The Annals of Probability
1982, Vol. 10, No. 3, 554–580

WANDERING RANDOM MEASURES IN THE FLEMING-VIOT MODEL

By Donald A. Dawson[1] and Kenneth J. Hochberg[2]

Carleton University and Case Western Reserve University

Fleming and Viot have established the existence of a continuous-state-space version of the Ohta-Kimura ladder or stepwise-mutation model of population genetics for decribing allelic frequencies within a selectively neutral population undergoing mutation and random genetic drift. Their model is given by a probability-measure-valued Markov diffusion process. In this paper, we investigate the qualitative behavior of such measure-valued processes. It is demonstrated that the random measure is supported on a bounded generalized Cantor set and that this set performs a "wandering" but "coherent" motion that, if appropriately rescaled, approaches a Brownian motion. The method used involves the construction of an interacting infinite particle system determined by the moment measures of the process and an analysis of the function-valued process that is "dual" to the measure-valued process of Fleming and Viot.

1. Introduction. In a recent paper, Fleming and Viot (1979) have introduced a probability-measure-valued stochastic process in a variation of a model for the distribution of allelic frequencies in a selectively neutral genetic population. In this paper, we introduce techniques for the study of the qualitative behavior of such processes. We then proceed to analyze in detail the local structure and qualitative behavior of the Fleming-Viot model.

The paper is organized as follows. In Sections 2 and 3, we outline the foundations of the theory of probability-measure-valued processes and introduce the qualitative notions of microscopic clustering and macroscopic coherence for such processes. Sections 4 and 5 are devoted to a summary of several relevant models in population genetics, including the Ohta-Kimura stepwise-mutation model and its continuous-state-space analogue, the Fleming-Viot model. Section 5 also contains a description of the function-valued process that is dual to the measure-valued Fleming-Viot process. In Section 6, we introduce two basic tools used in our analysis of these processes, namely, an infinite system of interacting particles which arises from the partial differential equations satisfied by the moment densities, and the empirical moments of the random distributions. Section 7 establishes the principal qualitative features of the Fleming-Viot model, including the tendency to cluster at microscopic scales and the long-term coherence of the random distributions. Finally, in Section 8 we complete the qualitative description by identifying the asymptotic behavior of the wandering random probability distribution at large time scales.

2. Mathematical formulation of probability-measure-valued processes. We begin by considering $S = R^d \cup \{\infty\}$, the one-point compactification of R^d, $\mathscr{B}(S)$, the σ-algebra of Borel subsets of S, and the space $M_1(S)$ of probability measures on S furnished with the topology of weak convergence of probability measures. $M_1(S)$ serves as the state-space for the family of probability-measure-valued Markov processes. Let $\Omega = C([0, \infty),$

Received February 1981; revised August 1981.

[1] Research supported in part by the Natural Sciences and Engineering Research Council of Canada. Part of the research was carried out while the author was a visitor at the University of Heidelberg and the University of Wisconsin at Madison.

[2] Research supported in part by U.S. National Science Foundation grant MCS78-02144 and Public Health Service grant GM28336-01.

Key words and phrases. Measure-valued Markov process, random measure, Hausdorff dimension, ladder or stepwise-mutation model, population genetics, Fleming-Viot model, wandering coherent distribution.

AMS 1980 *subject classifications.* Primary 60G57, 60J70, 60J25; secondary 60K35, 92A15.

$M_1(S)$) and $\Omega^D = D([0, \infty), M_1(S))$, the spaces of functions mapping $[0, \infty)$ into $M_1(S)$ that are, respectively, continuous and right continuous with limits from the left. We consider the canonical process $X:[0, \infty) \times \Omega \to M_1(S)$ (or alternatively, $X:[0, \infty) \times \Omega \times \mathscr{B}(S) \to [0, 1]$) defined by $X(t, \omega, A) \equiv \omega(t, A)$ for $A \in \mathscr{B}(S)$, $\omega \in \Omega$, $t \geqslant 0$. The *distribution* of a *probability-measure-valued diffusion process* is determined by a mapping $\mu \to P_\mu$ from $M_1(S)$ into $\Pi(\Omega)$, the space of probability measures on Ω.

The stochastic processes described below are characterized as the unique solutions of measure-valued martingale problems on Ω in the sense of Stroock and Varadhan (1979). A *martingale problem* on Ω is prescribed by a pair $(L, \mathscr{D}(L))$, where L is a linear operator defined on the linear subspace $\mathscr{D}(L)$ of $C(M_1(S))$. A *solution* is the distribution $\{P_\mu : \mu \in M_1(S)\}$ of a stochastic process which satisfies the conditions

$$P_\mu(X(0) = \mu) = 1, \quad \text{and} \tag{2.1}$$

$$\text{for} \quad \psi \in \mathscr{D}(L),\ \psi(X(t)) - \int_0^t L\psi(X(s))\, ds \quad \text{is a } P_\mu\text{-martingale of each } \mu \in M_1(S). \tag{2.2}$$

It can be shown (see Fleming and Viot, 1979, page 835) that a *unique* solution of such a martingale problem defines a Markov, Feller diffusion process with state-space $M_1(S)$.

The operators associated with diffusion processes with values in $M_1(S)$ have the form

$$L\psi(\mu) \equiv \int_S A(\delta\psi(\mu)/\delta\mu(x))\mu(dx) + \int_S \int_S (\delta^2\psi(\mu)/\delta\mu(x)\delta\mu(y))Q(\mu : dx \times dy), \tag{2.3}$$

where

$$\delta\psi(\mu)/\delta\mu x) \equiv \lim_{\varepsilon \downarrow 0}(\psi(\mu + \varepsilon\delta_x) - \psi(\mu))/\varepsilon,$$

$$Q : M_1(S) \to M_1(S \times S) \quad \text{(quadratic fluctuation functional).}$$

A is the infinitesimal generator of a strongly continuous Markov semigroup on $C_0(R^d)$, where $C_0(R^d)$ is the space of continuous functions on R^d which vanish at ∞, and δ_x represents a unit mass at the point $x \in R^d$.

3. Qualitative properties of $M_1(S)$-valued stochastic processes. At a fixed time t, an $M_1(S)$-valued stochastic process $X(t, dx)$ is described by a random probability measure. We introduce the notion of *clustering* to describe the structure of a random measure at microscopic scales. The concept of microscopic clustering can be intuitively described as follows. Consider a population of N_0 individuals distributed in a cube $V \subset R^d$ which is subdivided into Γ^d congruent disjoint subcubes. We assume that $N_0 \gg \Gamma^d$ and count the number of subcubes $N(\Gamma)$ that are occupied. If the distribution is uniformly random, that is, Poisson, then the number of occupied subcubes is of the order of Γ^d. If $N(\Gamma)/\Gamma^d \ll 1$, then the population exhibits a high degree of clustering. The phenomenon of clustering is seen dramatically by considering the continuous diffusion limit and identifying the size of the set on which the measure is concentrated. In order to give a precise formulation to the idea of the size of such carrying sets, we introduce the notion of the *Hausdorff-Besicovitch dimension of support* of a random measure.

Given a bounded Borel set $E \subset R^d$ and $\beta > 0$, $\delta > 0$, let

$$\Lambda_\delta^\beta(E) \equiv \inf_{\mathscr{S}} \textstyle\sum_i (d(S_i))^\beta, \tag{3.1}$$

where $d(S_i)$ is the diameter of the set S_i and $\mathscr{S} \equiv \{\{S_i\} : E \subset \cup_i S_i,\ d(S_i) < \delta \text{ for each } i\}$. Then the Hausdorff β-measure of E is defined by

$$\Lambda^\beta(E) = \lim_{\delta \to 0} \Lambda_\delta^\beta(E). \tag{3.2}$$

The *Hausdorff-Besicovitch dimension* of E is defined by

$$\dim E \equiv \inf\{\beta > 0 : \Lambda^\beta(E) = 0\} = \sup\{\beta > 0 : \Lambda^\beta(E) = \infty\}. \tag{3.3}$$

Note that $0 \leq \dim E \leq d$, and if E has positive Lebesgue measure, then $\dim E = d$. For comprehnsive references on Hausdorff measures and Hausdorff-Besicovitch dimension, refer to Federer (1969) and Rogers (1970).

Now consider a unit cube $V \subset R^d$ which for each $n \geq 1$ is subdivided into Γ_n^d equal subcubes of volume Γ_n^{-d}, where $\{\Gamma_n : n \geq 1\}$ is an increasing sequence of non-negative integers. The ratio of the diameter of the subcubes to that of the fixed cube V is Γ_n^{-1}.

Consider the set B obtained as follows:

$$B_0 = V, \qquad B_n \subset B_{n-1} \quad \text{for} \quad n \geq 1,$$

B_n is a union of N_n disjoint subcubes of volume Γ_n^{-d}, and

$$B \equiv \cap_{n=0}^{\infty} B_n.$$

Then B is said to be a *generalized Cantor set.*

LEMMA 3.1. *For the set B defined above,*

$$\dim B \leq \liminf_{n\to\infty}(\log N_n/\log \Gamma_n). \tag{3.4}$$

PROOF. If $\varepsilon > 0$ and

$$\beta \equiv (1+\varepsilon) \liminf_{n\to\infty}(\log N_n/\log \Gamma_n),$$

then

$$\log \Lambda^{\beta}_{\Gamma_n^{-1}}(B) \leq -\varepsilon \log N_n + c_d,$$

where c_d is a finite constant. Therefore, letting $n \to \infty$, we have

$$\Lambda^{\beta}(B) = 0.$$

Hence

$$\dim B \leq (1+\varepsilon) \liminf_{n\to\infty} (\log N_n/\log \Gamma_n).$$

Now let Y be a random measure on V. Given $\varepsilon > 0$, let

$$N_n^{\varepsilon}(Y) = \min\{n : \textstyle\sum_{i=1}^{n} Y(v_i) \geq Y(V) - \varepsilon\}, \tag{3.5}$$

and

$$K_n^{\varepsilon} \equiv \cup_{i=1}^{N_n^{\varepsilon}(Y)} v_i, \tag{3.6}$$

where $\{v_i : i = 1, 2, \cdots, N_n^{\varepsilon}(Y)\}$ is a disjoint cover consisting of the given subcubes of volume Γ_n^{-d} achieving the minimum in (3.5).

LEMMA 3.2. (Dawson and Hochberg, 1979). *Assume that*

$$P((\log N_n^{\varepsilon_n}(Y)/\log \Gamma_n) \leq \rho(1+\eta_n)) \geq 1 - \varepsilon_n', \tag{3.7}$$

where $\varepsilon_n \downarrow 0$, $\eta_n \downarrow 0$, *and* $\varepsilon_n' \downarrow 0$ *as* $n \to \infty$. *Then there exists a random generalized Cantor set* $B(\omega)$ *such that*

$$Y(\omega, B(\omega)) = Y(\omega, V), \quad \text{a.e. } \omega, \tag{3.8}$$

and

$$\dim B(\omega) \leq \rho, \quad \text{a.e. } \omega. \tag{3.9}$$

The smallest ρ satisfying (3.8) and (3.9) is said to be the *Hausdorff-Besicovitch dimension of support* of the random measure Y on the set V.

The long-term behavior of an $M_1(S)$-valued stochastic process $X(t, dx)$ can be classified as either *coherent* or *dispersive*. The process is said to be *coherent* if for every $\varepsilon > 0$ there exists t_0, $0 \leq t_0 < \infty$, with the property that for each $t \geq t_0$, there is a random sphere $S_\varepsilon(t)$ with center $x(t)$ and radius $R_\varepsilon(t)$ which satisfies

$$x(t) = \int_{R^d} xX(t, dx), \quad \text{and} \quad \int_{R^d} |x| X(t, dx) < \infty, \tag{3.10}$$

$$P(X(t, \omega, S_\varepsilon(t, \omega)) \geq 1 - \varepsilon) = 1, \tag{3.11}$$

and

$$R_\varepsilon(t) \quad \text{is a stationary stochastic process.} \tag{3.12}$$

The *wandering motion* of a coherent distribution can be described by the process $\{x(t) : t \geq t_0\}$. A process is said to be *compactly coherent* if (3.11) is also valid for $\varepsilon = 0$, with some centering, not necessarily that prescribed by (3.10). A process that is not coherent is said to be *dispersive*.

4. Stochastic models in population genetics. Consider a finite population of N individuals, with each individual having d observable numerical characteristics (e.g. height, weight, shade of eye color, shade of hair color, shoe size, etc.). We assume that these characteristics are measured in discrete units as integral multiples of some standard unit m. The "type" of an individual is determined by the set of the d characteristics; thus, an individual's type is represented by a vector $k = (k_1, \cdots, k_d) \in Z^d$. Let $n_k(t)$, $k \in Z^d$, be the number of individuals of type k in the population at time $t \geq 0$. Let $\bar{Z}^d$ denote the one-point compactification of Z^d, and $M_1(\bar{Z}^d)$ the set of probability measures on $\bar{Z}^d$. For $t \geq 0$, let $p(t) \in M_1(\bar{Z}^d)$ be defined by

$$p(t;k) \equiv n_k(t)/N, \quad k \in \bar{Z}^d. \tag{4.1}$$

We assume that $\{p(t) : t \geq 0\}$ is an $M_1(\bar{Z}^d)$-valued continuous-time Markov process with generator given by

$$L_N\psi(p) = \textstyle\sum_{i \neq j} g_{ij}(p)[\psi(p^{ij}) - \psi(p)], \tag{4.2}$$

where $p \in M_1(\bar{Z}^d)$ and $\psi \in C(M_1(\bar{Z}^d))$, the space of continuous functions on $M_1(\bar{Z}^d)$, and

$$\begin{aligned} p^{ij}(k) &= p(k) - 1/N, \quad \text{if} \quad k = i, \\ &= p(k) + 1/N, \quad \text{if} \quad k = j, \\ &= p(k), \qquad\qquad \text{if} \quad k \neq i, j. \end{aligned} \tag{4.3}$$

The coefficient $g_{ij}(p)$ denotes the rate at which an individual of type i is replaced by an individual of type j, given that the distribution of the population is given by p. The *continuous-time Ohta-Kimura model* for a randomly mating population of Moran type with stepwise mutation is given by the coefficients

$$g_{ij}(p) = [\gamma p(i)p(j) + Dp(i)\theta_{ij}], \tag{4.4}$$

where $\gamma > 0$, $D > 0$, and

$$\begin{aligned} \theta_{ij} &= 1, \qquad \text{if} \quad |i - j| = 1, \\ &= -2d, \quad \text{if} \quad i = j, \\ &= 0, \qquad \text{otherwise}. \end{aligned} \tag{4.5}$$

Note that the point ∞ is assumed to be an absorbing point for the mutation process but a regular point for the sampling process. In (4.4), γ represents the *sampling rate* and D represents the *mutation rate*.

Ohta and Kimura (1973, 1974) first introduced their model, known as the "ladder" or "stepwise-mutation" model, to describe the distribution of allelic frequencies in a large but finite population having a large number of possible genetic states. The model was formulated to describe the distribution of allelic types distinguishable as signed electrical charges in electrophoretic experiments. It should be noted that recently there has been some controversy regarding the reliability of electrophoretic techniques in classifying charge-states according to the stepwise-mutation model, as evidenced, for example, by the recent contrasting papers of Ramshaw, Coyne and Lewontin (1979) and Fuerst and Ferrell (1980) on electrophoretic detection of hemoglobin variants. The objections are aimed at the efficacy of electrophoresis for detection, however, and not at the mathematical formulation of the stepwise-mutation model itself. Additional studies and variations of the Ohta-Kimura stepwise-mutation model can be found in Brown, Marshall and Albrecht (1975), Chakraborty and Nei (1976), and Kimura and Crow (1975, 1978). For a complete discussion of stochastic models for selectively neutral allelic populations, refer to the book of Ewens (1979).

Moran (1975, 1976) studied the behavior of the Ohta-Kimura model in the limit of large population ($N \to \infty$) and small mutation rate ($D = O(1/N)$). He observed that although the Markov process has no stationary distribution, the relative genetic differences of randomly chosen pairs of individuals tends to a steady state. Kingman (1976) established that the joint distribution of the relative differences of the genetic states, measured from a randomly chosen one, does converge to a limit. Thus, the distribution of alleles is "coherent" in the sense that it tends to aggregate into a cluster rather than spread out like a pure diffusion process, and the cluster itself tends to wander throughout the set of allelic states. These results imply that for a fixed finite population size N, there is a limiting random number of types $\Lambda(N)$ as $t \to \infty$. Kesten (1980) proved that, in the case $D = O(1/N)$, $\Lambda(N)$ tends to ∞ with N but at an extremely slow rate. This phenomenon also appears in the "infinite alleles model" of Kimura and Crow (1964). In this model there is no spatial structure, and each mutation gives rise to a completely new type. A stationary distribution giving the limiting relative frequencies of countably many types exists and is known as the Poisson-Dirichlet distribution (refer to Kingman (1975)), Watterson (1976) and Ethier and Kurtz (1981)).

5. The Fleming-Viot model.

5.1 *The basic existence theorem.* In the studies of Moran, Kingman and Kesten on the Ohta-Kimura model, the limiting behavior was analyzed under the assumptions that the mutation rate is inversely proportional to the population size and the incremental effect on the numerical characteristics due to a single mutation remains constant. Fleming and Viot (1978, 1979) have introduced an alternative limiting form of the Ohta-Kimura stepwise mutation model in which the mutation rate is constant, but the incremental effect of a single mutation is assumed to decrease at a rate inversely proportional to the square root of the population size. In order to obtain convergence to a nondegenerate diffusion under the corresponding rescaling of the state space, the process must be observed at an appropriately adjusted time scale. The Ohta-Kimura process $\{p(t, \cdot): t \geq 0\}$ rescaled in space and time leads to the $M_1(S)$-valued process

$$Y_N(t, A) \equiv \textstyle\sum_{j/N^{1/2} \in A} p(N^2 t, j). \tag{5.1}$$

The generator L_n of the rescaled process Y_N has the following form:

(5.2) $\mathscr{D}(L_N)$ is the linear space of functions of the form $\psi(p) = f(\langle \phi, p \rangle)$ where $f \in C^3(R^1)$, $\phi \in C_K^3(R^d)$, $\langle \phi, p \rangle = \sum_j \phi(x_j) p(j)$, and $x_j \equiv j/N^{1/2}$,

$$\begin{aligned} L_N\psi(p) &= \textstyle\sum_{i,j} N^2[f(\langle \phi, p \rangle - \phi(x_i)/N + \phi(x_j)/N) - f(\langle \phi, p \rangle)] \\ &\qquad \cdot (\gamma p(i) p(j) + D p(i) \theta_{ij}) \\ &= N^2 D f'(\langle \phi, p \rangle)(\textstyle\sum_i p(i)[\sum_{|i-j|=1} \phi(x_j) - 2d\phi(x_i)]/N) \\ &\qquad + \tfrac{1}{2}\gamma f''(\langle \phi, p \rangle)(\textstyle\sum_{i,j} p(i)p(j)[\phi^2(x_i) + \phi^2(x_j) - 2\phi(x_i)\phi(x_j)]) \\ &\qquad + (1/N) R(\phi), \end{aligned} \tag{5.3}$$

where $R(\phi)$ is a bounded remainder term if $|f'''| \leq M < \infty$. Hence,

$$L_N\psi(p) = Df'(\langle \phi, p \rangle)(\langle \Delta\phi, p \rangle) + \gamma f''(\langle \phi, p \rangle)[\langle \phi^2, p \rangle - \langle \phi, p \rangle^2] + (1/N^{1/2}) R(\phi), \tag{5.4}$$

where Δ denotes the d-dimensional Laplace operator.

In order to identify the limiting Fleming-Viot process, we require an operator which is formally obtained as a limit of the operators L_N. Let $\mathscr{D}(L)$ be the set of functions of the form

$$\psi(\mu) = f(\langle\phi_1, \mu\rangle, \cdots, \langle\phi_n, \mu\rangle), \tag{5.5}$$

where $\langle\phi, \mu\rangle \equiv \int \phi(x)\mu(dx)$, $f(y_1, \cdots, y_n) \in C^2(R^n)$, and $\phi_i \in C_K^2(R^d)$, the space of twice continuously differentiable functions on R^d with compact support. The linear operator L is defined by

$$\begin{aligned} L\psi(\mu) = {} & D\textstyle\sum_{i=1}^n f_{y_i}(\langle\phi_1, \mu\rangle, \cdots, \langle\phi_n, \mu\rangle)\,\langle\Delta\phi_i, \mu\rangle \\ & + \gamma \textstyle\sum_{j=1}^n \sum_{i=1}^n f_{y_i y_j}(\langle\phi_1, \mu\rangle, \cdots, \langle\phi_n, \mu\rangle)[\langle\phi_i\phi_j, \mu\rangle - \langle\phi_i, \mu\rangle\langle\phi_j, \mu\rangle], \end{aligned} \tag{5.6}$$

where f_{y_i}, $f_{y_i y_j}$ denote the first and second partial derivatives of f. The operator L given by (5.6) is of the form (2.3) with

$$A = \Delta, \qquad \text{and} \tag{5.7}$$

$$Q(\mu{:}\, dx \times dy) = \mu(dx)\delta_x(dy) - \mu(dx)\mu(dy). \tag{5.8}$$

THEOREM 5.1. (*Fleming and Viot*, 1979).
(a) *The martingale problem associated with* $(L, \mathscr{D}(L))$ *given by* (5.5) *and* (5.6) *has a unique solution on* Ω. *Let the* $M_1(S)$*-valued Markov diffusion process determined by this solution be referred to as the Fleming-Viot process and be denoted by* $X(\cdot, \cdot)$.
(b) *Let* $Y_N(\cdot, \cdot)$ *be given as in* (5.1). *Then* $Y_N(\cdot, \cdot) \Rightarrow X(\cdot, \cdot)$ *as* $N \to \infty$, *where* $\Rightarrow$ *denotes convergence in the sense of weak convergence of probability measures on* Ω^D.

The remainder of this paper is devoted to the study of the qualitative features of the Fleming-Viot process and its associated random measure. It follows from Theorem 5.1(b) that the Fleming-Viot process can be interpreted as a *diffusion approximation* to the Ohta-Kimura model in the case of small incremental mutational effect observed in an appropriate time scale. In this respect, the qualitative results for the Fleming-Viot model have implications for the Ohta-Kimura model itself.

5.2. *The dual process.* In this section we introduce a function-valued Markov process which is the dual of the measure-valued Fleming-Viot process and which is used in the proof of Lemma 6.9. We first introduce an algebra $\mathscr{P}(M_1(S))$ of functions on $M_1(S)$ consisting of polynomials with bounded coefficients. $\mathscr{P}(M_1(S))$ is defined to be the smallest algebra of functions on $M_1(S)$ which contains all functions of the form

$$F_f(\mu) = \int_S \cdots \int_S f(x_1, \cdots, x_N)\mu(dx_1) \cdots \mu(dx_N) \tag{5.9}$$

where $f \in C(S^N)$, the space of continuous functions on S^N. Equation (5.9) also defines a family of functions $\mathscr{Q}(\mathscr{C})$ on $\mathscr{C} \equiv \cup_{N=1}^\infty C(S^N)$ as follows:

$$F_\mu(f) \equiv F_f(\mu).$$

The space $\mathscr{C}$ is furnished with the topology given by the inductive limit of the supremum norm topologies.

An alternative formulation of the Fleming-Viot martingale problem is given by the following pair $(L, \mathscr{D}_p(L))$:

$$\mathscr{D}_p(L) \equiv \{F_f{:} f \in \cup_{N=1}^\infty C^\infty(S^N)\} \tag{5.10}$$

where $C^\infty(S^N)$ denotes the set of infinitely differentiable functions on S^N, and, for $f \in C^\infty(S^N)$,

$$LF_f(\mu) \equiv \int \cdots \int Kf(x_1, \cdots, x_N)\mu(dx_1) \cdots \mu(dx_N), \tag{5.11}$$

where

$$Kf(x_1, \cdots, x_N) \equiv D \sum_{j=1}^N \Delta_j f(x_1, \cdots, x_N) + \gamma \sum_{j=1}^N \sum_{k=1, k\neq j}^N [\Phi_{jk}(f(x_1, \cdots, x_N)) - f(x_1, \cdots, x_N)]. \tag{5.12}$$

Here Δ_j denotes the action of the d-dimensional Laplacian acting on the variable x_j, and Φ_{jk} is a mapping from $C(S^N)$ to $C(S^{N-1})$ given by

$$(\Phi_{jk} f)(y_1, \cdots, y_{N-1}) \equiv f(x_1, \cdots, x_N) \tag{5.13}$$

where for $j \neq k$,

$$\begin{aligned} x_i &= y_i \quad \text{for} \quad i = 1, \cdots, (k-1), \\ x_k &= y_j \quad \text{if} \quad j < k, \\ x_k &= y_{j-1} \quad \text{if} \quad j > k, \\ x_i &= y_{i-1} \quad \text{if} \quad i = (k+1), \cdots, N. \end{aligned}$$

Equation (5.11) also defines a linear operator $L^{\#}$ on $\mathscr{D}(\mathscr{C})$ such that

$$L^{\#} F_\mu(f) \equiv L F_f(\mu). \tag{5.14}$$

The operator $L^{\#}$ agrees with the infinitesimal generator of a Markov process with state space $\mathscr{C}$, which we denote by $\{\eta_t : t \geq 0\}$ and is referred to as a *dual process* (not unique) (cf. Holley, Stroock and Williams (1977)). The process η_t evolves as follows:

(a) by jumps from $C(S^N)$ to $C(S^{N-1})$ if $N \geq 2$,
(b) at the time of a jump from $C(S^N)$ to $C(S^{N-1})$, a pair $\{j, k\}$ is picked at random from $\{1, \cdots, N\}$ and f is replaced by $\Phi_{jk} f$,
(c) between jumps it is deterministic on $C(S^N)$ and evolves according to the heat semigroup on $(R^d)^N$, $N \geq 1$,
(d) if it is in $C(S)$, that is, a function of one variable, no further jumps occur.

LEMMA 5.1. *Let* $\tau \equiv \inf\{t : \eta_t \in C(S)\}$. *Then*

$$P_f(\tau < \infty) = 1 \quad \textit{for any} \quad f \in \mathscr{C}. \tag{5.15}$$

PROOF. This follows from elementary facts about pure death processes.

Let $\{T_t : t \geq 0\}$ denote the semigroup of operators on $C(M_1(S))$ associated with the Fleming-Viot process and $\{U_t : t \geq 0\}$ denote the semigroup of operators on $\mathscr{D}(\mathscr{C})$ associated with the dual process $\{\eta_t : t \geq 0\}$. Then for $f \in C(S^N)$, the duality relationship is expressed as:

$$\begin{aligned} T_t F_f(\mu) &\equiv E_\mu\left(\int \cdots \int f(x_1, \cdots, x_N) X(t, dx_1) \cdots X(t, dx_N)\right) \\ &= E_f\left(\int \cdots \int \eta_t(x_1, \cdots, x_{N(t)}) \mu(dx_1) \cdots \mu(dx_{N(t)})\right) \\ &\equiv U_t F_\mu(f), \end{aligned} \tag{5.16}$$

where $N(t) \equiv n$ if $\eta_t \in C(S^n) \backslash C(S^{n-1})$.

REMARK 5.1. With the use of a truncation and limit argument, it can be verified that the relationship (5.16) remains valid for $f \in C_p((R^d)^N)$, the space of continuous functions with at most polynomial growth at infinity.

REMARK 5.2. The duality relationship (5.16) provides a proof of the uniqueness of the solution to the martingale problem $(L, \mathscr{D}_p(L))$.

6. The moment system.

6.1 *Moment measures and the canonical representation.* The distribution of a random probability measure X on S is given by a probability law P on the measure space $(M_1(S), \mathscr{B}(M_1(S)))$ where $\mathscr{B}(M_1(S))$ denotes the σ-algebra of Borel subsets of $M_1(S)$.

The *kth moment measure* $M_k(dx_1, \cdots, dx_k)$ is a probability measure on S^k which satisfies the equality

$$E(\textstyle\prod_{j=1}^{k} \langle \phi_j, X\rangle) = \displaystyle\int_S \cdots \int_S \phi_1(x_1) \cdots \phi_k(x_k) M_k(dx_1, \cdots, dx_k), \tag{6.1}$$

where E denotes the expectation operator and $\phi_i \in C(S)$.

Note that $\{M_k : k \geq 1\}$ form a consistent family of probability measures, that is,

$$M_k(dx_1, \cdots, dx_{j-1}, S, dx_{j+1}, \cdots, dx_k) = M_{k-1}(dx_1, \cdots, dx_{j-1}, dx_{j+1}, \cdots, dx_k). \tag{6.2}$$

Also, for each k, $M_k(\cdot, \cdots, \cdot)$ is an *exchangeable* probability law on S^k; that is,

$$M_k(A_1, \cdots, A_k) = M_k(A_{\pi(1)}, \cdots, A_{\pi(k)})$$

for every permutation π. It follows from Kolmogorov's extension theorem that there exists a probability measure P^* on $(\Omega^*, \mathscr{F}^*)$, where $\Omega^* = S^\infty$ and $\mathscr{F}^*$ denotes the P^*-completion of the product σ-algebra, such that

$$P^*(A_1 \times A_2 \times \cdots \times A_k) = M_k(A_1, A_2, \cdots, A_k) \tag{6.3}$$

for $A_1, \cdots, A_k \in \mathscr{B}(S)$. In turn, P^* is the probability law of a sequence of exchangeable S-valued random variables $\{Z_k : k \geq 1\}$. The random variables Z_k can be viewed as the locations of a countable collection of particles in S.

A permutation π of the non-negative integers is said to be finite if $\pi(n) = n$ for all but finitely many n. Given such a π, there is an induced mapping π^* defined on Ω^* by $\pi^*\omega^*(n) = \omega^*(\pi(n))$ for $\omega^* \in \Omega^*$. Let $Q = \{\pi^* : \pi \text{ finite}\}$. The sub-$\sigma$-algebra of $\mathscr{F}^*$ defined by

$$\mathscr{E} \equiv \{B : B \in \mathscr{F}^*, P^*(\pi^* B \underline{\Delta} B) = 0 \quad \text{for all} \quad \pi^* \in Q\},$$

where $\underline{\Delta}$ denotes the symmetric difference, is known as the *σ-algebra of exchangeable events.*

Let $\mathscr{A}$ denote the smallest algebra of subsets of S containing $\{\infty\}$ and the countable collection of rectangles of the form

$$\textstyle\prod_{j=1}^{d} [a_j, b_j) \quad \text{with} \quad a_j, b_j \text{ rational numbers.}$$

Recall that $\sigma(\mathscr{A}) = \mathscr{B}(S)$.

We now construct a canonical mapping $Y : \Omega^* \to M_1(S)$. For $\omega^* \in \Omega^*$, $A \in \mathscr{A}$, we define

$$\begin{aligned} Y(\omega^*, A) &\equiv \lim_{n\to\infty} n^{-1}[\textstyle\sum_{j=1}^{n} I_A(Z_j(\omega^*))], \quad \text{if the limit exists for all} \quad A \in \mathscr{A}, \\ &\equiv \delta_0, \quad \text{otherwise,} \end{aligned} \tag{6.4}$$

where I_A denotes the indicator function of the set A. The following lemma establishes the fact that $Y^*(\cdot, \cdot)$ can be extended to be a version of the random measure X. It is based on the circle of ideas contained in de Finetti's theorem on exchangeable random variables.

LEMMA 6.1.
(a) *The limit in* (6.4) *exists for all* $A \in \mathscr{A}$ *for* P^**-almost every* ω^*.
(b) *For every* $\omega^* \in \Omega^*$, $Y(\omega^*, \cdot)$ *has a unique extension to a probability measure on* S.

(c) *The mapping* $\omega^* \to Y(\omega^*, A)$ *is measurable with respect to the* σ-*algebra* $\mathscr{E}$ *for every set* $A \in \mathscr{B}(S)$.

(d) *The mapping* $Y: \Omega^* \to M_1(S)$ *is measurable with respect to the* σ-*algebra* $\mathscr{E}$.

(e) *For* $\omega^* \in \Omega^*, A \in \mathscr{B}(S)$,

(6.5) $$Y(\omega^*, A) = P^*(Z_1 \in A \mid \mathscr{E})(\omega^*), \quad P^*\text{-a.s.},$$

that is, $Y(\cdot\,,\,\cdot)$ *is a regular conditional probability given the* σ-*algebra* $\mathscr{E}$. *More generally, for* $A_k \in \mathscr{B}(S)$,

(6.6) $$P^*(Z_k \in A_k, k = 1, 2, 3, \cdots \mid \mathscr{E}) = \prod_{k=1}^{\infty} Y_k(\omega^*, A_k), \quad P^*\text{-a.s.},$$

where $Y_k(\omega^*, A) = Y(\omega^*, A)$ *for each* k.

(f) *For any set* $B \in \mathscr{B}(M_1(S))$,

(6.7) $$P^*(Y \in B) = P(X \in B),$$

that is, $Y(\cdot\,,\,\cdot)$ *is a random probability measure having the same distribution as* X.

PROOF. By de Finetti's theorem, the random variables Z_k, conditioned on $\mathscr{E}$, are identically distributed independent random variables (a complete discussion of de Finetti's theorem can be found in Chow and Teicher (1978, page 220)). Part (a) follows from the law of large numbers for bounded identically distributed independent random variables.

$Y(\omega^*, \cdot)$ as defined by (6.4) is finitely additive on the algebra $\mathscr{A}$. Furthermore, for any set $A \in \mathscr{A}$ and $\varepsilon > 0$, there exists a compact set $K \subset A$ which is a finite union of closed rectangles such that

$$\liminf_{n\to\infty} n^{-1}[\textstyle\sum_{j=1}^{n} I_K(Z_j(\omega^*))] \ge Y(\omega^*, A) - \varepsilon, \quad P^*\text{-a.s.}$$

This implies that if $\{A_k\}$ is a decreasing sequence of sets in $\mathscr{A}$ and $Y(\omega^*, A_k) \to \eta > 0$, then there exists a decreasing sequence of compact sets $K_k \subset A_k$ such that

$$\liminf_{n\to\infty} \liminf_{m\to\infty} m^{-1}[\textstyle\sum_{j=1}^{m} I_{K_n}(Z_j(\omega^*))] \ge \tfrac{1}{2}\eta$$

and hence $\cap_{k=1}^{\infty} A_k \ne \phi$. This implies that the additive set function $Y(\omega^*, \cdot)$ on the algebra $\mathscr{A}$ is continuous from above at ϕ. The existence of a countably additive measure $Y(\omega^*, \cdot)$, for almost every ω^*, on $\sigma(\mathscr{A}) = \mathscr{B}(S)$ then follows from the Carathéodory extension theorem. This completes the proof of (b) by noting that, if necessary, $Y(\omega^*, \cdot)$ can be redefined on a set of ω^* of P^*-zero probability so that $Y(\omega^*, \cdot)$ is a probability measure for every ω^*.

The measurability of the mapping $\omega^* \to Y(\omega^*, A)$ for $A \in \mathscr{A}$ with respect to the σ-algebra $\mathscr{E}$ is a consequence of (6.4). As defined, $Y(\omega^*, A)$ is a limit of $\mathscr{E}$-measurable functions and is clearly invariant under the action of $\pi^* \in Q$. Moreover, the class of A's for which $\omega^* \to Y(\omega^*, A)$ is $\mathscr{E}$-measurable is closed under monotone limits. Hence, $\omega^* \to Y(\omega^*, A)$ is $\mathscr{E}$-measurable for all $A \in \sigma(\mathscr{A}) = \mathscr{B}(S)$. This completes the proof of (c). Since $\mathscr{B}(M_1(S))$ is generated by functions of the form $\mu \to \mu(A)$ with $A \in \mathscr{B}(S)$, part (d) follows from part (c) and another standard monotone class argument.

Part (e) for the case of sets $A_k \in \mathscr{A}$ follows from (6.4) and the law of large numbers for exchangeable random variables. Since both sides of (6.5) or (6.6) describe probability measures which agree on generating algebras P^*-a.s., they are identical P^*-a.s.

To prove (f), it suffices to show that

$$E(\textstyle\prod_{j=1}^{m} X(A_j)) = E^*(\prod_{j=1}^{m} Y(A_j)),$$

where $A_1, \cdots, A_m \in \mathscr{B}(S)$. But

$$\begin{aligned} E(\textstyle\prod_{j=1}^{m} X(A_j)) &= M_m(A_1, \cdots, A_m) = P^*(A_1 \times \cdots \times A_m) \\ &= E^*(P^*(A_1 \times \cdots \times A_m \mid \mathscr{E})) = E^*(\textstyle\prod_{j=1}^{m} Y(\omega^*, A_j)), \end{aligned}$$

and the proof of the lemma is complete.

The probability space $(\Omega^*, \mathscr{F}^*, P^*)$ together with the mapping $Y:\Omega^* \to M_1(S)$ constructed above are referred to as the *canonical representation* of the random probability measure X.

6.2 *Moment equations for the Fleming-Viot Model.* Let $\{X(t):t \geq 0\}$ denote the Fleming-Viot process, and let $\phi_1, \cdots, \phi_n \in C_K^2(R^d)$. Then, from the defining martingale problem for the Fleming-Viot process, it follows that

$$\xi(t) = \prod_{j=1}^n \langle \phi_j, X(t)\rangle - \prod_{j=1}^n \langle \phi_j, X(s)\rangle$$

$$(6.8) \qquad - \int_s^t (D \textstyle\sum_{j=1}^n \prod_{k=1,k\neq j}^n \langle \phi_k, X(u)\rangle \langle \Delta\phi_j, X(u)\rangle$$

$$+ \gamma \textstyle\sum_{j=1}^n \sum_{i=1}^n \prod_{k\neq i,j} \langle \phi_k, X(u)\rangle [\langle \phi_j\phi_i, X(u)\rangle - \langle \phi_j, X(u)\rangle \langle \phi_i, X(u)\rangle]) \, du$$

is a P_μ-martingale for every μ and $0 \leq s \leq t$.

Let $M_k(s, \mu, t:dx_1, \cdots, dx_k)$ denote the kth moment measure for $X(t)$ given that $X(s) = \mu$; that is, the family $M_k(s, \cdot, t: \cdot)$ denote the transition function moment measures. They by (6.8),

$$\int \cdots \int \textstyle\prod_{j=1}^n \phi_j(x_j) M_n(s, X(s), t:dx_1, \cdots, dx_n)$$

$$= \textstyle\prod_{j=1}^n \langle \phi_j, X(s)\rangle$$

$$+ \int_s^t \int \cdots \int (D \textstyle\sum_{j=1}^n \{\prod_{k\neq j} \phi_k(x_k)\Delta\phi_j(x_j)\} - \gamma n(n-1) \prod_{j=1}^n \phi_j(x_j)) \, du$$

$$(6.9) \qquad \cdot M_n(s, X(s), u:dx_1, \cdots, dx_n)$$

$$+ \gamma \int_s^t \int \cdots \int (\textstyle\sum_{j=1}^n \sum_{i=1,i\neq j}^n \prod_{k=1,k\neq j}^n \phi_k(x_k)\delta(x_j - x_k))$$

$$\cdot M_{n-1}(s, X(s), u: \textstyle\prod_{k=1,k\neq j}^n dx_k) \, du.$$

(6.9) implies that $M_n(s, \mu, \cdot : \cdot)$ satisfies the following system of partial differential equations in the weak sense:

$$\partial M_n(t:dx_1, \cdots, dx_n)/\partial t = D \textstyle\sum_{i=1}^n \Delta_i M_n(t:dx_1, \cdots, dx_n)$$

$$(6.10) \qquad - \gamma n(n-1) M_n(t:dx_1, \cdots, dx_n)$$

$$+ \gamma \textstyle\sum_{i=1}^n \sum_{j=1,j\neq i}^n M_{n-1}(t: \prod_{p=1,p\neq i}^n dx_p)\delta(x_i - x_j)$$

with the initial condition

$$M_n(s; dx_1, \cdots, dx_n) = \textstyle\prod_{i=1}^n X(s, dx_i).$$

The initial-value problem given by (6.10) can be solved successively for $n = 1, 2, 3, \cdots$ and $t > s$, as follows:

$$M_n(s, X(s), t; dx_1, \cdots, dx_n) = k_t * X(s) + \gamma \textstyle\sum_{i=1}^n \sum_{j=1,j\neq i}^n$$

$$(6.11) \qquad \cdot \int_s^t k_{t-u} * [M_{n-1}(s, X(s), u; dx_1, \cdots dx_{j-1}, dx_{j+1}, \cdots, dx_n) \cdot \delta(x_i - x_j)] \, du,$$

where $*$ denotes convolution and

$$k_t(x_1, \cdots, x_n) \equiv (4\pi Dt)^{-nd/2} \exp(-\textstyle\sum_{i=1}^n |x_i|^2/4Dt) \exp(-\gamma n(n-1)t).$$

This implies that for $t > s$, $M(s, X(s), t; dx_1, \cdots, dx_n)$ is a probability measure which can be represented in the form:

$$M_n(s, X(s), t; dx_1, \cdots, dx_n) = \left[\int \cdots \int m_n(t-s; y_1, \cdots, y_n; x_1, \cdots, x_n) \cdot X(s, dy_1) \cdots X(s, dy_n)\right] dx_1 \cdots dx_n. \tag{6.12}$$

The probability transition density functions $m_n(u; \cdot; \cdot)$ are smooth off the diagonals $\{(x_1, \cdots, x_n) : x_i = x_j \text{ for some } i \neq j\}$ and satisfy a system of equations of the form (6.10) with the initial conditions

$$m_n(0; y_1, \cdots, y_n; x_1, \cdots, x_n) = \prod_{i=1}^n \delta(x_i - y_i). \tag{6.13}$$

Given the transition moment measures $M_k(\cdot, \cdot, \cdot : \cdot)$, we can compute joint moment measures at one or more times. For example, for $0 < s < t$,

$$\begin{aligned} &E_\mu\{\textstyle\prod_{i=1}^n \langle \phi_i, X(t)\rangle \prod_{j=1}^n \langle \psi_j, X(s)\rangle\} \\ &= \int \cdots \int \textstyle\prod_{i=1}^n \phi_i(x_i)\psi_i(y_i) m_{2n}(s; z_1, \cdots, z_{2n}; y_1, \cdots, y_n, v_1, \cdots, v_n) \\ &\quad \cdot m_n(t-s; v_1, \cdots, v_n; x_1, \cdots, x_n)\mu(dz_1) \cdots \mu(dz_{2n}) \textstyle\prod_{i=1}^n dv_i\, dx_i\, dy_i. \end{aligned} \tag{6.14}$$

Note that if $\mu(R^d) = 1$, $\mu(\{\infty\}) = 0$, then the same is true for $X(t)$, that is, $X(t, R^d) = 1$, $X(t, \{\infty\}) = 0$, a.s. for each $t \geq 0$. This follows since the system of equations (6.10) has solutions which are probability measures on $(R^d)^n$.

Equation (6.10), for fixed n, is the forward Kolmogorov equation for a Markov process with stationary transition mechanism on $(R^d)^n$. This Markov process is denoted by $(Z_1(t), \cdots, Z_n(t))$ and has probability transition density function $m_n(u; y_1, \cdots, y_n; x_1, \cdots, x_n)$. It can be interpreted as a process describing the motion of n particles in R^d as follows. Each particle performs an independent Brownian motion in R^d; however at constant rate, one particle disappears and another simultaneously splits into two, and the resulting particles continue to move as independent Brownian motions in R^d. Thus, for each $n \geq 1$, there is a right continuous Markov process defined on $(R^d)^n$ by the probability transition density function $m_n(s; y_1, \cdots, y_n; x_1, \cdots, x_n)$. We can construct a canonical version of this process on $(D^n, \mathscr{F}^n)$ where $D^n \equiv D([0, \infty), (R^d)^n)$, the space of right continuous functions having left limits from $[0, \infty)$ into $(R^d)^n$. The only discontinuities occurring are those described above, namely, the disappearance of one particle and the simultaneous binary fission of a second particle. Two particles resulting from a binary fission are said to be *siblings*, and the particle from which they split is said to be the *parent*. Note that the particle which disappeared has no descendents. The *descendents* of a particle are all particles which can be traced back through a sequence of binary fissions to that given particle which is also referred to as a *common ancestor* of the descendents.

REMARK 6.1. There is also a relationship between the dual process $\{\eta_t : t \geq 0\}$ defined in Section 5.2 and the N-particle process $\{Z_1(t), \cdots, Z_N(t)\}$ defined above. Recall that if $f \in C^\infty(S^N)$,

$$L^{\#}F_\mu(f) = F_\mu(Kf),$$

where K is defined in (5.12). Since K is the infinitesimal generator of the process $\{Z_1(t), \cdots, Z_N(t)\}$, this implies that for $f \in C(S^N)$

$$E_f(\eta_t)(x_1, \cdots, x_N) = E_{x_1,\cdots,x_N}(f(Z_1(t), \cdots, Z_N(t))).$$

6.3 *Construction of an infinite particle system.* In this section, we construct an infinite particle system whose evolution is determined by the moment equations (6.10).

We begin with the n-particle systems described in the last section. However, we now insist that the initial particles be indistinguishable. The weak sense partial differential equations (6.10) together with the symmetrized initial conditions

$$M_n^s(0; dx_1, \cdots, dx_n) = (n!)^{-1} \sum_\pi \sum_{i=1}^n \delta(x_i - y_{\pi(i)}), \tag{6.15}$$

where π denotes a permutation of $(1, \cdots, n)$ and $y_1, \cdots, y_n$ are n points in R^d, have probability-measure-valued solutions. The solutions $M_n^s(t; dx_1, \cdots, dx_n)$ possess densities $m_n^s(t; x_1, \cdots, x_n)$ for $t > 0$, which are smooth off the diagonals $\{(x_1, \cdots, x_n) : x_i = x_j$ for some $i \neq j\}$ (cf. (6.12)) and are symmetric functions of $x_1, \cdots, x_n$.

We now construct a probability measure P^n on $(D^n, \mathscr{F}^n)$ which describes the motion of an n-particle process $(Z_1(t), \cdots, Z_n(t))$ on R^d. $(Z_1(t), \cdots, Z_n(t))$ is assumed to be a Markov process on $(R^d)^n$ with transition probability density function $m_n(t; y_1, \cdots, y_n; x_1, \cdots, x_n)$ described in Section 6.2. The initial distribution is assumed to be the following:

$$P_\mu(Z_1(0) \in dx_1, \cdots, Z_n(0) \in dx_n) = (n!)^{-1} \sum_\pi (\prod_{i=1}^n \delta(x_i - y_{\pi(i)})\mu(dy_i)). \tag{6.16}$$

In other words, the n initial points are chosen to be a simple random sample from the probability law μ. Then the distribution at time t is given by

$$\begin{aligned} &P_\mu(Z_1(t) \in dx_1, \cdots, Z_n(t) \in dx_n) \\ &= [(n!)^{-1} \sum_\pi \int \cdots \int m_n(t; y_{\pi(1)}, \cdots, y_{\pi(n)}; x_1, \cdots, x_n)\mu(dy_1) \cdots \mu(dy_n)]\, dx_1 \cdots dx_n \\ &\equiv m_n^s(t; x_1, \cdots, x_n)\, dx_1 \cdots dx_n. \end{aligned} \tag{6.17}$$

In terms of $m_n^s(t;\cdot)$, (6.10) becomes

$$\begin{aligned} &\partial m_n^s(t; x_1, \cdots, x_n)/\partial t \\ &\quad = D \sum_{i=1}^n \Delta_i m_n^s(t; x_1, \cdots, x_n) - \gamma n(n-1) m_n^s(t; x_1, \cdots, x_n) \\ &\quad\quad + \gamma \sum_{i=1}^n \sum_{j=1, j\neq i}^n m_{n-1}^s(t; x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)\delta(x_i - x_j) \\ &\quad \equiv G_1 m_n^s(t; \cdot) + G_2 m_{n-1}^s(t; \cdot). \end{aligned} \tag{6.18}$$

Since $\{m_n^s(t; x_1, \cdots, x_n)\}$ are the moment densities of the Fleming-Viot process, (6.2) implies that for $n \geq 2$,

$$\int m_n^s(t; x_1, \cdots, x_n)\, dx_n = m_{n-1}^s(t; x_1, \cdots, x_{n-1}).$$

However, starting with the moment equations alone, the transition function for the Fleming-Viot process can be constructed as in Lemma 6.1 (without prior assumption of the existence of the process). It is therefore instructive to provide an independent proof of the consistency of the family of moment densities $\{m_n^s(t; x_1, \cdots, x_n)\}$.

LEMMA 6.2. *Let* $m_{n-1}^*(t; x_1, \cdots, x_{n-1}) \equiv \int m_n^s(t; x_1, \cdots, x_n)\, dx_n$. *Then*

$$m_{n-1}^*(t; x_1, \cdots, x_{n-1}) = m_{n-1}^s(t; x_1, \cdots, x_{n-1}) \quad \textit{for all} \quad t \geq 0.$$

PROOF. Integrating both sides of (6.16), we obtain $m_{n-1}^*(0; x_1, \cdots, x_{n-1}) = m_{n-1}^s(0; x_1, \cdots, x_{n-1})$. Next, we integrate both sides of Equation (6.18) to obtain the equations

$$\begin{aligned} &\partial m_{n-1}^*(t; x_1, \cdots, x_{n-1})/\partial t \\ &= D \sum_{i=1}^{n-1} \Delta_i m_{n-1}^*(t; x_1, \cdots, x_{n-1}) - \gamma n(n-1) m_{n-1}^*(t; x_1, \cdots, x_{n-1}) \\ &\quad + \gamma \sum_{i=1}^n \sum_{j=1, j\neq i}^n \int m_{n-1}^s(t; x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)\delta(x_i - x_j)\, dx_n. \end{aligned}$$

But the last term on the right-hand side of (6.19) can be rewritten as follows:

$$\gamma[\textstyle\sum_{i=1}^{n-1}\sum_{j=1,j\neq i}^{n-1} m_{n-2}^{*}(t;x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_{n-1})\delta(x_j-x_i)$$
$$+\textstyle\sum_{i=1}^{n-1} m_{n-1}^{s}(t;x_1,\cdots,x_{n-1})\displaystyle\int\delta(x_i-x_n)\,dx_n$$
$$+\textstyle\sum_{j=1}^{n-1}\displaystyle\int m_{n-1}^{s}(t;x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_n)\delta(x_j-x_n)\,dx_n]$$
$$=\gamma[\textstyle\sum_{i=1}^{n-1}\sum_{j=1,j\neq i}^{n-1} m_{n-2}^{*}(t;x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_{n-1})\delta(x_j-x_i)$$
$$+(n-1)m_{n-1}^{s}(t;x_1,\cdots,x_{n-1})+\textstyle\sum_{j=1}^{n-1} m_{n-1}^{s}(t;x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_{n-1},x_j)].$$

Using the symmetry of $m_{n-1}^{s}(t;x_1,\cdots,x_{n-1})$ and the fact that $n(n-1)-2(n-1)=(n-1)(n-2)$, we obtain

$$\begin{aligned}\partial(m_{n-1}^{s}(t;\cdot)-m_{n-1}^{*}(t;\cdot))/\partial t&=G_1(m_{n-1}^{s}(t;\cdot)-m_{n-1}^{*}(t;\cdot))\\&\quad+G_2(m_{n-2}^{s}(t;\cdot)-m_{n-2}^{*}(t;\cdot))\\&\quad-2(n-1)(m_{n-1}^{s}(t;\cdot)-m_{n-1}^{*}(t;\cdot))\end{aligned}\tag{6.20}$$

and

$$(m_{n-1}^{s}(0;\cdot)-m_{n-1}^{*}(0;\cdot))=0.$$

It can be verified by an induction argument that the unique solution of the initial value problem (6.20) is $(m_{n-1}^{s}(t;\cdot)-m_{n-1}^{*}(t;\cdot))=0$ for all $t\geq 0$ and $n\geq 1$.

We now confine our attention to the interval $[0, t_0]$. Let $D_{t_0}^{n}\equiv D([0,t_0],(R^d)^n)$, and $\mathscr{F}_t^n$ the completion of the σ-algebra generated by $\{Z(s):0\leq s\leq t\}$. Let τ_p denote the time of the pth discontinuity. Then τ_p is a stopping time. Let $Z(t)\equiv(Z_1(t),\cdots,Z_n(t))$ as above. Then $Z(\tau_p)$ and $Z(\tau_p-)$ are $\mathscr{F}_{t_0}$-measurable (refer to Dellacherie and Meyer, 1978; IV.33, IV.64).

Particles m and k are said to have a *common parent* m at time t_0 for the sample history ω if for some $p\geq 1$,

$$\tau_p(\omega)\leq t_0,$$
$$Z_m(\omega,\tau_p(\omega))=Z_k(\omega,\tau_p(\omega)),$$
$$Z_m(\omega,\tau_p(\omega)-)\neq Z_k(\omega,\tau_p(\omega)-),$$
$$Z_m(\omega,\tau_p(\omega)-)=Z_m(\omega,\tau_p(\omega)),$$
$$Z_m(\omega,\cdot),Z_k(\omega,\cdot)\quad\text{are continuous on}\quad[\tau_p(\omega),t_0].$$

At time t_0 for the sample history ω, we define recursively the *set of ancestors* of a particle m to consist of its parent and the set of ancestors of its parent. Two particles are said to have a *common ancestor* if their sets of ancestors have a common element. We now let

$B_{m,k}^{1}\equiv\{\omega$: particles m and k have a common parent at time t_0 for $\omega\}$,
$B_{m,k}\equiv\{\omega$: particles m and k have a common ancestor at time t_0 for $\omega\}$,
and, for $\omega\in B_{m,k}$,
$T_{m,k}(\omega)\equiv\sup\{0\leq t\leq t_0$: particles m and k have a common ancestor alive at time t for $\omega\}$.

LEMMA 6.3. *$B_{m,k}^{1}$, $B_{m,k}$ and $T_{m,k}(\cdot)$ are $\mathscr{F}_{t_0}$-measurable.*

PROOF. The measurability of $B_{m,k}^{1}$ follows from the measurability of τ_p, $Z(\tau_p)$ and $Z(\tau_p-)$. The proof of the measurability of $B_{m,k}$ follows in a similar manner. Finally, note that the event $\{T_{m,k}\geq t\}$ can be expressed in a form similar to that of $B_{m,k}$ on the restricted interval $[t, t_0]$. The proof then follows in a similar manner.

LEMMA 6.4. *For $n \geq 2$, there exists a measurable mapping*

$$\Phi_n: (D_{t_0}^n, \mathscr{F}_{t_0}^n) \to (D_{t_0}^{n-1}, \mathscr{F}_{t_0}^{n-1})$$

such that $P^{n-1}(A) = P^n(\Phi_n^{-1}(A))$ for $A \in \mathscr{F}_{t_0}^{n-1}$.

PROOF. Given $\omega \in D_{t_0}^n$, consider $\{Z_1(\omega, t_0), \cdots, Z_{n-1}(\omega, t_0)\} \in (R^d)^{n-1}$. Starting from this point, we can trace out an $(n-1)$ particle path *backwards* in time as follows. Except at times of discontinuity, we retrace the original paths backward in time. At a time of discontinuity τ involving two of the $(n-1)$ paths, we add the particle which disappeared at time τ, thus maintaining the total of $(n-1)$ particles. Continuing in this way, we reach a subset $\{Z_{i_1}(\omega, 0), \cdots, Z_{i_{n-1}}(\omega, 0)\}$ of particles at time zero which are either ancestors of $Z_1(\omega, t_0), \cdots, Z_{n-1}(\omega, t_0)$ or particles whose deaths occurred at the time of birth of an ancestor of $Z_1(\omega, t_0), \cdots, Z_{n-1}(\omega, t_0)$. We then take the path $\{Z_{i_1}(\omega, t) \cdots, Z_{i_{n-1}}(\omega, t): 0 \leq t \leq t_0\}$.Under an appropriate relabeling of $(i_1, \cdots, i_{n-1})$, we obtain a unique path $Z^*(\omega, \cdot) = \{Z_1^*(\omega, \cdot), \cdots, Z_{n-1}^*(\omega, \cdot): 0 \leq t \leq t_0\} \in D_{t_0}^{n-1}$ such that $Z_k^*(\omega, t_0) = Z_k(\omega, t_0)$, $k = 1, \cdots, (n-1)$. We then define

$$\Phi_n(\omega) \equiv Z^*(\omega, \cdot).$$

The mapping Φ_n is continuous in the Skorohod metric on $D_{t_0}^n$, and therefore it is measurable. The distribution of $Z^*(t)$ is given by $m_{n-1}^*(t; \cdot)$ as defined in Lemma 6.2. Moreover, the process $Z^*(t)$ is Markov since the evolution of an $(n-1)$ particle subsystem of $Z(t)$, conditioned on no binary fission of the remaining particle, is Markov. Lemma 6.2 then implies that $Z^*(t)$ has distribution P^{n-1}, and the proof is complete.

In view of Lemma 6.4, the system $(D_{t_0}^n, \mathscr{F}_{t_0}^n)$ together with the mappings Φ_n form a projective system. Therefore, there exists a projective limit, that is, a probability law P^∞ on $(D_{t_0}^\infty, \mathscr{F}_{t_0}^\infty)$, where $D_{t_0}^\infty \equiv D([0, t_0], (R^d)^\infty)$ where $(R^d)^\infty$ is furnished with the product topology, and mappings $\Xi_n: D_{t_0}^\infty \to D_{t_0}^n$ such that $P^n = \Xi_n(P^\infty)$. For a discussion of Prokhorov's theorem on the existence of projective limits and its extensions, refer to Dellacherie and Meyer (1978; III.52). The $(R^d)^\infty$-valued Markov process defined by the triple $(D_{t_0}^\infty, \mathscr{F}_{t_0}^\infty, P^\infty)$ is referred to as the *infinite particle system* associated with the Fleming-Viot process.

LEMMA 6.5. *Let $Z(t)$ denote the ∞-particle process constructed above on the interval $[0, t_0]$. For fixed $n < \infty$ let $\alpha(t_0, \omega)$ denote the number of ancestors of $(Z_1(\omega, t_0), \cdots, Z_n(\omega, t_0))$ alive at time $t = 0$. Then*

$$P(\alpha(t_0) \leq m) = P(T_m^\infty < t_0), \tag{6.21}$$

where $T_m^\infty \geq 0$ has Laplace transform:

$$\phi_{T_m^\infty}(s) \equiv E(\exp(-sT_m^\infty)) = \{\textstyle\prod_{k=m+1}^\infty (1 + (s/\gamma k(k-1)))\}^{-1}, \quad s \geq 0. \tag{6.22}$$

PROOF. We begin with a study of the n-particle subsystem. We ignore the locations of the particles and consider the process $\xi(t)$ which is defined by the number of jump discontinuities in $[0, t]$ for the n-particle process. According to Equation (6.10), this is a Poisson process with parameter $\gamma n(n-1)$. In fact, for each of the $n(n-1)$ ordered pairs of particles, there can be associated an exponential alarm clock with parameter γ such that each arrival of the Poisson process coincides with the ringing of a clock. Since a Poisson process reversed in time is also a Poisson process, it follows that the last discontinuity before time t_0 occurred at time $t_0 - \tau_{n,1}$ where $\tau_{n,1}$ is exponentially distributed with mean $1/\gamma n(n-1)$. This means that at time $t_0 - \tau_{n,1}$, two of the particles alive at time t_0 were produced by a binary fission. Viewing the time-reversed process at this point, there is a coalescence of two particles at their common birth place and time. Now we continue to follow backwards in time the histories of the $(n-1)$ ancestors alive at time $(t_0 - \tau_{n,1})-$. At this point there are $(n-1)(n-2)$ remaining exponential alarm clocks, none of which have rung prior to $(t_0 - \tau_{n,1})$ (in reverse time). Thus the no-memory property implies that the first of these rings after an exponential time $\tau_{n,2}$ with mean $1/\gamma(n-1)(n-2)$, independent of $\tau_{n,1}$. This clock rings at time $t_0 - \tau_{n,1} - \tau_{n,2}$. Continuing in this way, we reach the time $t_0 - T_m^n$, $T_m^n \equiv \tau_{n,1} + \cdots + \tau_{n,(n-m)}$, at which there are exactly m ancestors. Note that $\tau_{n,1}, \cdots, \tau_{n,(n-m)}$ are independent exponential random variables with means $1/\gamma n(n-1), \cdots, 1/\gamma m(m+1)$. Therefore,

$$\phi_{T_m^n}(s) \equiv E(\exp(-sT_m^n)) = \{\textstyle\prod_{k=(m+1)}^{n} (1 + s/\gamma k(k-1))\}^{-1}. \tag{6.23}$$

Recall that

$$t_0 - T_m^n = \sup\{t : Z_1(t_0), \cdots, Z_n(t_0) \text{ has } \leq m \text{ ancestors at time } t\}. \tag{6.24}$$

Since T_m^n is monotone increasing in n, we can take the limit

$$T_m^\infty(\omega) \equiv \lim_{n\to\infty} T_m^n(\omega). \tag{6.25}$$

Moreover,

$$\phi_{T_m^\infty}(s) = \lim_{n\to\infty}\phi_{T_m^n}(s) = \{\textstyle\prod_{k=(m+1)}^{\infty} (1 + s/\gamma k(k-1))\}^{-1}, \tag{6.26}$$

and the proof is complete.

Note that T_n^∞ is a finite random variable. In fact,

$$E(T_m^\infty) = \textstyle\sum_{k=(m+1)}^{\infty} (1/\gamma k(k-1)) < \infty, \tag{6.27}$$

and

$$\operatorname{Var}(T_m^\infty) = \textstyle\sum_{k=(m+1)}^{\infty} (1/\gamma k(k-1))^2 < \infty. \tag{6.28}$$

Note also that we can extend the process backwards in time to negative time, if necessary, to go back to the time $t_0 - T_1^\infty$ at which there is exactly one common ancestor. In this way the infinite system of particles alive at time t_0 are the descendents of exactly one ancestor at a finite time in the past. Thus we can speak of the "age" of the infinite system of particles alive at time t_0.

Extending this idea, we define the *age process* $\{A(t) : t \geq 0\}$ as follows:

$$\begin{aligned} A(t) &\equiv t - \sup\{s : 0 \leq s \leq t, Z(t) \text{ has exactly one ancestor alive at time } s\}, \\ &\qquad \text{if the set is non-empty,} \\ &\equiv t, \quad \text{otherwise.} \end{aligned} \tag{6.29}$$

LEMMA 6.6.
(i) $\mathscr{L}(A(t) \mid A(t) < t) = \mathscr{L}(T_1^\infty \mid T_1^\infty < t)$, *where* $\mathscr{L}(\cdot \mid \cdot)$ *denotes the conditional probability law.*
(ii) $P(A(t) = t) = P(T_1^\infty \geq t)$.
(iii) $A(t+s) \leq A(t) + s$.
(iv) $\mathscr{L}(A(t+s) \mid A(s) = r)$ is independent of s *if* $r < s$.

PROOF. (i), (ii), and (iii) follow from the definitions (6.29) and (6.24). (iv) follows from the definition (6.29) and the fact that $Z(\cdot)$ is a Markov process with stationary transition mechanism.

Note that $Z(0)$ can be assigned the same distribution as that of a random cluster whose age is distributed according to $\mathscr{L}(T_1^\infty)$. In this case, the age process $A^*(t)$ is a stationary stochastic process with

$$\mathscr{L}(A^*(t)) = \mathscr{L}(T_1^\infty), \quad \text{for} \quad -\infty < t < \infty. \tag{6.30}$$

Note that part (iii) of Lemma 6.6 implies that

$$\sup_{0 \leq t \leq T} A^*(t) < \infty, \quad \text{with probability one.} \tag{6.31}$$

LEMMA 6.7. *Given any* $T > 0$, $\mathscr{E}(A(t+s):0 \le s \le T) \Rightarrow \mathscr{E}(A^*(s):0 \le s \le T)$, *as* $t \to \infty$, *in the sense of weak convergence of probability measures on* $D([0, T], R^+)$ *furnished with the Skorohod topology.*

PROOF. As a consequence of Lemma 6.5, we can associate a renewal process $\{S_k : k \ge 0\}$ with the infinite particle system $Z(t)$ as follows:

$$S_0 \equiv 0,$$

$$S_{k+1} \equiv \inf\{t\text{: all particles alive at time } t \text{ are descendents of exactly one particle alive at time } S_k\}. \tag{6.32}$$

The random variables $(S_{k+1} - S_k)$ are independent, identically distributed non-arithmetic random variables with distribution $\mathscr{E}(T_1)$. Since they have finite means and variances, we can apply the standard results of renewal theory to the sequence $\{S_k\}$ (refer to Feller, 1966, Chapter 11, or Athreya et al, 1978, for the required renewal theory). The desired result is obtained by proving the existence, for any $\varepsilon > 0$, of an "ε-coupling" of the processes $A(\cdot)$ and $A^*(\cdot)$. We begin by assuming that $A(\cdot)$ and $A^*(\cdot)$ are independent. Then there exists an almost surely finite random integer K such that $|S_K - S_K^*| < \varepsilon$. We then modify the joint process $(A(\cdot), A^*(\cdot))$ by coupling the hierarchy of Poisson processes of jumps of the two processes after S_K, S_K^*. This does not change the marginal distributions of the two processes. Moreover,

$$P(\sup_{s \ge t+\varepsilon} \inf_{|s'-s| \le \varepsilon} |A(s') - A^*(s)| > 0) \le P(S_{K+1} > t) \tag{6.33}$$

and

$$\lim_{t\to\infty} P(S_{K+1} > t) = 0.$$

Thus

$$P(\rho_S(A(s), A^*(s) : s \ge t) > 2\varepsilon) \to 0 \quad \text{as} \quad t \to \infty,$$

where $\rho_S(\cdot, \cdot)$ denotes the Skorohod metric, and the proof is complete.

LEMMA 6.8. *The stationary stochastic process* $A^*(\cdot)$ *is ergodic.*

PROOF. This follows since it is a consequence of the proof of Lemma 6.7 that the distribution of $\{A^*(s) : s \ge t\}$ is almost independent of $\{A^*(s) : 0 \le s \le T\}$, for fixed T, for sufficiently large t.

Thus if $A(t) < t$, we can interpret $Z(t)$ as a random infinite cluster of particles which are the descendents of a single particle, referred to as the *founder*, which begins to branch at time $t - A(t)$. In order to describe the past history of the particles alive at time t, we consider the binary branching Brownian motion process $\{Z_t^*(s) : 0 \le s \le t\}$ obtained by deleting all particles at times $s < t$ which are not ancestors of particles alive at time t. Note that $Z_t^*(s)$ consists of a single particle for $s < t - A(t)$. Also, conditioned on the age of the cluster $A(t) = a$, the times between branches $\{\tau_k : k \ge 1\}$ are distributed with joint density function

$$p(\tau_1, \tau_2, \cdots) = c \exp(-\gamma \textstyle\sum_{k=1}^{\infty} k(k+1)\tau_k), \quad \sum_{k=1}^{\infty} \tau_k = a, \tag{6.34}$$

where c is a normalizing constant.

6.4 *Moment densities of the centered cluster.* Fleming and Viot (1978, 1979) considered the quantity

$$I_1(t : \xi) = \int_{R^d} m_2(t; x, x + \xi)\, dx, \tag{6.35}$$

which denotes the density of the relative genetic displacement of two randomly chosen individuals. $I_1(t;\ \cdot)$ satisfies the partial differential equation

$$\partial I_1/\partial t = D\Delta I_1 - 2\gamma I_1 + 2\gamma\delta(\xi). \tag{6.36}$$

As $t \to \infty$, $I_1(t;\ \cdot)$ tends to an equilibrium value

$$I_1(\infty;\ \xi) = \lim_{t\to\infty} I_1(t;\ \xi)$$

which can be obtained explicitly using Fourier or Laplace transforms:

$$\begin{aligned} I_1(\infty;\ \xi) &= 2(2\pi)^{-d/2}(c/|\xi|)^{d/2-1}K_{1-d/2}(c|\xi|),\ d \geq 2, \\ &= \tfrac{1}{2}c\exp(-c|\xi|), \quad \text{if} \quad d = 1, \quad c = (2\gamma/D)^{1/2}, \end{aligned} \tag{6.37}$$

where $K_{1-d/2}(\cdot)$ is the modified Bessel function (see Erdelyi, 1954, page 146).

In the same way, we can identify the k-particle displacements relative to a randomly chosen one or, equivalently, the moment densities for the Fleming-Viot cluster centered at a randomly chosen point. These are given by

$$I_k(t;\ \xi_1, \cdots, \xi_k) \equiv \int_{R^d} m_{k+1}(t;\ x, x + \xi_1, \cdots, x + \xi_k)\ dx. \tag{6.38}$$

To obtain the stated identification, let $\theta_y X(t;\ \cdot)$ denote the shift of $X(t;\ \cdot)$ by the displacement y, that is,

$$\theta_y X(t;\ A) \equiv X(t;\ \{a - y : a \in A\}).$$

We now assume that the displacement Y is a random point in R^d chosen according to the probability law $X(t, dx)$. Then

$$\begin{aligned} E[\textstyle\prod_{i=1}^n \langle \phi_i, \theta_Y X(t, \cdot)\rangle] &= \int \cdots \int m_{n+1}(t;\ y, x_1 + y, \cdots, x_n + y) \\ &\qquad \cdot\ \phi_1(x_1) \cdots \phi_n(x_n)\ dx_1 \cdots dx_n\ dy \\ &= \int \cdots \int I_n(t;\ x_1, \cdots, x_n)\phi_1(x_1) \cdots \phi_n(x_n)\ dx_1 \cdots dx_n. \end{aligned} \tag{6.39}$$

In this way we obtain an infinite sequence of exchangeable random variables describing the random measure centered at a randomly chosen point. The limiting values

$$I_k(\infty;\ \xi_1, \cdots, \xi_k) = \lim_{t\to\infty} I_k(t;\ \xi_1, \cdots, \xi_k)$$

can be obtained as in the example above. $I_k(\infty;\ \cdot, \cdots, \cdot)$ can be interpreted as the steady-state moment densities for the randomly centered cluster. We note that the existence of these limiting values is a reflection of the coherence of the Fleming-Viot process, which will be established in another way in the next section. It also represents the analogue of the result of Kingman for the Ohta-Kimura model, namely, that the joint distribution of the relative differences of the genetic states, measured from a randomly chosen one, converges to a limit.

6.5 *The empirical moment processes.* Consider the Fleming-Viot process $\{X(t) : t \geq 0\}$ in R^d. The *empirical mean process* is defined by $x(t) \equiv (x_1(t), \cdots, x_d(t))$, where

$$x_i(t) \equiv \int_{R^d} x_i X(t, dx), \quad i = 1, \cdots, d. \tag{6.40}$$

Similarly, the *empirical covariance process* is defined by

$$v_{ij}(t) \equiv \int_{R^d} x_i x_j X(t, dx) - x_i(t)x_j(t), \quad i, j = 1, \cdots, d. \tag{6.41}$$

The higher order empirical centered moments of $X(t)$ are defined by

$$R_{k_1,\cdots,k_d}(t) \equiv \int_{R^d} \prod_{i=1}^d (x_i - x_i(t))^{k_i} X(t, dx), \tag{6.42}$$

where $k_1, \cdots, k_d$ are non-negative integers. (6.42) can be rewritten as

$$R_{k_1,\cdots,k_d}(t) = \int \cdots \int g(x, x^{(1)}, \cdots, x^{(N_0)}) X(t, dx) X(t, dx^{(1)}) \cdots X(t, dx^{(N_0)}), \tag{6.43}$$

where

$$g(x, x^{(1)}, \cdots, x^{(N_0)}) = \prod_{\ell=1}^d \prod_{j=1}^{k_\ell} (x_\ell - x_\ell^{(j)}) \tag{6.44}$$

and

$$N_0 \equiv \sum_{\ell=1}^d k_\ell.$$

LEMMA 6.9. *Assume that*

$$\int |x|^{N_0} \mu\,(dx) < \infty. \tag{6.45}$$

Then

(a) *for* $0 \le t < \infty$ *and any non-negative integers* $k_1, \cdots, k_d$,

$$E_\mu(R_{k_1,\cdots,k_d}(t)) < \infty, \quad \textit{and}$$

(b)
$$\lim_{t\to\infty} E_\mu(R_{k_1,\cdots,k_d}(t)) \equiv r_{k_1,\cdots,k_d}$$

exists and is finite.

PROOF. (a) From the dual representation (5.16) applied to the function g (cf. Remark 5.1),

$$E_\mu(R_{k_1,\cdots,k_d}(t)) = E_g\left(\int \cdots \int \eta_t(x_1, \cdots, x_{N(t)}) \mu\,(dx_1) \cdots \mu\,(dx_{N(t)})\right). \tag{6.46}$$

Assumption (6.45) implies that

$$\int \cdots \int |g(x, x^{(1)}, \cdots, x^{(N_0)})|\, \mu\,(dx) \mu\,(dx_1) \cdots \mu\,(dx_{N_0}) < \infty. \tag{6.47}$$

Observing that the dual process preserves an initial polynomial growth condition, one can verify using (6.47) that

$$E_g\left(\int \cdots \int \eta_t(x_1, \cdots, x_{N(t)}) \mu\,(dx_1) \cdots \mu\,(dx_{N(t)})\right) < \infty, \tag{6.48}$$

which completes the proof of (a).
(b) Note that the function g is translation invariant; that is,

$$g(y_1 + a, y_2 + a, \cdots, y_{N_0+1} + a) = g(y_1, y_2, \cdots, y_{N_0+1}) \quad \text{for any} \quad a \in R^d. \tag{6.49}$$

Observe also that if $\eta(0)$ satisfies condition (6.49), then η_t satisfies (6.49) for all $t \ge 0$. In addition, any function g in $C(S)$—that is, a function of one variable—that satisfies (6.49) is a constant. Then (6.46), Lemma 5.1 and the fact that the exponential distribution has finite moments of all orders imply that

$$\lim_{t\to\infty} E_\mu(R_{k_1,\cdots,k_d}(t)) = E_g\left(\int \eta_\infty(x)\mu(dx)\right) = E_g(\eta_\infty) < \infty,$$

and the proof is complete.

REMARK 6.2. Using Lemma 6.9b, it is possible to prove an ergodic theorem for the random cluster centered at the empirical mean. Then the collection $\{r_{k_1,\cdots,k_d}\}$ form the joint moment system of the expected steady state distribution of the cluster centered at the empirical mean. As noted below in Remark 7.3, this is also a consequence of the results of Section 7, which are obtained by different methods. The method of moments and the analogue of Lemma 6.9b are used in a forthcoming paper of Shiga (1981) to obtain the ergodic theorem for the continuous-time Ohta-Kimura model.

LEMMA 6.10. *Assume that* $\int x_i^2 X(0, dx) < \infty$ *for* $i = 1, \cdots, d$. *Then the stochastic process* $\{x(t) : t \geq 0\}$ *is a d-dimensional square integrable martingale.*

PROOF. Apply the martingale problem requirement for the Fleming-Viot process to the function $\psi(\mu) \equiv \langle \phi_n, \mu \rangle$. Then

$$\langle \phi_n, X(t) \rangle - \int_0^t \langle \Delta\phi_n, X(s) \rangle \, ds \text{ is a } P_\mu\text{-martingale for each } \mu \in M_1(R^d).$$

Now choose a sequence $\phi_n(\cdot)$ such that

$$|\phi_n(x)| \leq 2c_n, \quad \phi_n(x) = x_i \quad \text{if} \quad |x| \leq c_n \quad \text{with} \quad c_n \to \infty, \quad \text{and}$$
$$|\Delta\phi_n(x)| \leq M < \infty.$$

Since $E[\int |x_i| X(t, dx)] = \int |x_i| m_1(t; x) \, dx < \infty$, and $\Delta\phi_n(x) \to 0$ pointwise as $n \to \infty$, the fact that $x_i(t) = \langle x_i, X(t) \rangle$ is a P_μ-martingale follos from the dominated convergence theorem. Moreover, by the hypothesis $\int x_i^2 X(0, dx) < \infty$ and Lemma 6.9, it follows that

$$E(x_i^2(t)) = \int_{R^d} \int_{R^d} x_i y_i m_2(t; x, y) \, dx \, dy < \infty,$$

which completes the proof.

According to the Doob-Meyer decomposition for continuous square-integrable martingales, there exist unique increasing processes $\langle\langle x_i \rangle\rangle_t$ such that

$$x_i(t)x_j(t) - \langle\langle x_i, x_j \rangle\rangle_t \quad \text{is a } P_\mu\text{-martingale,} \tag{6.50}$$

where

$$\langle\langle x_i, x_j \rangle\rangle_t = \tfrac{1}{2}[\langle\langle x_i + x_j \rangle\rangle_t - \langle\langle x_i \rangle\rangle_t - \langle\langle x_j \rangle\rangle_t], \quad i \neq j,$$

and

$$x_i^2(t) - \langle\langle x_i \rangle\rangle_t \quad \text{is a } P_\mu\text{-martingale.} \tag{6.51}$$

LEMMA 6.11. $\langle\langle x_i, x_j \rangle\rangle_t = 2\gamma \int_0^t v_{ij}(s) \, ds$, *and*

$$\langle\langle x_i \rangle\rangle_t = 2\gamma \int_0^t v_{ii}(s) \, ds.$$

PROOF. It suffices to prove the result for the case $i = j$. Consider the martingale problem requirement for the function $\psi(\mu) = \langle x_i, \mu \rangle^2$. Then

$$\langle x_i, X(t) \rangle^2 - 2\gamma \int_0^t [\langle x_i^2, X(s) \rangle - \langle x_i, X(s) \rangle^2] \, ds \quad \text{is a } P_\mu\text{-martingale.} \tag{6.52}$$

Hence $\langle\langle x_i \rangle\rangle_t = 2\gamma \int_0^t v_{ii}(s)\,ds$, and the proof is complete.

LEMMA 6.12. *Assume that* $\int x_i^4 X(0, dx) < \infty$. *Then*

(i) $v_{ii}(t) = (D/\gamma)(1 - e^{-2\gamma t}) + e^{-\gamma t} \int_0^t e^{\gamma s}\, dM_i(s)$, *where* $M_i(\cdot)$ *is a square-integrable martingale.*

(ii) *Let* $h_{ij}(t) \equiv E(v_{ij}(t))$. *Then*

$$\lim_{t\to\infty} h_{ii}(t) = D/\gamma, \quad \text{if} \quad \gamma > 0, \tag{6.53}$$

$$\lim_{t\to\infty} h_{ij}(t) = 0 \qquad \text{if} \quad i \neq j, \quad \gamma > 0.$$

(iii) $h_{ii}(t) = 2Dt$ *if* $\gamma = 0$.

PROOF. Applying the martingale problem requirement to the function $\psi(\mu) = \langle x_i x_j, \mu\rangle - \langle x_i, \mu\rangle\langle x_j, \mu\rangle$, we obtain that

$$v_{ij}(t) - 2D\delta_{ij} \int_0^t \langle 1, X(s)\rangle\, ds + 2\gamma \int_0^t v_{ij}(s)\, ds \quad \text{is a } P_\mu\text{-local martingale.}$$

Hence for the case $i = j$,

$$v_{ii}(t) + 2\gamma \int_0^t v_{ii}(s)\, ds = 2Dt + M_i(t), \tag{6.54}$$

where $M_i(\cdot)$ is a P_μ-local martingale. The increasing process associated with the local martingale $M_i(\cdot)$ defined by (6.54) is given by (cf. Fleming and Viot, 1979, (5.4))

$$\langle\langle M_i \rangle\rangle_t = \int_0^t q_i(s)\, ds, \tag{6.55}$$

where

$$q_i(s) \equiv (r_{4,i}^*(s) - (r_{2,i}^*(s))^2), \tag{6.56}$$

and

$$r_{k,i}^*(s) \equiv \int (x_i - x_i(s))^k X(s, dx).$$

Then under the assumption $\int x_i^4 X(0, dx) < \infty$,

$$\begin{aligned} E(\langle\langle M_i \rangle\rangle_t) &= E\left[\int_0^t \left\{\left[\int (x_i - x_i(s))^4 X(s, dx)\right]\right.\right. \\ &\qquad \left.\left. - \left[\int (x_i - x_i(s))^2 X(s, dx)\right]^2\right\} ds\right] \\ &< \infty \quad \text{by Lemma 6.9.} \end{aligned} \tag{6.57}$$

Since

$$E(M_i^2(t) - M_i^2(0)) = E(\langle\langle M_i \rangle\rangle_t) < \infty,$$

this implies that the local martingale $M_i(\cdot)$ is actually a P_μ-square integrable martingale. (i) then follows by solving the linear first-order differential equation obtained from (6.54).

Taking expectations, we obtain the differential equation

$$h'_{ij}(t) - 2\gamma h_{ij}(t) = 2D\delta_{ij}. \tag{6.58}$$

As a consequence of Lemma 6.8, we can assume without loss of generality that $v_{ij}(t)$ is an ergodic stationary stochastic process. In view of the erogodicity of $v_{ij}(\cdot)$, we can choose $\mu = \delta_0$. Then $h_{ij}(0) = 0$ for all i, j. Solving Equation (6.58), we obtain

$$h_{ii}(t) = (D/\gamma)(1 - e^{-2\gamma t}), \qquad h_{ij}(t) = 0 \quad \text{if} \quad i \neq j. \tag{6.59}$$

Parts (ii) and (iii) then follow immediately from (6.59).

7. Clustering and coherence in the Fleming-Viot model.

7.1 *The binary branching process with explosion at t_0.* Let $X(t, dx)$ denote the Fleming-Viot probability-measure-valued process. In this section, we establish the microscopic clustering and coherence of this process.

We begin by summarizing the relevant results from the last section. Starting with the moment equations, we have constructed an infinite system of particles in R^d, denoted by $\{Z(t): t \geq 0\}$. This in turn determines, for each $t_0 > 0$, a binary branching Brownian motion process $\{Z^*_{t_0}(s): 0 \leq s \leq t_0\}$. All particles involved in $Z^*_{t_0}(s)$ are descendents of a founder which has a first branch at time $t_0 - A(t_0)$. $Z^*_{t_0}(\cdot)$ behaves like a pure birth process with birth rates $\lambda_k = \gamma k(k+1)$ and with explosion time t_0. In addition, $Z^*_{t_0}(t_0)$ consists of an infinite collection of particles which provides a canonical representation of the random measure $X(t_0, \cdot)$ described in Lemma 6.1.

Recall that the binary branches occur at times $t_0 - T^\infty_k$, $k = 1, 2, 3, \cdots$. Let $\tau_k \equiv T^\infty_k - T^\infty_{k+1}$. Then τ_k has a negative exponential distribution with mean $1/\gamma k(k+1)$.

LEMMA 7.1. *$\{T^\infty_k\}$ satisfies the following inequalities:*

$$T^\infty_k \leq 2 \log k/\gamma k, \tag{7.1}$$

for all but finitely many k, with probability one.

PROOF. We have

$$P(\tau_k > \xi) = e^{-\gamma k(k+1)\xi}, \quad \xi \geq 0. \tag{7.2}$$

Therefore

$$P(\tau_k > 2 \log k/\gamma k(k+1)) \leq 1/k^2. \tag{7.3}$$

Then, by applying the Borel-Cantelli lemma, it follows that

$$\tau_k < 2 \log k/\gamma k^2, \tag{7.4}$$

for all but finitely many k, with probability one. Therefore,

$$T^\infty_m = \sum_{k=m}^\infty \tau_k < \sum_{k=m}^\infty 2 \log k/\gamma k^2 \leq 2 \log m/\gamma m, \tag{7.5}$$

for all but finitely many m, with probability one, and the proof of the lemma is complete.

7.2 *The hierarchy of subclusters.* The random measure $X(t_0)$, or equivalently $Z^*_{t_0}(t_0)$, can be decomposed into a hierarchy of subclusters as follows. The infinite system of particles $Z^*_{t_0}(t_0)$ can be decomposed into the n subclusters of particles consisting of all those particles which are descendents of each of the n ancestors alive at time $t_0 - T^\infty_n$. In order to determine the relative sizes of these clusters, note that Equation (6.10) implies that at the time of a binary fission, each existing particle has the same probability of being chosen as the parent of the new particle.

Therefore, the probability that a binary fission will add a new particle of a given one of the n types is proportional to the current proportion of particles of that type. Thus, the evolution of the numbers in each of the n types is described by an n-type Polya urn scheme. Let $(P_1, \cdots, P_n)$ denote the limiting proportions. Let $\sum_{n-1} \equiv \{(p_1, \cdots, p_n): p_i \geq 0, \sum p_i = 1\}$. Then the distribution of $(P_1, \cdots, P_n)$ is given by the uniform distribution on $\sum_{n-1}$ (refer to Dawson (1970) for a discussion of the limit behavior of Polya urn models).

7.3 *The Gaussian displacement process.* The locations of each of the infinite set of particles $Z^*_{t_0}(t_0)$ is determined at the time of its birth as follows. Because each of the resulting particles of a binary fission is performing a Brownian motion in R^d with diffusion constant D, the relative displacement performs a Brownian motion with diffusion constant $2D$. Since in the binary branching process we label the particles according to the order of their birth, in this setting we view one of the two resulting particles as the parent and the other as the offspring. Therefore, the location of the kth new particle is obtained by taking a Gaussian displacement D_k with mean zero and variance $\sigma_k^2 = 2D\, T_k^\infty$ from the location of its parent.

LEMMA 7.2. *The displacement D_k of the kth particle from its parent satisfies the inequality*

$$|D_k| \le 4(Dd/\gamma)^{1/2}(\log k/k^{1/2}) \tag{7.6}$$

for all but finitely many k with probability one, where $|\cdot|$ denotes Euclidean length.

PROOF. D_k is a d-dimensional Gaussian random variable with covariance matrix given by $\sigma_k^2 I$, where I is the identity matrix. According to Lemma 7.1, for all but finitely many k,

$$\sigma_k^2 \le (4D/\gamma)(\log k/k), \quad \text{with probability one.} \tag{7.7}$$

But for a d-dimensional Gaussian random variable $N_d(\sigma^2)$ with covariance matrix $\sigma^2 I$,

$$P(|N_d(\sigma^2)| \ge x) \le dP(N_1(\sigma^2) \ge x/d^{1/2}) \le (d^3/2\pi)^{1/2}\sigma x^{-1}\exp(-x^2/2d\sigma^2).$$

Therefore,

$$P(|D_k| > [(16Dd/\gamma)(\log k)^2/k]^{1/2}) \le c/k^2(\log k)^{1/2}, \tag{7.8}$$

where $c = d/2(2\pi)^{1/2}$. Then applying the Borel-Cantelli lemma, it follows that

$$|D_k| \le (16Dd/\gamma)^{1/2}(\log k/k^{1/2}) \tag{7.9}$$

for all but finitely many k, with probability one.

7.4 *Bounds on the subcluster radii.* Let the n ancestors of the infinite system $Z^*_{t_0}(t_0)$ alive at time $t_0 - T_n^\infty$ be denoted by $Z^*_{t_0,i_1}(t_0 - T_n^\infty), \cdots, Z^*_{t_0,i_n}(t_0 - T_n^\infty)$. To each of these is associated an infinite subcluster at time t_0, denoted by $\{Z^*_{t_0,j}: j \in I_k\}$, $k = 1, \cdots, n$. The subcluster radius is defined by:

$$R_n \equiv \max_{1\le k\le n}\sup_{j\in I_k}|Z^*_{t_0,j}(t_0) - Z^*_{t_0,i_k}(t_0 - T_n^\infty)|. \tag{7.10}$$

R_n can be viewed as the maximal spread of the "genealogical branches" emanating from $\{Z^*_{t_0,i_1}, \cdots, Z^*_{t_0,i_n}\}$ or, equivalently, the maximal length of the corresponding *genealogical chains.* A genealogical chain is a sequence of individual particles, each of which is the offspring of the preceding one. Let $\mathscr{C}$ denote the set of all genealogical chains.

LEMMA 7.3. *Given $\delta > 0$, there is a constant c such that*

$$R_{N_k} \le c/N_k^{1/2-\delta}, \quad \text{with} \quad N_k \equiv k^{2+\eta} \tag{7.11}$$

forall but finitely many k, with probability one, where η is the smallest integer strictly greater than $2 + 1/\delta$.

PROOF. Let $J_k \equiv [t_0 - T^\infty_{N_k}, t_0 - T^\infty_{N_{k+1}})$. Note that in this interval there are always at least N_k individuals alive in the binary branching process. Moreover $N_{k+1} - N_k \le c_1 k^{1+\eta}$ for some constant c_1. Let G_k be defined by

$$G_k \equiv \max\{r: \text{chain } C \text{ has } r \text{ births in } J_k \text{ for some } C \in \mathscr{C}\}. \tag{7.12}$$

Recall that, looking backwards, the parent of a particle is chosen at random with all existing particles having the same probability of being the parent. Therefore, the probability that a particle born in J_k has as its parent another particle also born in J_k is bounded above by $c_1 k^{1+\eta}/k^{2+\eta} = c_1/k$. Hence, the probability that a given genealogical chain has r births in J_k is bounded above by $(c_1/k)^r$. Next, $\mathscr{C}$ can be divided into N_{k+1} equivalence classes of chains having the same histories up to time $t_0 - T_{N_{k+1}}$. At most $c_1 k^{1+\eta}$ of these equivalence classes has at least one birth in J_k. Thus,

$$P(G_k \geq r) \leq c_2 k^{1+\eta-r}. \tag{7.13}$$

If $r = r_0 \equiv 3 + \eta$, then

$$\textstyle\sum_{k=1}^{\infty} P(G_k \geq r_0) < \infty. \tag{7.14}$$

Therefore, applying the Borel-Cantelli lemma,

$$G_k < r_0, \tag{7.15}$$

for all but finitely many k, with probability one. But according to Lemma 7.2, the genealogical displacement from its parent of an offspring born in J_k is bounded above by

$$|D_\ell| \leq 4(Dd/\gamma)^{1/2}(\log N_k/N_k^{1/2}),$$

for all but finitely many k. Therefore,

$$\begin{aligned} R_{N_m} &\leq \textstyle\sum_{k=m}^{\infty} 4(Dd/\gamma)^{1/2} r_0 (\log N_k/N_k^{1/2}) \\ &\leq c_2 \textstyle\sum_{k=m}^{\infty} (\log k/k^{(2+\eta)/2}) \leq c_2(\log m/m^{\eta/2}) \\ &\leq c_3(\log N_m/N_m^{\eta/2(2+\eta)}) \\ &\leq c/N_m^{1/2-\delta}, \quad \text{since} \quad \eta > 2 + 1/\delta, \end{aligned} \tag{7.16}$$

where c, c_1, c_2, c_3 are constants, and the proof is complete.

Thus the random cluster $Z_{t_0}^*$ consists of N_n infinite subclusters whose relative masses are given by the uniform distribution on $\sum_{N_n-1}$ and which are contained in spheres of radius R_{N_n}.

7.5. *Compact support and clustering of the random measure.*

THEOREM 7.1. *Let $X(t, \cdot)$ denote the Fleming-Viot process in R^d, and assume that $X(0)$ has compact support, with probability one.*

(a) *In any spatial dimension, $X(t, \cdot)$ has compact support, with probability one, for each fixed $t > 0$.*

(b) *For $d \geq 3$, $t > 0$, $X(t, \cdot)$ has Hausdorff-Besicovitch dimension of support bounded above by two, with probability one.*

PROOF. (a) According to Lemma 7.3, the infinite collection of particles Z_t^* can be contained in a large finite random cube $V \subset R^d$ which can be assumed to belong to the algebra $\mathscr{A}$ of subsets of R^d. Therefore, from the canonical representation given by (6.4), it follows that

$$X(t, V) = 1 \quad \text{with probability one,} \tag{7.17}$$

and the proof of (a) is complete.

(b) In view of (a), there is no loss of generality in restricting attention to the random measure $X(t, \cdot)$ restricted to a large cube V. Consider a subdivision of the cube V into congruent subcubes of edge $\Gamma_n^{-1} = N_n^{-1/2+\delta'}$, where δ' is assumed to be rational so that the subcubes can be assumed to belong to the algebra $\mathscr{A}$. From Lemma 7.3 it follows that each of the N_n subclusters is contained in at most $(c+3)^d$ of these subcubes, so that the total number of occupied subcubes is bounded above by $(c+3)^d N_n$. Let $B_n \in \mathscr{A}$ denote the union of the latter occupied subcubes. Lemma 7.3 implies that the infinite collection of particles is contained in

$$B \equiv \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} B_n. \tag{7.18}$$

Moreover the Hausdorff-Besicovitch dimension of B is bounded above by

$$\dim(B) \le \liminf_{n\to\infty}(\log N_n/\log(N_n^{1/2-\delta'})) \le 2 + \delta'', \tag{7.19}$$

where δ'' can be made arbitrarily small, by Lemma 3.1. According to the canonical representation (6.4), $\cap_{n=k}^{k_2} B_n \in \mathscr{A}$, and

$$X(t, \omega, \cap_{n=k}^{k_2} B_n) = 1, \quad k \le k_2 < \infty, \quad k \text{ sufficiently large}, \tag{7.20}$$

for almost every ω. Therefore, since $X(t, \cdot)$ is a probability measure,

$$X(t, \omega, B(\omega)) = 1, \quad \text{with probability one}. \tag{7.21}$$

Then the conclusion (b) follows from (7.21) together with (7.19).

COROLLARY 7.1. *The Fleming-Viot random measure $X(t, \cdot)$ is singular for $t > 0$, in R^d, $d \ge 3$, with probability one.*

PROOF. This follows from the fact that in R^d sets of positive Lebesgue measure have Hausdorff-Besicovitch dimension d. If $X(t)$ were absolutely continuous and carried on a set $B(\omega)$, then $\dim B(\omega) = d \ge 3$, in contradiction to Theorem 7.1.

REMARK 7.1. According to Theorem 7.1, the microscopic scale distribution of three or more genetic characteristics is subject to a high degree of clustering. Although this clustering effect is brought into sharp focus in the continuous limit, it is also present in the discrete case. For the original Ohta-Kimura model for $d \ge 3$ genetic characteristics we have the following implication. If the incremental effect of a single mutation is sufficiently small, then the type of clustering described in terms of nonuniformity of subcell occupation frequency will appear at the scale of mean interparticle distance.

7.6. *Coherence of the random measure.*

THEOREM 7.2. *The Fleming-Viot process $X(t, \cdot)$ is compactly coherent.*

PROOF. Let $X^*(t)$ denote the random probability measure centered at the location $Z_t^*(t - A(t))$ of the founder, that is,

$$X^*(t) \equiv \theta_{\xi(t)} X(t, \cdot), \quad \text{where} \quad \xi(t) \equiv Z_t^*(t - A(t)). \tag{7.22}$$

It is a consequence of Lemma 6.7 that, assuming the appropriate initial condition, $\{X^*(t): t \ge 0\}$ is a stationary probability-measure-valued process. In particular, $\mathscr{L}(X^*(t))$ is a random cluster whose age $A^*(t)$ is distributed according to (6.30). Moreover, by Theorem 7.1, there is a random sphere $S(t)$ of radius $R(t)$ such that

$$X^*(t, S(t)) = 1, \quad \text{with probability one}. \tag{7.23}$$

Since $X^*(\cdot)$ is a stationary stochastic process, it follows that $\{R(t): t \ge 0\}$ is also a stationary stochastic process. Thus $X^*(t)$ is compactly coherent.

REMARK 7.2. Since we have no regularity on $R(t)$, the previous theorem does *not* imply that

$$\sup_{t_1 \le t \le t_2} R(t) < \infty \quad \text{with probability one}.$$

However, the results to be proved in Section 8, in particular (8.13) and Chebyshev's inequality, do imply the existence of a process $\{R_\varepsilon(t): t \ge 0\}$ for $\varepsilon > 0$ satisfying (3.11) and such that

$$\sup_{t_1 \le t \le t_2} R_\varepsilon(t) < \infty \quad \text{if} \quad t_2 < \infty. \tag{7.24}$$

REMARK 7.3. As a consequence of Lemma 6.8, $X^*(t)$ is an ergodic stationary probability-measure-valued process. Therefore, the Birkhoff ergodic theorem implies that

$$\lim_{T\to\infty} T^{-1} \int_0^T X^*(t)\, dt = E(X^*(0)), \quad \text{a.s.,} \tag{7.25}$$

where $E(X^*(0)) \in M_1(R^d)$.

8. Scaling limit of the wandering random measure. In the last section we established that the Fleming-Viot process is coherent and thus describes a wandering random probability measure. In this section we identify the scaling limit of the wandering random measure. This is done by rescaling space and time as follows:

$$Z_K(t, dx) \equiv X(K^2 t, K\, dx). \tag{8.1}$$

THEOREM 8.1. *Assume that $X(0)$ has compact support. Then*

$$Z_K(t, dx) \to \delta_{W(ct)}, \quad as \quad K \to \infty,$$

where $c = 2D$ and $W(\cdot)$ is a standard Brownian motion in R^d. The convergence is weak convergence of probability measures on Ω. In other words, $\delta_{W(ct)}$ is a measure-valued process consisting of a single unit atom undergoing Brownian motion in R^d.

PROOF. Let the empirical mean and covariance processes $x_i(\cdot)$ and $v_{ij}(\cdot)$ be defined as in (6.40) and (6.41). Then $x_i(t)$ is a martingale with increasing process

$$\ll x_i \gg_t = 2\gamma \int_0^t v_{ii}(s)\, ds.$$

We will denote by $x_i^K(t)$ the corresponding empirical mean process for the process $Z_K(\cdot,\cdot)$. Thus,

$$x_i^K(t) = K^{-1} x(K^2 t). \tag{8.2}$$

Then $x_i^K(t)$ is also a square integrable martingale, with increasing process

$$\ll x_i^K \gg_t = 2\gamma K^{-2} \int_0^{K^2 t} v_{ii}(s)\, ds \equiv T_K(t). \tag{8.3}$$

In view of Lemma 6.8, without loss of generality, we can assume that $v_{ii}(s)$ is an ergodic stationary stochastic process. Therefore

$$T_K(t) = 2\gamma K^{-2} \int_0^{K^2 t} v_{ii}(s)\, ds \to 2Dt \quad \text{as} \quad K \to \infty, \quad \text{a.s.,} \tag{8.4}$$

by the ergodic theorem and Lemma 6.12, part (ii). By a standard time-change argument for continuous real martingales (refer to Stroock and Varadhan, 1979, (6.6)), we have

$$x_i^K(t) = W_i(T_K(t)) \tag{8.5}$$

where $W_i(\cdot)$ is a standard Brownian motion. But then it follows from a result of Rebolledo (1977) on the weak convergence of martingales that

$$x_i^K(t) \to W_i(ct), \quad \text{where } c = 2D, \quad \text{as} \quad K \to \infty; \tag{8.6}$$

that is, $x_i^K(t)$ converges weakly to a constant multiple of a standard one-dimensional Brownian motion.

Next we proceed to show that the components $\{x_i^K(\cdot); i = 1, \cdots, d\}$ are asymptotically independent. Recall that

(8.7) $$\ll x_i^K, x_j^K \gg_t = (2\gamma/K^2) \int_0^{K^2 t} v_{ij}(s)\, ds.$$

Therefore, using the ergodic theorem,

(8.8) $$\lim_{K\to\infty} \ll x_i^K, x_j^K \gg_t = 2\gamma \lim_{t\to\infty} h_{ij}(t) = 0 \quad \text{if} \quad i \neq j,$$

by Lemma 6.12, part (ii). Therefore, the d components $(x_1^K(\cdot), \cdots, x_d^K(\cdot))$ converge weakly to d independent real-valued Brownian motions as $K \to \infty$.

The last step consists in proving that the limiting measure-valued process consists of a single atom which, in view of the above results, must be located at $W(ct)$. Consider

(8.9) $$v_{ii}^K(t) = \int x_i^2 Z_K(t, dx) - \left(\int x_i Z_K(t, dx)\right)^2, \quad i = 1, \cdots, d.$$

Then from Lemma 6.12, part (i), we have

(8.10) $$v_{ii}^K(t) = (2D/K^2\gamma)(1 - e^{-K^2\gamma t}) + (e^{-K^2\gamma t}/K^2) \int_0^{K^2 t} e^{\gamma s}\, dM_i(s).$$

By (6.55),

(8.11) $$\ll M_i \gg_t = \int_0^t q_i(s)\, ds,$$

where $q_i(\cdot)$ is given by (6.56). Hence, in view of Lemma 6.9 and Birkhoff's ergodic theorem,

(8.12) $$t^{-1} \ll M \gg_t \to \lim_{t\to\infty} E(q_i(t)) = \beta < \infty, \quad \text{a.s.}$$

Therefore for large K, the second term on the right-hand side of (8.10) behaves like

$$(e^{-K^2\gamma t}/K^2) W_0(\beta(e^{2\gamma K^2 t} - 1)),$$

where $W_0(\cdot)$ is a standard one-dimensional Brownian motion. The law of the iterated logarithm for $W_0(\cdot)$ implies that, for fixed $T < \infty$,

(8.13) $$\sup_{0\le t\le T} v_{ii}^K(t) \to 0 \quad \text{in probability as} \quad K \to \infty.$$

Finally, (8.4), (8.5), (8.13) and Chebyshev's inequality imply that

(8.14) $$Z_K(\cdot,\cdot) \to \delta_{W(ct)} \quad \text{in the topology of} \quad C([0, T], M_1(R^d)), \quad \text{in probability.}$$

This completes the proof of the theorem.

Acknowledgments. We wish to thank Wendell Fleming for his many stimulating suggestions and discussions. We would also like to thank Thomas Kurtz for useful remarks about time-change and duality arguments and Tokuzo Shiga, whose correspondence stimulated the present exposition of Section 5.2.

REFERENCES

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

BROWN, A. H. D., MARSHALL, D. R., and ALBRECHT, L. (1975). Profiles of electrophoretic alleles in natural populations. *Genetics Res. Camb.* **25** 137–143.

CHAKRABORTY, R., and NEI, M. (1976). Hidden genetic variability within electromorphs in finite populations. *Genetics* **84** 385–393.

CHOW, Y. S., and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Springer, New York.

DAWSON, D. A., (1970). *Denumerable Markov Chains.* Canadian Mathematical Congress, Montreal.

DAWSON, D. A., and HOCHBERG, K. J. (1979). The carrying dimension of a stochastic measure diffusion. *Ann. Probability* **7** 693–703.

DELLACHERIE, C., and MEYER, P. A. (1978). *Probabilities and Potential.* North Holland and Hermann, Amterdam.

ERDÉLYI, A. (ed.), MAGNUS, W., OBERHETTINGER, F., and TRICOMI, F. G. (1954). *Table of Integral Transforms,* Vol. 1, Bateman Manuscript Project. McGraw-Hill, New York.

ETHIER, S. N., (1981). A class of infinite-dimensional diffusions occurring in population genetics. *Indiana Univ. Math. J.* **30** 925–935.

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

EWENS, W. J. (1979). *Mathematical Population Genetics. Biomathematics* **9.** Springer, New York.

FEDERER, H. (1969). *Geometric Measure Theory.* Springer, New York.

FELLER, W. (1966). *An Introduction to Probability Theory and Is Applications,* Vol. 2. Wiley, New York.

FLEMING, W. H. and VIOT, M. (1978). Some measure-valued population processes, in *Proc. International Conference on Stochastic Analysis, Northwestern University.* Academic, New York, 97–108.

FLEMING, W.H. and VIOT, M. (1979). Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J.* **28** 817–843.

FUERST, P. A. and FERRELL, R. E. (1980). The stepwise mutation model: an experimental evaluation utilizing hemoglobin variants. *Genetics* **94,** 185–201.

HOLLEY, R., STROOCK, D. and WILLIAMS, D. (1977). Applications of dual processes to diffusion theory. *Proc. Symp. Pure Math.* **31** Amer. Math. Soc., 23–36.

KESTEN, H. D. (1980a). The number of distinguishable alleles according to the Ohta-Kimura model of neutral mutation. *J. Math. Biology* **10** 167–187.

KESTEN, H. (1980b). The number of alleles in electrophoretic experiments. *Theor. Pop. Biol.* **18** 290–294.

KIMURA, M. and CROW, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49** 725–738.

KIMURA, M. and OHTA, T. (1975). Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. *Proc. Nat. Acad. Sci. U.S.A.* **72** 2761–2764.

KIMURA, M. and OHTA, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc. Nat. Acad. Sci. U.S.A.* **75** 2868–2872.

KINGMAN, J. F. C. (1975). Random discrete distribution, *J. Roy. Statist. Soc.* **B37** 1–22.

KINGMAN, J. F. C. (1976). Coherent random walks arising in some genetical models. *Proc. Roy. Soc. London A.* **351** 19–31.

KINGMAN, J. F. C. (1978). Uses of exchangeability, *Ann. Probability* **6** 183–197.

KURTZ, T. G. (1980). *Approximation of Population Processes,* CBMS-NSF Regional Conference Series in Applied Mathematics **36,** Society for Industrial and Applied Mathematics, Philadelphia.

MORAN, P. A. P. (1975). Wandering distributions and the electrophoretic profile. *Theor. Pop. Biol.* **8** 318–330.

MORAN, P. A. P. (1975). Wandering distributions and the electrophoretic profile, II. *Theor. Pop. Biol.* **10** 145–149.

OHTA, T. and KIMURA, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res. Camb.* **22** 201–204.

OHTA, T. and KIMURA, M. (1974). Simulation studies on electrophoretically detectable genetic variability in a finite population. *Genetics* **76** 614–624.

RAMSHAW, J. A., COYNE, J. A. and LEWONTIN, R. C. (1979). The sensitivity of gel electrophoresis as a detector of genetic variation. *Genetics* **93** 1019–1037.

REBOLLEDO, P. (1977). Remarques sur la convergence en loi des martingales vers des martingales continues. *C. R. Acad. Sci. Paris* **285** 517–520.

ROGERS, C. A. (1970). *Hausdorff Measures.* Cambridge University Press.

SHIGA, T. (1981). Wandering phenomena of an infinite dimensional diffusion model in population genetics—continuous time Ohta-Kimura model, personal communication.

STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes,* Grundlehren der mathematischen Wissenschaften **233.** Springer, Berlin.

WATTERSON, G. A. (1976). The stationary distribution of the infinitely many neutral alleles diffusion model. *J. Appl. Probability* **13** 639–651.

DEPARTMENT OF MATHEMATICS AND STATISTICS
CARLETON UNIVERSITY
OTTAWA, CANADA KIS 5B6.

DEPARTMENT OF MATHEMATICS AND STATISTICS
CASE WESTERN RESERVE UNIVERSITY
CLEVELAND, OH 44106