TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 347, Number 7, July 1995

EQUILIBRIA AND QUASI-EQUILIBRIA FOR INFINITE COLLECTIONS OF INTERACTING FLEMING-VIOT PROCESSES

DONALD A. DAWSON, ANDREAS GREVEN, AND JEAN VAILLANCOURT

ABSTRACT. In this paper of infinite systems of interacting measure-valued diffusions each with state space $\mathscr{P}([0, 1])$, the set of probability measures on $[0, 1]$, is constructed and analysed (Fleming-Viot systems). These systems arise as diffusion limits of population genetics models with infinitely many possible types of individuals (labelled by $[0, 1]$), spatially distributed over a countable collection of sites and evolving as follows. Individuals can migrate between sites and after an exponential waiting time a colony replaces its population by a new generation where the types are assigned by resampling from the empirical distribution of types at this site.

It is proved that, depending on recurrence versus transience properties of the migration mechanism, the system either clusters as $t \to \infty$, that is, converges in distribution to a law concentrated on the states in which all components are equal to some δ_u, $u \in [0, 1]$, or the system approaches a nontrivial equilibrium state. The properties of the equilibrium states, respectively the cluster formation, are studied by letting a parameter in the migration mechanism tend to infinity and explicitly identifying the limiting dynamics in a sequence of different space-time scales. These limiting dynamics have stationary states which are quasi-equilibria of the original system, that is, change only in longer time scales. Properties of these quasi-equilibria are derived and related to the global equilibrium process for large N. Finally we establish that the Fleming-Viot systems are the unique dynamics which remain invariant under the associated space-time renormalization procedure.

0. INTRODUCTION

(a) **Background and motivation.** In the present paper, we construct a system consisting of countably many interacting Fleming-Viot processes. Each component takes values in the space of probability measures on a compact space, say $[0, 1]$. This model arises as the diffusion limit of the following model from population genetics. The population is spatially distributed among a collection of colonies in which there are individuals of various genetic types and these types are labelled via values in $[0, 1]$. The types of individuals in the next generation in each colony are obtained by *sampling* according to the empirical frequency of current types within the colony. In addition individuals can *migrate* between colonies.

Received by the editors January 20, 1994.

1991 *Mathematics Subject Classification.* Primary 60K35; Secondary 60J70.

Key words and phrases. Fleming-Viot process, hierarchical group, stability-clustering dichotomy, renormalization.

Depending on the strength of the migration relative to the resampling, the system either *clusters*, that is, regions consisting of colonies of a single type grow unboundedly, or an equilibrium exists (*stability*), which preserves the initial spatial frequency of types even in the limit $t \to \infty$. It is this longterm behavior that we want to understand better and to analyse in more detail. There are two main guiding ideas behind our approach: multiple space-time analysis and renormalization. We now explain these two viewpoints.

(i) In general it is difficult to completely describe the *space-time structure* of interacting systems in *equilibrium* when the components are dependent. Previously, attempts have been made to get around this by averaging over components to obtain limiting Gaussian fields which are easier to describe. We shall take a different approach in this paper. It turns out that the system displays the feature that after long time spans a sequence of states occur which when viewed locally remain *almost stationary over long periods of time* and which are easier to describe since they are determined by simple macroscopic variables (block averages). The later we look at the system the more closely the corresponding *quasi-stationary system* resembles the real equilibrium space-time process.

We construct the quasi-equilibria by considering a sequence of space and time scales (multiple space-time scale analysis). This way we obtain a sequence of renormalized systems. After taking a limit of a parameter tending to infinity we obtain a sequence of limiting dynamics whose equilibria generate a sequence of distributions of components which we call the *interaction chain.* We read off many qualitative properties of the system from this object since it also gives a sufficiently precise description of the *space-time structure of the equilibrium process.* The interaction chain contains the essence of the space-time dependence structure of the system in equilibrium and therefore we view it as a central object in the study of interacting systems. (Compare Dawson and Grevan (1933b), Dawson and Greven (1993c) for applications of these ideas to interacting diffusions.) The quasi-equilibria can be described explicitly as equilibria for *independent Fleming-Viot processes with immigration* evolving in a *random medium* which provides the sources of immigration. As mentioned already, in order to carry out the above program, we consider a *limit* of a sequence of systems of countably many Fleming-Viot processes, in which the range of migration increases unboundedly; we refer to this as the mean-field limit even though all systems considered will have infinitely many components.

The multiple space-time scale analysis also allows us to study via the interaction chain some typical features of this system with infinite-dimensional components. In particular in the stable case we can make precise statements about the spatial density of a specific type observed in a certain colony and about the number of distinct types per volume which make up a significant part of the population in a large volume. In the case of clustering we find that the sizes of the monotype regions grow at different scales, which we also determine.

(ii) Next we turn to the renormalization viewpoint which is another important feature related to the *multiple space-time scale analysis.* Many qualitative properties of evolutions involving linearly interacting components are shared by an entire class of evolutions which is called a *universality class.* In order to explain this phenomenon we view the multiple space-time scale analysis as a *renormalization procedure.* We form averages over blocks (space-scaling) and change the time scale. This gives a new interacting system. The procedure is repeated and then we take the limiting case in which the range of migration tends to infinity. This yields a sequence of evolutions each belonging to a particular block average and time scale. If we start with Fleming-Viot systems, all elements of the sequence are also processes of that type. We prove that this property is characteristic for Fleming-Viot systems. We cannot yet prove that other dynamics lead to a sequence converging to the same fixed point in sufficiently strong topologies which would demonstrate the universality of all qualitative properties of the behavior for large times. However we do get partial results and for the simpler case of a two type system such an analysis has been carried out completely (Dawson and Greven (1993c), Baillon et al. (1994)). It is this perspective that motivates us to use methods of proof for the multiple space-time scale analysis which can be easily adapted to more general processes—we avoid using the special features of the Fleming-Viot process.

Let us outline the rest of this chapter. We continue in subsection (b) to construct the model (Theorem 0.0), in (c) we study the behavior as $t \to \infty$ (Theorem 0.1), in section (d) we introduce the multiple space-time scale analysis (Theorems 0.2, 0.3) which we extend in (e) to construct quasi-equilibria and finally in (f)–(h) we study the qualitative properties of the quasi-equilibria via the interaction chain (Theorems 0.5–0.10). In subsection (i) we discuss extensions and open problems.

(b) **The model.** We consider an interacting system with a countable number of components (subpopulations at spatially distributed sites), where the components shall take values in $\mathscr{P}([0,1])$, the space of probability measures on $[0,1]$. Thus the state space of our process is $\mathscr{P}([0,1])^I$ with I a countable set. Such processes arise as diffusion limits of discrete time genetic models with a continuum (namely $[0,1]$) of possible types and evolution driven by *resampling* within colonies (components) and *migration* between components (cf. Ethier and Kurtz (1986, Chapter 10) for an introduction to diffusion limits of genetic models and Vaillancourt (1990) for the diffusion limit of genetic models with migration). We introduce in (i)–(iv) below the ingredients needed for a formal definition of the model.

(i) There are two natural candidates for the *index set* I for the components of our interacting systems: one is $\mathbb{Z}^n$ and the other is the so-called hierarchical group Ω_N (where N is a fixed parameter), which is defined as follows:

(0.1)
$$\Omega_N = \{\xi | \xi = (\xi_1, \xi_2, \dots), \xi_i \in \mathbb{N}, 0 \le \xi_i \le N-1, \exists k_0 \text{ with } \xi_j = 0 \ \forall j \ge k_0\}.$$

We view Ω_N as an abelian group by defining addition as the componentwise addition of natural numbers modulo N. For $\xi, \xi' \in \Omega_N$ define the *hierarchical distance*

$$d(\xi, \xi') := \inf(k-1 | \xi'_j = \xi_j \ \forall j \ge k), \quad \text{where } k \in \{1, 2, \dots\}. \tag{0.2}$$

The element $\xi = (\xi_1, \xi_2, \dots)$ represents the ξ_1-individual in the ξ_2-family of the ξ_3-clan in the ξ_4-village.... In the genetics context this hierarchical structure is more natural than the structure of $\mathbb{Z}^n$ with the euclidean notion of distance. Note that each Ω_N can be viewed as a subset of Ω_M for $M \ge N$ and $\Omega_\infty = \bigcup_M \Omega_M \subseteq (\mathbb{N})^{\mathbb{N}}$.

(ii) In order to describe the interaction (migration) between the components we use a random walk transition kernel on I $(= \mathbb{Z}^n$ or $\Omega_N)$ denoted by $a(\cdot, \cdot)$, with the property

$$(0.3)\qquad \begin{gathered} a(\xi, \xi') \in [0, 1], \quad a(\xi, \xi') = a(0, \xi - \xi'), \\ \sum_{\xi} a(0, \xi) = 1, \quad \sum_{n=0}^{\infty} (a^n(0, \xi) + a^n(\xi, 0)) > 0 \quad \forall \xi. \end{gathered}$$

(iii) We shall define a Markov process $X = (X(t))_{t \geq 0}$ (the *Fleming-Viot process with interacting components*) on the state space (endowed with the product topology)

$$(0.4)\qquad \mathbf{E} = (\mathscr{P}([0, 1]))^I$$

via a *martingale problem*. We introduce the following notation: an element of $\mathbf{E}$ is written $x = (x_\xi)_{\xi \in I}$ with $x_\xi \in \mathscr{P}([0, 1])$ and variables in $[0, 1]$ are denoted by u, v. The collection of continuous functions from $[0, T]$ into $\mathbf{E}$ is denoted by $C([0, T], E)$, (pathspace).

(iv) In order to define the *generator* L of X we introduce three objects: the *algebra* $\mathscr{A}$ of functions F on $\mathbf{E}$ which are linear combinations of functions of the form

(0.5)

$$F(x) = \int_0^1 \cdots \int_0^1 f(u_1, \ldots, u_m) x_{\xi(1)}(du_1) \cdots x_{\xi(m)}(du_m), \qquad f \in C([0, 1]^m),$$

where $(\xi(i))_{i=1, \ldots, m}$ is any finite collection of elements in I. This includes, for example, functions of the form

$$F(x) = \prod_{i=1}^m \int_0^1 u^{a_i} x_{\xi(i)}(du), \qquad a_i \in \mathbb{N}.$$

Giving $E(F)$ for all $F \in \mathscr{A}$ determines uniquely a law on $(\mathscr{P}([0, 1]))^I$. We define the *diffusion coefficient*, $Q_{x_\xi}(\cdot, \cdot)$, to be the signed measure on $[0, 1] \times [0, 1]$ given by

$$(0.6)\qquad Q_{x_\xi}(du, dv) = x_\xi(du)\delta_u(dv) - x_\xi(du)x_\xi(dv).$$

For elements in $\mathscr{A}$ we define *partial derivatives* as follows:

$$(0.7)\qquad \frac{\partial F(x)}{\partial x_\xi}(u) = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}(F(x^\varepsilon(\xi, u)) - F(x))$$

with

$$(x^\varepsilon(\xi, u))_{\xi'} := \begin{cases} x_\xi, & \text{if } \xi' \neq \xi, \\ x_\xi + \varepsilon\delta_u(\cdot) & \text{if } \xi' = \xi. \end{cases}$$

Definition. For every $F \in \mathscr{A}$ define

$$(0.8)\qquad \begin{aligned} L(F)(x) = {} & \sum_{\substack{\xi' \in I \\ \xi \in I}} a(\xi, \xi') \int_{[0,1]} \left(\frac{\partial F(x)}{\partial x_\xi}(u) \right) (x_{\xi'}(du) - x_\xi(du)) \\ & + d \sum_{\xi \in I} \int_{[0,1]} \int_{[0,1]} \left(\frac{\partial^2 F(x)}{\partial x_\xi \partial x_\xi}(u, v) \right) Q_{x_\xi}(du, dv). \end{aligned}$$

The first term is the migration term and the second is the resampling term. Consider the cases $I = \Omega_N$ or $I = \mathbb{Z}^n$. Then we have the following.

Theorem 0.0 (Existence, Uniqueness, Path properties).

(a) *For every initial distribution* μ *on* $(\mathscr{P}([0,1]))^I$ *the martingale problem for* (L,μ) *is well posed in* $\mathscr{P}(C([0,\infty);(\mathscr{P}([0,1]))^I))$. *The resulting process is denoted* $(X(t))_{t\geq 0}$ *(system of interacting Fleming-Viot processes on* I*).*

(b) *This process is Markov and Feller.*

(c) *The values of* $X(t)$ *for* $t>0$ *are in the set of atomic measures; in fact,* $(X(t))_{t>0}\in C((0,\infty),(\mathscr{P}_a([0,1]))^I)$ *a.s. where* $\mathscr{P}_a$ *are the atomic probability measures on* $[0,1]$. □

Remark. In other words, according to (a) there exists a unique distribution on $\mathscr{P}(C([0,\infty);(\mathscr{P}([0,1]))^I))$ such that for the canonical process $(X_t)_{t\geq 0}$, the following holds: $\mathscr{L}(X(0))=\mu$, and

$$\left(F(X_t)-\int_0^t (LF)(X_s)\,ds\right)_{t\geq 0} \text{ is a martingale for every } F\in\mathscr{A}. \tag{0.9}$$

(c) **The long term behavior of the process.** The behavior of the process $(X(t))_{t\geq 0}$ is analogous to the case of one-dimensional components, that is, a system of interacting Fisher-Wright diffusions describing the frequencies of two types (compare (0.56)). For $t\to\infty$ the qualitative properties depend solely on the migration mechanism, that is, on the kernel $a(\cdot,\cdot)$ on the group I.

Before we state the theorem we introduce the following objects (i)–(iv):

(i) The appropriate class of *initial distributions* μ is the following:

$$\mathscr{T}=\{\mu\in\mathscr{P}(\mathscr{P}([0,1])^I)\,|\,(\alpha)\text{ and }(\beta)\text{ hold}\} \tag{0.10}$$

where

(α) μ is a homogeneous measure, that is, invariant under the map $(x_\xi)_{\xi\in I}\xrightarrow{T_\zeta}(x_{\xi+\zeta})_{\xi\in I}$ for all $\zeta\in I$.

(β) μ is ergodic (the σ-field of events invariant under all maps T_ζ, $\zeta\in I$, is trivial under μ).

We shall also denote the set of homogeneous measures on $(\mathscr{P}([0,1]))^I$, that is, satisfying (α), by $\widetilde{\mathscr{T}}$.

(ii) For every measure $\mu\in\mathscr{T}$ we can introduce the *mean measure* $\theta(\mu)\in\mathscr{P}([0,1])$ (abbreviated often simply by θ in the sequel), by the defining relation:

$$\langle\theta(\mu),f\rangle=E^\mu\langle x_\xi,f\rangle\quad\forall f\in L_\infty([0,1]). \tag{0.11}$$

This allows us to decompose the class $\mathscr{T}$ into disjoint subclasses $\mathscr{T}_\theta$, where

$$\mathscr{T}_\theta=\{\mu\in\mathscr{T}\,|\,\theta(\mu)=\theta\}. \tag{0.12}$$

(iii) The set $\mathscr{J}$ denotes the set of all *equilibrium states*, that is,

$$(\mu\in\mathscr{J},\ \mathscr{L}(X(0))=\mu)\Rightarrow(\mathscr{L}(X(t))=\mu,\ \forall t\geq 0).$$

The extremal elements of a convex set K are denoted by K_e.

(iv) Finally we need the following symmetric homogeneous kernel on $I\times I$ describing the distance process of two particles moving as $a(\cdot,\cdot)$-random walks:

$$\hat{a}(\xi,\zeta)=\frac{1}{2}(a(\xi,\zeta)+a(\zeta,\xi)). \tag{0.13}$$

Then there are two cases, namely, $\hat{a}$ recurrent and $\hat{a}$ transient, corresponding to whether $\sum_1^\infty \hat{a}^n=\infty$ or $<\infty$.

Theorem 0.1 (a) (invariant measures).

Case 1. *$\hat{a}$ is transient.*

(0.14) *For every $\theta \in \mathscr{P}([0, 1])$ the limit ν_θ of $\mathscr{L}(X(t)|X(0) = \theta^I)$ exists for $t \to \infty$ and satisfies:*

$$\nu_\theta \in (\mathscr{I} \cap \widetilde{\mathscr{T}})_e, \quad \langle \theta, f \rangle = E^{\nu_\theta}\langle x_\xi, f \rangle \ \forall \xi \in I;$$

ν_θ is ergodic, in particular

$$E^{\nu_\theta}(\langle x_\xi, f \rangle \langle x_\zeta, f \rangle) \underset{d(\xi, \zeta) \to \infty}{\to} \langle \theta, f \rangle^2 \quad \forall f \in L_\infty([0, 1]).$$

Furthermore

$$(\mathscr{I} \cap \widetilde{\mathscr{T}})_e = \{\nu_\theta\}_{\theta \in \mathscr{P}([0, 1])}. \tag{0.15}$$

Case 2. *$\hat{a}$ is recurrent.*

$$(\mathscr{I} \cap \widetilde{\mathscr{T}})_e = \{\delta_{(\delta_u)^I}\}_{u \in [0, 1]}. \tag{0.16}$$

(b) (long term behavior)

For every initial measure $\mu \in \mathscr{T}_\theta$ the following dichotomy holds:

$$\hat{a} \text{ transient: } \mathscr{L}(X(t)) \underset{t \to \infty}{\Rightarrow} \nu_\theta, \tag{0.17}$$

$$\hat{a} \text{ recurrent: } \mathscr{L}(X(t)) \underset{t \to \infty}{\Rightarrow} \int_{[0, 1]} \delta_{(\delta_u)^I} \theta(du). \quad \square \tag{0.18}$$

In the first case we have a situation in which the system allows for the local coexistence of types, while in the second case we see locally one type prevailing, as long as t becomes large. In the first case the system is called *stable* and in the second case we say that the system *clusters.* We shall now apply this result to the case where $I = \mathbb{Z}^n$, respectively $I = \Omega_N$.

Corollary 0.1. *Suppose that $I = \mathbb{Z}^n$ and $a(\cdot, \cdot)$ satisfies $\sum_\xi a(0, \xi)|\xi|^2 < \infty$. Then for*

$n = 1, 2$ *the system clusters,*

$n \geq 3$ *the system is stable.* $\square$

A *mean-field random walk* on Ω_N can be written in the form

$$a(0, \xi) = \sum_{j \geq k} \beta_j N^{-j} \quad \text{for } d(0, \xi) = k. \tag{0.19}$$

Here β_k denotes the probability to jump to an element at a distance of at most k and the number of elements ξ with $d(0, \xi) \leq k$ is N^k. We can also write this in the form

$$a(0, \xi) = \sum_{j \geq k} (c_{j-1} N^{-(j-1)}) N^{-j} \quad \text{for } d(0, \xi) = k, \ k \geq 1, \tag{0.20}$$

with $(c_k)_{k \in \mathbb{N}}$ as canonical parameters and by assumption the total jump rate $\sum_k c_k / N^k < \infty$.

Remark. We will show in Corollary 0.2 that the qualitative behavior of the system depends on the sequence c_k but provided that $\overline{\lim}_{k \to \infty} \sqrt[k]{c_k} < N$ it does not depend on N. Since we are primarily interested in the behavior of the systems for large N it would suffice to assume that $\sum_k c_k/N_0^k < \infty$ for some positive integer N_0.

Introduce

$$C_m = \sum_{k=m}^{\infty} \tilde{c}_k N^{-(k-m)} \quad \text{with } \tilde{c}_m = \sum_{k=m}^{\infty} c_k N^{-2(k-m)}. \tag{0.21}$$

Corollary 0.2. *Suppose that* $I = \Omega_N$ *and* $a(\cdot, \cdot)$ *is defined via* $(c_k)_{k\in\mathbb{N}}$ *as in* (0.20). *Then for*

$$\sum_m C_m^{-1} = \infty \text{ the system clusters,} \tag{0.22}$$

$$\sum_m C_m^{-1} < \infty \text{ the system is stable.} \tag{0.23}$$

$\sum_m c_k^{-1} < \infty$ *implies that* $\sum_k C_k^{-1} < \infty$. *Moreover, if* $\overline{\lim}_{k\to\infty} \sqrt[k]{c_k} < N$, *then* $\sum C_k^{-1} < \infty, = \infty$ *is equivalent to* $\sum c_k^{-1} < \infty, = \infty$. □

(d) **The multiple space-time scale behavior.** In the sequel we shall only consider the case where $I = \Omega_N$. Since we shall vary N later on, from now on we write

$$(X^N(t))_{t\geq 0} \tag{0.24}$$

for a system of interacting Fleming-Viot processes on Ω_N. As initial distribution we choose a measure which is the restriction to $\mathscr{P}([0, 1])^{\Omega_N}$ of a product measure μ in $\mathscr{T}_\theta$ on $\mathscr{P}([0, 1])^{\Omega_\infty}$. We denote this measure by μ as well. This special choice is merely to simplify the notation and to avoid the introduction of sequences of initial measures.

In Theorem 0.1 we gave a qualitative description of the long term behavior of our system. However relatively little has been said about the process with marginal ν_θ appearing there as equilibria in the stable case, or about the formation of clusters in the case of clustering. One observes however that for sufficiently large N the system, when observed in a whole *sequence* of combined *fast* and *slow time scales* (that means here at times of the form: $sN + t$, $sN^2 + t$, $sN^3 + t, \ldots$ with t varying, but small compared to N), will look almost stationary. Since the interaction is still local, these almost stationary states will be much easier to describe, compared to the stationary process (for fixed values of N). Here t is the microscopic (slow) and s the macroscopic (fast) time scale; it is in the latter that we see changes in the quasi-equilibria. Here we should understand what happens between times $sN, sN^2, \ldots$. For this purpose we introduce a *rescaling of space* by forming block averages over blocks of $N, N^2, N^3, \ldots$ components.

In order to make this analysis rigorous, we consider the limit $N \to \infty$, so that the various time scales $s, sN^2, sN^3, \ldots$ and block sizes $N, N^2, \ldots$ all *separate*, that is, become incomparable, and we can precisely describe the phenomenon of quasi-stationary states via appropriate limit theorems.

In order to carry out this program, we shall need the following ingredients of *multiple space-time scale analysis*: for $k = 1, 2, 3, \ldots$,

(i) *Time scales* $\beta_k(N) = N^k$.

(ii) *Block averages* on level k (space-scaling):

$$x_{\xi,k} = N^{-k} \sum_{\xi' : d(\xi,\xi') \leq k} x_{\xi'} \text{ and } x_{\xi,0} := x_\xi, \quad \forall \xi \in \Omega_N.$$

(iii) The *quasi-equilibrium* $\Gamma_\theta^k(\cdot)$ on the kth level, the kth level *associated measure-valued diffusion* on $\mathscr{P}([0,1])$ denoted by $(Z_t^{\theta,k})_{t\geq 0}$ and the kth *level fluctuation constant* d_k:

$-d_k$, the diffusion constants are defined for a given sequence $(c_k)_{k\in\mathbb{N}}$ via $d_{k+1} = d_k c_k/(c_k + d_k)$, $d_0 = 1$.

$-\Gamma_\theta^k(\cdot) \in \mathscr{P}(\mathscr{P}([0,1]))$ is the unique equilibrium state of $\tilde{Z}^{\theta,k}$.

$-(\tilde{Z}_t^{\theta,k})$ is a Fleming-Viot process with immigration from a source $\theta \in \mathscr{P}([0,1])$ and with initial state $\tilde{Z}_0^{\theta,k}$. The associated equilibrium process is denoted $(Z_t^{\theta,k})_{t\in\mathbb{R}}$. Formally $(\tilde{Z}_t^{\theta,k})$ is the solution of the well-posed (see Lemma 2.1 in section 2) martingale problem $(L_{\theta,k}, \delta_\theta)$ on $C([0,\infty), \mathscr{P}([0,1]))$, where $L_{\theta,k}$ is defined on an algebra $\mathscr{B}$ of functions on $\mathscr{P}([0,1])$ as follows: for $G \in \mathscr{B}$

$$L_{\theta,k}(G)(z) = c_k \int_{[0,1]} \frac{\partial G}{\partial z}(u)(\theta(du) - z(du))$$
$$+ d_k \int_{[0,1]} \int_{[0,1]} \frac{\partial^2 G}{\partial z \partial z}(u,v) Q_z(du,dv)$$

with $Q_z(du,dv) = z(du)\delta_u(dv) - z(du)z(dv)$ and $\mathscr{B}$ is the algebra generated by functions G from $\mathscr{P}([0,1])$ to $\mathbb{R}$ of the form

$$G(z) = \prod_{i=1}^m \int_{[0,1]} f_i(u_i) z(du_i) \quad \text{for } f_i \in C([0,1]),\ i = 1, \ldots, m.$$

(iv) The *level k-marginals* of the *interaction chain* at level j $(j \geq k)$:

$\mu_\theta^{j,k}$ is a measure on $\mathscr{P}([0,1])$ defined by

$$\mu_\theta^{j,k}(d\theta_k) = \int_{\mathscr{P}([0,1])} \cdots \int_{\mathscr{P}([0,1])} \Gamma_\theta^j(d\theta_j)\Gamma_{\theta_j}^{j-1}(d\theta_{j-1}) \cdots \Gamma_{\theta_{k+1}}^k(d\theta_k).$$

Notation. We shall denote by $\bar{\theta}$ the function $\mathbb{R}^+ \to \mathscr{P}([0,1])$ which is identically equal to θ. We denote by $\Rightarrow$ weak convergence in path space $D([0,\infty), \mathscr{P}([0,1])$ or $C([0,\infty), \mathscr{P}([0,1]))$ and we denote the associated canonical process by $(X(t))_{t\in\mathbb{R}^+}$.

We are now ready to formulate our result on the *multiple space-time scale behavior* of the process $X^N(t) = (x_\xi^N(t)_{\xi\in\Omega_N})$. We shall first focus on block averages around a single component ξ. The discussion of the behavior of a spatial collection of averages is deferred to Theorem 0.4 in the next subsection.

Theorem 0.2 (Multiple space-time scale behavior). *For $k, j \in \mathbb{N}$ we have the following behavior for every $\xi \in \Omega_\infty$ and for $\mu \in \mathscr{T}_\theta$:*

Case 1. $k > j$ $\mathscr{L}((x_{\xi,k}^N(t\beta_j(N)))_{t\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \delta_{\bar{\theta}}$, *Case* 2. $k = j$ $\mathscr{L}((x^N_{\xi,k}(t\beta_k(N)))_{t\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((\tilde{Z}^{\theta,k}_t)_{t\in\mathbb{R}^+})$,

Case 3. $k < j$ $\mathscr{L}((x^N_{\xi,k}(s\beta_j(N) + t\beta_k(N)))_{t\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((Z^{\theta^*,k}_t)_{t\in\mathbb{R}^+})$,

θ^* *and the evolution of the processes are independent, and*

$$\mathscr{L}(\theta^*) = \int \mu^{j-1,k}_\rho P(\tilde{Z}^{\theta,j}_s \in d\rho). \quad \square$$

Remark. In Case 3, if we let $s \to \infty$, $s = o(N)$, then we get the simple formula $\mathscr{L}(\theta^*) = \mu^{j,k}_\theta$.

Remark. This shows that in time scale $\beta_j(N)$ large blocks $\{\xi : d(\xi, 0) \le k\}$ with $k > j$ display an empirical distribution θ of types which for $N \to \infty$ equals the original distribution of types in the thermodynamic limit. On the other hand blocks of moderate or small size $k = j$, $k < j$, viewed after times of order $\beta_j(N)$ already display in the time scale $\beta_k(N)$ an empirical distribution which is random and fluctuates like a Fleming-Viot process with immigration distribution of types described by θ, respectively θ^*. For small k, that is, $k < j$, the process has already relaxed into equilibrium.

Remark. The point of the theorem is to combine two ideas: first to consider slow and fast time scales together $(s\beta_j(N) + t\beta_k(N))$ and second to look at a whole sequence of such pairs of time scales combined with a suitable sequence of rescalings of space.

The very nice thing about Theorem 0.2 is that we can now quite explicitly determine the ingredients which define the limiting dynamics. The crucial object is the measure, $\Gamma^k_\theta(\cdot)$, in $\mathscr{P}(\mathscr{P}([0, 1]))$ which can be represented using the following ingredients:

(i) let $(U_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence with marginal distribution $\theta \in \mathscr{P}([0, 1])$,

(ii) let $(V^k_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence with marginal $\mathrm{Beta}(1, \gamma_k^{-1})$ where $\gamma_k = d_k/c_k$,

(iii) construct $(U_i)_{i\in\mathbb{N}}$, $(V^k_i)_{i\in\mathbb{N}}$ as independent processes.

Theorem 0.3 (Identification and continuity of Γ^k_θ).

$$\Gamma^k_\theta = \mathscr{L}\left(\sum_{i=1}^{\infty}\left[V^k_i \prod_{j=1}^{i-1}(1 - V^k_j)\right]\delta_{U_i}(\cdot)\right); \tag{0.25a}$$

(0.25b) *the mapping* $\theta \to \Gamma^k_\theta$ *is continuous.* $\square$

Remark. This means that the measure Γ^k_θ is the law of a random measure, which is atomic. The positions of the atoms are given by the sequence (U_i), where the law θ enters. The weights are chosen independently of the positions and according to the sequence $(V^k_\ell \prod_{j=1}^{\ell-1}(1 - V^k_j))_{\ell\in\mathbb{N}}$, where the parameters (c_k) and (d_k) of the evolution enter. The law of the sequence of weights is called the $\mathrm{GEM}(\gamma_k^{-1})$ distribution.

(e) **Description of the quasi-equilibria.** This subsection completes the multiple space-time scale analysis. In the previous section we have studied the block average behavior around a single site of the process $(X^N(t))_{t\ge0}$ in various time scales. Our original objective went further: we wanted to determine the *quasi-equilibria* which the whole system $(X^N(t))_{t\geq 0}$ displays. This means we want to consider slow and fast time scales and to see whether for each k

$$(X^N(s\beta_k(N)+t))_{t\geq 0} \underset{N\to\infty}{\Rightarrow} (X^{\infty}_{k,s}(t))_{t\geq 0} \tag{0.26}$$

and whether it is possible to identify this limit evolution, $(X^{\infty}_{k,s}(t))_{t\geq 0}$ for the various time scales β_k and values of $s>0$. The limit $s\to\infty$ gives a process $(X^{\infty}_k(t))_{t\geq 0}$. The identification of these limit dynamics is the purpose of this section (see Theorem 0.4 below).

The evolution $(X^{\infty}_k(t))_{t\geq 0}$ will turn out to be a process on $(\mathscr{P}([0,1]))^{\Omega_\infty}$ which can be viewed as *evolution in a random medium*. We describe the process in three steps (i)–(iii). First in (i) we describe the evolution mechanism given a random field (the random medium): the latter is constructed formally in (ii) and is of the form

$$(\theta^{(k)}_\xi)_{\xi\in\Omega_\infty} \text{ with values in } \mathscr{P}([0,1]), \tag{0.27}$$

and has the structure of a *Markov random field on a tree*.

Finally we construct the initial state in (iii).

(i) *The dynamics for the given random field*: Conditioned on the random field in (0.27) the evolution of the components of $X^{(\infty)}_{k,s}$ are independent Fleming-Viot processes with immigration $\theta^{(k)}_\xi$ in equilibrium. This means that the law of $(X^{\infty}_{k,s}(t))_{t\geq 0}$ is a Fleming-Viot system in a random environment, the environment being the field $(\theta^{(k)}_\xi)_{\xi\in\Omega_\infty}$ of sources of immigration. The law of this field is s-dependent, but in a mild way and for this reason we suppress this in the notation.

Formally the evolution is defined as follows. For given random medium $(\theta^{(k)}_\xi)_{\xi\in\Omega_\infty}$ the components evolve independently and the evolution of a single component is defined as the solution of the martingale problem $(L_{\theta^{(k)}_\xi},\mu)$ on $\mathscr{P}(C([0,\infty):\mathscr{P}([0,1])))$ with

$$\begin{aligned} L_{\theta^{(k)}_\xi} &= c_0\int_{[0,1]}\frac{\partial G(z)}{\partial z}(u)(\theta^{(k)}_\xi(du)-z(du)) \\ &\quad + d_0\int_{[0,1]}\int_{[0,1]}\frac{\partial^2 G(z)}{\partial z\partial z}(u,v)Q_z(du,dv) \end{aligned} \tag{0.28}$$

where

$$Q(du,dv)=z(du)\delta_u(dv)-z(du)z(dv),$$

$$G(z)=\int_0^1\cdots\int_0^1 f(u_1,\dots,u_n)z(du_1)\cdots z(du_n), \qquad f\in C([0,1]^n),$$

and

$$\mu=\bigotimes_{\xi\in\Omega_\infty}\mu_\xi \quad \text{with } \mu_\xi=\Gamma^0_{\theta^{(k)}_\xi}.$$

(ii) *Construction of the random field*. Next we have to specify the law of the random field $(\theta^{(k)}_\xi)_{\xi\in\Omega_\infty}$. The components of this field will be dependent reflecting the interaction of the original system $X^N(t)$ before we take the limit

$N \to \infty$. In particular the range of dependence will grow with the index k since in larger time scales interactions over longer stretches have been built up in the original system. We obtain a sequence of *inhomogeneous Markov random fields on a tree* giving the distribution of the field. It is only the initial state at the root of the tree which will depend on s.

Here is the formal construction: Introduce

(0.30) $$\widetilde{\Omega}_\infty^{(k)} = \{\xi \in \Omega_\infty | \xi_1 = \cdots = \xi_k = 0\}.$$

Then we can write

(0.31) $$\Omega_\infty = \bigotimes_{\widetilde{\Omega}_\infty^{(k)}} \mathbb{N}^k.$$

We shall see that the dependence in time scale β_k in the limit $N \to \infty$ will extend only over hierarchical distance k. Define therefore

(0.32) $$\mathscr{L}((\theta_\xi^{(k)})_{\xi \in \Omega_\infty}) = \prod_{\widetilde{\Omega}_\infty^{(k)}} \mathscr{L}((\theta_\xi)_{\xi \in \mathbb{N}^k}).$$

In order to specify $\mathscr{L}((\theta_\xi)_{\xi \in \mathbb{N}^k})$ we use the following construction: Let $\hat{\theta}$ (initial element) be an element of $\mathscr{P}([0, 1])$ to be specified later. Then

$(\tilde{\theta}^i)_{i \in \mathbb{N}}$ are i.i.d. $\Gamma_{\hat{\theta}}^{k-1}$ distributed,

$(\tilde{\theta}^{i,j})_{i \in \mathbb{N}, j \in \mathbb{N}} : \mathscr{L}(\tilde{\theta}^{i,j} | \{\tilde{\theta}^{m,n} \text{ for } (m, n) \neq (i, j)\}; \tilde{\theta}^m, \text{ for } m \in \mathbb{N}) = \Gamma_{\tilde{\theta}^i}^{k-2}(\cdot).$

In particular, conditioned on $\tilde{\theta}^i$, $(\tilde{\theta}^{i,j})_{j \in \mathbb{N}}$ are i.i.d. Similarly we define random variables $\tilde{\theta}^{i_1, \dots, i_\ell}$, $i_j \in \mathbb{N}$ for $j \in \{1, \dots, \ell\}$, using the kernel defined by $\Gamma_\theta^{k-\ell}$. Next we define our random field. For $\xi = (\xi_1, \dots, \xi_k) \in \mathbb{N}^k$ simply set

(0.33) $$\theta_\xi^{(k)} = \tilde{\theta}^{\xi_k, \dots, \xi_1} \quad \text{(note reversed order, see (0.1), (0.2))}.$$

So far we have not specified the initial state $\hat{\theta}$ of the Markov random field on the tree $\mathbb{N}^k$. Since we are interested in the time scale $\beta_k(N)$ where $x_{\xi,k}(s\beta_k(N))$ has as limiting dynamics $(\widetilde{Z}_s^{k,\theta})$ according to Theorem 0.2, the suitable thing is to randomise $\hat{\theta}$ according to the law of $\widetilde{Z}_s^{k,\theta}$ (independent of the rest); i.e.

(0.34) $$\mathscr{L}(\hat{\theta}) = \mathscr{L}(\widetilde{Z}_s^{k,\theta}).$$

The construction means that the dependence structure building up over times of order N^k is captured via a Markov field on the tree induced by $\mathbb{N}^k$ (each node splits into $\mathbb{N}$ nodes on the next level) and with the transition kernel K_j at a node of level j (counted from the bottom to the top) given by

$$K_j(\theta, d\theta') = \Gamma_\theta^{j-1}(d\theta').$$

(iii) *The initial state of the evolution.* Given the random field the evolution in each component starts in its unique equilibrium state. Hence integrating over the distribution of the random field we get a unique measure $\nu_\theta^{(j)}(s)$ on $(\mathscr{P}([0, 1]))^{\Omega_\infty}$ (recall (0.34)), which is an equilibrium state of the evolution $(X_{j,s}^\infty(t))_{t \geq 0}$. Hence for each j and $s > 0$ we have constructed a stationary process. The limit as $s \to \infty$ exists as well. It is denoted by $(X_j^\infty(t))_{t \geq 0}$ and is obtained by setting $\mathscr{L}(\hat{\theta}) = \mathscr{L}(Z_\infty^{j,\theta})$ in (0.34).

Having now completed the construction of the limiting dynamics we can state our results as follows:

Theorem 0.4 (Quasi-equilibria). *Let $(X^{\infty}_{j,s}(t))_{t\geq 0}$ be the stationary process with values in $(\mathscr{P}([0,1]))^{\Omega_\infty}$ which was defined above. Then for every $s>0$, $j\in\mathbb{N}$ and every $s(N)\uparrow\infty$ with $s(N)=o(N)$*

$$\begin{aligned}&\mathscr{L}((X^N(s\beta_j(N)+t))_{t\geq 0}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((X^{\infty}_{j,s}(t))_{t\geq 0}),\\ &\mathscr{L}((X^N(s(N)\beta_j(N)+t))_{t\geq 0}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((X^{\infty}_{j}(t))_{t\geq 0}). \quad \square\end{aligned} \tag{0.35}$$

Hence we obtain a sequence of quasi-equilibria $\nu^{(j)}_\theta$ by setting

$$\nu^{(j)}_\theta = \mathscr{L}(X^{\infty}_j(t)), \tag{0.36}$$

where j indicates that deviations from stationarity occur only in times of order $\beta_j(N)=N^j$. Recall that θ is the mean measure of the initial state of our original system.

(f) **The interaction chain and the qualitative behavior of the process.** The goal of this section is to introduce the interaction chain and use it to study the *space-time* picture of the process $(X^N(t))_{t\geq 0}$. The interaction chain arises by viewing at the sequence of original systems $(X^N)_{N\in\mathbb{N}}$ renormalized in space and time, in slow and fast time scales. After introducing this $\mathscr{P}([0,1])$-valued Markov chain we study it first from the point of view of stability versus clustering (cf. Theorem 0.5). Then we consider the two regimes in more detail and investigate the pattern of cluster formation and spatial density of types respectively (cf. Theorems 0.6, 0.7). A precise formulation of the relation between these results and the corresponding behavior of the original systems as N becomes large is deferred to section (h) (Theorem 0.10).

Studying the space-time structure of our interacting system amounts mathematically to considering the following vector-valued field:

$$\begin{aligned}&\hat{x}^N_{\xi,j}(s_0(N),s_1,\dots,s_j)\\ &\quad=(x_{\xi,j}(s_0(N)\beta_j(N)),\ x_{\xi,j-1}(s_0(N)\beta_j(N)+s_1\beta_{j-1}(N)),\\ &\qquad\qquad \dots,\ x_{\xi,0}(s_0(N)\beta_j(N)+\dots+s_j)),\end{aligned}$$

with $s_0(N)\uparrow\infty$ but $s_0(N)/N\to 0$, $(s_1,\dots,s_j)\geq 0$.

We study stability and clustering of the above object by investigating properties of the limiting object of $\hat{x}^N_{\xi,j}(s_0,\dots,s_j)$ as $N\to\infty$, which defines what we shall call the interaction chain at level j. Compare subsection (h) for a discussion of the relevance of this approximation for the fixed N system.

(i) We start by introducing and explaining the concept of the *interaction chain.* In the previous section we saw that there is associated with each time scale $\beta_k(N)$, and each $\theta\in\mathscr{P}([0,1])$, a quasi-equilibrium $\nu^{(k)}_\theta$ of our process. For a given time-scale $\beta_j(N)$ it makes sense to investigate the behavior of the process on the level of a single component $x_{\xi,0}(t)$ or block averages $x_{\xi,k}(t)$ for $k\leq j$. (Recall Theorem 0.2.) In the limit $N\to\infty$ the behavior of each of these quantities is described using the collection $\{\nu^{(k)}_\theta,\ k\leq j\}$. The dependence structure for fixed ξ of the vector $\hat{x}_{\xi,j}\in(\mathscr{P}([0,1]))^{j+1}$ is therefore described by the following object for the limiting dynamics as $N\to\infty$. Let $\xi=(\xi_1,\dots,\xi_j,\xi_{j+1},\dots)$. Consider the random vector defined as follows (recall (0.33)):

$$(\theta,\hat{\theta},\tilde{\theta}^{\xi_j},\tilde{\theta}^{\xi_j,\xi_{j-1}},\dots,\tilde{\theta}^{\xi_j,\dots,\xi_1}), \tag{0.37}$$

with θ given by the expectation measure of the initial distribution μ of the process and $\mathscr{L}(\hat{\theta})$ given by (0.34), see Theorem 0.1. Relation (0.37) defines for every j a time-dependent Markov chain on $\mathscr{P}([0,1])$ starting at time -1 in the state θ with transition kernel at time $\ell-1$ given by $K_\ell^{(j)}(\cdot,\cdot)$ with

$$(0.38)\qquad K_\ell^{(j)}(\theta', A) = \Gamma_{\theta'}^{j-\ell}(A), \qquad A \in \mathscr{B}(\mathscr{P}([0,1])),\ \theta' \in \mathscr{P}([0,1]).$$

Here everything depends on the chosen level j and time scale $\beta_j(N)$.

Definition. (a) The Markov chain $(\theta_k^j)_{k=-1,\dots,j}$ on $\mathscr{P}([0,1])$ defined in (0.37) or (0.38) is called the *interaction chain at level* j.

(b) This chain is embedded in the Markov random field over Ω_∞ with values in $\mathscr{P}([0,1])$, defined by (compare (0.33) for $s\to\infty$)

$$(0.39)\qquad (\theta_\xi^{(j)})_{\xi\in\Omega_\infty}$$

and this field is called the *interaction field at level* j. □

Remark. Recall that according to Theorem 0.3 the state $\theta_k^{(j)}$ of the level j interaction chain is of the form $\sum_{\ell=0}^\infty M_{k,\ell}^{(j)}\delta_{U_\ell}$, that is, consists of at most countably many atoms for every j, k.

Since the components are our natural reference point and from the point of view of entrance laws, it is now more convenient to consider the chain shifted in time, that is, to introduce for some fixed (but arbitrary) $\xi\in\Omega$ (see (0.33))

$$(0.40)\qquad (\hat{\theta}_\ell^{(j)})_{\ell=-j-1,\dots,0} \text{ with } \hat{\theta}_\ell^{(j)} = \hat{\theta}^{\xi_j,\dots,\xi_{|\ell|}},\ \hat{\theta}_{-j}^{(j)} = \hat{\theta},\ \hat{\theta}_{-j-1}^{(j)} = \theta.$$

The marginal distributions of $\hat{\theta}_\ell^{(j)}$ are $\mu_\theta^{j,|\ell|}$ introduced in subsection 0(d),(iv).

The rest of this section, namely subsections (ii)–(iv), gives an analysis of the qualitative properties of the interaction chain. First we discuss the stability versus clustering dichotomy. Then we discuss properties corresponding to the cluster formation in the original system. Finally we investigate the number of different types found in large blocks of the interacting system, which is reflected in path properties of the interaction chain.

(ii) In order to state below our theorem on the stability-clustering dichotomy we introduce the following terminology.

Definition. Let $M=\{\delta_u | u\in[0,1]\}$.

We say that the system of interacting Fleming-Viot processes *clusters in the mean-field limit* if, for all $\theta\in\mathscr{P}([0,1])$,

$$(0.41)\qquad \mathscr{L}(\hat{\theta}_0^{(j)}) = \mu_\theta^{j,0} \underset{j\to\infty}{\Rightarrow} \mathscr{L}(\delta_U(\cdot)) \quad \text{with } \mathscr{L}(U)=\theta$$

and is called *stable in the mean-field limit* if, for every $\theta\in\mathscr{P}([0,1])$,

$$(0.42)\qquad \mathscr{L}(\hat{\theta}_0^{(j)}) = \mu_\theta^{j,0} \underset{j\to\infty}{\Rightarrow} \mu_\theta^{\infty,0} \quad \text{with } \mu_\theta^{\infty,0}(M)=0,\ \text{for } \theta\notin M. \quad □$$

Remark. In a system which eventually clusters, locally we shall see a state which is a δ-measure; i.e. only one type prevails locally. In a stable system with probability 1 we see more than one type coexisting locally, provided we do not start in a monotype state.

Remark. The above picture can be extended and viewed from the point of view of *entrance laws* for the chain given in (0.40). Recall that an entrance law for a Markov chain indexed by $\mathbb{Z}$ and with transition kernel P_k is a collection $(\alpha_k)_{k\in\mathbb{Z}}$ of measures on the state space such that $\alpha_{k+1} = \alpha_k P_k$ for all $k \in \mathbb{Z}$. In particular, we shall show later on (see section 6) that

> $\mathscr{L}((\hat{\theta}_k^{(j)})_{k\in\{-j-1,\dots,0\}}) \Rightarrow \mathscr{L}((\hat{\theta}_k^{\infty})_{k\in\mathbb{Z}^-})$, where $\mathscr{L}(\hat{\theta}_0^{\infty}) = \mu_\theta^{\infty,0}$ and $\hat{\theta}_k^{\infty} \underset{k\to-\infty}{\Rightarrow} \hat{\theta}_{-\infty}^{\infty}$ and $(\hat{\theta}_k^{\infty})_{k\in\mathbb{Z}^-}$ is a Markov chain with kernels $\Gamma_\theta^{-k}(dy)$.

Hence each $\mathscr{L}((\hat{\theta}_k^{\infty})_{k\in\mathbb{Z}^-})$ constitutes an extremal entrance law. Moreover each entrance law for the collection of kernels $\{\Gamma_\infty^{-k}(du)\}_{k\in\mathbb{Z}^-}$ (see section 0(d),(iii)) can be obtained as mixtures over the extremal entrance laws corresponding to the ones where the random variable $\hat{\theta}_{-\infty}^{\infty}$ is equal to a constant $\theta \in \mathscr{P}([0,1])$, that is, the ones constructed above. Then the case of clustering corresponds to the situation where $\hat{\theta}_{-\infty}^{\infty} = \delta_u$ for some u, and the only extremal entrance laws are the sequences $((\delta_u)_{k\in\mathbb{Z}^-})$ for some $u \in [0,1]$, while in the stable case every $\theta \in \mathscr{P}([0,1])$ is possible as a value of $\hat{\theta}_{-\infty}^{\infty}$ and the entrance laws are not constant sequences if $\theta \notin M$. We summarize:

clustering $\Leftrightarrow$ $\{(\delta_u)^{\mathbb{Z}^-} \mid u \in \mathscr{P}([0,1])\}$ are the extremal entrance laws of the Markov chain in (0.40),

stability $\Leftrightarrow$ for every $\theta \in \mathscr{P}([0,1])$ there exists an extremal entrance law which for $\theta \notin M$ is concentrated on $(\mathscr{P}([0,1])\backslash M)^{\mathbb{Z}^-}$ and has nonconstant paths.

Since the $(c_k)_{k\in\mathbb{N}}$ describe the range of the interaction over distance k, we expect that for weak interaction (c_k at most slowly growing), we shall see clusters, while for strong interaction of longer range (c_k rapidly growing), the initial density of types should be preserved. In other words, in the first case we should have only trivial entrance laws based on $\hat{\theta}_{-\infty}^{\infty} = \delta_u$, $u \in [0,1]$, while in the second one to every $\theta \in \mathscr{P}([0,1])$ there should correspond an extremal entrance law.

Recall that θ is the density measure of the initial law of our process and hence the initial state of the shifted interaction chain of (0.40).

Theorem 0.5 (Clustering versus stability).

$$\sum_k c_k^{-1} = \infty \Rightarrow \textit{ the system clusters in the mean-field limit and}$$
$$\mathscr{L}(\hat{\theta}_{-\infty}^{\infty}) = \int \delta_{\delta_u}\theta(du); \tag{0.43}$$

$$\sum_k c_k^{-1} < \infty \Rightarrow \textit{ the system is stable in the mean-field limit and}$$
$$\hat{\theta}_{-\infty}^{\infty} = \theta. \quad \square \tag{0.44}$$

Compare the above result with Corollary 0.2 which represents the analogous result for fixed N.

In the next two subsections (iii) and (iv) we discuss each of the two cases in Theorem 0.5 in more detail.

(iii) The next part of this subsection is focussed on a more detailed analysis of the *clustering case.* In order to motivate the definitions below let us return to the interacting system $X^N(t)$ for a moment. In the case of a clustering system one would like to identify the rate at which clusters, i.e. sets of components in a state close to δ_u for one and the same $u \in [0, 1]$, grow. (Note that due to migration, we do not hit the set $\{\delta_u | u \in [0, 1]\}$ in finite time if we do not start in this set.) If we have at time t a cluster of components close to state δ_u which contains the block of components within distance k of the point of observation, say ξ, then the state of $x_{\xi, k}(t)$ will also be close to δ_u (and vice versa). Hence we can get an idea about the size of clusters by studying $(x_{\xi, k}(t))_{k \in \mathbb{N}}$ in various time scales and by relating the block size k and the time scale. A simplified approach is to fix a time scale and then look at averages of various sized blocks. For the time scale $\beta_j(N)$ we get good information on possible cluster sizes by considering the vector of the successive averages

$$(x_{\xi, k}(t\beta_j(N)))_{k=0, \dots, j}.$$

Since the clusters grow as time tends to infinity, we should determine that size of a block in which we see a cluster close to a δ-measure. In particular we must find the relation between blocksize k at which such a cluster appears and the index j of the time scale. In other words investigate the behavior of

$$x_{\xi, f(j)}(t\beta_j(N)) \quad \text{for } j \to \infty$$

and try to determine, for suitable increasing function f, the probability (in terms of f) that the limiting object remains concentrated on states which are δ-measures in $[0, 1]$. Appropriate f's would be $f(j) = [\sigma j]$ for $\sigma \in (0, 1)$.

Instead of approaching this (quite difficult) problem directly, we pass immediately to the corresponding question for the limiting object as $N \to \infty$, which is described completely by the interaction chain. This approach has the advantage of not using duality and allows us later on to generalize the results to a larger class of processes (compare Theorem 0.9). A discussion of the two-type case is found in Fleischmann and Greven (1994).

In the case of a system clustering in the mean-field limit the interaction chains $(\hat{\theta}_\ell^{(j)})$ become degenerate for $\ell = -M, -M+1, \dots, 0$ for every finite M as $j \to \infty$. Degenerate means the state becomes a δ-measure. We shall then ask whether the *rescaled* process $(\hat{\theta}^{(j)}_{f_\alpha^{(j)}})_{\alpha \in [0,1]}$ can be made, via choice of $f_\alpha(\cdot)$, to converge to a nontrivial process $(\hat{\theta}_\alpha)_{\alpha \in [0,1]}$ on $\mathscr{P}([0, 1])$. (Here nontrivial means that $\hat{\theta}_\alpha$ is not constant and equal to a δ-measure.)

The cluster formation exhibits two principal regimes, which correspond to the distinction between dimensions 1 and 2 in lattice models. These regimes are distinguished by the property that $(\hat{\theta}^{(k)}_{-k+m})_{m \in \mathbb{N}}$ converges to the constant process (regime I), respectively to a random sequence (regime II). The distinction depends on how rapidly c_k decreases; for $c_k = c^{-k}$ with $c > 1$ we are in regime I while for c_k decaying to 0 at most polynomially we are in regime II. (Compare Dawson and Greven (1993a).) Indeed we shall see in the second regime that it is possible to find a family of scaling functions for the blocksize as a function of the order of the time scale:

Definition. (a) A set of functions $\{f_\alpha(\cdot)\}_{\alpha\in[0,1]}$, $f_\alpha(\cdot): Z^+ \to Z^-$, $f_0(k) = -k$, are called the *cluster scales* of the process if they satisfy two conditions:

$$f_\alpha(k) \text{ is an increasing function of } \alpha \text{ for every } k,$$
$$\mathscr{L}((\hat{\theta}^{(k)}_{f_\alpha(k)})_{\alpha\in[0,1]}) \underset{k\to\infty}{\Rightarrow} \mathscr{L}((\hat{\theta}_\alpha)_{a\in[0,1]}). \tag{0.45}$$

(b) The process $(\hat{\theta}_a)_{\alpha\in[0,1]}$ is called the *cluster process* if for $\theta \notin M$

$$\operatorname{Prob}(\hat{\theta}_a = \delta_u \text{ for some } u \in [0,1]) \in (0,1) \quad \forall \alpha \in (0,1). \tag{0.46}$$

(c) We classify the growth of clusters as follows:

$$|f_\alpha(k)|/k \underset{k\to\infty}{\to} 1 \textit{ rapidly growing clusters.} \tag{0.47}$$

$$|f_\alpha(k)|/k \underset{k\to\infty}{\to} 1-\alpha \textit{ diffusive clustering.} \tag{0.48}$$

$$f_\alpha(k)/k \underset{k\to\infty}{\to} 0 \textit{ slowly growing clusters.} \quad \square \tag{0.49}$$

We shall later identify as cluster processes those processes which are (deterministic) time changes of $(Y_t)_{t\geq 0}$ where $(Y_t)_{t\geq 0}$ is a Fleming-Viot process on $\mathscr{P}([0,1])$ starting at θ. Formally, (Y_t) is the solution to the martingale problem (L, δ_θ), where

$$L(G)(z) = \frac{1}{2}\int_{[0,1]}\int_{[0,1]} \frac{\partial^2 G(z)}{\partial z \partial z}(u, v) Q_z(du, dv)$$

with

$$Q_z(du, dv) = z(du)\delta_u(dv) - z(du)z(dv)$$

and $G(z)$ as in (0.28).

Recall that the path of the process Y_t is such that for all $t > 0$ the measure $Y_t \in \mathscr{P}([0,1])$ is atomic and consists of a finite (random) number of atoms located at points in $[0,1]$ (cf. Dawson (1993, Section 8.2)). Furthermore as $t\to\infty$ the path Y_t will hit the set of measures $\{\delta_u, u \in [0,1]\}$ and remain trapped in the state first hit in this set. The location, u, of the traps is distributed according to θ.

Theorem 0.6 (Cluster formation). *Assume that* $\sum_k c_k^{-1} = \infty$ *and* $\liminf_{k\to\infty} c_k^{1/k} = 1$.

(a) *Diffusive clustering. Assume that* $c_k \to c$ *as* $k \to \infty$ *with* $c \in (0, \infty)$. *Then*

$$f_\alpha(k) = -(1-\alpha)k \tag{0.50}$$

are cluster scales and the cluster process is given by

$$(Y_{\log(1/(1-\alpha))})_{\alpha\in[0,1]}. \tag{0.51}$$

(b) *Rapidly growing and slowly growing clusters. Define* $I(m) := \sum_{k=0}^m c_k^{-1}$. *We have rapidly growing clusters if, for all* $\alpha \in (0,1)$, $I(m)/I(\alpha m) \underset{m\to\infty}{\to} 1$. *We have slowly growing clusters if, for all* $\alpha \in (0,1)$, $I(m)/I(\alpha m) \underset{m\to\infty}{\to} \infty$.

(c) *Next assume that* $c_k = c^{-k}$ *with* $c > 1$. *Then*

$$\mathscr{L}((\hat{\theta}^{(k)}_{-k+m})_{m\in\mathbb{N}}) \underset{k\to\infty}{\Rightarrow} \mathscr{L}((\hat{\theta}^{(\infty)}_m)_{m\in\mathbb{N}})$$

where $(\hat{\theta}_m^{(\infty)})_{m\in\mathbb{N}}$ *is a time homogeneous Markov chain on* $\mathscr{P}([0,1])$ *with transition kernel* $\Gamma_\theta^{1/c}$. □

For a more detailed analysis of the fast and slowly growing clusters compare Dawson and Greven (1993c). These three cases of cluster formation occur also in lattice models, i.e. $I = \mathbb{Z}^d$. Under the assumptions of Corollary 0.1, $d = 1$ corresponds to regime I and $d = 2$ to regime II and the diffusive clustering case.

(iv) We conclude this subsection (f) by analysing the *stable case* further. Here we have to distinguish two qualitatively different situations. First, θ is atomic and second, θ is *nonatomic.* Questions for the first case are essentially questions about diffusions on finite-dimensional simplices (compare (0.56)). It is the second case which displays a number of phenomena specific to the *infinite-dimensional* state space of a component. We shall focus on this second case in this subsection.

Again to motivate our definitions we return to $X^N(t)$. The two basic questions we would like to ask about our original system $X^N(t)$ are the following. Take a realization of our process $X^N(t)$ in equilibrium. Recall that $x_\xi^N(t)$ for $t > 0$ is an atomic measure. Pick at a fixed time t a certain type which has positive mass in the ξ-component $x_\xi^N(t)$, say. What is the *spatial density* of that *particular type* in the realization? Furthermore how fast does the *number of types* grow as a function of increasing block sizes, when we consider only the number of the largest types which account for a nontrivial proportion of the population, in that realization? These questions are difficult to analyse for fixed N and we now investigate this in the limit $N \to \infty$.

We define first the process $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$ corresponding to an entrance law of the interaction chain (the interaction chain at level ∞):

$$\mathscr{L}((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}) = \operatorname*{w-lim}_{j\to\infty} \mathscr{L}((\hat{\theta}_k^{(j)})_{k=-j-1,\dots,0}). \tag{0.52}$$

This limit exists as we shall prove in section 6(c) and is nontrivial by Theorem 0.5. Since every random variable $\hat{\theta}_k^\infty$ takes values in the space of measures with countably many atoms, we can write it in the *size-ordered representation*

$$\begin{aligned} &\hat{\theta}_k^\infty = \sum_{j=1}^\infty M_j^{(k)} \delta_{U_j^{(k)}} \quad \text{with } (M_j^{(k)})_{j\in\mathbb{N}} \text{ nonincreasing (in } j) \text{ and} \\ &(U_j^{(k)})_{j\in\mathbb{N}} = (U_{\pi_k^{(j)}})_{j\in\mathbb{N}} \\ &\qquad \text{(here } (U_i)_{i\in\mathbb{N}} \text{ are i.i.d. and } \pi_k : \mathbb{N}\to\mathbb{N} \text{ is a random map).} \end{aligned} \tag{0.53}$$

In this setup the processes $(M_j^{(k)})_{j\in\mathbb{N}}$ with values in $[0,1]$ are now the basic objects for the further analysis.

Recall now that we are in the case in which θ is *nonatomic.* We shall show that the *asymptotic spatial density of a type* is 0 and that the *number of types per block* increases at a certain rate depending on the interaction strengths c_k. This behavior can be reformulated in terms of the following objects which are stochastic processes obtained by taking functionals of the interaction chain

$(\hat{\theta}_k^{\infty})_{k\in\mathbb{Z}^-}$ at level ∞, that is, the entrance law.

$$R^{(k)} = \sum_{j=0}^{\infty}(M_j^{(k)})^2, \qquad k \in \mathbb{Z}^-, \tag{0.54}$$

$$S_\ell^{(k)} = \sum_{j=0}^{\ell} M_j^{(k)}, \qquad \ell \in \mathbb{N}, k \in \mathbb{Z}^-. \tag{0.55}$$

Theorem 0.7 (Spatial density of types). *Assume that θ is nonatomic. Let (b_k) be an $\mathbb{N}$-valued sequence with $b_k \uparrow \infty$ as $k \to \infty$.*

(a) $R^{(k)} \underset{k\to-\infty}{\longrightarrow} 0$ *a.s.*

(b) $S_{b_{|k|}}^{(k)} \underset{k\to-\infty}{\longrightarrow} 1$ *in probability provided that* $b_{|k|}/c_{|k|} \to \infty$ *as* $k \to -\infty$,

$S_{b_{|k|}}^{(k)} \underset{k\to-\infty}{\longrightarrow} 0$ *in probability provided that* $b_{|k|} \sum_{\ell=|k|}^{\infty} (c_\ell)^{-1} \to 0$ *as* $k \to -\infty$. □

Remark. (a) implies that $\sup_j M_j^{(k)}$ converges to 0 in probability as $k \to -\infty$; that is, we have spatial density 0 of particular types.

(b) gives information about the number of types per block in the limit of large blocks. In particular in the case in which $c_k = c^k$ for $c > 1$ we obtain sharp estimates for the growth of b_k as $k \to \infty$.

Remark. Note that for θ atomic, the assertion of the theorem fails! In that case every type has asymptotic spatial density $\theta(\{u\})$.

(g) **A fixed point property of Fleming-Viot systems under renormalization.** One expects that the features displayed in Theorems 0.1–0.7 are shared by a large class of evolutions. In this section we investigate the reason behind such *universality properties* (see Dawson and Greven (1993c) for a discussion of interacting diffusions). We now view the rescaling of time and space as a *renormalization procedure*, which gives a new interacting system which is studied in the limit $N \to \infty$. The key to universality is now the following fact.

The Fleming-Viot systems have the remarkable property that if we perform the multiple time scale analysis of Theorem 0.2, we obtain as limiting processes Fleming-Viot processes with immigration on *all* levels and in all times scales β_k. This means that the coefficient of the second-order term in the generator preserves its form. If we specialise to the case $\theta = \alpha\delta_0 + (1-\alpha)\delta_1$, we obtain a system of interacting Fisher-Wright diffusions, which can alternatively be described by the weights $(\alpha_\xi(t))_{\xi\in\Omega^N}$ of the atom at type corresponding to the point 1. Then

$$d\alpha_\xi(t) = \sum_{\xi'} a(\xi, \xi')(\alpha_{\xi'}(t) - \alpha_\xi(t))\,dt + \sqrt{2d\alpha_\xi(t)(1-\alpha_\xi(t))}\,dw_\xi(t) \tag{0.56}$$

with $(w_\xi(t))_{\xi\in\Omega_N}$ i.i.d. standard Wiener processes.

It could then be shown (cf. Dawson and Greven (1993c)) that among all systems of interacting diffusions in which $\alpha(1-\alpha)$ in the diffusive term is replaced by a nonnegative Lipschitz function g on $[0, 1]$ vanishing exactly at the boundary, the Fisher-Wright system is the only system where on the various levels of the multiple scale analysis we obtain diffusions with diffusion coefficients of the same form.

In order to formulate an analogous result in our context consider processes defined as in (0.8) but in the diffusion term the object

$$Q_{x_\xi}(du, dv) = x_\xi(du)\delta_u(dv) - x_\xi(du)x_\xi(dv)$$

is replaced by

$$Q^G_{x_\xi}(du, dv) = \left(\int G(x_\xi, u, z)x_\xi(dz)\right) x_\xi(du)\delta_u(dv) - G(x_\xi, u, v)x_\xi(du)x_\xi(dv). \tag{0.57}$$

Here $G \in \mathscr{G}$, where $\mathscr{G}$ is the class of nonnegative measurable functions on $\mathscr{P}([0, 1]) \times [0, 1]^2$ such that for each $\mu \in \mathscr{P}([0, 1])$, $G(\mu, \cdot, \cdot)$ is symmetric and

$$\|G\|_\infty := \sup_\mu \iint G(\mu, u, v)\mu(du)\mu(dv) < \infty.$$

Let $L^{G,c}_\theta$ denote the operator of the form (0.28) with immigration source θ at rate c and diffusion operator Q^G_z. The form of Q^G_x guarantees that the state space of the measure-valued process is $\mathscr{P}([0, 1])$. Such systems arise as diffusion limits of genetic models in which the reproduction of a colony produces a covariance of the frequency of different types which may depend on the state and the type involved and not only from the constraint that the total mass is one.

Assume for the moment that the martingale problem for $(L^{G,c}_\theta, \mu)$ has a solution and that $\Gamma^{G,c}_\theta$ is an equilibrium state of the process on $\mathscr{P}([0, 1])$ (see 0(d), (iii) and Lemma 2.1 for a formal construction of processes of this type).

For $G \in \mathscr{G}$ define

(0.58)

$\mathscr{F}_c Q^G(\theta, u, v)$ to be the signed measure on $[0, 1]^2$ given by

$$\mathscr{F}_c Q^G(\theta, du, dv) := E^{\Gamma^{G,c}_\theta}(Q^G_x)(du, dv) = \int_{\mathscr{P}([0,1])} Q^G_x(du, dv)\Gamma^{G,c}_\theta(dx).$$

We write Γ^G_θ instead of $\Gamma^{G,c}_\theta$ and $\mathscr{F}$ instead of $\mathscr{F}_c$ if the parameter $c > 0$ is fixed.

We will prove in Lemma 7.1 in section 7 that

$$\int \Gamma^G_\theta(dx)G(x, u, v)x(du)x(dv) \ll \theta(du)\theta(dv). \tag{0.59}$$

Let $H(\theta, u, v)$ denote the Radon-Nikodym derivative of the l.h.s. with respect to the right-hand side. of (0.59). Then we also show in Lemma 7.1 that

(0.60)

$$\mathscr{F}_c Q^G(\theta, du, dv) = \int H(\theta, u, z)\theta(dz)\theta(du)\delta_u(dv) - H(\theta, u, v)\theta(du)\theta(dv);$$

that is,

$$\mathscr{F}_c Q^G(\theta, du, dv) = Q^H_\theta(du, dv)$$

and

$$\|H\|_\infty \le \|G\|_\infty. \tag{0.61}$$

We write $H = \mathscr{F}_c(G)$.

In order to be able to iterate the map $\mathscr{F}$, respectively $\mathscr{F}_c$, we need some more properties of the functions G we consider. Define for some fixed $c > 0$

$$\widetilde{\mathscr{G}} = \{G \in \mathscr{G} | G \text{ satisfies (i), (ii), (iii)}\}:$$

(i) The martingale problem $(L_\theta^{G,c}, z)$ has a unique solution for every $\theta \in \mathscr{P}([0,1])$, $z \in \mathscr{P}([0,1])$.

(ii) The system corresponding to $(L_\theta^{G,c}, z)$ is ergodic for all $\theta \in \mathscr{P}([0,1])$ and $c > 0$.

(iii) The map $\theta \to \Gamma_\theta^G$ is continuous.

Now consider the largest subset $\widehat{\mathscr{G}} \subseteq \widetilde{\mathscr{G}}$ such that

$$\mathscr{F}_c(\widehat{\mathscr{G}}) \subseteq \widehat{\mathscr{G}}.$$

The set $\widehat{\mathscr{G}}$ is nonempty since it contains at least the functions $c1$ with $c > 0$. In the case where the type space consists of two points instead of $[0,1]$ it has been proved that $\widehat{\mathscr{G}} = \mathscr{G}$ (Dawson and Greven (1993c) and Baillon et al. (1994)). It is an open problem to prove this in our more general context. Since all of the three points in the definition of $\widetilde{\mathscr{G}}$ above are already very challenging problems, we do not discuss this topic any further at this point.

Theorem 0.8 (Fixed-point property). *Let $c > 0$. Then the function $G(x_\xi, u, v) = \text{const}$ is the only function in $\widehat{\mathscr{G}}$ with the property that there exists a function $h : \mathbb{R}^+ \to \mathbb{R}^+$ such that for all $a \in \mathbb{R}^+$:*

$$Q_\theta^{\mathscr{F}_c(aG)} = h(a) Q_\theta^G. \tag{0.62}$$

Furthermore, for every $G \in \widehat{\mathscr{G}}$ and with $\mathscr{F}^n(G) := \mathscr{F} \circ \mathscr{F} \circ \cdots \circ \mathscr{F}(G)$,

$$\begin{aligned} &\text{(i) } (n/c) Q_\theta^{\mathscr{F}_c^n(G)} \underset{n\to\infty}{\longrightarrow} Q_\theta. \\ &\text{(ii) } \|(n/c)\mathscr{F}_c^n(G) - 1\|_\infty \underset{n\to\infty}{\longrightarrow} 0, \text{ provided } G(x, u, v) \geq \varepsilon > 0. \quad \square \end{aligned} \tag{0.63}$$

Note: The question whether the corresponding martingale problems still have a unique solution and whether Theorems 0.1, 0.2, 0.4, 0.5, 0.6 hold here as well is not yet completely settled (cf. Dawson and March (1993)). However the above statements make sense without having solved these problems.

With the same argument as in Dawson and Greven (1993c), Baillon et al. (1993) we can derive Theorems 0.5, 0.6 from (0.63). Hence relation (0.63) is the reason behind the universality in the longterm behavior of the system:

Theorem 0.9 (Universality). *Assume that $G \in \widehat{\mathscr{G}}$.*

(a) *For every solution of the martingale problem (L, μ) with Q in* (0.6) *replaced by Q^G, the conclusion of Theorem* 0.5 *holds.*

(b) *Assume that $Q^G \geq \varepsilon Q$ for some $\varepsilon > 0$. Then for every solution of the martingale problem (L, μ) with Q replaced by Q^G, Theorem* 0.6 *holds.* $\square$

(h) **Space-time structure of the process for large N and the interaction chain.** The role of this section is to formulate a precise answer to two questions:

(i) To what extent does the interaction chain $(\hat{\theta}_k^j)_{k=-j-1,\dots,0}$ and the realizations of the entrance laws $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$ give a good approximation of the *space-time structure* of the *process* $X^N(t)$ for fixed but large N and large times?

(ii) What do the quasi-equilibria $\nu_\theta^{(k)}$, i.e. $\mathscr{L}(\hat{\theta}_0^\infty)$, tell us about the spatial structure of the equilibrium measures ν_θ^N of the interacting system for N fixed?

We first discuss question (i). In order to understand the space-time structure of the equilibrium process, we look at the process at times of moderate size $\beta(N)$, say $\beta(N) \gg N^j$ but $\beta(N) \ll N^{j+1}$ and at very large times, i.e. $\beta(N) \gg N^j$ for every j. To capture the spatial structure consider a vector of successive averages over blocks of different sizes. To formulate this idea precisely consider $\hat{t} = (t_1, \dots, t_j)$, $t \geq 0$, $i = 1, \dots, j$, and $s(N)$ with $s(N) \uparrow \infty$ and $s(N) = o(N)$. Define chains

$$(0.64) \qquad (Z_k^{j,N}(s, \hat{t}))_{k=-j-1,\dots 0}, \qquad j \in \mathbb{N},$$

by setting

$$(0.65) \quad Z_k^{j,N}(s, \hat{t}) = x_{\xi,|k|}^N \left(s(N)N^j + \sum_{i=1}^{j-|k|} t_i N^{j-i} \right), \qquad k = -j-1, \dots, 0.$$

Hence the process $(Z_k^{j,N})_{k=-j-1,\dots,0}$ describes in which way the various block averages of the system $X^N(t)$ fluctuate when observed at time points $(t_1, \dots, t_j)$ of moderate size. Next consider the time scale $\sigma(N)$ with $\sigma(N) \uparrow \infty$ and $\sigma(N)/N^j \to \infty$ for every $j \in \mathbb{N}$. Define a process $(Z_k^{\infty,N})_{k\in\mathbb{Z}^-}$ by setting

$$(0.66) \qquad Z_k^{\infty,N} = x_{\xi,|k|}^N(\sigma(N)), \qquad k \in \mathbb{Z}^-.$$

This object describes the spatial dependence structure of the process in a large time scale.

Finally we come to question (ii). Namely, in order to analyse the structure of the equilibrium measure ν_θ^N we want to know what is the distribution of the vectors $(x_{\xi,0}, x_{\xi,1}, \dots)$ for fixed $\xi \in \Omega_N$. It is convenient to view this object in the form $(x_{\xi,|k|})_{k\in\mathbb{Z}^-}$. Denote the distribution of this random variable on $(\mathscr{P}([0,1]))^{\mathbb{Z}^-}$ by $\hat{\nu}_\theta^N$; i.e.

$$(0.67) \qquad \mathscr{L}((x_{\xi,|k|}^N)_{k\in\mathbb{Z}^-}) = \hat{\nu}_\theta^N.$$

Before stating the next result we recall some ideas from statistical physics. For this purpose let us briefly review some facts about the infinite systems and the hierarchical mean-field systems which we need to explain some concepts appearing in the answer to the above questions. First according to Theorem 0.1 and Corollary 0.2, under condition (0.23) for each $N \in \mathbb{N}$ and $\theta \in \mathscr{P}([0,1])$ the infinite system on Ω_N has a spatially homogeneous ergodic invariant random measure with marginal mean measure θ, ν_θ^N. On the other hand for each $J \in \mathbb{N}$ the (finite) system truncated at level J, i.e. $c_k \equiv 0$ for $k > J$, and with $x_{\xi,J+\ell}^N$, $\ell = 1, 2, \dots$, replaced by a fixed source θ has a unique invariant measure $\nu_\theta^{N,J}$. It can be shown that $\nu_\theta^{N,J} \Rightarrow \nu_\theta^N$ as $J \to \infty$ in the sense that the restrictions of $\nu_\theta^{N,J}((x_\xi : \xi \in B) \in \cdot) \Rightarrow \nu_\theta^N((x_\xi : \xi \in B) \in \cdot)$ for every finite subset $B \subset \Omega_N$. This is the analogue of the *thermodynamic limit* of statistical physics (cf. Ruelle (1969)). On the other hand, according to Theorem 0.5, under condition (0.44) $\mathscr{L}((\hat{\theta}_k^{(J)})_{k\in\{-J-1,\dots,0\}})$ converges to the *entrance law* $\mathscr{L}((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-})$ as $J \to \infty$. Moreover the conditions (0.23) and (0.44) coincide provided that $\limsup_{k\to\infty}(c_k)^{1/k} < N$. Thus the entrance law of the hierarchical mean-field limit is the analogue of the thermodynamic limit of the infinite system in the stable case. To complete the picture we will now formulate the result that the joint distribution of the block averages under $\hat{\nu}_\theta^N$ converge to $\mathscr{L}((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-})$ as $N\to\infty$. Thus the entrance law constructed above can be viewed as the hierarchical mean-field limit of the sequence of ergodic invariant measures of the corresponding infinite systems $\{x_{\xi,k}^N : \xi\in\Omega_N\}$ with mean measure θ (compare (0.71) below).

Theorem 0.10 (Approximation property of the interaction chain). *Consider the stable case. Then*

(a) *Assume that the initial measure is homogeneous with site mean measure θ and that for any $g\in C([0,1])$*

$$\mathrm{Cov}(\langle x_{\xi'}, g\rangle\langle x_{\xi''}, g\rangle)\to 0 \quad \textit{as } d(\xi',\xi'')\to\infty \textit{ uniformly in } N.$$

Then

$$\mathrm{Var}(\langle x_{\xi,|k|}(t), g\rangle)\to 0 \quad \textit{as } k\to\infty \textit{ uniformly in } t \textit{ and } N>N_0. \tag{0.68}$$

(b) *For every $\hat{t}\in(0,\infty)^j$:*

$$\mathscr{L}((Z_k^{j,N}(\hat{t}))_{k=-j-1,\dots,0}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((\hat{\theta}_k^j)_{k=-j-1,\dots,0}), \tag{0.69}$$

$$\mathscr{L}((Z_k^{\infty,N})_{k\in\mathbb{Z}^-}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}), \tag{0.70}$$

$$\hat{\nu}_\theta^N \underset{N\to\infty}{\Rightarrow} \mathscr{L}((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}). \quad \square \tag{0.71}$$

Remark. A particular case of interest is $t\equiv 0$ in which we get that the vector $(x_{\xi,|k|})_{k\in\mathbb{Z}^-}$ under the equilibrium law ν_θ^N is approximately distributed as $N\to\infty$ as the entrance law of the interaction chain associated with $\hat{\theta}_{-\infty}^\infty = \theta$. In particular the approximation property says that for large fixed N we cannot cluster in the case $\sum c_k^{-1}<\infty$. (However, the opposite conclusion, namely that $\sum c_k^{-1}=\infty$ implies clustering for large N, cannot immediately be drawn.)

Remark. The analogous result can be obtained by considering in (0.71) not just a fixed $\xi\in\Omega_N$ but the whole field in ξ.

Remark. The weak convergence statement (0.71) is not sufficiently strong to yield results on the spatial distribution of types from Theorem 0.7 and more generally it is not always a priori possible to interchange the limits $J\to\infty$ and $N\to\infty$. However it can be shown that many properties of the systems remain the same under the interchange of the limits; for example this is the case in Corollary 0.2 on the stability versus clustering dichotomy. Also the finer properties of the cluster formation are stable under letting $N\to\infty$ at least in the two-type case (see Dawson and Greven (1993), Fleischmann and Greven (1994)).

(i) **Open problems and extensions. Outline of paper.** The central open question is to obtain Theorems 0.0–0.10 for more general systems of the type which was introduced in (0.57) in connection with (0.8) and to obtain theorems on the interchange of the limits $N\to\infty$ and $J\to\infty$ with respect to Theorems 0.5–0.7. By analogy with the theory for the case of two types, i.e. $[0,1]$ is replaced by $\{0, 1\}$, which leads to interacting diffusion systems, we expect that the theorems we have generalize. (Compare Dawson and Greven, (1993b,c), Cox, Greven and Shiga (1993) and Fleischmann and Greven (1994). In these papers coupling techniques were essential which do not immediately work in the case of "general" Fleming-Viot systems.)

A second problem is to analyse the stable case further. For example the limiting behavior of the process $(\widetilde{M}_k^{(\infty)})_{k\in\mathbb{Z}^-}$, after rescaling, should be studied (this describes the density of a particular type over blocks of different sizes). In equilibrium critical branching systems in $\mathbb{R}^n$, $n = 4$ is the critical dimension for the transcience-recurrence dichotomy for the motion of each of the countably many surviving types (cf. Dawson (1994)). The analogue of this for the system of interacting Fleming-Viot processes should be found.

A third round of questions is to obtain Theorems 0.1, 0.5, 0.6, 0.7 without taking the limit $N \to \infty$. Again in the case of interacting diffusion systems partial results are known.

The rest of this paper is organised as follows. In sections 1 and 2 we study existence and uniqueness of the model and related questions. In section 3 we treat the behavior of the system as $t \to \infty$ using duality (which is introduced in this section). The sections 4 and 5 systematically develop the multiple space-time scale analysis. Section 6 treats the qualitative properties of the interaction chain and finally section 7 is devoted to the study of the map $\mathscr{F}$ and the question of universality.

1. Existence, uniqueness and path properties of the system of infinitely many interacting Fleming-Viot processes

The purpose of this section is to establish the existence of the model described in section 0 and in particular to give the proof of Theorem 0.0. As in the theory of infinite particle systems the proof of such a result consists of two steps. First one has to ensure that the systems with finitely many components are uniquely described by the given generator. Then, in a second step, one has to show that one can obtain the infinite system by taking a limit in the system size. The first step has been carried out in Vaillancourt (1990).

The proof of Theorem 0.0. We shall prove a slightly stronger result. Let G be a continuous symmetric function $\mathscr{P}([0, 1]) \times [0, 1]^2 \to [0, \infty)$. Define (recall (0.57) for Q_z^G)

$$L^G(F)(x) = \sum_{\xi,\xi'\in I} a(\xi, \xi') \int_{[0,1]} \frac{\partial F(x)}{\partial x_\xi}(u)(x_{\xi'}(du) - x_\xi(du)) + \sum_{\xi\in I} \int_{[0,1]} \int_{[0,1]} \frac{\partial^2 F(x)}{\partial x_\xi \partial x_\xi}(u, v) Q_{x_\xi}^G(du, dv). \tag{1.0}$$

Proposition 1.0. *For every initial distribution $\mu \in \mathscr{P}(E)$, the martingale problem for (L^G, μ) has a solution in $\mathscr{P}(C([0, \infty), E))$. If there exists a $d > 0$ with $G(\mu; u, v) \equiv d$, then the martingale problem for (L^d, μ) is well-posed.* □

Remark. To prove that the (L^G, μ) martingale problem is well-posed for non-constant G in general is difficult. However if $G(\theta, u, v)$ has a representation of the form

$$G(\theta, u, v) = d + \sum_{n=0}^{\infty} \int_0^1 \cdots \int_0^0 G_n(z_1, \dots, z_n, u, v)\theta(dz_1)\cdots\theta(dz_n)$$

and

$$\sum_{n=0}^{\infty} b^n \|G_n(z_1, \dots, z_n, u, v)\|_\infty < \infty \quad \text{for some } b > 0,$$

then it is proved in Dawson and March (1993, Theorem 5.8) that there exists a d^* such that if $d \geq d^*$, then the corresponding Fleming-Viot (single site) martingale problem is well-posed.

Proof. Define an increasing sequence of finite subsets of I, say $I_1 \subset I_2 \subset \cdots$ such that $\bigcup_{m\geq 1} I_m = I$; denote by L_m^G the mth truncated operator approximating L^G, (that is, L_m^G is defined as L is but using sums over I_m in place of I).

By Theorem 4.2 of Vaillancourt (1990), the martingale problem for (L^d, μ), $d \in (0, \infty)$, is well-posed in $\mathscr{P}(D([0, \infty), \mathscr{P}([0, 1])^{I_m}))$ for every distribution $\mu \in \mathscr{P}(\mathscr{P}([0, 1])^{\Omega_m})$. The existence of a continuous version can be obtained as in Dawson (1993, Proposition 7.3.1). Given this unique solution, let us denote by P_m $(= P_m^\mu)$ its unique extension to $\mathscr{P}(C([0, \infty), E))$, obtained by freezing the components $\notin I_m$. In the case where $G(\mu, u, v)$ is not constant we have to use techniques from Dawson and March (1993, Theorem 5.6) to construct a solution.

Let F be a function in $\mathscr{A}$ (see (0.5)). The sequence of P_m-semimartingales $\{F(X_t^m)\}$, where $\{X_t^m\}$ here denotes the canonical process in $\mathscr{P}(C([0, \infty), E))$ associated with probability measure P_m, has uniformly bounded local characteristics. The tightness of $\{P_m \circ F^{-1} : m \geq 1\} \subset \mathscr{P}(C([0, \infty), \mathbb{R}))$ for F in $\mathscr{A}$ follows from the criteria of Joffe and Métivier (1986, Proposition 3.2.3). We have to check the hypotheses made there on the drift and diffusion coefficients of the semimartingale $F(X_t^m)$, denoted by $b^m(t, \omega)$, and $a^m(t, \omega)$, respectively. The condition needed is

$$\sup_m E \sup_{t\leq T} (b^m(t, \cdot)^2 + a^m(t, \cdot)) < \infty. \tag{1.1}$$

Since $F^2 \in \mathscr{A}$ if $F \in \mathscr{A}$ it suffices to show that

$$\sup_m E \sup_{t\leq T} (b^m(t, \cdot))^2 < \infty. \tag{1.2}$$

This follows from the bound

$$\sup_m \sup_x L_m^G(F)(x) \leq C_{F,G} < \infty, \tag{1.3}$$

which is elementary.

Since $\mathscr{A} \subseteq C(\xi)$ contains a countable dense algebra, the tightness of $\{P_m : m \geq 1\}$ in $\mathscr{P}(D([0, \infty), E))$ follows from that of $\{P_m \circ F^{-1} : m \geq 1\}$.

For each $F \in \mathscr{A}$, and integer $M \geq 1$ there is a $\delta_M > 0$ such that

$$|L^G(F) - L_m^G(F)| \leq \delta_M C_F \quad \text{for every } m \geq M \text{ with } \delta_M \to 0 \text{ as } M \to \infty. \tag{1.4}$$

Therefore all weak limits of $\{P_m\}$ are solutions to the martingale problem for L^G. For $G \not\equiv d$, we pick any such solution. To complete the proof—and obtain that P_m converges weakly in the case $G \equiv d$, $d \in (0, \infty)$, as a bonus—we need only prove the uniqueness of solutions to the martingale problem for (L^d, μ) when $\mu \in \mathscr{P}(E)$. Since we are now talking about the case $G \equiv d$, we use the concept of duality. This will be discussed extensively in section 3. The reader is referred to Lemma 3.1 therein, which immediately gives uniqueness. Furthermore, the continuity of the limit follows from the fact that $P_m(C([0, \infty), E)) = 1$ for each m (cf. Ethier and Kurtz (1986, Chapter 3, Theorem 10.2)).

Proposition 1.1. *The unique solution of the martingale problem* (L^G, μ) *is for* $G \equiv d$, $d > 0$, *a Markov process which is Feller.* □

Proof. See Ethier and Kurtz (1986, Chapter 10, section 4.5).

Proposition 1.2. *The unique solution of the martingale problem* (L^G, μ) *has for* $G \equiv d$ *the property that each component of the process takes values in the set of atomic measures on* $[0, 1]$ *for all* $t > 0$, *a.s.*

Proof. It suffices to show for a fixed $\xi_0 \in I$, and $0 < T < \infty$, that $x_{\xi_0}(t)$ is pure atomic for $0 < t \leq T$, a.s. The idea is to reduce the question first to a statement for Fleming-Viot systems with finitely many interacting components and then in a second step to an ordinary Fleming-Viot process with mutation, for which results are known. We define a finite sequence of systems of interacting Fleming-Viot processes on a common probability space as follows. Let I_m be a finite subset of I which is given in the case $I = \mathbb{Z}^n$ by $[-m, m]$ and in the case Ω_N by $\{\xi | d(0, \xi) \leq m\}$. We next replace $[0, 1]$ in (0.8) by $[0, 1] \cup \{\Delta\}$ and for $m \in \mathbb{N}$ we set

$$
\begin{aligned}
L_m^G(F)(x) = &\sum_{\substack{\xi \in I_m \\ \xi' \in I}} a(\xi, \xi') \int_{[0,1]\cup\{\Delta\}} \frac{\partial F(x)}{\partial x_\xi}(u)(x_{\xi'}(du) - x_\xi(du)) \\
&+ \sum_{\xi \in I} \int_{[0,1]} \int_{[0,1]} \frac{\partial^2 F(x)}{\partial x_\xi \partial x_\xi}(u, v) Q_{x_\xi}^G(du, dv).
\end{aligned}
\tag{1.5}
$$

This means that the process associated with $L_m^G(F)$ has the property that if $x_\xi(0) = \delta_\Delta$ for $\xi \in I_m^c$, then $x_\xi(t) \equiv \delta_\Delta$ for all $t \geq 0$, $\xi \in I_m^c$. Moreover it is possible to construct a sequence of processes associated with $\{L_m^G(F), m \in \mathbb{N}\}$ on a common probability space such that for each $\xi \in I$, $x_\xi^m | [0, 1]$ is non-decreasing in m. For each m, $\{x_\xi^m, \delta_\Delta | \xi \in I_m\}$ is a finite system of interacting Fleming-Viot processes on with state space $(\mathscr{P}([0, 1] \cup \Delta))^{m+1}$ and with bounded migration rates. A simple modification of the proof of Theorem 5.2 in Donnelly and Kurtz (1991) yields that $\{x_\xi^m(\cdot), \delta_\Delta | \xi \in I_m\}$ has with probability one paths in $C([0, \infty), (\mathscr{P}_a([0, 1] \cup \Delta))^{m+1})$ where $\mathscr{P}_a$ denotes the purely atomic probability measures. (The proof is a direct application of their "look-down process" representation of the Fleming-Viot process and their construction can easily be extended to represent the system $\{x_\xi^m(\cdot), \delta_\Delta | \xi \in I_m\}$.) (An alternative proof using stochastic calculus can also be obtained by modifying the arguments in Ethier and Kurtz (1986, p. 441) or Dawson (1993, Theorem 8.2.1).) Moreover a first moment calculation shows that $x_\xi^m \uparrow x_\xi$ as

$m \to \infty$ where $\{x_\xi\}$ is a solution to the martingale problem for L^G. Therefore to complete the proof it suffices to note that the increasing limit of pure atomic measures is pure atomic.

2. Fleming-Viot process with immigration source

From now on, the migration mechanism is a random walk on the hierarchical group Ω_N; i.e. the stochastic matrix $a(\cdot, \cdot)$ is actually a random walk kernel on $I = \Omega_N$. To be specific we consider a mean-field random walk (see (0.19), (0.20)) given by a sequence $\{c_k : k \in \mathbb{N}\}$ of nonnegative parameters such that, for every choice of ξ and ζ in Ω_N,

$$a(\xi, \zeta) := \sum_{j \geq k} c_{j-1} N^{-2j+1} \quad \text{where } k = d(\xi, \zeta).$$

The probability of a jump to a randomly chosen element within distance k is of order N^{-k}. This means that the strength of interaction between elements depends on their hierarchical distance and this will give rise to multiple time scale phenomena. The quasi-stationary behavior in each time scale will be approximated (for large times, order N^k) by a measure-valued diffusion on $\mathscr{P}([0, 1])$, denoted by $(Z_t^{\theta, k})_{t \geq 0}$, which is a Fleming-Viot process with immigration from a source $\theta \in \mathscr{P}([0, 1])$. We next describe processes of this type.

(a) **The Fleming-Viot martingale problem.** Consider the following martingale problem. Let $G(\cdot, \cdot, \cdot)$ be a continuous symmetric function on $\mathscr{P}([0, 1]) \times [0, 1] \times [0, 1] \to [0, \infty)$, and $\theta(\cdot)$ a measurable function $[0, \infty) \to \mathscr{P}([0, 1])$. Consider the operator (recall (0.57) for Q_z^G)

$$\begin{aligned} L_t(F)_{(z)} &= c \int_0^1 \frac{\partial F}{\partial z}(u)(\theta(t, du) - z(du)) \\ &\quad + \int_0^1 \int_0^1 \frac{\partial^2 F}{\partial z \partial z}(u, v) Q_z^G(du, dv) \end{aligned} \tag{2.1}$$

acting on functions

$$F(z) = \prod_{i=1}^{k} \left(\int_0^1 f_i(u) z(du) \right), \qquad f_i \in C([0, 1]),\ i = 1, \ldots, k,\ k \in \mathbb{N}.$$

Lemma 2.1. *The martingale problem (L_t, z) has a solution. If $G(z, u, v) \equiv d$, then the solution is unique.* □

Proof. See for example [Dawson (1993, Theorem 10.3.2.1), or Dawson and March (1993)].

Corollary 2.1. *The martingale problem for (L_θ^k, μ) is well-posed in $\mathscr{P}(C([0, \infty) : \mathscr{P}([0, 1])))$ for every $\theta \in \mathscr{P}([0, 1])$ and every $\mu \in \mathscr{T}_\theta$, with*

$$\begin{aligned} L_\theta^k(F)(z) &:= c_k \int_{[0, 1]} \frac{\partial F}{\partial z}(u)(\theta(du) - z(du)) \\ &\quad + d_k \int_{[0, 1]} \int_{[0, 1]} \frac{\partial^2 F}{\partial z \partial z}(u, v) Q_z(du, dv), \end{aligned} \tag{2.2}$$

where

$$d_k = \frac{c_{k-1}d_{k-1}}{c_{k-1}+d_{k-1}} \quad \textit{for } k \geq 1,$$

$$d_0 = 1, \; F(z) = \prod_{i=1}^{k}\left(\int_0^1 f_i(u)z(du)\right), \quad f_i \in C([0,1]),$$

$i = 1, \ldots, k, \; k \in \mathbb{N}$. □

(b) **Representation of equilibrium states: Proof of Theorem 0.3.** We must show that, for each $\theta \in \mathscr{P}([0,1])$, the above martingale problem for L_θ^k possesses a unique stationary distribution, namely the law $\Gamma_\theta^k \in \mathscr{P}(\mathscr{P}([0,1]))$ of

$$\sum_{i=1}^{\infty}\left[V_i^k \prod_{j=1}^{i-1}(1-V_j^k)\right]\delta_{U_i}.$$

Ethier and Kurtz (1991, Theorem 3.8) proved that there is only one stationary distribution, namely the law of $\sum_{i\geq 1}\eta_i\delta_{U_i}$ where $\{U_i\}$ are i.i.d. (θ) and the vector $(\eta_1, \eta_2, \ldots)$ has the Poisson-Dirichlet distribution with parameter γ_k^{-1}. This law was originally constructed by Kingman (1975) as the weak limit of the reversed order statistics of a sequence of Dirichlet distributions. Patil and Taillie (1977) showed that rearranging the coordinates of a GEM in decreasing order yields a vector with Poisson-Dirichlet law (see Ethier (1992, Corollary, §3) and Ethier (1990, Theorem 3.2) for an elegant proof). Combining these yields our assertion.

We conclude this section by deriving a formula needed later on in the section on renormalization. Consider the solution of (2.1) with $G(z,u,v) \equiv d$, $\theta(t) \equiv \theta$.

Lemma 2.2. *Let* $Q_z(du,dv) = z(du)\delta_u(dv) - z(du)z(dv)$ *and let* Γ_θ *be the measure on the right-hand side of* (0.25) *with* $\gamma_k = \gamma$. *Then*

$$E_{\Gamma_\theta} Q_z(du,dv) = \frac{1}{\gamma+1} Q_\theta(du,dv). \quad □ \tag{2.3}$$

Proof. We begin by reviewing some properties of the ingredients defining Γ_θ. The Beta(a,b) distribution on $[0,1]$ has density given by

$$f(x) = c(a,b)x^{a-1}(1-x)^{b-1}, \quad a>0, b>0,$$
$$c(a,b) = \frac{(a+b-1)!}{(a-1)!(b-1)!}.$$

A Beta(a,b) distributed random variable X satisfies

$$E[X] = \frac{a}{a+b}, \quad \mathrm{Var}[X] = \frac{ab}{(a+b)^2(a+b+1)}.$$

The sequence of $[0,1]$-valued random variables, $\{M_j\}$, is said to have the GEM(β) joint distribution iff

$$M_j := V_j \prod_{i=1}^{j-1}(1-V_i) \quad \text{where } V_1, V_2, \ldots \text{i.i.d. Beta}(1,\beta).$$

Then using the above formulas for the beta distribution

$$E[M_j] = \frac{1}{1+\beta}\left(\frac{\beta}{1+\beta}\right)^{j-1}.$$

Note that specializing to the case of interest we get

$$\Gamma_\theta := \mathscr{L}\left(\sum_{j=1}^{\infty} M_j \delta_{U_j}\right) \quad \text{where } \{U_j\} \text{ are i.i.d. } \theta \text{ and } \{M_j\} \text{ are } \mathrm{GEM}(\gamma^{-1}).$$

We are now ready to begin with the proof of (2.3).

$$\begin{aligned}
\int Q_z(du, dv)\Gamma_\theta(dz) &= E\left[\sum_{j=1}^{\infty} M_j \delta_{U_j}(du)\delta_u(dv)\right. \\
&\qquad \left. - \left(\sum_{j,k=1}^{\infty} M_j M_k \delta_{U_j}(du)\delta_{U_k}(dv)\right)\right] \\
&= \theta(du)\delta_u(dv) - \sum_{\substack{j,k=1\\ j\neq k}}^{\infty} E[M_j M_k]\theta(du)\theta(dv) \qquad (2.4)\\
&\quad - \sum_{\ell=1}^{\infty} E[M_\ell^2]\theta(du)\delta_u(dv) \\
&= \eta[\theta(du)\delta_u(dv) - \theta(du)\theta(dv)],
\end{aligned}$$

where

$$\eta = 1 - \sum_{j=1}^{\infty} E[M_j^2] \quad \left(\text{since } \sum_{j=1}^{\infty} M_j = 1\right).$$

We shall now verify that $\sum_{j=1}^{\infty} E[M_j^2] = \gamma/(1+\gamma)$. If we define

$$A := \frac{2\gamma^2}{(\gamma+1)(2\gamma+1)}, \qquad B := 1 + A - \frac{2\gamma}{(1+\gamma)},$$

then we obtain

$$E[M_j^2] = E\left[V_j^2 \prod_{i=1}^{j-1}(1-V_i)^2\right] = AB^{j-1},$$

$$\sum_j E[M_j^2] = A/(1-B).$$

Hence

$$\sum_{j=1}^{\infty} E[M_j^2] = \frac{\gamma}{1+\gamma}.$$

Therefore in (2.4) $\eta = \frac{1}{1+\gamma}$ and the proof is complete. □

Remark. (2.3) immediately yields for the equilibrium state of (2.2) with $c_k = c$, $d_k = d$ that

$$E_{\Gamma_\theta}(d \cdot Q_z(du, dv)) = \frac{cd}{c+d} \cdot Q_\theta(du, dv). \qquad (2.5)$$

3. Duality and behavior of the infinite system: Proof of Theorem 0.1

(a) **Duality.** A central tool in proving theorems about the behavior of systems of interacting Fleming-Viot processes is the duality relation between these systems and systems of *delayed coalescing random walks* (For other methods compare 4(c).) The purpose of this subsection is to formulate the appropriate dual process for our system. In order to reduce the notation we shall write down everything for the case $I = \Omega_N$. Since N is a fixed parameter, in section 3 we will simply write Ω instead of Ω_N. We construct the duality in three steps.

(i) In the first step we introduce a set of functions which determines a probability law on $(\mathscr{P}([0, 1]))^\Omega$ and for which the duality relation will later be formulated. Define for $X(t) = \{x_\xi(t)\}_{t\geq 0}$ the following "characteristic" functions:

$$F_{(\xi^i), f}(t) = E\left(\int_0^1 \cdots \int_0^1 f(u_1, \dots, u_m) x_{\xi^1}(t)(du_1) \cdots x_{\xi^m}(t)(du_m)\right) \tag{3.1}$$

with $f \in C([0, 1]^m)$, $(\xi^i)_{i=1}^m \in \Omega^m$, $m \in \mathbb{N}$.

Remark. Note that by choosing n distinct points $\tilde{\xi}^i$, $i = 1, \dots, n$, in Ω and considering $(\tilde{\xi}^1, \dots, \tilde{\xi}^2, \dots, \tilde{\xi}^{n-1}, \dots, \tilde{\xi}^n)$, where $\tilde{\xi}^1$ appears n_1 times, etc. and setting $f \in C([0, 1]^{n_1+\cdots+n_n})$ to be equal to

$$f(u_1, \dots, u_m) = \prod_{i=1}^m g_i(u_i), \quad m = \sum_{i=1}^n n_i, \qquad g_i \in C([0, 1]), \ i = 1, \dots, m,$$

we prescribe the mixed moment measures of all orders of the field $\{x_\xi\}_{\xi\in\Omega}$ and hence determine its distribution uniquely. Since $C([0, 1])$ separates points in $\mathscr{P}([0, 1])$, the law of $X(t)$ is determined once we know

$$\left\{F_{(\xi^i), f}(t) | (\xi^i)_{i=1}^m \in \Omega^m, \text{ with } f(u_1, \dots, u_m) = \prod_{i=1}^m g_i(u_i), \right.$$
$$\left. g_i \in C([0, 1]), \ i = 1, \dots, m; \ m \in \mathbb{N} \right\}. \quad \square$$

Since the next objective is to determine $\{F_{(\xi^i), f}(t), t \geq 0\}$ via a suitable dual process, we first rewrite $\{F_{(\xi^i), f}(t), t \geq 0\}$ in the standard form for duality relations. For every $f \in C([0, 1]^n)$, $f \geq 0$, we set

$$H_f(X, \eta) : (\mathscr{P}([0, 1]))^\Omega \times \Omega^n \to \mathbb{R}^+,$$
$$H_f(X, \eta) = \int f(u_1, \dots, u_n) x_{\eta^1}(du_1) \cdots x_{\eta^n}(du_n), \tag{3.2}$$
$$X = (x_\xi)_{\xi\in\Omega} \in (\mathscr{P}([0, 1]))^\Omega, \ \eta = (\eta^1, \dots, \eta^n) \in \Omega^n.$$

Then

$$F_{(\xi^i), f}(t) = EH_f(X(t), \eta), \quad \text{where } \eta = (\xi^1, \dots, \xi^n). \tag{3.3}$$

(ii) The next step is to define the dual process, a *delayed coalescing random walk*. Denote by

$$\zeta_n = \text{ set of partitions of } \{1, \dots, n\}. \tag{3.4}$$

For $\pi \in \zeta_n$, let $|\pi|$ denote the number of nonempty partition elements and represent π as a mapping from $\{1, \dots, n\}$ onto $\{1, \dots, |\pi|\}$; that is, π^{-1} gives the indices of the particles in a particular position. Each partition element is a subset of $\{1, \dots, n\}$. For every element $\eta \in \Omega^n$, written as $\eta = (\eta^1, \dots, \eta^n)$, we define a Markov process

$$(\hat{\eta}_t)_{t\in\mathbb{R}^+} = (\eta_t^1, \dots, \eta^{|\pi_t|}; \pi_t) \tag{3.5}$$

with state space $\mathscr{Z}_n := \{(\eta^1, \dots, \eta^{|\pi|}; \pi) | \eta^j \in \Omega,\ j = 1, \dots, |\pi|,\ \pi \in \zeta\}$, initial state $\hat{\eta}_0 = (\eta^1, \dots, \eta^n; \{\{1\}, \dots, \{n\}\})$ and in which η_t^j denotes the location of the jth partition element at time t. The time evolution consists of two types of transition given by the rules

(3.6)
(a) The locations of the partition elements perform independent continuous time *random walks* on Ω with transition rates $a(\xi, \xi')$ (we can think of this as a set of particles moving together).
(b) All pairs of partition elements in π whose locations in Ω coincide *coalesce* at a fixed rate, d, independently of each other yielding a partition with one fewer elements.

(iii) We now extend H_f to a function $\widehat{H}_f$ on $(\mathscr{P}([0, 1]))^\Omega \times \mathscr{Z}_n)$, in order to be able to formulate the duality relation. Set

$$\begin{aligned} \widehat{H}_f(X, \hat{\eta}) &= \widehat{H}_f(X, (\eta, \pi)) \\ &= \int \cdots \int f(u_{\pi(1)}, \dots, u_{\pi(n)}) x_{\eta^1}(du_1) \cdots x_{\eta^{|\pi|}}(du)_{|\pi|}, \\ \eta &= (\eta^1, \dots, \eta^{|\pi|}). \end{aligned} \tag{3.7}$$

Now we are ready for:

Lemma 3.1 (Duality). *For every* $\eta \in \Omega^n$, $f \in C([0, 1]^n)$, $\pi \in \zeta_n$, $n \in \mathbb{N}$:

$$E\widehat{H}_f(X(t), \hat{\eta}_0) = E\widehat{H}_f(X(0), \hat{\eta}_t), \tag{3.8}$$

where $\hat{\eta}_0 = (\eta, \pi)$. □

Proof. Denote by G the generator of the system of interacting Fleming-Viot processes and by G^* the generator of the system of coalescing random walks with delay. Then by inspection

$$\begin{gathered} G\widehat{H}_f(\cdot, \hat{\eta}_0)(X, \hat{\eta}_0) = (G^* H_f(X, \cdot))(X, \hat{\eta}_0), \\ X \in (\mathscr{P}([0, 1]))^\Omega, \text{ and } \hat{\eta}_0 = (\eta, \pi) \text{ with } \eta \in \Omega^{|\pi|},\ \pi \in \zeta_n. \end{gathered} \tag{3.9}$$

Furthermore the linear hull of the set

(3.10)
$$\{\widehat{H}_f(\cdot, \hat{\eta}_0) : f \in C([0, 1]^n), \hat{\eta}_0 = (\eta, \{\{1\}, \dots, \{n\}\}),\ \eta \in \Omega^n,\ n \in \mathbb{N}\}$$
is dense in $C(\mathscr{P}([0, 1])^\Omega)$,

and finally

$$\begin{gathered} \{\widehat{H}_f(X, \cdot) : f \in C([0, 1]^n), X \in (\mathscr{P}([0, 1]))^\Omega, n \in \mathbb{N}\} \\ \text{is dense in } L_\infty(\mathscr{Z}^n) \text{ for all } n \in \mathbb{N}. \end{gathered} \tag{3.11}$$

The first statement was shown in the remark following (3.1). The second statement follows since these functions separate points of the countable set $(\Omega^n \times \zeta_n)$.

The relations (3.9)–(3.11) then imply the relations (3.8) as follows. For $t = 0$ and left-hand side and the right-hand side of (3.8) are equal and the left-hand side of (3.8) is a solution to a linear system of countably many differential equations which is given by G^*. Since G^* is a bounded operator on $L_\infty(\mathcal{Z}^n)$, the solution of this system, that is, the right-hand side of (3.8), is (uniquely) given by $\exp(tG^*)[H_f(X(0), \cdot)]$.

(b) **Proof of Theorem 0.1.** The proof is based on the duality from subsection 3(a). We start with some preparatory remarks on coalescing random walks. The first observation is the following. Suppose $Z_1(t)$, $Z_2(t)$ are independent continuous time random walks on Ω with transition kernel $\hat{a}(\xi, \xi')$; then $(Z_1(t) - Z_2(t))_{t\geq 0}$ is a random walk on Ω with transition kernel $\hat{a}(\xi, \xi')$ and jump rate 2, where

$$\hat{a}(\xi, \xi') = \frac{1}{2}(a(\xi, \xi') + a(\xi', \xi)). \tag{3.12}$$

Next since $|\pi_t|$ is monotone decreasing, we can define the random variable π_∞:

$$\pi_\infty = \lim_{t\to\infty} \pi_t. \tag{3.13}$$

Finally we discuss the behavior of η_t for $t \to \infty$. We have to distinguish between the two cases, $\hat{a}$ recurrent, respectively transient. Denote by $Z(t)$ a random walk on Ω with transition rate $a(\xi, \xi')$ starting at 0 and by $\{Z_i(t)\}_{i\in\mathbb{N}}$ a sequence of independent copies of $Z(t)$. We can assume that the system of coalescing random walks with delay is constructed on a probability space on which both the sequence $\{Z_i(t)\}_{i\in\mathbb{N}}$ and an independent collection of exponentially distributed random variables are defined. Then the behavior, as $t \to \infty$, of the system of coalescing random walks with delay is as follows:

Lemma 3.2 (Coalescing systems).

$$\begin{gathered}\hat{a} \text{ recurrent}: \\ \mathcal{L}(\hat{\eta}_t) - \mathcal{L}((Z(t); \{1, \dots, n\})) \underset{t\to\infty}{\Rightarrow} 0.\end{gathered} \tag{3.14}$$

For the case in which $\hat{a}$ is transient we write π_∞ as

$$\pi_\infty = \{\pi_\infty(1), \dots, \pi_\infty(n)\}. \tag{3.15}$$

$\hat{a}$ *transient*:

$$\mathcal{L}(\eta_t \,|\, |\pi_\infty| = k) - \mathcal{L}(Z_1(t), \dots, Z_k(t)) \underset{t\to\infty}{\Rightarrow} 0 \quad \forall k \in \{0, 1, \dots\}, \tag{3.16}$$

$$P(|\pi_\infty| = 1) < 1, \text{ provided } |\pi_0| \neq 1. \quad \square$$

Proof. Consider first the case in which $\hat{a}$ is recurrent. Since $Z_1(t) - Z_2(t)$ is a random walk with jump rate 2 and transition kernel $\hat{a}$, we know that with probability one all particles coalesce. This implies (3.14).

In the case in which $\hat{a}$ is transient, there exists a random time σ, $\sigma < \infty$ a.s., such that at σ the last collision of particles (in particular the last coalescence) occurs in the system $\hat{\eta}_s$. Denote by $\xi^1(t), \dots, \xi^{|\pi_\infty|}(t)$ the position of the partition elements at time $\sigma + t$. The system η_u behaves for times $u = s + t + \sigma$ like a system of $|\pi_\infty|$ random walks in s, starting at $\xi(t), \dots, \xi^{|\pi_\infty|}(t)$ and conditioned on never meeting. Since for every pair i, j with $i \neq j$, $\xi^i(t) - \xi^j(t) \underset{t\to\infty}{\to} \infty$, the event that ξ^i and ξ^j never meet after time t has probability tending to one as $t \to \infty$. Hence it suffices to know the following:

Let $(Z_1^1(t), Z_2^1(t), \dots, Z_m^1(t))$, $(Z_1^2(t), Z_2^2(t), \dots, Z_m^2(t))$ be independent collections of m independent copies of $Z(t)$, starting at $(0, \dots, 0)$, respectively a vector $(\xi^1, \dots, \xi^m)$. Then there exist random times S, T such that

$$\mathscr{L}((Z_1^1(S), \dots, Z_m^1(S))|S) = \mathscr{L}((Z_1^2(T), \dots, Z_m^2(T))|T),$$
$$\text{Prob}(S < \infty) = \text{Prob}(T < \infty) = 1,$$
$$\text{Prob}(S = n) = \text{Prob}(T = n).$$

According to Corollary 1 in Greven (1987) this follows for discrete time Markov processes if all bounded space-time harmonic functions are constant. The latter fact follows for aperiodic random walks from the lemma of Choquet and Deny (cf. Spitzer (1964, Chapter 6, T1)) (aperiodicity is true in our case since $a(0, 0) > 0$). Since the jump times are independent of the state, the continuous time statement follows immediately from the discrete time result as follows: the difference of the two processes counting the jumps is a random walk on $\mathbb{Z}$ with expected jump size, 0 hence a recurrent process.

Now we are ready for the proof of Theorem 0.1.

Case 1. $\hat{a}$ recurrent.

We proceed in three steps. First we show that the components concentrate on δ-measures and then in a second step that we have consensus among the components and finally we combine these facts.

Step 1. Choose in (3.2) $m = 2$, $f(u_1, u_2) = g(u_1)g(u_2)$, $\eta = (\xi, \xi)$ and get from (3.7)

$$\begin{aligned} &E\left(\int_0^1 g(u)x_\xi(du)\right)^2 \\ &= E\left(\int g^2(u)x_{\eta_t^1}(du)\mathbf{1}(\pi_t = \{1, 2\})\right. \\ &\qquad \left. + \int g(u)x_{\eta_t^1}(du) \cdot \int g(u)x_{\eta_t^2}(du)\mathbf{1}(\pi_t = \{\{1\}, \{2\}\})\right). \end{aligned} \tag{3.17}$$

Since

$$\text{Var}_{x_\xi}(g) = \int g^2(u)x_\xi(du) - \left(\int g(u)x_\xi(du)\right)^2, \tag{3.18}$$

we have using (3.14) in (3.17) that

$$E(\text{Var}_{x_\xi(t)}(g)) \underset{t\to\infty}{\longrightarrow} 0. \tag{3.19}$$

Since $\mathscr{L}(x_\xi(t)) \in \mathscr{P}(\mathscr{P}([0, 1]))$, the set

$$\{\mathscr{L}(x_\xi(t))\}_{t\geq 0} \tag{3.20}$$

is weakly relatively compact. A weak limit point $\mathscr{L}(x_\xi(\infty))$ of the sequence must according to (3.19) be concentrated on

$$M = \{\delta_u | u \in [0, 1]\}; \tag{3.21}$$

i.e. $\mathscr{L}(x_\xi(\infty)) = \int_0^1 \delta_{(\delta_u)}H_\xi(du)$ for some $H_\xi(\cdot) \in \mathscr{P}([0, 1])$. But we know from (3.8) that for $\mathscr{L}(\{x_\xi(0)\}_{\xi\in\Omega}) = \mu$

$$E\langle x_\xi(t), f\rangle = \langle \theta, f\rangle \tag{3.22}$$

so that for such a limit point $\mathscr{L}(\{x_\xi(\infty)\}_{\xi\in\Omega})$

$$E\langle x_\xi(\infty), f\rangle = \langle \theta, f\rangle \quad \forall f \in C([0, 1]). \tag{3.23}$$

We conclude from (3.23) and (3.21) that $H_\xi = \theta$ so that

$$\mathscr{L}(x_\xi(\infty)) = \int_0^1 \delta_{(\delta_u)}\theta(du) \tag{3.24}$$

which finishes the first step.

Step 2. In order to show that consensus of the components occurs for $t \to \infty$ set in (3.2) $m = 2$, $f(u_1, u_2) = g(u_1)g(u_2)$ but use $\eta = (\xi^1, \xi^2)$ with $\xi^1 \neq \xi^2$. Then we get from (3.8)

$$\begin{aligned} &E\left(\int g(u)x_{\xi^1}(t)(du)\int g(u)x_{\xi^2}(t)(du)\right)\\ &\quad = E\left(\int g^2(u)x_{\eta_t^1}(du)\mathbf{1}(\pi_t = \{1, 2\})\right.\\ &\qquad \left.+\int g(u)x_{\eta_t^1}(du)\int g(u)x_{\eta_t^2}(du)\mathbf{1}(\pi_t = \{\{1\}, \{2\}\})\right). \end{aligned} \tag{3.25}$$

By (3.14), (3.24) the right-hand side (cf. (3.25)) converges to

$$\int g^2(u)\theta(du) \tag{3.26}$$

and the left-hand side to

$$E\left(\int g(u)x_{\xi^1}(\infty)(du)\int g(u)x_{\xi^2}(\infty)(du)\right). \tag{3.27}$$

Since indicators of open or closed intervals can be pointwise approximated by bounded continuous functions, we have (3.25)–(3.27) for such indicator functions. Hence choosing $g(\cdot) = \mathbf{1}_A$, A an interval, we conclude from (3.24) and (3.25)–(3.27) that

$$\begin{aligned} \theta(A) &= Ex_{\xi^i}(\infty)(A), \qquad i = 1, 2,\\ \theta(A) &= E(x_{\xi^1}(\infty)(A)x_{\xi^2}(\infty)(A)). \end{aligned} \tag{3.28}$$

This implies with (3.24) ($x_\xi(\infty)(A)$ is either 0 or 1 !):

$$x_{\xi^1}(\infty) = x_{\xi^2}(\infty) \quad \forall \xi^1, \xi^2 \in \Omega \tag{3.29}$$

so that consensus occurs as $t \to \infty$.

Step 3. Combining (3.29) and (3.24) yields

$$\mathscr{L}((x_\xi(t))_{\xi\in\Omega}) \underset{t\to\infty}{\Rightarrow} \int \delta_{(\delta_u)^\Omega}\theta(du). \tag{3.30}$$

This proves the assertion (0.16) and (0.18).

Case 2. $\hat{a}$ transient.

In order to establish the convergence of $\mathscr{L}(X(t))$ for $t \to \infty$ we shall need below the following two facts (here μ is again a homogeneous ergodic law on $(\mathscr{P}([0,1]))^\Omega$ with marginal mean measure θ).

$$\pi_t \underset{t\to\infty}{\rightarrow} \pi_\infty, \tag{3.31}$$

$$\sum_{(\xi^1,\dots,\xi^n)} \langle x_{\xi^1}(0), f\rangle \cdots \langle x_{\xi^n}(0), f\rangle P(Z_1(t)=\xi^1,\dots,Z_n(t)=\xi^n) \rightarrow \langle\theta, f\rangle^n, \quad \mu\text{-probability}. \tag{3.32}$$

The last relation follows from the fact that $\langle x_\xi(0), f\rangle_{\xi\in\Omega}$ under μ is a homogeneous ergodic field with mean-value $\langle\theta, f\rangle$ (cf. abstract L^2-ergodic theorem in Fleischmann (1978)). The first relation is (3.13). Furthermore note that due to the transience of $\hat{a}$ (see (3.12)) it is easy to show using characteristic functions that the probability of the event that a random walk with transition kernel $\hat{a}$ starting in ξ hits $\xi^0 = (0,0,\dots)$ tends to zero as $d(\xi,\xi^0)$ tends to infinity. Hence

$$\begin{aligned} P(\pi_\infty = \{\{1\},\dots,\{n\}\} \mid \eta^1 = \xi^1,\dots,\eta^n=\xi^n, \\ \pi_0 = \{\{1\},\dots,\{n\}\}) \underset{\substack{|\xi^i-\xi^j|\to\infty \\ i\neq j}}{\rightarrow} 1. \end{aligned} \tag{3.33}$$

Now we are ready to start the proof. Observe that in (3.16) the distribution of a single $Z_{k(i)}(t)$ is that of the random walk $Z(t)$. Then combine (3.31), (3.32) with (3.15), (3.16) and insert this result in the right-hand side of (3.8). Choose $f(u_1,\dots,u_n) = g(u_1)\cdots g(u_n)$ and let $j_1,\dots,j_{|\pi_\infty|}$ denote the cardinality of the elements of π_∞. Then

$$E\widehat{H}_f(X(t),\hat{\eta}) \underset{t\to\infty}{\rightarrow} E\prod_{i=1}^{|\pi_\infty|}\langle\theta, g^{j_i}\rangle. \tag{3.34}$$

This proves that, for every $\mu\in\mathscr{T}_\theta$, the process starting in μ, $X^\mu(t)$, satisfies: there is a $\nu_\infty \in \mathscr{P}(\mathscr{P}([0,1]))$ with the two properties:

$$\mathscr{L}(X^\mu(t)) \Rightarrow \nu_\theta \quad \text{as } t\to\infty, \tag{3.35}$$

$$E^{\nu_\theta}\langle x_\xi, g\rangle = \langle\theta, g\rangle. \tag{3.36}$$

Due to the relation (3.33) we know furthermore

$$E^{\nu_\theta}\prod_{i=1}^n \langle x_{\xi^i}, g\rangle \underset{\substack{|\xi^i-\xi^j|\to\infty \\ i\neq j}}{\rightarrow} \langle\theta, g\rangle^n. \tag{3.37}$$

Hence the limit measure ν_θ is homogeneous, spatially ergodic and has mean measure θ.

It remains to show that ν_θ is an invariant measure for the dynamics. (The extremality is already a consequence of the spatial ergodicity.) Denote by $S(t)$ the semigroup of the process $X(t)$ and by μ_θ the measure concentrated on the state $(\delta_\theta)^{\Omega_N}$. Then by the Feller property (recall Theorem 0.0)

$$\nu_\theta S(t) = \operatorname*{w-lim}_{s\to\infty} \mu_\theta S(s+t) = \operatorname*{w-lim}_{u\to\infty}\mu_\theta S(u). \tag{3.38}$$

Since μ_θ is in $\mathscr{T}(\theta)$, the right-hand side of (3.38) equals ν_θ. This shows that ν_θ is an invariant measure. This finishes the proof of (0.14), (0.15) and (0.17).

(c) **Proof of (0.68).** Note that

$$\begin{aligned}
&\mathrm{Var}(\langle x_{\xi,|k|}(t), g\rangle)\\
&\quad = N^{-2k}\Bigg\{\sum_{\xi' : d(\xi,\xi')\le k} \mathrm{Var}(\langle x_{\xi'}(t), g\rangle)\\
&\qquad + \sum_{\substack{d(\xi,\xi'),d(\xi,\xi'')\le k\\ d(\xi',\xi'')\le[k/2]}} \mathrm{Cov}(\langle x_{\xi'}(t), g\rangle, \langle x_{\xi''}(t), g\rangle)\\
&\qquad + \sum_{\substack{d(\xi,\xi'),d(\xi,\xi'')\le k\\ d(\xi',\xi'')>[k/2]}} \mathrm{Cov}(\langle x_{\xi'}(t), g\rangle, \langle x_{\xi''}(t), g\rangle)\Bigg\}.
\end{aligned} \tag{3.39}$$

Since $\mathrm{Var}(\langle x_{\xi'}(t), g\rangle)$ are uniformly bounded, the first and second terms go to zero as $k \to \infty$ uniformly in $N \ge 2$. But using the duality relation

$$\begin{aligned}
&\mathrm{Cov}(\langle x_{\xi'}(t), g\rangle, \langle x_{\xi''}(t), g\rangle)\\
&\quad = E\left(\int g(u)x_{\xi'}(t)(du)\int g(u)x_{\xi''}(t)(du)\right) - \left(\int g(u)\theta(du)\right)^2\\
&\quad = E\Bigg(\int g^2(u)\theta(du)\mathbf{1}(\pi_t = \{1, 2\})\\
&\qquad\qquad + \int g(u)x_{\eta_t^1}(du)\int g(u)x_{\eta_t^2}(du)\mathbf{1}(\pi_t = \{\{1\}, \{2\}\})\Bigg)\\
&\quad - \left(\int g(u)\theta(du)\right)^2.
\end{aligned} \tag{3.40}$$

But for the coalescing random walks satisfying $\sum_k 1/c_k < \infty$ (cf. Dawson and Greven (1993c, 1.40 and Lemma 6.1)), $P(\pi_t = \{1, 2\}|\eta_0^1 = \xi', \eta_0^2 = \xi'') \to 0$ (uniformly for $N > N_0$ for some N_0) as $d(\xi', \xi'') \to \infty$ and conditioned on $\pi_t = \{\{1\}, \{2\}\}$, $d(\eta_t^1, \eta_t^2) \to \infty$ in probability as $d(\xi', \xi'') \to \infty$. Hence $\mathrm{Cov}(\langle x_{\xi'}(t), g\rangle, \langle x_{\xi''}(t), g\rangle) \to 0$ as $d(\xi', \xi'') \to \infty$ and therefore

$$\mathrm{Var}(\langle x_{\xi,|k|}(t), g\rangle) \to 0 \quad \text{as } k \to \infty \text{ uniformly in } t \text{ and } N > N_0.$$

4. The multiple time scale behavior for finite systems without hierarchy in the mean-field limit $N \to \infty$

The major issues we deal with in this section and the next are the following two problems which arise in connection with the formulation and proof of the multiple space-time scale behavior. First we have to deal with a sequence of *infinite* systems of interacting components, indexed by N, which we analyse in the limit $N \to \infty$, in contrast to the usual situations in which one considers McKean-Vlasov limits of finite systems. Second we work in *fast time scales* $\beta(N)t$, not for fixed t. In our treatment we shall separate these two problems. First (in this section 4) we shall investigate an analogous system with finitely many (namely N) components without hierarchy in a fast time scale Nt, as $N \to \infty$. In the second step in section 5 we shall then extend these results to treat the system with infinitely many components.

(a) **Definition and properties of the finite system without hierarchy in fast time scales.** We want to reduce the complexity of our system. The idea is to remove the interaction between components at a hierarchical distance of more than k. This interaction vanishes if $c_j = 0 \ \forall j > k$. These "truncated" systems are essentially *finite* systems since all components can be grouped in finite blocks, which evolve independently. In order to analyse these finite systems we start in this subsection with the simplest one where $k = 1$, i.e. $c_j = 0$ for $j = 1, 2, 3, \dots$. Then we might as well consider only the evolution of one single block $\{\{\xi, 0, 0, \dots\} | \xi \in \{0, \dots, N-1\}\}$ and then view this system as one which has state space $(\mathscr{P}([0,1]))^N$ and write c, d instead of c_0, d_0:

Definition. $(Y^N(t))_{t\geq 0}$ is the (unique) solution of the martingale problem on $C([0,\infty), (\mathscr{P}([0,1]))^N)$ given by $(\widetilde{L}_{N,\mu_N})$. Here with $y = (y_1, \dots, y_N) \in (\mathscr{P}([0,1]))^N$:

$$\widetilde{L}_N(F)(y) = \sum_{\xi=0}^{N-1}\left(\int_0^1 \frac{c}{N}\sum_{\xi'=0}^{N-1}\frac{\partial F(y)}{\partial y_{\xi'}}(u)(y_{\xi'}(du) - y_\xi(du))\right)$$

$$+\sum_{\xi=0}^{N-1}\int_0^1\int_0^1 d\frac{\partial F^2(y)}{\partial y_\xi \partial y_\xi}(u, v)Q_{y_\xi}(du, dv) \tag{4.1}$$

$$\text{where } Q_{y_\xi}(du, dv) = (y_\xi(du)\delta_u(dv) - y_\xi(du)y_\xi(dv)),$$

$$F(z) = \prod_{i=1}^m \int_0^1 u^{a_i} z(du) \quad \text{with } a_i \in \mathbb{N},\ z \in \mathscr{P}([0,1]).$$

The initial law μ_N is an *N-exchangeable* measure on $(\mathscr{P}([0,1]))^N$ with marginal mean measure $\theta \in \mathscr{P}([0,1])$; that is, $\{y_\xi(0) : \xi \in \{0, \dots, N-1\}\}$ are N-exchangeable $\mathscr{P}([0,1])$-valued random variables whose joint distribution is invariant under permutations of $\{0, \dots, N-1\}$. □

Remark. Similarly, when extending the results of the present section to the hierarchical system with state space $(\mathscr{P}([0,1]))^{\Omega_N}$ we will consider laws μ_N which are *multi-N-exchangeable*, that is, invariant under all maps of the form

$$(x_\xi)_{\xi\in\Omega_N} \to (x_{\tilde{\pi}\xi})_{\xi\in\Omega_N},$$

where $\tilde{\pi}$ has the form

$$\tilde{\pi}(\xi_1, \xi_2, \dots, \xi_k, \dots) = (\xi_1, \xi_2, \dots, \pi\xi_k, \dots) \quad \text{for some } k \in \mathbb{N}$$

and π is a permutation of $\{0, \dots, N-1\}$.

It then follows that, for every $t \geq 0$, $\{y_\xi(t) : \xi \in \{0, \dots, N-1\}\}$ are N-exchangeable $\mathscr{P}([0,1])$-valued random variables. It will turn out that the behavior of the above system in fast time scales Nt $(N \to \infty)$ is determined by the random variable θ_t^N with values in $\mathscr{P}([0,1])$ defined as follows (empirical distribution of types):

$$\langle \theta_t^N, f\rangle = \left\langle \frac{1}{N}\sum_{\xi=0}^{N-1} y_\xi(t), f\right\rangle, \quad \forall f \in L_\infty([0,1]). \tag{4.2}$$

The crucial results of this section are the following two propositions, which provide a complete picture of the *finite process in the fast time scale* Nt. Recall that Γ^0_θ denotes the equilibrium of a Fleming-Viot process on $\mathscr{P}([0,1])$ with immigration given by θ (see (0.25)) at rate $c = c_0$.

Proposition 4.1. *Suppose that $\{N_k\} \subset \mathbb{N}$ is a subsequence such that the following two statements hold for some fixed $s \in [0,\infty)$:*

(i)

$$\mathscr{L}((\theta^{N_k}_{tN_k})_{t\in\mathbb{R}^+}) \underset{k\to\infty}{\Rightarrow} \mathscr{L}((\theta_t)_{t\in\mathbb{R}^+})$$

where $(\theta_t)_{t\in\mathbb{R}^+}$ is a stochastic process with values in $\mathscr{P}([0,1])$ and continuous paths;

(ii)

$$\mathscr{L}(Y^{N_k}(sN_k)) \underset{k\to\infty}{\Rightarrow} \nu(s) \in \mathscr{P}((\mathscr{P}([0,1]))^{\mathbb{Z}^+}).$$

Define $P_t = \mathscr{L}(\theta_t)$ and $\nu_\theta = \bigotimes_{\theta=0}^{\infty}\Gamma^0_\theta \in \mathscr{P}((\mathscr{P}([0,1]))^{\mathbb{Z}^+})$. Then the measure $\nu(s)$ is the law of an exchangeable sequence of $\mathscr{P}([0,1])$-valued random variables with de Finetti representation:

$$\nu(s) = \int_{\mathscr{P}([0,1])} P_s(d\rho)\nu_\rho. \quad \square \tag{4.3}$$

Proposition 4.2. *Let $(\theta_s)_{s\in\mathbb{R}^+}$ be the solution of the well-posed martingale problem (L, δ_θ) on $C([0,\infty): \mathscr{P}([0,1]))$ with*

$$\begin{gathered} L(F)(z) = \frac{cd}{c+d}\int_0^1\int_0^1 \frac{\partial^2 F(z)}{\partial z \partial z}(u,v)Q_z(du,dv) \\ \text{with } Q_z(du,dv) = z(du)\delta_u(dv) - z(du)z(dv) \text{ and} \\ F(z) = \prod_{i=1}^m \int_0^1 u^{a_i} z(du), \qquad a_i \in \mathbb{N},\ m \in \mathbb{N},\ z \in \mathscr{P}([0,1]). \end{gathered} \tag{4.4}$$

Then

$$\mathscr{L}((\theta^N_{sN})_{s\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((\theta_s)_{s\in\mathbb{R}^+}). \quad \square \tag{4.5}$$

Combining these two propositions above we obtain the key result of section 4:

Corollary. *Let μ be a product measure on $(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$ and μ_N the restriction to $(\mathscr{P}([0,1]))^N$. Choose this μ_N as initial law of $Y^N(\cdot)$. Furthermore, $(\theta_t)_{t\geq 0}$ is the process defined in* (4.4) *and ν_θ the measure introduced in Proposition* 4.1. *Then for every $t \in [0,\infty)$:*

$$\mathscr{L}(Y^N(tN)) \underset{N\to\infty}{\Rightarrow} \mathscr{L}(\nu_{\theta_t}) = \int_{\mathscr{P}([0,1])} \mathrm{Prob}(\theta_t \in d\theta)\nu_\theta. \quad \square \tag{4.6}$$

The basic tool for the first proposition is the *finite system scheme argument* discussed in 4(b) below based on *coupling* which we discuss in 4(c), and Proposition 2 is derived via the techniques of martingale problems in 4(d).

(b) **Mean-field finite system scheme: Proof of Proposition 4.1.** The heuristic idea is that the density (averages over the components) changes slowly compared to the evolution of the individual components. Hence we expect the system to relax into the equilibria corresponding to a system in which the density $N^{-1}\sum_\xi x_\xi(sN+t)$ is replaced by the fixed state, namely the measure

$N^{-1}\sum_\xi x_\xi(sN)$, on the original time scale t. To turn this into a rigorous proof the key idea is to view via periodic continuation $(Y^N(t))$ as a system indexed by $\mathbb{Z}^+$ rather than by $\{0, 1, \ldots, N-1\}$ and compare it with the auxiliary infinite system $(Y^\infty(t))_{t\geq 0}$ on $(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$ defined below which has independent components and each component evolves as Fleming-Viot process with immigration from source $\rho \in \mathscr{P}([0,1])$ where ρ is obtained as the average over the initial state of the components.

The formal definition of $Y^\infty(t)$ is as follows. Fix a $\rho \in \mathscr{P}([0,1])$. We take as initial distributions exchangeable measures μ on $\mathscr{P}([0,1]))^{Z^+}$ with the properties:

(i) μ is $L_2(\mu)$-ergodic; i.e. for every $f:[0,1]\to\mathbb{R}$ with $E^\mu f^2(y_0)<\infty$, the field $\{\langle y_i, f\rangle\}_{i\in\mathbb{Z}^+}$ has the property that

$$\frac{1}{M+1}\sum_{i=0}^{M}\langle y_i, f\rangle \underset{M\to\infty}{\to} E\langle y_0, f\rangle \quad L_2(\mu) \text{ and } \mu\text{-a.s.}$$

(ii) $E\langle y_0, f\rangle = \langle \rho, f\rangle\ \forall f\in L_\infty([0,1])$.

Definition. $(Y^\infty(t))_{t\geq 0}$ is defined to be the solution of the martingale problem $(\widetilde{L}_\infty, \mu^\infty)$ on $C([0,\infty):(\mathscr{P}([0,1]))^{\mathbb{Z}^+})$ with μ^∞ satisfying (i) and (ii) and $y\in(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$, $y=(y_i)_{i\in\mathbb{Z}^+}$:

$$\begin{aligned}\widetilde{L}_\infty(F)(y) = \sum_{i\in\mathbb{Z}^+}\Bigg[&c\int_0^1 \frac{\partial F(y)}{\partial y_i}(u)(\rho(du)-y_i(du)) \\ &+ d\int_0^1\int_0^1 \frac{\partial^2 F(y)}{\partial y_i \partial y_i}(u,v)Q_{y_i}(du,dv)\Bigg]\end{aligned} \tag{4.7}$$

where

$$F(y)=\prod_{i\in\mathbb{Z}^+}\int_0^1 u^{a_i}y_i(du), \qquad a_i\in\mathbb{N}, \sum_{i\in\mathbb{Z}^+}a_i<\infty, \text{ and}$$

$$Q_{y_k}(du,dv) = y_k(du)\delta_u(dv) - y_k(du)y_k(dv). \quad \square$$

We can obviously define the process as well for periodic initial states.

The argument is now based on an abstract scheme developed in Dawson and Greven (1993b) presented in step (i). So as not to obscure the main ideas of this argument we use five properties, Lemmas 4.1–4.5 below, which will only be subsequently reformulated and proved in step (ii).

(i) **Abstract scheme.** The space $\mathscr{P}(\mathscr{P}([0,1]))$ is compact in the weak topology of measures. Hence, given some sequence $L(N)$ with $L(N)\leq s'N$, with $s'<s$, we can choose a subsequence $\widetilde{N}_k$ of N_k (of Proposition 1) such that:

$$\mathscr{L}(Y^{\widetilde{N}_k}_{(s\beta(\widetilde{N}_k)-L(\widetilde{N}_k))}) \text{ converges weakly as } k\to\infty, \tag{4.8}$$

where $\beta(N):=N$. Denote the limit in (4.8) by μ. Note that μ is exchangeable since the initial distributions of Y^N were N-exchangeable measures. In the sequel we choose $L(N)$ such that

$$L(N)\uparrow +\infty \quad \text{as } N\to\infty, \quad L(N)=o(N). \tag{4.9}$$

Next by de Finetti's theorem (cf. Dawson (1993), section 11.2)) we can decompose μ as follows (condition on the density):

$$\mu = \int_{\mathscr{P}([0,1])} \mu_\rho \, d\Lambda(\rho) \tag{4.10}$$

with $\Lambda \in \mathscr{P}(\mathscr{P}([0,1]))$ and for each ρ, μ_ρ an i.i.d. measure on $(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$, satisfying

$$E^{\mu_\rho}[f(y_0)] = \langle \rho, f \rangle \quad \forall f \in L_\infty([0,1]).$$

By our assumption on (θ_t^N) we know that:

$$\mathscr{L}(\theta^{N_k}_{s\beta(N_k)}) \underset{k\to\infty}{\Rightarrow} P_s(\cdot), \qquad \mathscr{L}(\theta^{N_k}_{s\beta(N_k)-L(N_k)}) \underset{k\to\infty}{\Rightarrow} P_s(\cdot) \tag{4.11}$$

by the continuity of P_s. The measure Λ of (4.10) can of course be easily identified via the L_2-ergodic theorem:

$$\Lambda(\cdot) = \mathscr{L}\left(\lim_{n\to\infty}\left(\frac{1}{n+1}\sum_{i=0}^{n} y_i^\infty\right)\right) \quad \text{with } \mathscr{L}((y_i^\infty)_{i\in\mathbb{Z}^+}) = \mu. \tag{4.12}$$

We shall show in (4.26) of Lemma 4.3 below that

$$\Lambda(\cdot) = P_s(\cdot). \tag{4.13}$$

Hence we have from (4.10) that

$$\mu = \int_{\mathscr{P}([0,1])} \mu_\rho P_s(d\rho). \tag{4.14}$$

For a system $(Y^\infty(t))_{t\geq 0}$ with $\mathscr{L}(Y^\infty(0)) = \mu_\rho$ we shall show in (4.18) of Lemma 4.1 below that

$$\mathscr{L}(Y^\infty(t)) \Rightarrow \nu_\rho \quad \text{as } t\to\infty. \tag{4.15}$$

Next we shall see in (4.21) that $\theta \to \langle \nu_\theta, f\rangle$, $f \in C([0,1])$, is continuous and bounded. Furthermore $(Y^\infty(t))_{t\geq 0}$ is a Feller process. Hence we can summarize the last two relations to conclude that for $\mathscr{L}(Y^\infty(0)) = \mu$ we have

$$\mathscr{L}(Y^\infty(t)) \underset{t\to\infty}{\Rightarrow} \int \nu_\rho P_s(d\rho). \tag{4.16}$$

To finish the proof we introduce the measure $\tilde{\mu}_N$ on $(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$ as the distribution of the random element obtained by periodic continuation of $Y^N(s\beta(N) - L(N))$ to $(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$. Denote the process $Y^\infty(t)$ starting in $\tilde{\mu}_N$ by $Y_N^\infty(t)$ and by $Y_\infty^\infty(t)$ the process starting in μ, defined in (4.7). Write (abusing notation somewhat)

(4.17)

$$\begin{aligned}\mathscr{L}(Y^{\widetilde{N}_k}(s\beta(\widetilde{N}_k))) &= [\mathscr{L}(Y^{\widetilde{N}_k}(s\beta(\widetilde{N}_k))) - \mathscr{L}(Y^\infty_{\widetilde{N}_k}(L(\widetilde{N}_k)))] \\ &\quad + [\mathscr{L}(Y^\infty_{\widetilde{N}_k}(L(\widetilde{N}_k))) - \mathscr{L}(Y^\infty_\infty(L(\widetilde{N}_k)))] + \mathscr{L}(Y^\infty_\infty(L(\widetilde{N}_k))).\end{aligned}$$

We shall prove below (in (4.28) of Lemma 4.4 and (4.29) of Lemma 4.5) that the first two signed measures converge to the 0-measure weakly as $k\to\infty$, while according to (4.16) the third measure converges to $\int \nu_\rho P_s(d\rho)$ which finishes the proof.

(ii) **Key lemmas.** It remains now to derive and formulate the properties used in the argument above. This is the content of the next five lemmata. Some of the proofs involve coupling techniques. We shall devote section 4(c) to a systematic development of this method.

The longterm behavior of Fleming-Viot processes on $\mathscr{P}([0,1])$ with immigration is well known (cf. Ethier (1990), Ethier and Griffith (1987), Shiga (1982)). We can summarize the results in the following statement about the process $(Y^\infty(t))_{t\geq 0}$:

Lemma 4.1. *Let μ be a shift invariant and ergodic measure on $(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$ with mean measure ρ. For $\mathscr{L}(Y^\infty(0)) = \mu$, the system $(Y^\infty(t))$ behaves as follows:*

$$\begin{aligned}&\mathscr{L}(Y^\infty(t)) \underset{t\to\infty}{\Rightarrow} \nu_\rho,\\ (4.18)\qquad &\nu_\rho = \bigotimes_{x\in\mathbb{Z}^+} \Gamma_\rho, \quad \textit{with } \Gamma_\rho \textit{ as given in } (0.25) \textit{ and with } \gamma = d/c. \quad \square\end{aligned}$$

Some properties of the dependence of the measures above on the parameter ρ will be of importance:

Lemma 4.2. *Endow $\mathscr{P}([0,1])$ with the weak topology of measures and define the map $I : \mathscr{P}([0,1]) \to (\mathscr{P}([0,1]))^{\mathbb{Z}^+}$ by*

$$I(\theta) = \nu_\theta, \quad \theta \in \mathscr{P}([0,1]), \tag{4.19}$$

and the map $\mathscr{F} : C(\mathscr{P}([0,1])) \to L_\infty(\mathscr{P}([0,1]))$ by

$$\mathscr{F}(f)_{(\theta)} = E^{\nu_\theta}(f(y_i)). \tag{4.20}$$

Then

$$I \textit{ is continuous.} \tag{4.21}$$

Define

(4.22)

$$\|\nu - \mu\| := \inf\left[\int |u - v|\Gamma(du, dv) : \Gamma \in \mathscr{P}([0,1]^2),\ \pi_1(\Gamma) = \nu,\ \pi_2(\Gamma) = \mu\right].$$

If we assume that $f \in L_\infty(\mathscr{P}([0,1]))$ has the property $|f(\theta) - f(\theta')| \leq L(f)\|\theta - \theta'\|$, then

$$|\mathscr{F}(f)_{(\theta)} - \mathscr{F}(f)_{(\theta')}| \leq L(f)\|\theta - \theta'\|. \quad \square \tag{4.23}$$

Proof. If the Lipschitz continuous functions on $\mathscr{P}([0,1])$ (in the sense of (4.22)) were dense in $C(\mathscr{P}([0,1]))$, it would be enough to show the second relation (4.22). Assume this property of $C(\mathscr{P}([0,1]))$ for the moment. Now a coupling argument which also yields the result of Lemma 4.1 is essential. If we start $Y^\infty(t)$ in the states θ, and θ' respectively, we can define both processes on one probability space such that

$$\limsup_{t\to\infty} |\langle Y_1^\infty(t), f\rangle - \langle Y_2^\infty(t), f\rangle| \leq L(f)\|\theta - \theta'\|$$

which proves the assertion. The construction of such a bivariate dynamic can be found as a special case of that which will be given in section 4(c) (cf. Lemma 4.7).

In order to finish the argument we must verify that the Lipschitz functions on $\mathscr{P}([0, 1])$ are dense in $C(\mathscr{P}([0, 1]))$. To do this first note that a Lipschitz continuous function on $\mathscr{P}([0, 1])$ (in the sense of (4.22)) is continuous, and that $\mathscr{P}([0, 1])$ is compact in the weak topology. We then apply the Stone-Weierstrass theorem (Yosida (1980, p. 9)). If f_1, f_2 are Lipschitz functions, then f_1+f_2 is Lipschitz with $L(f_1+f_2) \leq L(f_1)+L(f_2)$, and $f_1 f_2$ is Lipschitz with $L(f_1 f_2) \leq \|f_2\|_\infty L(f_1) + \|f_1\|_\infty L(f_2)$. Clearly the collection of Lipschitz functions separate points. Hence the assumptions of that theorem are fulfilled.

The next step is to verify the stability property (4.13) of the estimator of θ with respect to weak convergence. This can be done in either of two ways. The first method is to consider μ_N as the restriction to $(\mathscr{P}([0, 1]))^{2N+1}$ of a translation invariant measure, μ, on $(\mathscr{P}([0, 1]))^{\mathbb{Z}}$ and then to apply the analogue of Lemma 2.3 of Cox, Greven and Shiga (1993). The second method which we will follow here is based on the property of exchangeability.

Lemma 4.3. *Let $(y_i^N)_{i\in[0,N-1]}$ be an exchangeable family of $\mathscr{P}([0, 1])$-valued random variables distributed according to μ_N. Assume that $\mu_N \Rightarrow \mu$ where μ is an exchangeable measure on $(\mathscr{P}([0, 1]))^{\mathbb{Z}^+}$ in the sense that, for every $k \in \mathbb{Z}^+$, $(y_0^N, \dots, y_k^N) \Rightarrow (y_0, \dots, y_k)$ in distribution as $N \to \infty$.*

Define random variables ϕ, ϕ^N on the probability spaces $(\mu, (\mathscr{P}([0, 1]))^{\mathbb{Z}^+})$, respectively $(\mu_N, (\mathscr{P}([0, 1]))^{[0,N]})$, by setting:

$$\langle \phi, f\rangle = \lim_{n\to\infty}\left(\frac{1}{n+1}\sum_{i=0}^{n}\langle y_i, f\rangle\right), \qquad f \in C([0, 1]), \tag{4.24}$$

$$\langle \phi^N, f\rangle = \frac{1}{N}\sum_{i=0}^{N-1}\langle y_i^N, f\rangle, \qquad f \in C([0, 1]). \tag{4.25}$$

Then

$$\mathscr{L}(\phi^N) \underset{N\to\infty}{\Rightarrow} \mathscr{L}(\phi). \quad \square \tag{4.26}$$

Proof. Choose first a countable dense subset M of $C([0, 1])$. For $f \in M$ a result of Aldous (1985, Proposition 7.20) states that

$$\mathscr{L}\left(\frac{1}{N+1}\sum_{i=0}^{N}\delta_{\langle y_i^N, f\rangle}\right) \Rightarrow \mathscr{L}\left(\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}\delta_{\langle y_i, f\rangle}\right) \quad \text{as } N \to \infty.$$

Since the mapping $\mu \to \int_{[0,1]} x\mu(dx)$ is continuous on $\mathscr{P}([0, 1])$, and M is convergence determining the result follows.

The proofs of the next two lemmas will be deferred to a separate section, since they are based on coupling techniques which we shall systematically develop there. The first of these lemmas formalises the fact that an infinite interacting system and a large finite one look similar for long times provided that the interaction is local enough. We first need some notation.

Let θ_s^N be the density process corresponding to $Y^N(s)$ according to (4.2). Continue the configuration of $Y^N(s\beta(N) - L(N))$ to $(\mathscr{P}([0, 1]))^{\mathbb{Z}^+}$ periodically. The distribution thus obtained is denoted by $\tilde{\mu}_N$. Let $Y_N^\infty(t)$ be the system starting with the distribution $\tilde{\mu}_N$ and with dynamics as introduced in the beginning of section 4(b), in (4.7), where we set $\rho = \theta^N_{s\beta(N)-L(N)}$.

Lemma 4.4 (Comparison of finite-infinite systems). *Fix some $s \in \mathbb{R}^+$ and a sequence $L(N) \uparrow +\infty$ as $N \to \infty$ but with $L(N) = o(N)$. Assume that (with $\|\cdot\|$ as defined in (4.22))*

$$\sup_{t \le L(N)} \|\theta^N_{s\beta(N)} - \theta^N_{s\beta(N)-t}\| \underset{N\to\infty}{\Rightarrow} 0 \text{ in probability.} \tag{4.27}$$

Then for every function $f \in C((\mathscr{P}([0,1]))^{\mathbb{Z}^+}, \mathbb{R})$ which depends only on finitely many components we have (for $N \ge N_0(f)$) that:

$$|E[f(Y^\infty_N(L(N)))] - E[f(Y^N_{s\beta(N)})]| \underset{N\to\infty}{\to} 0. \quad \square \tag{4.28}$$

Finally some *uniformity* in the ergodic theorem for the infinite system $(Y^\infty(t))_{t\ge 0}$ is needed, which roughly says that the limits of the laws of the process for $t \to \infty$ are a continuous function of the initial configuration so that we can apply to the situation in which the initial measure changes as a function of time.

Again we need some notation for the precise statement. Let μ_N be a sequence of measures on $(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$ such that, for each N, μ_N is concentrated on N-periodic configurations, or is exchangeable, and furthermore under μ_N, $\{y_\xi : \xi = 0, \dots, N-1\}$ is N-exchangeable. Let μ be an exchangeable measure on $(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$. Assume that $\mu_N \Rightarrow \mu$ as $N \to \infty$. Denote by θ^N, θ the density corresponding to μ_N (according to (4.2)), μ respectively. Denote by $Y^\infty_N(t)$, $Y^\infty_\infty(t)$ the process $Y^\infty(t)$ defined in (4.7) with $\rho = \theta^N$, respectively $\rho = \theta$, and started with initial distribution μ_N, respectively μ.

Lemma 4.5. *Let $L(N)$ be a sequence with $L(N) \to \infty$ as $N \to \infty$. Then for every $f \in C((\mathscr{P}([0,1]))^{\mathbb{Z}^+}, \mathbb{R})$ we have*

$$|E[f(Y^\infty_N(L(N)))] - E[f(Y^\infty_\infty(L(N)))]| \underset{N\to\infty}{\Rightarrow} 0. \quad \square \tag{4.29}$$

(c) **Coupling.** In the arguments needed to prove our various lemmata, it is essential to compare processes starting in different initial states or processes using different sources θ for the immigration term. A technique very useful for these purposes is the so-called *coupling* of two processes. In the context of interacting particle systems or interacting diffusions these coupled processes can easily be defined. In our infinite-dimensional context this is slightly more complex. We start by introducing a "coupling" of two measures, a concept which is then used to define a coupling between two Fleming-Viot processes with possibly different immigration term.

For two elements $\theta_1, \theta_2 \in \mathscr{P}([0,1])$ we shall define a *coupling* $\hat\theta$ to be a measure on $(\mathscr{P}([0,1]))^2$ with (π_i denotes projection on the ith coordinate, $i = 1, 2$)

$$\begin{gathered}\hat\theta(A \times [0,1]) = \theta_1(A) \quad \forall A \in \mathscr{B}([0,1]),\\ \hat\theta([0,1] \times A) = \theta_2(A) \quad \forall A \in \mathscr{B}([0,1]),\\ \iint |u-v|\hat\theta(du, dv) = \inf\left(\iint |u-v|\theta(du,dv) | \theta : \pi_i(\theta) = \theta_i,\ i = 1, 2\right).\end{gathered}$$

Since $\mathscr{P}([0,1])$ is compact in the weak topology and since $(u, v) \to |u-v|$ is continuous and bounded on $[0,1]^2$ such a measure $\hat\theta$ exists for every pair (θ_1, θ_2) with $\theta_i \in \mathscr{P}([0,1])$. In fact we need more, namely that we can find a *measurable map* χ on $(\mathscr{P}([0,1]))^2$ such that $\chi((\theta_1, \theta_2))$ is a coupling of (θ_1, θ_2). This can be proved using the measurable selection theorem (cf. Ethier and Kurtz (1986, Appendix 10)). Denote this map by χ and define

$$\hat{\theta} = \hat{\theta}((\theta_1, \theta_2)) = \chi((\theta_1, \theta_2)) \tag{4.30a}$$

throughout the remainder of this subsection. Then the Wasserstein distance between θ_1 and θ_2 is given by

$$\|\theta_1 - \theta_2\| := \int_{[0,1]^2} |u-v| \chi(\theta_1, \theta_2)(du, dv).$$

The previous construction extends to the case in which $[0,1]$ is replaced by any compact set. However in the case of $[0,1]$ the Wasserstein distance between all laws on $[0,1]$ can be realized simultaneously by random variables, W_θ, with laws, θ, on $[0,1]$ with Lebesgue measure, namely,

$$W_\theta(u) := \inf\{v : \theta([0,v]) \geq u\}, \qquad 0 < u < 1$$

(cf. Dudley (1989, page 333)). Moreover this realization induces a *continuous mapping* χ, from $\mathscr{P}([0,1]) \times \mathscr{P}([0,1])$ to $\mathscr{P}([0,1]^2)$, given by

$$\chi((\theta_1, \theta_2)) := \mathscr{L}((W_{\theta_1}, W_{\theta_2})). \tag{4.30b}$$

Construction of the coupled dynamics. We begin by coupling the evolutions on $(\mathscr{P}([0,1]))^N$. In the sequel we need two types of immigration sources. In the first case we have constant (in time) immigration source; in the second case they are time dependent and are functionals of the process they are driving.

We begin by constructing the coupled process when the two immigration source processes $\tilde{\theta}_1(s), \tilde{\theta}_2(s)$ are *autonomous* (that is, they do not depend on the processes they are driving).

The processes $Y_i^N(t)$, $i = 1, 2$, which we are interested in, solve the martingale problems $(L_{\tilde{\theta}_1}, \nu)$, $(L_{\tilde{\theta}_2}, \mu)$ defined below on $C([0,\infty), (\mathscr{P}([0,1]))^N)$. Here for $i = 1, 2$ and $Y = (y_\xi)_{\xi \in \{0,1,\dots,N-1\}} \in (\mathscr{P}([0,1]))^N$, Q as in (0.28):

$$\begin{aligned} L_{\tilde{\theta}_i} F(Y(s)) = \sum_{\xi=0}^{N-1} \bigg[& c \int_0^1 \frac{\partial F(Y(s))}{\partial y_\xi}(u)(\tilde{\theta}_i(s) - y_\xi(s))(du) \\ & + d \int_0^1 \int_0^1 \frac{\partial^2 F(Y(s))}{\partial y_\xi \partial y_\xi}(u, v) Q_{y_\xi(s)}(du, dv) \bigg], \end{aligned} \tag{4.31}$$

$$F(Y) = \prod_{i=1}^m \int u^{a_i} y_{\xi_i}(du), \qquad \alpha_i \in \mathbb{N},\ \xi_i \in \{0, 1, \dots, N-1\}.$$

The existence of a unique solution to the martingale problems follows from Vaillancourt (1990, Theorem 4.2) or Handa (1990).

Given the measurable $\mathscr{P}([0,1])$-valued processes, $(\tilde{\theta}_i(s))_{s\geq 0}$, $i = 1, 2$, from above, the $\mathscr{P}([0,1]^2)$-valued process $(\hat{\theta}(s))_{s \geq 0}$ is defined via an immigration source given by the coupling (4.30); that is,

$$\hat{\theta}(s) := \chi(\tilde{\theta}_1(s), \tilde{\theta}_2(s)).$$

We then define the *coupled dynamics* $(\overline{Y}(t)) = (\bar{y}_\xi(t))_{\xi \in \{0,1,\dots,N-1\}}$ on the state space $(\mathscr{P}([0,1]^2))^N$ to be a solution of the martingale problem $(L_{\hat{\theta}}, \nu \otimes \mu)$:

$(u = (u_1, u_2), v = (v_1, v_2) \in [0, 1]^2)$, with
(4.32)

$$L_{\hat{\theta}} F(\overline{Y}(s)) = \sum_{\xi=0}^{N-1} \left[c \int_{[0,1]^2} \left(\frac{\partial F(\overline{Y}(s))}{\partial \bar{y}_\xi}(u) \right) (\hat{\theta}(s) - \bar{y}_\xi(s))(du) \right.$$

$$\left. + d \int_{[0,1]^2} \int_{[0,1]^2} \left(\frac{\partial^2 F(\overline{Y}(s))}{\partial \bar{y}_\xi \partial \bar{y}_\xi}(u, v) \right) \widetilde{Q}_{\bar{y}_\xi(s)}(du, dv) \right]$$

and

$$F(\overline{Y}) = \prod_{i=1}^{k} \int u_1^{m_i} u_2^{n_i} \bar{y}_{\xi_i}(du_1, du_2), \qquad m_i, n_i \in \mathbb{N},$$

$$\widetilde{Q}_{\bar{y}_\xi}(du, dv) = \bar{y}_\xi(du_1, du_2)\delta_{(u_1, u_2)}(dv_1, dv_2) - \bar{y}_\xi(du_1, du_2)\bar{y}_\xi(dv_1, dv_2)$$

with $\bar{y} \in \mathscr{P}([0, 1]^2)$ and $\overline{Y}(t) = (\bar{y}_\xi(t))_{\xi \in \Omega_N}$. □

The existence of a solution of the above martingale problem follows by viewing it as a Fleming-Viot process with time inhomogeneous immigration on $\mathscr{P}([0, 1]^2)$ and following the existence argument of Dawson and March (1993, sections 4 and 5). Although we do not require the uniqueness, it can be proved using the time inhomogeneous dual process as in Dawson (1993, section 5.7).

Now we turn to the nonautonomous case in which each process $\tilde{\theta}_i(s) = \tilde{\theta}_i(s, Y_i^N(s))$, $i = 1, 2$, is either autonomous (i.e. does not depend on $Y_i^N(s)$) or is of the form

$$\tilde{\theta}_i(s, Y_i^N(s)) = \theta(Y_i^N(s)) \tag{4.33}$$

with $\theta(Y^N)$ defined according to

$$\langle \theta(Y^N), f \rangle = \left\langle \frac{1}{N} \sum_{i=0}^{N-1} y_i, f \right\rangle \quad \forall f \in C([0, 1])$$

and where $Y_1^N(s)$, $Y_2^N(s)$ are processes on $(\mathscr{P}([0, 1]))^N$.

We now consider the corresponding coupled martingale problem for a process $(\overline{Y}(t))$ with state space $(\mathscr{P}([0, 1]^2))^N$ associated to the operator of the form (4.32) but in which

$$\hat{\theta}(s) := \chi(\tilde{\theta}_1(s), \tilde{\theta}_2(s))$$

with $\tilde{\theta}_i$ now either autonomous or given by (4.33) but with Y_i^N replaced by $\overline{Y}_i := \pi_i \circ \overline{Y}^N$. We can then obtain the existence of a solution to the corresponding coupled martingale problem (4.32) by following the method of "local freezing" as in Dawson and March (1993, section 5). To do this we construct a sequence of approximating solutions $\overline{Y}^{N,M}(t)$, $M \in \mathbb{N}$, $t \in [0, T]$. The process $\overline{Y}^{M,N}$ is obtained by first solving the martingale problem of the form (4.32) on the time interval $[0, T/M)$, $M \in \mathbb{N}$, using the constant coupled immigration source $\chi(\tilde{\theta}_1(0, \overline{Y}_1^N(0)), \tilde{\theta}_2(0, \overline{Y}_2^N(0)))$ which can be done as above since these are autonomous in this interval. The resulting fixed coupled solution at time T/M, $\chi(\tilde{\theta}_1(T/M, \overline{Y}_1^{N,M}(T/M)), \tilde{\theta}_2(T/M, \overline{Y}_2^{N,M}(T/M)))$, is then used as the coupled immigration source on the interval $[T/M, 2T/M)$ and so on. Taking the limit as $M \to \infty$, noting that the mapping $\overline{Y} \to \chi(\tilde{\theta}(\pi_1 \circ \overline{Y}), \tilde{\theta}(\pi_2 \circ \overline{Y}))$ can be chosen to be uniformly continuous (recall (4.30b)), and using a compactness argument as in Dawson and March (1993, section 5), we obtain a solution to the coupled martingale problem associated with (4.32), (4.33).

Remark. The construction can be extended to $(\mathscr{P}([0,1]))^{\mathbb{Z}^+}$ by using the method of section 1. Furthermore, Lemmas 4.6 and 4.7 below will still hold. We do not elaborate on the straightforward details here.

Lemma 4.6. *The "marginal" processes $\overline{Y}_i(t)$, $i = 1, 2$, associated to the solution, $\overline{Y}(t)$, of the coupled martingale problem* (4.32) *satisfy*

$$\mathscr{L}((\overline{Y}_i(s))_{s\geq 0}) = \mathscr{L}((Y_i(s))_{s\geq 0}), \qquad i = 1, 2, \tag{4.34}$$

where $Y_i(t)$ is the unique solution to the martingale problem (4.31) *in which $\tilde{\theta}_i(s)$ is either autonomous or of the form* (4.33).

Proof. Note that $\widetilde{Q}_{\bar{y}}(du, [0,1]) = Q_{\overline{Y}_1}(du)$. Then if in (4.32) we consider only the functions F corresponding to the case $n_i = 0$, then $L_{\hat{\theta}}F(\overline{Y}(s)) = L_{\tilde{\theta}_1}F(Y_1)(s)$. This implies that $\pi_1 \circ \overline{Y}$ solves the martingale problem $L_{\tilde{\theta}_1}$. The result then follows from the uniqueness of the solutions to the martingale problem (4.31). (The uniqueness for either choice of $\tilde{\theta}_i(s)$ follows by duality.)

Remark. Note that for $\theta_1 = \theta_2$ the measure $\hat{\theta}$ is concentrated on the diagonal. The next objective is to investigate how far $\bar{y}_\xi(t)$ is from a measure concentrated on the diagonal. If we choose ν and μ as *N-exchangeable* measures on $(\mathscr{P}([0,1]))^N$, then the law of $\overline{Y}(t)$ is N-exchangeable for every $t \geq 0$ and it suffices to study the associated empirical measures:

$$E^N(s) := \frac{1}{N}\sum_{\xi=0}^{N-1} \delta_{y_\xi}(s\beta(N)) \in \mathscr{P}(\mathscr{P}([0,1])),$$

and

$$E^\infty(s) := \Gamma^0_{\theta_s} \in \mathscr{P}(\mathscr{P}([0,1])).$$

In order to obtain a measure of how far $\bar{y}_\xi(t)$ is from a measure concentrated on the diagonal we define

$$\begin{aligned} h(t) &= E\left[\frac{1}{N}\sum_{\xi=0}^{N-1}\int_{[0,1]^2} |u_1 - u_2|\bar{y}_\xi(t)(du_1, du_2)\right], \\ g(t) &= E\left[\int_{[0,1]^2} |u_1 - u_2|\hat{\theta}(t)(du_1, du_2)\right]. \end{aligned} \tag{4.35}$$

Lemma 4.7. *Under the assumptions described above*:

$$h(t) = h(0)e^{-t} + \int_0^t e^{(s-t)} g(s)\, ds. \quad \square \tag{4.36}$$

Proof. Define the function H_ξ on $\mathscr{P}([0,1]^2)$ for fixed $\xi \in \{0, \ldots, N-1\}$ as follows:

$$H_\xi(\bar{y}) = \int_{[0,1]^2} |u_1 - u_2|\bar{y}_\xi(du_1, du_2). \tag{4.37}$$

Then

$$\frac{\partial H}{\partial \bar{y}_\xi}(\bar{y})(u) = |u_1 - u_2|. \tag{4.38}$$

Now use (4.32) for $F = H_\xi$ to get the differential equation (recall (4.35)):

$$\dot{h}(t) = g(t) - h(t), \tag{4.39}$$

which has the unique solution given in (4.36) for every initial point.

Now we are ready to prove the two lemmas still open in section 4(b).

Proof of Lemma 4.4. We shall apply our coupling results. For each N we set in (4.31)–(4.32) $\tilde{\theta}_1(t) = \theta^N_{s\beta(N)-L(N)}$, $\nu = \tilde{\mu}_N$ and $\tilde{\theta}_2(t) = \theta_2(Y^N_{s\beta(N)-L(N)+t})$, $\mu = \tilde{\mu}_N$. Then we get by (4.32) a coupled dynamics with marginals $Y^\infty_N(t)$ and $Y^N(t)$. Denote by $g_N(t)$, $h_N(t)$ the quantities introduced in (4.35) for this process. Then the assumption (4.27) gives in (4.35) that for $N \geq N_0(\varepsilon)$

$$g_N(t) \leq \varepsilon \quad \forall t \leq L(N). \tag{4.40}$$

Hence according to (4.36)

$$h_N(L(N)) \leq h_N(0)e^{-L(N)} + \varepsilon \quad \forall N \geq N_0(\varepsilon) \tag{4.41}$$

so that (using $L(N) \to \infty$ as $N \to \infty$)

$$\overline{\lim_{N\to\infty}}\, h_N(L(N)) \leq \varepsilon. \tag{4.42}$$

Since ε is arbitrary we can obtain the final result that for all fixed ξ and functions $f \in C(\mathscr{P}([0, 1]))$, which are Lipschitz as well,

$$|E(f(y^\infty_{N,\xi}(L(N))) - E(f(y^N_\xi(s\beta(N))))| \underset{N\to\infty}{\longrightarrow} 0.$$

Then the assertion follows by a standard argument. (Use the fact that Lipschitz functions on $\mathscr{P}([0, 1])$ are dense in $C(\mathscr{P}([0, 1]))$; see proof of Lemma 4.2.)

Proof of Lemma 4.5. Again apply the coupling result with $\tilde{\theta}_1(t) \equiv \theta^N$, $\tilde{\theta}_2(t) \equiv \theta$, $\nu = \mu_N$, $\mu = \mu$ in (4.32). Compare the remark following (4.33). Then we get from the fact that θ^N, θ are constant in time and $\mathscr{L}(\theta^N) \Rightarrow \mathscr{L}(\theta)$ (the latter follows as in Lemma 4.3), the relations:

$$\begin{aligned} g_N(t) &= g_N(0) \quad \forall t \in \mathbb{R}^+, \\ g_N(0) &\underset{N\to\infty}{\Rightarrow} 0. \end{aligned} \tag{4.43}$$

As in (4.40)–(4.42) the proof above this yields the assertion.

(d) **Convergence of the density measure process: Proof of Proposition 4.2.** The proof is based on Proposition 4.1. The heuristic idea is the following. Since the density changes only slowly in time, the components relax before that into the unique equilibrium (spatially ergodic) corresponding to the fixed density. Hence the density process becomes Markov in the limit and its mean square displacement can be obtained by averaging the mean square displacement of a single component with respect to the equilibrium corresponding to the actual value of the density.

The proof proceeds as usual with such statements, in three steps. We first establish tightness; then we derive a martingale problem which any weak limit points must satisfy and finally show that this martingale problem has a unique solution on $C([0, \infty), \mathscr{P}([0, 1]))$.

We start by formulating the key results of each of these three steps in three lemmas which imply Proposition 4.2 and will be proved one by one. Recall the definition of $E^N(s)$ above preceding (4.35).

Lemma 4.8. (a) *For* $f \in C([0, 1])$, *the sequence*

$$\left\{\mathscr{L}\left(\left(\int_0^s \frac{1}{N}\sum_\xi \operatorname{Var}_{y_\xi(u\beta(N))}(f)\, du, \langle\theta^N_{s\beta(N)}, f\rangle\right)_{s\in\mathbb{R}^+}\right)\right\}_{N\in\mathbb{N}}$$

is tight in $C([0, \infty), \mathbb{R}^2)$.

(b) *The sequence* $\{\mathscr{L}((\theta^N_{s\beta(N)})_{s\in\mathbb{R}^+})\}_{N\in\mathbb{N}}$ *is tight in* $C([0, \infty), \mathscr{P}([0, 1]))$.

(c) *The sequence* $\{\mathscr{L}(\int_0^t(E^N(s)\, ds)_{t\in\mathbb{R}^+})\}_{N\in\mathbb{N}}$ *is tight in* $C([0, \infty)$, $\mathscr{P}(\mathscr{P}([0, 1])))$.

Lemma 4.9. *Suppose that* $(N_k)_{k\in\mathbb{N}}$ *is a subsequence such that*

$$\mathscr{L}((\theta^{N_k}_{s\beta(N_k)})_{s\in\mathbb{R}^+}) \underset{k\to\infty}{\Rightarrow} \mathscr{L}((\theta_s)_{s\in\mathbb{R}^+}). \tag{4.44}$$

Then for every $f \in C([0, 1])$:

(i) $(\langle\theta_s, f\rangle)_{s\in\mathbb{R}^+}$ *is a continuous, square integrable martingale.*

(ii) $(\langle\theta_s, f\rangle^2 - \int_0^s \Phi_f(\theta_t)dt)_{t\geq 0}$ *is a martingale, where*

$$\Phi_f(\theta) = \frac{cd}{c+d}2\int_{[0,1]}\int_{[0,1]} f(u)f(v)Q_\theta(du, dv) = 2\frac{cd}{c+d}\operatorname{Var}_\theta(f)$$

and

$$Q_\theta(du, dv) = \theta(du)\delta_u(dv) - \theta(du)\theta(dv). \quad \square$$

Lemma 4.10. (a) *The following martingale problem on* $C([0, \infty) : \mathscr{P}([0, 1]))$ *has a unique solution* (*weak*) (*with the notation of the previous lemma*). *For any* $f \in C([0, 1])$

(i) $(\langle\theta_s, f\rangle)_{s\in\mathbb{R}^+}$ *is a martingale with continuous paths, and* $(\langle\theta, f\rangle^2 - \int_0^s \Phi_f(\theta_t)\, dt)_{s\geq 0}$ *is a martingale.*

(ii) *The solution of the above martingale problem is the Fleming-Viot process on* $\mathscr{P}([0, 1])$ *defined in* (4.4). $\square$

The combination of Lemmas (4.8)(b), (4.9) and (4.10) proves Proposition 4.2.

The crucial property in proving the results in the above lemmas are the following facts (1) and (2) which allow the reduction to real valued processes (diffusions).

(1) Let A_t be the process on $\mathscr{P}([0, 1])$ solving uniquely the martingale problem (L, δ_θ) with L defined on functions F of the form

$$F_{(A)} = \prod_{i=1}^m \int_0^1 u^{a_i} A(du), \quad A \in \mathscr{P}([0, 1]), \ m \in \mathbb{N},$$

by

$$L(F)_{(A)} = \int_0^1\int_0^1 \frac{\partial F(A)}{\partial A \partial A}(u, v)Q_A(du, dv), \qquad A \in \mathscr{P}([0, 1]), \tag{4.45}$$
$$Q_A(du, dv) = A(du)\delta_u(dv) - A(du)A(dv).$$

Then for every $f \in C([0, 1])$ consider

$$F_f^2(A) = \left(\int f(u)A(du)\right)^2, \quad F_f^1(A) = \int f(u)A(du).$$

One calculates, for every $f \in C([0, 1])$,

$$\frac{\partial F_f^1}{\partial A \partial A} \equiv 0, \quad \frac{\partial F_f^2}{\partial A \partial A}(u, v) = 2f(u)f(v). \tag{4.46}$$

Hence (4.45) implies

(4.47) $\quad (\langle A_t, f\rangle)_{t\geq 0}$ is a martingale with continuous paths;

(4.48) $\quad \left((\langle A_t, f\rangle^2) - \int_0^t \Phi_f(A_s)\, ds\right)_{t\geq 0}$ is a martingale,

where

$$\Phi_f(A) = 2\int\int f(u)f(v)Q_A(du, dv).$$

(2) Let $X, X_N \in C([0, \infty) : \mathscr{P}([0, 1]))$. Then

(4.49a) $\quad X_n \to X$ as $n \to \infty$ iff $(\langle X_n(s), f\rangle)_{s\geq 0} \to (\langle X(s), f\rangle)_{s\geq 0}$ in the topology of $C([0, \infty) : \mathbb{R})$ for every $f \in C([0, 1])$.

Similarly, if $X, X_N \in C([0, \infty) : \mathscr{P}(\mathscr{P}([0, 1])))$, then

(4.49b) $\quad X_n \to X$ as $n \to \infty$ iff for every integer k and $f_1, \dots, f_k \in C([0, 1])$

$$\left(\left\langle X_n(s), \prod_{i=1}^k \langle \mu, f_i\rangle\right\rangle\right)_{s\geq 0} \to \left(\left\langle X(s), \prod_{i=1}^k \langle \mu, f_i\rangle\right\rangle\right)_{s\geq 0}$$

in the topology of $C([0, \infty) : \mathbb{R})$.

Both facts (1) and (2) together allow us to "reduce" the problem to real-valued processes, which are square integrable martingales and related questions were treated in Dawson and Greven (1993c).

We proceed now by successively proving Lemmas 4.8–4.10.

Proof of Lemma 4.8. Using (4.49) the proof of (c) is straightforward, and (b) follows immediately from (a) which we now verify.

For $\theta \in \mathscr{P}([0, 1])$, $f \in C([0, 1])$ define:

$$F_f^1(\theta) = \langle \theta, f\rangle, \quad F_f^2(\theta) = (\langle \theta, f\rangle)^2. \tag{4.50}$$

By the definition of (θ_s^N) and (4.1) we obtain the following expression for the generator (which is denoted by G_N) of θ_s^N applied to F_f^i, $i = 1, 2$ (recall (4.1)):

$$G_N F_f^1(\theta_s^N) = \frac{1}{N}\sum_{\xi=0}^{N-1} \widetilde{L}_N(F_f^1)_{(y_\xi)} = 0 \tag{4.51}$$

where we used (4.45). With this result we calculate
(4.52)

$$G_N F_f^2(\theta_s^N) = \frac{1}{N^2}\sum_{\xi=0}^{N-1} \widetilde{L}_N(F_f^2)_{(y_\xi)}$$
$$= \frac{1}{N^2}\sum_{\xi=0}^{N-1} d(\langle y_\xi, f^2\rangle - (\langle y_\xi, f\rangle)^2) = \frac{1}{N}\left(\frac{1}{N}\sum_\xi d\cdot \mathrm{Var}_{y_\xi}(f)\right).$$

Hence we have

$$NG_N(F^1(\theta_s^N)) = 0, \quad NG_N(F^2(\theta_s^N)) = \frac{1}{N}\sum_\xi d\cdot \mathrm{Var}_{y_\xi}(f). \tag{4.53}$$

Therefore $(\langle \theta_{s\beta(N)}^N, f\rangle)_{s\geq 0}$ is a square integrable martingale with increasing process given by

$$\left(\int_0^s \frac{1}{N}\sum_\xi d\cdot \mathrm{Var}_{y_\xi(u\beta(N))}(f)\,du\right)_{s\geq 0}. \tag{4.54}$$

This sequence of increasing processes is uniformly equicontinuous and hence relatively compact in $C([0,\infty))$ by Arzela-Ascoli. This implies the tightness of

$$\left(\int_0^s \frac{1}{N}\sum_\xi \mathrm{Var}_{y_\xi(u\beta(N))}(f)\,du, (\langle \theta_{s\beta(N)}^N, f\rangle)\right)_{s\in\mathbb{R}^+}$$

via the representation (4.55) below. Namely the relations (4.52) and (4.53) imply that we can represent

$$\langle \theta_{s\beta(N)}^N, f\rangle = w\left(2\int_0^s \frac{1}{N}\sum_\xi \mathrm{Var}_{y_\xi(u\beta(N))}(f)\,du\right) \tag{4.55}$$

where $w(\cdot)$ is a suitable Brownian motion.

Proof of Lemma 4.9. It suffices to prove for a subsequence $\{N_k\}$ satisfying (4.44) and $f\in C([0,1])$ that

$$\left(\int_0^s \frac{1}{N_k}\sum_\xi \mathrm{Var}_{y_\xi(u\beta(N_k))}(f)\,du, (\theta_{s\beta(N_k)}^{N_k})\right)_{s\geq 0} \underset{k\to\infty}{\Rightarrow} \left(\int_0^s E^{\Gamma_{\theta_u}}(\mathrm{Var}_{y_0}(f))\,du, (\theta_s)\right)_{s\geq 0}. \tag{4.56}$$

Tightness of the pair has been proved in Lemma 4.8. Moreover Proposition 4.1 proves that

$$\mathscr{L}\left(\frac{1}{N_k}\sum_\xi \mathrm{Var}_{y_\xi(u\beta(N_k))}(f)\right) \Rightarrow \mathscr{L}(E^{\Gamma_{\theta_u}}(\mathrm{Var}_{y_0}(f)));$$

that is, the one-dimensional marginal distributions of the integrand in the first component converge to the appropriate limit. Once we have established (4.56) it will remain to show that ($x \in \mathscr{P}([0, 1])$ is the integration variable)

$$E_{\Gamma_\theta}(\mathrm{Var}_x(f)) = \frac{cd}{c+d}\mathrm{Var}_\theta(f). \tag{4.57}$$

This will be done in section 6; see equation (6.8).

Hence it remains to verify (4.56). In order to do this we could follow Dawson and Greven (1993b, section 3). There is however another approach pointed out by T. Shiga and worked out in Cox, Greven, and Shiga (1993) for lattice models. Roughly speaking, the idea is to prove that the process of empirical measures $(\int_0^s E^N(u)\,du)_{s\geq 0}$ converges to the process $(\int_0^s \Gamma^0_{\theta_u}\,du)_{s\geq 0}$ and then to use the fact that the increasing process of $(\theta_s^N)_{s\geq 0}$ is a continuous functional of the process of $(\int_0^s E^N(u)\,du)_{s\geq 0}$. Formally, we develop this approach in two steps.

Step 1. Define the empirical measures

$$E^N(s) := \frac{1}{N}\sum_{\xi=0}^{N-1}\delta_{y_\xi(s\beta(N))} \in \mathscr{P}(\mathscr{P}([0, 1])), \tag{4.58}$$

and

$$E^\infty(s) := \Gamma^0_{\theta_s} \in \mathscr{P}(\mathscr{P}([0, 1])), \tag{4.59}$$

and note that, for $f \in C([0, 1])$, we have the representation

$$\begin{aligned} \langle \theta^N_{s\beta(N)}, f\rangle &= \int_{\mathscr{P}([0,1])}\langle \nu, f\rangle E^N(s)(d\nu), \\ \frac{1}{N}\sum_{\xi=0}^{N-1}\mathrm{Var}_{y_\xi(s\beta(N))}(f) &= \int_{\mathscr{P}([0,1])}[\langle \nu, f^2\rangle - \langle \nu, f\rangle^2]E^N(s)(d\nu). \end{aligned} \tag{4.60}$$

Furthermore for $\Phi \in \mathscr{P}(\mathscr{P}([0, 1]))$, $(\Phi_s)_{s\geq 0} \in C([0, \infty), \mathscr{P}(\mathscr{P}([0, 1])))$, $g \in C([0, 1])$ and $T \in \mathbb{R}^+$ the maps

$$\Phi \to \int \langle \nu, g\rangle \Phi(d\nu), \text{ and } \Phi \to \int_0^T \int [\langle \nu, g^2\rangle - \langle \nu, g\rangle^2]\Phi(d\nu)\,du \tag{4.61}$$

are continuous.

Hence by combining (4.60) with (4.61), the relations (4.56) will be implied by the (stronger) result

$$\mathscr{L}\left(\left(\int_0^s E^{N_k}(u)\,du, \theta^{N_k}_{s\beta(N_k)}\right)_{s\geq 0}\right) \underset{k\to\infty}{\Rightarrow} \mathscr{L}\left(\left(\int_0^s E^\infty(u)\,du, \theta_s\right)_{s\geq 0}\right). \tag{4.62}$$

Step 2. The verification of (4.62) is achieved with coupling techniques. Since tightness of the sequence $(\int_0^s E^N(u)\,du)_{s\geq 0}$ is known from Lemma 4.8, it suffices therefore to identify the finite-dimensional distributions. Suppose that for the given subsequence $\{N_k\}$ we could define $(E^{N_k}(s))_{s\geq 0}$ and $(E^\infty(s))_{s\geq 0}$ on one probability space such that for every $F \in C(\mathscr{P}(\mathscr{P}([0, 1])))$, $s \in [0, \infty)$,

$$E|F(E^{N_k}(s)) - F(E^\infty(s))| \underset{k\to\infty}{\to} 0; \tag{4.63}$$

then we would be done. In fact, it is enough to verify (4.63) for a dense subset of $C(\mathscr{P}([0, 1]))$, in particular the set of Lipschitz functions. The latter is now achieved by bringing into play the coupling results of section 4(c).

Define a probability space on which $(\{(\theta^{N_k}_{s\beta(N_k)})_{s\geq 0}\}, (\theta_s)_{s\geq 0})$ are defined and $\theta^{N_k}_{s\beta(N_k)}$ converges a.s. to θ_s as elements of $C([0, \infty), \mathscr{P}([0, 1]))$. (Recall (4.44).) Now in order to check (4.63) for some fixed $s \in \mathbb{R}^+$ we introduce a comparison process $(\widetilde{Y}^N(t))_{t\geq 0}$. To construct this process we need a sequence $T(N)$ increasing to infinity as $N \to \infty$ and $T(N) = o(N)$. Then let $\widetilde{Y}^N(t)$ be the following modification of $Y^N(t)$. In the time interval $[s\beta(N) - T(N), s\beta(N)]$ we observe first $Y^N(s\beta(N) - T(N))$ and then we replace the process in this time interval by the solution to the martingale problem $(L_{\tilde{\theta}(N)}, \mu)$ (see (4.32)), where μ is the δ-measure on $Y^N(s\beta(N) - T(N))$ and

$$\tilde{\theta}(N)_t := \theta^N_{s\beta(N)-T(N)}, \quad \text{for } s\beta(N) - T(N) \leq t < s\beta(N). \tag{4.64}$$

Define $\widetilde{E}^N(s)$ by the analogue of (4.58) for $\widetilde{Y}^N(t)$. Using the existence of the coupled dynamics established above we can in fact construct $Y^N(t)$ and $\widetilde{Y}^N(t)$ on one probability space. The idea is now to compare the pair $E^N(s)$, $E^\infty(s)$ by comparing the pairs $\widetilde{E}^N(s)$, $E^N(s)$ and $\widetilde{E}^N(s)$, $E^\infty(s)$. To execute this comparison we first observe that since on our basic probability space

$$(\theta^{N_k}_{s\beta(N_k)})_{s\geq 0} \underset{k\to\infty}{\Rightarrow} (\theta_s)_{s\geq 0} \quad \text{in } C([0, \infty), \mathscr{P}([0, 1])), \quad \text{a.s.}, \tag{4.65}$$

we know that

$$\sup_{t\in I(N_k)} \|\theta^{N_k}_t - \tilde{\theta}(N_k)_t\| \underset{k\to\infty}{\to} 0, \quad \text{a.s.}, \tag{4.66}$$

where $I(N_k) = [s\beta(N_k) - T(N_k), s\beta(N_k))$. Let $\{\bar{y}_\xi(t)\}_{\xi\in\{0,\dots,N_k-1\}}$ denote the coupled version of $Y^{N_k}(t)$, $\widetilde{Y}^{N_k}(t)$ constructed with generator as in (4.32) but with $\hat{\theta}(t) = \chi(\theta^{N_k}_t, \tilde{\theta}(N_k)_t)$, $t \in I(N_k)$. Then by our coupling result (Lemma 4.7) we conclude with (4.66) that for every Lipschitz function F on $\mathscr{P}(\mathscr{P}([0, 1]))$:

$$\begin{aligned} &E|F(\widetilde{E}^{N_k}(s)) - F(E^{N_k}(s))| \\ &\quad \leq \text{const}\cdot E\left[\frac{1}{N_k}\sum_{\xi=0}^{N_k-1}\int_{[0,1]^2} |u_1 - u_2| \bar{y}_\xi(s\beta(N_k))(du_1, du_2)\right] \\ &\quad \leq \text{const}\cdot \sup_{t\in I(N_k)} E\|\theta^{N_k}_t - \tilde{\theta}(N_k)_t\| \underset{k\to\infty}{\to} 0. \end{aligned} \tag{4.67}$$

On the other hand since $\Gamma^0_{\tilde{\theta}(N)}(\cdot)$ is the equilibrium of the process defined by the martingale problem $(L_{\tilde{\theta}(N)}, \mu)$, and since $\Gamma^0_{\tilde{\theta}(N_k)} \Rightarrow \Gamma^0_{\theta_s}$ by (4.65) and Lemma 4.2, we can prove with the same coupling techniques that for every Lipschitz function F on $\mathscr{P}(\mathscr{P}([0, 1]))$

$$E|F(\widetilde{E}^{N_k}(s)) - F(E^\infty(s))| \underset{k\to\infty}{\to} 0. \tag{4.68}$$

Finally, note that (4.67) and (4.68) can be extended to all $F \in \mathscr{P}(\mathscr{P}([0, 1]))$ (using the fact that the Lipschitz functions are dense). This shows that, conditioned on θ, $E^{N_k}(s)$ converges in probability to $\Gamma^0_{\theta_s}$. Therefore for every

$f \in C(\mathscr{P}([0,1]))$, $\int f(\mu) E^{N_k}(s, d\mu) \to \int f(\mu) E^{\infty}(s, d\mu)$ in probability as $k \to \infty$ and

$$E\left|\int f(\mu) \int_0^t E^{N_k}(s, d\mu)\, ds - \int f(\mu) \int_0^t E^{\infty}(s, d\mu)\, ds\right| \underset{k\to\infty}{\longrightarrow} 0.$$

This yields (4.62) and therefore (4.56). This completes the proof of Lemma 4.9.

Proof of Lemma 4.10. There exists a unique solution to the martingale problem (i), namely the Fleming-Viot process on $\mathscr{P}([0,1])$ (see (4.44)) which is time changed according to $t \to \frac{cd}{(c+d)} t$ (cf. Dawson (1993, Theorem 5.7.1)). This completes the proof of Lemma 4.10.

(e) **Extension.** At this stage we would like to point out that a generalization of the results of this section can be obtained by a slight modification of the arguments above in order to apply them to systems with any finite number of levels in section 5.

Let f denote a bounded continuous function on $(\mathscr{P}([0,1]))^N$ and $B(N)$ a sequence of bounded operators on $C([0,1]^2)$. Consider the operator R_N defined as follows:

$$R_N(F)(y) = \sum_{\xi=0}^{N-1} \int_0^1 \int_0^1 \left[B(N)\left(\frac{\partial^2 F(y)}{\partial y_\xi \partial y_\xi}\right)\right](u, v) Q_{y_\xi}(du, dv), \tag{4.69}$$

$$y = (y_1, \dots, y_N).$$

Consider now the perturbed system defined by replacing the generator in (4.1) by

$$\tilde{\tilde{L}}_n := \tilde{L}_N + R_N. \tag{4.70}$$

Denote the corresponding system by $(\tilde{Y}^N(t))$.

Corollary. *Assume that* $\|B(N)\| \to 0$ *as* $N \to \infty$. *Then the assertions of Propositions* 4.1 *and* 4.2 *still hold for the perturbed system* $(\tilde{Y}^N(t))$.

5. The multiple time scale behavior of infinite interacting systems in the mean-field limit

In order to derive the multiple time scale behavior (Theorem 0.2) for the infinite hierarchy of Fleming-Viot processes we proceed in three steps. We first use the results from the previous section to treat the *finite model* with *two levels.* In a second major step we treat *finite systems* with an *arbitrary number of hierarchical levels* and thirdly we build on this a treatment of the *infinite hierarchy.* This program is carried out through a sequence of three propositions, namely, Propositions 5.1, 5.2, 5.3. Furthermore in subsection 5(d) we derive in Proposition 5.4 the tools to prove Theorem 0.4. In section 5(e) we finally complete the proof of Theorems 0.2, 0.4 and in 5(f) prove Theorem 0.10.

(a) **Finite two level models.** We define now a process $(Y^{N,2}(t))_{t\geq 0}$, the two level finite system, on the state-space $(\mathscr{P}([0,1]))^{M(2,N)}$, $M(2,N) = \{0, \dots, N-1\}^2$ via the martingale problem $(\tilde{L}_{N,2}, \mu_N)$ defined as follows.

We introduce the notation $y = (y_\xi)_{\xi \in M(2,N)}$, $\xi = (\xi_1, \xi_2)$, and $\xi^0 := (0, 0)$, $\xi[k] := \{\xi' \in M(2, N) | d(\xi, \xi') \le k\}$ and $y_{\xi,k} := \frac{1}{N^k} \sum_{\xi' \in \xi[k]} y_{\xi'}$.

(5.1)

$$\widetilde{L}_{N,2}(F)(y) = \sum_{\xi \in M(2,N)} \left[\left(\int_0^1 \frac{\partial F(y)}{\partial y_\xi}(u) \left(\sum_{k=1}^{2} \frac{c_{k-1}}{N^{k-1}} (y_{\xi,k}(du) - y_\xi(du)) \right) \right) + d_0 \int_0^1 \int_0^1 \frac{\partial^2 F(y)}{\partial y_\xi \partial y_\xi}(u, v) Q_{y_\xi}(du, dv) \right]$$

with $Q_{y_\xi}(du, dv) = y_\xi(du)\delta_u(dv) - y_\xi(du)y_\xi(dv)$, and

$$F(y) = \prod_{i=1}^{n} \int_0^1 f_{\xi(i)}(u) y_{\xi(i)}(du),$$

where $f_{\xi(i)} \in C([0, 1])$, $\xi(1), \dots, \xi(n) \in M(2, N)$.

The existence and uniqueness of that martingale problem is an immediate consequence of Theorem 0.0, since it can be viewed as a special case where $c_3 = c_4 = \cdots = 0$. If $\xi \in M(2, N)$, then

$$y^N_{\xi,1} = \frac{1}{N} \sum_{\xi_1=0}^{N-1} y_{(\xi_1,\xi_2)} \quad \text{and} \quad y^N_{\xi,2} = \frac{1}{N} \sum_{\xi_2=0}^{N-1} y^N_{\xi,1}. \tag{5.2}$$

Recall that (ξ_1, ξ_2) identifies the ξ_1th individual in the ξ_2th family and therefore $y^N_{\xi,1}$ denotes the average of the ξ_2th family. Finally recall the processes $Z^{\theta,k}_t$, the measures $\mu^{k,j}_\theta$ and the numbers $(d_k)_{k\in\mathbb{N}}$ introduced in section 0(d).

Now we are able to formulate the result for the two level model, which is the extension of (4.6) in the last section to two levels. Recall the definition of the stationary process $Z^{\theta,k}_t$ in (0.d).

Proposition 5.1. *Assume that the initial measure μ_N satisfies*

$$\mu_N = \bigotimes_{\xi \in M(2,N)} \mu_\xi \quad \textit{with} \int_{\mathscr{P}([0,1])} y\mu_\xi(dy) = \theta.$$

Then

$$\mathscr{L}(y^N_{\xi,2}(t)_{t\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \delta_{\{Y_t \equiv \theta\}}, \tag{5.3}$$

$$\mathscr{L}((y^N_{\xi,2}(sN^2))_{s\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((\widetilde{Z}_s)_{s\in\mathbb{R}^+}) \tag{5.4}$$

where $\widetilde{Z}_t$ is the solution of the martingale problem (L, δ_θ) with

$$L(F)_{(z)} = d_2 \int_0^1 \int_0^1 \frac{\partial^2 F(z)}{\partial z \partial z}(u, v) Q_z(du, dv).$$

$$\mathscr{L}((y^N_{\xi,1}(sN^2 + tN))_{t\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((Z^{\theta^*_1(s),1}(t))_{t\in\mathbb{R}^+}), \tag{5.5}$$

$$\mathscr{L}((y^N_{\xi,0}(sN^2 + t))_{t\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((Z^{\theta^*_0(s),0}(t))_{t\in\mathbb{R}^+}). \tag{5.6}$$

In (5.5)–(5.6), *conditioned on $\theta^*_k(s) = \theta$, $\{Z^{\theta,k}(t)\}_{t\in\mathbb{R}^+}$ is a stationary Fleming-Viot process with immigration source θ (as defined in section* (0.d)) *independent* *of* $\{\theta_k^*(s')\}_{s'\in\mathbb{R}^+}$. *In addition* $\mathscr{L}(\theta_1^*(s)) = \nu_s$, $\mathscr{L}(\theta_0^*(s)) = \int \mu_\rho^{1,1}\nu_s(d\rho)$, $\nu_s = \mathscr{L}(\widetilde{Z}_s)$ *and* $\mu_\theta^{1,1}$ *is as defined in section* (0.d). □

Remark. The statements (5.4)–(5.6) remain valid if on the right-hand side we replace sN^2 by $s\beta_2(N)$ where $\lim_{N\to\infty}\beta_2(N)/N^2 = 1$. We shall write $\beta_2(N)$ for such a function and $\beta_1(N)$ for N in this subsection.

Proof of (5.3). the analogous argument to that in (4.49)–(4.53) yields that for every $f \in C([0, 1])$:

(5.7)
$(\langle y_{\xi,2}^N(t\beta_1(N)), f\rangle)_{t\in\mathbb{R}^+}$ is a continuous martingale with increasing process

$$\frac{1}{N}\int_0^s \left[\frac{1}{N^2}\sum_{\zeta\in M(2,N)} \mathrm{Var}_{y_\zeta(u\beta_1(N))}(f)\right] du. \tag{5.8}$$

Since by assumption on μ_N and by (5.7)

$$\mathscr{L}(\langle y_{\xi,2}^N(t\beta_1(N)), f\rangle) \underset{N\to\infty}{\Rightarrow} \delta_{\langle\theta, f\rangle}$$

and since furthermore the expression in (5.8) converges to 0 uniformly in $s \in [0, T]$ for every $T > 0$, we know that for all $f \in C([0, 1])$ (recall the representation (4.55)):

$$\mathscr{L}((\langle y_{\xi,2}^N(t\beta_1(N)), f\rangle)_{t\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \delta_{\{y(t)\equiv\langle\theta, f\rangle\}}. \tag{5.9}$$

This implies (5.3).

Proof of (5.4). As above we have that $(\langle y_{\xi,2}^N(t\beta_2(N)), f\rangle)_{t\in\mathbb{R}^+}$ is a martingale with increasing process

$$\int_0^s \left[\frac{1}{N^2}\sum_{\zeta\in M(2,N)} \mathrm{Var}_{y_\zeta(u\beta_2(N))}(f)\right] du. \tag{5.10}$$

As in the proof of Lemma 4.8, (4.54)–(4.55) we conclude that, for every $f \in C([0, 1])$, the sequence $\{\mathscr{L}((\langle y_{\xi,2}^N(t\beta_2(N)), f\rangle)_{t\geq 0})\}$ is tight.

It then remains to show that any limit point of $\{\mathscr{L}((\langle y_{\xi,2}^N(t\beta_2(N)), f\rangle)_{t\geq 0})\}$ solves a well-posed martingale problem. This is achieved by showing that each weak limit point solves the martingale problem (L, δ_θ) defined in (4.4) where we set $c = c_2$ and $d = d_2$ (defined in 0(d), (iii)). The fact that this martingale problem is well-posed was established in Lemma 4.10.

The first step in the proof that a weak limit along a subsequence $\{N_k\}$ satisfies the martingale problem is to prove the weak convergence of

$$(Y^{N,2}(s\beta_2(N) + t))_{t\geq 0}$$

along the same subsequence. However the latter result will be shown in the proof of (5.5) using only the tightness part of the current proof.

It then remains to show a weak limit point of $\mathscr{L}((y_{\xi,2}^{N_k}(s\beta_2(N_k)))_{s\geq 0})$ as $k \to \infty$ solves the martingale problem (L, δ_θ) using the convergence of $Y^{N,2}(s\beta_2(N) + t)$ along the subsequence N_k. The processes

$$\{y_{\xi,2}^{N_k}(s\beta_2(N_k))_{s\geq 0}\}_{k\in\mathbb{N}}$$

are martingales with increasing processes

$$\left\{ \left(\int_0^s \frac{1}{N_k^2} \sum_{\xi \in M(N,2)} \mathrm{Var}_{y_{\xi,0}(u\beta_2(N))}(f)\, du \right)_{s\geq 0} \right\}_{k\in\mathbb{N}}.$$

In Lemma 4.10 we gave a reformulation of the martingale problem (L, δ_θ) in terms of a martingale problem involving the increasing process. According to this reformulation it suffices to show that the pair (increasing process, process) converges in distribution to the pair (x is the random variable over which we take the expectations)

$$\left(\int_0^s E_{\Gamma^2_{Z_u}}(\mathrm{Var}_x(f))\, du, \langle Z_s, f\rangle \right)_{s\geq 0}.$$

The key tool here is again to consider the empirical process of Y^N, to view the above pair as two continuous functionals of this empirical process and then to construct a coupling between the empirical processes for fixed N and the conjectured limit process as $N \to \infty$. Exactly this construction was given in the proof of Lemma 4.9 ((4.58)–(4.68)) and we refer to that proof for details. This establishes (5.4).

Proof of (5.5). This is concerned with the process describing the average over a family $y^N_{\xi,1}(t)$ with $\xi = (\xi_1, \xi_2)$. Due to the homogeneity of the process we can focus on $y^N_{\xi,1}(t)$ for some fixed ξ. When considering a single family we will choose the point $\xi = \xi^0$ and define the set $I(2, N) := \xi^0[1] = \{\xi \in M(2, N) | \xi_2 = 0\}$.

We use here the characterization of our processes as solutions of martingale problems. The proof is broken into steps. In steps 1, 2 we give two key lemmas, Lemmas 5.1, 5.2, giving tightness, respectively the martingale problem characterization of the limit. The proof of Lemma 5.2 is involved and proceeds in steps 3–6.

Step 1. Note that for each N the process $\{y^N_{\xi,1}(t)\}_{t\geq 0}$ is a continuous $\mathscr{P}([0,1])$-valued process. The first step is to verify that:

Lemma 5.1. *For every $s \in \mathbb{R}^+$ and for a fixed $\xi_2 \in \{0, \dots, N-1\}$ the sequence*

$$\{\mathscr{L}(y^N_{\xi,1}(s\beta_2(N) + t\beta_1(N))_{t\in\mathbb{R}^+})\}_{N\in\mathbb{N}} \tag{5.11}$$

is relatively weakly compact in $C([0,\infty), \mathscr{P}([0,1]))$. □

Proof. To prove this is suffices to show that $\{F(y^N_{\xi,1}(s\beta_2(N)+t\beta_1(N))_{t\in\mathbb{R}^+})\}_{N\in\mathbb{N}}$ are relatively weakly compact in $C([0,\infty), \mathbb{R})$ for continuous functions F : $\mathscr{P}([0,1]) \to \mathbb{R}$ of the form $F(z) := (\langle z, f\rangle)^k$, $f \in C([0,1])$, $k \in \{0, 1, \dots\}$.

Define $K_N F$ to be the function from $(\mathscr{P}([0,1]))^{I(2,N)}$ to $\mathbb{R}$ defined by

$$(K_N F)(y_{(0,0)}, \dots, y_{(N-1,0)}) := F(y_{\xi^0,1}).$$

For the function $F(y) = (\langle y, f\rangle)^k$ this becomes

$$F(y^N_{\xi^0,1}) = (K_N F)((y_\xi)_{\xi\in I(2,N)}) = \frac{1}{N^k} \sum_{\xi_1(1),\dots,\xi_1(k)=0}^{N-1} \prod_{i=1}^k \langle y_{(\xi_1(i),0)}, f\rangle. \tag{5.12}$$

Next observe that the process $(y^N_{\xi^0,1}(t))_{t\geq 0}$ is a functional of the two level system defined as the solution of the martingale problem $(\widetilde{L}_{N,2}, \mu_N)$ defined in (5.1). Therefore, the process $(y^N_{\xi^0,1}(t))_{t\in\mathbb{R}^+}$ satisfies the martingale problem $(\widetilde{L}_{N,2} \circ \pi_N \circ K_N, \mu_N)$ with $\mu_N = \mathscr{L}(y^N_{\xi,1}(0))$ on the probability space of the interacting system and where π_N is the embedding of $I(2, N)$ in $M(2, N)$. More precisely, for every F and fixed $\xi = (\xi_1, 0)$,

$$\left(F(y^N_{\xi^0,1}(t)) - \int_0^t \widetilde{L}_{N,2} \circ \pi_N(K_N F)(y^N_\xi(s))_{\xi\in M(2,N)}\, ds\right)_{t\geq 0}$$

is a continuous martingale.

For every F of the form considered it is easy to check by explicit calculation (compare (5.16), (5.17) for an expression for the first and second derivatives of F) that

$$(5.13)\quad \beta_1(N)\widetilde{L}_{N,2} \circ \pi_N \circ K_N(F)_{(y)} \leq C_k \|f\|^k_\infty \quad \forall y \in \mathscr{P}([0, 1]) \text{ and all } N \in \mathbb{N}.$$

We will insert the following two special choices for F : $F(z) = \langle z, f\rangle$ and $F(z) = \langle z, f\rangle^2$ with f running through a dense countable subset of $C([0, 1])$. We see that the local characteristics of the semimartingale $(F(y^N_{\xi^0,1}(t\beta_1(N))))_{t\geq 0}$ are bounded on compact time intervals uniformly in N. Hence again (see (1.1) in section 1) by the tightness criteria of Joffe and Métivier (1986, Proposition 3.2.3) the claim in (5.11) follows since the functions chosen form a countable dense subset of $C([0, 1])$. This completes the proof of Lemma 5.1.

Step 2. Recall that $(Z^{\theta,1}_t)_{t\in\mathbb{R}^+}$ is the unique solution of the martingale problem $(L_{\theta,1}, \Gamma^1_\theta)$. In this step we will verify that (5.5) would follow from the two convergence statements in the next lemma.

Lemma 5.2. (i) *For every* $\rho, \tilde{\rho} \in \mathscr{P}([0, 1])$ *we have that every weak limit point of the sequence*

(5.14a)

$$\{\mathscr{L}(y^N_{\xi,1}(s\beta_2(N) + t\beta_1(N))_{t\geq 0}) | y^N_{\xi,2}(s\beta_2(N)) = \rho, y^N_{\xi,1}(s\beta_2(N)) = \tilde{\rho}\}_{N\in\mathbb{N}}$$

(*compare* (5.11)) *is a solution of the well-posed martingale problem* $(L_{\rho,1}, \delta_{\tilde{\rho}})$.

(ii) *Let* $\nu(s) = \mathscr{L}(\widetilde{Z}_s)$ *where* $(\widetilde{Z}_s)$ *is defined in* (5.4). *Then*

$$(5.14b)\qquad \mathscr{L}(y^N_{\xi,2}(s\beta_2(N))) \underset{N\to\infty}{\Rightarrow} \nu(s). \quad \square$$

We will now complete the proof of (5.5) assuming the results of Lemma 5.2. This involves proving that $\mathscr{L}(y^N_{\xi,1}(s, \beta_2(N))) \underset{N\to\infty}{\Rightarrow} \int \Gamma^1_\rho \nu_s(d\rho)$ and that, conditioned on $\theta^*_k(s) = \rho$, $\{Z^{\theta,k}(t)\}_{t\in\mathbb{R}^+}$ is a stationary Fleming-Viot process with immigration source ρ independent of $\{\theta^*_k(s')\}_{s'\in\mathbb{R}^+}$. Not too much detail will be provided, since we shall follow closely the proof of Proposition 4.1, in particular (4.8)–(4.17). (Note that (5.14b) plays the role of assumption (i) in the statement of Proposition 4.1.) Pick a sequence $L(N) \uparrow \infty$, $L(N) = o(N)$ as $N \to \infty$, and consider the times $s\beta_2(N) - L(N)\beta_1(N)$. The first observation is that by (5.14b) and the continuity of $\nu(s)$ (recall Z_s has continuous paths) it follows that

$$\sup_{0\leq t\leq 1} |y^N_{\xi,2}(s\beta_2(N)) - y^N_{\xi,2}(s\beta_2(N) - L(N)\beta_1(N)t)|$$

will tend to 0 in probability. Next consider a weak limit point of

$$\{\mathscr{L}(y^N_{\xi,1}(s\beta_2(N) - \beta_1(N)L(N)))\}_{N\in\mathbb{N}}.$$

Now proceed as in section 4, and use the Markov property to write

$$\mathscr{L}(y^N_{\xi,1}(s\beta_2(N)))$$

as the result of the evolution through time 0 up to time $\beta_1(N)L(N)$ starting in the configuration $y^N_{\xi,1}(s\beta_2(N) - \beta_1(N)L(N))$. Now follow the argument of section 4(b), (i) (abstract scheme) found in (4.8)–(4.17) to show that $\mathscr{L}(y^N_{\xi,1}(s\beta_2(N)))$ converges to $\int \Gamma^1_\rho \nu_s(d\rho)$ (and note that this is independent of $y^N_{\xi,1}(s\beta_2(N)-\beta_1(N)L(N))$. The conditional independence of $\{Z^{\theta,k}(t)\}_{t\in\mathbb{R}^+}$ follows from the uniqueness of the Fleming-Viot martingale problem with immigration source ρ. This implies that the joint law of $y^N_{\xi,2}(s\beta_2(N))$, $y^N_{\xi,1}(s\beta_2(N))$ converges to $\int \delta_\rho \otimes \Gamma^1_\rho \nu_s(d\rho)$. In following the proofs of section 4, we replace Lemma 4.4 and Lemma 4.5 in the argument preceding (4.17) by (5.14a) (with $s\beta_2(N)$ replaced by $s\beta_2(N)-L(N)\beta_1(N)$). This completes step 2 and the proof of (5.5).

Step 3. We now begin the verification of Lemma 5.2. Since (5.14b) was proved in the proof of (5.4), it therefore suffices to prove (5.14a) assuming that (5.14b) is true. This will be done in steps 4–6 below. We first reformulate the assertion (5.14a) in a form more suitable for calculation (see e.g. Ethier and Kurtz, (Theorem 8.10)). The assertion (5.14a) is implied by: (recall that $L_{\rho,1}$ generates $Z^{\rho,1}(t)$)

(5.15)
$$\left|\int_0^T [\widetilde{L}_{N,2}\circ\pi_N\circ(K_NF)(y^N_\xi)_{\xi\in M(2,N)}) - L_{\rho,1}F(y^N_{\xi_0,1})](s\beta_2(N)) + t\beta_1(N)))\,dt\right|$$
$$\underset{N\to\infty}{\longrightarrow} 0 \text{ in probability.}$$

The proof of (5.15) is given in various steps. First, in step 4 we give an explicit representation which we then analyse in the limit $N\to\infty$ in steps 5 and 6.

Step 4. We start by calculating $\widetilde{L}_{N,2}\circ\pi_N\circ(K_NF)$ in explicit form. Let $\xi(1),\ldots,$ $\xi(k)$ be a fixed collection of elements of $M(2,N)$. First observe that with $u,v\in[0,1]$:

$$(5.16)\quad \left[\frac{\partial}{\partial y_\xi}\left(\prod_{i=1}^k \langle y_{\xi(i)}, f\rangle\right)\right]_{(y)}(u) = \begin{cases} 0 \quad \text{if } \xi \notin \{\xi(1),\ldots,\xi(k)\}, \\ nf(u)\left(\displaystyle\prod_{\substack{i=1\\ i\neq j(\xi)}}^k \langle y_{\xi(i)}, f\rangle\right) \\ \qquad\qquad \text{if } \xi \in \{\xi(1),\ldots,\xi(k)\}, \end{cases}$$

where $j = j(\xi)$ and $n = n(\xi)$ are defined by $j = \min\{i : \xi = \xi(i)\}$, $n = |\{i : \xi(i) = \xi\}|$.

(5.17)

$$\left[\frac{\partial^2}{\partial y_\xi \partial y_\xi}\left(\prod_{i=1}^{k}\langle y_{\xi(i)}, f\rangle\right)\right]_{(y)}(u, v)$$
$$= \begin{cases} 0 & \text{if } \xi \notin \{\xi(1), \dots, \xi(k)\}, \\ 0 & \text{if } \xi = \xi(j), \xi(i) \neq \xi(j), j \neq i, \\ n(n-1)f(u)f(v)\left(\prod_{\substack{i\neq j,\ell \\ i=1}}^{k}\langle y_{\xi(i)}, f\rangle\right), & \text{otherwise,} \end{cases}$$

where again $j = j(\xi)$, $\ell = \ell(\xi)$ are the two smallest indices with $\xi = \xi(j) = \xi(\ell)$, and $n = n(\xi) = |\{i : \xi(i) = \xi\}|$.

Recall that we are using $\xi = (\xi_1, 0)$ and $I(2, N) = \{\xi \in M(2, N) | \xi_2 = 0\}$. Using (5.16), (5.17) we get the following explicit representation of the generator for the process in the fast time scale:

(5.18)

$$\begin{aligned} N \cdot \widetilde{L}_{N,2} &\circ \pi_N \circ (K_N F) \\ &= \frac{1}{N^k} \sum_{\xi(1),\dots,\xi(k)\in I(2,N)} \\ &\cdot \Bigg[\Bigg\{ \sum_{\xi'\in I(2,N)} \sum_{\xi\in I(2,N)} I_{\{\xi\in\{\xi(1),\dots,\xi(k)\}\}} \\ &\quad \cdot N \cdot \left(\int_0^1 n\frac{c_0}{N}(y_{\xi'}(du) - y_\xi(du))f(u)\right)\left(\prod_{i=1}^{k}\langle y_{\xi(i)}, f\rangle / \langle y_\xi, f\rangle\right)\Bigg\} \\ &\quad + \Bigg\{\sum_{\xi\in I(2,N)} \sum_{\xi'\in M(2,N)} I_{\{\xi\in\xi(1),\dots,\xi(k)\}\}} \\ &\qquad \cdot N \cdot \int_0^1 \frac{c_1}{N^3}(y_{\xi'}(du) - y_\xi(du))f(u)\left(\prod_{i=1}^{k}\langle y_{\xi(i)}, f\rangle / \langle y_\xi, f\rangle\right)\Bigg\} \\ &\quad + d_0\Bigg\{\sum_{\xi\in I(2,N)} N \cdot \int_0^1\int_0^1 n(n-1)f(u)f(v)\mathbf{1}_{\{\xi\in\{\xi(1),\dots,\xi(k)\}\}} Q_\xi(du, dv) \\ &\qquad \cdot \left(\prod_{i=1}^{k}\langle y_{\xi(i)}, f\rangle / \langle y_\xi, f\rangle^2\right)\Bigg\}\Bigg]. \end{aligned}$$

Recall $n = n(\xi)$ and that $n(n-1)$ is equal to 0 if ξ does not appear at least twice in the collection $\{\xi(i)\}_{i=1,\dots,k}$. The expression on the right-hand side of

(5.18) equals the quantity I_N defined below (the c_0-sum is equal to 0!),

(5.19)

$$I_N = \frac{1}{N^k} \sum_{\xi(1),\dots,\xi(k)\in I(2,N)} \cdot \Bigg[\Bigg\{ \sum_{\xi\in I(2,N)} \left(\int_0^1 \left(\frac{c_1}{N^2} \sum_{\xi'\in M(2,N)} (y_{\xi'}(du) - y_\xi(du)) \right) \mathbf{1}_{\{\xi\in\{\xi(1),\dots,\xi(k)\}\}} f(u) \right) \cdot \left(\left(\prod_{i=1}^k \langle y_{\xi(i)}, f\rangle \right) / \langle y_\xi, f\rangle \right) \Bigg\}$$

$$+ \Bigg\{ N \sum_{\xi\in I(2,N)} \left(n(n-1) \int_0^1 f(u)f(v) \mathbf{1}_{\{\xi\in\{\xi(1),\dots,\xi(k)\}\}} Q_\xi(du, dv) \right) \cdot \left(\left(\prod_{i=1}^k \langle y_{\xi(i)}, f\rangle \right) / (\langle y_\xi, f\rangle)^2 \right) \Bigg\} \Bigg].$$

Step 5. The quantity I_N can be replaced for $N \to \infty$ by J_N defined as below, where the terms $n(n-1)$ are dropped. (Simply use that $|\langle y_\xi, f\rangle| \le \|f\|_\infty$ and that in the sum we can drop groups of terms as long as their number is small compared to the power of N^{-1} in front of the sum.)

$$J_N = \Bigg[\int_0^1 c_1 \left(\left\{ \frac{1}{N^2} \sum_{\xi'\in M(2,N)} y_{\xi'}(du) \right\} - \left\{ \frac{1}{N} \sum_{\xi\in I(2,N)} y_\xi(du) \right\} \right) \cdot \left(\frac{1}{N^{k-1}} \sum_{\substack{\xi(1),\dots,\xi(k-1)\\ \in I(2,N)}} f(u) \left(\prod_{i=1}^{k-1} \langle y_{\xi(i)}, f\rangle \right) \right) \Bigg]$$

$$+ d_0 \Bigg[\int_0^1 \int_0^1 \Bigg\{ 2f(u)f(v) \left(\frac{1}{N} \sum_{\xi\in I(2,N)} Q_{y_\xi}(du, dv) \right) \cdot \left(\frac{1}{N^{k-2}} \sum_{\substack{\xi(1),\dots,\xi(k-2)\\ \in I(2,N)}} \prod_{i=1}^{k-2} \langle y_{\xi(i)}, f\rangle \right) \Bigg\} \Bigg].$$

Recall

$$F : \mathscr{P}([0, 1]) \to \mathbb{R} : F(y) = (\langle y, f\rangle)^k \quad \text{for some } f \in C([0, 1])$$

and the formulas for the first and second derivatives of F with respect to y (see (6.3)). Then we can continue (5.19) by

(5.20)

$$\begin{aligned}
J_N = & \int_0^1 c_1(y_{\xi^0,2}(du) - y_{\xi^0,1}(du)) \left(\frac{\partial F}{\partial y_{\xi^0,1}}\right)_{(y_{\xi^0,1})} (u) \\
& + d_0 \int_0^1 \int_0^1 \left(\frac{\partial^2 F}{\partial y_{\xi^0,1} \partial y_{\xi^0,1}}\right)_{(y_{\xi,1})} (u, v) \left(\frac{1}{N} \sum_{\xi \in I(N,2)} Q_{y_\xi}(du, dv)\right) \\
= & \int_0^1 c_1(y_{\xi^0,2}(du) - y_{\xi^0,1}(du)) \left(\frac{\partial F}{\partial y_{\xi^0,1}}\right)_{(y_{\xi^0,1})} (u) \\
& + d_0 \int_0^1 \int_0^1 \left(\frac{\partial^2 F}{\partial y_{\xi^0,1} \partial y_{\xi^0,1}}\right)_{(y_{\xi^0,1})} (u, v) y_{\xi^0,1}(du) \delta_u(dv) \\
& - d_0 \left\{\int_0^1 \int_0^1 \left(\frac{\partial^2 F}{\partial y_{\xi^0,1} \partial y_{\xi^0,1}}\right)_{(y_{\xi^0,1})} (u, v) \left(\frac{1}{N} \sum_{\xi \in I(2,N)} y_\xi(du) y_\xi(dv)\right)\right\}.
\end{aligned}$$

We have already proved that $y^N_{\xi^0,2}(t)$ does not vary in time intervals of length $O(\beta_1(N))$ in the limit $N \to \infty$ (relation (5.3)). Furthermore we work on the event $y^M_{\xi^0,2}(s\beta_2(N)) = \theta$ according to the assumptions of the lemma. Therefore we can replace J_N for $N \to \infty$ by:

(5.21)

$$\begin{aligned}
& \int_0^1 c_1(\theta - y_{\xi^0,1}(du)) \left(\frac{\partial F}{\partial y_{\xi^0,1}}\right)_{(y_{\xi^0,1})} (u) \\
& + d_0 \int_0^1 \left(\frac{\partial^2 F}{\partial y_{\xi^0,1} \partial y_{\xi^0,1}}\right)_{(y_{\xi^0,1})} (u, v) y_{\xi,1}(du) \delta_u(dv) \\
& - d_0 \left\{\int_0^1 \int_0^1 \left(\frac{\partial^2 F}{\partial y_{\xi^0,1} \partial y_{\xi^0,1}}\right)_{(y_{\xi^0,1})} (u, v) \left(\frac{1}{N} \sum_{\xi \in I(2,N)} y_\xi(du) y_\xi(dv)\right)\right\}.
\end{aligned}$$

Hence we are left with analysing the third term in (5.21).

Step 6. For the purpose of analysing the third term in (5.21) we need the behavior of the following expression as $N \to \infty$ along the path of $Y^{N,2}(t)$:

$$\frac{1}{N} \sum_{\xi \in I(2,N)} \int_0^1 \int_0^1 f(u) f(v) y_\xi(du) y_\xi(dv) = \frac{1}{N} \sum_{\xi \in I(2,N)} (\langle y_\xi, f\rangle)^2. \tag{5.22}$$

By the same argument as in section 4 (proof of (4.56)), or alternatively based on the analysis in Dawson and Greven (1993c), it can be shown that given $y^N_{\xi,2}(s\beta_2(N)) = \theta$ we have

$$\begin{aligned}
& \int_0^T \left(\int_0^1 \int_0^1 \frac{1}{N} \sum_{\xi \in I(2,N)} y^N_\xi(t\beta_1(N))(du) y^N_\xi(t\beta_1(N))(dv) f(u) f(v)\right) dt \\
& \underset{N\to\infty}{\longrightarrow} \int_0^T E_{\Gamma^1_\theta}(\langle x, f\rangle^2)\, ds.
\end{aligned} \tag{5.23}$$

Since Γ^1_θ is the equilibrium of the Fleming-Viot process with immigration from the source θ, generated by $L_{\theta,1}$ (see section 0(d),(iii)), we have the equations

(equilibrium condition—for details see proof of Lemma 6.1)

$$
\begin{aligned}
E_{\Gamma^1_\theta}(\langle x, f\rangle^2) &= \left(1+\frac{d_0}{c_0}\right)^{-1} E_{\Gamma^1_\theta}\left(\langle\theta, f\rangle^2 + \frac{d_0}{c_0}\langle x, f^2\rangle\right) \\
&= \left(1+\frac{d_0}{c_0}\right)^{-1} \langle\theta, f\rangle^2 + \frac{d_0/c_0}{1+d_0/c_0}\langle\theta, f\rangle^2.
\end{aligned} \tag{5.24}
$$

From (5.21)–(5.24) we obtain

$$
\begin{aligned}
&\left|\int_0^T \left[\widetilde{L}_{N,2}\circ\pi_N\circ(K_N F)((y^N_\xi)_{\xi\in M(2,N)}) - L_{\theta,1}F(y^N_{\xi^0,1})\right](s\beta_2(N)+t\beta_1(N))\,dt\right| \\
&\quad= \left|\int_0^1\int_0^1 \left(\frac{\partial^2 F}{\partial y_{\xi^0,1}\partial y_{\xi^0,1}}\right)_{(y_{\xi^0,1})}(u,v)\right. \\
&\qquad\cdot\left\{d_0\left[y^N_{\xi^0,1}(s\beta_2(N)+t\beta_1(N))(du)\delta_u(dv)\right.\right. \\
&\qquad\quad\left.- \left(\frac{1}{N}\sum_{\xi\in I(2,N)} y_\xi(s\beta_2(N)+t\beta_1(N))(du)y_\xi(s\beta_2(N)+t\beta_1(N))(dv)\right)\right] \\
&\qquad - \frac{c_0 d_0}{c_0+d_0}\left[y_{\xi^0,1}(s\beta_2(N)+t\beta_1(N))(du)\delta_u(dv)\right. \\
&\qquad\qquad\qquad - y_{\xi^0,1}(s\beta_2(N)+t\beta_1(N))(du)y_{\xi^0,1} \\
&\qquad\qquad\qquad\qquad \left.\left.\left.\cdot(s\beta_2(N)+t\beta_1(N))(dv)\right]\right\}dt\right| \\
&\quad\to 0 \quad\text{as } N\to\infty.
\end{aligned}
$$

This completes the proof of (5.15).

Proof of (5.6). Consider a new system where, in (5.1), $y_{\xi,1}$ is replaced by $\theta^{*,1}(s)$ and c_2 is set equal to 0. The law of this new system will converge according to Lemma 4.1 to $\Gamma^0_{\theta^*,1}$ as $t\to\infty$. Hence if we can show that $Y^{N,2}(s\beta_2(N) - L(N)\beta_1(N))$ and the new system are close for $L(N)\to\infty$ but $L(N)=o(N)$ we are done. The latter can be obtained from (5.5) and (5.29). We have proved (5.6) and completed the proof of Proposition 5.1.

Again as at the end of section 4, namely in 4(e) we can generalize the results to a system $\widetilde{Y}^{N,2}_t$ which is obtained by adding the perturbation

$$
R_N(F)_{(y)} = \sum_{\xi_1,\xi_2=1}^{N} B(N)\int_{[0,1]}\int_{[0,1]}\frac{\partial^2 F}{\partial y_\xi\partial y_\xi}(u,v)Q_{y_\xi}(du,dv),
$$
$$
\xi=(\xi_1,\xi_2),\; y=(y_\xi),\; \xi\in\{0,\dots,N-1\}^2,
$$

to the generator $\widetilde{L}_{N,2}$ in (5.1) where $B(N)$ are bounded operators on $C([0,1]^2)$.

Corollary. *Assume that* $\|B(N)\|\to 0$ *as* $N\to\infty$. *Then the assertions of Proposition* 5.1 *also hold for the perturbed process* $(\widetilde{Y}^{N,2}(t))$.

(b) **Finite ℓ-level models** ($\ell = 3, 4, \dots$). The purpose of this subsection is to investigate general ℓ-level models based on the knowledge of two level models.

These are still *finite* systems. We define the process $(Y^{N,\ell}(t))_{t\geq 0}$ on the state space $(\mathscr{P}([0,1]))^{M(\ell,N)}$, $M(\ell,N)=\{0,1,\dots,N-1\}^\ell$, via the martingale problem $(\widetilde{L}_{N,\ell},\mu_N)$, defined as follows. We write $y\in(\mathscr{P}([0,1]))^{M(\ell,N)}$ in the form $y=(y_\xi)_{\xi\in M(\ell,N)}$, $\xi=(\xi_1,\dots,\xi_\ell)$. Let

(5.25)

$$\widetilde{L}_{N,\ell}(F)(y)=\sum_{\xi\in M(\ell,N)}\left[\int_0^1\frac{\partial F(y)}{\partial y_\xi}(u)\left(\sum_{i=1}^{\ell}\frac{c_{i-1}}{N^{i-1}}(y_{\xi,i}(du)-y_\xi(du))\right)\right.$$
$$\left.+d_0\int_0^1\int_0^1\frac{\partial^2F(y)}{\partial y_\xi\partial y_\xi}(u,v)Q_{y_\xi}(du,dv)\right],$$

$$F(y)=\prod_{i=1}^{n}\int_0^1 f_{\xi(i)}(u)y_{\xi(i)}(du),\qquad f_{\xi(i)}\in C([0,1]),\ n\in\mathbb{N},\ \xi(i)\in M(\ell,N).$$

Denote by

$$y_{\xi,j}=\frac{1}{N^j}\sum_{\xi_1,\dots,\xi_j=0}^{N-1}y_{(\xi_1,\dots,\xi_\ell)},\qquad \xi\in M(\ell,N),\ 1\leq j\leq\ell.$$

Then we get the following multiple space-time scale result, using the definitions from section 0(d) for the stationary process $(Z^{\theta,k}(t))_{t\geq 0}$.

Proposition 5.2. *Let $j,k\in\{0,1,\dots,\ell\}$. For $j<k$:*

$$\mathscr{L}((y_{\xi,k}(tN^j))_{t\in\mathbb{R}^+})\underset{N\to\infty}{\Rightarrow}\delta_{\{Y_t\equiv\theta\}}\tag{5.26}$$

and

$$\mathscr{L}((y_{\xi,\ell}(tN^\ell))_{t\in\mathbb{R}^+})\underset{N\to\infty}{\Rightarrow}\mathscr{L}((\widetilde{Z}^\ell(t))_{t\in\mathbb{R}^+}),\tag{5.27}$$

where $(\widetilde{Z}^\ell_s)_{s\geq 0}$ solves the martingale problem in (5.4) *with d_2 replaced by d_ℓ. Moreover, if $k<\ell$, $s(N)\to\infty$ and $s(N)/N\to 0$ as $N\to\infty$, then*

$$\mathscr{L}((y_{\xi,k}(s(N)N^k))_{t\in\mathbb{R}^+})\underset{N\to\infty}{\Rightarrow}\Gamma^k_\theta.\tag{5.27′}$$

For $0\leq k<j<\ell$:

(5.28) $\mathscr{L}((y_{\xi,k}(sN^j+tN^k))_{t\in\mathbb{R}^+})\underset{N\to\infty}{\Rightarrow}\mathscr{L}((Z^{\theta^*_k(s),k}(t))_{t\in\mathbb{R}^+})$ *with* $\mathscr{L}(\theta^*_k(s))=\int\mu^{j-1,k}_\theta\nu(s)(d\theta)$, *$\theta^*_k(s)$ is independent of the evolution and $\nu(s)=\mathscr{L}(\widetilde{Z}^j_s)$ for $s\in\mathbb{R}^+$ and $\mu^{j-1,k}_\theta$ is as defined in* 0(d). □

Proof. The proof of (5.26) is exactly the same argument as in the proof of (5.3).

The relation (5.27) follows from relation (5.28) in the same fashion as in the proof of relation (5.4).

The relation (5.28) is proved by induction over k, starting with $k=2$ which was treated in subsection 5(a). The induction step is very similar to the argument there in which the level $(k-1)$ systems now play the role played by the level 1 systems in the proof of (5.4). However if $k>2$, then the second order term is a perturbed Fleming-Viot operator with a perturbation (due to the level $(k-1)$ approximation) that disappears in the $N\to\infty$ limit. The proof involves using an analogue of the corollary formulated at the end of subsection 5(a). We omit the details here.

Corollary 5.2. *Let $s(N)$ be a sequence with $s(N) \to \infty$ but $s(N)/N \to 0$ as $N \to \infty$. Then (5.28) can be modified to give*

$$\begin{aligned}(5.28') \qquad &\mathscr{L}((y_{\xi,k}(s(N)N^j + tN^k))_{t\in\mathbb{R}^+}) \underset{N\to\infty}{\Rightarrow} \mathscr{L}((Z^{k,\theta_k^*}(t))_{t\in\mathbb{R}^+}) \\ &\qquad\qquad \text{with } \mathscr{L}(\theta_k^*) = \lim_{s\to\infty} \mathscr{L}(\theta_k^*(s)) = \mu_\theta^{j,k}.\end{aligned}$$

Proof. Here one observes first that Γ_θ^k is the equilibrium measure of the process $\widetilde{Z}^{k,\theta}(t)$. Since the latter process is ergodic, we have in fact that $\mathscr{L}(\widetilde{Z}^{k,\theta}(t))$ converges as $t \to \infty$ to Γ_θ^k. Since we proved in section 4(c) that we can couple $y_{\xi,k}(sN^j+t)$ and $\widetilde{Z}^{k,\theta_1(N,s)}(t)$ where we set $\theta_1(N,s) = y_{\xi,k}(sN^j)$ so that the expected difference is nonincreasing in t, the result follows from (5.28).

(c) **Approximation of infinite by finite systems.** Finally we need to prove that our infinite system $(X^N(t))_{t\geq 0}$ can locally be approximated by ℓ-level systems if we choose a level increasing with time in a suitable way. Due to homogeneity properties it suffices in fact to compare one component. Let ξ^0 be the element $(0, 0, 0, \ldots) \in \Omega_\infty$ and $\xi^{0,\ell} = (0, \ldots, 0) \in \mathbb{N}^\ell$. Denote by

$$X^N(t) = (x_\xi^N(t))_{\xi\in\Omega_N}, \qquad Y^{N,\ell}(t) = (y_\xi^{N,\ell}(t))_{\xi\in\{0,\ldots,N-1\}^\ell},$$

the infinite system of interacting Fleming-Viot processes, respectively the finite ℓ-level system, introduced in the last section. As starting measure of the finite system use the "restriction" of the infinite initial state. (View $(\xi_1, \ldots, \xi_\ell)$ as an element of Ω_∞ by setting $(\xi_1, \ldots, \xi_\ell) \leftrightarrow (\xi_\ell, \xi_{\ell-1}, \ldots, \xi_1, 0, 0, \ldots)$.) The key to this approximation property is *coupling*. By continuation (set y_ξ equal for all ξ with the same first ℓ components) of $(Y^{N,\ell}(t))_{t\in\mathbb{R}^+}$, which is a process with state space $(\mathscr{P}([0,1]))^{M(\ell,N)}$ we obtain a process $(\widehat{Y}^{N,\ell}(t))_{t\in\mathbb{R}^+}$ with state space $(\mathscr{P}([0,1]))^{\Omega_N}$ contained in $(\mathscr{P}([0,1]))^{\Omega_N}$. For the state space $(\mathscr{P}([0,1]))^{\Omega_\infty}$ we can carry through the whole procedure of constructing a coupling, which we gave in section 4(c). (Replace everywhere the index set $\{0, 1, \ldots, N-1\}$ by $\Omega_\infty \subseteq \{0, 1, \ldots\}^{\mathbb{N}}$.) So let us assume $(\widehat{Y}^{N,\ell}(t))_{t\geq 0}$ and $(Y^N(t))_{t\geq 0}$ are constructed on one probability space. Recall the definition in (4.22) of the norm $\|\cdot\|$.

Proposition 5.3. *For every $\ell \in \mathbb{N}$,*

$$(5.29) \qquad E\left(\sup_{t\leq\beta_{\ell-1}(N)} \|x_\xi^N(t) - \hat{y}_{\xi^\ell}^{N,\ell}(t)\|\right) \underset{N\to\infty}{\to} 0.$$

Moreover the convergence is uniform in ℓ and θ. □

Remark. This means that for systems starting in homogeneous distributions we can replace the infinite hierarchical system observed in time scales of order $\beta_j(N)$ or less by a $(j+1)$-level system, at least in the limit $N \to \infty$.

Proof. We proceed in two steps. First we want to apply the relation (4.36) to estimate $\|x_\xi^N(t) - \hat{y}_{\xi^\ell}^{N,\ell}(t)\|$ with t running through a fixed sequence t_N. Define $\Delta(t_N)$ by

$$(5.30) \qquad \Delta(t_N) := E\|x_{\xi^0}^N(t_N) - \hat{y}_{\xi^0}^{N,\ell}(t_N)\|.$$

Now recall the notation from section 4(c), in particular (4.35). We have for a constant C independent of N and s:

$$h(s) \leq CsN^{-\ell} \tag{5.31}$$

and therefore if $t_N \leq \beta_{\ell-1}(N)$

$$\lim_{N\to\infty} \Delta(t_N) = 0. \tag{5.32}$$

The second step is to use the martingale structure to obtain from (5.32) the desired relation (5.29). For every $f \in C([0, 1])$ the process

$$(\langle x_{\xi^0}^N(t) - \hat{y}_{\xi^0,\ell}^{N,\ell}(t), f\rangle)_{t\geq 0} \tag{5.33}$$

is a semimartingale where the bounded variation part is bounded almost surely by the constant $DN^{-\ell}T$ uniformly on the time interval $[0, T]$. Choose $N \geq N_0(\varepsilon)$ such that $DN^{-\ell}T \leq \varepsilon/2$. Then by the martingale inequality one gets

$$P\left(\sup_{t\leq T}|\langle x_{\xi^0}^N(t) - \hat{y}_{\xi^0,\ell}^{N,\ell}(t), f\rangle| \geq \varepsilon/2\right) \leq \frac{2}{\varepsilon}E|\langle x_{\xi^0}^N(T) - \hat{y}_{\xi^0,\ell}^{N,\ell}(T), f\rangle|. \tag{5.34}$$

Suppose now f is a Lipschitz function on $[0, 1]$; that is, for some $C_f < \infty$

$$|f(u) - f(v)| \leq C_f|u - v| \quad \forall u, v \in [0, 1]. \tag{5.35}$$

Then the right-hand side of (5.34) can be bounded by (see (4.22) for $\|\cdot\|$)

$$\frac{2}{\varepsilon}C_f\|x_{\xi^0}^N(T) - \hat{y}_{\xi^0,\ell}^{N,\ell}(T)\|. \tag{5.36}$$

To complete the proof of Proposition 5.3, we insert (5.36) in (5.34), set $T = T_N = \beta_{\ell-1}(N) = o(N^\ell)$ and then use (5.32).

(d) **Extension to the spatial finite-dimensional distributions.** Up to this point we have studied the behavior of single components or averages over blocks in various time scales. In this subsection we extend the analysis to the study of the behavior of tuples of components, tuples of averages over blocks, etc. These results will be needed in order to prove Theorem 0.4, which involves the complete dependence structure of the model.

Proposition 5.4. *Pick two elements ζ, ξ with $d(\zeta, \xi) \geq \ell$. Then as $N \to \infty$ the laws $\mathscr{L}((y_\xi^N(t \wedge \beta_{\ell-1}(N)))_{t\geq 0}|Y^N(0))$ and $\mathscr{L}((y_\zeta^N(t \wedge \beta_{\ell-1}(N)))_{t\geq 0}|Y^N(0))$ are asymptotically independent. Moreover this independence is uniform in the sense: For all $K_1, K_2 \in [0, \infty)$, $n \in \mathbb{N}$,*

(5.37)
$$\sup_{f,g\in\mathscr{C}_{K_1,K_2,n}} (|E[f(X_1^N)g(X_2^N)|Y^N(0)] - E(f(X_1^N)|Y^N(0))E(g(X_2^N)|Y^N(0))|) \underset{N\to\infty}{\longrightarrow} 0$$

where $X_1^N = y_\xi^N(t \wedge \beta_{\ell-1}(N))$, $X_2^N = y_\zeta^N(t \wedge \beta_{\ell-1}(N))$,

$$\mathscr{C}_{K_1,K_2,n} = \{h : C([0, \infty), \mathscr{P}([0, 1])) \to \mathbb{R}|h(X) = f(X(t_1), \dots, X(t_n)) \text{ for some } f,\ f : (\mathscr{P}([0, 1]))^n \to \mathbb{R},\ C_f \leq K_1,\ \|f\|_\infty \leq K_2\}$$

(recall (5.35)). □

Proof. We use again the characterization of the involved process as solutions to a martingale problem and compare with the system where $c_k = 0\ \forall k \geq \ell - 1$.

In the latter model the independence property holds by the very definition of the process for every N. Then the result follows from Proposition 5.3.

(e) **Proof of Theorems 0.2, 0.4.** Theorem 0.2 will now follow by combining Proposition 5.2 with Proposition 5.4. Namely, we check the three cases, $k > j$, $k = j$, $k < j$ in the assertion of Theorem 0.2 as follows:

Cases $k > j$, $k = j$: In these cases we choose $\ell > k + 1$. Hence applying Proposition 5.2 relation (5.26), respectively (5.27), for $k = \ell$, that is, a system with ℓ levels, gives in combination with Proposition 5.3 the assertions.

Case $k < j$: Here choose $\ell > j + 1$ and then apply Corollary 5.2 (5.28) for the ℓ-level system to get the assertion.

In order to prove Theorem 0.4 we shall use Theorem 0.2 in combination with Proposition 5.4. For that purpose pick $\ell = j+1$ and then use Proposition 5.4 to see that sites which are at distance at least ℓ apart will in fact evolve independently in the time scale $\beta_j(N)$ considered in the Theorem 0.4. Then use Theorem 0.2 as follows. Given the time scale $\beta_j(N)$ in (0.35) we apply Case 3 of Theorem 0.2 with this same time scale but with $k = 0$; that is, we apply Theorem 0.2 at the individual component level. Then we get the first assertion of (0.35). The second one is obtained using the remark following Theorem 0.2.

(f) **Approximation properties of the interaction chain: Proof of Theorem 0.10.** Since we now look at the system at a sequence of time points spreading apart but which always start beyond $s(N)N^j$, resp. $\sigma(N)$, we will use Theorem 0.2 to control the behavior right at these large times $s(N)N^j$, resp. $\sigma(N)$. We will then give additional arguments in order to analyse what happens at later time points defined via $\hat{t}$.

(0.68) was proved in section 3 and the ground work for (0.69)–(0.71) has already been carried out in sections 5(b) and 5(c).

Proof of (0.69). The key ingredients here are Proposition 5.2 combined with Propositions 5.3 and 5.4. The first step is however to use the argument of the proofs of Lemmas 4.8, 4.9 (given in (4.49)–(4.53)) to conclude that

$$\mathscr{L}((x^N_{\xi,k+i}(t\beta_{k-1}(N)))_{t\in\mathbb{R}^+}) \to \delta_{\{X_t\equiv\theta\}}, \qquad i = 0, 1, 2, \ldots. \tag{5.38}$$

This proves that (recall (0.65))

$$\mathscr{L}(Z^{j,N}_{-j-1}(\hat{t})) \underset{N\to\infty}{\Rightarrow} \delta_\theta. \tag{5.39}$$

The next step is to choose $\ell = j+2$ in Proposition 5.3, which then tells us that

$$\begin{aligned}&\mathscr{L}\left(y^{N,\ell}_{\xi,k}\left(s(N)N^j + \sum_{i=1}^{j-k} t_i N^{j-i}\right)\right)\\ &\qquad - \mathscr{L}\left(x^N_{\xi,k}\left(s(N)N^j + \sum_{i=1}^{j-k} t_i N^{j-i}\right)\right) \underset{N\to\infty}{\Rightarrow} 0.\end{aligned} \tag{5.40}$$

However for the first expression we can apply Corollary 5.2 after observing that we can write

$$\begin{aligned}&s(N)N^j + \sum_{i=1}^{j-k} t_i N^{j-i} = s'(N)N^j + t_{j-k}(N)N^k,\\ &s'(N)/N \to 0, \quad s'(N) \to \infty \quad \text{and} \quad t_{j-k}(N) \to t_{j-k}.\end{aligned} \tag{5.41}$$

This will then give (recall (0.38))

$$\mathscr{L}(Z_{-k}^{j,N}(\hat{t})) \underset{N\to\infty}{\Rightarrow} \mu_\theta^{j,k} = \mathscr{L}(\hat{\theta}_k^j), \qquad k = j, j-1, \dots, 0. \tag{5.42}$$

The next step is to verify that

$$\mathscr{L}(Z_{-k}^{j,N}(\hat{t}), Z_{-k+1}^{j,N}(\hat{t})) \underset{N\to\infty}{\Rightarrow} \mathscr{L}(\hat{\theta}_k^j, \hat{\theta}_{k-1}^j), \qquad k = j, j-1, \dots, 1.$$

To do this we first use an argument similar to that of step 2 of the proof of (5.5) to show that

$$\begin{gathered}\sup_{-1\le t\le 1} |(x_{\xi,k}^N(S_j(N) + tL(N)\beta_{k-1}(N))) - (x_{\xi,k}^N(S_j(N)))| \\ \text{tends to 0 in probability as } N \to \infty\end{gathered} \tag{5.43}$$

where $S_j(N) := s(N)N^j - L(N)N^{k-1}$, $L(N) \to \infty$, $L(N)/N \to 0$ as $N \to \infty$, and $s(N) \to \infty$, $s(N)/N \to 0$ as $N \to \infty$. Then by an argument similar to that in step 2 in the proof of (5.5) we obtain

$$\begin{gathered}\mathscr{L}((x_{\xi,k-1}^N(S_j(N) + (t + L(N)N^{k-1}))|x_{\xi,k}^N(S_j(N))) = \rho) \\ \underset{N\to\infty}{\Rightarrow} \Gamma_\rho^{k-1}.\end{gathered} \tag{5.44}$$

Continuing in this way and also using (5.39) we obtain that

$$\mathscr{L}(Z_{-k}^{j,N}(\hat{t}), \dots, Z_0^{j,N}(\hat{t})) \underset{N\to\infty}{\Rightarrow} \mathscr{L}(\hat{\theta}_k^j, \dots, \hat{\theta}_0^j). \tag{5.45}$$

This yields relation (0.69).

Proof of (0.70). At this point we will have to make use of some properties of the interaction chain which will be proved later in section 6. Let $\Lambda_\rho^{j,k}$ be the following law on $(\mathscr{P}([0,1]))^{k+1}$:

$$\Lambda_\rho^{j,k} = \mathscr{L}(\hat{\theta}_0^j, \hat{\theta}_1^j, \dots, \hat{\theta}_k^j | \hat{\theta}_{j+1}^j = \rho). \tag{5.46}$$

Then by (6.24) we know that

$$\Lambda_\rho^{j,k} \underset{j\to\infty}{\Rightarrow} \mathscr{L}(\hat{\theta}_0^\infty, \hat{\theta}_1^\infty, \dots, \hat{\theta}_k^\infty | \hat{\theta}_\infty^\infty = \rho) = \Lambda_\rho^{\infty,k} \tag{5.47}$$

uniformly in ρ. Furthermore from (0.25b) and (0.44) we know that the map

$$\rho \to \Lambda_\rho^{j,k} \text{ is continuous (uniformly in } j!). \tag{5.48}$$

Next recall the definition of the time scale $\sigma(N)$ in (0.67) and introduce a modified time scale $\Delta(N)$ (allowing for relaxation on j-levels)

$$\Delta(N) = (\sigma(N) - s(N)N^j)^+, \quad s(N) \uparrow \infty, s(N)/N \to 0 \text{ as } N \to \infty. \tag{5.49}$$

Now observe that we can use the Markov property of our interacting system $(X^N(t))_{t\in\mathbb{R}^+} = (x_\xi(t))_{\xi\in\Omega_N}$ together with (0.69) to conclude that

$$\mathscr{L}(x_{\xi,0}(\sigma(N)), \dots, x_{\xi,k}(\sigma(N))|x_{\xi,j}(\Delta(N)) = \rho) \underset{N\to\infty}{\Rightarrow} \Lambda_\rho^{j,k}, \tag{5.50}$$

and note that this convergence is uniform in ρ.

Let $\mathbf{d}$ be a metric for the weak convergence on $\mathscr{P}(\mathscr{P}([0,1]))$. Then by the triangle inequality

(5.51)

$$\begin{aligned}
&\mathbf{d}(\mathscr{L}(x_{\xi,0}(\sigma(N)),\ldots,x_{\xi,k}(\sigma(N))),\Lambda_\theta^{\infty,k})\\
&\quad\le \mathbf{d}\Bigg(\Bigg[\int\mathscr{L}(x_{\xi,0}(\sigma(N)),\ldots,x_{\xi,k}(\sigma(N))|x_{\xi,j}(\Delta(N))=\rho)\\
&\qquad\qquad\cdot P(x_{\xi,j}(\Delta(N))\in d\rho),\ \int\Lambda_\rho^{j,k}P(x_{\xi,j}(\Delta(N))\in d\rho)\Bigg]\Bigg)\\
&\qquad+\mathbf{d}\Bigg(\Bigg[\int\Lambda_\rho^{j,k}P(x_{\xi,j}(\Delta(N))\in d\rho),\ \int\Lambda_\rho^{\infty,k}P(x_{\xi,j}(\Delta(N))\in d\rho)\Bigg]\Bigg)\\
&\qquad+\mathbf{d}\Bigg(\Bigg[\int\Lambda_\rho^{\infty,k}P(x_{\xi,j}(\Delta(N))\in d\rho),\ \int\Lambda_\theta^{\infty,k}\Bigg]\Bigg).
\end{aligned}$$

To show that this is less than a given $\varepsilon>0$, first choose $j>k$ so that each of the second and third terms is less than $\frac{\varepsilon}{3}$. This can be done for the second term by (6.24) and for the third term by (0.68) and the continuity in ρ of $\Lambda_\rho^{\infty,k}$. Finally given this j, choose N sufficiently large so that the first term is also less than $\frac{\varepsilon}{3}$. This can be done by (5.50). This implies that $\mathbf{d}(\mathscr{L}(x_{\xi,0}(\sigma(N)),\ldots,x_{\xi,k}(\sigma(N))),\Lambda_\theta^{\infty,k})\to 0$ as $N\to\infty$ thus completing the proof.

Proof of (0.71). Recall the definition (0.67). If we take ν_θ^N as the initial state of the system $X^N(t)$ (which gives the stationary system), then (0.71) is proved exactly as (0.70). The only additional observation required is to note that (0.68) and the definition (recall (0.14)) of ν_θ^N allow us to verify that $\mathrm{Var}_{\nu_\theta^N}(x_{\xi,|k|})\to 0$ as $k\to\infty$ uniformly in N.

6. Analysis of the interaction chain

The main purpose of this section is to prove Theorems 0.5, 0.6 and 0.7, which make assertions about the behavior of the interaction chains $(\hat\theta_\ell^k)_{\ell=-k-1,\ldots,0}$ as the parameter k tends to infinity. Recall that these chains were obtained by picking a point $\bar\xi$ and considering averages over blocks of components of the type $\{\xi\,|\,d(\bar\xi,\xi)\le|\ell|\}$ with $\ell\in\{-k-1,\ldots,0\}$ in the time scale $\beta_k(N)$ as $N\to\infty$. This section is divided into three subsections dealing with Theorems 0.5, 0.6 and 0.7 respectively.

(a) **The stability versus clustering dichotomy, Entrance laws.** The principal tool in this section will be an identity which we formulate and prove in 6.1 below. This will be used to prove Propositions 6.1–6.3 which will then yield Theorem 0.5.

Lemma 6.1. *Let f be a function in $C([0,1])$ and denote by x, θ elements of $\mathscr{P}([0,1])$, where x is a random element, θ a constant. Then*

$$\mathrm{Var}_{\mu_\theta^{k,j}}(\langle x,f\rangle)=\left(\sum_{n=j}^{k}\frac{1}{c_n}\right)E_{\mu_\theta^{k,0}}(\mathrm{Var}_x(f)).\quad\square \tag{6.1}$$

Proof. Step 1. Define a function $F_f : \mathscr{P}([0,1]) \to \mathbb{R}$ by setting

$$F_f(x) = \langle x, f\rangle^2. \tag{6.2}$$

A straightforward calculation shows

$$\frac{\partial F_f}{\partial x}(u) = 2f(u)\langle x, f\rangle, \quad \frac{\partial^2 F_f}{\partial x \partial x} = 2f(u)f(v). \tag{6.3}$$

We shall use this to calculate $E_{\Gamma_\theta^c}(E_f)$ which is the important term in determining $E_{\mu_\theta^{k,j}}(\langle x, f\rangle^2)$. Recall (0.7) for definitions and notation. Let Γ_θ denote the unique equilibrium of the Fleming-Viot process with immigration from the source θ at rate c and resampling constant d. (See section 0(d),(iii).) Denote this process by x_t for the moment. Then we must have

$$E_{\Gamma_\theta}(F_f(x_t)) = E_{\Gamma_\theta}(F_f(x_0)) \quad \forall t \in \mathbb{R}^+ \tag{6.4}$$

which implies, using the characterization of the process via a martingale problem (see Lemma 4.10) and relation (6.3):

$$\begin{aligned} 0 = E_{\Gamma_\theta}\Big(c\langle x, f\rangle \int_{[0,1]} f(u)(\theta(du) - x(du)) \\ + d\int_{[0,1]}\int_{[0,1]} f(u)f(v)Q_x(du, dv)\Big). \end{aligned} \tag{6.5}$$

Evaluation of the terms in (6.5) gives (recall $E^{\Gamma_\theta}\langle x, f\rangle = \langle \theta, f\rangle$):

$$\begin{aligned} 0 &= c\langle \theta, f\rangle^2 - cE_{\Gamma_\theta}\langle x, f\rangle^2 + dE_{\Gamma_\theta}\langle x, f^2\rangle - dE_{\Gamma_\theta}\langle x, f\rangle^2 \\ &= -c \cdot \mathrm{Var}_{\Gamma_\theta}(\langle x, f\rangle) + d \cdot E_{\Gamma_\theta}(\mathrm{Var}_x(f)) \end{aligned} \tag{6.6}$$

and consequently we have the following two useful identities:

$$\begin{aligned} \mathrm{Var}_{\Gamma_\theta}(\langle x, f\rangle) &= \frac{d}{c} E_{\Gamma_\theta}(\mathrm{Var}_x(f)), \\ E_{\Gamma_\theta}(\langle x, f\rangle^2) &= \langle \theta, f\rangle^2 + \frac{d}{c}E_{\Gamma_\theta}(\mathrm{Var}_x(f)). \end{aligned} \tag{6.7}$$

Step 2. In section 2 (2.5) we calculated that

$$E_{\Gamma_\theta}(d\,\mathrm{Var}_x(f)) = d'\,\mathrm{Var}_\theta(f) \quad \text{with } d' = \frac{cd}{c+d}. \tag{6.8}$$

Combining this with the two relations in (6.7) yields

$$\mathrm{Var}_{\Gamma_\theta}(\langle x, f\rangle) = d'\frac{1}{c}\,\mathrm{Var}_\theta(f), \tag{6.9}$$

$$E_{\Gamma_\theta}(\langle x, f\rangle^2) = \langle \theta, f\rangle^2 + d'\frac{1}{c}\,\mathrm{Var}_\theta(f). \tag{6.10}$$

In order to exploit the latter relations we now apply the definition of $\mu_\theta^{k,j}$ and write
(6.11)

$$E_{\mu_\theta^{k,j}}(\langle x, f\rangle^2) = \int_{\mathscr{P}([0,1])} \cdots \int_{\mathscr{P}([0,1])} \Gamma_\theta^k(d\theta_1)\Gamma_{\theta_1}^{k-1}(d\theta_2)\cdots\Gamma_{\theta_{k-j}}^j(dx)\langle x, f\rangle^2.$$

Now we successively use (6.10), (6.8) and the recursion formula for d_k to get

$$\operatorname{Var}_{\mu_\theta^{k,j}}(\langle x, f\rangle) = \left(\sum_{n=j}^{k} \frac{1}{c_n}\right) d_k \operatorname{Var}_\theta(f). \tag{6.12}$$

By the definition of d_k and (2.5):

$$E_{\mu_\theta^{k,0}}(\operatorname{Var}_x(f)) = d_k \operatorname{Var}_\theta(f). \tag{6.13}$$

Inserting (6.13) in the right-hand side of (6.12) gives the assertion of Lemma 6.1. □

Next we come to a sequence of Propositions 6.1, 6.2 and 6.3. We obtain the convergence of $\mu^{k,0}$ as $k \to \infty$ in Propositions 6.1 and 6.2 while Proposition 6.3 studies the whole process $(\hat{\theta}_j^k)_{j=-k-1,\dots,0}$ as $k \to \infty$.

We start with the case in which $\hat{a}$ is recurrent.

Proposition 6.1. *If* $\sum c_k^{-1} = +\infty$, *then*

$$\mu_\theta^{k,0} \underset{k\to\infty}{\Rightarrow} \int_{[0,1]} \delta_{(\delta_u)} \theta(du) := \mu_\theta^{\infty,0}. \quad \square \tag{6.14}$$

Proof. Since f in (6.1) is bounded we can use (6.1) to estimate as follows:

$$2\|f\|_\infty^2 \geq \left(\sum_{n=0}^{k} \frac{1}{c_n}\right) E_{\mu_\theta^{k,0}}(\operatorname{Var}_x(f)). \tag{6.15}$$

This implies that

$$E_{\mu_\theta^{k,0}}(\operatorname{Var}_x(f)) \underset{k\to\infty}{\to} 0 \tag{6.16}$$

and hence since f is bounded and $(\mathscr{P}(\mathscr{P}([0, 1])))$ is compact, we conclude that for every weak limit point $\tilde{\mu}_\theta^{\infty,0}$ of $\{\mu_\theta^{k,0}\}$ we have

$$E_{\tilde{\mu}_\theta^\infty}(\operatorname{Var}_x(f)) = 0 \quad \forall f \in C([0, 1]). \tag{6.17}$$

This implies that $\tilde{\mu}_\theta^\infty$ is concentrated on point measures; that is, for $M = \{\delta_u \mid u \in [0, 1]\}$ we know that

$$\tilde{\mu}_\theta^{\infty,0}(M) = 1. \tag{6.18}$$

But we have more information available. Since $\int \langle x, f\rangle \Gamma_\theta^k(dx) = \langle \theta, f\rangle$ for all $k \in \mathbb{N}$:

$$E_{\mu_\theta^{k,0}}(\langle x, f\rangle) = \langle \theta, f\rangle \quad \forall f \in C([0, 1]), \ k \in \mathbb{N}. \tag{6.19}$$

Hence

$$E_{\tilde{\mu}_\theta^{\infty,0}}(\langle x, f\rangle) = \langle \theta, f\rangle \quad \forall f \in C([0, 1]). \tag{6.20}$$

Using (6.18) we can conclude that there exists a law H_θ on $[0, 1]$ such that

$$\tilde{\mu}_\theta^{\infty,0} = \int \delta_{(\delta_u)} H_\theta(du). \tag{6.21}$$

With (6.20) we obtain the following condition on $H_\theta(\cdot)$:

$$\langle \theta, f\rangle = \int H_\theta(du) f(u) \quad \forall f \in C([0, 1]) \tag{6.22}$$

so that

$$H_\theta = \theta. \tag{6.23}$$

Therefore in particular $\tilde{\mu}_\theta^{\infty,0}$ is independent of the convergent subsequence we choose from $\{\mu_\theta^{k,0}\}_{k\in\mathbb{N}}$.

Hence we can conclude that convergence takes place and we have proved the Proposition 6.1.

In the case in which $\hat{a}$ is transient we prove (recall $M = \{\delta_u \mid u \in [0, 1]\}$)

Proposition 6.2. *Assume that* $\sum c_k^{-1} < \infty$ *and* $\theta \notin M$. *Then*

$$\mu_\theta^{j,k} \underset{j\to\infty}{\to} \mu_\theta^{\infty,k} \text{ uniformly in } \theta, \tag{6.24}$$

$$\mu_\theta^{\infty,k}(M) = 0, \tag{6.25}$$

$$E_{\mu_\theta^\infty}\langle x, f\rangle = \langle \theta, f\rangle \quad \forall f \in C([0, 1]). \quad \square \tag{6.26}$$

Proof. Step 1. The first observation is that $(\hat{\theta}_\ell^j)_{\ell=-j-1,\dots,0}$ is such that the transition kernels at times $-k, \dots, 0$ are independent of j (if $j > k$). Furthermore $\theta \to \Gamma_\theta$ is a continuous map (see (4.21)). In fact even the composition map $\theta \to (\Gamma^k \circ \Gamma^{k-1} \circ \cdots \circ \Gamma^0)(\theta, \cdot)$ is continuous uniformly in k (compare relation (4.23)). Therefore in order to show (6.24) it suffices to prove that $\hat{\theta}_k^j$ for k fixed does not change much if $j \to \infty$. Since $E_{\mu_\theta^{j,k}}(\langle x, f\rangle) = \langle \theta, f\rangle$, this is true if the following relation holds:

$$\left[\sup_{\theta\in\mathscr{P}([0,1])} \sup_{j\ge k} \mathrm{Var}_{\mu_\theta^{j,k}}(\langle x, f\rangle)\right] \underset{k\to\infty}{\to} 0 \quad \forall f \in C([0, 1]). \tag{6.27}$$

But by (6.1)

$$\sup_{\theta\in\mathscr{P}([0,1])} \sup_{j\ge k} \mathrm{Var}_{\mu_\theta^{j,k}}(\langle x, f\rangle) \le \|f\|_\infty^2 \sum_{n=j}^\infty \frac{1}{c_n}. \tag{6.28}$$

By the assumption $\sum c_k^{-1} < \infty$ we know that the right-hand side of (6.28) tends to 0 as $k \to \infty$ which finishes the proof of (6.24).

Step 2. To see why (6.25) holds note first that by the result (0.25) on Γ_θ we know that

$$\forall \theta \notin M : \Gamma_\theta^j(M) = 0 \quad \forall j \in \mathbb{N}. \tag{6.29}$$

From here one can use (6.27) to get the assertions by contradiction using that if $\theta \notin M$ then for every $y \in [0, 1]$ there exists an $f \ge 0$, $f \in C([0, 1])$ and an interval $(y - \varepsilon, y + \varepsilon)$ with $\varepsilon > 0$ such that

$$|\langle \theta, f\rangle - \langle \delta_x, f\rangle| \ge \delta > 0 \quad \forall x \in (y - \varepsilon, y + \varepsilon).$$

Consequently by (6.27) $\lim_{k\to\infty}(\text{w-}\lim_{j\to\infty} \mu_\theta^{j,k})(M) = 0$. Then use (6.29) to conclude $\mu_\theta^{j,0}(M) = 0$. (See also Dawson and Greven (1993b).) This proves (6.25). Relation (6.26) follows immediately from the bounded convergence theorem and (6.19). This finishes the proof of Proposition 6.2.

Next we rephrase the previous result in terms of entrance laws. We define the law of $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$ as the following limit:

$$\mathscr{L}((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}) = \operatorname*{w-lim}_{j\to\infty} \mathscr{L}((\hat{\theta}_k^j)_{k\in[-j-1,\dots,0]}).$$

This limit exists since we have a sequence of time inhomogeneous chains with the same transition kernels only differing by the condition $\hat{\theta}_{-k-1}^{(k)} = \theta$. Then we argue as in the proof of (6.24) to obtain the existence of the limit above. We prove

Proposition 6.3. (a) $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$ *is a Markov chain with transition kernel* $K_k(\theta, d\theta') = \Gamma_\theta^{|k|}(d\theta')$.

(b) *If* $\sum_k c_k^{-1} = +\infty$, *then* $\mathscr{L}((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-})$ *is concentrated on the paths which are identically* δ_U *where* U *is distributed according to* θ.

(c) *If* $\sum_k c_k^{-1} < +\infty$, *then* $\mathscr{L}((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-})$ *is a martingale and* θ_k^∞ *converges to* θ *as* $k \to -\infty$. □

Proof. (a) This follows immediately from the fact that $(\hat{\theta}_k^j)_{k=-j-1,\dots,0}$ has this property which is given in the proof of (6.24).

(b) This follows immediately from the fact that, according to Proposition 6.1, $\mathscr{L}(\hat{\theta}_0^\infty) = \int \delta_u \theta(du)$ and since $\Gamma_{\delta_u}^k = \delta_u$, $\Gamma_x^k(M) = 0$ for $x \notin M$.

(c) The martingale property follows from the Markov property of $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$ and $E_{\Gamma_\theta^k}(\langle x, f\rangle) = \langle \theta, f\rangle$. Finally since $\hat{\theta}_{-j-1}^j = \theta$, the assertion on convergence follows from (6.28) and the reverse martingale convergence theorem.

(b) **The rescaled interaction chain (Proof of Theorem 0.6).** The object of our investigation is now the rescaled interaction chain $(\hat{\theta}_{f_\alpha(k)})_{\alpha\in[0,1]}$. We abbreviate

$$\tilde{\theta}_\alpha^{(k)} = \hat{\theta}_{f_\alpha(k)}^{(k)}, \qquad \alpha \in [0, 1],\ k \in \mathbb{N}. \tag{6.30}$$

We shall proceed in five steps. Steps 1–3 prove Theorem 0.6(a), and step 4 part (b) and step 5 part (c). We shall show in step 1 how to derive the result of Theorem 0.6(a) from the following lemmata which are proved in step 3, after technical preparation in step 2.

Step 1. We start the analysis with the case

$$c_k \equiv c, \quad f_\alpha(k) = (1-\alpha)k. \tag{6.31}$$

Lemma 6.2. *The sequence of measures* $\{\mathscr{L}((\tilde{\theta}_\alpha^{(k)})_{\alpha\in[0,1]})\}_{k\in\mathbb{N}}$ *is tight in* $\mathscr{P}(D([0, 1); \mathscr{P}([0, 1])))$. □

As a second tool we prove a property of the weak limit points which characterizes them uniquely. Introduce for this purpose the time change $\alpha \to \log(1-\alpha)^{-1}$; that is, define the new process

$$\tilde{\tilde{\theta}}_\alpha^k = \tilde{\theta}_{1-\exp(-\alpha)}^k, \quad \alpha \in [0, \infty). \tag{6.32}$$

Lemma 6.3. *Every weak limit point as* $k \to \infty$ *of the sequence* $\mathscr{L}((\tilde{\tilde{\theta}}^{k}_\alpha)_{\alpha\in[0,\infty)})$ *has the property*:

For every $f \in C([0,1])$ *the following two processes are martingales with continuous paths*:

$$((\langle \tilde{\tilde{\theta}}^{(\infty)}_\alpha, f\rangle)_{\alpha\in[0,\infty)}: \tag{6.33}$$

$$(\langle \tilde{\tilde{\theta}}^{(\infty)}_\alpha, f\rangle^2 - \int_0^\alpha \mathrm{Var}_{\tilde{\tilde{\theta}}^{(\infty)}_\beta}(f)\, d\beta)_{\alpha\in[0,\infty)}. \quad \square$$

With these two lemmas the proof is completed as follows. According to Lemma 6.2 we can pick a convergent subsequence, and we denote the limit by $(\tilde{\theta}^\infty_\alpha)_{\alpha\in[0,1]}$. Applying the (invertible) time transformation of (6.32) we get via Lemma 6.3 a process $(\tilde{\tilde{\theta}}^{(\infty)}_\alpha)_{\alpha\in[0,\infty)}$ such that (6.33) holds. Since by Lemma 4.10 the martingale problem in (6.33) has a unique solution, namely, the Fleming-Viot process, we can conclude that $(\tilde{\tilde{\theta}}^{(k)}_\alpha)_{\alpha\in[0,\infty)}$ converges in law to the Fleming-Viot process on $\mathscr{P}([0,1])$, which finishes the proof of Theorem 0.6(a).

Step 2. Before we start the actual proof of Lemmas 6.2, 6.3, we prepare some tools. First of all note that, by Lemma 6.1, for $d_0 = 1$ we have

(6.34)

$$\begin{aligned}\mathrm{Var}(\langle \hat{\theta}^{(k)}_\ell, f\rangle) &= \left(\sum_{m=|\ell|}^{k} c_m^{-1}\right) d_k \,\mathrm{Var}_\theta(f)\\ &= (k+1-|\ell|)c^{-1}\frac{c}{c+k}\mathrm{Var}_\theta(f) = (1-|\ell|/(k+1))\,\mathrm{Var}_\theta(f)\frac{k+1}{k+c}.\end{aligned}$$

The latter equality follows from the fact that for $c_k = c$ we can solve the recursion formula $d_{n+1} = c_n d_n/(c_n + d_n)$ explicitly, namely $d_n = c/(c+n)$. The second observation is that $(F(\langle \theta^{(k)}_\ell, f\rangle))_{\ell=-k-1,\dots,0}$ is a submartingale for every convex function F on $\mathbb{R}$, and hence $E(F(\langle \hat{\theta}^{(k)}_\ell, f\rangle))$ is increasing in ℓ. This allows us to represent using (6.13) the interaction chain via a Brownian motion. (See Dawson and Greven (1993c, Lemma 5.2) for a proof.) In connection with (6.9) we can conclude

Lemma 6.4. *There exists for each* k *a sequence* $0 \le S_0 \le S_1 \le S_2 \le \cdots$ *of stopping times of a Brownian motion* $(W(t))_{t\ge 0}$ *such that the increments* $(S_{i+1}-S_i)_{i\in\mathbb{N}}$ *are independent and the following relations hold for* $\ell = -k, -k+1, \dots, 0$:

$$\langle \hat{\theta}^{(k)}_\ell, f\rangle = W(S_{k+\ell}), \tag{6.35}$$

$$E(S_{k+\ell+1} - S_{k+\ell}) = E(\langle \hat{\theta}^{(k)}_{\ell+1}, f\rangle - \langle \hat{\theta}^{(k)}_\ell, f\rangle)^2, \tag{6.36}$$

$$\begin{aligned} E(S_{k+\ell+1} - S_{k+\ell} | \hat{\theta}^{(k)}_\ell = \theta) &= E((\langle \hat{\theta}^{(k)}_{\ell+1}, f\rangle - \langle \theta, f\rangle)^2 | \hat{\theta}^{(k)}_\ell = \theta)\\ &= \frac{1}{c+\ell}\mathrm{Var}_\theta(f). \quad \square\end{aligned} \tag{6.37}$$

Since

$$\sum_{\ell=n}^{m} \frac{1}{c+\ell} \sim (\log m - \log n) \quad \text{for } n \to \infty,$$

using (6.37) and (6.34) we can obtain that, for $\beta_1, \beta_2 \in [0, 1]$ $(\beta_2 > \beta_1)$, we have

$$\lim_{k\to\infty} E(S_{[\beta_1 k]} - S_{[\beta_2 k]} | \hat{\theta}^k_{[(1-\beta_1)k]} = \theta) = \frac{\beta_2 - \beta_1}{1 - \beta_1} \operatorname{Var}_\theta(f) + o(\beta_2 - \beta_1). \tag{6.38}$$

This relation will be the major tool in the proof of Lemmas 6.2, 6.3 which are given in the next step.

Step 3. In this step we prove Lemmas 6.2 and 6.3.

Proof of Lemma 6.2. In Dawson and Greven (1993c, Lemma 5.2) it was shown that the processes $(S_{[\beta k]})_{\beta\in[0,1]}$, $k = 1, 2, \dots$, are tight (one uses the increasing property of $(S_k)_{k\in\mathbb{N}}$, relation (6.38) together with properties of the construction leading to Lemma 6.4). Therefore the processes $(W(S_{[\beta k]}))_{\beta\in[0,1]}$ are tight since $W(\cdot)$ has continuous paths. We can conclude that, for every $f \in C([0, 1])$, the processes $(\langle\tilde{\theta}^{(k)}_{[\beta k]}, f\rangle)_{\beta\in[0,1]}$ are tight. This implies the tightness of the sequence of processes $(\tilde{\theta}^{(k)}_{[\beta k]})_{\beta\in[0,1]}$.

Proof of Lemma 6.3. Denote by $(S^\infty_\beta)_{\beta\in[0,1]}$ a weak limit point of $(S_{[\beta k]})_{\beta\in[0,1]}$. Then $(W(S^\infty_\beta))_{\beta\in[0,1]}$ is a martingale with continuous paths. This process depends of course on the function $f \in C([0, 1])$ we have chosen for the construction. We need to vary f in the sequel and therefore we write $(S^{\infty,f}_\beta)_{\beta\in[0,1]}$ for this process generating the time transformation. Next choose a set $\mathscr{F} \subset C([0, 1])$ which is dense and countable. For every $f \in \mathscr{F}$, $(W(S^{\infty,f}_\beta))_{\beta\in[0,1]}$ is a martingale. Define the law of the process $(\tilde{\theta}^{(\infty)}_\beta)_{\beta\in[0,1]}$ by setting

$$\mathscr{L}(\langle\tilde{\theta}^{(\infty)}_\beta, f\rangle) = \mathscr{L}(W(S^{\infty,f}_{1-\beta})) \quad \forall f \in \mathscr{F}. \tag{6.39}$$

From the relation (6.38) we now obtain

$$\begin{aligned} &\operatorname{Var}(W(S^{\infty,f}_{\beta_2}) - W(S^{\infty,f}_{\beta_1}) | \tilde{\theta}^{(\infty)}_{1-\beta_1} = \theta) \\ &\quad = \frac{\beta_2 - \beta_1}{1 - \beta_1} \operatorname{Var}_\theta(f) + o(\beta_2 - \beta_1), \quad \mathscr{L}(\tilde{\theta}^{(\infty)}_{1-\beta_1})\text{-a.s.} \end{aligned} \tag{6.40}$$

In order to obtain a time homogeneous expression we use the time transformation

$$\beta = 1 - e^{-s} \tag{6.41}$$

and define

$$\tilde{\tilde{\theta}}^{(\infty)}_s = \tilde{\theta}^{(\infty)}_{\exp(-s)}. \tag{6.42}$$

Using (6.39) with the new time parameter the relation (6.40) becomes

$$\begin{aligned} &\operatorname{Var}(\langle\tilde{\tilde{\theta}}^{(\infty)}_{s_2} - \tilde{\tilde{\theta}}^{(\infty)}_{s_1}, f\rangle | \tilde{\tilde{\theta}}^{(\infty)}_{s_1} = \theta) \\ &\quad = \operatorname{Var}(W(S^{\infty,f}_{1-\exp(-s_2)}) - W(S^{\infty,f}_{1-\exp(-s_1)}) | \tilde{\tilde{\theta}}^{(\infty)}_{s_1} = \theta) \\ &\quad = (1 - \exp(-(s_2 - s_1))) \operatorname{Var}_\theta(f) \\ &\quad \underset{s_2\to s_1}{\sim} (s_2 - s_1) \operatorname{Var}_\theta(f). \end{aligned} \tag{6.43}$$

Therefore the process $(\tilde{\tilde{\theta}}_s^{(\infty)})_{s\in[0,\infty)}$ satisfies

$$(6.44) \qquad (\langle \tilde{\tilde{\theta}}_s^{(\infty)}, f\rangle)_{s\in[0,\infty)} \text{ is a martingale } \quad \forall f \in C([0,1]),$$

$$(6.45) \qquad \left((\langle \tilde{\tilde{\theta}}_s^{(\infty)}, f\rangle^2) - \int_0^s \operatorname{Var}_{\tilde{\tilde{\theta}}_t^{(\infty)}}(f)\, dt \right)_{s\in[0,\infty)} \text{ is a martingale.}$$

It is clear by (6.39) that the process $(\tilde{\tilde{\theta}}_s^{(\infty)})_{s\in[0,\theta)}$ has continuous paths. With the help of Lemma 4.10 then relations (6.44) and (6.45) finish the proof of Lemma 6.3.

Finally observe that replacing $c_k \equiv c$ by a sequence of $c_k \to c$ does not change the asymptotic relations we use in the proof. This completes the proof of Theorem 0.6(a).

Step 4. It remains to prove Theorem 0.6(b), that is, consider the case where (c_k) is not asymptotically constant. Observe that the martingales $\langle \hat{\theta}_\ell^{(k)}, f\rangle$, $\ell = -k-1, \dots, 0$, hit the set $M = \{\delta_u \mid u \in [0,1]\}$ with probability tending to one at time $\ell = [\alpha k]$ for all $f \in C[0,1]$ as $k \to \infty$ if $\operatorname{Var}\langle \hat{\theta}_\ell^{(k)}, f\rangle$ tends to $\operatorname{Var}_\theta(f)$ as $k \to \infty$, which is easily seen using Jensen's inequality on $(\int f(u)\hat{\theta}_\ell^{(k)}(du))^2$. On the other hand if $\operatorname{Var}\langle \hat{\theta}_\ell^{(k)}, f\rangle$ tends to 0 for $\ell = [\alpha k]$, the martingales remain constant with probability tending to 1 as $k \to \infty$. Applying these two facts we see from the formula (6.34) that we have for $n/m \sim \alpha \in (0,1)$:

$$(6.46) \qquad \left.\begin{array}{l}\text{fast growing clusters}\\ \text{slow growing clusters}\end{array}\right\} \text{ iff } \left(\sum_n^m c_k^{-1}\right) d_m \begin{cases} \underset{m\to\infty}{\longrightarrow} 0, \\ \underset{m\to\infty}{\longrightarrow} 1. \end{cases}$$

It remains to investigate the behavior of d_m as a function of $c_1, \dots, c_m$. One proves by induction

$$(6.47) \qquad d_{n+1} = \frac{d_0}{1 + d_0 \sum_0^n c_k^{-1}}.$$

Then we conclude that we have for $m/n \sim \alpha \in (0,1)$ and the assumptions on (c_k):

$$(6.48) \qquad \begin{array}{l} \text{fast growing clusters iff } \sum_m^n c_k^{-1} \Big/ \sum_0^{n-1} c_k^{-1} \to 0, \\[2ex] \text{slow growing clusters iff } \sum_m^n c_k^{-1} \Big/ \sum_0^{n-1} c_k^{-1} \to 1. \end{array}$$

This proves Theorem 0.6(b).

Step 5. Since we now have that $c_k = c^k$ for $c < 1$ we immediately get from (6.47) that

$$\frac{d_n}{c_n} \underset{n\to\infty}{\longrightarrow} c^{-1}.$$

Then by construction of the interaction chain we get that $(\theta_{-j-1+m}^{(j)})_{m\in\mathbb{N}}$ converges as $j \to \theta$ to a time homogeneous Markov chain on $\mathscr{P}([0,1])$ with transition kernel $\Gamma_\theta^{1/c}$. This proves part (c) of Theorem 0.6.

(c) **Spatial density of types: Proof of Theorem 0.7.** The proof proceeds in three steps. We start in step 1 to collect a number of properties of the kernels Γ_θ^k and of the paths of $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$. In steps 2, 3 we then prove part (a), respectively part (b) of the theorem.

Step 1. *Preliminaries.* Throughout this section we assume that $\sum_{k=0}^\infty c_k^{-1} < \infty$. We shall derive various properties of the process $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$ corresponding to an *entrance law.*

We first need some facts about the random variables corresponding to the law $\Gamma_\theta^k(\cdot)$ (recall (0.25)) which gives the transition kernel of $(\hat{\theta}_j^\infty)_{j\in\mathbb{Z}^-}$ at time $-k-1$. According to Theorem 0.3 we have the representation

$$\Gamma_\theta^k = \mathscr{L}\left(\sum_{j=1}^\infty M_j^*\delta_{U_j}\right) \tag{6.49}$$

where $(U_j)_{j\in\mathbb{N}}$ are i.i.d. θ-distributed and $M_j^* = V_j\prod_{i=1}^{j-1}(1-V_i)$ with $(V_j)_{j\in\mathbb{N}}$ independent Beta$(1, \gamma_k^{-1})$ distributed $(\gamma_k = d_k/c_k)$. On the right-hand side of (6.49) we have suppressed the k-dependence in the notation. Since our goal is to study the "concentration" of the M_j^*, it is useful to analyse the ℓ^2-norm of the sequence (M_j^*) of weights and the sums $\sum_{j=\ell}^\infty M_j^*$. Define the random variables $T^{(k)}$ by

$$T^{(k)} = \sum_{j=1}^\infty (M_j^*)^2 \tag{6.50}$$

and its moment generating function

$$\Phi_k(\lambda) = E(\exp(\lambda T^{(k)})). \tag{6.51}$$

Lemma 6.5. $\Phi_k(\lambda)$ *satisfies the equation*

$$\Phi_k(\lambda) = \int (e^{\lambda u^2}\Phi_k(\lambda(1-u)^2)(1-u)^{(\gamma_k^{-1}-1)}\gamma_k^{-1})\,du, \tag{6.52}$$

$$E(T^{(k)}) = \frac{\gamma_k}{1+\gamma_k}, \tag{6.53}$$

$$E[(T^{(k)})^2] = \frac{(\gamma_k^2 + 6\gamma_k^3)}{(1+\gamma_k)(1+2\gamma_k)(1+3\gamma_k)}, \tag{6.54}$$

$$\left(1 - \sum_{j=1}^n M_j^*\right) = \prod_{i=1}^n (1-V_i). \quad \square \tag{6.55}$$

Proof. Write

$$T^{(k)} = V_1^2 + (1-V_1)^2[V_2^2 + V_3^2(1-V_2)^2 + \cdots] \tag{6.56}$$

to obtain for two independent copies $T^{(k)}$, $\widetilde{T}^{(k)}$ the relation

$$T^{(k)} = V_1^2 + (1-V_1)^2\widetilde{T}^k. \tag{6.57}$$

Taking moment generating functions in (6.57) yields (6.52). By differentiating $\Phi(\lambda)$ we obtain the expressions

$$E(T^{(k)}) = \frac{E(V_1^2)}{2E(V_1) - E(V_1^2)},$$
$$E[(T^{(k)})^2] = \frac{E(V_1^4) + 2E(V_1^2) \cdot E(T^{(k)})}{1 - E((1-V_1)^4)}.$$

Substituting the moments of V_1 we obtain (6.53), (6.54). The relation (6.55) follows immediately by induction over n from the representation of M_j^* as $V_j \prod_{i=1}^{j-1}(1-V_i)$.

Now we come to some martingales related to $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$. In order to formulate the next lemma we write $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$ in the form

$$\hat{\theta}_k^\infty = \sum_{j=1}^\infty \widetilde{M}_j^{(k)} \delta_{U_j} \quad \text{with } (U_j)_{j\in\mathbb{N}} \text{ i.i.d. } \theta\text{-distributed.} \tag{6.58}$$

This representation is possible for the following reason. Using the fact that the kernels Γ_θ^k have the form given in (6.49), which gives immediately such a representation for $(\hat{\theta}_\ell^\infty)_{\ell=-k-1,\dots,0}$. Furthermore since $\hat{\theta}_{-k}^\infty$ is a chain with transition kernels given by Γ_θ^k, we know that the states of this chain are atomic measures, where no new atoms appear as k varies. It remains to show that these atoms arise from i.i.d. sampling from θ. Since we can construct for every k $(\hat{\theta}_\ell^k)$ with one sequence $(U_j)_{j\in\mathbb{N}}$ of i.i.d. θ-distributed random variables in the first time step (at $-k-1 \to -k$), we know from (6.28) that letting $k \to \infty$ preserves the form of the random measures forming the states of the chain $\hat{\theta}_k^\infty$. (Now recall the definitions (0.53), (0.54) for $R^{(k)}$, $M_j^{(k)}$.) Note that $M_j^{(k)}$ is nonincreasing in j whereas $\widetilde{M}_j^{(k)}$ is the mass at level k assigned to the atom located at U_j where the $\{U_j\}$ is the same sequence for all k.)

Lemma 6.6. *For every* $j \in \mathbb{N}$

$$(\widetilde{M}_j^{(k)})_{k\in\mathbb{Z}^-} \textit{ is a martingale with respect to the filtration generated by } ((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}). \tag{6.59}$$

$$(R^{(k)})_{k\in\mathbb{Z}^-} \textit{ is a submartingale (with respect to the filtration generated by } ((\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-})) \textit{ satisfying} \tag{6.60}$$

$$E(R^{(k)}|R^{(k-1)}) = \frac{1}{1+\gamma_k} R^{(k-1)} + \frac{\gamma_k}{1+\gamma_k}, \quad \gamma_k := \frac{d_{|k|}}{c_{|k|}}. \quad \square \tag{6.61}$$

Proof. We start with the following representation of $\widetilde{M}_k^{(k)}$ which follows from (6.49) and the fact that $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$ is a Markov chain with transition kernels $K_{-k}(u, dv) = \Gamma_u^{k-1}(dv)$ at time $-k$. For every $j \in \mathbb{N}$:

$$\widetilde{M}_j^{(k)} = \sum_i \mathbf{1}_{U_j}(\widetilde{U}_i) M_i^*, \quad \{M_i^*\}_{i\in\mathbb{N}} \text{ as in (6.49) and } (\widetilde{U}_i)_{i\in\mathbb{N}} \text{ i.i.d. } \hat{\theta}_{k-1}^\infty\text{-distributed, independent of } \{M_i^*\}_{i\in\mathbb{N}}. \tag{6.62}$$

Consequently

$$E(\widetilde{M}_j^{(k)} | \hat{\theta}_{k-1}^{\infty}, \hat{\theta}_{k-2}^{\infty}, \ldots) = E(\widetilde{M}_j^{(k)} | \hat{\theta}_{k-1}^{\infty}) = \widetilde{M}_j^{(k-1)}, \tag{6.63}$$

where we use the facts that $\{M_i^*\}_{i\in\mathbb{N}}$ and $(\widetilde{U}_i)_{i\in\mathbb{N}}$ are independent, that $\sum_i M_i^* = 1$, and that from the representation (6.49) we know $E(\mathbf{1}_{U_j}(\widetilde{U}_i)) = \widetilde{M}_j^{(k-1)}$ where the expectation is taken with respect to the $(\widetilde{U}_i)_{i\in\mathbb{N}}$. This proves (6.59).

Next we use the fact that $((\widetilde{M}_j^{(k)})^2)_{k\in\mathbb{Z}^-}$ is a submartingale (cf. (6.59)) for each j; hence the sum over j is a submartingale. Furthermore $\sum_j (\widetilde{M}_j^{(k)})^2 = \sum_j (M_j^{(k)})^2$ (recall (0.53) for definitions) which gives (6.60).

The relation (6.61) follows from the calculation below:

$$E(R^{(k)} | \hat{\theta}_{k-1}^{\infty}, \hat{\theta}_{k-2}^{\infty}, \ldots) = E(R^{(k)} | \hat{\theta}_{k-1}^{\infty}) = \sum_j \sum_{i,\ell} E(\mathbf{1}_{U_j}(\widetilde{U}_i)\mathbf{1}_{U_j}(\widetilde{U}_\ell) M_i^* M_\ell^*) \tag{6.64}$$

where expectation is taken over $((\widetilde{U}_k)_{k\in\mathbb{N}}, (M_k^*)_{k\in\mathbb{N}})$, which is distributed as in (6.62). Here we again used (6.49).

$$\begin{aligned}
\text{r.h.s. } (6.64) &= \sum_j \left(\sum_{i,\ell} E[\mathbf{1}_{U_j}(\widetilde{U}_i)\mathbf{1}_{U_j}(\widetilde{U}_\ell)] \cdot E[M_i^* M_\ell^*] \right) \\
&= \sum_j \left((\widetilde{M}_j^{(k-1)})^2 \sum_{i\neq\ell} E[M_i^* M_\ell^*] + (\widetilde{M}_j^{(k-1)}) \sum_i E[M_i^{*2}] \right) \\
&= \frac{1}{1+\gamma_k} \sum_j (\widetilde{M}_j^{(k-1)})^2 + \frac{\gamma_k}{1+\gamma_k} \\
&= \frac{1}{1+\gamma_k} R^{(k-1)} + \frac{\gamma_k}{1+\gamma_k}.
\end{aligned} \tag{6.65}$$

The last equality uses the fact that $\sum_i E[M_i^{*2}] = \gamma_k/(1+\gamma_k)$ which was shown at the end of the proof of Lemma 2.2, and $\sum_j \widetilde{M}_j^{(k-1)} = 1$, $\sum_i M_i^* = 1$.

As an application of the previous lemmas we obtain the useful formula:

Lemma 6.7.

$$ER^{(k)} = \left(1 - \prod_{\ell=k}^{-\infty} (1+\gamma_\ell)^{-1}\right). \tag{6.66}$$

Proof. From the relation (6.61) we get

$$ER^{(k+1)} = \frac{1}{1+\gamma_{k+1}} ER^{(k)} + \frac{\gamma_{k+1}}{1+\gamma_{k+1}} \tag{6.67}$$

so that we get the recursion relation (recall $k \in \mathbb{Z}^-$!)

$$ER^{(k)} = (1+\gamma_{k+1}) ER^{(k+1)} - \gamma_{k+1}. \tag{6.68}$$

By iteration this leads to

$$ER^{(k)} - 1 = \left[\prod_{j=k+1}^{0} (1+\gamma_j)\right] (ER^{(0)} - 1). \tag{6.69}$$

Since we know that, as $k \to -\infty$, $ER^{(k)}$ tends to zero, we obtain a formula for $R^{(0)}$ and conclude that

$$ER^{(k)} = \left(1 - \prod_{\ell=k}^{-\infty}(1+\gamma_\ell)^{-1}\right). \tag{6.70}$$

Finally we come to the path properties of $(\hat{\theta}_k^\infty)_{k\in\mathbb{Z}^-}$.

Lemma 6.8. *Assume that* $\sum c_k^{-1} < \infty$.

(a) *If* $\theta \in \mathscr{P}([0, 1])$ *is nonatomic or has countably many atoms, then a.s.* $\hat{\theta}_k^\infty$ *has countably many atoms and* $\hat{\theta}_k^\infty$, $\hat{\theta}_j^\infty$ *have the same atoms for every* $k, j \in \mathbb{Z}^-$.

(b) *If* $\theta \in \mathscr{P}([0, 1])$ *has exactly* n *atoms, then* $\hat{\theta}_k^\infty$ *has* n *atoms for every* $k \in \mathbb{Z}^-$. □

Proof. We first prove the second part (b). Start with the case in which θ has two atoms. Then by Theorem 0.5 we know that $\hat{\theta}_k^\infty$ attaches positive weight to both atoms ($\mu_\theta^{\infty,0}(M) = 0!$). To finish by reducing the case of n atoms to the simple case of two atoms we have to use the following decomposition property of the kernels $\Gamma_\theta^k(\cdot)$: If we divide the type space in the form $\sum_{i\in J} A_i = [0, 1]$, $A_i \cap A_j = \varnothing$ for $i \neq j$, and define the map $\Lambda : \mathscr{P}([0, 1]) \to \mathscr{P}(J)$ by

$$\Lambda(x)(\{j\}) = x(A_j),$$

then $(\Lambda(\hat{\theta}_k^\infty))_{j\in\mathbb{Z}^-}$ is the interaction chain for the system of interacting Fleming-Viot processes where the components have state space J. This however follows immediately from the duality relation in section 3(a) read for indicator functions. Namely from here we can read off first that the evolution of the new interaction system $\{\Lambda(x_\xi(t))\}_{\xi\in\Omega_N}$ is a system of interacting Fleming-Viot processes on $(\mathscr{P}(J))^{\Omega_N}$. Then the above statement about the interaction chain is immediate.

The part (a) follows again using the above decomposition property. First of all no atom in $\hat{\theta}_k^{(\infty)}$ can disappear as time progresses (recall 0.25(a)). On the other hand we know from (6.58) that $\hat{\theta}_k^{(\infty)}$ is atomic, which finishes the proof.

Step 2. Proof of Theorem 0.7, part (a). According to (6.60), $(R^{(k)})_{k\in\mathbb{Z}^-}$ is a bounded submartingale. Therefore by the reverse submartingale convergence theorem we know that

$$R^{(k)} \to R^{(-\infty)} \text{ as } k \to -\infty, \text{ and in } L_1. \tag{6.71}$$

The proof is now by contradiction. Denote by $j(k)$ the index of the maximal weight in $(M_j^{(k)})_{j\in\mathbb{N}}$. Assume for the moment that $P(R^{(-\infty)} > \varepsilon) > \varepsilon$ for some $\varepsilon > 0$. Then since $\sum_j M_j^{(k)} = 1$ this implies that there exists k_0 such that

$$P(M_{j(k)}^{(k)} > \varepsilon/2) > \varepsilon/2 \quad \forall k \geq k_0. \tag{6.72}$$

Since we know (see Proposition 6.3) that $\hat{\theta}_k^\infty \Rightarrow \theta$ in distribution as $k \to -\infty$, therefore for any $0 \leq a < b \leq 1$,

$$\sum_{j\,:\,U_j\in[a,b)} M_j^{(k)} \underset{k\to-\infty}{\longrightarrow} \theta([a, b)) \text{ in probability.}$$

Now divide $[0, 1]$ into a finite number of subintervals $[a_i, b_i)$ with $\theta([a_i, b_i)) < \varepsilon^2/16$. Then

$$E\left[\max_{\ell} \sum_{j\,:\,U_j\in[a_\ell, b_\ell)} M_j^{(k)}\right] \le \varepsilon^2/8 \quad \text{for all sufficiently large } k. \tag{6.73}$$

But (6.72) implies that

$$\operatorname{Prob}\left(\max_{\ell}\sum_j M_j^{(k)}\mathbf{1}_{[a_\ell, b_\ell)}(U_j) > \varepsilon/2\right) > \varepsilon/2 \quad \forall k > k_0$$

and hence

$$E\left[\max_{\ell} \sum_{j\,:\,U_j\in[a_\ell, b_\ell)} M_j^{(k)}\right] > \varepsilon^2/4. \tag{6.74}$$

(6.73) and (6.74) yield a contradiction thus completing the proof that $R^{(-\infty)} = 0$ a.s.

Step 3. Proof of Theorem 0.7, part (b).

Step i. Upper bound for the number of types.
It suffices to find a lower bound for $S_\ell^{(k)}$ such that for $\ell = \ell(k)$ the lower bound converges to 1 in probability for suitably chosen $\ell(k)$. To get such a lower bound take the representation (6.62) and estimate as follows: (We suppress the *k-dependence* of $(\widetilde{U}_i)_{i\in\mathbb{N}}$ and $(M_i^*)_{i\in\mathbb{N}}$ in the notation; recall (6.49) for definitions.)

$$\begin{aligned}\sum_{j=1}^n M_j^{(k)} \ge \sum_{\ell=1}^n \widetilde{M}_{j_\ell}^{(k)} &= \sum_i \sum_{\ell=1}^n \mathbf{1}_{U_{j_\ell}}(\widetilde{U}_i) M_i^* \quad \text{for any } \{j_1, \dots, j_n\} \\ &\ge \sum_{i=1}^n M_i^*,\end{aligned} \tag{6.75}$$

where for the last inequality we choose $U_{j_\ell} = \widetilde{U}_\ell$ for $\ell = 1, \dots, n$. Therefore we need to prove that for $\ell(k)/c_k \to \infty$:

$$\sum_{j=1}^{\ell(k)} M_j^* \underset{k\to-\infty}{\longrightarrow} 1 \text{ in probability.} \tag{6.76}$$

This we achieve as follows. Apply relation (6.55) to write the generating function of $(1 - \sum_{j=1}^n M_i^*) := Q_n$ for $\lambda > \gamma_k^{-1}$ as follows:

$$E[(Q_n)^\lambda] = (E(1 - V_1)^\lambda)^n = (E\exp(\lambda\log(1 - V_1)))^n = \left(1 - \frac{\gamma_k\lambda}{1 + \gamma_k\lambda}\right)^n \tag{6.77}$$

(the last equality follows by explicit calculation from the fact that $(1 - V_1)$ is distributed as $\operatorname{Beta}(\gamma_k^{-1}, 1)$ since V_1 is $\operatorname{Beta}(1, \gamma_k^{-1})$). We know that (see (6.47))

$$\gamma_k \sim c_{|k|}^{-1}\left(d_0^{-1} + \sum_{j=0}^\infty c_j^{-1}\right)^{-1}, \qquad \gamma_k \to 0 \text{ as } k \to -\infty. \tag{6.78}$$

Hence the right-hand side of (6.77) converges to zero for all $\lambda \geq 0$ provided that $n = \ell(k)$ satisfies $\ell(k)/c_{|k|} \underset{k\to-\infty}{\to} \infty$. This proves the claim.

Step ii. Lower bound for the number of types.
We need to find a sequence $\ell(k)$ such that $S^k_{\ell(k)}$ converges to zero in probability as $k \to -\infty$. It suffices to prove that $E(S^k_{\ell(k)})$ converges to zero as $k \to -\infty$. By Cauchy-Schwarz one has the inequality

$$E(S^k_{\ell(k)}) = E\left(\sum_{j=1}^{\ell(k)} M_j^{(k)}\right) \leq (\ell(k))^{1/2}\left(E\sum_{j=1}^{\ell(k)}(M_j^{(k)})^2\right)^{1/2} \leq (\ell(k)ER^{(k)})^{1/2}.$$

By Lemma 6.6
(6.79)

$$ER^{(k)} = \left(1 - \prod_{\ell=k}^{\infty}(1+\gamma_k)^{-1}\right) = \left(1 - \exp\left(-\sum_{\ell=k}^{\infty}\log(1+\gamma_\ell)\right)\right) \leq \sum_{\ell=k}^{\infty}\gamma_\ell.$$

Combining the last two relations together with (6.78) leads to

$$E(S^k_{\ell(k)}) \leq \left(\ell(k)\sum_{\ell=k}^{\infty}\gamma_\ell\right)^{1/2} \leq C\left(\ell(k)\sum_{\ell=k}^{\infty}c_\ell^{-1}\right)^{1/2} \tag{6.80}$$

which proves the assertion.

7. Fixed point properties of interacting Fleming-Viot processes. Universality

Let Q^G_θ and Q^H_θ be defined as in (0.57), (0.58). The objective of this section is to prove relations (0.59)–(0.61) as well as Theorems 0.8 and 0.9

(a) Proof of (0.59)–(0.61).

Lemma 7.1. *For every* $G \in \mathscr{G}$, $\mathscr{F}Q^G_x(\theta, du, dv)$ *satisfies*:

$$\int \Gamma^G_\theta(dx)G(x, u, v)x(du)x(dv) \ll \theta(du)\theta(dv). \tag{7.1}$$

Let $H(\theta, u, v)$ *denote the Radon-Nikodym derivative of the left-hand side with respect to the right-hand side of* (7.1). *Then*

(7.2)
$$\mathscr{F}_c Q^G_x(\theta, du, dv) = \int H(\theta, u, z)\theta(dz)\theta(du)\delta_u(dv) - H(\theta, u, v)\theta(du)\theta(dv),$$

$$\|H\|_\infty \leq \|G\|_\infty, \quad \text{and} \tag{7.3}$$

$$\int (\langle x, f\rangle)^2\Gamma^{G,c}_\theta(dx) = \frac{1}{c}\int Q^G_\mu(du, dv)f(u)f(v)\Gamma^{G,c}_\theta(d\mu) + \langle\theta, f\rangle^2. \tag{7.4}$$

Proof. Recall that $\Gamma^G_\theta(dx)$ is the probability law of a random probability measure and since it is assumed to be a stationary measure for the martingale problem associated with $L^{G,c}_\theta$ it follows that $\int \mu\Gamma^G_\theta(d\mu) = \theta$. From this it follows that

$$\int_{A\times B} G(x, u, v)x(du)x(dv) = 0, \quad \Gamma^G_\theta\text{-a.e. } x$$

if $\theta(A) = 0$ or $\theta(B) = 0$. From this we can conclude (7.1). Given any $f \in C([0, 1]^2)$ we then have

$$
\begin{aligned}
&\iint f(u,v)\mathscr{F}_c Q_x^G(\theta, du, dv)\\
&= \iint (f(u,u) - f(u,v)) \int G(x,u,v)x(du)x(dv)\Gamma_\theta^G(dx)\\
&= \iint (f(u,u) - f(u,v))H(\theta,u,v)\theta(du)\theta(dv)\\
&= \iint f(u,v)Q_\theta^H(du,dv),
\end{aligned}
\tag{7.5}
$$

which yields (7.2). (7.3) follows immediately from the definition of H and $\|\cdot\|$.

It remains to derive the relation (7.4). The proof is parallel to the proof of equation (6.1). For every solution of the martingale problem in (2.1) in section 2, with $\theta(t) \equiv \theta$ and any of its versions stationary in time, we must have the following relation ($x \in \mathscr{P}([0,1])$ denotes the random element with respect to which the expectation E is taken):

$$
\begin{aligned}
0 &= c\langle\theta, f\rangle^2 - cE_{\Gamma_\theta^{G,c}}\langle x, f\rangle^2 + E_{\Gamma_\theta^{G,c}}\int_0^1\int_0^1 f(u)f(v)Q_x^G(du,dv)\\
&= c\langle\theta, f\rangle^2 - cE_{\Gamma_\theta^{G,c}}\langle x, f\rangle^2 + \int_0^1\int_0^1 f(u)f(v)Q_\theta^{\mathscr{F}(G)}(du,dv).
\end{aligned}
\tag{7.6}
$$

(For the last identity we use Fubini and the definition of $Q_\theta^{\mathscr{F}(G)}$.) The last equation results in the following formula for the second moment of $\langle x, f\rangle$:

$$
E_{\Gamma_\theta^{G,c}}\langle x, f\rangle^2 = \langle\theta, f\rangle^2 + \frac{1}{c}\int_0^1\int_0^1 f(u)f(v)Q_\theta^{\mathscr{F}(G)}(du,dv). \quad \square \tag{7.7}
$$

Remark. Note that this formula is valid regardless of whether the underlying martingale problem has a unique solution or a unique equilibrium.

(b) **Proof of Theorem 0.8.** We prove first the second assertion of the theorem, the pointwise convergence of $\mathscr{F}^n(G)$ stated in (0.63). The crucial tool is the relation (7.4) which holds for every equilibrium state of a process solving the martingale problem in (2.1) with $\theta(t) \equiv \theta$. Rewrite (7.4) in terms of variances and iterate to obtain

$$
\mathrm{Var}_{\mu_\theta^{k,0}}(\langle x, f\rangle) = \left(\sum_{j=0}^{k} c_j^{-1}\right) \cdot \int\int Q_\theta^{\mathscr{F}_c^k(G)}(du,dv)f(u)f(v) \tag{7.8}
$$

with $\mathscr{F}_c^k = \mathscr{F}^k = \mathscr{F} \circ \cdots \circ \mathscr{F}(G)$, the k-fold composition. From this relation we can read off first the following implication:

$$
\left(\left(\sum_{k=0}^{\infty} c_k^{-1}\right) = \infty\right) \Rightarrow \left(\mu_\theta^{k,0} \underset{k\to\infty}{\Rightarrow} \int_0^1 \theta(du)\delta_{(\delta_u)}\right). \tag{7.9}
$$

(Simply repeat the argument from the proof of Theorem 0.5, (0.43).) Inserting (7.9) with $c_k \equiv c \in (0,\infty)$ into the left-hand side of (7.8) we obtain

$$(7.10)\quad \iint (n/c) Q_\theta^{\mathscr{F}_c^n}(du, dv) f(u) f(v) \underset{N\to\infty}{\longrightarrow} \mathrm{Var}_\theta(f) = \int Q_\theta(du, dv) f(u) f(v) \qquad \forall f \in C([0, 1]),\ \theta \in \mathscr{P}([0, 1]).$$

This finishes the proof of (0.63,(i)).

The second part, (0.63(ii)), now follows from (7.4) along the lines of Baillon et al. (1994, section 1). We refer to that paper for details.

The proof of the relation (0.62) is now easy. Without loss of generality we can take $c = 1$ and suppose that

$$(7.11)\qquad Q_\theta^{\mathscr{F}(a\cdot G)} = h(a) Q_\theta^G \quad \text{for some } h : [0, \infty) \to [0, \infty).$$

Then

$$(7.12)\qquad Q_\theta^{\mathscr{F}^n(G)} = h \circ \cdots \circ h(a) Q_\theta^G$$

and hence according to (7.10) above

$$(7.13)\qquad nh \circ \cdots \circ h(a) Q_\theta^G \underset{N\to\infty}{\Rightarrow} Q_\theta.$$

This means first of all that

$$(7.14)\qquad nh \circ \cdots \circ h(a) \underset{N\to\infty}{\Rightarrow} a^* \in (0, \infty).$$

Combining (7.14) and (7.13) we get

$$(7.15)\qquad G(\theta, u, v) = (a^*)^{-1},$$

which finishes the proof of (0.62).

Proof of Theorem (0.9). Theorem 0.9 follows from Theorem 0.8 exactly along the same lines as in Dawson and Greven (1993c, Corollary to Theorems 3, 4, 5). The modification needed is to replace the expression $\theta(1 - \theta)$ in (4.5), respectively (5.26), (5.27) of that paper, by

$$\int_0^1 \int_0^1 f(u) f(v) Q_\theta^G(du, dv), \qquad f \in C([0, 1]),$$

and $F_{k+1}(\theta)$ by $\mathscr{F}_{c_k} \circ \cdots \circ \mathscr{F}_{c_0}(G)(\theta)$ where we define $\mathscr{F}_c(G)$ by the relation $\int \Gamma_\theta^{G,c}(dx) Q_x^G = Q_\theta^{\mathscr{F}_c(G)}$. We refer to the above reference for further details.

INDEX OF NOTATION

δ_u single atom measure at u

$\mathscr{P}([0, 1]) =$ set of Borel probability measures on $[0, 1]$

$\theta =$ a fixed element of $\mathscr{P}([0, 1])$

$\theta(\mu) = \theta$ (mean measure of initial state)

$C([0, T], E)$ continuous functions from $[0, T]$ into E

$\|\cdot\| =$ supremum norm

$\ll$ absolute continuity of measures

$\Omega_N =$ hierarchical group

$\xi^0 = (0, 0, 0, \dots) \in \Omega_N$

$\frac{\partial F(x)}{\partial x_\xi}(u) =$ variational derivative

$\mathscr{L} =$ probability law

$\mathscr{P}_a =$ atomic probability measures

$\beta_k(N) = N^k$ time scales
$\Gamma_\theta^k(\cdot) =$ level k equilibrium measure
$\tilde{Z}_t^{\theta,k} =$ level k measure-valued diffusion
$Z_\theta^{\theta,k} =$ level k stationary measure-valued diffusion in equilibrium
$\mu_\theta^{j,k} =$ multistep transition probabilities for interaction chain
$d_k =$ diffusion constants
$x_{\xi,k} = N^{-k} \sum_{\xi' : d(\xi,\xi') \le k} x_{\xi'}$
Beta$(a, b) =$ beta probability distribution with parameters a, b
GEM$(a) =$ Griffiths-Engen-McCloskey distribution with parameter a
$\mathbb{Z} =$ integers
$\mathbb{N} = \{1, 2, 3, \ldots\}$
$\mathbb{Z}^- = \{0, -1, -2, -3, \ldots\}$
$\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$
$\alpha! = \int_0^\infty s^\alpha e^{-s}\, ds, \alpha > -1$
$f_\alpha(k) =$ cluster scales
$\{c_k\} =$ sequence of multilevel migration rates
$Q_{x_\xi}(du, dv), Q_{x_\xi}^G(du, dv)$ (Section 0(g))
$\mathcal{G}, \tilde{\mathcal{G}}, \hat{\mathcal{G}}$ (Section 0(g))
$\mathcal{F}_c$ ((7.8))
$M_j^{(k)}, R^{(k)}, S_\ell^{(k)}$ ((0.54),(0.55))
$M = \{\delta_u | u \in [0, 1]\}$
w-lim $=$ weak limit
$\Rightarrow$ weak convergence

References

D. J. Aldous, *Exchangeability and related topics*, École d'Été de Probabilités de Saint Flour XIII, Lecture Notes in Math., vol. 1117, Springer-Verlag, 1985, pp. 1–198.

J. Baillon, P. Clement, A. Greven, and F. den Hollander, *On the attracting orbit of a nonlinear transformation arising from renormalization of hierarchically interacting diffusions*, Canad. J. Math. (1995) (to appear).

J. T. Cox, A. Greven, and T. Shiga, *Finite and infinite systems of interacting diffusions*, preprint, 1993.

D. A. Dawson, *Measure-valued Markov processes*, École d'Été de Probabilités de Saint Flour XXI, Lecture Notes in Math., vol. 1541, Springer-Verlag, 1993, pp. 1–261.

D. A. Dawson and A. Greven, *Multiple time scale analysis of hierarchically interacting systems*, A Festschrift to Honor G. Kallianpur, Springer-Verlag, 1993a, pp. 41–50.

D. A. Dawson and A. Greven, *Multiple time scale analysis of interacting diffusions*, Probab. Theory Related Fields **95** (1993b), 457–508.

D. A. Dawson and A. Greven, *Hierarchical models of interacting diffusions: multiple time scale phenomena, phase transition and pattern of cluster-formation*, Probab. Theory Related Fields **96** (1993c), 435–473.

D. A. Dawson and P. March, *Resolvent estimates for Fleming-Viot operators and uniqueness of solutions to related martingale problems*, J. Funct. Anal. (1995) (to appear).

P. Donnelly and T. G. Kurtz, *A countable representation of the Fleming-Viot measure-valued diffusion*, preprint, 1991.

R. M. Dudley, *Real analysis and probability*, Wadsworth and Brooks/Cole, Pacific Grove, CA, 1989.

S. N. Ethier, *The distribution of frequencies of age-ordered alleles in a diffusion model*, Adv. Appl. Probab. **22** (1990a), 519–532.

S. N. Ethier, *On the stationary distribution of the neutral one-locus diffusion model in population genetics*, Ann. Appl. Probab. **2** (1990b), 24–35.

S. N. Ethier, *Equivalence of two descriptions of the ages of alleles*, J. Appl. Probab. **29** (1992), 185–189.

S. N. Ethier and Griffiths, *The infinitely-many-sites model as a measure-valued diffusion*, Ann. Probab. **15** (1987), 515–545.

S. N. Ethier and T. G. Kurtz, *The infinitely-many-neutral-alleles diffusion model*, Adv. Appl. Probab. **13** (1981), 429–452.

S. N. Ethier and T. G. Kurtz, *Markov processes, characterization and convergence*, Wiley, New York, 1986.

S. N. Ethier and T. G. Kurtz, *The infinitely-many-alleles-model with selection as a measure-valued diffusion*, Lecture Notes in Biomath., vol. 70, Springer-Verlag, 1987, pp. 72–86.

S. N. Ethier and T. G. Kurtz, *Convergence to Fleming-Viot processes in the weak atomic topology*, Stochastic Process. Appl. **54** (1995), 1–27.

K. Fleischmann, *Mixing properties of cluster invariant distributions*, Litovsk. Mat. Sb. **18** (1978), 191–199.

K. Fleischmann and A. Greven, *Diffusive clustering in an infinite system of hierarchically interacting diffusions*, Probab. Theory Related Fields. **98** (1994), 517–566.

A. Greven, *Couplings of Markov chains by randomized stopping times. Part* I: *Couplings, harmonic functions and the Poisson equation*, Probab. Theory Related Fields **75** (1987), 195–212.

K. Handa, *A measure-valued diffusion process describing the stepping stone model with infinitely many alleles*, Stochastic Process. Appl. **36** (1990), 269–296.

A. Joffe and M. Métivier, *Weak convergence of sequences of semimartingales with application to multitype branching processes*, Adv. Appl. Probab. **18** (1986), 20–65.

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

G. P. Patil and C. Taillie, *Diversity as a concept and its implications for random communities*, Bull. Internat. Statist. Inst. **47** (1977), 497–515.

D. Ruelle, *Statistical mechanics: Rigorous results*, Benjamin, 1969.

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

T. Shiga, *Continuous time multi-allelic stepping stone models in population genetics*, J. Math. Kyoto Univ. **22** (1982), 1–40.

F. Spitzer, *Principles of random walk*, Van Nostrand, Princeton, NJ, 1964.

J. Vaillancourt, *Interacting Fleming-Viot processes*, Stochastic Process. Appl. **36** (1990), 45–57.

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

K. Yosida, *Functional analysis*, 6th ed., Springer-Verlag, 1980.

Department of Mathematics & Statistics, Carleton University, Ottawa K1S 5B6, Canada

E-mail address: `ddawson@math.carleton.ca`

Institut für Stochastik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany

E-mail address: `greven@namu01.gwdg.de`

Département de Mathématiques et d'Informatique, Université de Sherbrooke, Québec, Canada J1K 2R1

E-mail address: `jean.vaillancourt@DMI.USherb.CA`