *successful* Markov coupling for B if, in addition, for each solution X of the $D_{\tilde{E}}[0, \infty)$ martingale problem for $\tilde{B}$, there exists a random time τ such that $P\{\tau < \infty,\ X_1(\tau + t) = X_2(\tau + t) \text{ for all } t \geq 0\} = 1$.

Typically, $\tilde{E} = E \times E$, ρ_1 and ρ_2 are the projections on $E \times E$ [i.e., $\rho_i(x_1, x_2) = x_i$ for $i = 1, 2$] and ρ is the identity operator. For example, given $\nu \in P(E)$ and $\theta > 0$, define

$$Bf(x) = \tfrac{1}{2}\theta \int_E (f(y) - f(x))\,\nu(dy) \tag{3.1}$$

and

$$\tilde{B}f(x_1, x_2) = \tfrac{1}{2}\theta\int_E (f(y, y) - f(x_1, x_2))\nu(dy). \tag{3.2}$$

Then $\tilde{B}$, ρ_1, ρ_2, and ρ determine a successful Markov coupling for B, with τ being the hitting time of the diagonal of $E \times E$ by the solution X of the $D_{E\times E}[0,\infty)$ martingale problem for $\tilde{B}$.

For an example in which $\tilde{E}$ is not $E \times E$, see the paragraph following the next one.

Returning to the general case, suppose $\tilde{B}$, ρ_1, ρ_2 and ρ determine a Markov coupling for B and let $D = \{x \in \tilde{E}: \rho_1(x) = \rho_2(x)\}$. Given a transition function η from $E \times E$ to E, define the transition function $\tilde{\eta}$ from $\tilde{E} \times \tilde{E}$ to $\tilde{E}$ by

$$\begin{aligned}\tilde{\eta}(x, y, G) = I_{D\times D}(x, y)\int_E I_G(\rho(z, z))\eta(\rho_1(x), \rho_1(y), dz)\\ + I_{(D\times D)^c}(x, y)\int_{E\times E} I_G(\rho(x_1, x_2))\\ \eta(\rho_1(x), \rho_1(y), dx_1)\eta(\rho_2(x), \rho_2(y), dx_2).\end{aligned} \tag{3.3}$$

Note that $\rho(z, z) \in D$, so if x and y are in D, then $\tilde{\eta}(x, y, D) = 1$. Also, for all $\Gamma \in B(E)$ and $x, y \in \tilde{E}$, we have $\tilde{\eta}(x, y, \rho_i^{-1}(\Gamma)) = \eta(\rho_i(x), \rho_i(y), \Gamma)$, so that if $m \ge 1$, $g \in B(E^m)$, $i \in \{1, 2\}$ and $f(x_1, \dots, x_m) \equiv g(\rho_i(x_1), \dots, \rho_i(x_m))$, then

$$\tilde{H}_j^{(m)}f(x_1, \dots, x_{m+1}) = H_j^{(m)}g(\rho_i(x_1), \dots, \rho_i(x_{m+1})) \tag{3.4}$$

for all $x_1, \dots, x_{m+1} \in \tilde{E}$ and $j = 1, \dots, m$.

Define $\hat{\rho}_i$: $P(\tilde{E}) \mapsto P(E)$ for $i = 1, 2$ by $\hat{\rho}_i(\mu) = \mu\rho_i^{-1}$ and define $\hat{\rho}$: $P(E) \times P(E) \mapsto P(\tilde{E})$ by $\hat{\rho}(\mu_1, \mu_2) = (\mu_1 \times \mu_2)\rho^{-1}$. Let A be the generator for the neutral ($\sigma \equiv 0$) Fleming–Viot process in $P(E)$ with mutation operator B and recombination given by α and η, and let $\tilde{A}$ be the generator for the neutral Fleming–Viot process in $P(\tilde{E})$ with mutation operator $\tilde{B}$ and recombination given by α and $\tilde{\eta}$. [We are implicitly assuming that B generates a Feller semigroup on $\hat{C}(E)$ and that $\tilde{B}$ generates a Feller semigroup on $\hat{C}(\tilde{E})$.] Let $\tilde{Z}$ be a solution of the martingale problem for $\tilde{A}$. Then, for $i = 1, 2$, $Z_i \equiv \hat{\rho}_i \circ \tilde{Z}$ is a Fleming–Viot process with mutation operator B and recombination given by α and η, that is, $\tilde{A}$, $\hat{\rho}_1$, $\hat{\rho}_2$ and $\hat{\rho}$ determine a Markov coupling for A. More importantly, we will show that if $\tilde{B}$, ρ_1, ρ_2 and ρ determine a successful Markov coupling for B, then, with some additional assumptions, $\tilde{A}$, $\hat{\rho}_1$, $\hat{\rho}_2$ and $\hat{\rho}$ determine a successful Markov coupling for A.

LEMMA 3.1. *Let Z be a neutral Fleming–Viot process with type space E, mutation operator B with corresponding semigroup $\{T(t)\}$ on $B(E)$ as in Section 1 and recombination given by α and η. Let $D \subset E$ be closed. Suppose that $T(t)I_D \ge I_D$ for all $t \ge 0$, $\lim_{t\to\infty} T(t)I_D(x) = 1$ for each $x \in E$ and $\eta(x, y, D) = 1$ for all $x, y \in D$. Define $\tau = \inf\{t \ge 0: Z(t, D) = 1\}$. Then the* *following conclusions hold*:

(a) $P\{\tau < \infty\} > 0$ *and, a.s. on the event* $\{\tau < \infty\}$, $Z(\tau + t, D) = 1$ *for all* $t \geq 0$.

(b) *If* $\alpha = 0$ (*i.e., there is no recombination*), *then* $P\{\tau < \infty\} = 1$.

(c) *If there exist* $t > 0$ *and* $\delta > 0$ *such that* $T(t)I_D(x) \geq \delta$ *for all* $x \in E$, *then* $P\{\tau < \infty\} = 1$.

PROOF. Since Z has continuous sample paths (in the weak topology) we can assume that Z is a solution of the martingale problem for A with respect to a right continuous filtration $\{F_t\}$. Let $\lambda > 0$, $g \in B(E)$ and $h_{\lambda, g} = \int_0^\infty \lambda e^{-\lambda t} T(t) g\, dt$. If $h_{\lambda, g} \in D(B)$, then $Bh_{\lambda, g} = \lambda(h_{\lambda, g} - g)$. In general, the collection of pairs $\{(h_{\lambda, g}, \lambda g)\colon g \in B(E)\}$ is the bounded pointwise closure of $\{(h, \lambda h - Bh)\colon h \in D(B)\}$ [since $B(E)$ is the bounded pointwise closure of $R(\lambda - B)$] and (setting $h = h_{\lambda, g}$) it follows that

$$\begin{aligned} M_h(t) = {}& \langle h, Z(t)\rangle - \langle h, Z(0)\rangle \\ & - \int_0^t \Big(\langle \lambda(h - g), Z(s)\rangle \\ & \qquad + \alpha\Big(\big\langle H_1^{(1)} h, Z(s)^2\big\rangle - \langle h, Z(s)\rangle\Big)\Big)\, ds \end{aligned} \tag{3.5}$$

is an $\{F_t\}$-martingale for all $g \in B(E)$. If g extends to a continuous function on the one-point compactification of E, h will be continuous, so M_h will have continuous sample paths. If M_h has continuous sample paths, its quadratic variation is

$$\int_0^t \Big(\langle h^2, Z(s)\rangle - \langle h, Z(s)\rangle^2\Big)\, ds, \tag{3.6}$$

and for each $f \in C^2(\mathbf{R})$,

$$\begin{aligned} & f(\langle h, Z(t)\rangle) - f(\langle h, Z(0)\rangle) \\ & \quad - \int_0^t \Big(\tfrac{1}{2}\Big(\langle h^2, Z(s)\rangle - \langle h, Z(s)\rangle^2\Big) f''(\langle h, Z(s)\rangle) \\ & \qquad + \Big(\langle \lambda(h - g), Z(s)\rangle \\ & \qquad\quad + \alpha\Big(\big\langle H_1^{(1)} h, Z(s)^2\big\rangle - \langle h, Z(s)\rangle\Big)\Big) f'(\langle h, Z(s)\rangle)\Big)\, ds \end{aligned} \tag{3.7}$$

is an $\{F_t\}$-martingale. However, the collection of g for which (3.7) is a martingale is closed under bounded pointwise convergence, so (3.7) is a martingale for all g and $h = h_{\lambda, g}$. Let $g = I_D$ and $h_\lambda = \int_0^\infty \lambda e^{-\lambda t} T(t) I_D\, dt$. Note that $T(t)I_D(x)$ is a nondecreasing function of t for each $x \in E$, and hence $h_\lambda(x)$ is a nonincreasing function of λ for each $x \in E$. The fact that $T(t)I_D(x) \to I_D(x)$ for each $x \in E$ as $t \to 0$ implies that $\lim_{\lambda \to \infty} h_\lambda(x) = I_D(x)$

for each $x \in E$. Assume that $f' \geq 0$. Then, since $\lambda(h_\lambda - I_D) \geq 0$,

$$\begin{aligned} V_\lambda(t) &\equiv f(\langle h_\lambda, Z(t)\rangle) - f(\langle h_\lambda, Z(0)\rangle) \\ &\quad - \int_0^t \Big(\tfrac{1}{2}\big(\langle h_\lambda^2, Z(s)\rangle - \langle h_\lambda, Z(s)\rangle^2\big) f''(\langle h_\lambda, Z(s)\rangle) \\ &\qquad + \alpha\big(\langle H_1^{(1)} h_\lambda, Z(s)^2\rangle - \langle h_\lambda, Z(s)\rangle\big) f'(\langle h_\lambda, Z(s)\rangle)\Big)\, ds \end{aligned} \tag{3.8}$$

is an $\{F_t\}$-submartingale. Now, letting $\lambda \to \infty$ and observing that $H_1^{(1)} I_D \geq I_{D\times D}$, we see that

$$\begin{aligned} V(t) &\equiv f(Z(t, D)) - f(Z(0, D)) \\ &\quad - \int_0^t \Big(\tfrac{1}{2}\big(Z(s, D) - Z(s, D)^2\big) f''(Z(s, D)) \\ &\qquad + \alpha\big(Z(s, D)^2 - Z(s, D)\big) f'(Z(s, D))\Big)\, ds \end{aligned} \tag{3.9}$$

is an $\{F_t\}$-submartingale. Since $\{F_t\}$ is assumed right continuous, $\tilde{V}(t) \equiv \lim_{s\in \mathbf{Q},\, s\to t+} V(s)$ defines a right continuous process satisfying $\tilde{V}(t) \geq V(t)$ a.s. for each $t \geq 0$ [since $\tilde{V}(t) = E[\tilde{V}(t)|F_t] = \lim_{s\in\mathbf{Q},\, s\to t+} E[V(s)|F_t] \geq V(t)$ a.s.]. By taking f to be strictly increasing in (3.9), we see that $\tilde{Z}(t) \equiv \lim_{s\in\mathbf{Q},\, s\to t+} Z(s, D)$ must exist and be right continuous and satisfy $\tilde{Z}(t) \geq Z(t, D)$ a.s. for each $t \geq 0$. Since D is closed, the continuity of Z implies that $Z(\cdot, D)$ is upper semicontinuous and hence

$$P\big\{Z(t, D) \geq \tilde{Z}(t) \text{ for all } t \geq 0\big\} = 1. \tag{3.10}$$

It follows that $\tilde{Z}(t) = Z(t, D)$ a.s. for each $t \geq 0$, that is, $\tilde{Z}$ is a modification of $Z(\cdot, D)$. Take $f(z) = e^{2\alpha z}$ if $\alpha > 0$ and $f(z) = z$ if $\alpha = 0$. Then $f(\tilde{Z}(t))$ is a right continuous submartingale that is a modification of $f(Z(t, D))$.

Let $\psi_\varepsilon = \inf\{t \geq 0: \tilde{Z}(t) \notin (\varepsilon, 1)\}$. If ψ_ε is finite a.s., then, with f as above,

$$E\big[f(Z(\psi_\epsilon, D))\big] \geq E\Big[f\big(\tilde{Z}(\psi_\varepsilon)\big)\Big] \geq E\Big[f\big(\tilde{Z}(0)\big)\Big] = E\big[f(Z(0, D))\big] \tag{3.11}$$

and hence, using (3.10),

$$\begin{aligned} &P\big\{Z(\psi_\varepsilon, D) = 1\big\} \\ &\quad \geq P\big\{\tilde{Z}(\psi_\varepsilon) = 1\big\} \\ &\quad \geq \begin{cases} \big(E[e^{2\alpha Z(0,D)}] - e^{2\alpha\varepsilon}\big)/\big(e^{2\alpha} - e^{2\alpha\varepsilon}\big), & \text{if } \alpha > 0, \\ \big(E[Z(0, D)] - \varepsilon\big)/(1 - \varepsilon), & \text{if } \alpha = 0. \end{cases} \end{aligned} \tag{3.12}$$

The finiteness a.s. of ψ_ε also follows from (3.9) by letting g be the solution on $[0, 1]$ of

$$\tfrac{1}{2}(z - z^2) g''(z) + \alpha(z^2 - z) g'(z) = z, \qquad g'(0) = 0,\ g(0) = 0, \tag{3.13}$$

and observing that g is bounded and

$$E[\psi_\varepsilon] \le \frac{E[g(Z(\psi_\varepsilon, D))] - E[g(Z(0,D))]}{\varepsilon}. \tag{3.14}$$

Finally, the right-hand side of (3.12) is less than or equal to zero unless $E[e^{2\alpha Z(0,D)}] > e^{2\alpha\varepsilon}$. If we know that $\sup_{t\ge 0} Z(t,D) > 0$, then the strong Markov property and (3.12) imply that there is positive probability of $Z(\cdot, D)$ hitting 1. However, the assumptions on $\{T(t)\}$ and η imply that the dual process Y with $Y(0) = I_D$ satisfies $Y(t) \ge I_{D^{M(t)}}$ for all $t \ge 0$ and hence that

$$E[Z(t,D)] \ge E\Big[\big\langle T(t-s_t) I_D, Z(0)\big\rangle^{M(t)}\Big], \qquad t \ge 0, \tag{3.15}$$

where s_t is the time of the last jump in the dual process before time t. Since $T(t)I_D(x) \to 1$ as $t \to \infty$ for each x, the right-hand side of (3.15) will be positive for some t sufficiently large, and hence $P\{Z(t,D) > 0\} > 0$, giving the first conclusion in part (a). The second conclusion in part (a) uses (3.9). Under the assumptions of part (b), the right-hand side of (3.15) converges to 1, so given $0 < \varepsilon < 1$, there exists t such that $E[Z(t,D)] \ge 1 - \varepsilon$ and hence $P\{Z(t,D) \ge 1 - \sqrt{\varepsilon}\} \ge 1 - \sqrt{\varepsilon}$. Consequently, by (3.12),

$$P\{\tau < \infty\} \ge \begin{cases} \big(e^{2\alpha(1-\sqrt{\varepsilon})}(1-\sqrt{\varepsilon}) - e^{2\alpha\varepsilon}\big)/(e^{2\alpha} - e^{2\alpha\varepsilon}), & \text{if } \alpha > 0, \\ \Big(\big(1-\sqrt{\varepsilon}\big)^2 - \varepsilon\Big)/(1-\varepsilon), & \text{if } \alpha = 0, \end{cases} \tag{3.16}$$

and since ε is arbitrary, part (b) follows.

Finally, under the uniformity assumption in part (c), the above calculations give a lower bound on $P\{\tau < \infty\}$ that is independent of the initial distribution. Consequently, the strong Markov property and a renewal argument give the desired conclusion. □

If there is a successful Markov coupling $(\tilde{B}, \rho_1, \rho_2, \rho)$ for the mutation operator B with $D = \{x \in \tilde{E}\colon\ \rho_1(x) = \rho_2(x)\}$ a closed set satisfying the conditions of parts (b) or (c) of the lemma (with respect to the semigroup $\{\tilde{T}(t)\}$ corresponding to $\tilde{B}$), it is clear how to construct a successful coupling for the neutral Fleming–Viot process. Clearly, these conditions hold for the example defined in (3.1) and (3.2) [but in this case the resulting ergodic theorem has already been proved by Shiga (1990); see also Ethier and Griffiths (1993)]. See Lindvall and Rogers (1986) for a coupling for multidimensional diffusion processes.

THEOREM 3.2. *Suppose that the martingale problem for $\tilde{A}$, the neutral Fleming–Viot generator with mutation operator B and recombination given by α and η, is well posed. Assume that $(\tilde{B}, \rho_1, \rho_2, \rho)$ determines a successful Markov coupling for B and that $\tilde{\eta}$ is defined by* (3.3). *Suppose that the martingale problem for $\tilde{A}$, the neutral Fleming–Viot generator with mutation* *operator $\tilde{B}$ and recombination given by α and $\tilde{\eta}$, is well posed. Then the following conclusions hold.*

(a) *There exists at most one stationary distribution for the Fleming–Viot process with generator A.*

(b) *If $\alpha = 0$, then $(\tilde{A}, \hat{\rho}_1, \hat{\rho}_2, \hat{\rho})$ determines a successful Markov coupling for A, where $\hat{\rho}_1$, $\hat{\rho}_2$ and $\hat{\rho}$ are defined as in the paragraph preceding Lemma* 3.1.

(c) *If there exist $t > 0$ and $\delta > 0$ such that $\tilde{T}(t) I_D(x) \geq \delta$ for all $x \in \tilde{E}$, then $(\tilde{A}, \hat{\rho}_1, \hat{\rho}_2, \hat{\rho})$ determines a successful Markov coupling for A.*

PROOF. Given $\mu_1, \mu_2 \in P(E)$, let $\gamma = \hat{\rho}(\mu_1, \mu_2) \in P(\tilde{E})$, so that $\mu_i = \hat{\rho}_i(\gamma)$ for $i = 1, 2$. Let $\tilde{Z}$ be a solution of the martingale problem for $\tilde{A}$ starting from γ and let $Z_i = \hat{\rho}_i \circ \tilde{Z}$ for $i = 1, 2$. By Lemma 3.1, $\tau \equiv \inf\{t \geq 0: \tilde{Z}(t, D) = 1\}$ is finite with positive probability and, a.s. on $\{\tau < \infty\}$, $\tilde{Z}(\tau + t, D) = 1$ for all $t \geq 0$. Consequently, a.s. on the event $\{\tau \leq t\}$, $Z_1(t, G) \equiv \tilde{Z}(t, \rho_1^{-1} G) = \tilde{Z}(t, D \cap \rho_1^{-1} G) = \tilde{Z}(t, D \cap \rho_2^{-1} G) = \tilde{Z}(t, \rho_2^{-1} G) = Z_2(t, G)$ for all $G \in B(E)$, and hence

$$P\{Z_1(t) = Z_2(t)\} \geq P\{\tau \leq t\}, \qquad t \geq 0. \tag{3.17}$$

It follows that for t satisfying $P\{\tau \leq t\} > 0$, the distributions of $Z_1(t)$ and $Z_2(t)$ cannot be mutually singular. However, if there exists more than one stationary distribution for A, then there exist two mutually singular stationary distributions. (See Lemma 5.3 below.) This contradiction verifies part (a). Under the assumptions of parts (b) and (c), $P\{\tau < \infty\} = 1$, so (3.17) implies that the coupling is successful. □

The usual application of a successful coupling is to prove an ergodic theorem for the coupled process. If there is a stationary distribution for the process, then existence of a successful coupling ensures that the stationary distribution is unique and implies asymptotic stationarity for the process starting from any initial distribution. Since for the neutral Fleming–Viot process, duality gives ergodicity under much more general conditions, our primary interest in the above coupling is as a first step in constructing a coupling for the Fleming–Viot process with selection, a setting in which duality appears to give ergodicity results in only very limited cases. Even in the neutral case, however, when the above coupling exists, it gives stronger ergodicity results than the duality argument does.

THEOREM 3.3. *Suppose that the conditions of Theorem* 3.2(*b*) *or* (*c*) *are satisfied and that there exists a stationary distribution* Π *for the Fleming–Viot process. (If $\alpha = 0$, the existence of a stationary distribution for the mutation* *process implies the existence of* Π.) *Then, for each solution* Z *of the martingale problem for* A,

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t F(Z(s))\,ds=\int_{P(E)}F(\mu)\Pi(d\mu)\quad a.s.,\quad F\in B(P(E)) \tag{3.18}$$

and

$$\lim_{t\to\infty}\sup_{G\in B(P(E))}\left|P\{Z(t)\in G\}-\Pi(G)\right|=0. \tag{3.19}$$

PROOF. Suppose that Z_1 and Z_2 are solutions of the martingale problem for A with initial distributions $\mu_1,\mu_2\in P(P(E))$. The coupling inequality (3.17) implies

$$\left|P\{Z_1(t)\in G\}-P\{Z_2(t)G\}\right|\le P\{\tau>t\},\qquad t\ge 0, \tag{3.20}$$

where τ is the coupling time for a solution $\tilde{Z}$ of the martingale problem for $\tilde{A}$ with $P\{\hat{\rho}_i(\tilde{Z}(0))\in\cdot\}=\mu_i$ for $i=1,2$. Taking $\mu_2=\Pi$ in (3.20) gives (3.19), and by the successful coupling and the ergodicity of Z_2 (in the stationary process sense, by virtue of the uniqueness of stationary distributions), we have

$$\begin{aligned}\lim_{t\to\infty}\frac{1}{t}\int_0^t F(Z_1(s))\,ds&=\lim_{t\to\infty}\frac{1}{t}\int_0^t F(Z_2(s))\,ds\\&=\int_{P(E)}F(\mu)\Pi(d\mu)\quad\text{a.s.}\end{aligned} \tag{3.21}$$

and (3.18) follows. □

4. An ergodic theorem for the unlabeled infinitely-many-alleles diffusion model with symmetric overdominance. A Fleming–Viot process Z with type space E and mutation operator B of the form

$$Bf(x)=\tfrac{1}{2}\theta\int_E(f(y)-f(x))Q(x,dy), \tag{4.1}$$

where $\theta>0$ and Q is a one-step transition function for a Markov chain in E such that

$$Q(x,\{y\})=0,\qquad x,y\in E, \tag{4.2}$$

is sometimes referred to as the *infinitely-many-alleles diffusion model*, since by (4.2) every mutant allele is new. It is known [Ethier and Kurtz (1987)] that

$$P\{Z(t)\in P_{\mathrm{a}}(E)\text{ for all }t>0\}=1, \tag{4.3}$$

where $P_{\mathrm{a}}(E)=\{\mu\in P(E):\ \mu\text{ is purely atomic}\}$.

Let

$$\nabla_\infty=\left\{p=(p_1,p_2,\dots):\ p_1\ge p_2\ge\cdots\ge 0,\ \sum_{i=1}^\infty p_i=1\right\} \tag{4.4}$$

and

$$\overline{\nabla}_\infty = \left\{ p = (p_1, p_2, \dots): p_1 \ge p_2 \ge \cdots \ge 0,\ \sum_{i=1}^\infty p_i \le 1 \right\}, \tag{4.5}$$

so that $\overline{\nabla}_\infty$ is the closure of ∇_∞ in $[0,1]^\infty$ and is therefore compact. Define γ: $P(E) \mapsto \overline{\nabla}_\infty$ by letting $\gamma(\mu) = p$ if p_i is the size (or mass) of the ith largest atom of μ (or 0 if μ has fewer than i atoms) for each $i \ge 1$, and note that $\gamma(P_a(E)) = \nabla_\infty$. In some applications it is sufficient to consider the process $\gamma \circ Z$, which keeps track of the allele frequencies but not the alleles to which they correspond. (For this reason we call it the *unlabeled* infinitely-many-alleles diffusion model.)

In the neutral case with no recombination, Ethier and Kurtz (1987) showed that $\gamma \circ Z$ can be characterized as the unique solution of the $C_{\overline{\nabla}_\infty}[0,\infty)$ martingale problem for

$$G = \frac{1}{2} \sum_{i,j=1}^\infty p_i(\delta_{ij} - p_j) \frac{\partial^2}{\partial p_i\, \partial p_j} - \frac{1}{2}\theta \sum_{i=1}^\infty p_i \frac{\partial}{\partial p_i} \tag{4.6}$$

with $D(G)$ taken to be the subalgebra of $C(\overline{\nabla}_\infty)$ generated by $1, \varphi_2, \varphi_3, \dots,$ where $\varphi_m(p) \equiv \sum_{i=1}^\infty p_i^m$. [By convention, sums in (4.6) are evaluated on ∇_∞ and extended to $\overline{\nabla}_\infty$ by continuity.]

In the presence of selection, however, $\gamma \circ Z$ is ordinarily non-Markovian. Nevertheless, in one important case, namely, heterozygote advantage [known as symmetric overdominance or heterosis; see Watterson (1977)] or disadvantage, the Markov property of $\gamma \circ Z$ is preserved. Fix a real constant σ, let D be the diagonal of $E \times E$ and consider the selection intensity function

$$\tilde{\sigma}(x, y) = \sigma I_D(x, y); \tag{4.7}$$

$\sigma < 0$ means heterozygote advantage and $\sigma > 0$ means the opposite. We begin by showing that $\gamma \circ Z$ can be characterized as the unique solution of the $C_{\overline{\nabla}_\infty}[0,\infty)$ martingale problem for

$$G_\sigma = G + \sigma \sum_{i=1}^\infty p_i(p_i - \varphi_2(p)) \frac{\partial}{\partial p_i} \tag{4.8}$$

with $D(G_\sigma) = D(G)$.

LEMMA 4.1. *G_σ is closable in $C(\overline{\nabla}_\infty)$ and its closure generates a Feller semigroup $\{T_\sigma(t)\}$ on $C(\overline{\nabla}_\infty)$. In particular, the $C_{\overline{\nabla}_\infty}[0,\infty)$ martingale problem for G_σ is well posed.*

PROOF. Denote by G_σ^+ the action of (4.8) on D^+, the subspace of $C(\overline{\nabla}_\infty)$ consisting of all functions of the form $F \circ (\varphi_{m_1}, \dots, \varphi_{m_k})$, where $k \ge 1$, m_1,

$\dots, m_k \geq 2$, and $F \in C^2(\mathbf{R}^k)$, and note that

$$G_\sigma^+\left[F \circ (\varphi_{m_1}, \dots, \varphi_{m_k})\right]$$

$$= \tfrac{1}{2} \sum_{i,j=1}^{k} m_i m_j (\varphi_{m_i + m_j - 1} - \varphi_{m_i}\varphi_{m_j})\left[F_{z_i z_j} \circ (\varphi_{m_1}, \dots, \varphi_{m_k})\right] \tag{4.9}$$

$$+ \sum_{i=1}^{k} G_\sigma \varphi_{m_i}\left[F_{z_i} \circ (\varphi_{m_1}, \dots, \varphi_{m_k})\right]$$

and $G_\sigma^+ \subset \overline{G}_\sigma$. Letting $a(p)$ denote for each $p \in \overline{\nabla}_\infty$ the infinite-dimensional square matrix whose (i, j)th entry is $p_i(\delta_{ij} - p_j)$ for $i, j = 1, 2, \dots,$ we find that

$$
\begin{aligned}
G_\sigma^+&\left(\exp(-\sigma\varphi_2/2) f\right)\\
&= \exp(-\sigma\varphi_2/2) G_\sigma f + G_\sigma^+\left(\exp(-\sigma\varphi_2/2)\right) f\\
&\quad + \left\langle \operatorname{grad} \exp(-\sigma\varphi_2/2), a \operatorname{grad} f\right\rangle\\
&= \exp(-\sigma\varphi_2/2)\left\{Gf + \tfrac{1}{2}\sigma\left\langle \operatorname{grad} \varphi_2, a \operatorname{grad} f\right\rangle\right\}\\
&\quad + G_\sigma^+\left(\exp(-\sigma\varphi_2/2)\right) f\\
&\quad - \tfrac{1}{2}\sigma \exp(-\sigma\varphi_2/2)\left\langle \operatorname{grad} \varphi_2, a \operatorname{grad} f\right\rangle\\
&= \exp(-\sigma\varphi_2/2)(G + c) f
\end{aligned} \tag{4.10}
$$

for each $f \in D(G)$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in l^2 and $c = \exp(\sigma\varphi_2/2) G_\sigma^+ \exp(-\sigma\varphi_2/2) \in C(\overline{\nabla}_\infty)$. We know that $R(\lambda - G)$ is dense in $C(\overline{\nabla}_\infty)$ for every $\lambda > 0$ [cf. Ethier and Kurtz (1981)]. Let $\lambda > \|c\|_\infty$. Then $R(\lambda - G - c)$ is dense in $C(\overline{\nabla}_\infty)$ by the bounded perturbation theorem, hence $R(\lambda - G_\sigma^+)$ is dense in $C(\overline{\nabla}_\infty)$ by (4.10) and finally $R(\lambda - G_\sigma)$ is dense in $C(\overline{\nabla}_\infty)$ since $G_\sigma^+ \subset \overline{G}_\sigma$. Since $D(G_\sigma)$ is dense in $C(\overline{\nabla}_\infty)$ by Stone–Weierstrass and G_σ is dissipative [cf. Ethier and Kurtz (1981)], the first conclusion of the lemma follows by the Hille–Yosida theorem. The second conclusion is a consequence of this [Ethier and Kurtz (1986), Theorem 4.4.1]. □

PROPOSITION 4.2. *Let Z be a Fleming–Viot process with type space E, mutation operator B satisfying* (4.1) *and* (4.2), *no recombination, and selection intensity function $\tilde{\sigma}$ given by* (4.7). *Then $\gamma \circ Z$ solves the $C_{\overline{\nabla}_\infty}[0, \infty)$ martingale problem for G_σ.*

PROOF. Denote the generator of Z by A [see (1.9)]. Let $k \geq 1$, $m_1, \dots, m_k \geq 2$ and $m = m_1 + \cdots + m_k$. Define $f \in B(E^m)$ to be the indicator function of the set of $(x_1, \dots, x_m) \in E^m$ for which $x_1 = \cdots = x_{m_1}$, $x_{m_1+1} = \cdots = x_{m_1+m_2}$, $\dots, x_{m_1+\cdots+m_{k-1}+1} = \cdots = x_{m_1+\cdots+m_k}$. Then

$$F_f(\mu) \equiv \langle f, \mu^m \rangle = (\varphi_{m_1} \cdots \varphi_{m_k})(\gamma(\mu)) \tag{4.11}$$

and

$$
\begin{aligned}
AF_f(\mu) = &\sum_{i=1}^{k} \binom{m_i}{2} \varphi_{m_i-1}(\gamma(\mu)) \prod_{l:\, l\neq i} \varphi_{m_l}(\gamma(\mu)) \\
&+ \sum_{1\le i<j\le k} m_i m_j \varphi_{m_i+m_j-1}(\gamma(\mu)) \prod_{l:\, l\neq i,\, j} \varphi_{m_l}(\gamma(\mu)) \\
&- \left[\binom{m}{2} + \frac{m\theta}{2}\right] \prod_{l=1}^{k} \varphi_{m_l}(\gamma(\mu)) \\
&+ \sigma \sum_{i=1}^{k} m_i \varphi_{m_i+1}(\gamma(\mu)) \prod_{l:\, l\neq i} \varphi_{m_l}(\gamma(\mu)) \\
&- \sigma m \varphi_2(\gamma(\mu)) \prod_{l=1}^{k} \varphi_{m_l}(\gamma(\mu)) \\
= {}& G_\sigma(\varphi_{m_1} \cdots \varphi_{m_k})(\gamma(\mu))
\end{aligned}
\tag{4.12}
$$

for all $\mu \in P(E)$, and the result follows. □

It is known [Ethier and Kurtz (1981)] that the diffusion in $\overline{\nabla}_\infty$ with generator G has a unique stationary distribution $\pi_0 \in P(\overline{\nabla}_\infty)$ and is reversible with respect to π_0. The probability measure π_0 is known as the Poisson-Dirichlet distribution with parameter θ [Kingman (1975)], and $\pi_0(\nabla_\infty) = 1$. Let us define $\pi_\sigma \in P(\overline{\nabla}_\infty)$ by

$$
\pi_\sigma(dp) = \exp(\sigma\varphi_2(p))\pi_0(dp) \Big/ \int_{\overline{\nabla}_\infty} \exp(\sigma\varphi_2)\, d\pi_0. \tag{4.13}
$$

The main result of this section is the following strong ergodic theorem. A more general result can be proved using the methods of Section 5, but the present approach has the advantage of simplicity.

THEOREM 4.3. *Let $P_\sigma(t, p, dq)$ denote the transition function corresponding to the semigroup $\{T_\sigma(t)\}$ of Lemma* 4.1. *Then π_σ is stationary for $\{T_\sigma(t)\}$ and*

$$
\lim_{t\to\infty} \sup_{\Gamma \in B(\overline{\nabla}_\infty)} |P_\sigma(t, p, \Gamma) - \pi_\sigma(\Gamma)| = 0, \qquad p \in \overline{\nabla}_\infty. \tag{4.14}
$$

In particular, π_σ is the unique stationary distribution for $\{T_\sigma(t)\}$.

PROOF. As already noted in (4.10),

$$
G_\sigma f = Gf + \tfrac{1}{2}\sigma\langle \operatorname{grad} \varphi_2, a \operatorname{grad} f\rangle \tag{4.15}
$$

for all $f \in D(G)$, so the stationarity of π_σ (in fact, reversibility) follows by a result of Fukushima and Stroock (1986). [Alternatively, see Ethier and Kurtz (1994), Section 4.] The strong ergodicity (4.14) will follow from a lemma of

Shiga (1981), provided we can show that, for some $t_0 > 0$,

$$P_\sigma(t_0, p, \cdot) \simeq P_0(t_0, p, \cdot) \simeq \pi_0 \simeq \pi_\sigma, \qquad p \in \overline{\nabla}_\infty, \tag{4.16}$$

where $\simeq$ denotes mutual absolute continuity.

Again appealing to Fukushima and Stroock (1986) (or just using the Girsanov transformation), we have

$$\begin{aligned} &P_\sigma(t, p, \cdot) \\ &\quad = E_p\left[I_{\{X(t)\in\cdot\}} \exp\left\{ h(X(t)) - h(X(0)) - \int_0^t g(X(s))\, ds \right\}\right] \end{aligned} \tag{4.17}$$

for all $t \geq 0$ and $p \in \overline{\nabla}_\infty$, where X is the diffusion in $\overline{\nabla}_\infty$ with generator G, $h = \frac{1}{2}\sigma\varphi_2$ and $g = Gh + \frac{1}{2}\langle \text{grad}\, h, a\, \text{grad}\, h\rangle$. [Alternatively, see Dawson (1978).] This gives the first equivalence in (4.16) for all $t_0 \geq 0$. The third equivalence is immediate from (4.13), and the second holds for all t_0 sufficiently large, because

$$P_0(t, p, dq) = \Xi(t, p, q)\pi_0(dq) \tag{4.18}$$

for a function Ξ on $(0,\infty) \times \overline{\nabla}_\infty \times \overline{\nabla}_\infty$ with the property that $\Xi(t, p, q) \to 1$ as $t \to \infty$, uniformly in p and q; see Ethier (1992) [or Griffiths (1979)]. $\square$

5. Coupling and an ergodic theorem for Fleming–Viot processes with selection. Construction of a coupling for the Fleming–Viot process with selection does not seem to be as simple as in the neutral case. In fact, we use the neutral coupling and Dawson's Girsanov-type formula to give conditions under which a successful coupling (not necessarily Markov) exists. The following lemma is suggested by work of Athreya and Ney (1978) and Nummelin (1978).

LEMMA 5.1. *Let $\mu(x, \Gamma)$ be a one-step transition function on a measurable state space $(S, \mathcal{S})$ and let $\nu(x, y, \Gamma)$ be transition function from $(S \times S, \mathcal{S} \times \mathcal{S})$ to $(S, \mathcal{S})$. Let ε: $S \times S \mapsto [0,1]$ be $\mathcal{S} \times \mathcal{S}$-measurable and satisfy*

$$\mu(x, \Gamma) \wedge \mu(y, \Gamma) \geq \varepsilon(x, y)\nu(x, y, \Gamma), \qquad x, y \in S, \Gamma \in \mathcal{S}. \tag{5.1}$$

Let $\{X_k\}$ and $\{Y_k\}$ be independent Markov chains with one-step transition function μ. If

$$\sum_{k=0}^{\infty} \varepsilon(X_k, Y_k) = \infty \quad a.s., \tag{5.2}$$

then there exists a probability space on which is defined a Markov chain $\{(\tilde{X}_k, \tilde{Y}_k)\}$ such that $\{\tilde{X}_k\}$ has the same distribution as $\{X_k\}$, $\{\tilde{Y}_k\}$ has the same distribution as $\{Y_k\}$ and there exists a random variable $0 \leq \kappa < \infty$ a.s. such that $k \geq \kappa$ implies $\tilde{X}_k = \tilde{Y}_k$.

PROOF. Assume without loss of generality that $\nu(x, x, \Gamma) = \mu(x, \Gamma)$, $\varepsilon(x, x) = 1$, $\nu(x, y, \Gamma) = \nu(y, x, \Gamma)$, $\varepsilon(x, y) = \varepsilon(y, x)$ and, for $x \neq y$, $\varepsilon(x, y) \leq \frac{1}{2}$. [Replacing $\varepsilon(x, y)$ by $\varepsilon(x, y) \wedge \frac{1}{2}$ if $x \neq y$ does not affect the validity of

(5.2).] For each $y \in S$, define a transition function η_y from $(S, \mathcal{S})$ to $(S, \mathcal{S})$ by

$$\eta_y(x, \Gamma) = \begin{cases} (\mu(x,\Gamma) - \varepsilon(x, y)\nu(x, y, \Gamma))/(1 - \varepsilon(x, y)), & \text{if } x \neq y, \\ \mu(x,\Gamma), & \text{if } x = y. \end{cases} \tag{5.3}$$

Define the transition function $\tilde{\mu}$ from $(S \times S, \mathcal{S} \times \mathcal{S})$ to $(S \times S, \mathcal{S} \times \mathcal{S})$ by

$$\begin{aligned} \tilde{\mu}(x, y, \Gamma_1 \times \Gamma_2) &= \varepsilon(x, y)\nu(x, y, \Gamma_1 \cap \Gamma_2) \\ &\quad + (1 - \varepsilon(x, y))\eta_y(x, \Gamma_1)\eta_x(y, \Gamma_2). \end{aligned} \tag{5.4}$$

Let $\{(\tilde{X}_k, \tilde{Y}_k)\}$ be a Markov chain with one-step transition function $\tilde{\mu}(x, y, \Gamma)$ and with initial distribution that of $\{(X_k, Y_k)\}$. Then $\{\tilde{X}_k\}$ and $\{\tilde{Y}_k\}$ are Markov chains with one-step transition function μ. Intuitively, at the kth transition a coin is flipped which is heads with probability $\varepsilon(\tilde{X}_{k-1}, \tilde{Y}_{k-1})$. If heads comes up, then $\tilde{X}_k = \tilde{Y}_k$ and both have conditional distribution $\nu(\tilde{X}_{k-1}, \tilde{Y}_{k-1}, \cdot)$. If tails comes up, $\tilde{X}_k$ and $\tilde{Y}_k$ are conditionally independent with conditional distributions $\eta_{\tilde{Y}_{k-1}}(\tilde{X}_{k-1}, \cdot)$ and $\eta_{\tilde{X}_{k-1}}(\tilde{Y}_{k-1}, \cdot)$, respectively. Let $\kappa = \min\{k \geq 0: \tilde{X}_k = \tilde{Y}_k\}$. We want to show that $P\{\kappa < \infty\} = 1$.

First note that, for each $x, y \in S$, $\eta_y(x, \cdot) \ll \mu(x, \cdot)$ and, letting $f_y(x, z)$ denote the Radon–Nikodym derivative, $f_y(x, z) \leq 1/(1 - \varepsilon(x, y))$ if $x \neq y$, and $f_y(y, z) = 1$. By the Markov property and the definition of $\tilde{\mu}$, we have for each $m \geq 1$,

$$\begin{aligned} &P\{\kappa > m + 1\} \\ &\quad = E\left[I_{\{\kappa > m+1\}}\right] \\ &\quad = E\left[1 - \varepsilon\left(\tilde{X}_m, \tilde{Y}_m\right)\right] \\ &\quad = E\Bigg[\left(1 - \varepsilon\left(\tilde{X}_{m-1}, \tilde{Y}_{m-1}\right)\right) \\ &\qquad\quad \times \int_E\int_E (1 - \varepsilon(x_m, y_m))\eta_{\tilde{Y}_{m-1}}\left(\tilde{X}_{m-1}, dx_m\right)\eta_{\tilde{X}_{m-1}}\left(\tilde{Y}_{m-1}, dy_m\right)\Bigg] \\ &\quad = E\Bigg[\left(1 - \varepsilon\left(\tilde{X}_{m-1}, \tilde{Y}_{m-1}\right)\right) \\ &\qquad\quad \times \int_E\int_E (1 - \varepsilon(x_m, y_m)) f_{\tilde{Y}_{m-1}}\left(\tilde{X}_{m-1}, x_m\right) f_{\tilde{X}_{m-1}}\left(\tilde{Y}_{m-1}, y_m\right) \\ &\qquad\qquad\qquad \mu\left(\tilde{X}_{m-1}, dx_m\right)\mu\left(\tilde{Y}_{m-1}, dy_m\right)\Bigg]. \end{aligned} \tag{5.5}$$

Continuing in this way, we see that

$$P\{\kappa > m+1\} = E\Bigg[\prod_{k=0}^{m}\big(1-\varepsilon(X_k, Y_k)\big) \times \prod_{k=0}^{m-1} f_{Y_k}(X_k, X_{k+1}) f_{X_k}(Y_k, Y_{k+1})\Bigg], \tag{5.6}$$

where $\{(X_k, Y_k)\}$ is as in the statement of the lemma. Observe that

$$L_m \equiv \prod_{k=0}^{m-1} f_{Y_k}(X_k, X_{k+1}) f_{X_k}(Y_k, Y_{k+1}) \tag{5.7}$$

is a martingale (empty products are 1) and

$$L_m \le \exp\left(4\sum_{k=0}^{m-1}\varepsilon(X_k, Y_k)\right), \qquad m \ge 0. \tag{5.8}$$

[Here, we are using the assumption that $\varepsilon(x, y) \le \frac{1}{2}$ when $x \ne y$, together with the inequality $1/(1-u) \le e^{2u}$ for $0 \le u \le \frac{1}{2}$.]

For $c > 0$, let $\tau_c = \min\{m \ge 0: \sum_{k=0}^{m}\varepsilon(X_k, Y_k) > c\}$ and note that by (5.2), $\tau_c < \infty$ a.s. By (5.6),

$$\begin{aligned} P\{\kappa > m+1\} &\le E\left[L_m \sum_{k=0}^{m\wedge\tau_c}\big(1-\varepsilon(X_k, Y_k)\big)\right] \\ &= E\left[L_{m\wedge\tau_c}\prod_{k=0}^{m\wedge\tau_c}\big(1-\varepsilon(X_k, Y_k)\big)\right]. \end{aligned} \tag{5.9}$$

Since by (5.8), $L_{m\wedge\tau_c} \le e^{4c}$, letting $m \to \infty$, we can interchange the limit and expectation on the right to obtain

$$\limsup_{m\to\infty} P\{\kappa > m+1\} \le E\left[L_{\tau_c}\prod_{k=0}^{\tau_c}\big(1-\varepsilon(X_k, Y_k)\big)\right] \le e^{-c}. \tag{5.10}$$

Since c is arbitrary, the lemma follows. □

Let B, $\tilde{B}$, ρ_1, ρ_2, ρ, α, η and $\tilde{\eta}$ satisfy the conditions of Theorem 3.2, and let $\sigma \in B_{\mathrm{sym}}(E \times E) \cap D(B^{(2)})$. Let $\tilde{A}_0$ be the generator for the neutral Fleming–Viot process in $P(\tilde{E})$ with mutation operator $\tilde{B}$ and recombination given by α and $\tilde{\eta}$, and let $\tilde{Z}$, defined on some (Ω, F, Q), be a solution of the martingale problem for $\tilde{A}_0$ with respect to a filtration $\{F_t\}$. Define $\tilde{\sigma}_i \in$

$B_{\mathrm{sym}}(\tilde{E} \times \tilde{E}) \cap D(\tilde{B}^{(2)})$ by $\tilde{\sigma}_i(x, y) = \sigma(\rho_i(x), \rho_i(y))$. Following Dawson (1978),

$$
\begin{aligned}
L_i(t) = \exp\Bigg\{\tfrac{1}{2}\Bigg(&\langle \tilde{\sigma}_i, \tilde{Z}(t)^2\rangle - \langle \tilde{\sigma}_i, \tilde{Z}(0)^2\rangle \\
&- \int_0^t \Big(\langle \Phi_{12}^{(2)}\tilde{\sigma}_i, \tilde{Z}(s)\rangle + \langle \tilde{B}^{(2)}\tilde{\sigma}_i, \tilde{Z}(s)^2\rangle \\
&\qquad + 2\alpha\langle H_1^{(2)}\tilde{\sigma}_i, \tilde{Z}(s)^3\rangle \\
&\qquad + 2\bar{\sigma}\langle K_1^{(2)}\tilde{\sigma}_i, \tilde{Z}(s)^4\rangle \\
&\qquad - (1 + 2\alpha + \bar{\sigma})\langle \tilde{\sigma}_i, \tilde{Z}(s)^2\rangle\Big)\, ds\Bigg)\Bigg\}
\end{aligned} \tag{5.11}
$$

is an $\{F_t\}$-martingale, and defining

$$
P_i(A) = \int_A L_i(t)\, dQ, \qquad A \in F_t, \tag{5.12}
$$

and extending P_i to $G = \bigvee_t F_t$ (enlarging Ω if necessary), $\tilde{Z}$ on (Ω, G, P_i) is a solution of the martingale problem for $\tilde{A}_i$, the generator of the Fleming–Viot process in $P(\tilde{E})$ with mutation operator $\tilde{B}$, recombination given by α and $\tilde{\eta}$, and selection intensity function $\tilde{\sigma}_i$. Note that for each t, there are constants $c(t)$ and $C(t)$ depending on σ, $B^{(2)}\sigma$ and α such that $c(t) \le L_i(t) \le C(t)$ a.s. for $i = 1, 2$.

Set $Z_i = \hat{\rho}_i \circ \tilde{Z}$ for $i = 1, 2$. Then, on (Ω, G, P_i), Z_i is a Fleming–Viot process with mutation operator B, recombination given by α and η and selection intensity function σ. Letting τ be the coupling time for Z_1 and Z_2,

$$
P_i\{Z_i(T) \in G\} \ge c(T)Q\{Z_i(T) \in G\} \ge c(T)Q\{Z_i(T) \in G, \tau \le T\}. \tag{5.13}
$$

Note that the right-hand side does not depend on i and is a nonzero measure in G for T sufficiently large. Let $\mu_i = Z_i(0)$ and define

$$
\nu_T(\mu_1, \mu_2, G) = Q\{Z_i(T) \in G | \tau \le T\} \tag{5.14}
$$

for such T and

$$
\varepsilon_T(\mu_1, \mu_2) = c(T)Q\{\tau \le T\} \tag{5.15}
$$

for all T. The following result is an immediate consequence of Lemma 5.1 and the estimate in (5.13).

PROPOSITION 5.2. *Let U and V be independent Fleming–Viot processes in $P(E)$ with mutation operator B, recombination given by α and η and selection intensity function σ. Suppose that a successful Markov coupling exists for the neutral process with mutation operator B and recombination given by α and η, that $\sigma \in B_{\mathrm{sym}}(E \times E) \cap D(B^{(2)})$ and that $\varepsilon_T(\mu_1, \mu_2)$ is defined by*

(5.15) *for all choices of* $\mu_1, \mu_2 \in P(E)$. *If for some* $T > 0$,

$$\sum_{k=0}^{\infty} \varepsilon_T(U(kT), V(kT)) = \infty \quad a.s., \tag{5.16}$$

then there exists a probability space on which are defined processes $\tilde{U}$ *and* $\tilde{V}$ *such that* U *and* $\tilde{U}$ *have the same distribution and* V *and* $\tilde{V}$ *have the same distribution, and there exists a random variable* λ *such that* $\tilde{U}(t) = \tilde{V}(t)$ *for all* $t \geq \lambda$.

PROOF. By Lemma 5.1, the "skeletons" $\{\tilde{U}(kT)\}$ and $\{\tilde{V}(kT)\}$ can be constructed. Then, using the Markov property, the skeletons can be interpolated. [See, e.g., Lemma 4.5.15 of Ethier and Kurtz (1986).] □

The condition in (5.16) is immediate under the assumptions of Theorem 3.2(c) [e.g., in the case defined by (3.1) and (3.2)], but in general this would appear to be a difficult problem. We can, however, use the estimates to give a very general proof of uniqueness of stationary distributions. First, we need the following lemma, which is part of the mythology of Markov chain theory.

LEMMA 5.3. *Let* $P(x, \Gamma)$ *be a one-step transition function on* E. *Suppose that* π_1 *and* π_2 *are stationary distributions for* P. *Then either* $\pi_1 = \pi_2$ *or there exist two mutually singular stationary distributions for* P.

REMARK. A referee has pointed out that a proof can be based on Theorem 12.1 and Remark 12.2 of Dynkin (1978). Instead, we give an essentially self-contained argument.

PROOF OF LEMMA 5.3. Let $\{X_k,\ k \geq 0\}$ denote the canonical process on E^∞. Let E_ν and P_ν denote expectations and probabilities on E^∞ for the Markov chain corresponding to $P(x, \Gamma)$ with initial distribution ν. Define T on $B(E)$ by $Tg(x) = \int_E g(y) P(x, dy)$.

Suppose that $\pi_1 \neq \pi_2$ and let $f \in B(E)$ satisfy $\int f\, d\pi_1 \neq \int f\, d\pi_2$. By the ergodic theorem,

$$H = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \tag{5.17}$$

exists P_{π_i}-a.s. for both $i = 1$ and $i = 2$. Let $\pi = \frac{1}{2}(\pi_1 + \pi_2)$. Then π is another stationary distribution. Note that H cannot be constant P_π-a.s., since

$$\langle f, \pi_i \rangle = E_{\pi_i}[H] = E_\pi\left[H \frac{d\pi_i}{d\pi}(X_0)\right], \tag{5.18}$$

and if H were constant P_π-a.s., the right-hand side would not depend on i. We claim that, given a Borel set B with $P_\pi\{H \in B\} > 0$, the measure π_B defined by

$$\pi_B(\Gamma) = \frac{E_\pi\left[I_\Gamma(X_0) I_{\{H \in B\}}\right]}{P_\pi\{H \in B\}} \tag{5.19}$$

is a stationary distribution for $P(x, \Gamma)$. Let

$$H^* = \limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k). \tag{5.20}$$

[Of course $P_\pi\{H = H^*\} = 1$.] Define

$$C_B = \left\{(x_0, x_1, \dots) \in E^\infty \colon \limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(x_k) \in B\right\} \tag{5.21}$$

and $L_B(x_0, x_1, \dots) = I_{C_B}(x_0, x_1, \dots)/P_\pi\{H \in B\}$. Note that C_B is invariant in the sense that $(x_0, x_1, \dots) \in C_B$ if and only if $(x_1, x_2, \dots) \in C_B$. We want to show that the invariance implies

$$E_\pi\left[L_B(X_0, X_1, \dots) | X_0\right] = E_\pi\left[L_B(X_0, X_1, \dots) | X_1\right]. \tag{5.22}$$

To see that this is the case, denote the left-hand side by $h(X_0)$ and observe that the right-hand side satisfies

$$E_\pi\left[L_B(X_0, X_1, \dots) | X_1\right] = E_\pi\left[L_B(X_1, X_2, \dots) | X_1\right] = h(X_1). \tag{5.23}$$

Then, using the invariance of C_B and the Markov property,

$$h(X_0) = E_\pi\left[L_B(X_1, X_2, \dots) | X_0\right] = E_\pi\left[h(X_1) | X_0\right] = Th(X_0). \tag{5.24}$$

Using the stationarity and (5.24), we have

$$\begin{aligned} &E_\pi\left[\left(h(X_1) - h(X_0)\right)^2\right] \\ &\quad = E_\pi\left[h^2(X_1)\right] + E_\pi\left[h^2(X_0)\right] - 2E_\pi\left[h(X_1)h(X_0)\right] \\ &\quad = 2E_\pi\left[h^2(X_0)\right] - 2E_\pi\left[Th(X_0)h(X_0)\right] \\ &\quad = 0 \end{aligned} \tag{5.25}$$

and hence (5.22). To see that π_B is a stationary distribution, note that $d\pi_B/d\pi = h$ and therefore, for each $g \in B(E)$,

$$\begin{aligned} E_{\pi_B}\left[g(X_1)\right] &= E_\pi\left[g(X_1)h(X_0)\right] \\ &= E_\pi\left[g(X_1)h(X_1)\right] \\ &= E_\pi\left[g(X_0)h(X_0)\right] \\ &= \langle g, \pi_B\rangle. \end{aligned} \tag{5.26}$$

Finally, observe that

$$
\begin{aligned}
L_B(X_0, X_1, \dots) &= \lim_{k\to\infty} E_\pi[L_B(X_0, X_1, \dots) | X_0, \dots, X_k] \\
&= \lim_{k\to\infty} E_\pi[L_B(X_k, X_{k+1}, \dots) | X_k] \\
&= \lim_{k\to\infty} h(X_k) \\
&= h(X_0), \qquad P_\pi\text{-a.s.}
\end{aligned}
\tag{5.27}
$$

Let c satisfy $0 < P_\pi\{H > c\} < 1$ and define $B_1 = (c, \infty)$ and $B_2 = (-\infty, c]$. Then C_{B_1} and C_{B_2} are disjoint, so $(L_{B_1} L_{B_2})(X_0, X_1, \dots) = 0$ a.s. By (5.27), there exist bounded measurable functions h_1, h_2 on E such that $h_1(X_0) = L_{B_1}(X_0, X_1, \dots)$ and $h_2(X_0) = L_{B_2}(X_0, X_1, \dots)$, P_π-a.s., which must satisfy $h_1 h_2 = 0$, π-a.s. It follows that π_{B_1} and π_{B_2} are mutually singular stationary distributions. □

LEMMA 5.4. *Let $P(x, \Gamma)$ be a one-step transition function on E and suppose π is a stationary distribution. Then, either π is ergodic in the sense that*

$$
\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \langle f, \pi \rangle, \qquad P_\pi\text{-}a.s., \ f \in B(E),
\tag{5.28}
$$

where $\{X_n\}$ is the canonical process on E^∞ and P_π is the Borel probability measure on E^∞ under which $\{X_n\}$ is a Markov chain with transition function P and initial distribution π, or there are mutually singular stationary distributions, π_1 and π_2, and $0 < \alpha < 1$ such that $\pi = \alpha\pi_1 + (1 - \alpha)\pi_2$.

REMARK. Note that (5.28) implies $\{X_k\}$ is ergodic under P_π in the usual sense of ergodic theory (invariant sets have probability 0 or 1).

PROOF OF LEMMA 5.4. If (5.28) fails to hold, then there exists f such that H^* defined in (5.20) is not constant P_π-a.s., and hence π_{B_1} and π_{B_2} constructed in the proof of Lemma 5.3 provide the desired stationary distributions with $\alpha = P_\pi\{H^* > c\}$. □

THEOREM 5.5. *Suppose that the conditions on B, α and η of Theorem 3.2 are satisfied. Let $\sigma \in B_{\mathrm{sym}}(E \times E) \cap D(B^{(2)})$ be a selection intensity function. Then there is at most one stationary distribution for the Fleming–Viot process determined by B, α, η and σ.*

PROOF. Suppose $\Pi_1, \Pi_2 \in P(P(E))$ are both stationary distributions. By Lemma 5.3, we can assume that Π_1 and Π_2 are mutually singular. However, constructing the coupling so that $Z_i(0)$ has distribution Π_i for $i = 1, 2$, (5.13) implies that

$$\Pi_i(G) \geq c(T) Q\{Z_i(T) \in G, \tau \leq T\}, \tag{5.29}$$

where the right-hand side is independent of i and is a nonzero measure for T sufficiently large, contradicting the singularity of Π_1 and Π_2. □

REFERENCES

ATHREYA, K. and NEY, P. E. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501.

ATHREYA, K., MCDONALD, D. and NEY, P. E. (1978). Coupling and the renewal theorem. *Amer. Math. Monthly* **85** 809–814.

DAWSON, D. (1978). Geostochastic calculus. *Canad. J. Statist.* **6** 143–168.

DAWSON, D. A. and HOCHBERG, K. J. (1982). Wandering random measures in the Fleming–Viot model. *Ann. Probab.* **10** 554–580.

DOEBLIN, W. (1940). Éléments d'une théorie générale des chaines simple constantes de Markoff. *Ann. Sci. École Norm. Sup.* (3) **57** 61–111.

DYNKIN, E. B. (1978). Sufficient statistics and extreme points. *Ann. Probab.* **6** 705–730.

ETHIER, S. N. (1992). Eigenstructure of the infinitely-many-neutral-alleles diffusion model. *J. Appl. Probab.* **29** 487–498.

ETHIER, S. N. and GRIFFITHS, R. C. (1990). The neutral two-locus model as a measure-valued diffusion. *Adv. in Appl. Probab.* **22** 773–786.

ETHIER, S. N. and GRIFFITHS, R. C. (1993). The transition function of a Fleming–Viot process. *Ann. Probab.* **21** 1571–1590.

ETHIER, S. N. and KURTZ, T. G. (1981). On the infinitely-many-neutral-alleles diffusion model. *Adv. in Appl. Probab.* **13** 429–452.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

ETHIER, S. N. and KURTZ, T. G. (1987). The infinitely-many-alleles model with selection as a measure-valued diffusion. *Stochastic Models in Biology. Lecture Notes in Biomath.* **70** 72–86. Springer, Berlin.

ETHIER, S. N. and KURTZ, T. G. (1993). Fleming–Viot processes in population genetics. *SIAM J. Control Optim.* **31** 345–386.

ETHIER, S. N. and KURTZ, T. G. (1994). Convergence to Fleming–Viot processes in the weak atomic topology. *Stochastic Process. Appl.* **54** 1–27.

FUKUSHIMA, M. and STROOCK, D. W. (1986). Reversibility of solutions to martingale problems. In *Probability, Statistical Mechanics, and Number Theory. Advances in Mathematics Supplementary Studies* (G.-C. Rota, ed.) **9** 107–123. Academic Press, Orlando, FL.

GRIFFEATH, D. (1976). Partial coupling and loss of memory for Markov chains. *Ann. Probab.* **4** 850–858.

GRIFFEATH, D. (1978). Coupling methods for Markov processes. In *Studies in Probability and Ergodic Theory* 1–43. Academic Press, New York.

GRIFFITHS, R. C. (1979). A transition density expansion for a multi-allele diffusion model. *Adv. in Appl. Probab.* **11** 310–325.

KINGMAN, J. F. C. (1975). Random discrete distributions. *J. Roy. Statist. Soc. Ser. B* **37** 1–22.

LIGGETT, T. M. (1985). *Interacting Particle Systems.* Springer, Berlin.

LINDVALL, T. (1983). On coupling diffusion processes. *J. Appl. Probab.* **20** 82–93.

LINDVALL, T. and ROGERS, L. C. G. (1986). Coupling of multidimensional diffusions by reflection. *Ann. Probab.* **14** 860–872.

NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete* **43** 309–318.

NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-negative Operators.* Cambridge Univ. Press.

SHIGA, T. (1981). Diffusion processes in population genetics. *J. Math. Kyoto Univ.* **21** 133–151.

SHIGA, T. (1982). Wandering phenomena in infinite-allelic diffusion models. *Adv. in Appl. Probab.* **14** 457–483.

SHIGA, T. (1990). A stochastic equation based on a Poisson system for a class of measure-valued diffusion processes. *J. Math. Kyoto Univ.* **30** 245–279.

VASERSHTEIN, L. N. (1969). Markov processes over denumerable products of spaces describing large systems of automata. *Probablemy Peredachi Informatsii* **5** 64–73.

WATTERSON, G. A. (1977). Heterosis or neutrality? *Genetics* **85** 789–814.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
SALT LAKE CITY, UTAH 84112
E-MAIL: ethier@math.utah.edu

DEPARTMENTS OF MATHEMATICS AND STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53706
E-MAIL: kurtz@math.wisc.edu