Electronic Journal of Probability

Vol. 4 (1999) Paper no. 4, pages 1–81.

Journal URL
http://www.math.washington.edu/~ejpecp/

Paper URL
http://www.math.washington.edu/~ejpecp/EjpVol4/paper4.abs.html

Hierarchically interacting Fleming-Viot processes with selection and mutation: Multiple space time scale analysis and quasi-equilibria.

Donald A. Dawson
Fields Institute, 222 College Street, Toronto M5A 4B8, Canada.
don@fields.utoronto.ca

Andreas Greven
Mathematisches Institut, Universität Erlangen-Nürnberg
Bismarckstraße 1 1/2, D-91054 Erlangen, Germany.
greven@mi.uni-erlangen.de

Abstract: Genetic models incorporating resampling and migration are now fairly well-understood. Problems arise in the analysis, if both selection and mutation are incorporated. This paper addresses some aspects of this problem, in particular the analysis of the long-time behaviour before the equilibrium is reached (quasi-equilibrium, which is the time range of interest in most applications).

Keywords: Interacting Fleming-Viot processes, Renormalization analysis, Selection, Mutation, Recombination.

Research supported by NSERC, DFG and Max Planck Award for International Cooperation.

Submitted to EJP on January 2, 1998. Final version accepted on March 4, 1999.

Hierarchically interacting Fleming-Viot processes with selection and mutation: Multiple space time scale analysis and quasi-equilibria

Donald A. Dawson[1,2] **Andreas Greven**[2,3]

March 9, 1999

Abstract

Genetic models incorporating resampling and migration are now fairly well-understood. Problems arise in the analysis, if both selection and mutation are incorporated. This paper addresses some aspects of this problem, in particular the analysis of the long-time behaviour before the equilibrium is reached (quasi-equilibrium, which is the time range of interest in most applications).

The first model we use is a countable system of Fleming-Viot processes with selection and interaction between components via migration. Types are in principle described by the set $[0, 1]$ or a hierarchically structured countable subset, but we relabel the countable subset occuring in our universe during the evolution via lexicographical ordering by $\mathbb{N}$ and each component of the system takes values in $\mathcal{P}(\mathbb{N})$, the probability measures on $\mathbb{N}$ and represents the frequency of genetic types which occur in this colony. Furthermore we discuss in a second model the effect of adding mutation and recombination to the system already incorporating selection.

Of particular interest in such evolutionary models is the nonequilibrium behaviour. The latter can be studied rigorously by exhibiting quasi-equilibria. These are obtained by letting a parameter in the evolution tend to infinity and observing the system in various different time scales expanding with this parameter in different orders of magnitude. In these limits the analysis simplifies and the behaviour of the space-time rescaled system is described by a sequence of time dependent Markov chains with values in $\mathcal{P}(\mathbb{N})$ and initial points, which vary with the time scale used. The properties of these collections of chains correspond to properties in the longtime behaviour of the original system.

The resulting nonequilibrium behaviour displays a very rich structure, we focus on the following phenomena: influence of the migration on the quasi-equilibria, the increased variety of species under selection, the influence of migration on the speed of selection, and the role of mutation.

Finally, in this paper we set up the framework needed to formulate on a rigorous level a number of unresolved issues in evolutionary models which will be discussed in future work. At the same time the analysis provides an interesting example of a rigorous renormalisation analysis.

Keywords: Interacting Fleming-Viot processes, Renormalization analysis, Selection, Mutation, Recombination.

[1] Fields Institute, 222 College Street, Toronto M5A 4B8, Canada.

[2] Research supported by NSERC, DFG and Max Planck Award for International Cooperation.

[3] Mathematisches Institut, Universität Erlangen-Nürnberg, Bismarckstraße 1 1/2, D-91054 Erlangen, Germany.

Contents

Part A
Introduction, model and results

0 Introduction and motivation

The present paper investigates a system of countably many interacting measure valued diffusions on $\mathcal{P}(\mathbb{N})$, the set of probability measures on $\mathbb{N}$, which arise from models in population genetics or from evolutionary models and include three driving forces for evolution: *migration, resampling* and *selection.* $\mathbb{N}$ occurs as type space by labeling a countable universe of hierarchically structured types in a specific way. Furthermore in a second part of this paper we begin the discussion of the influence of *mutation* and *recombination* in such a model. In our models the resampling drives the diffusive term and migration, selection, mutation and recombination give the drift term. Some aspects of the model in the *two-type* case are discussed by Shiga (1982) and by Shiga and Uchiyama (1986). Some multitype situations are treated by Ethier and Kurtz (1987). Related particle models of this type are discussed by Kang, Krone and Neuhauser (1995).

It is the *nonlinear drift term* induced by the selection which is responsible for the *new features* the model exhibits (compare theorems 4-6). Furthermore the *combination* of migration, selection and mutation leads to effects which cannot be reduced to one of those mechanisms. In particular this makes the effects of mutation in the model difficult to analyse and gives rise to new effects. On the other hand with our approach of *multiple space-time scale analysis* we can build up the proper mathematical framework, in which to discuss a number of principal questions arising in evolutionary models and which until now are not resolved neither on the empirical nor on the theoretical level.

A bit more needs to be said on the choice of models we investigate in this paper. Evolutionary models are based on three key ingredients: a source of *diversity*, a mechanism of *competition* in fitness reducing diversity since the least fit types are replaced by types of higher fitness and finally a *spatial* structure of the population and evolution. For the spatial structure we use the hierarchical group first introduced to mathematical biology by Sawyer and Felsenstein (1983), which reflects varying degrees of kinship between geographically spread out colonies.

We first consider in chapter 1 an evolutionary model which includes *resampling, migration* and *selection.* Here we get the diversity since the system starts with a large repertoire of initial types. The behaviour for large times is now dictated by the selection, namely, as time proceeds only the fittest types with the highest selective value can remain in the system. However at earlier times there will be large time stretches where types with only slightly different selective value will remain in the system and coexist already in an almost stable fashion. We call the corresponding distributions *quasi-equilibria* of the system and these are the objects of primary interest. An example of this phenomenon of selection of highly fit types from a large repertoire of initial types, which provides the basic model in this paper, arises in the theory of *neuronal group selection* as described by G.M. Edelman (1987).

In the second model which we consider in chapter 2, the diversity is produced by *mutation* which together with *recombination* is the source of diversity in population genetics. Here initially we have few different types of primitive nature, i.e. low fitness, but during the evolution the diversity is produced by the mechanism of mutation and recombination. We choose the mutation mechanism such that it reflects Wright's ideas of *adaptive landscapes* (for a modern description see [W]).

In this paper we shall not cover every question of interest for application but shall focus mainly on the understanding of the following five phenomena:

(1) *Nonequilibrium behaviour:* The selection will typically make the system ergodic (or allow only for a small number of equilibria). However relaxation into equilibrium will occur late and we are therefore interested in the *quasi-stationary states* which we see on the way to equilibrium if the observation is in time scales of interest (to an observer of the evolution process, who is part of that process himself).

(2) *Mutation:* What are the effects, if mutation is incorporated in the model with selection? In particular we ask,

can the fact that selection and mutation are competing forces, with respect to diversity, lead to new phenomena? What role does the migration play in this context?

(3) *Mathematical framework for evolution theory:* We shall develop with the tool of measure valued diffusions and multiple scale analysis the framework to discuss some problems arising in evolution theory, this will be the main thrust of our chapter 2.

(4) *Nonlinear perturbation of linear systems:* A system with migration and resampling is well-understood due to the fact that the interaction between components arises from the migration, which is a linear mechanism (the "self-interaction" of a component due to resampling is of course also nonlinear but this is a nonlinearity in the diffusive term). For suitable migration mechanism, interacting systems allow for a large set of extremal equilibrium states indexed by the mean frequency distribution of "types" i.e. elements of $\mathcal{P}(\mathbb{N})$. The introduction of a nonlinear drift term can lead to a restricted manifold needed to index the extremal invariant states. This general phenomenon can be studied here in a tractable example, since the diffusion coefficient and the nonlinear drift coefficient vanish on the same part of the state space.

(5) *Renormalization analysis:* In the evolution of spatial systems we can rescale time and space and define this way new renormalized systems. The models under consideration allow for a mathematically rigorous renormalization analysis in the mean-field limit of a migration parameter tending to infinity and provides therfore an interesting class of examples where the heuristics of renormalization analysis can be turned into mathematics. This is of interest for mathematical physics as well.

Let us now discuss below some of the new features related to the above points (1) - (5) in more detail, for (2) and (3) see also section 2(a).

In order to model the phenomenon of quasi-equilibria rigorously we need sequences of *separating space-time scales* and this is achieved by taking a whole sequence of (infinite) interacting systems and by a mean-field type limit of the interaction. Passing to this limit we are however able to capture the most important phenomena and construct a sequence of quasi-equilibria corresponding to a *sequence of time scales*. That is, we obtain probability measures $(\tau_k)_{k \in \mathbb{N}}$ on the type space, i.e. elements of $\mathcal{P}(\mathbb{N})$, which describe the mean type distribution of the k-th quasi-equilibrium, where all types are still present, whose selective value differs only by an amount negligible for the k-th time scale. Corresponding to the (τ_k) we construct the associated quasi-equilibrium, which is an element of $\mathcal{P}((\mathcal{P}(\mathbb{N}))^{\mathbb{N}})$, where $\mathcal{P}$ denotes probability measures. This law appears as the equilibrium of an evolution in a *random medium*. The latter is given as a functional of a *Markov random field* on a tree, which arises from a *space-time renormalization scheme* with a sequence of time scales and matching spatial averaging.

There are now two different types of behaviour possible. If the highest selective value is *attained* by a set of types, then $\tau_k \Rightarrow \tau_\infty$ as $k \to \infty$ and τ_∞ will be concentrated on the optimal types. More interesting, from the point of view of an adequate evolutionary model as well as mathematically, is the second case in which the highest selective value is *not attained*. Then the structure of the quasi-equilibria becomes more rich and reflects properties of the migration kernel. That is, on the one hand we can have the situation where the types which currently globally exist due to having similar selective values are present in every local window in which we might observe the process, while in other situations we see locally (in the limit) only monotype configurations, and the selection acts only on the boundaries of these monotype clusters. This implies in particular that the *speed of selection* will depend strongly on the properties of migration.

Another interesting phenomenon is the fact that in the situation where the maximal fitness is not attained the variety of species might increase under the influence of selection. That is, even though many types disappear in the competition, the frequencies of the successful types in a colony become more and more uniformly distributed since their proportion is randomly changing as a result of competition and no longer a conserved quantity as it was when only undergoing migration and resampling.

In general the role of *mutation* is to be a source of noise in the system which creates new types which may not have been represented in the initial state or which may have accidentally died out. As a consequence the effects of the initial state are to a large extent washed out. However, whereas in a model without the selection term the addition of mutation is almost trivial to analyse, this is no longer the case in the presence of selection. This is due to the nonlinear character of the selection mechanism considered. In fact there is a very subtle interplay between migration, mutation and the *competing effects* of selection and we begin in this paper with the analysis of this complicated phenomenon. As before the effect of the selection is reduced, if the migration is recurrent, this holds with and without mutation. Most of these phenomena depend very much on the particular nature of the mutation mechanism. We shall consider in this paper a particular choice which is motivated by qualitative biological considerations and is designed to reflect the ideas behind *Wright's adaptive landscapes*. These ideas are described in chapter 2(a).

Outline In the next two chapters we collect the results. For the reason of transparency we shall discuss first in the next chapter 1 the model with selection giving in subsections a) - d) the basis of the analysis and in e) - g) the new effects due to selection. We add then in a second chapter 2 the results on mutation. The proofs are in part B of this paper.

Some words on the methods used are appropriate. The analyses carried out in this paper will use ideas from Dawson and Greven 1993 a) b) c), 1996 and of Dawson, Greven and Vaillancourt 1995, in particular, the *multiple space-time scale analysis* and the *interaction chain*. However due to the selection mechanism the latter object will now be of a more complicated nature. Finally including mutation and recombination a whole set of new mathematically quite challenging problems arises. Nevertheless the methods, *coupling* and *duality* are flexible enough to allow for the needed extension to study some new phenomena due to selection and the effects of mutation.

1 Main results on selection

(a) The model

We want to introduce in this section 1(a) formally a system with infinitely many $\mathcal{P}(\mathbb{N})$-valued components, each describing the *frequencies of* (genetic) *types* within a certain colony, where $\mathbb{N}$ labels the possible types. The interaction between the components comes from the *migration*. The evolution of a single component is a diffusion with a fluctuation term representing *resampling* and a drift term representing *selection* (both corresponding to the change of generation). In order to be able to use in a transparent way earlier work and important techniques from the literature we will use the framework of measure valued diffusions and not ℓ^1-valued diffusions.

The components form a collection with a hierarchically structured system of neighborhoods. It will be convenient to code the space of types in different ways. When we deal with selection there will be a countable set of fitness levels and we choose $\mathbb{N}$ as type space in order to describe the course of the evolutionary process in a transparent fashion. Alternatively, the model could be defined on the type space $[0,1]$. However in this case the Fleming-Viot process for times $t > 0$ almost surely has atomic states due to the fluctuations during the change of generations. (See [D]). In the presence of mutation we deal with a universe of types, which arise from hierarchically structured perturbations of simple types. Here we can use the lexicographic order to also bring everything onto the type space $\mathbb{N}$.

Before we can give a formal definition of our process in point (2) below, we need to define in point (1) some ingredients.

(1) *Ingredients of the basic model:*

<u>(0)</u> The state space of a single component (describing frequencies of types) will be

(1.1) $\mathcal{P}(\mathbb{N}) = \text{set of probability measures on}\, \mathbb{N}.$

The set of components will be indexed by a set Ω_N, which is countable and specified below. The *state space* $\mathcal{X}$ of the system is therefore

(1.2) $\mathcal{X} = (\mathcal{P}(\mathbb{N}))^{\Omega_N}$

and a typical element is written

(1.3) $\quad X = (x_\xi)_{\xi \in \Omega_N}.$

(i) The hierarchial group Ω_N indexing the *colonies*:

(1.4) $$\Omega_N = \Big\{ \xi = (\xi_i)_{i \in \mathbb{N}} \,|\, 0 \le \xi_i \le N-1, \xi_i \in \mathbb{N}, \exists k_0 : \quad \xi_j = 0 \quad \forall j \ge k_0 \Big\},$$

with addition defined as componentwise addition modulo N. Here N is a parameter with values in $\{2, 3, 4, ...\}$.

An element $(\xi_0, \xi_1, \xi_2, \ldots)$ of Ω_N should be thought of as the ξ_0-th village in the ξ_1-th county in the ξ_2-th state, etc. In other words we think of colonies as grouped according to degrees of neighborhood.

The hierarchical distance $d(\cdot, \cdot)$ is defined accordingly as

(1.5) $$d(\xi, \xi') = \inf(k \,|\, \xi_j = \xi'_j \quad \forall j \ge k).$$

(ii) The transition kernel $a(\cdot, \cdot)$ on $\Omega_N \times \Omega_N$, modeling *migration*.

We shall only consider *homogeneous* transition kernels, which have in addition the *exchangeability property* between colonies of equal distance. That is $a(\cdot, \cdot)$ has the form:

(1.6) $$\begin{aligned} a(\xi, \xi') &= a(0, \xi' - \xi) \\ a(0, \xi) &= \sum_{k \ge j} (c_{k-1} N^{-(k-1)}) N^{-k} \qquad \text{if } d(0, \xi) = j \ge 1 \end{aligned}$$

where

$$c_k > 0 \quad \forall k \in \mathbb{N}, \quad \sum_k c_k N^{-2k} < \infty.$$

The kernel $a(\cdot, \cdot)$ should be thought of as follows. With rate $c_{k-1} N^{-(k-1)}$ we choose a distance k, and then each point within distance k is picked with equal probability as the new location. Note that the scaling factor N^k in the rate $c_k N^{-k}$ represents the volume of the k-ball in the hierarchical metric.

(iii) The *selection* matrix $(V(u, v))_{u, v \in \mathbb{N}}$.

We choose a special form of this matrix corresponding to so-called *additive selection* (haploid case), where fitness of a pair (u, v) is derived from a fitness function of a single type. However in chapter 2, we prove some results for the general case. For a discussion of this assumption compare [D][1993] chapter 10.2 and [EK3]. To describe the fitness of a type u consider the sequence $\{\varphi_0^u, \varphi_1^u, \varphi_2^u, \ldots\}$ with $\varphi_r^u \in [0, M] \cap \mathbb{N}$ and all but finitely many φ_r^u equal to zero. Here M is a fixed parameter and N is as in the definition of Ω_N in (i). As long as N is fixed, M may be thought of as $N-1$, once the limit $N \to \infty$ is considered M is however a fixed parameter. Then set

(1.7) $$V(u, v) = \chi_N(u) + \chi_N(v)$$

with

$$\chi_N(u) = \sum_{r=0}^{\infty} \frac{1}{N^r} (\varphi_r^u), \qquad \varphi_r^u \in [0, M] \cap \mathbb{N}, \quad M \le N-1.$$

We usually suppress N in this notation and write χ for χ_N. We can then identify the type u with with the sequence $(\varphi_r^u,)$.

Before we continue we briefly motivate the choice of V and χ_N in two ways. One way to think about the special structure of χ is the following. Suppose all possible types in the universe would be labeled by the interval $[0, 10] \cap \mathbb{Q}$. Then draw independently types from this reservoir according to some probability law. Assume furthermore all types are ordered according to fitness and the fitness function χ has the form $\chi(x) = x$. Then pick $N = 10$ and we can write down for a given type the decimal expansion and the decimals give the φ^u_r. This is a fairly typical situation whenever the types are all linearly ordered according to fitness.

On a biological level the structure of the fitness function χ corresponds to the following idea:
Consider a collection of genes $z_1, z_2, \dots$ such that a string of those genes characterizes types. Then each of these genes have a certain value for the fitness of a type, if they are present. On the other hand on a smaller scale the appearance of certain combinations (z_i, z_j) may add to this fitness. Continue this picture with higher order effects related to combination (z_i, z_j, z_k) etc. This will lead to a grouping of the genetic types in the form given above and which represents the idea that higher fitness requires more complex structures.

Remark **(Limit $N \to \infty$)** Later we discuss the case where $N \to \infty$. Then for selection and migration, the limit $N \to \infty$ *separates* those times at which qualitative changes (start of competition of types differing in level k fitness or leaving a k-ball to a $(k+1)$-ball), occur. □

(iv) Initial distribution μ.
We call a measure $\nu \in \mathcal{P}((\mathcal{P}(\mathbb{N}))^{\Omega_N})$ homogeneous, if for all $\eta \in \Omega_N$ and $\mathcal{B}$ denoting Borel algebra,

(1.8) $$\nu((x_{\xi+\eta})_{\xi \in \Omega_N} \in B) = \nu((x_\xi)_{\xi \in \Omega_N} \in B) \quad \forall B \in \mathcal{B}((\mathcal{P}(\mathbb{N}))^{\Omega_N}).$$

The class of all homogeneous elements of $\mathcal{P}((\mathcal{P}(\mathbb{N}))^{\Omega_N})$ is denoted by $\mathcal{T}$ and the ergodic ones (with respect to shifts) with $\mathcal{T}_e$. For every $\mu \in \mathcal{T}_e$ we define $\Theta \in \mathcal{P}(\mathbb{N})$ by

(1.9) $$\Theta = \Theta(\mu) \ : \ \Theta(\{u\}) = E^\mu(x_\xi(\{u\})) \quad u \in \mathbb{N}.$$

The measure Θ describes the *spatial density of types* under the law μ.

(2) *The interacting Fleming-Viot processes with selection*
The process will be defined as solution to a *martingale problem*. We need some objects to be able to define the associated generator L.

We start by defining a second order partial differential operator on functions on $(\mathcal{M}(\mathbb{N}))^{\Omega_N}$, with $\mathcal{M}$ denoting finite measures. For a function

(1.10) $$F : (\mathcal{M}(\mathbb{N}))^{\Omega_N} \longrightarrow \mathbb{R},$$

we define the derivative in the point $x = (x_\xi)_{\xi \in \Omega_N}$ with respect to x_η and in direction δ_u as

(1.11) $$\frac{\partial F(x)}{\partial x_\eta}(u) = \lim_{\varepsilon \to 0} \frac{F(x^{\eta,\varepsilon,u}) - F(x)}{\varepsilon}, \qquad u \in \mathbb{N},$$

where

$$x^{\eta,\varepsilon,u}_\xi = \begin{cases} x_\xi + \varepsilon \delta_u & \text{for} \quad \xi = \eta \\ x_\xi & \text{for} \quad \xi \neq \eta. \end{cases}$$

Since convex combinations of point masses approximate finite measures in the weak topology it suffices to consider directions δ_u as long as F has a weakly continuous derivative.

The second derivative $\frac{\partial^2 F}{\partial x_\xi \partial x_\xi}(u,v)$ is obtained by using (1.11) on $\frac{\partial F(x)}{\partial x_\xi}(u)$ as function of x.

Let $\mathcal{A}$ be the algebra of functions on $(\mathcal{M}(\mathbb{N}))^{\Omega_N} \supseteq (\mathcal{P}(\mathbb{N}))^{\Omega_N}$ generated by functions F of the form (we shall write for integrals over $\mathbb{N}$ often $\int\limits_0^\infty$):

$$(1.12) \qquad F(x) = \left[\int\limits_0^\infty \dots \int\limits_0^\infty f(u_1, \dots u_\ell) x_{\xi_1}(du_1) \dots x_{\xi_\ell}(du_\ell) \right]$$
$$f \in C_0(\mathbb{N}^\ell) \quad (\xi_1, \dots, \xi_\ell) \in (\Omega_N)^\ell, \quad \ell \in \mathbb{N}.$$

These functions are at least twice differentiable (with weakly continuous derivatives). Then we can define the generator L on functions $F \in \mathcal{A}$ in $x \in (\mathcal{P}(\mathbb{N}))^{\Omega_N}$:

$$\begin{aligned}(1.13) \qquad L(F)(x) = \sum_{\xi \in \Omega_N} \Bigg[& \sum_{\xi' \in \Omega_N} a(\xi, \xi') \int\limits_0^\infty \frac{\partial F(x)}{\partial x_\xi}(u) \, (x_{\xi'} - x_\xi)(du) \\ & + d \int\limits_0^\infty \int\limits_0^\infty \frac{\partial^2 F(x)}{\partial x_\xi \partial x_\xi}(u, v) Q_{x_\xi}(du, dv) \\ & + s \int\limits_0^\infty \int\limits_0^\infty \int\limits_0^\infty \frac{\partial F(x)}{\partial x_\xi}(u) V(v, w) x_\xi(dv) Q_{x_\xi}(du, dw) \Bigg],\end{aligned}$$

where d and s are positive real numbers and

$$(1.14) \qquad Q_{x_\xi}(du, dv) = x_\xi(du)\delta_u(dv) - x_\xi(du)x_\xi(dv).$$

The first term in the r.h.s. of (1.13) models *migration*, the second one *resampling* and the third one *selection*. These diffusion models are the large population per colony diffusion limits of particle models. Compare for example [S2], [EK2], [D] for material showing this in non spatial situations.

Remark For our special choice of the selection matrix (additive selection) the last term in (1.13) can be simplified and this will be used occasionally so we write this out here. If V is of the form in (1.7) an explicit calculation gives that the selection term has the form:

$$\begin{aligned}(1.15) \qquad & s \int\limits_0^\infty \frac{\partial F(x)}{\partial x_\xi}(u) \left[\chi(u) - \int_0^\infty \chi(v) x_\xi(dv) \right] x_\xi(du) \\ & = s \int\limits_0^\infty \int\limits_0^\infty \frac{\partial F(x)}{\partial x_\xi}(u) \chi(v) Q_{x_\xi}(du, dv).\end{aligned}$$

This means that types u which have higher than average fitness in the colony ξ , that is

$$\chi(u) > \int_0^\infty \chi(v) x_\xi(dv)$$

will grow in frequency and those having lower than average fitness will decrease in frequency.

Now we are ready for the basic existence and uniqueness statement.

Theorem 0 (Existence and Uniqueness)

(a) *For every probability law μ on $(\mathcal{P}(\mathbb{N}))^{\Omega_N}$, the martingale problem (L, μ) is well-posed, i.e. there exists a unique* *law on $D([0,\infty);(\mathcal{P}(\mathbb{N}))^{\Omega_N})$, under which for every $F \in \mathcal{A}$ the canonical process with initial law μ has the property that*

$$(1.16)\qquad \left(F(X(t)) - \int_0^t LF(X(s))ds\right)_{t\geq 0} \quad \text{is a martingale.} \qquad \square$$

Definition 1.1 *The system of interacting Fleming-Viot processes with selection and initial distribution μ is the canonical process $X(t)$ on $D([0,\infty),(\mathcal{P}(\mathbb{N}))^{\Omega_N})$ with law given by the unique solution of the (L,μ) martingale problem.* $\square$

It follows from standard theorems that we can conclude from Theorem 0 (a) the following:

Corollary 0 *The Fleming-Viot process with selection can be realized on $C([0,\infty),\ (\mathcal{P}(\mathbb{N}))^{\Omega_N})$ and this process is a strong Markov process.* $\square$

Extension In order to generalize the resampling and selection mechanism of the model we first look at a special case. In the case of just two types we can identify x_ξ with $x_\xi = x_\xi(\{0\})$, if 0 and 1 label the types and 0 is the superior type. In that case $X(t) = (x_\xi(t))_{\xi\in\Omega_N}$ solves the following system of stochastic differential equations

$$(1.17)\qquad \begin{aligned} dx_\xi(t) &= \sum_{\xi'\in\Omega_N} a(\xi,\xi')(x_{\xi'}(t) - x_\xi(t))dt \\ &+ \sqrt{2d\{x_\xi(t)(1-x_\xi(t))\}}dw_\xi(t) \\ &+ sVx_\xi(t)(1-x_\xi(t))dt, \qquad \xi\in\Omega_N, \end{aligned}$$

where $(\omega_\xi(t))_{\xi\in\Omega_N}$ are independent Brownian motions and $V = \chi(0) - \chi(1)$.

For two types we can generalize the system (1.17) by:

$$\begin{aligned} dx_\xi(t) &= \sum a(\xi,\eta)(x_\eta(t) - x_\xi(t))dt + \sqrt{2h(x_\xi(t))\{x_\xi(t)(1-x_\xi(t))\}}dw_\xi(t) \\ &+Vh(x_\xi(t))x_\xi(t)(1-x_\xi(t))dt \end{aligned}$$

for some bounded non-negative continuous function h on [0,1]. In this case we can view h has a function modulating the speed of reaction induced by "pairs" of opposite types. In the multitype situation (more than 2 types) we must have a diffusion on $\mathcal{P}(\mathbb{N})$, which induces a specific form of the diffusion coefficient, namely for the multitype situation we would replace $Q_z(du,dv),\ z\in\mathcal{P}([0,1])$ in (1.13) by

$$(1.18)\qquad Q^H_z : Q^H_z(du,dv) = \Big(\int H(z,u,w)z(dw)\Big)z(du)\delta_u(dv) - H(z,u,v)z(du)z(dv)$$

for some continuous bounded function $H:\mathcal{P}(\mathbb{N})\times\mathbb{N}\times\mathbb{N}\to[0,\infty)$. Then again the martingale problem has a solution however for $H\not\equiv d\in[0,\infty)$ the question of *uniqueness* of the solution is subtle and results are available only in two cases, for H close to the constant function or $H(z,u,v)$ is independent of z. (See [DM].) Furthermore we can only prove some statements under additional assumptions on H.

Theorem 0 (b) (Existence and uniqueness general model)
Let the function $H(\cdot,\cdot,\cdot)$ be independent of the state (i.e. the first variable) and bounded on $\mathbb{N}\times\mathbb{N}$ and $\mu\in\mathcal{T}$. If we replace in (1.13) $Q_{x_\xi}(du,dv)$ by $Q^H_{x_\xi}(du,dv)$ then the martingale problem (L,μ) is well-posed. $\square$

Throughout the rest of the paper we shall impose the *assumption* of Theorem 0(b) on H and for sections c) - g) we shall even assume $H\equiv d$.

(b) Preliminaries: Basic Ergodic Theory

The first preliminary step is to determine the behavior of $(X(t))_{t\geq 0}$ for $t \to \infty$. In the following sections we then discuss the essential question, namely, the problem of quasi-equilibria.

We start by recalling the behavior for the model without selection (see [DGV, 1995], Theorem 0.1). If $V(\cdot,\cdot) \equiv 0$, then for $\hat{a}(\xi,\eta) = \frac{1}{2}(a(\xi,\eta) + a(\eta,\xi))$:

case 1 : $\hat{a}(\cdot,\cdot)$ transient
For every $\Theta \in \mathcal{P}(\mathbb{N})$ there exists an extremal equilibrium state ν_Θ such that

(1.19) ν_Θ is homogeneous and spatially ergodic with $E^{\nu_\Theta}(x_\xi(\{u\})) = \Theta(\{u\})$.

case 2 : $\hat{a}(\cdot,\cdot)$ recurrent.
The only extremal equilibrium states are given by

(1.20) $\{\delta_{\{x_\xi(\{i\})=1\,\forall\xi\}}\}_{i\in\mathbb{N}}$.

In case 1 we say the system is *stable* since different types coexist in each colony, while in case 2 the system *clusters* by forming larger and larger monotype regions.

In order to describe the longterm behavior of the process in the presence of selection we define for every $\Theta \in \mathcal{P}(\mathbb{N})$:

$$(1.21) \qquad \mathcal{N}^\Theta = \left\{ u \in \mathbb{N} \left| \sum_{r=0}^{\infty} \frac{1}{N^r}\varphi_r^u = \sup\left(\sum_{r=0}^{\infty} \frac{1}{N^r}\varphi_r^v \right| v : \Theta(\{v\}) > 0 \right) \right\}.$$

The set $\mathcal{N}^\Theta$ consists of all the types with maximal degree of fitness among all those competitors given by the support of Θ in type space. Define the potential (depending on the support) maximal level sets of χ:

$$(1.22) \qquad \mathcal{N} = \{H \subseteq \mathbb{N} \,|\, \exists\Theta \in \mathcal{P}(\mathbb{N}) \text{ with } H = \mathcal{N}^\Theta\}.$$

There are now two principal cases to distinguish. We fix an initial measure μ with intensity measure $\Theta \in \mathcal{P}(\mathbb{N})$, i.e. $E^\mu(x_\xi(\{u\})) = \Theta(u)$, which is translation invariant and ergodic. Then either

$$|\mathcal{N}^\Theta| \geq 1 \quad \text{or} \quad |\mathcal{N}^\Theta| = 0.$$

We split the next theorem into two parts related to these two cases and start analyzing the first case. We need the following notion. If under any permutation of the labels of the type which have equal fitness, a law remains invariant, we call it *φ-exchangeable.*

Theorem 1, Part I (Ergodic theorem, equilibria in case $|\mathcal{N}^\Theta| \geq 1$)

a) case 1: $\hat{a}(\cdot,\cdot)$ *transient*
Exactly for those $\Theta \in \mathcal{P}(\mathbb{N})$ concentrated on level sets of φ, i.e. satisfying

$$supp(\Theta) \in \mathcal{N},\ supp(\Theta) = \mathcal{N}^\Theta,$$

there exists a unique homogeneous extremal equilibrium state, with

(1.23) $E^{\nu_\Theta}(x_\xi(\{u\})) = \Theta(\{u\})$ *and* ν_Θ *is spatially ergodic.*

case 2: $\hat{a}(\cdot,\cdot)$ *is recurrent*
All extremal homogeneous equilibrium states are given by monotype states:

(1.24) $\{\delta_{\{x_\xi(\{u\})=1\,\forall\xi\}}\}_{u\in\mathbb{N}}$.

In case 2 define $\nu_\Theta = \sum_{u=0}^{\infty} \Theta(u)\delta_{\{x_\xi(\{u\})=1,\forall\xi\}}$.

b) Assume that $H(u,v;x) \equiv d$. For every $\mu \in \mathcal{T}_e$ with $|\mathcal{N}^\Theta| \geq 1$ there exists a unique $\Theta_\infty(\mu)$ such that

$$(1.25) \qquad \mathcal{L}((X(t))) \underset{t\to\infty}{\Longrightarrow} \nu_{\Theta_\infty(\mu)}, \text{ with supp}\,[\Theta_\infty(\mu)] = \mathcal{N}^{\Theta(\mu)}.$$

The limiting intensity measure $\Theta_\infty(\mu)$ conditioned on $\mathcal{N}^{\Theta(\mu)}$ can be identified if μ is φ-exchangeable as $\Theta(\mu)$ (recall (1.9)). □

Remark For general H, the expression for $\Theta_\infty(\mu)$ is subtle. In the case where the values of $\{\sum_r \frac{1}{N^r}\varphi_r^u, u \in \mathbb{N}\}$ are strictly ordered $\Theta_\infty(\mu)$ is always determined by the fact that $\mathcal{N}^\Theta$ consists of exactly one element. In the other cases there are no simple expressions for $\Theta_\infty(\mu)$, since the latter is a *complicated nonlinear function* of the distribution μ.

In the second case where μ is such that $\mathcal{N}^\Theta = \emptyset$ a complete analysis would need more detailed assumptions on the structure of the fitness function. We state and prove here only:

Theorem 1, Part II (Ergodic theorem, equilibria in case $\mathcal{N}^\Theta = \emptyset$)
Assume that $\mu \in \mathcal{T}_e$ has intensity measure Θ and $\mathcal{N}^\Theta = \emptyset$.

c) $$(1.26) \qquad \mathcal{L}(X(t)\,|_A) \underset{t\to\infty}{\Longrightarrow} \delta_{\underline{0}} \quad \text{for every}\, A \subseteq \mathbb{N}\, \text{with} |A| < \infty.$$

where for $z \in (\mathcal{P}(\mathbb{N}))^{\Omega_N}$ we mean by

$$Z\,|_A = \{z_\xi(A \cap \cdot),\ \xi \in \Omega\}$$

and $\underline{0}$ the measure $(0)^{\Omega_N}$.

d) Assume that $H(x,u,v) \equiv d$. Furthermore for any two initial laws ν and μ in $\mathcal{T}_e$ with the property that they are φ-exchangeable and that their intensity measures are the same and on their support the maximum of the fitness function is not attained (i.e. $\mathcal{N}^{\Theta(\nu)} = \mathcal{N}^{\Theta(\mu)} = \emptyset$) satisfy

$$(1.27) \qquad \Big[\mathcal{L}(X^\mu(t)) - \mathcal{L}(X^\nu(t))\Big] \underset{t\to\infty}{\Longrightarrow} 0. \qquad \square$$

(c) Multiple space-time scale analysis

This section prepares the construction of quasi-equilibria in subsection 1(d). The important tool needed there for the rigorous study of quasi-equilibria of the corresponding interacting systems is a *renormalization scheme*, which we call multiple space-time scale analysis. That is, in this section we shall study the *spatially rescaled* systems in a *combination* of fast and slow *time scales.*

In other words, we are interested in the behavior of the system after having first of all evolved for times of order $S(N)N^j, S(N) \uparrow \infty, S(N)/N \to 0$ and then we observe it for subsequent time spans of order tN^k, in a block of size k. That is, we look at (we display now the dependence of $X(t)$ on N by writing $X(t) = ((x_\xi^N(t))_{\xi\in\Omega_N}))$ the following collection of space-time rescaled systems:

$$(1.28) \qquad \big(x_{\xi,k}^N(S(N)N^j + tN^k)\big)_{t\geq 0} \quad j,k \in \mathbb{N},$$

where $x_{\xi,k}$ denotes the average over the $x_{\xi'}$ with ξ' such that $d(\xi,\xi') \leq k$.

This defines for every $(j,k) \in \mathbb{N}^2$ a new system again indexed by $\xi \in \Omega_N$ and again with $\mathcal{P}(\mathbb{N})$-valued components, which is obtained by rescaling both space and time in the original system. In the limit of $N \to \infty$ the limiting object will allow us to describe asymptotically the various states through which the system passes on its way to equilibrium.

The important properties of the *renormalization scheme* given above is the following effect on selection and migration:

(i) In times of order $S(N)N^j$ the selection will not distinguish between those types u, v such that in the representation (1.7) of the selection matrix $V(u,v)$ the coefficients φ_r^u and φ_r^v are equal for $r \leq j$.
(ii) In time $S(N)N^j$ the migration is not able to reach components which are in distance $j+1$ or further.

Thus we expect that we see a j-dependent subset of types which do not interact by selection during times of order $tN^k, k < j$ and whose density in space is (asymptotically as $N \to \infty$) preserved during that time span. At the same time the time evolution exhibits migration only within blocks of size k during time spans of order N^k. It is therefore this time scale in which we see a diffusion on the level of k-block averages. This will lead for every pair $j, k \in \mathbb{N}$ to a *limiting dynamics* of the renormalized system defined in (1.28), which has a simpler selection matrix and a simpler migration mechanism, which we shall identify explicitly below.

We proceed now in two steps, first in step 1 we develop the *ingredients* and *concepts* needed for the multiple space-time scale analysis, in step 2 we give the *basic limit result.*

First we assume that we are in the case

$$(1.29) \qquad H(z;u,v) \equiv d \quad \text{on } \mathcal{P}([0,1]) \times [0,1]^2,$$

and later on we discuss the question of an extension to more general resampling mechanisms.

Step 1 In order to describe the limiting dynamics mentioned above we have to introduce a number of concepts. Define first

$$(1.30) \qquad \Omega_\infty = \{(\xi_i)_{i \in \mathbb{N}} \,|\, \xi_i \in \mathbb{N}, \exists k \quad \text{with} \quad \xi_i = 0 \quad \forall i > k\}$$

and note that (as sets)

$$(1.31) \qquad \Omega_N \subseteq \Omega_M \subseteq \Omega_\infty \quad \forall N \text{ and } M \text{ with} : N, M \in \mathbb{N}, N \leq M.$$

Here are now the basic ingredients of the *multiple space-time scale analysis.*

(i) *time scales*

$$(1.32) \qquad \beta_k(N) = N^k \qquad k = 1, 2, ...$$

(ii) spatial scaling given by the *block averages.*
For every $x \in (\mathcal{P}(\mathbb{N}))^{\Omega_N}$, $\xi \in \Omega_N$ and $k > 0$ define:

$$(1.33) \qquad x_{\xi,k}(t) = N^{-k} \sum_{\xi' : d(\xi,\xi') \leq k} x_{\xi'}(t).$$

(iii) The *quasi-equilibrium* of level j : $\Gamma_\Theta^j(\cdot)$, the *associated measure valued diffusions* at level j: $Z_\Theta^j(t)$ and $\tilde{Z}_\Theta^j(t)$, and the level j *diffusion coefficient:* d_j.

They are defined as follows: Consider the measure valued diffusion on $\mathcal{P}(\mathbb{N})$ denoted $Y(t)$ which solves the (L_Θ^k, δ_z) martingale problem (Fleming-Viot process with emigration and immigration from the source $\Theta \in \mathcal{P}(\mathbb{N})$). To define L_Θ^k introduce as domain the algebra generated by functions of the type (M_f denotes finite signed measures):

$$(1.34) \qquad F(z) = \left(\int_0^\infty \ldots \int_0^\infty f(u_1, \ldots, u_\ell) z(du_1) \ldots z(du_\ell) \right) \quad z \in \mathcal{M}_f(\mathbb{N}), \ell \in \mathbb{N}, \ f \in C_0(\mathbb{N}).$$

Then with $Q_z(du,dv) = z(du)\delta_u(dv) - z(du)z(dv)$ we define:

$$(1.35) \qquad L^k_\Theta(F)(z) = c_k \int_0^\infty \frac{\partial F(z)}{\partial z}(u)(\Theta(du) - z(du)) + d_k \int_0^\infty \int_0^\infty \frac{\partial^2 F(z)}{\partial z \partial z}(u,v) Q_z(du,dv),$$

with

$$(1.36) \qquad d_{k+1} = \frac{c_k d_k}{c_k + d_k}, \quad d_0 = 1.$$

Then define for fixed $k \in \mathbb{N}$ and $\Theta \in \mathcal{P}(\mathbb{N})$:

(1.37) • $\Gamma^k_\Theta \in \mathcal{P}(\mathcal{P}(\mathbb{N}))$ is the unique equilibrium distribution for the martingale problem $(L^k_\Theta, \circ)$

(1.38) • $\widetilde{Z}^k_\Theta(t)$ is the unique solution of the $(L^k_\Theta, \delta_\Theta)$ martingale problem

(1.39) • $Z^k_\Theta(t)$ is the unique solution of the $(L^k_\Theta, \Gamma^k_\Theta)$ martingale problem.

The following objects are the additional pieces we need to handle the selection.

(iv) The *k-th level competition process* $(Y^k_t)_{t\geq 0}$ and the *sequence of mean survivor frequencies* $(\tau_k)_{k\in\mathbb{N}}$ starting in Θ. If necessary we display for the latter the dependence on Θ as τ^Θ_k.

The sequences Y^k_t, τ_k are defined interdependent and recursively:

$$(1.40) \qquad \tau_{-1} := \Theta.$$

Denote by $(Y^k_t)_{t\geq 0}$ the solution of the $(\tilde{L}_k, \Gamma^k_{\tau_{k-1}})$ *nonlinear martingale problem* (i.e. the generator depends on the current law of the process, whose expectation is denoted E), which is well-posed (see section 5) and where s, d_0 are the parameters appearing in (1.35) (recall (1.16)):

$$(1.41) \qquad \tilde{L}_k(F)(z) = c_k \int_{\mathbb{N}} \frac{\partial F(z)}{\partial z}(u)([E(z(du))] - z(du)) + d_k \int_{\mathbb{N}} \int_{\mathbb{N}} \frac{\partial^2 F(z)}{\partial z \partial z}(u,v) Q_z(du,dv) + (s/d_0) d_k \int_{\mathbb{N}} \int_{\mathbb{N}} \int_{\mathbb{N}} \frac{\partial F(z)}{\partial z}(u)(\varphi^v_k + \varphi^w_k) z(dv) Q_z(du,dw).$$

Define for $k = 0, 1, 2, \dots$ the mean survivor frequencies τ_k and their supports $\mathcal{N}_k$:

$$(1.42) \qquad \tau_k(\{i\}) = EY^k_\infty(\{i\}) = \tau_{k-1}(\{i\}))$$

$$+ \quad (s/d_0)d_k \int\limits_0^\infty \left[\int\limits_{\mathbb{N}} \int\limits_{\mathbb{N}} \int\limits_{\mathbb{N}} 1_{\{i\}}(u)\,(\varphi^v_k + \varphi^w_k) E\Big(Y^k_t(dv) Q_{Y^k_t}(du, dw) \Big) \right] dt$$

$$= \quad \tau_{k-1}(\{i\}) + (s/d_0)d_k \int\limits_0^\infty \left[E(Y^k_t(\{i\})(\varphi^i_k - \int\limits_{\mathbb{N}} Y^k_t(du)\varphi^u_k)) \right] dt,$$

$$(1.43) \qquad \mathcal{N}_k = \{u \in \mathbb{N} | \tau_k(\{u\}) > 0\}.$$

Occasionally we will need the average *gain function* $(A^k_t)_{t \geq 0}$ in time scale k defined by

$$(1.44) \qquad A^k_t = EY^k_t.$$

(v) The *interaction chain* at *level* j associated with Θ, denoted $(Z^j_k)_{k=-j-1,-j,\dots,0}$ and their *marginals* $\mu^{j,k}_\Theta$.

The process $(Z^j_k)_{k=-j-1,\dots,0}$ is a time inhomogeneous Markov chain on $\mathcal{P}(\mathbb{N})$ starting at time $-j-1$ in τ^Θ_j, i.e.

$$(1.45) \qquad \mathcal{L}(Z^j_{-j-1}) = \tau^\Theta_j$$

and with transition kernel K_{-k} at time $-k, k \in \mathbb{N}$, given by (recall (1.37))

$$(1.46) \qquad K_{-k}(\Theta, d\rho) = \Gamma^{k-1}_\Theta(d\rho).$$

The $k-$th marginal distribution $\mu^{j,k}_\Theta$ of the level-j chain is given by:

$$(1.47) \qquad \mu^{j,k}_\Theta(\cdot) = \int\limits_{\mathcal{P}(\mathbb{N})} \dots \int\limits_{\mathcal{P}(\mathbb{N})} \Gamma^j_{\tau^\Theta_j}(d\Theta_j)\Gamma^{j-1}_{\Theta_j}(d\Theta_{j-1}) \dots \Gamma^k_{\Theta_{k+1}}(\cdot).$$

Remark Consider the equilibrium $\Gamma^{c,d}_\Theta$ of the process generated by L^k_θ (see (1.35) for $c_k = c$ and $d_k = d$. This law can in the case of unweighted sampling ($H \equiv 1$) be constructed explicitly as follows. Take i.i.d. random variables $(U_i)_{i \in \mathbb{N}}$ with marginals Θ and independently of this other collection $(V_i)_{i \in \mathbb{N}}$, which are i.i.d. Beta$(1, c/d)$ distributed random variables. Then

$$(1.48) \qquad \Gamma^{c,d}_\Theta = \mathcal{L}\left(\sum_{i=1}^\infty \left(V_i \prod_{j=1}^{i-1} (1 - V_j) \right) \delta_{U_i} \right).$$

Remark The reader may note that in contrast to the situations previously studied in [DG1] - [DG3] and [DGV, 95], the *initial laws* of the interaction chain at level j are now *j-dependent.* This j-dependence is highly nontrivial if we leave the simple resampling context and discuss the *weighted sampling* case. We therefore extend our analysis to more general models, even though the reader might skip the next paragraph in a first reading.

Extension Finally we indicate the way the ingredients of the analysis look like in the case where we allow the more general resampling/selection terms induced by Q^H_z. Here the problem of the uniqueness of the martingale problem

(1.16) modified according to (1.18) causes problems as well as the proof of an ergodic theorem for such processes. We will therefore in the sequel consider only the case where H is independent of the state and only a function of the types.

In the case $H \not\equiv$ const the resampling term changes at every level its form and hence (1.35) has to be replaced. The definitions must be modified as follows: Suppose that the process associated with the martingale problem below has a unique solution. (This is an *open problem*). Let $\Gamma^{c,H}_\Theta$ be this unique equilibrium state of the martingale problem associated with the generator

$$(1.49) \qquad L^{c,H}_\Theta(F)(z) = c\int_0^\infty \frac{\partial F(z)}{\partial z}(u)(\Theta(du) - z(du)) + \int_0^\infty \int_0^\infty \frac{\partial^2 F(z)}{\partial z \partial z}(u,v) Q^H_z(du,dv).$$

Define $H_A = H|_{\mathcal{P}(A)\times A\times A}$ for $A \subseteq \mathbb{N}$. Furthermore for $c \in [0,\infty)$:

$$(1.50) \qquad \mathcal{F}_{c,A}(H_A)(z,v,u) = \int \Gamma^{c,H}_z(dx) Q^H_x(u,v) \text{ for } u,v \in A,\ z \in \mathcal{P}(A) \text{ with } A \subseteq \mathbb{N}.$$

Then let $\mathcal{F}^{(n)}$ denote (recall (1.43) for the definition of $\mathcal{N}_k$) the map

$$(1.51) \qquad \mathcal{F}^{(n)} = \mathcal{F}_{c_{n-1},\mathcal{N}_{n-1}} \circ \ldots \circ \mathcal{F}_{c_0,\mathcal{N}_0}.$$

Occasionally we shall need $\bar{\mathcal{F}}^{(n)}$ which is obtained by extending the functions obtained as images under $\mathcal{F}^{(n)}$ to $\mathcal{P}(\mathbb{N}) \times \mathbb{N} \times \mathbb{N}$ by setting it equal to zero outside $\mathcal{P}(\mathcal{N}_n) \times \mathcal{N}_n \times \mathcal{N}_n$. Having $\bar{\mathcal{F}}^{(n)}$ at hand we define then

- $\tilde{Z}^k_\Theta(t)$ is the unique solution of the $(L^{c_k,\bar{\mathcal{F}}^{(k)}(H_k)}_\Theta, \delta_\Theta)$ martingale problem, with $H_k = H|_{\mathcal{N}_k}$.
- $Z^k_\Theta(t)$ is the unique solution of the $(L^{c_k,\bar{\mathcal{F}}^{k}(H_k)}_\Theta,\ \Gamma^{c_k,\bar{\mathcal{F}}^{(k)}(H_k)}_\Theta)$ martingale problem.

The defining relations (1.41) and (1.42) are analoguesly replaced by the expression where $Q_x(du,dv)$ becomes Q^H_x.

Step 2 We are now ready to state below the basic *limit theorem* (recall the paragraph before (1.28)) describing the behavior of the *renormalized* process $(x^N_{\xi,k}(S(N)\beta_j(N) + t\beta_k(N)))_{t\geq 0}$ for fixed ξ and the various combinations of j and k. Afterwards we give an heuristic interpretation of this limit theorem. Here $S(N) \uparrow +\infty$ as $N \to \infty$ but $S(N) = o(N)$.

In order to reduce notation throughout the rest of the paper we consider only initial distributions of the process satisfying the following:

Hypotheses $(x^N_\xi(0))_{\xi\in\Omega_N}$ are restrictions of a homogeneous product measure on $(\mathcal{P}(\mathbb{N}))^{\Omega_\infty}$ to $(\mathcal{P}(\mathbb{N}))^{\Omega_N}$. We denote by $\Theta(\{i\}) = E^\mu(x_\xi(i))$ and refer to Θ as the intensity measure.

The basic theorem reads (recall (1.32) - (1.47) for notation.):

Theorem 2 (Multiple space-time scale behavior)
Assume that $H(z;u,v) \equiv d$. Then the following relations hold for the different cases for (j,k).
case $j < k$:

$$(1.52) \qquad \mathcal{L}((x^N_{\xi,k}(s\beta_j(N)))_{s\geq 0}) \underset{N\to\infty}{\Longrightarrow} \delta_{\{(A^k_s)_{s\geq 0}\}}$$

case $j = k$:

(1.53) $\mathcal{L}\big((x^N_{\xi,k}(s\beta_j(N)))_{s\geq 0}\big) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Y^k_s)_{s\geq 0})$

case $j > k$: If $S(N) \to \infty$ but $S(N)/N \to 0$ as $N \to \infty$ then

(1.54) $\mathcal{L}((x^N_{\xi,k}(S(N)\beta_j(N) + t\beta_k(N)))_{t\geq 0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Z^k_{\Theta^*_{j,k}}(t))_{t\geq 0}), \quad \mathcal{L}(\Theta^*_{j,k}) = \mu^{j,k}_\Theta.$ □

We shall now give an interpretation of the results (1.52) - (1.54) in the theorem above.

(1.52): Averages over very big blocks of components (N^k-components) behave in comparably small time scales, i.e. tN^j with $k > j$, like a *deterministic dynamical system* reflecting only the level j-competition. This system reduces some of the weights to 0 and increases others. This corresponds to the disappearance of types not competitive on level j, which in turn leads to an increase in the density of the survivors.

(1.53): In this space-time scale we "see" the level j-selection taking place. Namely in the case, where the averaging in space and time match i.e. blocks with N^j components are averaged and viewed in time scale sN^j, we see a *measure valued diffusion* involving *two drift terms*. One term is the drift towards the current mean measure of types (the effect from the migration) and a second term, a selection term with selection matrix $\{V_j(u,v)\}_{u,v}$ (reflecting the competition on the j-th level). Since inferior types, inferior is meant here in term of k-level selective value φ^u_k with $j > k$, have already long disappeared, the φ^u_k terms with $k > j$ are not yet felt and hence we see *only* the *level j competition*.

(1.54): Looking finally at the case $j > k$, we see the same picture we would see for a process without selection but starting in a reduced initial state, namely with type intensity given by a measure with $Ex_\xi(i) = \tau^\theta_j(i)$ rather than with $\Theta(\{i\})$: looking at time scales $S(N)\beta_j(N) + t\beta_k(N)$ with $j > k$ and with t varying, means that the selection has, for $S(N)$ large, basically removed inferior (w.r.t. level j) types and we see effectively an *evolution* involving *only resampling* and *migration* of the remaining types, which survived the lower ($< j$) level competition, as well as the level j competition itself.

Remark The next observation is that in the presence of selection we obtain, compared to a model involving only resampling and migration, two additional objects namely $(Y^k_t)_{t\geq 0}$ and $(A^k_t)_{t\geq 0}$ resp. $(\tau_k)_{k\in\mathbb{N}}$ which are difficult to handle due to the *nonlinear character* of the interaction between different types.

Open Problem It should be correct that we can remove the assumption on H in the Theorem 2 entirely, if we use the modified ingredients of the multiple space-time scale analysis. We will see in section 5 however what the technical obstacles are in proving this.

(d) Construction of the Quasi-equilibria

In the previous subsection we studied the behavior of block averages of $(x^N_\xi(t))_{\xi\in\Omega_N}$ in blocks centered at the "origin" and observed in various time scales in the limit $N \to \infty$. Our original interest went further, we wanted to exhibit *quasi-equilibria*, that is states which are stable over long (but not infinite) periods of time. This means in precise mathematical terms that we want to know whether for $S(N) \uparrow \infty$ but $S(N)/N \to 0$ as $N \to \infty$ the (time-rescaled) complete spatial process

(1.55) $(X^N(S(N)\beta_j(N) + t))_{t\geq 0},$

has as $N \to \infty$ a limit in law and we want to know what it looks like. We will denote the limit process of (1.55) as $N \to \infty$ by $(X^{\infty,j}(t))_{t\geq 0}$ *(j-th order quasi-equilibrium)*. The next task is now to actually define for every $j \in \mathbb{N}$ the candidate for the above mentioned limiting process.

The evolutions $(X^{\infty,j}(t))_{t\geq 0}$ will be processes with state space (see (1.30) for Ω_∞)

(1.56) $(\mathcal{P}(\mathbb{N}))^{\Omega_\infty}.$

More specifically, they will be represented as *Fleming-Viot processes with emigration and immigration* evolving in a *random medium* where the latter is provided by the immigration sources. The random medium of immigration sources is for every fixed j constituted by a random field

$$(1.57) \qquad (\Theta^{(j)}_\xi)_{\xi \in \Omega_\infty}, \; \Theta^{(j)}_\xi \in \mathcal{P}(\mathbb{N}).$$

This random field will turn out to be a functional of a *Markov random field on a tree*, where the tree structure reflects the *dependence structure* of the process *induced* by the *migration*.

We shall proceed in two steps, first we construct the evolution for a given realization of the random field and then in a second step we actually construct this random field.

Step 1 (The evolution mechanism of $(X^{\infty,j}(t))_{t\geq 0}$ for given medium.)
Fix a realization of the random field $(\Theta^{(j)}_\xi)_{\xi\in\Omega_\infty}$. Then the evolution of $X^{\infty,j}(t) = \{x^{\infty,j}_\xi(t)\}_{\xi\in\Omega^\infty}$ is a Markov process on $(\mathcal{P}(\mathbb{N}))^{\Omega_\infty}$ evolving according to the following rules:

- All components $x^{\infty,j}_\xi(t)$ evolve independent of each other.
- For every $\xi \in \Omega_\infty$, the law of the component process $(x^{\infty,j}_\xi(t))_{t\geq 0}$ is the unique solution of the $(L^0_{\Theta^{(j)}_\xi}, \mu^{j,0}_\Theta)$, martingale problem, where the generator L^0_Θ is defined in (1.35) and the initial distribution is $\mu^{j,0}_\Theta$ from (1.47).

In other words every component is a Fleming-Viot process (on the state space $\mathcal{P}(\mathbb{N})$) with both emigration and immigration, the latter from the (time independent) source $\Theta^{(j)}_\xi$. In particular all components in a j-ball are "coupled" after averaging over the medium due to the fact that they use the same immigration source.

Step 2 (Construction of the random field $(\Theta^{(j)}_\xi)_{\xi\in\Omega^\infty}$.)
Introduce

$$(1.58) \qquad \tilde\Omega^j_\infty = \{\xi \in \Omega_\infty \,|\, \xi_1 = ... = \xi_j = 0\}$$

and write

$$(1.59) \qquad \Omega_\infty = ((\mathbb{N})^j)^{\tilde\Omega^j_\infty}.$$

Now define the random field as an i.i.d. collection of fields indexed by $\mathbb{N}^j$, namely:

$$(1.60) \qquad \mathcal{L}((\Theta^{(j)}_\xi)_{\xi\in\Omega_\infty}) = \bigotimes_{\tilde\Omega^j_\infty} \mathcal{L}((\hat\Theta^j_\eta)_{\eta\in\mathbb{N}^j}).$$

It now suffices to define the marginal laws appearing in the r.h.s. of (1.60) and which are laws of the smaller fields indexed by $(\mathbb{N})^j$. This law corresponds to a *Markov field* on a *rooted tree of depth* $j+1$, which we now construct.

Let Θ^* be an element of $\mathcal{P}(\mathbb{N})$ to be specified later. Define the following collection $\{(\tilde\Theta^{(i)})_{i\in\mathbb{N}}$, $(\tilde\Theta^{(i,\ell)})_{i,\ell\in\mathbb{N}}, \ldots, (\tilde\Theta^{(i_1,\ldots,i_j)})_{i_k\in\mathbb{N},k=1,\ldots,j}\} = (w_1, w_2, \ldots, w_j)$. For this string of length j we require the *Markov property*. We specify the distribution of this j-tuple by specifying $\mathcal{L}(w_1), \mathcal{L}(w_2|w_1)$, $\mathcal{L}(w_3|w_2,w_1), \ldots$ of $\mathcal{P}(\mathbb{N})$-valued random variables:

$$(1.61) \qquad (\tilde\Theta^{(i)})_{i\in\mathbb{N}} \; : \; \text{i.i.d. } \Gamma^j_{\Theta^*}\text{-distributed}$$

$$(1.62) \qquad (\tilde\Theta^{(i,\ell)})_{i\in\mathbb{N},\ell\in\mathbb{N}} \; : \; \mathcal{L}[\tilde\Theta^{(i,\ell)} \,|\, (\tilde\Theta^{(k)})_{k\in\mathbb{N}}, \tilde\Theta^{(m,n)} \text{ for } (m,n)\neq(i,\ell)] = \Gamma^{j-1}_{\tilde\Theta^{(i)}}$$

$$\vdots$$

Then set for $\eta = (\eta_1, ..., \eta_j) \in \mathbb{N}^j$ and $\widehat\Theta^{(j)}_\eta$ from the r.h.s. of (1.60):

$$(1.63) \qquad \widehat\Theta^{(j)}_\eta = \tilde\Theta^{(\eta_j,\ldots,\eta_1)},$$

which specifies then the law on the r.h.s. of (1.60), up to the choice of Θ^*. We therefore complete the definition by setting (recall (1.40) - (1.42)):

(1.64) $$\Theta^* = \tau_j^{\Theta}.$$

Together with the construction of step 1 this completes our definition of $(X^{\infty,j}(t))_{t\geq 0}$.

Now we are ready to prove the existence of quasi-equilibria and to identify them explicitly via the object $(X_t^{\infty,j})_{t\geq 0}$ just constructed. Namely:

Theorem 3 (Quasi equilibria)

Assume $S(N)\uparrow\infty$ with $S(N)/N\to 0$ as $N\to\infty$. Then for every $j\in\mathbb{N}$:

(1.65) $$\mathcal{L}((X^N(S(N)\beta_j(N)+t))_{t\geq 0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((X_t^{\infty,j})_{t\geq 0}).$$

In particular for all $\xi\in\Omega_\infty$:

(1.66) $$\mathcal{L}(x_\xi^N(S(N)\beta_j(N))) \underset{N\to\infty}{\Longrightarrow} \mu_\Theta^{j,0}. \qquad \square$$

Remark Since we think of a typical observer of our system having an observation time *negligible* compared to the *age of the system*, it is the quasi-equilibria which are of interest to him and *not* the eventual equilibrium state.

(e) The stability versus clustering dichotomy

In this section we shall explore the influence of the migration mechanism on the longterm behavior of our system and we shall prove that there is a dichotomy between the case of strong and weak interaction through migration. To exhibit this dichotomy we introduce the *interaction chain* and its *entrance law*. To this end we consider a *combination of several space and time scales* and consider the *joint* distribution in *time* of the rescaled systems. In the limit $N\to\infty$ this will define the interaction chain $(Z_k^j)_{k=-j-1,\dots,0}$, see (1.45) - (1.47). The structure of the interaction chain will be qualitative different for strong and weak interaction by migration.

We begin with the time scales. We shall observe the process after a very long time, this means here after time $S(N)\beta_j(N)$, at a *sequence of time points* spaced with differences of decreasing order of magnitude as follows:

$$S(N)\beta_j(N),\quad S(N)\beta_j(N)+t_1\beta_{j-1}(N),\quad S(N)\beta_j(N)+t_1\beta_{j-1}(N)+t_2\beta_{j-2}(N),\dots.$$

This allows us to capture the *time structure* of our process in quasi-equilibrium.

At these time points we now consider in addition a *spatial rescaling* by building averages over blocks of a suitable size, that is, look at $x_{\xi,k}^N$ after time $\beta_j(N)$ for a certain time interval for which we choose the time scale such that it allows for fluctuations over this time interval which then has to have size $t_{j-k}\beta_k(N)$, where $k\leq j$ and $t_i\in[0,\infty)$.

We *combine* the time and the spatial aspect from above and consider the $(j+1)$-vector:

(1.67) $$\{x_{\xi,j+1}^N(S(N)\beta_j(N)),\ x_{\xi,j}^N(S(N)\beta_j(N)),$$
$$x_{\xi,j-1}^N(S(N)\beta_j(N)+t_1\beta_{j-1}(N)),\dots,x_{\xi,0}^N(S(N)\beta_j(N)+\sum_{i=1}^{j}t_i\beta_{j-i}(N))\}.$$

This way we have obtained a description of the *space-time structure* on various macroscopic space and time scales.

In order to understand the object defined in (1.67) we pass to the limit ($N\to\infty$) and study the limiting object the socalled interaction chain. Indeed it is a corollary of our Theorem 2 (compare [DGV] Theorem 0.10), that in the limit of $N \to \infty$ the random vector in (1.67) converges in law if $S(N) \uparrow \infty$ with $S(N)/N \to 0$ as $N \to \infty$ to the interaction chain:

(1.68) $\quad (Z^j_k)_{k=-j-1,\dots,0}.$

We shall therefore in the sequel study directly the object $(Z^j_k)_{k=-j-1,-j,\dots,0}$ for various j, but of course with the implications for (1.67) in mind.

Now we have to distinguish two cases for the fitness function leading to different behavior. Assume here that we are in the case where (recall (1.21)) the initial law $\mu \in \mathcal{T}_e$ with intensity measure Θ satisfies:

(1.69) $\quad \mathcal{N}^\Theta = \emptyset,$

that is, the maximum of the fitness function is *not attained*.

In the other case where the maximum of the fitness function is attained i.e. where $\mathcal{N}^\Theta = \{v_1, v_2, ..., v_k\}$, the system behaves in the same way as the system with state space $\{v_1, ..., v_k\}^{\Omega_N}$ and *without* selection, the only question we need to settle is what will be the intensity measure $\Theta(\mu)$ of the final equilibrium on $\{v_1, v_2, \dots, v_k\}$. Given the solution to that problem the results for this situation can be found in [DGV]. We shall use the multiple space-time scale analysis to say something about the question what $\Theta(\mu)$ looks like in part g) of this chapter.

To obtain something meaningful in the situation $\mathcal{N}^\Theta = \emptyset$, where types keep disappearing as we increase the level j of the quasi-equilibria considered, we have to relabel our types by size ordering them. Let φ_j be a one-to-one map $\mathbb{N} \hookrightarrow \mathbb{N}$ satisfying

(1.70) $\quad \tau_j(\varphi_j(i)) \geq \tau_j(\varphi_j(i')) \qquad \forall i < i'\,, i, i' \in \mathbb{N}.$

In other words we *size order* the types on every level $0, 1, \dots, j, j+1$ according to the mean frequency distribution, associated with the time scale β_j. Define:

(1.71) $\quad \hat{\tau}_j = \varphi_j(\tau_j).$

Furthermore introduce the *size ordered interaction chain* (with respect to level j) at level j by:

(1.72) $\quad (\hat{Z}^j_k)_{k=-j-1,\dots,0} \qquad$ has transition kernels K_{-k} at time $-k$
but starts with $\hat{Z}^j_{-j-1} = \hat{\tau}^\Theta_j$.

This object now allows us to exhibit the influence of the migration mechanism on the non equilibrium behavior of our system. It turns out that this influence depends also on the structure of the initial intensity measure Θ, since the behavior of $\hat{\tau}^\Theta_j$ for $j \to \infty$ depends on Θ. We have to distinguish the monotype limiting situations with $\widehat{\tau}^\Theta_j \Rightarrow \delta_0$ from the ones with multitype limiting structure. Since we shall establish later in Theorem 6 that $\{\hat{\tau}^\Theta_j\}_{j\in\mathbb{N}}$ is tight under certain conditions or "spreads out" (if $\hat{\Theta}$ does not have a superexponentially vanishing tail) we can in many cases, in fact the interesting ones, obtain a nice limiting object $(\widehat{Z}^\infty_k)_{k\in\mathbb{Z}^-}$.

The following *dichotomy* between strong and weak interaction occurs depending on Θ, which will show up again in Theorem 5 once we look at speeds of decay of inferior types. Here $\Longrightarrow$ denotes again *weak* convergence of laws.

Theorem 4 (stability versus clustering)

(a) Assume that $\mathcal{N}^\Theta = \emptyset$. Recall (1.71) for $\widehat{\tau}_j$ and (1.72) for $\widehat{Z}^j_k$.

case 1: $\sum\limits_0^\infty c_k^{-1} = \infty.$

Then we have clustering:

(1.73) $$\left[\mathcal{L}((\hat{Z}^j_k)_{k=-j-1,\dots,0}) - \int\limits_{\mathbb{N}} \delta_{\{Z_k=u, \forall k\in\mathbb{Z}^-, k\geq -j-1\}} \hat{\tau}_j(du)\right] \underset{j\to\infty}{\Longrightarrow} 0.$$

case 2: $\sum_0^\infty c_k^{-1} < \infty$.
Then we have pseudo-stability:

$$\text{(1.74)} \qquad \Big[\mathcal{L}((\hat{Z}_k^j)_{k=-j-1,\dots,0}) - \Lambda^j_{\hat{\tau}_j^\Theta}\Big] \underset{j\to\infty}{\Longrightarrow} 0$$
$$\text{with } \Lambda_\rho^j := \mathcal{L}((Z_k^\infty(\varrho))_{k=-j-1,\dots,0}),$$

where $\mathcal{L}((Z_k^\infty(\rho))_{k\in\mathbb{Z}^-}) = w - \lim_{j\to\infty}\mathcal{L}((Z_k^j)_{k=-j-1,\dots,0}|Z^j_{-j-1} = \rho)$ *and* Λ_ρ^j *is supported on paths whose states put positive weight on every type* u *with* $\rho(\{u\}) > 0$ *a.s. and satisfy* $Z_k^\infty(\rho) \to \rho$ *as* $k \to -\infty$.

(b) *Assume that* $\sum c_k^{-1} < \infty$ *and that* $(\widehat{\tau}_j^\Theta)_{j\in\mathbb{N}}$ *converges weakly as* $j\to\infty$ *to* $\widehat{\tau}_\infty^\Theta \in \mathcal{P}(\mathbb{N})$. *Then there exists an entrance law* $(\widehat{Z}_k^\infty)_{k\in\mathbb{Z}^-}$ *which satisfies* $\hat{Z}_{-\infty}^\infty = \widehat{\tau}_\infty^\Theta$ *and*

$$\text{(1.75)} \qquad \mathcal{L}((\widehat{Z}_k^\infty)_{k\in\mathbb{Z}^-}) = \mathcal{L}\Big(w - \lim_{j\to\infty}\mathcal{L}((\widehat{Z}_k^j)_{k=-j-1,\dots,0}|\widehat{Z}^j_{-j-1} = \widehat{\tau}_j^\Theta)\Big). \qquad \square$$

Remark The theorem says that for large range interaction through migration ($\sum c_k^{-1} < \infty$) the system preserves every type which is globally present in times of order $\beta_j(N)$ in every local window of observation in *coexistence* with all other types. This holds if $\widehat{\tau}_j^\Theta$ does not converge to a monotype situation, i.e. δ_0 (compare Theorem 6).
For small range interaction ($\sum c_k^{-1} = \infty$) large regions develop where one *single type prevails locally* and this phenomenon occurs independent of the Θ chosen.

Remark The effect of the migration can be annihilated if the initial state is such that $\widehat{\tau}_j^\Theta$ converges to δ_0 as $j\to\infty$. This will happen if initially the system is strongly concentrated on few types (see (1.91). But in this case the whole system will consist almost entirely of a certain type which first of all is with respect to the time scale N^j the fittest one and second had the highest intensity among those which are equally fit with respect to that time scale.

(f) Selection speed and migration

The influence of the migration on the system manifests itself also in the *speed* at which inferior types are removed from the system. This removal happens at exponential scales and we can define as the speed of removal of a type the exponential rate of decay of the mean intensity of that type. A pseudo-stable system removes inferior types faster, since the migration brings everybody together rapidly and competition starts. On the other hand in the clustering situation inferior types can hold out longer by living in monotype regions. So for fixed N one would expect to see in the case of transient (symmetrized) migration exponential decay of the intensity of inferior types and in the recurrent case subexponential decay.

Mathematically this phenomenon is described in the mean field limit $N\to\infty$ as follows:

Theorem 5 (migration and effectiveness of competition)
Assume that $\mathcal{N}^\Theta = \emptyset$ *and the intensity* $\Theta \in \mathcal{P}(\mathbb{N})$ *is strictly positive.*

a) *Now fix a type* $u\in\mathbb{N}$ *with* $\varphi_j^u > 0$. *Define:*

$$\text{(1.76)} \qquad \big[\lim_{N\to\infty}\frac{1}{t}logE(x^N_{\xi,j}(t\beta_j(N)))(u)\big] = \lambda_j(t,u) \quad u\in\mathbb{N},$$

$$\text{(1.77)} \qquad \lambda_j^+(u) := \limsup_{t\to\infty}\lambda_j(t,u), \qquad \lambda_j^-(u) := \liminf_{t\to\infty}\lambda_j(t,u).$$

Then with d_j as defined in (1.36) and φ_j^u as defined in (1.7):

$$\begin{aligned}(1.78)\qquad \lambda_j^{+,-}(u) &\geq -d_j \cdot \max_v(\varphi_j^v - \varphi_j^u)\cdot s,\\ \lambda_j^{+,-}(u) &> -d_j \max_v(\varphi_j^v - \varphi_j^u)\cdot s \quad \text{for } s \leq s_0 \quad \text{for some } s_0 > 0.\end{aligned}$$

b) For $j\to\infty$ the decay rate $\lambda_j^{+,-}(\cdot)\leq 0$ of the j-block density of type v in time scales of order $S(N)\beta_j(N)$ depends on (c_k) as follows: Define $I_j=\{u|\varphi_j^u<M\}$. Then

$$\begin{aligned}(1.79)\qquad \sum_0^\infty c_k^{-1}<\infty &\;\Rightarrow\; (\sup_{u\in I_j}\lambda_j^+(u)) \underset{j\to\infty}{\longrightarrow} \lambda_\infty<0\\ \sum_0^\infty c_k^{-1}=\infty &\;\Rightarrow\; (\inf_{u\in I_j}\lambda_j^-(u)) \underset{j\to\infty}{\longrightarrow} 0. \qquad \square\end{aligned}$$

Remark To show that actually a decay rate $\lambda_j(u)=\lambda_j^+(u)=\lambda_j^-(u)$ exists amounts to a very complex *large deviation problem* via the duality relation. We do not have the space here to discuss this question and therefore just exhibit the phenomenon that selection speed is depending heavily on the migration mechanism.

(g) The sequence of survivor frequencies $(\tau_k)_{k\in\mathbb{N}}$ and abundance of species

In this section we present two new phenomena induced by the selection: the survivor chain and its evolution and the increasing variety of species under selection. If we consider the evolutionary properties of our system now on a more macroscopic scale, a natural question is: *how does the spatial density of types change from one evolutionary epoch to the next*? That is, we want to understand how the sequence $(\tau_k)_{k\in\mathbb{N}}$, or the size ordered one denoted $(\hat{\tau}_k)_{k\in\mathbb{N}}$ (recall (1.71)) evolves, as a function of k. We need some preparation.

We shall need the following maps Π_B acting on elements ρ of $\mathcal{P}(\mathbb{N})$ and producing new probability measures on $\mathbb{N}$ according to the following rule (using the convention $\frac{0}{0}=0$)

$$(1.80)\qquad [\Pi_B(\rho)](A) = \rho(A/B).$$

We can view Π_B in a natural way also as a map

$$(1.81)\qquad \Pi_B:\mathcal{P}(\mathbb{N})\longrightarrow\mathcal{P}(B)$$

and we will *not* use a different notation if we do so.

Next we introduce what will turn out to be the sequence of supports of the $(\tau_k^\Theta)_{k\in\mathbb{N}}$:

$$(1.82)\qquad \mathcal{N}_0^\Theta = supp\,\Theta \qquad \Theta\in\mathcal{P}(\mathbb{N})$$

$$(1.83)\qquad \mathcal{N}_j^\Theta = \{u\in\mathbb{N}\,|\,\varphi_j^u = \max_{v\in\mathcal{N}_{j-1}^\Theta}\varphi_j^v\}.$$

Apparently $\mathcal{N}_j^\Theta$ is a nonincreasing sequence of subsets of $\mathbb{N}$ which describes the types which are maximal if we replace the fitness function φ by the expansion in (1.7) up to the j-th term. Hence $\mathcal{N}_j^\Theta$ consists of the types which where *present at time* 0 and *survive* until times of order $\beta_j(N)$.

The crucial object needed to describe the evolution of the sequence $(\tau_k^\Theta)_{k\in\mathbb{N}}$ are the following maps $\mathcal{H}_j^\Theta$. They will tell us how to recursively determine the sequence $(\tau_k^\Theta)_{k\in\mathbb{N}}$ of survivor frequencies. Recall from relation (1.44) the quantity A_t^j. Then we define

$$(1.84)\qquad \begin{aligned} \mathcal{H}_j^\Theta \quad &: \quad \mathcal{P}(\mathcal{N}_j^\Theta) \hookrightarrow \mathcal{P}(\mathcal{N}_{j+1}^\Theta) \\ \mathcal{H}_j^\Theta(\rho) \quad &= \quad \lim_{t\to\infty} A_t^j \quad \text{if } A_\infty^{j-1} = \rho, \qquad j \in \{0,1,2,\dots,\}. \end{aligned}$$

Note that by definition we then have

$$(1.85)\qquad \tau_j^\Theta = \mathcal{H}_j^\Theta o \dots o \mathcal{H}_0^\Theta(\Theta)$$

In general it is not easy to analyze the action of the nonlinear map $\mathcal{H}_j^\Theta$. We will now consider a special case in order to point out some very interesting effects in the simplest possible setting. In the sequel we shall need certain hypotheses on φ and on the intensity measure Θ.

Recall that χ_N is the fitness function $\chi_N(u) = \sum_r \frac{1}{N^r}\varphi_r^u$. Assume that for every N the set

$$(1.86)\qquad \chi_N(\{u|\Theta(\{u\})>0\}) \text{ has at most one limit point namely the supremum.}$$

For example, this hypothesis is satisfied if φ_r^u takes only the values 0 and M.

Then it is possible to consider the relabeling of types according to the size of the fitness function namely index 0 is assigned to the type of *lowest* fitness and then types are consecutively labeled in the order of their fitness value. The unique image of Θ under this size ordering map is called $\tilde{\Theta}$. We can apply this map $\tilde{}$ also to τ_j^Θ. Note that $\tilde{\tau}_j^\Theta$ orders then types according to fitness on the set $\{i\in\mathbb{N}|\tau_j^\Theta(i)>0\} = \mathcal{N}_j^\Theta$, that is the ordering is j-dependent.

We now formulate a *hypotheses* involving two assumptions on the initial intensity Θ, where $\tilde{\Theta}$ again denotes the "reversed" order statistics of Θ induced by φ as introduced above.

(i) $\tilde{\Theta}$ is monotone.

(ii) We are in one of the following three different cases of intensity measures Θ:

$$(1.87)\qquad \begin{aligned} A \quad &: \quad \underline{\lim}_{u\to\infty}\tilde{\Theta}(u+v)/\tilde{\Theta}(u) = 1 \\ B \quad &: \quad \lim_{u\to\infty}\tilde{\Theta}(u+v)/\tilde{\Theta}(u) = q^v \quad 0<q<1, \quad \forall v = 1,2,\dots \\ C \quad &: \quad \overline{\lim}_{u\to\infty}\tilde{\Theta}(u+v)/\tilde{\Theta}(u) = 0. \end{aligned}$$

In this special situation we can exhibit below an interesting phenomenon.

Theorem 6 (survivor frequencies, increasing variety of species)
Denote $\mathcal{L}(x_\xi(0))$ by μ and by Θ the intensity measure of the distribution μ.
Assume in (a) - (c) below that $H(z,v,u)\equiv d$, $\mathcal{N}^\Theta = \emptyset$.

(a) Then the map $\mathcal{H}_j^\Theta$ is given by

$$(1.88)\qquad \mathcal{H}_j^\Theta = \Pi_{\mathcal{N}_{j-1}^\Theta} \qquad j = 1,2,\dots$$

(b) $\tau_j^\Theta \Rightarrow 0$ as $j\to\infty$ in the vague topology.

(c) *Suppose that the hypotheses (1.87) in addition to (1.86) is satisfied. Then the size ordered (1.70) $\widehat{\tau}_j$, satisfy depending on Θ:*

$$(1.89) \qquad \textit{Case A:} \quad \hat{\tau}_j(\cdot)/\hat{\tau}_j(\{0\}) \underset{j\to\infty}{\overset{vague}{\Longrightarrow}} n(\cdot), \ n(\cdot) \textit{ is counting measure on } \mathbb{N},$$

$$(1.90) \qquad \textit{Case B:} \quad \hat{\tau}_j(\cdot) \underset{j\to\infty}{\Longrightarrow} \textit{ geometric distribution with parameter } q,$$

$$(1.91) \qquad \textit{Case C:} \quad \hat{\tau}_j(\cdot) \underset{j\to\infty}{\Longrightarrow} \delta_0(\cdot).$$

(d) *Assume that $\mathcal{N}^\Theta = \mathcal{N} \neq \emptyset$. Then*

$$(1.92) \qquad \tau_j^\Theta \underset{j\to\infty}{\Longrightarrow} \tau_\infty^\Theta, \quad \tau_\infty^\Theta(\{u\}) > 0 \quad \forall u \in \mathcal{N}. \qquad \square$$

Remark The paradoxical situation that selection increases the variety of species (as expressed in (1.89)) comes from the fact that the frequencies conquered from inferior species introduces an additional symmetrizing effect for the surviving types. This effect is not present once we have that the maximum of the fitness function is attained in the support of the intensity measure Θ of the initial law $\mu \in \mathcal{T}_e$.

Remark It remains to exhibit the effect described above for more general initial state and for weighted sampling (instead of $H \equiv$ const), however we do not have the space here to carry that out.

2 The combined effect of selection, mutation and recombination

(a) Motivation and introduction to mutation

In this section we formulate the basic mechanisms of population genetics and evolutionary biology in the setting of multiple space-time scale analysis for probability measure-valued diffusions. In addition to the diversity produced by the (spatially structured) initial state analyzed in the previous section 1, we now introduce the *mutation* and *recombination* mechanisms as new sources of *diversity* in the population.

We introduce these new mechanisms in order to approach the basic problem of evolutionary biology, namely to understand how these different mechanisms *combine* to produce evolutionary development over long time scales and how the *competition* between the selection mechanism on the one hand and mutation and recombination on the other hand produces new phenomena. The first problem is to actually model the mutation appropriately. We next briefly review the background of the model, in order to be able to explain the specific structure of mutation we consider.

Starting with the seminal contributions of Fisher and Wright in the 1930's there have been competing views of the relative importance to biological evolution of the different genetic mechanisms such as selection, recombination, sampling replacement and migration. The task of the mathematical analysis would be to give a framework in which such questions could be rigorously discussed. For example *Wright's shifting balance theory* of evolution emphasized the importance of the spatial isolation of small subpopulations. However based on the empirical evidence which has been studied since then, there appears to be no conclusive evidence for either this or the competing viewpoints (cf. Wade (1992)). Current research in this field is largely devoted to obtaining new empirical evidence which can discriminate between competing theories (which itself involves new methodological questions).

A new perspective has also arisen from *optimization problems*, in particular, with the advent of the development of *g*enetic algorithms (Holland (1992)). These are Monte Carlo type stochastic optimization algorithms which are based on the mechanisms of finite population sampling, mutation, recombination and selection. This provides motivation for studying variations and combinations of these mechanisms even if there is no evidence that they occur in the biological setting.

On the theoretical and mathematical side because these models involve *spatially* distributed and *nonlinear* stochastic models, new mathematical tools for such an analysis must be developed. One of the motivations of the present work is to develop an approach to this problem. In particular the framework of multiple time scale analysis based on the hierarchical mean-field limit provides a simplified setting in which to investigate the interplay of the above mentioned mechanisms mutation, selection migration and recombination in different time scales.

Outline In the rest of this subsection we discuss in 2(b) the appropriate mathematical set up for the mutation and recombination mechanism in our context, in section 2(c) we construct formally the process via a martingale problem and in 2(d) and 2(e) we discuss properties of the longterm behavior and of related mean-field limits.

(b) Description of the mutation and recombination mechanisms

In this section we will extend the model of section 1 by adding the mechanisms of *mutation* and *recombination*. We shall now discuss in paragraphs (i) and (ii) the structure of these two new mechanisms for which we need to assume *specific properties* so that they are suitable for the purpose mentioned in the section on motivation above. We will explain this in some detail in this subsection, since already setting up the model is subtle.

(i) A *mutation* occurs when an offspring has a type different from that of the parents. In our model the set of types will be assumed to be structured and the mutation mechanism we shall consider here will explicitly depend on this structure.

The basic idea of the model is that mutation in the space of types proceeds as follows: during the evolutionary process new types are created out of the existing ones in a random manner and the new types have two properties, they represent *s*mall perturbations of the types present in the system at that time and the fitter types are more *stable* against mutation. We make this precise and we will explain this concept from a biological point of view below (Wright's adaptive landscapes). Since fitness represents differential reproduction rates and higher fitness means more complex organization, this means that the successive fitness levels are achieved via mutations in successively longer time scales. This fits the set up of chapter 1 for the selection term and the mutation rates have now to be chosen such that they also fit this structure.

A key point for the longtime behavior is that in such a situation *both* fitness levels and mutation rates determine the longterm success of different types. More precisely, in order for a type of given fitness to survive over long time scales, it is essential that its mutation rate to types of lower fitness be small relative to the difference in fitness. In fact in the model we consider there is a hierarchy of fitness levels that could be viewed as levels of organization that arise over *successively longer time periods* and which are in addition *progressively more stable* in the sense that the rate at which they degenerate to lower order types is smaller. In order to model a system of this type we will formulate a specific mutation structure in (2.2) below. First note that such a mutation structure is also a key component of stochastic optimization algorithms, namely, when close to the optimum the randomly generated element in the algorithm should be inferior.

We next describe the main idea concerning the biological content of the notion of a stable type which was introduced by Wright.

This idea of stable type is related to Wright's notion of peaks in the *adaptive landscape*. In his model the space of types is structured and the fitness function over this type space has local maxima separated by valleys. The stability of a peak is determined by the depth of the valley separating it from neighboring peaks. In terms of the gene structure the reason for these high peaks or equivalently the complexity of a type is due to *epistasis* - this refers to interaction between gene loci and in particular the situation in which to go from one peak to another requires changes at each of a set of gene loci - the successive peaks would then involve more and more loci. In addition it is assumed that the intermediate types have lower fitness.

To give a precise formulation to the above concept in our framework it is useful to think of the types as labeled by the heights of the corresponding adaptive peaks so that the type space also gets the structure of some hierarchical group. The mutation will then be modeled by a particular Markov chain on this hierarchical group, which is adapted to the hierarchical structure used to model selections.

On a qualitative level this leads to a system of mutation rates as follows. The ideas described above mean, following the results of chapter 1, that at high fitness levels the evolution occurs over times of the order of N^k and consequently if we model the mutation rate such that it is comparable to selection on that level, this rate should be of order $1/N^k$. This means in particular, if a mutant of lower fitness is produced say the fitness difference is $1/N^{k-1}$ or more, then in time scale N^k the difference becomes $O(N)$ and it immediately disappears. On the other hand if the difference is $1/N^k$ it persists in time scale N^k.

In order to define the model mathematically we need the following ingredients.
Recall that we defined the fitness of a type $u \in \mathbb{N}$ to be of the form $\chi(u) = \sum_{k=0}^{\infty} \varphi_k^u N^{-k}$ with $\varphi_k^u \in \{0, \dots, M\}$ so that we can also identify types with strings $(\varphi_0, \varphi_1, \dots)$. In this part where the diversity is created by mutation, from one initial type, which we choose as the 0-string, we consider the case where $\varphi_r^u \neq 0$ only for finitely many r (the number of those r's depends on u though). In that context it is useful to introduce the notion of a *level-j-type*.

Definition 2.1 (the level of a type, level set E_K)
(a) We define the *level of type* u as follows (recall M is the maximal value of φ_j^u): Type u is said to be of level k if and only if

$$\varphi_j^u = M \quad \forall\, j < k, \quad \varphi_k^u \neq 0, \quad \varphi_j^u = 0 \quad \forall j > k. \tag{2.1}$$

Note that the fitness of a level k type is (strictly) greater than the fitness of any lower level type.

(b) Let E_K denote the set of types of at most level K, that is types u corresponding to strings $(\varphi_k^u)_{k \in \mathbb{N}}$ with $\varphi_k^u = 0 \;\forall\, k > K$. □

Then we introduce the *mutation transition rates* $m(u,v)$ based on matrices for transition rates of level k types denoted $m_{k,k}(\cdot,\cdot)$ and rates $m_{k,\ell}(\cdot,\cdot)$ between different level types:
Definition 2.2 (Mutation rates)

$$\begin{aligned} m(u,v) &= \tfrac{m_{k,k}(u,v)}{N^k} && \text{if } u,v \text{ are both level } k \\ &= \tfrac{m_{k,k+1}(u,v)}{N^{k+1}} && \text{if } u \text{ is level } k \text{ and } v \text{ is level } k+1 \\ &\leq \tfrac{m_{k,\ell}(u,v)}{N^k} && \text{if } u \text{ is level } k \text{ and } v \text{ is level } \ell < k \\ &\quad 0 && \text{otherwise} \end{aligned} \tag{2.2}$$

where $0 \leq m_{k,\ell}(u,v) \leq M$ and $M \leq N-1$. □

Here $m(u,v)$ is the transition rate that a type u mutates and produces a type v. The form of $m(u,v)$ implies that level k types are more stable than lower level types and mutation to higher level types proceed always by one step. We have not introduced a special notation for the mutation rates to lower order types, since in the further analysis we need only that they are bounded as stated. Later M will be fixed and $N \to \infty$.

Let us return to the qualitative aspects and interpretation for evolutionary models of a mutation rate given by $m(u,v)$ in (2.2) in order to see they fit the requirements posed in the introduction (Wright adaptive landscape).

In the processes of interest for population genetics we start originally with a very simple "organism", in our formulation a type with $\varphi_r^u = 0$ for all r. Then the evolution process starts first driven by mutations and then proceeds to produce more complex and fitter types. The mutation mechanism described above satisfies two qualitative requirements: (i) a type at a certain level of complexity can only mutate to types of comparable or lower levels of complexity at the natural time scale of his level, (ii) mutations to lower level types occur at a rate much smaller than the selective disadvantage of this mutant, (iii) a type can mutate to the next higher level of complexity but this occurs with a lower order rate.

This means that the more complex a type is the more stable it is against random perturbations, since in the level j-timescale lower order mutants, even though they occur, will be wiped out in shorter time scales. In other words, the more complex types evolve at a slower rate than the simpler ones, both as far as selection and mutation goes.

Let us look closer at the connections with *Wright's theory.* The mutation matrix $m(u,v)$ fits Wright's adaptive landscapes, which is seen as follows. Consider a sequence of adaptive peaks, labeled by $k \in \mathbb{N}$, in the adaptive landscape in which to get to the $(k+1)$st level one must climb through the previous k levels. This resembles in our model for mutation rates, the situation of a *random walk in a linear random potential.* One can pass to the next highest level only by going through a valley in the landscape of the random potential of height u_k with $e^{-u_k} = \frac{m_{k,k+1}}{N^{k+1}}$, that is, $u_k = (k+1)\log N - \log m_{k,k+1} \approx (k+1)\log N$ (recall $\log m_{k,k+1} \in (0, \log M)$). On the other hand in order to dominate in the time scale in which the $k+1$-adaptive peak is reached, a fitness difference of only $\frac{1}{N^k}$ is needed.

Finally we introduce some *notation* used frequently later on. Let m denote the total mutation rate and let $M(u,dv)$ denote the probability distribution of the type, which results if a type u undergoes mutation. We will sometimes want to consider the case of, at least partly, *state independent mutation* distribution.

Definition 2.3 (state independent mutation part)
If there exists $\varrho \in \mathcal{P}(\mathbb{N})$, $\bar{m} \in (0,1)$ with $M(u,dv) \geq \bar{m}\varrho(dv)$, then we define $\bar{M}(u,dv)$ by

$$M(u,dv) = \bar{m}\varrho(dv) + (1-\bar{m})\bar{M}(u,dv) \tag{2.3}$$

On a finite type space we can often pick the uniform distribution for ϱ. □

(ii) We now turn to the mechanism of *recombination.* This corresponds to the *diploid case* of reproduction, in which the genetic information of an offspring is determined by two genetic strings. On reproduction the two strings of the offspring are obtained by taking one string from each of two parents. Recombination refers to the case in which a string of the offspring results from cutting one string from each parent at a site and then pasting together one piece from each - this is referred to as *crossing over* (cf. J. Maynard Smith chapt. 12). The mathematical form of recombination used below is a generalization of this mechanism due to Ethier and Kurtz (1996, section 1).

Definition 2.4 (Recombination)
By recombination two types v_1 and v_2 combine to give a new type u at a rate based on the ingredients

$$\alpha(v_1,v_2) \quad \text{and} \quad R(v_1,v_2,du),$$

where $\alpha(v_1,v_2)$ is $\mathbb{R}^+$-valued bounded and symmetric on $\mathbb{N}\times\mathbb{N}$ and $R(v_1,v_2;du)$ is a probability transition kernel on $\mathbb{N}^2\times\mathbb{N}$.

At rate $\alpha(v_1,v_2)$ types v_1 and v_2 associated with the parent particles undergo recombination and the kernel $R(v_1,v_2;du)$ tells us, with which probability the type u results from recombining strings from v_1 and v_2. □

(c) Characterization of the process via a martingale problem

Since in this section the main emphasis is on selection, mutation and recombination we will again consider as diffusive term only unweighted sampling that is (recall (1.18)) $H \equiv d$. The new process incorporating mutation and recombination will again be defined via a *martingale problem.* At the point where we only define the process, we can be more general than in the discussion of the longtime behavior, where we need a more specific structure of the mutation. Let M be a transition kernel on $\mathbb{N}\times\mathbb{N}$ and R a transition kernel on $\mathbb{N}^2\times\mathbb{N}$, c,d,s,m,r are nonnegative numbers and $a(\cdot,\cdot)$, $V(\cdot,\cdot)$, $Q_x(\cdot,\cdot)$, F, $\frac{\partial F}{\partial x}$ are as in section 1. Recall (1.1) - (1.14).

We obtain the new generator $L_{M,R}$ of the martingale problem replacing L from (1.13), by including now also the *mutation* and *recombination* in the drift terms (two last expressions):

$$
\begin{aligned}
(2.4)\qquad L_{M,R}(F)(x) = \sum_{\xi\in\Omega_N}\Bigg[& c\sum_{\xi'\in\Omega_N} a(\xi,\xi')\int\limits_{\mathbb{N}} \frac{\partial F(x)}{\partial x_\xi}(u)(x_{\xi'}-x_\xi)(du) \\
& + d\int\limits_{\mathbb{N}}\int\limits_{\mathbb{N}} \frac{\partial^2 F(x)}{\partial x_\xi \partial x_\xi}(u,v) Q_{x_\xi}(du,dv) \\
& + s\int\limits_{\mathbb{N}}\int\limits_{\mathbb{N}}\int\limits_{\mathbb{N}} \frac{\partial F(x)}{\partial x_\xi}(u) V(v,w) x_\xi(dv) Q_{x_\xi}(du,dw) \\
& + m\int\limits_{\mathbb{N}} \left\{\left\{\int\limits_{\mathbb{N}} \frac{\partial F(x)}{\partial x_\xi}(v) M(u,dv)\right\} - \frac{\partial F(x)}{\partial x_\xi}(u)\right\} x_\xi(du) \\
& + r\int\limits_{\mathbb{N}}\int\limits_{\mathbb{N}} \left[\alpha(v_1,v_2)\left\{\int\limits_{\mathbb{N}} \frac{\partial F(x)}{\partial x_\xi}(u) R(v_1,v_2,du) - \frac{\partial F(x)}{\partial x_\xi}(v_1)\right\} x_\xi(dv_1) x_\xi(dv_2)\right]\Bigg].
\end{aligned}
$$

Without loss of generality we assume that $|V| \le 1$, by choosing s appropriately. Here V will have the additive structure of (1.7) for most of the statements and we will indicate, when greater generality is possible.

We first establish that our model is well defined:

Theorem 7 (Existence and Uniqueness including mutation, recombination)
Let ν be a probability measure on $(\mathcal{P}(\mathbb{N}))^{\Omega_N}$. Then the $(L_{M,R};\nu)$-martingale problem on $C([0,\infty), (\mathcal{P}(\mathbb{N}))^{\Omega_N})$ is well-posed. For a fixed value of the parameter N the resulting canonical stochastic process is denoted $(X_t^N)_{t\ge 0}$. □

Corollary *The statement of Theorem 7 holds for general bounded selection matrices V.* □

The following observation is useful since it often allows us to restrict consideration to the finite type situation. Namely, if we form equivalence classes of types according to fitness and if $u \to M(u,\cdot)$, $(u_1,u_2) \to \alpha(u_1,u_2)$, $(u_1,u_2) \to R(u_1,u_2;\cdot)$ is constant on those equivalence classes and $R(u_1,u_2,\cdot)$ respects them, then the proportions of types within such an equivalence class follow the Fleming-Viot dynamic with migration only and their behavior is known from [DGV].

(d) Ergodic theory of systems with mutation and recombination

The next task would be to investigate the behavior of the process (X_t^N) for $t\to\infty$ for *fixed* N. Here we face still some mathematical problems not yet solved and we have to restrict ourselves here to a simplified situation, where we have only finitely many types and for some results even finitely many colonies. In that case however we shall get a complete ergodic theory of the model.

We introduce now *spatially finite* systems indexed by Ω_N^K which in addition have only K-fitness levels and hence have *finite type space.*

Definition 2.5 (K-level system)
Let $\Omega_N^K = \{(\xi_i)_{i\in\mathbb{N}} \in \Omega_N : \xi_i = 0\ \forall i > K\}$. Put in (2.4) for all $k,\ell > K$:

$$(2.5)\qquad c_k = 0, \quad \varphi_k^u = 0, \quad m_{k,\ell} = 0$$

furthermore require:

(i) $\alpha(u,v) > 0$ only if u,v are of level at most K.
(ii) $R(u,v,\cdot)$ is concentrated on types of level at most K.
Then identify the type space with $E_K = \{0,1,\ldots,M-1\}^K$ and restrict space to Ω_N^K. The resulting processes are called *K-level systems*: $(X_t^{N,K})_{t\geq 0}$. □

Remark The systems above are not just caricatures, but are important because they can be viewed as dynamics on equivalence classes of countably many types, where certain rates have been truncated, so that many types become *neutral* with respect to selection, mutation or recombination. Due to the symmetry in the resampling mechanism, the results we obtain can be reformulated for systems with *countably many types*. We do not have the space here to elaborate on that.

We can prove the following about K-level systems:

Theorem 8 (Ergodic theorem for K-level spatially finite and spatially infinite systems)
Assume that in (2.3) the selection matrix is of the form

$$V(u,v) = \chi(u) + \chi(v) \quad \textit{with}\ \chi(u) = \sum_{j=0}^{\infty} \varphi_j^u N^{-j}, \tag{2.6}$$

where χ is bounded, the mutation kernel M is as described in (2.2) with $m_{k,\ell} \geq \bar{m} > 0 \quad \forall\, k,\ell \leq K$ and $m_{k,\ell} = 0$ if $k > K$ or $\ell > K$, and the recombination is as in Definition 2.4 and Definition 2.5.

(a) *(Spatially finite K-level system)*

(i) Consider the K-level spatially finite system indexed by Ω_N^K. Then there is a unique equilibrium measure $P_{eq}^{N,K}$ on $(E_K)^{\Omega_N^K}$.

(ii) For every initial state $X_0^{N,K} \in (E_K)^{\Omega_N^K}$:

$$\mathcal{L}(X_t^{N,K}) \underset{t\to\infty}{\Longrightarrow} P_{eq}^{N,K}. \tag{2.7}$$

(iii) In the special case of $K = 0$, $r = 0$, $s = 1$ and $m_0(u,v) = \varrho(v)$, we have a Gibbs representation of the equilibrium measure:

$$\begin{aligned} P_{eq}^{N,0} &= \prod_{\Omega_N^0} P_{eq}^*, \\ P_{eq}^*(d\mu) &= \left[Z^{-1}\exp(2\langle\mu,\varphi_0\rangle)\right] \cdot \Gamma_\varrho^{m,d}(d\mu), \end{aligned} \tag{2.8}$$

where $\Gamma_\varrho^{m,d}$ is the GEM distribution on $\{0,\ldots,M\}$ with mean measure ϱ. (See (1.48) for a definition).

(b) *(K-level system with infinite space)*
If we now allow in Definition 2.5 more general c_k and assume in fact that $c_k > 0$ for all k, and in addition for fixed s,r the parameter $\bar{m}$ is sufficiently large, then the conclusion of (i) and (ii) in part (a) remains valid for the system induced on $(E_K)^{\Omega_N}$. □

Remark In the special case of two types, (b) also follows from Shiga and Uchiyama (1986, Theorem 1.2) under the hypothesis $m > 0$ alone because in this case there is a simpler dual.

Remark In the two type example Shiga and Uchiyama give an example of the infinite system with two equilibrium measures. This is in the case in which there is for the mutation process one ergodic class and one transient class but which has sufficiently large fitness.

(e) The McKean-Vlasov limit and elements of the multiple time scale analysis

The goal of the multiple space-time scale analysis is to describe in the context of the model defined in subsection 2(c), how the *interplay* of *selection and mutation* can lead to an evolving population structure described in terms of a series of *quasi-equilibria* analogous to the prescription of subsections 1(c) and 1(d). Each quasi-equilibrium state is stable in the sense that although mutations to types of lower order of fitness can occur, due to selection they are removed from the population and the fittest types represented in the quasi-equilibrium remain dominant. However on a longer time scale types which are both *fitter and more stable* but which were not initially present eventually arise by mutation and then spread in space. This results in the displacement of the quasi-equilibrium by a new quasi-equilibrium in which the newly created types with a higher order of fitness are now dominant.

Ideally our program would now be to verify this qualitative picture rigorously by carrying out the analysis of chapter 1 for the process with mutation and recombination in particular to generalize Theorem 2 and 3. However the fact that *both* selection and mutation terms occur in a competing way in the transition from one quasi-equilibria to the next, the analysis faces serious and *new mathematical problems* and for this reason our last Theorem (Theorem 10) restricts to a special situation. Therefore in this paper, we shall only present a first step of the multiple space-time scale analysis and we defer the rest of the program to a future paper.

The most basic important ingredient of the multiple time scale analysis is to consider the limiting dynamics in the slow time scale in the $N \to \infty$ limit (this is known as the *McKean-Vlasov limit*) and to find the equilibrium state of this limit equation on the component level (compare (1.41) and chapter 5 later on). This is needed, since the transition kernel of the *interaction chain*, which produces the chain of quasi-equilibria, is given in terms of that equilibrium. Therefore the McKean-Vlasov limit is the point, we now generalize to the model with mutation in Theorem 9 and 10 below. However the analyses of the longtime behavior (Theorem 10) leaves many challenging open problems for future research.

(i) The McKean-Vlasov limit In order to formulate the limit theorem for X_t^N as $N \to \infty$ (Mc Kean-Vlasov limit) we start by formulating a *nonlinear martingale problem* for a single component of the limiting process (nonlinear means the generator shall involve a functional of the law of the process).

Consider for a function F on $\mathcal{M}(\mathbb{N})$ of the form $F(z) = \prod_{k=1}^n \langle z, \Phi_k\rangle,\ \Phi_k(u) = e^{-\lambda_k u}$ for some $\lambda_k > 0$ and let π be a probability measure on $\mathcal{P}(\mathcal{P}(\mathbb{N}))$. Define the operator $L^{0,\pi}_{M,R}$ by:

$$
\begin{aligned}
(2.9)\qquad L^{0,\pi}_{M,R}(F)(z) &= c_0 \int\limits_{\mathbb{N}} \frac{\partial F(z)}{\partial z}(u)\left(\left[\int\limits_{\mathcal{P}(\mathbb{N})} z'\pi(dz')\right](du) - z(du)\right)\\
&+d_0 \int\limits_{\mathbb{N}}\int\limits_{\mathbb{N}} \frac{\partial^2 F(z)}{\partial z \partial z}(u,v) Q_z(du,dv)\\
&+s_0 \int\limits_{\mathbb{N}}\int\limits_{\mathbb{N}}\int\limits_{\mathbb{N}} \frac{\partial F(z)}{\partial x}(u)(\varphi_0^v + \varphi_0^w) z(dv) Q_z(du,dw)\\
&+m_0 \int\limits_{\mathbb{N}}\left[\int\limits_{\mathbb{N}} \frac{\partial F(z)}{\partial z}(v) M_0(u,dv) - \frac{\partial F(z)}{\partial z}(u)]\right] z(du)\\
&+r_0 \int\limits_{\mathbb{N}}\int\limits_{\mathbb{N}} \alpha(v_1,v_2)\left\{\int\limits_{\mathbb{N}} \frac{\partial F(z)}{\partial z}(u) R(v_1,v_2,du) - \frac{\partial F(z)}{\partial z}(v_1)\right\} z(dv_1) z(dv_2),
\end{aligned}
$$

where $M_0(u,dv)$ is obtained from $M(u,dv)$ by setting $m_{k,\ell} = 0$ for $k > 0$ or $\ell > 0$.

We now define the notion of solution of a nonlinear martingale problem. Let P be a measure on $C([0,\infty],\ \mathcal{P}(\mathbb{N}))$ and denote by $\Pi_t P$ the time t-marginal distribution of P.

Definition 2.6 (Limiting process)
Given an initial measure ν on $\mathcal{P}(\mathbb{N})$, the probability measure P_ν on $C([0,\infty),\mathcal{P}(\mathbb{N}))$ is a solution to the $(L^{0,\pi_t}_{M,R},\nu)$ *nonlinear martingale problem.* This means with Π_t denoting projection on the time t marginal that the following properties hold:

(2.10) $$\Pi_0 P = \nu$$
$$\left(F(z(t)) - \int_0^t L^{0,\Pi_s P_\nu}_{M,R}(F)(z(s))ds\right)_{t\geq 0} \text{ is a } P_\nu - \text{martingale, } \forall F \text{ in } C^2(\mathcal{P}(\mathbb{N})). \quad \square$$

Remark The marginal law can be characterized as a consequence of (2.10) as follows. Define $\pi_t := \Pi_t P_\nu$. The measure π_t is a weak solution of the following equation (called McKean-Vlasov equation):

(2.11) $$\langle \pi_t, F\rangle - \langle \pi_0, F\rangle = \int_0^t \langle \pi_s, (A_{\nu_s} F)\rangle ds, \quad \nu_s = \int_{\mathcal{P}(\mathbb{N})} z\pi_s(dz),$$

where A_{ν_t} is defined as the time-dependent generator, which is obtained by replacing in (2.9) the term $\int z\pi(dz)$ by ν_t at time t.

The first basic result of this section is the following generalization of Theorem 2 in the presence of mutation, but only for the lowest level, i.e. $j = k = 0$ and under the assumptions which we now specify.

Assume that different types have different fitness. Define $E = \{\chi(u)|u \in \mathbb{N}\}$ and let $\mathcal{P}(E)$ be equipped with the Vasherstein metric (to recall the definition see 5.55). We assume that either

(2.12) $$|E| < \infty,$$

or the following three regularity conditions are satisfied, namely the conditions:

(2.13) $$\sup_{X\in\mathcal{P}(\mathbb{N})} \sup_v \sum_w \alpha(v,w)X(w) < \infty$$

(2.14) M is a Lipschitz kernel from $E \to \mathcal{P}(E)$

(2.15) R is a Lipschitz kernel from $E \times E \to \mathcal{P}(E)$.

Recall that a map $H : E_1 \to E_2$ is called Lipschitz from E_1 to E_2 if

(2.16) $$||H(x) - H(y)||_{E_2} \leq L(H)||x-y||_{E_1}, \quad L(H) < \infty.$$

Now we are able to identify the McKean-Vlasov limit as follows.

Theorem 9 (McKean-Vlasov limit and propagation of chaos)
Under the above assumptions:
(a) *The $(L^{0,\pi_t}_{M,R},\Lambda)$-nonlinear martingale problem has a unique solution P_Λ for every initial law Λ on $\mathcal{P}(\mathbb{N})$.*
(b) *Let $X^*(t) = (x^*_\xi(t))_{\xi\in\Omega_\infty}$, where the $(x^*_\xi(t))_{t\geq 0}$ are independent solutions of the $(L^{0,\pi_t}_{M,R},\Lambda)$ nonlinear martingale problem defined in (2.10)-(2.11) . Furthermore consider the process $(X^N(t))_{t\geq 0}$ defined in (2.4), with initial law given by $\mathcal{L}(X^N(0)) = \bigotimes_{\Omega_N}\Lambda$.*
Then the following holds:

(2.17) $$\mathcal{L}((X^N(t))_{t\geq 0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((X^*(t))_{t\geq 0}) = \bigotimes_{\xi\in\Omega_\infty} P_\Lambda. \quad \square$$

Corollary *The statement of Theorem 9a) holds for general selection matrices V.* $\square$

(ii) Longtime behavior of the Mc-Kean-Vlasov limit. The next step is now to examine the behavior of the limit process for $N \to \infty$, which we called $X^*(t)$ as $t \to \infty$. Here we restrict to a mutation matrix of the special form in which a mutation from u to v appears at a rate depending only on u. In this case we can identify the limiting state in the form of a Gibbs measure.

Theorem 10 (Equilibrium states of McKean-Vlasov limit)
Assume that $r = 0$ (no recombination), and $m_{0,0}(u, dv) = \bar{m}_0 M_0(u, dv)$ with $M_0(u, dv) = \varrho(dv)$.
(a) *The equilibrium state of the McKean-Vlasov limit satisfy:*

(i) There is a one-to-one correspondence between equilibrium states for the solution of the McKean-Vlasov nonlinear martingale problem of (2.10) and fixed points $\nu \in \mathcal{P}(\mathbb{N})$ of the equation $\nu = G(\nu)$, where

$$G(\nu) = Z^{-1} \int \mu e^{V(\mu)} \Gamma^{\bar{m}_0 + c_0, d_0}_{(\alpha\nu + (1-\alpha)\varrho)}(d\mu) \tag{2.18}$$

with

$$V(\mu) = \int_{\mathbb{N}} \varphi_0^u \mu(du), \qquad \alpha = \frac{c_0}{c_0 + \bar{m}}. \tag{2.19}$$

(ii) Given a fixed point ν_* of G, the corresponding invariant measure of the process defined in (2.10) has the "Gibbs" representation

$$P^{\nu_*}_{eq}(d\mu) = Z^{-1} \left[\exp(2 \int_{\mathbb{N}} \varphi_0^u \mu(du)) \right] \Gamma^{\bar{m}_0 + c_0, d_0}_{(\alpha\nu_* + (1-\alpha)\varrho)}(d\mu) \tag{2.20}$$

and the process $(X^*_t)_{t \geq 0}$ of (2.17) has an equilibrium state $\bigotimes_{\Omega_\infty} P^{\nu_*}_{eq}$.

(b) *For sufficiently large mutation rate $\bar{m}$, there is a unique fixed point ν_*.*

(c) *Assume that $\bar{m}$ is sufficiently large. For every initial measure ν one has*

$$\Pi_t P_\nu \underset{t \to \infty}{\Longrightarrow} P^{\nu_*}_{eq}. \qquad \square \tag{2.21}$$

Conjecture If $\bar{m} > 0$ then the fixed point of the nonlinear map is unique. A proof of this result can be based on the genealogical process associated with the model by showing a "$(\bar{m}, \varrho)$-mutation" occurs on each line of descent. Similarly (2.21) should hold just under $\bar{m} > 0$.

The result of Theorem 10 allows us, at least in the situation of state independent mutation to define the ingredients of the multiple space-time scale analysis as we did in chapter 1 (c) in the case without mutation. We defer the generalization to the more complicated mutation structure needed in evolutionary models for the discussion of the multiple space-time analysis including selection *and* mutation to a future paper.

Part B
Proofs

3 Dual process, Existence and Uniqueness (Proof Theorems 0, 7)

The purpose of this section is to prove the Theorems 0 and 7 and the Corollary to Theorem 7. The proof proceeds in three steps. We show *existence* of the process and we construct a *dual process*, better a Feynman-Kac function valued dual driven by a particle system, which is used to get *uniqueness* of the solution to the martingale problem. This dual will be also of use in the proofs of the Theorems of chapter 2.

(a) The dual process

Step 1. (Dual process)
In this step we will construct a process which will serve as the dual process to $(X_t^N)_{t\geq 0}$ as defined in (2.4) and which will be used in the next subsection to prove that the martingale problem has a unique solution. For a duality relation we need two things, a class of functions determining laws on the state space of our process $(X_t)_{t\geq 0}$ and a dynamic of a dual process. We now introduce these two objects.

The key observation is that in order to determine the law of the system $X(t) = (x_\xi(t))_{\xi\in\Omega_N}$ it suffices to calculate for every initial point $X_0 \in (\mathcal{P}(\mathbb{N}))^{\Omega_N}$ and for every $n \in \mathbb{N}$, $\hat{\xi} = (\xi_1, \ldots, \xi_n) \in (\Omega_N)^n$, $f \in C_0((\mathbb{N})^n, \mathbb{R})$ the function

$$(3.1) \qquad F_t((\hat{\xi}, f), X_0) := E_{X_0}\left(\int_{\mathbb{N}} \cdots \int_{\mathbb{N}} f(u_1, \ldots, u_n) x_{\xi_1}(t)(du_1) \ldots, x_{\xi_n}(t)(du_n)\right).$$

We will calculate the r.h.s. by defining a new evolution semigroup of a *bivariate process* $(\hat{\eta}_t, \mathcal{F}_t)$, one component namely $\widehat{\eta}_t$ is a *particle system* and the other component $\mathcal{F}_t$ has values in the function space

$$(3.2) \qquad \bigcup_{m=1}^{\infty} L_\infty((\mathbb{N})^m)$$

(the functions with finitely many variables running in $\mathbb{N}$ which are bounded). This bivariate process, the so-called *function-valued dual process*, will allow to write F_t as expectation over the evolution of $(\hat{\eta}_t, \mathcal{F}_t)$.

The major building block of this dual evolution is a *particle system* with *migration, coalescence* and *birth* of new particles. In this particle system we keep track of the history of the individual particles (at least to some extent). A transition in this particle system will induce a transition of the function valued process. Finally we will express $F_t((\hat{\xi}, f), X_0)$ as an expectation over an X_0-dependent functional of the function-valued process starting in f based on a particle system starting with particles on $\hat{\xi}$.

The point in introducing the new process is that in the function-valued process in a finite time interval only a finite number of transitions occur, which allows to calculate expectations, while in $X(t)$ changes occur in countably many components in any finite time interval.

We begin defining the *particle system* driving the evolution of the dual process. The particle system $(\hat{\eta}_t)_{t\geq 0}$ will have the form

$$(3.3) \qquad \hat{\eta}_t = (\eta_t^1, \ldots, \eta_t^{|\pi_t|};\ \pi_t),$$

where $\eta_t^1, \eta_t^2, \ldots$ are elements of Ω_N and π_t is an partition (ordered tuple of subsets) of the basic set of the form $\{1, 2, \ldots, n\}$ where n is going to vary as well as a random process. We order partition elements by its smallest element and then label them by 1,2,3 in increasing order. Every partition element has a location in Ω_N, namely the i-th partition element η_t^i. (The interpretation is that we have n particles which are grouped in $|\pi_t|$-subsets and each of these subsets has a position in Ω_N).

The evolution of $(\hat{\eta}_t)$ is defined as follows (recall (2.4) and (1.1) - (1.7) for the meaning of the parameters):

- every partition element performs independent of the other partition elements a continuous time *random walk* on Ω_N with transition rate $a(\cdot,\cdot)$ (see (1.6)).
- all pairs of partition elements which occupy the same site in Ω_N *coalesce* into one partition element at rate d
- at rate $(s+r)$ every partition element creates a new element at its location (*birth*) which forms its own partition element, now of a set increased by one element.

Note that this particle system is well-defined for all times since the total number of partition elements is dominated by an ordinary linear birth process. We will need the total number of particles which we denote by N_t (as opposed to the total number of partition elements), i.e.

(3.4) $$N_t = |\pi_t^1| + \ldots + |\pi_t^{|\pi_t|}|, \text{ with } \pi_t^i \text{ being the i-th partition element at time } t.$$

We want to define a bivariate process $(\hat{\eta}_t, \mathcal{F}_t)$ and we have already described the evolution mechanism of the first component. Now we are able to define for a given path of the *particle process* $(\hat{\eta}_t)_{t\geq 0}$ the *function valued process* $(\mathcal{F}_t)_{t\geq 0}$ with values in $\bigcup_1^\infty L_\infty((\mathbb{N})^m)$. The evolution of $(\mathcal{F}_t)_{t\geq 0}$ has two components a deterministic component driven by a semigroup $(\tilde{M}_t)_{t\geq 0}$ derived below from the semigroup M_t of the (pure) mutation process and a stochastic component driven by $(\hat{\eta}_t)_{t\geq 0}$.

Let M_t denote the semigroup on $L_\infty(\mathcal{P}(\mathbb{N}))$ induced by the mutation process, i.e. the process we obtain from (2.4) by letting the index set for the components consist of one point and putting all other coefficients d, s, r, c in (2.4) equal to 0.

The evolution starts with a function f of $|\pi_0|$-variables and the evolves according to the rules:

(i) If no transition occurs in $(\hat{\eta}_t)_{t\geq 0}$ then $\mathcal{F}_t$ follows the evolution given by $\bigotimes_{i=1}^n M_t$, if $\mathcal{F}_t$ after the last transition in $(\hat{\eta}_s)$ became an element of $L_\infty((\mathbb{N})^n)$.

(ii) If a birth occurs in the process $(\hat{\eta}_s)$ due to the partition to which the element i of the basic set belongs, then with probabilities $s/(s+r)$ respectively $r/(s+r)$ the following transitions occur from an element in $L_\infty((\mathbb{N})^m)$ to elements in $L_\infty((\mathbb{N})^{m+1})$:

(3.5) $$f(u_1, \ldots, u_m) \longrightarrow \varphi(u_i) f(u_1, \ldots, u_m) - \varphi(u_{m+1}) f(u_1, \ldots, u_m)$$

(3.6) $$f(u_1, \ldots, u_m) \longrightarrow \int_{\mathbb{N}} f(u_1, \ldots, u_{i-1}, y, u_{i+1}, \ldots, u_m)\, \alpha(u_i, u_{m+1}) \cdot R(u_i, u_{m+1}; dy).$$

(iii) If a coalescence of two partition elements occurs, then the corresponding variables of $\mathcal{F}_t$ are set equal, so that changes from an element of $L_\infty(\mathbb{N}^m)$ to one of $L_\infty(\mathbb{N}^{m-1})$.

In this fashion, $\mathcal{F}_t$, starting with an element $f \in L_\infty(\mathbb{N}^m)$ for some m, we have defined uniquely a (bivariate) process $(\hat{\eta}_t, \mathcal{F}_t)$ (where the randomness sits essentially in $(\hat{\eta}_t)_{t\geq 0}$).

Extension In order to incorporate the general, not necessarily additive, selection which we need later we replace the transition in $\hat{\eta}_t$ as follows, at rate s instead of one particle *two* new particles are born and (3.5) is replaced by

$$f(u_1,\dots,u_m) \to (V(u_i,u_{m+1}) - V(u_i,u_{m+2}))f(u_1,\dots,u_m). \tag{3.7}$$

Step 2 *(Feynman-Kac duality relation)*
In this step we use the terminology introduced in step 1. Duality relations are used frequently in the theory of interacting systems. Here we use a more general version, a so-called *Feynman-Kac duality relation* which involves also an exponential expression due to the fact that besides the migration the evolution involves other drift terms.

Proposition 3.1 (Duality relation)
Let $(X_t)_{t\geq 0}$ be a solution of the $(L_{M,R}, X)$-martingale problem with $X = (x_\xi)_{\xi\in\Omega_N}$. Choose $\hat{\xi} \in (\Omega_N)^n$ and $f \in L_\infty((\mathbb{N})^n)$ for some $n \in \mathbb{N}$. Let $\pi(i)$ denote the label of the partition element to which the i-th of the n particles $\{1,2,\dots,n\}$ belongs. (Recall (3.2), (3.4)).

Then $(\widehat{\eta}, \mathcal{F}_t)$ is the Feynman-Kac dual of (X_t) i.e.:

$$F_t((\hat{\xi}, f), X) = E\Bigg\{\Bigg[\exp(s\int_0^t |\pi_u| du)\Bigg] \cdot \tag{3.8}$$

$$\cdot \Bigg[\int_{\mathbb{N}} \dots \int_{\mathbb{N}} \mathcal{F}_t(u_{\pi(1)},\dots,u_{\pi(N_t)}) x_{\eta_t^1}(du_1)\dots x_{\eta_t^{|\pi_t|}}(du_{|\pi_t|})\Bigg]\Bigg\}$$

with

$$\begin{aligned} \hat{\eta}_0 &= [\xi_1,\xi_2,\dots,\xi_n;(\{1\},\{2\},\dots,\{n\})] \\ \mathcal{F}_0 &= f. \quad \Box \end{aligned} \tag{3.9}$$

Proof The case of migration and countably many colonies but without selection, mutation and recombination is given in [DGV], section 3 a). We have therefore to incorporate the additional mechanisms of mutation, selection and recombination. Here we can use [D], chapter 5.6 and chapter 10.2, where the duality relation is verified for the one colony case provided the r.h.s. of the duality relation is defined (the exponential expression has to be finite). Assuming the finiteness of the r.h.s. of (3.8), combining these two techniques (from [DGV] and [D]) is straightforward.

However there is a technical issue due to the Feynman Kac term (i.e. the exponential) appearing in addition to the terms which can be interpreted via the generator of the dual process. We must check the uniform integrability condition of [D], Cor. 5.5.3. in order to guarantee that the r.h.s. of (3.8) is well-defined.

First note that the term involving the integral is bounded so that we can restrict our attention to the integrability of the exponential term - note that this does not depend on the initial function except through $|\pi_0|$. But we next observe that the particle system $\widehat{\eta}_t$ as described below (3.3) is dominated (by excluding the quadratic death terms) by a linear birth system. Expectation with respect to such a system is denoted by $\tilde{E}$ and by $\tilde{E}^1$ the expectation with respect to a particle system starting with only one element. But for such a branching system

$$\tilde{E}[\exp(s\int_0^t |\pi_u| du)] = \Bigg[\tilde{E}^1[\exp(s\int_0^t |\pi_u| du)]\Bigg]^{|\pi_0|} \tag{3.10}$$

by the branching property.

To complete the proof we must show that there exists $t_0 > 0$ such that

$$(3.11) \qquad \tilde{E}^1[\exp(s\int_0^t |\pi_u|du)] < \infty \qquad \forall\, 0 < t < t_0.$$

To do this we note that

$$(3.12) \qquad \tilde{E}^1[\exp(s\int_0^t |\pi_u|du)] = \lambda(t),$$

where the function λ satisfies the differential equation

$$(3.13) \qquad \begin{aligned} \frac{d\lambda}{dt} &= [(s+r)\lambda^2 - (s+r)\lambda] + s\lambda \\ \lambda(0) &= 1. \end{aligned}$$

The relations (3.12) and (3.13) can be verified as follows. Replace the integral $\int_0^t |\pi_u|du$ by the Riemann-Stieltjes sum $\sum_i |\pi_{t_i}|(t_{i+1} - t_i)$. Apply repeatedly the branching property, which reads on the level of Laplace transforms:

$$(3.14) \qquad \frac{d}{dt}u = (s+r)u^2 - (s+r)u, \quad u(t) := \tilde{E}^1([\exp(\theta|\pi_t|)])$$

and then taking limit $\max(t_{i+1} - t_i) \to 0$ gives the desired differential equation.

But since the differential equation (3.13) has a local solution (3.11) holds.

(b) Proof of Theorems 0 and 7

Step 1. (Existence)

In the first step we give the proof that the martingale problem has a solution. The proof consists of two parts. First one shows there is a solution of the martingale problem if the index set if finite and from here we construct the solution for countably many components. We can use existing results to carry out this step of the program, therefore not many details will be given.

First define an increasing sequence of finite subsets I_m of Ω_N

$$(3.15) \qquad I_m = \{\xi \in \Omega_N | d(\tilde{\xi}, \xi) \le m\}, \quad \tilde{\xi} = (0, 0, \ldots)$$

and consider the system which is obtained by freezing the dynamics off I_m. Then we obtain essentially a finite system, which can be viewed as diffusion limit of a particle system. That is, the solution is obtained by taking a weak limit of discrete Wright-Fisher models as in [EK2].

Existence for the infinite system is then obtained by verifying that the collection of these modified systems are tight as $I_m \uparrow \Omega_N$ and that the weak limits satisfy the martingale problem. This part follows the arguments in [DGV] page 2300 (proof of Proposition 1.0).

Step 2 (Uniqueness)

We have now proved that the r.h.s. of (3.8) is well-defined and hence we know the evolution of $F_t((\hat{\xi}, f), X)$ for every $\hat{\xi} \in (\Omega_N)^n$, $f \in C_0(\mathbb{N}^n)$ and $n \in \mathbb{N}$. Since the law of our process is determined uniquely by these quantities, this completes the proof of the uniqueness of the solution to the martingale problem.

4 Ergodic theorem (Proof of Theorem 1)

This section studies the longtime behavior of $X(t)$ and gives the proof of Theorem 1 based on a Lyapunov function and on coupling.

We begin with part a) of Theorem 1. Under the condition $\text{supp}(\Theta) = \mathcal{N}^\Theta$ we know that every type u which appears in the system (i.e. $\text{Prob}(x_\xi(t)(u) > 0$ for some $\xi, t) > 0)$ has maximal fitness among all appearing types. Since in this situation $\chi(u) - \int \chi(v) x_\xi(dv) = 0$ the selection term in the generator (by (1.15)) disappears, we have simply a system of Fleming-Viot processes, which interact by migration. This allows us to apply [DGV] Theorem 0.1, which yields the assertions of part a).

Next we come to the parts b) and c):
The key idea here is to compare the system with a system without selection defined on the set of types of maximal fitness. Precisely we have to distinguish two cases, if this set of maximal fitness is empty we want to show that in probability

$$x_\xi(t)(u) \underset{t\to\infty}{\longrightarrow} 0 \quad \forall u \in \mathbb{N} \tag{4.1}$$

and if the set is not empty we want to find a system $\{z_\xi(t)\}_{\xi\in\Omega_N}$ without selection (i.e. $\chi \equiv \text{const}$ (compare below (1.7)) concentrated on types of maximal fitness, which satisfies

$$\mathcal{L}((x_\xi(t))_{\xi\in\Omega_N}) - \mathcal{L}((z_\xi(t))_{\xi\in\Omega_N}) \underset{t\to\infty}{\Longrightarrow} 0. \tag{4.2}$$

Essential for this purpose are the following two differential equations, which are obtained from (1.15) by explicit calculation using the defining martingale problem (for details see (5.37)):

If the system $X(t)$ has initially a translation invariant distribution then:

$$\frac{d}{dt} E(\langle x_\xi(t), \chi\rangle) = E\ (\text{Var}_{x_\xi(t)}(\chi)) \geq 0, \tag{4.3}$$

which says that the mean-fitness in a colony has nondecreasing expectation in fact is strictly increasing as long as χ is not constant almost surely under the current distribution.

Define $I_\varepsilon = \{u|\chi(u) \geq \sup_v \chi(v) - \varepsilon\}$. Then for all $\varepsilon > 0$, (for the proof see (5.37)):

$$\begin{cases} \frac{d}{dt} E(\sum\limits_{u\in I_\varepsilon} x_\xi(t)(u)) & > & 0 & \text{for} \;\; x_\xi(t)(\mathcal{C}I_\varepsilon) > 0 \\ \frac{d}{dt} E(\sum\limits_{u\in I_\varepsilon} x_\xi(t)(u)) & \geq & 0 & \text{for} \;\; x_\xi(t)(\mathcal{C}I_\varepsilon) = 0, \end{cases} \tag{4.4}$$

which says that the expected mass of types of ε-maximal fitness respectively of maximal fitness is nondecreasing and is even strictly increasing if there are still types at ξ at time t which are not ε-maximal, respectively maximal.

Both relations (in 4.4) together prove immediately that if $\Theta(\{u\}) > 0 \quad \forall u \in \mathbb{N}$ then for every fixed u:

$$E x_\xi(t)(u) \underset{t\to\infty}{\longrightarrow} 0 \quad \text{if } \chi(u) < \sup(\chi(v)|v : \Theta(v) > 0). \tag{4.5}$$

This proves relation (1.26).

Next we come to (1.25), that is u does not satisfy the r.h.s. of (4.5). Begin by noting that if u satisfies $\chi(u) = \sup(\chi(v)|v : \Theta(v) > 0)$, then $E(x_\xi(t)(u))$ is nondecreasing in t and hence we can define in order to prove (1.25):

$$\Theta_\infty(\mu)(u) = \lim_{t\to\infty} (E(x_\xi(t)(u))). \tag{4.6}$$

Since it is well-known (see [DGV], Theorem 0.1) how a system with $\chi \equiv \text{const}$ behaves we can proceed by comparing our system with a system with $\chi \equiv 0$ which starts in a translation invariant ergodic initial state with mean measure $\Theta_\infty(\mu)$. Since we can control the total mass of types which is concentrated on those u which do not have maximal fitness we need to construct a *coupling* of the two processes on the set of maximal types to show that indeed the relation (1.25) holds.

This coupling we now have to construct. A coupling is a bivariate process $(X^1(t), X^2(t))$, where $X^1(t)$ and $X^2(t)$ are versions of the two processes we want to compare but now they are defined on a *common* probability space. So in addition to the process $X(t) = (x_\xi(t))_{\xi\in\Omega_N}$ we consider the process which starts in the equilibrium $\nu_{\Theta_\infty(\mu)}$ and we shall denote this process by $(z_\xi(t))_{\xi\in\Omega_N}$. We want to compare these processes. We need a notion of distance for which we can calculate its expectation as a function of time. Define the distance $||x_1 - x_2||$ between two measures $x_1, x_2 \in \mathcal{P}(\mathbb{N})$ by taking an injective function $\Psi : \mathbb{N} \to [0,1]$ and defining

$$(4.7) \qquad ||x_1 - x_2|| \quad = \quad \inf\Big(\int |u_1 - u_2| y(du_1, du_2) | y \in \mathcal{P}([0,1]^2),\ \pi_1 y = \Psi(x_1),\ \pi_2 y = \Psi(x_2)\Big)$$

where π_i denotes projections on first and second component.

The proof is completed by showing that for any such Ψ, which we choose to define $||\cdot||$, the mean distance $E||x_\xi(t) - z_\xi(t)||$ between the two components of the coupled processes can be bounded by a multiple of $||Ex_\xi(t) - \Theta_\infty(\mu)||_{\mathrm{Var}(\mathcal{N}^\Theta)}$, which goes by construction of $\Theta_\infty(\mu)$ to zero as $t \to \infty$. The construction of a coupling will occur in a more subtle context in section 5, so that we refer the reader to section 5 (f) step 2 for details.

It remains to identify $\Theta_\infty(\mu)$ under the given assumptions on the initial state μ. The point is that since the initial state is φ-exchangeable we know that the relative proportions of the types of maximal fitness are preserved. This completes the proof of (1.25).

For part d) of Theorem 1 we use again the method of coupling of the two processes, but first we identify all types u with $\chi(u) \geq \sup_v \chi(v) - \varepsilon$ and then we are back in the situation discussed in part b) of our Theorem 1 and we can conclude that the process concentrates on the types of ε-maximal fitness as time tends to infinity. Now we let $\varepsilon \to 0$ and get the result.

5 Mean-field finite system scheme

The proof of the Theorems 2 and 3 requires as a main tool the analysis of the large *finite* systems in *large time scales*. We collect the needed results in a systematic fashion in this chapter 5 before we then exploit these results in chapter 6 to actually prove Theorems 2, 3. For an outline of chapter 5 see the end of the subsection 5(a).

(a) A mean-field model

In this chapter we discuss the longtime-large scale behavior of a much simpler process, which is nevertheless an essential building block for the multiple space-time scale analysis. The main simplification will be that this auxiliary system will be essentially (i.e. if we forget the frozen components) a *finite system* but we will later analyze it in *large time scales* adapted to the system size. The latter is the main idea of the socalled *finite system scheme* [DG2], [CGS]. The main result of chapter 5 is in (5.20) and (5.21) below.

Consider the following mean-field model with N *active* components:

$$(5.1) \qquad Y^N(t) = (y_i^N(t))_{i\in\mathbb{N}} \in (\mathcal{P}(\mathbb{N}))^{\mathbb{N}}$$

where $Y^N(t)$ is the unique solution of the (G_N, π_Θ)-martingale problem which we define now. The initial law π_Θ is a homogeneous product measure on $(\mathcal{P}(\mathbb{N}))^{\mathbb{N}}$ with

$$(5.2) \qquad E_{\pi_\Theta}(y_i) = \Theta \in \mathcal{P}(\mathbb{N}),$$

and G_N is the operator acting on twice differentiable functions F on $(\mathcal{P}(\mathbb{N}))^{\mathbb{N}}$ as follows: $(z = (z_i)_{i\in\mathbb{N}} \in \mathcal{P}(\mathbb{N})^{\mathbb{N}}$ and all integrals extend over $\mathbb{N}$.)

$$(5.3) \qquad G_N(F)_{(z)} = \sum_{i=1}^{N} \Big[c \int \frac{\partial F(z)}{\partial z_i}(u)\Big(\frac{1}{N}\sum_{j=1}^{N} z_j(du) - z_i(du)\Big) + s \int\int\int \frac{\partial F(z)}{\partial z_i}(u)(\varphi(v)+\varphi(w)) z_i(dv) Q_{z_i}(du,dw) + sN^{-1} \int\int\int \frac{\partial F(z)}{\partial z_i}(u)(\chi(v)+\chi(w)) z_i(dv) Q_{z_i}(du,dw) + d \int\int \frac{\partial^2 F(z)}{\partial z_i \partial z_i}(u,v) Q_{z_i}(du,dv)\Big].$$

Here c, d, s are nonnegative real numbers and φ and χ are bounded functions on $\mathbb{N}$, which attain their maxima.

The *existence and uniqueness* of a solution to the (G_N, π_Θ)-martingale problem is established by noting that it can be viewed as a special case of the general model introduced in (1.13).

Outline We briefly outline what happens in the rest of the chapter 5. In order to keep the analysis of the longtime behavior of this system $Y^N(t)$ transparent we first state in 5 (b) – (c) the main results and then later present the details of the proofs in the subsequent subsections 5 (d) – (f). Before we analyze this system $Y^N(t)$ in large time scales we discuss first in 5 (b) the behavior as $N \to \infty$ for fixed times t (McKean-Vlasov limit), since this will lead to a number of objects needed for the analysis in large time scales. In the subsequent subsection 5 (c) we shall formulate properties of the large time scale behavior and give in Corollary 5.1 equation (5.21) the main result of this chapter. The proofs of these results will be in subsections 5 (d) - 5 (f) respectively.

(b) The McKean-Vlasov limit-process and its longtime behavior

This section studies $(Y^N(t))_{t\geq 0}$ as $N \to \infty$ and subsequently the limit $t \to \infty$ in the limiting object; the proofs will be in section 5 (d). In order to study $Y^N(t)$ for large N, we shall need to define first the candidate for the limiting object $Y(t) = (y_i(t))_{i\in\mathbb{N}} \in (\mathcal{P}(\mathbb{N}))^{\mathbb{N}}$.

This $Y(t)$ is a process on $C([0,\infty), (\mathcal{P}(\mathbb{N}))^{\mathbb{N}})$ with exchangeable marginal distributions given by the (G_t, π) martingale problem, where the initial law π is an exchangeable law on $(\mathcal{P}(\mathbb{N}))^{\mathbb{N}}$ and for a twice differentiable function F on $(\mathcal{P}(\mathbb{N}))^{\mathbb{N}}$ which depends only on finitely many components the action of G_t at time t is given by $(z = (z_i)_{i\in\mathbb{N}} \in (\mathcal{P}(\mathbb{N}))^{\mathbb{N}})$:

$$(5.4) \qquad G_t(F)_{(z)} = \sum_{i=1}^{\infty} \Big[c \int \frac{\partial F(z)}{\partial z_i}(u)(\Theta_t(du) - z_i(du)) + s \int\int\int \frac{\partial F(z)}{\partial z_i}(u)(\varphi(v)+\varphi(w)) z_i(dv) Q_{z_i}(du,dw) + d \int\int \frac{\partial^2 F(z)}{\partial z_i \partial z_i}(u,v) Q_{z_i}(du,dv),$$

where

$$(5.5) \qquad \Theta_t = \Theta_t(z) = \lim_{N\to\infty} \left[\frac{1}{N} \sum_{i=1}^{N} z_i(t) \right].$$

Since in the martingale problem formulation we need to consider only the action of G_t on a set of measure one (with respect to an exchangeable probability law) of sequences $(z = (z_i)_{i\in\mathbb{N}} \in (\mathcal{P}(\mathbb{N}))^{\mathbb{N}})$ the limit (5.5) exists and the resulting martingale problem makes sense. The system is then governed by the global signal $\Theta(\mathcal{L}(Y_t))$ and has no local interaction between components.

We now state the three important facts, we shall need throughout. The first point is to make sure $Y(t)$ is well-defined.

Lemma 5.0 *The (G_t, π)-martingale problem is well-posed.* □

The system $Y(t)$ is the McKean-Vlasov limit of $Y^N(t)$ (defined in (5.1)) as $N \to \infty$:

Lemma 5.1 *Choose $\pi = \pi_\Theta$ (see (5.1 - 5.2)), then*

$$\mathcal{L}((Y^N(t))_{t\geq 0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Y(t))_{t\geq 0}). \qquad \square \tag{5.6}$$

The behavior of the process $Y(t)$ as $t \to \infty$ is as follows:

Lemma 5.2 *For every initial measure μ, where μ is a homogeneous product measure, there exists a unique probability measure τ on $\mathbb{N}$ such that:*

$$\mathcal{L}(Y(t)) \underset{t\to\infty}{\Longrightarrow} \Lambda^{c,d}_\tau, \tag{5.7}$$

with

$$\Lambda^{c,d}_\Theta = \bigotimes_{i=1}^{\infty} \Gamma^{c,d}_\Theta, \tag{5.8}$$

where $\Gamma^{c,d}_\Theta$ is the unique equilibrium measure of the $\mathcal{P}(\mathbb{N})$-valued process defined in (1.35) with $c_k = c$ and $d_k = d$, which is the Fleming-Viot process on $\mathbb{N}$, with immigration-emigration rate c, immigration source Θ, resampling rate d and without selection. □

We conclude this subsection with a property needed later on. The equilibrium measures $\Gamma^{c,d}_\Theta$ have the following property which will be crucial analyzing in chapter 6 the whole hierarchy of models of the above type:

Lemma 5.3

$$\int_{\mathcal{P}(\mathbb{N})} \Gamma^{c,d}_\Theta(d\varrho) Q_\varrho(du, dv) = \frac{c}{c+d} Q_\Theta(du, dv) \tag{5.9}$$

Put $b = c/d$. Then

$$\begin{aligned}\int_{\mathcal{P}(\mathbb{N})} \Gamma^{c,d}_\Theta(dz)\varrho(du)Q_\varrho(dv,dw) \quad = \quad & \frac{b}{1+b}(\Theta(du)\Theta(dw)\delta_u(dv) \\ & + \frac{1}{1+b}\Theta(du)\delta_u(dv)\delta_u(dw) \\ & - \left\{ \frac{b^2}{(2+b)(1+b)}\Theta(du)\Theta(dv)\Theta(dw) \right. \\ & + \frac{2}{(2+b)(1+b)}\Theta(du)\delta_u(dv)\delta_u(dw) \\ & \left. + \frac{3b}{(2+b)(1+b)}\Theta(du)\Theta(dv)\delta_v(dw) \right\}. \qquad \square\end{aligned} \tag{5.10}$$

This lemma shows that (recall (1.15)) with *additive selection* in Y_t we obtain by averaging the selection term over the equilibrium law again an expression of the same structure (as the second term in (5.3)) and this is *not* true for a general selection matrix V. We shall see that this means that the renormalization analysis works only for additive selection in the simple form described.

(c) Large time scale behavior of mean-field model

The main point of this chapter is to analyze $Y^N(t)$ in *large time scales* rather than passing as in part (b) to the McKean-Vlasov limit for fixed time t, that is we look at the rescaled process

$$(5.11) \qquad (Y^N(tN))_{t\geq 0}.$$

The spatial average over the N-active components, $\bar{Y}^N(t)$, defined by

$$(5.12) \qquad \bar{Y}^N(t) = \frac{1}{N}\sum_{i=1}^N y_i(t).$$

is important for this new object as $N \to \infty$ We will denote the *time*, respectively, *time-space rescaled processes* by

$$(5.13) \qquad \tilde{Y}^N(t) = Y^N(tN), \quad \tilde{\bar{Y}}^N(t) = \bar{Y}^N(tN).$$

The idea now is to show the convergence of $(\tilde{Y}^N(t))_{t\geq 0}$ by proving that $\tilde{\bar{Y}}^N(t)$ for $N \to \infty$ converges to a limiting process and that as a consequence $\mathcal{L}(\tilde{Y}^N(t)|\tilde{\bar{Y}}^N(t) = \Theta)$ converges as $N \to \infty$ to $\Lambda^{c,d}_\Theta$. The reason being that $\bar{Y}^N(t)$ varies slowly compared to the components, so that in the time scale where $\bar{Y}^N$ fluctuates the components have relaxed into an equilibrium state of the system where the center of drift is fixed and equal to the value of $\bar{Y}^N$ and furthermore no selection occurs, since over time spans $o(N)$ no φ-selection takes place (recall (5.3)) because the inferior types with respect to φ have already disappeared at time tN while, on the other hand, the χ-selection needs time spans of order N to work.

We want to find the limiting dynamics of $(\tilde{\bar{Y}}^N(t))_{t\geq 0}$. Note first that the process $\bar{Y}^N(t)$ is a functional of the system $Y^N(t)$ and solves (not necessarily uniquely) the $(\bar{G}_N, \pi_\Theta)$-martingale problem on $\mathcal{P}(\mathbb{N})$ on the probability space generated by the component process. For a function F on $\mathcal{P}(\mathbb{N})$ (in $\mathcal{A}$) consider the function $\tilde{F} : (\mathcal{P}(\mathbb{N}))^{\mathbb{N}} \to \mathbb{R}$ given by

$$(5.14) \qquad \tilde{F}(z) = F\left(N^{-1}\sum_{i=1}^N z_i\right)$$

and express the action of G_N on $\tilde{F}$ in terms of F to get (here $z = (z_i)_{i\in\Omega_N}$, $\bar{z} = N^{-1}\sum_1^N z_i$ and integrals extend over $\mathbb{N}$) and define:

$$\begin{aligned}(5.15) \qquad \bar{G}_N(F)_{(\bar{z})} \quad &= \quad G_N\tilde{F}(z) = \sum_{i=1}^N \Big[\frac{s}{N^2}\int\int\int \frac{\partial F(\bar{z})}{\partial \bar{z}}(u)(\chi(v)+\chi(w))z_i(dv)Q_{z_i}(du,dw) \\ &+ \frac{s}{N}\int\int\int \frac{\partial F(\bar{z})}{\partial \bar{z}}(u)(\varphi(v)+\varphi(w))z_i(dv)Q_{z_i}(du,dw) \\ &+ \frac{d}{N}\int\int \frac{\partial^2 F(\bar{z})}{\partial \bar{z}\partial \bar{z}}(u,v)Q_{z_i}(du,dv)\Big].\end{aligned}$$

Define $\bar{G}_\infty$, which is in view of (5.15) with (5.9) (use the additivity of the selection, see (1.15)) the natural candidate for the limit of $\bar{G}_N$, by:

$$\begin{aligned}(5.16) \qquad \bar{G}_\infty(F)_{(\bar{z})} \quad &= \quad (\frac{cs}{c+d})\int\int\int \frac{\partial F(\bar{z})}{\partial \bar{z}}(u)(\chi(u)+\chi(w))\bar{z}(dv)Q_{\bar{z}}(du,dw) \\ &+ (\frac{cd}{c+d})\int\int \frac{\partial^2 F(\bar{z})}{\partial \bar{z}\partial \bar{z}}(u,v)Q_{\bar{z}}(du,dv).\end{aligned}$$

It is well-known that the $(\bar{G}_\infty, \delta_{\bar{z}})$-martingale problem is well-posed for $\bar{z} \in \mathcal{P}(\mathbb{N})$. Namely the process generated by $\bar{G}_\infty$ is the following Fleming-Viot process with selection on $\mathbb{N}$: The resampling rate is $cd/(c+d)$ the selection rate $sc/(c+d)$ and the selection matrix is given by $V(u,v) = (\chi(u) + \chi(v))$. (See [D], chapter 2.7 and 5).

The key arguments in analyzing the longtime behavior are now the following two Propositions 5.1 and 5.2 which determine the behavior of $\tilde{Y}^N(t)$ (recall 5.13) as $N \to \infty$ and combine to the main result (5.21) of chapter 5, which is given in Corollary 5.1 below. The first key point is the following implication (recall (5.13) for $\tilde{Y}$, $\tilde{\tilde{Y}}$):

Proposition 5.1 (Local equilibria)
Suppose that $(N_k)_{k \in \mathbb{N}}$ is a sequence with $N_k \uparrow \infty$ as $k \to \infty$ and such that on $C([0,\infty), \mathcal{P}(\mathbb{N}))$:

$$\mathcal{L}((\tilde{\tilde{Y}}^{N_k}(t))_{t>0}) \underset{k\to\infty}{\Longrightarrow} \mathcal{L}((\bar{Y}^*(t))_{t>0}). \tag{5.17}$$

We define (recall Θ is the initial mean frequency of types and τ is defined in Lemma 5.2, via the relation 5.7):

$$Q_t^*(\tau, \cdot) = \mathcal{L}(\bar{Y}^*(t)). \tag{5.18}$$

Then the following holds for every $t \in [0,\infty)$, d and $\Lambda_\varrho^{c,d}$ as in (5.8):

$$\mathcal{L}(\tilde{Y}^{N_k}(t)) \underset{k\to\infty}{\Longrightarrow} \int Q_t^*(\tau, d\varrho)\Lambda_\varrho^{c,d}. \qquad \square \tag{5.19}$$

To complete the analyses of the behavior of $\tilde{Y}_t^N$ as $N \to \infty$, according to (5.19) we need information on the average $\tilde{\tilde{Y}}^N(t)$ and two facts are crucial:

Lemma 5.4 *The sequence $\mathcal{L}((\tilde{\tilde{Y}}^N(t))_{t>0})$ is tight.* □

Lemma 5.5 *A weak limit point of $\mathcal{L}((\tilde{\tilde{Y}}^N(t))_{t>0})$ as $N \to \infty$ satisfies the well-posed $(\bar{G}_\infty, \tau)$-martingale problem (recall (5.16) for $\bar{G}_\infty$) where τ is defined in (5.7).* □

Now combining Lemma 5.4 and 5.5. we obtain immediately (recall (5.12), (5.13)).

Proposition 5.2 (Convergence of the mean process)
On $C([0,\infty), \mathcal{P}(\mathbb{N}))$:

$$\mathcal{L}((\tilde{\tilde{Y}}^N(t))_{t>0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((\bar{Y}(t))_{t>0}), \tag{5.20}$$

where $\bar{Y}(t)$ is the Fleming-Viot process on $\mathbb{N}$ with selection generated by $\bar{G}_\infty$ and with $\bar{Y}(0) = \tau$. □

As a Corollary of the above two Propositions 5.1 and 5.2 we obtain the main result of chapter 5:

Corollary 5.1 (Finite system scheme)
For every $t \in (0,\infty)$ and τ defined by (5.7) respectively $\Lambda_\varrho^{c,d}$ in (5.8):

$$\mathcal{L}(\tilde{Y}^N(t)) \underset{N\to\infty}{\Longrightarrow} \int Q_t(\tau, d\varrho)\Lambda_\varrho^{c,d}, \tag{5.21}$$

where $Q_t(\cdot,\cdot)$ is the transition kernel of $\bar{Y}(t)$, the Fleming-Viot process with selection, where the resampling rate equals $cd/(c+d)$ the selection rate equals $sc/(c+d)$ and the selection matrix $V(u,v) = (\chi(u) + \chi(v))$ (recall 5.16). □

The proof of the key results of this section namely Proposition 5.1 and Lemma 5.5 are in section 5 (e) respectively 5 (f).

(d) Proof of the Lemmata 5.0 - 5.4

In this section we collect the proofs of those Lemmata, which do not rely on longer arguments.

Proof of Lemma 5.0 Since all components develop independently it will turn out that it suffices to consider the induced $\mathcal{P}(\mathbb{N})$-valued process of one single component which is obtained by replacing $\Theta_t(z)$ in the definition of G_t by the expression $Ez(t)$, if $z(t)$ denotes the $\mathcal{P}(\mathbb{N})$-valued process. Denote this new generator by $\widehat{G}$. We have to show that the $(\widehat{G}, \mu)$-martingale problem is for every $\mu \in \mathcal{P}(\mathcal{P}(\mathbb{N}))$ well-posed. Since $\widehat{G}$ includes a term which is a functional of the current distribution rather than the current state we are in the set up of socalled *nonlinear Markov processes*. In this context we can proceed by showing two things. For every continuous function $\lambda : [0, \infty) \to \mathcal{P}(\mathbb{N})$ there exists a unique solution of the $(\widehat{G}_\lambda, \mu)$-martingale problem, where $\widehat{G}_\lambda$ is obtained from $\widehat{G}$ by replacing the term $E(z(t))$ by $\lambda(t)$. Furthermore let $\mathcal{L}((z_\lambda(t))_{t\geq 0})$ be the solution of the $(\widehat{G}_\lambda, \mu)$-martingale problem and define $\widehat{\lambda}(t) = E(z_\lambda(t))$. Then we need to show that the map $\lambda \to \widehat{\lambda}$ has a unique fixed point in $C([0, \infty), \mathcal{P}(\mathbb{N}))$. Since we proved this in chapter 3 for the more general case including mutation we refer the reader to that section.

To conclude we note that the law of $(Y(t))_{t\geq 0}$ is an exchangeable law on $(\mathcal{C}([0, \infty),\ \mathcal{P}(\mathbb{N})))^{\mathbb{N}}$ and each component is a separable Banach space. However due to the uniqueness of the mean curve the spatial average is deterministic hence equal to the mean of a component. This completes the proof.

Proof of Lemma 5.1 Note that both $[Y^N(t),\ \Theta^N(t)]$ with $\Theta^N(t)$ the average over the active components 1 to N of the system extended to all of $\mathbb{N}$ by the frozen indices and $[Y(t),\ \Theta(Y(t))]$ are Markov processes on the same state space $(\mathcal{P}(\mathbb{N}))^{\mathbb{N}} \times \mathcal{P}(\mathbb{N})$ and hence uniquely determined by the corresponding semigroups (note that $Y(t)$ is a stationary process even exchangeable (in space for all t).

The operator G_N generates a unique semigroup on $\mathcal{P}(\mathbb{N})^{\mathbb{N}}$. The same is true according to Lemma 5.0 for G once we extend the state space to $(\mathcal{P}(\mathbb{N}))^{\mathbb{N}} \times \mathcal{P}(\mathbb{N})$ where the second component is the overall average of the system. Note that since we start and hence stay in exchangeable initial laws (hence the spatial intensity is well-defined for every t) we can view the process $Y^N(t)$ also as a Markov process on $(\mathcal{P}(\mathbb{N}))^{\mathbb{N}} \times \mathcal{P}(\mathbb{N})$ by adding the total intensity to the description of the current state. The generators on these extended state spaces we denote by $\tilde{G}_N$, $\tilde{G}$. Consider now functions F depending only on the first coordinate of these extended processes.

Let $E_{N,t}$ denote expectation with respect to z distributed according to $\mathcal{L}(Y^N(t))$. Then by construction we have that for the function F depending only on finitely many components that is, $\frac{\partial}{\partial z_i} F = 0$ for all but finitely many i and the property that $\frac{\partial F}{\partial z_i}$ is bounded and continuous for all $i \in \mathbb{N}$:

$$
\begin{aligned}
(5.22) \qquad E_{N,t}|(G_N - G)(F)_{(z)}| &= E|(s/N) \sum_{i=1}^{N} \int\int\int \frac{\partial F(z)}{\partial z_i}(u)(\chi(v) + \chi(w)) z_i(dv) Q_{z_i}(du, dw) \\
&+ \quad c \int \frac{\partial F(z)}{\partial z_i}(u) \Big\{ \Theta(du) - \frac{1}{N} \sum_{i=1}^{N} z_i(du) \Big\} | \underset{N\to\infty}{\longrightarrow} 0,
\end{aligned}
$$

for every t as can be seen as follows.

The first term is of the form $\frac{1}{N} O(1)$ if $\frac{\partial F}{\partial z_i}$ is bounded. In the second term we have to show that $N^{-1} \sum z_i(du) \Longrightarrow \Theta(du)$ as $N \to \infty$, if z is distributed according to $\mathcal{L}(Y^N(t))$. For this purpose we notice that the distribution of $Y^N(t)$ restricted to $(\mathcal{P}(\mathbb{N}))^N$ are exchangeable for every $t \geq 0$. Then according to Lemma 4.3 in [DGV] along every convergent subsequence of $\mathcal{L}(Y^N(t))$ the quantity $N^{-1} \sum_{i=1}^{N} z_i$ converges to $\lim_{N\to\infty}(N^{-1} \sum_{i=1}^{N} z_i)$ under the limit law. Since $\{\mathcal{L}(Y^N(t))\}_{N\in\mathbb{N}}$ is tight, this gives the desired convergence.

In order to use the information above in (5.22) we proceed as follows. There exists a set of functions F depending on finitely many coordinates with $\frac{\partial F}{\partial z_i}$ continuous and bounded which are dense in $C_0(\mathcal{P}(\mathbb{N})^{\mathbb{N}})$ and this set of F is preserved under the evolution defined by G for a prescribed path of the second component $\Theta(Y(t))$. (Take for example the orbit of functions of the form $F(z) = \prod_{i\in I}\langle z_i, f\rangle$ for $|I| < \infty$ and f bounded). Hence we get (5.6) immediately from (5.22), via the formula of partial integration for semigroups which reads a follows:

$$\tilde{U}_{\tilde{G}_N}(t) = \tilde{U}_{\tilde{G}}(t) + \int_0^t U_{\tilde{G}_N}(t-s)(\tilde{G}_N - \tilde{G})U_{\tilde{G}}(s)ds. \tag{5.23}$$

Proof of Lemma 5.2 Since all components evolve independently it suffices to study the component process. If we consider a process $Y(t)$ starting already with intensity measure τ, concentrated on the types of maximal fitness we can apply a Theorem by Ethier and Kurtz [EK5] which says that as $t \to \infty$ a Fleming-Viot process with immigration-emigration converges to a unique equilibrium state $\Lambda_\tau^{c,d}$.

On the other hand using the fact that $Y(t)$ is the McKean-Vlasov limit of $Y^N(t)$ as $N \to \infty$ we know from a Lemma, (see (5.36)) that we shall prove below, that the mean fitness increases to the maximal value. Hence the process $Y(t)$ concentrates on maximal types as $t \to \infty$ and the expected weight of those maximal types converges to a measure τ. We need to show that the process restricted to the maximal types has law $\Gamma_\tau^{c,d}$.

Again passing to the McKean-Vlasov limit in a statement which we shall prove for $Y^N(t)$ in all detail, we get that we can *couple* the Y-processes starting in $\mathcal{L}(Y(0))$, respectively in a state with mean measure τ for example $\Gamma_\tau^{c,d}$, that is we can construct the process $(Y^1(t), Y^2(t))$ whose marginals are the laws of our two given processes and which satisfy for all f in a measure determining class:

$$E(|\langle Y^1(t) - Y^2(t), f\rangle|) \text{ converges to } 0 \text{ as } t \to \infty, \tag{5.24}$$

which gives the assertion. Since we carry out this coupling argument in great detail for $Y^N(t)$ in (e) step 2 and (f) step 2 we omit at this point further details of this (easier) case.

Proof of Lemma 5.3 First we observe that in the limit $t \to \infty$ only the types u with $\varphi(u) = \max_v \varphi(v)$ will survive (compare (5.35), (5.36) and (5.40) for a rigorous proof). Therefore the assertion is a statement about an ordinary Fleming-Viot process with initial measures μ satisfying $E^\mu(z(u)) = \tau(u)$ for $u \in \mathbb{N}$. Therefore we have an explicit representation of the equilibrium measure (see [DGV] Theorem 0.3) and we know that (5.9) holds from [DGV] (equation (2.3)).

The second equation follows by explicit calculation using the GEM representation of the measure $\Gamma_0^{c,d}$ (compare [DGV]).

Proof of Lemma 5.4 We use here the semi-martingale structure of the process $\bar{Y}^N(t)$ which was defined via the generator $\bar{G}_N$ of (5.15). This allows to apply a standard tightness criterion. First observe that the second term in the operator $\bar{G}_N$ in (5.15) vanishes if the initial state is concentrated on types u satisfying

$$\varphi(u) = \max_v \varphi(v). \tag{5.25}$$

Since we consider the time scale Nt we can assume this without loss of generality as we shall see later in (5.40). Let us therefore here assume that (5.25) holds.

The process $\bar{Y}^N(t)$ is a semi-martingale and the martingale part corresponds to the third term in the operator $\bar{G}_N$, while the finite variation part corresponds to the first term (drift term). Under the assumption (5.25) the first term in the definition of $\bar{G}_N$ has uniformly in N, $t \le T$ bounded coefficients of $\frac{\partial F}{\partial z_i}$ and the second term vanishes. Hence the finite variation parts of $\bar{Y}^N(tN)$ has uniformly in N and $t \le T$ bounded local characteristics. Since we know from [DGV] (4.50)-(4.55) that the martingale part of $(\tilde{\bar{Y}}^N(t))_{t\ge 0}$ has also uniformly in N and $t \le T$ bounded local characteristics we are finished using a standard tightness criterion (Prop. 3.2.3) in [JM] (see (8.48), (8.49) for a quotation).

(e) Proof of Proposition 5.1

In order to prove this Proposition we shall proceed in two steps. In the first step we shall consider initial distributions of the process with the additional property that

$$(5.26) \qquad y^N_\xi(0)(u) = 0 \text{ a.s., if } \varphi(u) < \sup_v \varphi(v).$$

This has the advantage that only selection of order N^{-1} is at work, which is what we then study on large time scales of order Nt. In the second step we then remove (5.26) and shall incorporate the selection of order 1. Here we shall use a Lyapunov function which allows us to control the distance of $Y^N(t)$ from the manifold of types which satisfy $\varphi(u) = \sup_v(\varphi(v))$ and furthermore we can control via coupling the distance between the systems where one starts with intensity measure τ and the other with Θ.

Step 1 We assume (5.26) throughout this step. The first fact we will establish is that under this assumption over time intervals of length $L(N)$ with $L(N) \uparrow \infty$ as $N \to \infty$ and $L(N)/N \to 0$ as $N \to \infty$ the selection term plays no role and the process can be compared with a finite system of interacting Fleming-Viot processes with (mean-field) migration only. The latter process has been studied in [DGV] section 4 in great detail.

Let $Y^N(t)$ be the solution of (5.3), the martingale problem with generator G_N but with $\varphi \equiv 0$ so that only $N^{-1}\chi$ determines the fitness level and let $Z^N(t)$ be the corresponding Fleming-Viot process with migration, i.e. the process defined by the generator in (5.3) with $s = 0$. Both processes start with the same initial state. Then we shall prove below:

Lemma 5.6 *For $L(N) \uparrow \infty$, $L(N) = o(N)$, uniformly in the initial distribution,*

$$(5.27) \qquad \mathcal{L}(Y^N(L(N))) - \mathcal{L}(Z^N(L(N))) \underset{N\to\infty}{\Longrightarrow} 0. \qquad \square$$

Hence if we want to study $Y^N(tN)$, consider the law of $Y^N(tN - L(N))$ first instead. By the above Lemma together with assumption (5.26) this can be approximated by the $\mathcal{L}(Z^N(L(N)))$ if we set $\mathcal{L}(Z^N(0)) = \mathcal{L}(Y^N(tN - L(N)))$. Note that by the tightness of $\tilde{\tilde{Y}}^N$ the random process $(\bar{Y}^N(tN - L(N)))_{t\geq 0}$ behaves asymptotically exactly as $(\bar{Y}^N(tN))_{t\geq 0}$. However for the process $Z^N(t)$ we can use the results of [DGV], section 4, which say that (recall N_k is the subsequence along which $\tilde{\tilde{Y}}^N$ converges):

$$(5.28) \qquad \mathcal{L}(Z^{N_k}(L(N_k))) \underset{k\to\infty}{\Longrightarrow} \int \Lambda^{c,d}_\varrho Q^*_t(\Theta, d\varrho)$$

where (recall (5.8) and (5.17))

$$(5.29) \qquad Q^*_t(\Theta, \cdot) = \mathcal{L}(\bar{Y}^*(t) | \bar{Y}^*(0) = \Theta).$$

Since under the assumption (5.26) $\Theta = \tau$, the claim is proved at least under the assumption (5.26) once we have verified Lemma 5.6.

Proof of Lemma 5.6 Here we have to prove that if we have a generator of the form $G^N_1 + N^{-1}G^N_2$ such that G^N_1 and $G^N_1 + N^{-1}G^N_2$ generate contraction semigroups on the same L_∞ (for every N), then the corresponding semigroups $U^{i,N}_t$, $i = 1, 2$ satisfy for a constant $C_f < \infty$ the inequality:

$$(5.30) \qquad |(U^{1,N}_t - U^{2,N}_t)(f)| \leq C_f \frac{t}{N}$$

for at least a dense subset of L_∞ *provided* G^N_1, G^N_2 satisfy some *extra conditions.*

To find these conditions use that by the formula of partial integration for every f such that $U^{1,N}_t f \in D(G^N_2)$:

$$(5.31) \qquad (U^{1,N}_t - U^{2,N}_t)(f) = N^{-1} \int U^{2,N}_{t-s}(G^N_2 U^{1,N}_s)(f) ds.$$

We see that we need exactly the property: There exists a subset $\mathcal{L} \subseteq L_\infty$ with $\mathcal{L} = \bigcup_{a \in \mathbb{R}^+} \mathcal{L}_a$ and which satisfies that for fixed t and finite function h and for every $N \in \mathbb{N}$:

$$(5.32) \qquad \mathcal{L} = \text{ is measure determining}, \quad U_t^{1,N}(\mathcal{L}_a) \subseteq \mathcal{L}_{h(a)}, \quad ||G_2^N(f)||_\infty \leq C_a ||f||_\infty, \ \forall f \in \mathcal{L}_a.$$

Next we have to verify that we can satisfy these conditions, which will complete the proof.

We can choose for $\mathcal{L}_a$ those Lipschitz continuous functions with Lipschitz constant less than a which are also contained in the domain of G_1^N and G_2^N. Note that the domains of G_i^N are increasing in N since less and less components are frozen. A function f on $\mathcal{P}(\mathbb{N})^N$ is called Lipschitz continuous with Lipschitz constant $L(f)$ if

$$(5.33) \qquad \sup_z \left\{ |f(z) - f(z')| \, ||z - z'||^{-1} \right\} = L(f) < \infty,$$

with $||\cdot||$ denoting the *Vasherstein distance* between measures (see r.h.s. of (5.55) for a definition) with respect to some metric on $\mathbb{N}$, which is bounded. By varying the metric used in the Vasherstein distance we can cover all functions, which depend on the weights of finitely many types in such a way that we can view it as a Lipschitz function as function on $\mathbb{R}^n$ (with n labeling the components f depends on).

We are now left with two tasks: we have to show first that $U_t^{1,N}$ preserves the Lipschitz constant and second the preservation of the smoothness under $U_t^{1,N}$ (G_2^N is a differential operator of order 1), that is, the fact that a function, which is differentiable as function of the components preserves this property in time. The first fact follows via coupling techniques, which imply that $L(U_t^{1,N}(f)) \leq L(f)$, where L denotes the Lipschitz constant. This is found in [DGV] relations (4.35) - (4.39). The second fact is a property of the generator G_1^N in the above context and follows from $G_1^N U_t^{1,N} = U_t^{1,N} G_1^N$ together with G_1^N is differential operator and $\partial/\partial z$ is bounded on $\mathcal{L}_a$.

Step 2 In this step we have to show that the types with $\varphi(u) < \sup_v \varphi(v)$ have "disappeared“ by time Nt (with $t > 0$) and the effective intensity of the remaining types is τ. In other words, the assumption (5.26) will be justified, and moreover, we can use a modified initial intensity τ instead of Θ. Without loss of generality we can assume that

$$(5.34) \qquad \Theta(u) > 0 \text{ for all } u \in \mathbb{N}.$$

The central tool here is Lemma 5.7 below stating the fact that the *mean fitness* and the *mean intensity* of types of *maximal* fitness are both *Lyapunov functions* that is monotone (nondecreasing) functions in time under the evolution. Define for abbreviation

$$(5.35) \qquad \lambda(u) = \varphi(u) + N^{-1}\chi(u).$$

Lemma 5.7 *For every exchangeable initial state, $i \in \{0, 1, \dots, N-1\}$ and $t \geq 0$:*

$$(5.36) \qquad \begin{aligned} \frac{d}{dt} E(\langle y_i^N(t), \lambda \rangle) &= E(\ Var_{y_i^N(t)}(\lambda)) \geq 0, \\ \frac{d}{dt} E(y_i^N(t)(u)) &= E\Big(\lambda(u) y_i^N(u) - \Big(\sum_v \lambda(v) y_i^N(v)\Big) y_i^N(u)\Big) \cdot s \\ &\geq 0 \quad \text{if } \lambda(u) = \sup_v \lambda(v). \qquad \square \end{aligned}$$

Proof Start with the first relation and note that the function $F(z) = \langle z_i, \lambda \rangle, \ z \in (\mathcal{P}(\mathbb{N}))^{\mathbb{N}}$ is in the domain of the generator of the martingale problem (5.3). Furthermore:

$$(5.37) \qquad \frac{\partial F(z)}{\partial z_i}(u) = \lambda(u) \text{ and } \frac{\partial^2 F(z)}{\partial z_i \partial z_i}(u, v) \equiv 0,$$

so that an explicit calculation using (5.3) gives the first relation of (5.36), if we use that due to the exchangeable initial state $E\langle y_i^N(t), \lambda\rangle$ is independent of i and hence the migration term does not contribute. For the second relation of (5.36) consider $F(z) = \langle z_i, 1_{\{u\}}\rangle$ and proceed as before. This completes the proof of Lemma 5.7.

The last lemma implies immediately that the *mean-fitness* at site $i \in \{0, 1, \dots, N-1\}$ is strictly increasing as long as λ is not constant $\mathcal{L}(Y^N(t))$-a.s. on the support of the measure $y_i^N(t)$. From the second relation of (5.36) we know that the mean mass of a type of maximal fitness if nondecreasing. Hence since we assume that $\Theta(u) > 0$ for all $u \in \mathbb{N}$ (and this remains true for all $t > 0$ as is seen from the differential equation for $Ez_i(t)(u)$ immediately) we obtain strong conclusions from Lemma 5.7 which we formulate now.

Define:

$$I = \{u \in \mathbb{N} | \varphi(u) < \sup_v \varphi(v)\}. \tag{5.38}$$

We know that for all $N \in \mathbb{N}$, $0 \le s \le L(N)$:

$$E\left(\sum_{u\in I} y_i^N(tN-s)(u)\right) \le E\left(\sum_{u\in I} y_i^N(tN-L(N))(u)\right). \tag{5.39}$$

Furthermore since by assumption φ and χ are *bounded* on $\mathbb{N}$ and *attain* their maximum, we can conclude that at the time $tN - L(N)$ (which tends to ∞ as $N \to \infty$) the total mass on the set I tends to 0 as $N \to \infty$, i.e.

$$E\left(\sum_{u\in I} y_i^N(tN-L(N))(u)\right) \underset{N\to\infty}{\longrightarrow} 0, \tag{5.40}$$

by using (recall (5.35)) that φ is independent of N and hence by (5.36) this mean value actually decreases to zero.

In order to remove assumption (5.26) we need in addition to (5.40) and (5.39) another tool, namely a specific form of coupling. We introduce a map on $\mathcal{P}(\mathbb{N})$, denoted $\hat{f}$ which maps an element ν of $\mathcal{P}(\mathbb{N})$ into one with support $\mathcal{C}I$ (recall (5.38)) by:

$$\widehat{f}(\nu)(u) = \nu(u) / \sum_{u\in\mathcal{C}I} \nu(u),\ u \in \mathcal{C}I, \qquad \widehat{f}(\nu)(u) = 0 \qquad u \in I. \tag{5.41}$$

Such a map allows to redistribute mass of a general configuration in such a way that we get a state *concentrated on maximal states* i.e. satisfying (5.26). The above construction will help us together with (5.39) to establish an estimate on the effect of the types in I (defined in (5.38)) and get the analog of Lemma 5.6 of step 1, since we can control the effect of the selection of order 1 during the time stretch from time $tN - L(N)$ till tN. Namely, with the same notation as in step 1, (see text proceeding (5.27)) we can prove the following Lemma - the proof is deferred to section 5(f) using the coupling techniques which are developed in that subsection.

Lemma 5.8 *Let $\widehat{Y}^N(t)$ be a special solution of (5.3) satisfying (5.26) namely the one with initial configuration $\widehat{Y}^N(0) = \widehat{f}(Y^N(tN - L(N)))$. We can then define $Y^N(t)$ and $\widehat{Y}^N(t)$ on one probability space such that for $t > 0$:*

$$E||y_i^N(tN) - \widehat{y}_i^N(L(N))|| \le 2E(\sum_{u\in I} y_i^N(tN-L(N))(u)) \tag{5.42}$$

and $||\cdot||$ is the total variation norm of a (signed) measure on the space of types and I was defined in (5.38). □

To complete the proof of Proposition 5.1. we simply combine (5.40), (5.39), (5.42) and then apply to $\widehat{Y}^N(t)$ (5.27) from lemma 5.6 in step 1, to get via (5.28) and (5.29) the result, up to the statement that the initial point of $\bar{Y}^*$ is really τ. By initial point we mean the following: We know that the process $(\bar{Y}_t^*)_{t>0}$ has continuous path and (the process is time homogeneous with bounded mean) the limit $\bar{Y}_t^*$ for $t \to 0$ exists and the process is the solution to the $(\bar{G}_\infty, \nu)$-martingale problem with $\nu = \mathcal{L}(\bar{Y}_0^*)$. The fact $\nu = \delta_\tau$ will be shown in (5.84 - 5.89) after introducing the coupling techniques in the next subsection.

(f) Proof of Lemma 5.5, 5.8 and construction of a coupling

The purpose of this section is twofold namely to prove Lemma 5.5, 5.8 and to develop the *coupling-techniques* needed for this purpose and at many other places in this paper.

We begin in step 1 to prove Lemma 5.5 up to two assertions (Lemma 5.9, Lemma 5.10) which will be proved in step 3 together with Lemma 5.8. For the estimates needed in this step we develop *coupling techniques* in step 2.

Step 1 The fact that the $(\bar{G}_\infty, \tau)$-martingale problem is well-posed follows from [D], chapter 10. Hence we have to prove the remaining assertions and hence the proof will consist of two parts. First we shall show that the weak limit points of $\mathcal{L}((\bar{Y}^N(tN))_{t\geq\delta})$ is the law $\mathcal{L}((\bar{Y}(t))_{t\geq\delta})$ for a suitable initial measure $\mathcal{L}(\bar{Y}^*(\delta))$ which is then in the second part shown to converge to τ as $\delta \downarrow 0$.

Observe first the following fact which is implied by the construction of the process $(Y_t^N(t))$. Define (recall $\lambda(u) = \varphi(u) + N^{-1}\chi(u)$ and (5.13) for the definition of $\sim$):

$$(5.43) \qquad M_t^{N,f} = \langle \tilde{\tilde{Y}}^N(t), f\rangle - s\int_0^t \left(\int_{\mathbb{N}}\int_{\mathbb{N}}\int_{\mathbb{N}} f(u)(\lambda(v)+\lambda(w))\Big\{\frac{1}{N}\Big[\sum_{i=1}^N(\tilde{y}_i^N(s))(dv)Q_{\tilde{y}_i^N(s)}(du,dw)\Big]\Big\}\right)ds.$$

Here f is a bounded function on $\mathbb{N}$. Then for all $f \in L_\infty(\mathbb{N})$ the following two properties hold:

(5.44) $\quad (M_t^{N,f})_{t\geq 0}$ is a martingale with continuous path.

$$(5.45) \qquad \left((M_t^{N,f})^2 - d\int_0^t \frac{1}{N}\sum_{i=1}^N \operatorname{Var}_{\tilde{y}_i^N(s)}(f)ds\right)_{t\geq 0} \quad \text{is a martingale.}$$

This result follows by applying the generator in (5.15) to the functions $z \to \langle f, \frac{1}{N}\sum_{i=1}^N z_i\rangle^\ell$ on $(\mathcal{P}(\mathbb{N}))$ with $\ell = 1, 2$ and using that $\int f(u)f(v)Q_z(du,dv) = \operatorname{Var}_z(f)$.

To continue, define the analogous quantity M_t^f for the expected limit process (recall (5.16) for the definition of $\bar{Y}_t^*$ and recall that we have proved in Lemma 5.4 the tightness of the sequence $\tilde{\tilde{Y}}^N$). For that purpose abbreviate (recall that $\bar{G}_\infty$ involves only χ not φ):

$$(5.46) \qquad S_t^f = \langle \bar{Y}_t^*, f\rangle.$$

and define:

$$(5.47) \qquad M_t^f = S_t^f - \frac{c}{c+d}\int_0^t \left(\int_{\mathbb{N}}\int_{\mathbb{N}}\int_{\mathbb{N}} f(u)(\chi(v)+\chi(w))\bar{Y}_s^*(dv)Q_{\bar{Y}_s^*}(du,dw)\right)ds.$$

The key result for this latter object consists now of the following two Lemmata:

Lemma 5.9

(a) For every $f \in L_\infty(\mathbb{N})$ the following two properties hold for the process $(\bar{Y}_t^*)_{t\geq 0}$:

(i) (5.48) $\quad (M_t^f)_{t\geq 0}$ *is a martingale with continuous paths,*

(ii) (5.49) $$\left[(M_t^f)^2 - \frac{cd}{c+d}\int_0^t Var_{\bar{Y}^*(s)}(f)ds\right]_{t\geq 0} \quad \text{is a martingale.}$$

(b) $\bar{Y}^*(0) = \tau$ *with* τ *defined in (5.7).* □

Lemma 5.10 *The process* $(\bar{Y}(t))_{t\geq 0}$ *defined below (5.20) can be characterized as follows: for given initial point* $\bar{Y}(0)$ *it is the unique process satisfying for all* $f \in L_\infty(\mathbb{N})$ *the relations (5.48) and (5.49).*

The last two Lemmata combined prove that $\bar{Y}^*(t)$ must have the asserted evolution mechanism and hence be equal to $\bar{Y}(t)$.

The proof of the three key Lemmata 5.8 - 5.10 above follows ideas first used in a simpler context in [D] and [DGV]. A main point is the construction of a *coupling* of two processes which start in different states or which are driven by different drift terms. Before we start with the remaining proofs in step 3 we generalize the concept of coupling in [DGV] to processes involving *selection*, the latter complicates the matter quite a bit.

Step 2 (Coupling) We will have to couple two processes where the drift terms are different while the diffusion term is of the same form, that is, we have *two different dynamical systems* perturbed by the *same form of noise*. Coupling means here that we realize the two given processes on *one probability space* in such a way that they are in a suitable sense as close together as possible. In other words we have to define a new dynamic on $\mathcal{P}(\mathbb{N}\times\mathbb{N})$ where each component process (i.e. the $\mathcal{P}(\mathbb{N})$-valued projections) follows the given mechanism but the two evolutions are driven closer together. Here we will have to define a suitable notion of distance between two elements of $\mathcal{P}(\mathbb{N})$.

We proceed as follows, first in α) the problem is formally stated and mapped onto the coupling of $\mathcal{P}([0,1])$-valued processes, then in β) the coupled dynamics is constructed and finally in γ) we investigate how close the components are to each other.

α) For the coupling it is useful to map the problem back into a problem for probability measures on [0,1] instead of $\mathbb{N}$. The suitable way for this is to use the fitness function λ for that purpose since we can w.l.o.g. identify all types which have equal fitness (this is true in the case of $H(z,v,u) \equiv$ const only!). Hence we use the fitness function λ and define a map ϕ

(5.50) $$\phi : \mathbb{N} \to [0,1]$$
$$\phi(u) = \lambda(u)/||\lambda||_\infty$$

and applying ϕ to the type space of our system on $\mathcal{P}(\mathbb{N})$ we obtain a system on $\mathcal{P}([0,1])$ with the same form of the generator of the martingale problem. However we have a new fitness function to use, which we call ψ, where

(5.51) $$\psi : [0,1] \to \mathbb{R}^+,\ \psi(u) = ||\lambda||_\infty u.$$

The two processes which we want to couple have the following form. We have two processes defined by time-inhomogeneous martingale problems with generators of the form ($Y = (y_k)_{k\in\{1,\dots,N\}}$ and $i = 1,2$):

(5.52) $$L_{\tilde{\Theta}_i}(F(Y(s))) = \sum_{k=1}^N \Big[c\int_0^1 \frac{\partial F(Y(s))}{\partial y_k}(u)(\tilde{\Theta}_i(s) - y_k(s))(du)$$
$$+ s\int_0^1 \frac{\partial F(Y(s))}{\partial y_k}(u)(\psi(v)+\psi(w))y_k(s)(dv)Q_{y_k(s)}(du,dw)$$

$$+ d \int_0^1 \int_0^1 \frac{\partial^2 F(Y(s))}{\partial y_k \partial y_k}(u,v) Q_{y_k(s)}(du,dv)\big].$$

Here $\tilde{\Theta}_i = \tilde{\Theta}_i(s,(Y_k(s))_{k=1,\dots,N})$ $i = 1,2$ is assumed to have one of two special structures: either it is autonomous, i.e. does not depend on Y or $\tilde{\Theta}_i = \Theta(Y_i^N(s))$ with Θ defined by

$$(5.53) \qquad \langle \Theta(Y^N), f\rangle = \langle \frac{1}{N}\sum_{k=1}^N y_k, f\rangle.$$

In the case where we consider $N = \infty$ we view $(y_k(s))_{s\geq 0}$ as element of $C([0,\infty),\ \mathcal{P}([0,1]))$ and the relation (5.53) is replaced by:

$$(5.54) \qquad \langle \Theta(Y), f\rangle = \lim_{N\to\infty}(\frac{1}{N}\sum_{k=1}^N \langle y_k, f\rangle).$$

<u>$\beta)$</u> In order to couple the two processes corresponding to $L_{\tilde{\Theta}_1}$ and $L_{\tilde{\Theta}_2}$, we first want to couple the two driving terms $\tilde{\Theta}_1$ and $\tilde{\Theta}_2$, the immigration sources, as much as possible and then we want to use this coupled driving drift-term to define a $\mathcal{P}(\mathbb{N}\times\mathbb{N})$-valued dynamic producing the coupled process $\bar{\mathcal{Y}}(t)$. We proceed to carry out this construction in the following two steps.

We begin with the first object, that is, for two measures Θ_1 and Θ_2 on [0,1] we want to construct the *coupled measure* $\widehat{\Theta}$ *on* $[0,1]^2$ such that the marginals are Θ_1 and Θ_2, respectively, and the measure is supported as close as possible to the diagonal in the sense that:

$$(5.55) \qquad \int_0^1\int_0^1 |u-v|\widehat{\Theta}(du,dv) \quad = \inf \quad \Big[\int_0^1\int_0^1 |u-v|\Theta(du,dv);\ \Theta\in\mathcal{P}([0,1]^2),$$
$$\pi_i(\Theta) = \Theta_i \quad i=1,2\Big],$$

where π_1, π_2 denotes the projection on the first and second component. The quantity given by (5.55) defines the Vasherstein distance $||\Theta_1 - \Theta_2||$.

This measure $\widehat{\Theta}$ can be found by a continuous mapping $C : \mathcal{P}([0,1]) \times \mathcal{P}([0,1]) \to \mathcal{P}([0,1]^2)$ as follows: Consider the measurable space $([0,1], \mathcal{B}[0,1])$ equipped with the uniform distribution and put

$$(5.56) \qquad C((\Theta_1,\Theta_2)) := \mathcal{L}((W_{\Theta_1}, W_{\Theta_2}))$$

where on the event $\{U = u\}$, with U uniformly distributed, we defined the values of W_{Θ_1} and W_{Θ_2} by

$$(5.57) \qquad W_\Theta(u) := \inf\{v : \Theta([0,v]) \geq u\} \quad \text{for} \quad 0 \leq u \leq 1.$$

Now we define a second object, the *bivariate dynamics* $(\bar{\mathcal{Y}}(s))_{s\geq 0}$ as the unique solution of the $(L_{\widehat{\Theta}},\ \nu\otimes\mu)$-martingale problem with $L_{\widehat{\Theta}}$ given in (5.60). We write for an element $\bar{\mathcal{Y}}$ of $(\mathcal{P}([0,1]^2))^N$, $\bar{\mathcal{Y}} = (\bar{y}_k)_{k=1,\dots,N}$, and for $\bar{u}\in[0,1]^2$ put $\bar{u} = (u_1,u_2)$. We define the bivariate selection

$$(5.58) \qquad \bar{\psi}(\bar{u}) = \psi(u_1) + \psi(u_2)$$

and the coupled center of drift (coupled immigration source)

$$(5.59) \qquad \widehat{\Theta}(s) = C(\tilde{\Theta}_1(s)\tilde{\Theta}_2(s)).$$

Then $L_{\hat{\Theta}}$ and the initial law are defined by:

$$(5.60) \qquad L_{\hat{\Theta}} F(\bar{\mathcal{Y}}(s)) = \sum_{k=0}^{N} \Big[c \int\limits_{[0,1]^2} \frac{\partial F(\bar{\mathcal{Y}}(s))}{\partial \bar{y}_k}(\bar{u})(\hat{\Theta}(s) - \bar{y}_k(s))(d\bar{u})$$

$$+ s \int\limits_{[0,1]^2} \frac{\partial F(\bar{\mathcal{Y}}(s))}{\partial \bar{y}_k}(\bar{u})(\bar{\psi}(\bar{v}) + \bar{\psi}(\bar{w}))\bar{y}_k(d\bar{v}) Q_{\bar{y}_k}(d\bar{u}, d\bar{w})$$

$$+ d \int\limits_{[0,1]^2} \frac{\partial^2 F(\bar{\mathcal{Y}}(s))}{\partial \bar{y}_k \partial \bar{y}_k}(\bar{u}, \bar{v}) Q_{\bar{y}_k}(d\bar{u}, d\bar{v}) \Big].$$

$$(5.61) \qquad \mathcal{L}(\bar{\mathcal{Y}}(0)) = \nu \otimes \mu.$$

In words: $\bar{\mathcal{Y}}(s)$ is the Fleming-Viot process with immigration-emigration and selection but now on the new type space $[0,1]^2$, where the fitness of a type (u_1, u_2) is the sum of the fitness corresponding to u_1, respectively u_2, and where the immigration at time s is from a source $\widehat{\Theta}(s)$ given by the coupled sources of the individual processes at time s.

The *existence* and *uniqueness* of the solution to the above martingale problem is obtained by first using [DGV] page 2319 to establish this for the case where ψ is constant (no selection) and then use a Girsanov transformation to obtain the result with selection. For the latter compare the Proof of Lemma 5.10 and for the $N = +\infty$ case the Proof of Lemma 5.0. We do not repeat this argument here. Note that what we need in the sequel is the existence not the uniqueness of the solution to the martingale problem.

The process $\bar{y}(t)$ constructed in (5.60) - (5.61) has the *basic coupling property*, which reads as follows:

Lemma 5.11 *The solution $(\bar{\mathcal{Y}}(t))_{t \geq 0}$ to the $(L_{\hat{\Theta}},\ \nu \otimes \mu)$-martingale problem has the two marginals $(Y_1(t),\ Y_2(t))$ of the current state in $\mathcal{P}(\mathbb{N} \times \mathbb{N})$. The laws of $(Y_1(t))_{t \geq 0}$ resp. $(Y_2(t))_{t \geq 0}$ are solutions to the $(L_{\tilde{\Theta}_1}, \nu)$ resp. $(L_{\tilde{\Theta}_2}, \mu)$-martingale problems in (5.52).* □

Proof The key is that the bivariate selection is *additive* in the components of $\bar{u}$ and in the resampling term $H \equiv$ const (recall (1.18)). By considering functions F which depend only on the first respectively second marginal of the states $\bar{y}_k$, we check by explicit calculation that the action of the generator in (5.60) agrees with the action defined by the one in (5.52). Since the latter martingale problems were well-posed the assertion follows.

γ) The final step is to study the question of how close (in the mean sense) the coupled process $(\bar{\mathcal{Y}}(t))_{t \geq 0}$ is to a measure *concentrated* on the *diagonal* of $[0,1]^2$ provided that the driving terms Θ_i become close to the diagonal in the sense of (5.55). Define these two mean distances to the diagonal of $\bar{Y}^N(t)$ and $\widehat{\Theta}(t)$ by:

$$(5.62) \qquad h(t) = E\left[\frac{1}{N}\sum_{k=1}^{N} \int\limits_{[0,1]^2} |u_1 - u_2| \bar{y}_k(t)(du_1, du_2)\right]$$

$$(5.63) \qquad g(t) = E\left[\int\limits_{[0,1]^2} |u_1 - u_2| \hat{\Theta}(t)(du_1, du_2)\right].$$

To relate h with g we also need a term coming from the selection (let $\bar{v} = (v_1, v_2)$ etc.):

$$(5.64) \qquad f(t) = E\left(\int\limits_{[0,1]^2} \int\limits_{[0,1]^2} \int\limits_{[0,1]^2} |u_1 - u_2| \bar{V}(\bar{v}, \bar{w}) \bar{y}_k(t)(d\bar{v}) Q_{\bar{y}_k(t)}(d\bar{u}, d\bar{w})\right)$$

$$\bar{V}(\bar{v},\bar{w}) = V(v_1,w_1) + V(v_2,w_2).$$

We shall now derive differential inequalities which will allow us to estimate how small we can make $h(t)$ as t becomes larger, when $g(t)$ becomes sufficiently small. We obtain

Lemma 5.12

$$h(t) = h(0)e^{-t} + \int_0^t e^{-(t-s)}(g(s)+f(s))ds. \qquad \square \tag{5.65}$$

Proof Note that the function $H(\bar{\mathcal{Y}}) = \int_{[0,1]^2} |u_1-u_2|\bar{y}_k(du_1,du_2)$ has the property

$$\frac{\partial H(\bar{\mathcal{Y}})}{\partial \bar{y}_k}(\bar{u}) = |u_1-u_2|, \quad \frac{\partial^2 H(\bar{\mathcal{Y}})}{\partial \bar{y}_k \partial \bar{y}_k}(\bar{u},\bar{v}) = 0 \tag{5.66}$$

and hence using the equation (5.60) gives that

$$\dot{h}(t) = g(t) + f(t) - h(t). \tag{5.67}$$

This differential equation for h can be explicitly solved as stated.

Remark The quantity $f(t)$ in the formula (5.65) is of course not so easy to control since it involves higher moment measures of the process. However for additive selection we can estimate this part in the sequel, since then only second moments are involved.

Define the function $\hat{\psi}(z)$ on $\mathcal{P}([0,1]^2)$ of the average fitness of a state by

$$\hat{\psi}(z) = \int_{[0,1]^2} \bar{\psi}(\bar{u})z(d\bar{u}) \quad \text{where } \bar{\psi}(\bar{u}) = \psi(u_1) + \psi(u_2). \tag{5.68}$$

Then

$$\begin{aligned} f(t) &= E\left[\int_{[0,1]^2} \bar{\psi}(\bar{u})|u_1-u_2|(\bar{y}_k(t)(d\bar{u})) - \hat{\psi}(\bar{y}_k(t)) \cdot \int_{[0,1]^2} |u_1-u_2|\bar{y}_k(t)(d\bar{u})\right] \\ &= E(\text{cov}_{\bar{y}_k(t)}(\bar{\psi},\Delta)), \quad \text{with } \Delta(\bar{u}) = |u_1-u_2|. \end{aligned} \tag{5.69}$$

This means in particular that

$$f(t) = 0, \text{ if either: } \bar{\psi} \equiv \text{ const on supp}(\bar{y}_k(t)) \text{ or: supp } (\bar{y}_k(t)) \subseteq \text{ diag } ([0,1]^2) \text{ a.s.} \tag{5.70}$$

Furthermore from (5.69) we can conclude

$$|f(t)| \le E([\text{Var}_{\bar{y}_k(t)}(\bar{\psi}) \cdot \text{Var}_{\bar{y}_k(t)}(\Delta)]^{\frac{1}{2}}). \tag{5.71}$$

If we use furthermore that by assumption on ψ, $2|\psi(u)| \le D$ then the last equation implies the estimate:

$$|f(t)| \le D \cdot E[(\text{Var}_{\bar{y}_k(t)}(\Delta))^{\frac{1}{2}}]. \tag{5.72}$$

By inserting (5.72) in (5.65) we obtain the key relation of this part γ) namely:

Lemma 5.13 *If $|\psi(u)| \leq D$ for all $u \in [0,1]$, then with $\Delta(\bar{u}) = |u_1 - u_2|$:*

$$(5.73)\qquad h(t) \leq h(0)e^{-t} + \int_0^t e^{-(t-s)}(g(s) + \hat{f}(s))ds, \quad \hat{f}(s) = D \cdot E(\mathrm{Var}_{\bar{y}_k(s)}(\Delta))^{\frac{1}{2}}. \qquad \square$$

In the sequel, with this estimate we can proceed with many arguments as in [DGV] section 4.

Step 3 (Proof of Lemmata 5.8, 5.9, 5.10)

Proof of Lemma 5.8 The key to the proof is a *coupling argument*, with the coupling dynamics as defined in step 2 above, but this time we need to start the coupled dynamics in a special initial state which is suitable for our purposes. We now define this.

Recall $\hat{f}$ defined below (5.39). Given an element $(y_k)_{k=1,\dots,N} \in (\mathcal{P}(\mathbb{N}))^N$ and given the set I (see 5.38) of types, we define for the two configurations $(y_k)_{k=1,\dots,N}$ and $(\hat{f}(y_k))_{k=1,\dots,N}$ in $(\mathcal{P}(\mathbb{N}))^N$ the bivariate configuration $\bar{y}^N \in (\mathcal{P}(\mathbb{N}\times\mathbb{N}))^N$ as follows. Set $\bar{y}^N = (\bar{y}_k)_{k=1,\dots,N}$ with $\bar{y}_k \in \mathcal{P}(\mathbb{N}\times\mathbb{N})$ and put for every k:

$$(5.74)\qquad \bar{y}_k((i_1,i_2)) = \begin{cases} 0 & \text{if} \quad i_2 \in I \\ y_k(i_1) & \text{if} \quad i_1,i_2 \in \mathcal{C}I,\ i_1 = i_2 \\ 0 & \text{if} \quad i_1,i_2 \in \mathcal{C}I,\ i_1 \neq i_2 \\ \frac{y_k(i_2)}{\sum\limits_{j\in\mathcal{C}I} y_k(j)} y_k(i_1) & \text{if} \quad i_2 \in \mathcal{C}I,\ i_1 \in I. \end{cases}$$

Now note that with $\pi_i,\ i = 1,2$ denoting projections on the respective marginals we calculate:

$$(5.75)\qquad \pi_1\bar{y}_k = y_k \qquad \pi_2\bar{y}_k = \hat{f}(y_k).$$

Furthermore with $\bar{i} = (i_1,i_2)$ and $\mathrm{Diag} = \{(i,j) \in \mathbb{N}\times\mathbb{N} | i = j\}$:

$$(5.76)\qquad ||y_k - \hat{f}(y_k)||_{\mathrm{Var}} = 2\sum_{i\in I} y_k(i) = 2\sum_{\bar{i}\notin \mathrm{Diag}} \bar{y}_k(\bar{i}).$$

With the special initial state constructed above we now run our bivariate dynamics of (5.60). Observe that the fitness of all points in $I \times I$ or $I \times \mathcal{C}I$ is *not maximal*, while the fitness of all points in $\mathcal{C}I \times \mathcal{C}I$ is maximal. We know therefore, via Lemma 5.7 (5.36) that the total mass in $I \times \mathcal{C}I$ is decreasing and the mass on $\mathrm{Diag}(\mathcal{C}I \times \mathcal{C}I)$ increasing and all other points carry no mass over the whole evolution. This implies then via (5.76) immediately the assertion of the Lemma 5.8.

Proof of Lemma 5.9 Here we shall adapt the technique used in [DGV] section 4 (page 2325-2328) to prove Lemma 4.9 therein. The essential tool to achieve this adaptation is the coupling and the related estimates introduced above in (5.65) and (5.73). We have to show two things, first that the limiting local characteristics are given by $\bar{G}_\infty$ and second that the initial point is τ.

(i) Limiting local characteristics In order to show that the local characteristics are given by $\bar{G}_\infty$ proceed as follows. We use for $\bar{Y}^N(t)$ and $\bar{Y}^*(t)$ in this step the notation from [DGV] section 4, i.e. Θ_t^N and Θ_t. Define the empirical measures for finite and "infinite" N (recall (1.37)):

$$(5.77)\qquad E^N(s) = \frac{1}{N}\sum_{i=1}^N \delta_{y_i^N(s\beta(N))} \in \mathcal{P}(\mathcal{P}(\mathbb{N})),$$

$$(5.78)\qquad E^\infty(s) = \Gamma^0_{\Theta_s} \in \mathcal{P}(\mathcal{P}(\mathbb{N})).$$

Now for every function f on $\mathbb{N}$ which is bounded we can write the process Θ_s^N and the two different parts of the infinitesimal characteristics appearing in (5.43) and (5.45) as follows:

$$(5.79) \qquad \langle \Theta^N_{s\beta(N)}, f\rangle = \int\limits_{\mathcal{P}(\mathbb{N})} \langle \nu, f\rangle E^N(s)(d\nu)$$

$$(5.80) \qquad \frac{1}{N}\sum_{i=1}^N \mathrm{Var}_{y_i(s\beta(N))}(f) = \int\limits_{\mathcal{P}(\mathbb{N})} [\langle \nu, f^2\rangle - \langle \nu, f\rangle^2] E^N(s)(d\nu)$$

$$(5.81) \qquad \frac{1}{N}\sum_{i=1}^N \left\{\mathrm{cov}_{y_i(s\beta(N))}(\lambda, f)\right\} = \int\limits_{\mathcal{P}(\mathbb{N})} \mathrm{cov}_\nu(\lambda, f) E^N(s)(d\nu).$$

Furthermore for $\Phi \in \mathcal{P}(\mathcal{P}(\mathbb{N})),\ \Phi_s \in C([0,\infty), \mathcal{P}(\mathcal{P}(\mathbb{N}))),\ f \in L_\infty(\mathbb{N})$ and $T < \infty$, the maps:

$$(5.82) \qquad \Phi \rightarrow \int\limits_{\mathcal{P}(\mathbb{N})} \langle \nu, f\rangle \Phi(d\nu)$$

$$(\Phi)_{s\le T} \rightarrow \int\limits_0^T \Big[\int_{\mathcal{P}(\mathbb{N})} (\langle \nu, f^2\rangle - \langle \nu, f\rangle^2)\Phi_s(d\nu)\Big] ds$$

$$(\Phi)_{s\le T} \rightarrow \int\limits_0^T \Big[\int\limits_{\mathcal{P}(\mathbb{N})} \mathrm{cov}_\nu(\lambda, f)\Phi_s(d\nu)\Big] ds$$

are *continuous* on the spaces where they are defined.

Hence combining the relations (5.79) - (5.81) with (5.82) the result, that the limiting local characteristics are given by the expressions given in (5.48) and (5.49), will be implied by the stronger result:

$$(5.83) \qquad \mathcal{L}\left[\Big(\int\limits_0^s E^{N_k}(u)du, \quad \Theta^{N_k}_{s\beta(N_k)}\Big)_{s\ge 0}\right] \underset{k\to\infty}{\Longrightarrow} \mathcal{L}\left(\left(\int\limits_0^s E^\infty(u)du, \ \Theta_s\right)_{s\ge 0}\right).$$

This result however is obtained following the argument in [DGV] section 4, Proof of Lemma 4.9, step 2 on page 2336, where the only obstacle is to replace the coupling for the neutral case, which is given there by the one including selection. But this is now easy having developed the coupling results in Step 2. We only need to replace the estimate (4.36) in that paper by our (5.73) which takes care of the selection term. For details we refer the reader to the quoted pages in [DGV].

(ii) Limiting initial point Finally in order to show that $\bar{Y}^*(t) \to \tau$ as $t \downarrow 0$ we note that $\bar{Y}^*$ has continuous paths. We shall see that it suffices to prove that for every choice of $L(N)$ which satisfies $L(N) \uparrow \infty$ as $N \to \infty$ but $L(N) = o(N)$:

$$(5.84) \qquad \mathcal{L}(\bar{Y}^N(L(N)) \underset{N\to\infty}{\Longrightarrow} \tau.$$

By the Markov property of $Y^N(t)$ write the pair $(\tilde{Y}^N(t), \tilde{\tilde{Y}}^N(t))$ as evolution over the time interval $tN - L(N)$ starting in $(\tilde{Y}^N(L(N)), \bar{Y}^N(L(N)))$, we can obtain the result via the Lemma 5.4 together with relation (5.39), which implies that $\tilde{\tilde{Y}}^N(t_N + s)$ is *uniformly* in N not too far away from $\tilde{\tilde{Y}}^N(s)$ if s is small and $t_N = -L(N)/N$. To complete the proof we now turnto the proof of (5.84).

We proceed by showing that the l.h.s. of (5.84) has in the limit $N \to \infty$ the same law, as is obtained by performing the following three limits: first take the McKean-Vlasov limit of $(Y^N(t))_{t\geq 0}$, then let $t \to \infty$ and then finally form the spatial average in the infinite volume limit. If this *interchange of limits* is possible we have in particular using $E(y_i(t)) \to \tau$ as $t \to \infty$ (compare (5.7)):

$$E(y_i^N(L(N))(u)) \underset{N\to\infty}{\longrightarrow} \tau(u) \qquad \forall u \in \mathbb{N}. \tag{5.85}$$

Furthermore the possibility to interchange the limits as described implies that $\mathcal{L}((Y_i^N)(L(N)))_{i\in\mathbb{N}})$ has along every convergent subsequence N_k a limit which has components with mean measure τ and hence by forming the spatial average over an independent collection, each of whose members has mean measure τ, we obtain τ as the actual limit, as we wanted to show in (5.84).

In order to show that we can exchange the spatial averaging with the limits $t \to \infty$ and $N \to \infty$, we note first that for an exchangeable family of $\mathcal{P}(\mathbb{N})$-valued random variables $\{x_i^N,\ i \in \{0,1,\ldots,N\}\}$ with $\mathcal{L}((x_i^N,\ i \in \{0,1,\ldots,N\}))$ converging to $\mu \in \mathcal{P}(\mathcal{P}(\mathbb{N}))^{\mathbb{N}}$, the limit $N \to \infty$ and the averaging can be exchanged by [DGV], Lemma 4.3.

Hence it suffices to establish (5.39). Note first that the argument following (5.22) implies that we can assume without loss of generality $\chi \equiv 0$. Now we proceed in two steps for inferior and maximal types separately.

First consider $I = \{u|\varphi(u) < \max_v \varphi(v)\}$. By the first equation in (5.36) we know that

$$E\Big(\sum_{u\in I} y_t^N(t)(u)\varphi(u)\Big),\ E\Big(\sum_{u\in I} y_i(t)(u)\varphi(u)\Big) \tag{5.86}$$

are *monotone* decreasing to 0 in t and by (5.6):

$$Ey_i^N(t)(u) \underset{N\to\infty}{\longrightarrow} Ey_i(t)(u) \quad i \in \mathbb{N},\ t > 0. \tag{5.87}$$

Therefore since w.l.o.g. $\varphi > 0$ we know

$$Ey_i^N(L(N))(u) \underset{N\to\infty}{\longrightarrow} 0 = \tau(u) \quad \text{for } u \in I. \tag{5.88}$$

Next consider the maximal types $u \in \mathcal{C}I$. Then by (5.36) second equation

$$Ey_i^N(t)(u), \quad Ey_i(t)(u) \tag{5.89}$$

are increasing in t and (5.87) holds. As above the relation (5.88) holds also for $u \in \mathcal{C}I$. This completes the proof.

Proof of Lemma 5.10 We simply have to combine Theorem 7.2.2. and Lemma 7.2.1 from [D]. (The key point there is to prove a Girsanov type theorem which states that the martingale problem with selection is well-posed if this holds for the one without selection). We can conclude that the martingale problem is well-posed if this is the case for the problem without selection. The latter is proved via duality ([D], section 5).

6 Multiple space-time scale analysis and Proof of Theorems 2, 3

In this chapter we use the results of chapter 5 to analyze mean field models with a hierarchy of l-levels in the combination of time scales of different order. We shall derive approximation theorems of our infinite system defined in subsection 1 (a) by such l-level finite models which will allow us to prove our theorem 2 and 3. In section 6 (a) we discuss a two-level model and we use these results to study the general l-level models. The subsection 6 (b) contains the approximation results for infinite systems by l-level systems in different time scales and the proof of Theorem 2. The last subsection 6 (c) contains the proof of Theorem 3.

The basis of the present analysis is chapter 5 in [DGV] where the system without selection was treated. We shall therefore focus in this chapter on how the selection term is incorporated explaining into the analysis and we do not repeat the details of the remaining parts of the analysis.

(a) The two level model and ℓ-level generalizations

(i) We begin with a *two level model* $Y^N(t)$. In contrast to section 5(a) we shall now only consider the active components. Denote $M(2,N) = \{1,\dots,N\}^2$. Define the process $(Y^{N,2}(t))_{t\geq 0}$ on $(\mathcal{P}(\mathbb{N}))^{M(2,N)}$ by the $(G^2_N, \pi_\Theta \times \pi_\Theta)$-martingale problem. The components are denoted $(y^{N,2}_\xi)_{\xi\in\mathbb{N}\times\mathbb{N}}$. Here τ_Θ is again a homogeneous product measure on $\mathcal{P}(\mathbb{N})^N$ with mean measure Θ. We define for $z=(z_\xi)_{\xi\in M(2,N)}$ and $\xi=(\xi_1,\xi_2)$ with $\xi_i\in\{1,\dots,N\}$:

$$
\begin{aligned}
(6.1)\qquad G^2_N(F)_{(z)} \quad = \sum_{\xi_1,\xi_2=1}^{N} \quad & \Bigg[\Bigg\{\int \frac{\partial F(z)}{\partial z_\xi}(u)\Big[c_0(z_{\xi,1}(du)-z_\xi(du)) \\
& + \frac{c_1}{N}(z_{\xi,2}(du)-z_\xi(du))\Big]\Bigg\} \\
& + s\Bigg\{\int \frac{\partial F(z)}{\partial z_\xi}(u)[((\chi_0(v)+\chi_0(w)) + \frac{1}{N}(\chi_1(v)+\chi_1(w)) \\
& + \frac{1}{N^2}(\chi_2(v)+\chi_2(w))]z_\xi(dv)Q_{z_\xi}(du,dw)\Bigg\} \\
& + d_0\int \frac{\partial^2 F(z)}{\partial z_\xi \partial z_\xi}(u)Q_{z_\xi}(du,dv)\Bigg],
\end{aligned}
$$

where

$$
(6.2)\qquad z_{\xi,1} = \frac{1}{N}\sum_{\xi_1=1}^{N} z_{(\xi_1,\xi_2)}, \quad z_{\xi,2} = \frac{1}{N^2}\sum_{\xi_1,\xi_2=1}^{N} z_{(\xi_1,\xi_2)}.
$$

We assume that the selection function χ satisfies

(6.3) $\quad \chi_i,\ i=0,1,2$ are bounded functions on $\mathbb{N}$ attaining their maximum.

First note that, since the system is exchangeable, with arguments similar to those in section 5 (c) we obtain (after relabeling the indices) that for fixed t as $N\to\infty$ the system $Y^{N,2}(t)$ has as McKean-Vlasov limit the process $Y(t)$ with $c=c_0$ and $d=d_0$. (Recall (5.4) - (5.6)).

The question is now what happens in the time scales Nt respectively N^2t. Since in the time scale Nt the terms carrying the coefficient c_1 or χ_2 are of order $\frac{1}{N^2}$ and hence irrelevant, we get nothing new compared to the analysis of section 5. The rigorous argument for this fact is as the proof of Lemma 5.6 and we do not repeat this here. So we see that the task is to analyze the model as function of t in the time scale N^2t respectively N^2s+Nt. This has been carried out in [DGV] chapter 5 for a simpler model, where no selection term appeared.

Hence it suffices here to point out how we can incorporate the selection term into the analysis of the two level model viewed in the fast time scale N^2t. The results are formulated in Proposition 6.1 below. Recall the definitions (1.40) - (1.42).

Proposition 6.1 (Two level finite system scheme)
Assume that the initial law of $(Y^{N,2}(t))_{t\geq 0}$ *satisfies:*

$$\mathcal{L}(Y^{N,2}(0)) = \bigotimes_{i,j=1}^{N} \mu_{i,j}, \text{ with } \mu_{i,j} \equiv \mu_{1,1}, \int\limits_{\mathcal{P}(\mathbb{N})} y\mu_{1,1}(dy) = \tau^1_\Theta. \tag{6.4}$$

(a) *Let* $S(N)/N$
$\to \infty$, $S(N) = o(N^2)$ *and* $Y^{N,2} = (y^{N,2}_\xi)_{\xi \in M(2,N)}$ *where* $M(2,N) = \{1,\dots,N\}^2$. *Then the level two block averages of the level-two system satisfy (for* $y_{\xi,k}$ *recall (6.2)) :*

$$\mathcal{L}(y^{N,2}_{\xi,2}(S(N))) \underset{N\to\infty}{\Longrightarrow} \tau^1_\Theta, \tag{6.5}$$

$$\mathcal{L}((y^{N,2}_{\xi,2}(sN^2))_{s>0}) \Longrightarrow \mathcal{L}((\bar{Y}_s)_{s>0}), \tag{6.6}$$

where $\bar{Y}$ *is the unique solution to the* $(\bar{G}^2_\infty,\ \delta_{\tau^1_\Theta})$*-martingale problem on* $C([0,\infty),\mathcal{P}(\mathbb{N}))$ *with (for* d_k *see (1.36))*

$$\begin{aligned}\bar{G}^2_\infty(F)_{(z)} &= (s/d_0)d_2 \int \frac{\partial F(z)}{\partial z}(u)(\chi_2(v) + \chi_2(w))z(dv)Q_z(du,dw)\\ &+ d_2 \int \frac{\partial^2(z)F}{\partial z \partial z}(u,v)Q_z(du,dv).\end{aligned} \tag{6.7}$$

(b) *The level one block averages, respectively the components of the level two model, satisfy for every* $s > 0$ *and with* $Z^k_\Theta(t)$ *as defined in (1.39):*

$$\mathcal{L}((y^{N,2}_{\xi,1}(sN^2 + tN))_{t\geq 0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Z^1_{\Theta^*_1(s)}(t))_{t\geq 0}) \tag{6.8}$$

$$\mathcal{L}((y^{N,2}_{\xi,0}(sN^2 + t))_{t\geq 0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Z^0_{\Theta^*_0(s)}(t))_{t\geq 0}), \tag{6.9}$$

where $\Theta^*_k(s)$, $k = 0,1$ *is independent of the evolution and its law is defined as follows. Let* $\nu_s = \mathcal{L}(\bar{Y}_s | \bar{Y}_0 = \tau^1_\Theta)$ *and put*

$$\mathcal{L}(\Theta^*_1(s)) = \nu_s, \quad \mathcal{L}(\Theta^*_0(s)) = \int\limits_{\mathcal{P}(\mathbb{N})} \mu^{1,0}_\varrho \nu_s(d\varrho). \tag{6.10}$$

with $\mu^{j,k}_\Theta$ *as defined in (1.47).* □

Proof With the techniques we already used in section 5, the proof of the result of this subsection reduces to showing that by the time $N^2 t$ all types which are not χ_0 and χ_1 maximal have disappeared in the limit $N \to \infty$. More precisely we have to prove that for $L(N) \to \infty$ with $L(N)/N^2 \to 0,\ L(N)/N \to \infty$ as $N \to \infty$ (recall (1.40) and (1.42)):

$$\frac{1}{N^2} \sum_{\xi_1,\xi_2=1}^{N} y^{N,2}_{(\xi_1,\xi_2)}(L(N)) \underset{N\to\infty}{\longrightarrow} \tau^1_\Theta, \tag{6.11}$$

where τ^1_Θ is as defined in section 1 in (1.42). However, since the proof is only a routine modification of the arguments we presented in the proof of Lemma 5.8 we will in this subsection only give a few hints how to proceed after we formulated the key result and refer the reader to [DGV] section 5 (a) for the details of the remaining arguments.

For this purpose pick $L(N) \to \infty$ as $N \to \infty$ with $L(N) = o(N),\ L(N) \le S(N)$, write $T(N) = S(N) - L(N)$, and apply the Markov property

$$S(N) = L(N) + T(N), \qquad Q_s^\Theta = \mathcal{L}(Y^{N,2}(s)) \tag{6.12}$$

$$\mathcal{L}(Y^{N,2}(S(N))) = \int\limits_{\mathcal{P}(\mathbb{N})^{N2}} \mathcal{L}(Y^{N,2}(T(N))|Y^{N,2}(0) = z)Q^\Theta_{L(N)}(dz). \tag{6.13}$$

Now we proceed in two steps. First using the technique of the proof of Lemma 5.6 we see that the law $Q^\Theta_{L(N)}$ is approximated in the limit $N \to \infty$ by the dynamics of (6.1) with $\chi_2 \equiv 0,\ c_1 = 0$. Using the exchangeable structure of the dynamics and the initial law we are back in the situation of chapter 5 and we conclude that,denoting the induced law of the average $y^{N,2}_{\xi,2}$ by $\bar{Q}$:

$$Q^\Theta_{L(N)} \underset{N\to\infty}{\Longrightarrow} \Gamma^0_{\tau^0_\Theta}, \quad \bar{Q}^\Theta_{L(N)} \underset{N\to\infty}{\Longrightarrow} \tau^0_\Theta. \tag{6.14}$$

On the other hand it it clear that the system $Y^{N,2}(T(N))$ started in an exchangeable initial state (on $i, j \in \{1, \dots, N\}$ only), which is asymptotically (as $N \to \infty$) ergodic with intensity measure τ^0_Θ, then by the analysis of chapter 5 the law $\mathcal{L}(y^{N,2}_{\xi,2}(T(N)))$, with these random initial states, will converge to the point mass on τ^1_Θ.

The only point remaining is then to show that from here together with (6.13) we get the result (6.11). This follows, using *coupling techniques*, the same way as was done in the proof of Lemma 5.8.

At this stage the proof of the remaining two assertions (6.6) and (6.9) follows the line of argument which was given in [DGV] section 5 and hence further details are omitted here.

(ii) *The ℓ-level model* Essentially by induction we can extend the two-level analysis to *ℓ-level models* in a canonical way. We only give the definitions and the key result extending Proposition 6.1. to ℓ-levels. We need to define *finite* systems $(Y^{N,\ell}(t))$ on the state space $(\mathcal{P}(\mathbb{N}))^{M(\ell,N)}$ where $M(\ell, N) = \{1, 2, \dots, N\}^\ell$. This is done via the well-posed $(G^\ell_N,\ \delta_\tau)$-martingale problem. Define for $z = (z_\xi)_{\xi\in M(\ell,N)}$:

$$\begin{aligned} G^\ell_N(F))(z) \quad = \quad & \sum_{\xi\in M(\ell(N))} \Bigg[\int\limits_{\mathbb{N}} \frac{\partial F(z)}{\partial z}(u)(\sum_{i=1}^{\ell} \frac{c_{i-1}}{N^{i-1}}(z_{\xi,i} - z_\xi)(du)) \\ & +s \int\limits_{\mathbb{N}} \frac{\partial F(z)}{\partial z}(u) \left(\sum_{i=0}^{\ell} \frac{1}{N^i}(\chi_i(v) + \chi_i(w)) \right) z_\xi(dv)Q_{z_\xi}(du, dw) \\ & +d \int\limits_{\mathbb{N}} \int\limits_{\mathbb{N}} \frac{\partial^2 F(z)}{\partial z \partial z}(u, v)Q_{z_\xi}(du, dv) \Bigg] \end{aligned} \tag{6.15}$$

with

$$z_{\xi,i} = \frac{1}{N^i} \sum_{\zeta_1,\dots,\zeta_i=1}^{N} z_{(\zeta_1,\dots,\zeta_i,\xi_{i+1},\dots,\xi_\ell)}, \qquad \xi \in M(\ell, N),\ 1 \le i \le \ell.$$

We shall denote by $Y^{N,\ell}(t) = (y^{N,\ell}_\xi(t))_{\xi\in M(\ell,N)}$ the process defined by the $(G^\ell_N,\ \delta_\tau)$-martingale problem and get the following limit results.

Proposition 6.2 **(ℓ-level finite system scheme)**
Consider the process $(Y^{N,\ell}(t))_{t\ge0}$, with i.i.d. initial state and intensity measure $\Theta(= Ez_\xi)$. For $j \in \mathbb{N},\ k \in \{0, 1, \dots, \ell\}$

the limiting behavior as $N \to \infty$ is as follows:
case $j < k$:

$$(6.16) \qquad \mathcal{L}(y_{\xi,k}(S(N))) \underset{N\to\infty}{\Longrightarrow} \tau_\Theta^{j-1}, \text{ if } S(N)/N^{j-1} \to \infty \text{ and } S(N)/N^j \underset{N\to\infty}{\longrightarrow} 0,$$

$$(6.17) \qquad \mathcal{L}((y_{\xi,k}(tN^j))_{t>0}) \underset{N\to\infty}{\Longrightarrow} \delta_{\{y_t \equiv A_t^j\}} \quad \text{with } A_t^j \text{ as in (1.44)}$$

case $j = k$:

$$(6.18) \qquad \mathcal{L}((y_{\xi,k}(tN^k)_{t>0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((\bar{Y}(t))_{t>0})$$

where $\bar{Y}_t$ solves the $(\bar{G}^k_\infty,\ \delta_{\tau_\Theta^{k-1}})$-martingale problem on $C([0,\infty),\mathcal{P}(\mathbb{N}))$. $\bar{G}^j_\infty$ is obtained by replacing in (6.7) d_2 by d_k, and χ_2 by χ_k.
case $0 \le k < j < \ell$:

$$(6.19) \qquad \mathcal{L}((y_{\xi,k}(sN^j + tN^k))_{t>0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Z^k_{\Theta_k^*(s)}(t))_{t>0}),$$

where $\Theta_k^*(s)$ is independent of the evolution and has law:

$$(6.20) \qquad \mathcal{L}(\Theta_k^*(s)) = \int \mu_\rho^{j-1,k}\nu(s)(d\varrho) \quad \text{with} \quad \nu(s) = \mathcal{L}(\bar{Y}(s)|\bar{Y}(0) = \tau_\Theta^{j-1})$$

(recall (1.47) for $\mu_\varrho^{j,k}$).

If we replace in (6.19) the time $sN^j + tN^k$ by $S(N)N^j + tN^k$ with $S(N) \to \infty$ and $S(N)/N \to 0$ as $N \to \infty$ then (6.19) holds with $\Theta_k^*(s)$ replaced by $\Theta_k^*(\infty)$ where

$$(6.21) \qquad \mathcal{L}(\Theta_k^*(\infty)) = \mu^{j,k}_{\tau_\Theta^j}. \qquad \square$$

(b) Proof of Theorem 2

The main idea here is to reduce the assertions of the Theorem 2 to the statements which we obtained in Proposition 6.2 using that for times $t = t(N)$ not too large the process $(X^N(t))_{t\ge 0}$ can in the limit $N \to \infty$ be approximated by ℓ-level systems in which, as we recall from subsection 6 (a), the migration over distances larger than ℓ and selection with respect to φ_r^u with $r > \ell$ is suppressed.

This approximation works formally as follows. First fix a time scale $t(N)$ of at most polynomial growth. Choose ℓ such that for the given time scale $t(N)$ for *all* N large enough: $t(N) \le N^{\ell-1}$. We embed $M(\ell,N)$ in Ω_∞ by associating with $(\xi_1,\dots,\xi_\ell) \in M(\ell,N)$ the sequence $(\xi_\ell,\dots,\xi_1,0,0,\dots)$
$\in \Omega_\infty$. If we now want to investigate the behavior of $x_\xi^N(t(N))$ with $\xi \in \Omega_N \subseteq \Omega_\infty$, we simply consider the process $y_{\bar{\xi}}^{N,\ell}(t(N))$ with $\bar{\xi} = (\xi_1,\dots,\xi_\ell)$, defined in section 6(a).

Next note that if we view $(Y^{N,\ell}(t))_{t\ge 0}$ as a system with state space $(\mathcal{P}(\mathbb{N}))^{\Omega_N}$ by simply freezing all the additional coordinates, that is, those at distance more than ℓ from ξ and then compare the generators of the two martingale problems G_N^∞ (associated with $X^N(t)$) and $\tilde{G}_N^\ell$ of $Y^{N,R}(t)$ (the $\sim$ added to G_N^ℓ indicating that we extended the system beyond the active components in $M(\ell,N)$ to Ω_∞) we see that

$$(6.22) \qquad G_N^\infty - \tilde{G}_N^\ell = \frac{1}{N^\ell} B_N, \quad ||B_N(F)||_\infty \le C_F \quad \forall N \in \mathbb{N}$$

for functions $F \in \mathcal{A}$ (recall (1.12)). Therefore by denoting with $U_t^N,\ \tilde{U}_t^{N,\ell}$ the semigroups corresponding to G_N^∞ resp. $\tilde{G}_N^\ell$ we get by the formula of partial integration for all $t \ge 0$:

$$(6.23) \qquad U_t^N(F) = \tilde{U}_t^{N,\ell}(F) + \frac{1}{N^\ell}\Big(\int_0^t \tilde{U}_s^{N,\ell} B_N U_{t-s}^N ds\Big)(F).$$

Now note that $F(z) = \langle z, f\rangle^k$ is for $f \in L_\infty(\mathbb{N})$ a Lipschitz function on $\mathcal{P}(\mathbb{N})$ and the Lipschitz constant is preserved under U_t^N. This fact gives a bound on $\frac{\partial F}{\partial z}$ applied to $U_t(F)$ for $F \in \mathcal{A}$. Using this fact we derive by explicit calculation that we can even choose C_F such that

$$(6.24) \qquad ||B_N U_t^N(F)||_\infty \leq C_F.$$

We know therefore from (6.23) and B_N that $||B_N U_s^N((F))||_\infty \leq C_F \quad \forall N \in \mathbb{N},\ s \geq 0$ for $F \in \mathcal{A}$ that:

$$(6.25) \qquad ||U_{t(N)}^N(F) - \tilde{U}_{t(N)}^{N,\ell}(F)||_\infty \leq \frac{1}{N^\ell} t(N) \cdot C_F.$$

This implies for $F(z) = \langle z, f\rangle^k$ the crucial estimate (for N large enough)

$$(6.26) \qquad |E[F(X^N(t(N)))] - E[F(Y^{N,\ell}(t(N)))]| \leq C_F \cdot \frac{t(N)}{N^\ell}$$

where both processes start in the same initial state.

Since the r.h.s. of (6.26) converges to 0 as $N \to \infty$ we see that for a given time scale $t(N)$ the l-level approximation becomes arbitrarily accurate for every $F \in \mathcal{A}$. By combining (6.26) and Proposition 6.2 and using that $\mathcal{A}$ determines distributions on $(\mathcal{P}(\mathbb{N}))^{\mathbb{N}}$ it is straightforward to derive Theorem 2.

(c) Proof of Theorem 3

The Theorem 3 is an extension of the assertion (1.54) of Theorem 2 specialized to $k = 0$, to the whole collection

$$\{(x_{\xi,0}^N(S(N)\beta_j(N) + t))_{t\geq 0}\}_{\xi\in\Omega_\infty},$$

rather than for a fixed value of ξ as in Theorem 2. We want to show that for ξ and ξ' at distance at least ℓ and $t(N) \leq N^{\ell-1}$ the components evolve in the limit $N \to \infty$ *independently* so that together with the approximation property (6.26) we can again simply invoke Proposition 6.2 (6.19) and (6.20). Here the key is an asymptotic independence property, which we now state, we omit the proof since the dependence between components in the model is due to migration and the selection is here completely irrelevant so that the argument can be taken from [DGV] section 5 (c) and 5 (d).

Consider for a fixed $\ell \in \mathbb{N}$ and for $\xi, \zeta \in \Omega_\infty$ with $d(\xi,\zeta) \geq \ell$ the following two path-valued (i.e. $C([0,\infty),\ \mathcal{P}(\mathbb{N}))$-valued) random variables:

$$(6.27) \qquad X_1^N = \left(x_\xi^N(t \wedge \beta_{\ell-1}(N))\right)_{t\geq 0}, \quad X_2^N = \left(x_\zeta^N(t \wedge \beta_{\ell-1}(N))\right)_{t\geq 0}.$$

We need to show that the above random variables become asymptotically independent as $N \to \infty$ in a uniform way. To define the latter pick a injective map $\Psi : \mathbb{N} \to [0,\infty)$ and set

$$\begin{aligned}(6.28) \qquad \mathcal{F}_{K_1,K_2,n} \quad = \quad & \{h : C([0,\infty),\mathcal{P}(\mathbb{N})) \to \mathbb{R} | h(X) = f(X(t_1),\ldots,X(t_n)) \\ & \text{for some } f : (\mathcal{P}(\mathbb{N}))^n \to \mathbb{R},\ c_f \leq K_1,\ ||f||_\infty \leq K_2\},\end{aligned}$$

where c_f denotes the Lipschitz constant of f based on the chosen injective embedding of $\mathbb{N}$ into $[0,1]$ compare (Proof of Lemma 5.6).

Then by comparing the system $X^N(t)$ with a system where we put $c_k = 0$ for all $k \geq \ell$, we obtain along the same lines as in the derivation of (6.26) that:

$$(6.29) \qquad \sup_{f,g\in\mathcal{F}_{K_1,K_2,n}} |E[f(X_1^N)\ g(X_2^N)|X^N(0)] - [Ef(X_1^N|X^N(0)][Eg(X_2^N)|X^N(0)]| \underset{N\to\infty}{\longrightarrow} 0.$$

This fact used for times $S(N)\beta_j(N) + \sum_{k=j}^{l} S(N)\beta_k(N) + t$ successively from $l = 0$ up to $j-1$ together with Theorem 2, (1.54) proves Theorem 3.

7 Analysis of the interaction chain

The point of this chapter is to prove Theorems 4 - 6. In this chapter we are therefore only concerned with the object $(Z_k^j)_{k=-j-1,\dots,0}$ (interaction chain of level j) and no longer with the interacting system.

Compared with [DGV] the new aspect in the present analysis of the interaction chain, is that now

$$(7.1) \qquad Z^j_{-j-1} = \tau_j^\Theta$$

rather than $\tau_j^\Theta = \Theta$ for all $j \in \mathbb{N}$ where Θ is a fixed measure. This is a problem due to the fact that $(\tau_j^\Theta)_{j\in\mathbb{N}}$ is not always tight. On the other hand having the results for fixed τ_j^Θ this means that we can now focus in this chapter on two things, namely on the analysis of the impact of this changing initial condition on the behavior of $(Z_k^j)_{k=-j-1,\dots,0}$ and on the behavior of τ_j^Θ itself both as $j \to \infty$.

(a) Proof of Theorem 4

We begin by recalling Theorem 0.5, in [DGV] which says that we have to distinguish the two cases $\sum c_k^{-1} < \infty$ and $= \infty$ and get:

$$(7.2) \qquad \sum_{k=0}^{\infty} c_k^{-1} < \infty : \qquad \mathcal{L}((Z_k^j)_{k=-j-1,\dots,0} | Z^j_{-j-1} = \Theta) \underset{j\to\infty}{\Longrightarrow} \mathcal{L}((Z_k^\infty)_{k\in\mathbb{Z}^-})$$

where Z^∞ is an extremal entrance law and satisfies $Z_k^\infty \underset{k\to-\infty}{\longrightarrow} \Theta$ a.s., while in case

$$(7.3) \qquad \sum_{k=0}^{\infty} c_k^{-1} = +\infty : \qquad \mathcal{L}((Z_k^j)_{k=-j-1,\dots,0} | Z^j_{-j-1} = \Theta) \underset{j\to\infty}{\Longrightarrow} \int_{\mathbb{N}} \delta_{\{Y_k = u,\ k\in\mathbb{Z}^-\}} \Theta(du).$$

In order to prove our Theorem 4 we need to show that the relations (7.2) and (7.3) hold in a suitable sense *uniformly* in $\Theta \in \mathcal{P}(\mathbb{N})$ and that τ_j^Θ does in general not converge to a δ-measure as $j \to \infty$. The latter is a consequence of our analysis in the proof of Theorem 6 so we do not discuss this point here. The proof of the uniformity in Θ will be different in cases (7.2) and (7.3). The latter case is very simple and based on moment-calculations. In the case (7.2) the proof of uniformity will be achieved by constructing a *coupling* between two versions of the Markov chain with transition kernels K_{-k} at time $-k$ and two different initial states Θ and Θ'. Here K_{-k} is as defined in (1.46) and (1.37) and Θ, Θ' are elements of $\mathcal{P}(\mathbb{N})$.

We begin now with the proof in *case* $\sum c_k^{-1} < \infty$ (7.2) based on coupling. As we saw in the section 5 (f) on coupling, the construction of such a coupling works well only on compact type spaces with a *bounded* distance function. Therefore we shall here as well map our problem into one on such a state space. If we map the type space $\mathbb{N}$ onto $[0,1]$ we can also connect up with the work in [DGV]. Later we will show how to return to the initial picture.

Let ψ be any map

$$(7.4) \qquad \psi : \mathbb{N} \to [0,1],\ \psi \text{ is injective.}$$

Then $(Z_k^j)_{k=-j-1,\dots,0}$ is mapped onto a new chain. From the explicit representation of the transition kernel (see (1.48)) we read off immediately that the new chains have initial state $\psi(\Theta)$ and transition kernels $\tilde{K}_{-k}$ on $\mathcal{P}([0,1]) \times \mathcal{P}([0,1])$ given by

$$(7.5) \qquad \tilde{K}_{-k}(\varrho, d\Theta) = \tilde{\Gamma}_\varrho^{k-1}(\cdot),$$

where $\tilde{\Gamma}_\varrho^{k-1}(\cdot)$ is the distribution defined by the modified construction of (1.48), which is obtained, if we replace $\Theta \in \mathcal{P}(\mathbb{N})$ by $\psi(\Theta) \in \mathcal{P}([0,1])$.

This is exactly the transition kernel studied in [DGV] arising from renormalizing Fleming-Viot processes with migration. In the sequel we shall omit the $\sim$ on K_{-k} and Γ^k_ϱ to reduce the complexity of the notation.

We shall next derive the crucial estimate (7.9) below. For preparation we define a subclass of $C(\mathcal{P}([0,1]))$. Consider the function f on $\mathcal{P}([0,1])$ satisfying

$$(7.6) \qquad |f(\Theta_1) - f(\Theta_2)| \leq L(f)||\Theta_1 - \Theta_2||,$$

with $L(f) < \infty$, the socalled Lipschitz functions on $\mathcal{P}([0,1])$. Here $||\cdot||$ on the r.h.s. is defined as follows. For $i = 1, 2$; π_i denotes the projections of a measure π on $[0,1]^2$ onto the first and second components. Set

$$(7.7) \qquad ||\Theta_1 - \Theta_2|| = \inf\left[\int_0^1\int_0^1 |u_1 - u_2|\tilde{\Theta}(du_1, du_2);\ \tilde{\Theta} \in \mathcal{P}([0,1]^2),\ \pi_i(\tilde{\Theta}) = \Theta_i \quad i = 1, 2\right].$$

The Lipschitz-functions are dense in $C(\mathcal{P}([0,1]))$.

These ingredients allow us to formulate the following estimate:
For any given $\varepsilon > 0$ we can find a number $m = m(\varepsilon) \in \mathbb{N}$ such that if we define the two chains $\tilde{Z}^m_k$, $\tilde{\tilde{Z}}{}^{m,j}_k$ by using the same transition kernels K_{-k} but by putting the following different initial conditions:

$$(7.8) \qquad \begin{aligned} \mathcal{L}(\tilde{Z}^m_{-m-1}) &= \delta_\Theta \\ \mathcal{L}(\tilde{\tilde{Z}}{}^{m,j}_{-m-1}) &= \mathcal{L}(Z^j_{-m-1}|Z^j_{-j-1} = \Theta) \quad \text{for} \quad j > m, \end{aligned}$$

then for fixed $k \in \mathbb{Z}^-$, f as above and $j > m = m(\varepsilon)$

$$(7.9) \qquad |Ef(\tilde{Z}^m_k) - Ef(\tilde{\tilde{Z}}{}^{m,j}_k)| \leq \varepsilon.$$

Before we conclude the argument we have to construct the coupling and to use it to verify the relation (7.9). In order to construct the coupling recall that the transition mechanism of the interaction chain is the same as in the case without selection, hence we can use the nice properties of couplings of Fleming-Viot processes with immigration from source Θ_1 and Θ_2 to obtain a coupling of the equilibrium states with very good stability properties with respect to perturbations of Θ_1, Θ_2. Namely we just consider the equilibrium state of the coupled process introduced in step 2 of section 5 (f) and use the coupling estimates in this dynamical situation to obtain automatically estimates on the coupled measures $\Gamma^k_{\Theta_1}$, $\Gamma^k_{\Theta_2}$. Note that for our purposes here we can set the selection term equal to 0 in the dynamical picture. The next step below is to use the coupling estimates together with the assumption on $(c_k)_{k\in\mathbb{N}}$ to get to (7.9).

Recall we are in the case $\sum c_k^{-1} < \infty$. Then for every $\varepsilon > 0$ we can find an m such that

$$(7.10) \qquad \sum_{k=n}^{\infty} c_k^{-1} \leq \varepsilon \quad \text{for all} \quad n \geq m.$$

We know by explicit calculation using the generator (see [DGV] formula (6.1)) that for a bounded measurable function on $[0,1]$:

$$(7.11) \qquad \mathrm{Var}(\langle Z^j_k, g\rangle) = (\sum_{n=k}^{j} c_n^{-1})(\int_{\mathcal{P}([0,1])} (\mathrm{Var}_x(g))\mu^{j,k}_\Theta(dx)).$$

This implies that for the interaction chain (resp. entrance law) on $[0,1]$ and for a bounded function g on $[0,1]$:

$$(7.12) \qquad \mathrm{Var}(\langle Z^\infty_{-m}, g\rangle | Z^\infty_{-\infty} = \Theta) \leq \varepsilon \cdot (||g||_\infty)^2.$$

Recall that $E(Z^j_k|Z_{-j-1} = \Theta) = \Theta$. This implies with (7.12) that since all involved measures are atomic and in addition $\text{supp} Z^\infty_{-m} \subseteq \text{supp}\ \Theta$:

$$(7.13) \qquad E(||\Theta - Z^\infty_{-m}||_{\text{Var}})^2 \leq \varepsilon.$$

On the other hand based on the Vasherstein distance $||\cdot||$ on $\mathcal{P}([0,1])$, recall (7.7), we can define the induced Vasherstein distance $||\cdot||$ on $\mathcal{P}(\mathcal{P}([0,1]))$. With this definition the coupling result from (5.73) says using the induced coupling for the equilibria (recall (7.8) second line):

$$(7.14) \qquad ||\Gamma^{c,d}_{\Theta_1}(\cdot) - \Gamma^{c,d}_{\Theta_2}(\cdot)|| \leq ||\Theta_1 - \Theta_2||.$$

Hence every function f on $\mathcal{P}([0,1])$ which has the Lipschitz property defined in (7.6) satisfies:

$$(7.15) \qquad L(K_{-k}(f)) \leq L(f) \quad \forall k \in \mathbb{N}.$$

We get (7.9) as follows. Combine the two relations (7.13) and (7.15), apply it to the situation given in (7.8). Then use in (7.13) that the variational distance on $\mathcal{P}([0,1])$ dominates the $||\cdot||$-distance so that we can conclude from (7.15) for a fixed Lipschitz function f:

$$(7.16) \qquad E\langle Z^j_{-k}(\Theta), f\rangle \underset{j\to\infty}{\longrightarrow} E\langle Z^\infty_{-k}(\Theta), f\rangle \text{ uniformly in } \Theta,$$

where $Z^j_k(\Theta)$ denotes the interaction chain with $Z^j_{-j-1} = \Theta$.

Since the Lipschitz functions are dense we obtain the desired result (1.74) but *only on* $\mathcal{P}([0,1])$ and it remains to see that we can transfer this back to $\mathcal{P}(\mathbb{N})$. Due to the fact that all involved measures are atomic with support contained in the support of Θ and that the estimate (7.13) is in the variation norm (and hence invariant under ψ, which is injective) and (7.15) holds for every map ψ this is immediate.

Next we come to the *case* $\sum c_k^{-1} = +\infty$ (7.3). The argument in this case is based on the relation (7.11) which holds for arbitrary bounded functions g and is completely independent of the type space which we use. In particular the fact $|\langle Z^j_k, g\rangle| \leq ||g||_\infty$ implies that

$$(7.17) \qquad \int \mu^{j,k}_\Theta(dx)\text{Var}_x(g) \leq \frac{||g||_\infty}{(\sum_{n=k}^j c_n^{-1})}.$$

Hence since $\sum_{n=k}^\infty c_n^{-1} = +\infty$, $\mu^{j,k}_\Theta$ as $j \to \infty$ must concentrate uniformly in Θ (recall $\mu^{j,k}_\Theta \in \mathcal{P}(\mathcal{P}(\mathbb{N}))$) on point masses on measures concentrated on one type. This gives immediately (1.73) using that $\int \mu^{j,k}_\Theta(dx)x = \Theta$.

(b) Proof of Theorem 5

Here we first use Theorem 2 to carry out the limit $N \to \infty$ and pass to the limiting dynamics z^j_t, where $(z^j_t)_{t\geq 0}$ is a $\mathcal{P}(\mathbb{N})$-valued process defined by the $(\tilde{L}_j, \cdot)$-martingale problem (see (1.41)). For the limiting dynamics we now have to study the rate of decay of the expected mass of a type u which does not have maximal fitness if measured by φ^u_j. Hence we derive next a differential equation for the quantity $E(z^j_t(u))$, which we then use to get lower and upper bounds on the rate of decay.

Applying the martingale problem defining $(z^j_t)_{t\geq 0}$ results after an explicit calculation using (1.41) in the following differential equation

$$(7.18) \qquad \frac{d}{dt}E(z^j_t(\{u\})) = sd_jE\Big[z^j_t(\{u\})(\varphi^u_j - \sum_v z^j_t(\{v\})\varphi^v_j)\Big].$$

Now suppose that u satisfies

$$(7.19) \qquad \varphi_j^u < \max_w \varphi_j^w =: \hat{\varphi}_j.$$

We introduce now the following abbreviations in order to rewrite this equation:

$$(7.20) \qquad f_t^j = Ez_t^j(u) \quad g_t^j = E(z_t^j(u))^2.$$

Furthermore abbreviate by v the type with maximal fitness and define

$$(7.21) \qquad z_t^* = \sum_{w \neq u,v} z_t^j(\{w\})(\hat{\varphi}_j - \varphi_j^w)$$

and

$$(7.22) \qquad h_t^j = E(z_t^* z_t^j(u)).$$

Then we can rewrite (7.18) as follows:

$$(7.23) \qquad \frac{d}{dt} f_t^j = \left\{-d_j(\hat{\varphi}_j - \varphi_j^u)s\right\} f_t^j + \left\{d_j(\hat{\varphi}_j - \varphi_j^u)s\right\} g_t^j + \{d_j s\}\, h_t^j.$$

This allows us to obtain the following *lower bound* on the rate of decay. We have

$$(7.24) \qquad \frac{d}{dt} f_t^j \geq \left\{-sd_j(\hat{\varphi}_j - \varphi_j^u)\right\} f_t^j$$

and hence since $f_t^j \geq f_0^j \exp(-sd_j(\hat{\varphi}_j - \varphi_j^u)t)$ we conclude

$$(7.25) \qquad \liminf_{t \to \infty} \left(\frac{1}{t} \log f_t^j\right) \geq -d_j(\hat{\varphi}_j - \varphi_j^u)s.$$

This already implies that in the case $\sum c_k^{-1} = +\infty$

$$(7.26) \qquad \inf_{\{u\} \in I_j} \left\{ \liminf_{t \to \infty} \lim_{N \to \infty} \frac{1}{t} \log E\big[x_{\xi,j}^N(tN^j)(u)\big] \right\} \longrightarrow_{j \to \infty} \quad 0.$$

Note that in this case d_j converges to 0 as $j \to \infty$, since the recursion formula for d_k in (1.36) has the explicit solution:

$$(7.27) \qquad d_k = \frac{d_0}{1 + d_0 \sum_{j=0}^{k-1} c_j^{-1}}.$$

In the duality analyses below we shall see that for $s \leq 1 (= d_0)$

$$(7.28) \qquad \liminf_{t \to \infty} (g_t^j / f_t^j) \geq C \frac{d_j}{2c_j + (s+1)d_j}.$$

Inserting this in (7.23) gives

$$(7.29) \qquad \frac{d}{dt} f_t^j \geq sd_j(\widehat{\varphi}_j - \varphi_j^u)\Big(-1 + C\frac{d_j}{2c_j + (s+1)d_j}\Big) f_t^j$$

which proves that

$$(7.30) \qquad \lambda_j^{\pm}(u) > sd_j(\widehat{\varphi}_j - \varphi_j^u).$$

Note that $(\lambda_j^{\pm}(u) - sd_j(\widehat{\varphi}_j - \varphi_j^u)) \to 0$ as $j \to \infty$, for both cases $\sum c_k^{-1} < \infty$ and $= +\infty$ if in addition c_k^{-1} decays slower than exponential since either $d_j \to 0$ or $c_j \to \infty$ as $j \to \infty$.

The *upper bound* on the decay rate is more involved. Notice that by Jensen and by $z_t^j(u) \leq 1$:

(7.31) $$(f_t^j)^2 \leq g_t^j \leq f_t^j$$

and furthermore the upper bound is asymptotically attained if the random variable $z_t^j(u)$ takes for large t only values that are either close to 0 or close to 1. It turns out that this is actually happening and that indeed $f_t^j \cong a g_t^j$ as $t \to \infty$ with some $a \in (0,1)$. Hence we really have to analyze the quantity g_t^j in detail, the key tool being a duality relation, (recall (3.8).

We turn now to the upper bound for the decay rate, which we need in order to show that in the case $\sum c_k^{-1} < \infty$ the decay rate stays bounded away from 0. We shall for this case prove the inequality:

(7.32) $$\limsup_{t\to\infty}(g_t^j/f_t^j) \leq D\frac{d_j(1+s)}{2c_j+d_j(1-s)}$$

where D is independent of j.

We can conclude from (7.32) that

(7.33) $$\frac{d}{dt}f_t^j \leq -\left\{(\widehat{\varphi}_j-\varphi_j^u)sd_j\left[1-D\frac{d_j(1+s)}{2c_j+d_j(1-s)}\right]^+\right\} f_t^j + \{d_j s\}h_t^j.$$

In the case where $\sum_0^\infty c_k^{-1} < \infty$ we must have that

(7.34) $$c_j \to \infty \quad \text{and} \quad d_j \to d_\infty > 0 \quad \text{as } j \to \infty.$$

Note that in addition $h_t^j/f_t^j(\hat{\varphi}_j - \varphi_j^u)$ is bounded away from 1 if the type u has minimal fitness and hence:

(7.35) $$\limsup_{j\to\infty}\left\{\sup_{\{u\}\in I_j}\limsup_{t\to\infty}\lim_{N\to\infty}\left[\frac{1}{t}\log E(x_{\xi,j}^N(tN^j)(u))\right]\right\} < 0,$$

which completes the proof of the theorem once we have established the two relations (7.28) and (7.32) and for that we need the dual process. We have to extend our construction of the dual process for the finite system (5.3) to the system appearing as the McKean-Vlasov limit for $N \to \infty$, which is a nonlinear Markov process recall (5.4). Hence it is the easiest approach to actually think of an $\mathbb{N}$-indexed collection of independent versions of that nonlinear Markov process. Then in the definition of the dual process given in (3.3) we only have to modify the migration mechanism of the dual as follows:

(7.36) for every particle jumps occur at rate c_j to the free site of the smallest index.

To see that this way we obtain a dual process with respect to the function, we can either do a generator calculation or we can consider the limit of the dual process of the finite system (5.3) in the limit $N \to \infty$. We refer for details to the section on dual processes.

We return to the proof of (7.28) and (7.32). Note that we obtain $Ez_t^j(u)$ and $E(z_t^j(u))^2$ from the function valued dual process, if we choose as initial functions to start the dual process:

(7.37) $$f_1(v) = 1_{\{u\}}(v) \qquad f_2(v,w) = 1_{\{u\}}(v)1_{\{u\}}(w)$$

and if we start the driving particle system with one respectively two particles at the site 1. (Recall the site space is now $\mathbb{N}$). Note that we have in the two particle situation by putting $A = \{$a coalescence occurs as first event$\}$ and $B = \{$a jump occurs at first event$\}$:

(7.38) $$P(A) = \frac{d_j}{2c_j+(s+1)d_j} \qquad P(B) = \frac{2c_j}{2c_j+d_j(s+1)}.$$

On the event B the dual particle system will never have less than two particles, in fact if we denote with T the jump time truncated at t we can say that the expectation of the duality function at time t on the event B is given by

$$E\Big[\Big(e^{2sd_jT}\mathbb{I}_B\Big)\Big(f^j_{t-T}\Big)^2\Big] \tag{7.39}$$

while on the event A it is given with U denoting the coalescing time:

$$E\Big(e^{2sd_jU}\mathbb{I}_A f^j_{t-U}\Big). \tag{7.40}$$

Now for $s \leq 1$ we will have to show that the convolution given above has the correct asymptotics, such that the terms $(f^j_t)^2$ resp. f^j_t determine the decay of the resulting expression in t up to order one.

First it is easy to see (7.28) since by the duality relation a lower bound for g^j_t is given by (7.40). This expression can be written as:

$$\frac{d_j}{2c_j+(1+s)d_j}\int\limits_0^t \big\{(2c_j+(1+s)d_j)\exp(-u(2c_j+(1-s)d_j))\big\}\big\{f_{t-u}\big\}du \tag{7.41}$$

which together with the fact that f_t decays with at most exponential rate gives the desired bound.

The inequality (7.32) is more subtle. Here we note first that the expression in (7.39) is $o(f^j_t)$ as $t \to \infty$. Hence it suffices in order to estimate the l.h.s. of (7.32). to consider the expression in (7.40) and the corresponding expression for the event $(A \cup B)^c$, when a new particle is created as the first event of the dual process.

We now simply estimate the contribution on the latter two events, namely coalescence first. This contribution is using that $E(z^j_t(u))^3 \leq f^j_t$ less or equal to

$$\begin{aligned} &\int\limits_0^t \big\{e^{2sd_ju}(d_j(1+s))e^{-d_j(1+s)u}e^{-2c_ju}\big\}\big\{f_{t-u}\big\}du \\ &= \frac{d_j(1+s)}{2c_j+d_j(1-s)}\int \big\{(2c_j+d_j(1-s))\; e^{-(d_j(1-s)+2c_j)u}\big\}\big\{f_{t-u}\big\}du. \end{aligned} \tag{7.42}$$

This implies the assertion (7.32) using $s < 1$ and the fact that $c_j \to \infty$ and $d_j \to d_\infty$ in the regime where $\sum c_k^{-1} < \infty$.

(c) Proof of Theorem 6

(a) We begin with a deterministic initial state, where every weight is a multiple of the number δ. Now we introduce additional types such that every type has weight 0 or δ. This is achieved by splitting a type u with weight $k\delta$ initially in k subtypes with the same fitness. Consider first this model.

The relation (1.88) follows from the fact that the dynamics in (1.42) has with this special initial states the exchangeability property that all the types u with $\varphi^u_j = \max_v \varphi^v_j$ give the same value in the terms $E(Y_t(u)Y_t(v))$ and consequently in the limit we obtain the same expectation for every type.

Observe that splitting the mass of types at time 0 into artificial subtypes, running the dynamics and later recombining them again gives the same distribution. This is a basic property of the Fleming-Viot process and can for example be read off from the dual process (compare [DGV], section 3 (a)). Recombining the subtypes then gives the assertion for the original model. Then a limit procedure gives the assertion in the general situation, or random initial states.

(b) Since the maximum of the fitness function is not attained, the series defining $\chi(u)$ is less than the supremum and hence there exists a type u with $\Theta(u) > 0$ such that the expansion up to the j-th value exceeds $\chi(u)$. Hence by (4.4) every fixed type u will disappear in some time scale $\beta_j(N)$ and hence for some $j, \tau_j^{\Theta}(\{u\}) = 0$ holds. This implies then the assertion of part (b).

(c) Note first that under the size-ordering and under $\prod_{\mathcal{N}_j}$ the order between the remaining types is always preserved under the assumptions of the theorem. In fact, at each step the labels of the remaining types are changed according to a map of the form

$$u \to u - m$$

where m depends on the index j of the step. Now the assertion is immediate from the hypotheses (1.87).

8 The effects of mutation

In this chapter we prove Theorems 8 - 10, which deal with the role of mutation and recombination. The main tools will be an extension of the duality relation of subsection 3(a), a modified coupling extending the one from subsection 5(e) and finally the Girsanov formula for measure valued processes.

(a) Proof of Theorem 8.

Part (a) **(i)** The proof proceeds by tightness arguments for the existence of an equilibrium and with coupling and Girsanov formulas for the uniqueness.

We first establish the *existence* of an equilibrium. Under the restriction to the case of a K-level system with $K < \infty$ we have a *finite* number of different types. Then our system is a Markov process on a compact state space. Hence the *existence* of an invariant measure can be proved by the following standard argument. Since the sequence $\{\mathcal{L}(X^{N,K}(t))\}_{t \geq 0}$ is *tight* in $\mathcal{P}((\mathcal{P}(E_K))^{\Omega_N^K})$ we can choose a sequence $t_k \to \infty$ such that

$$\text{(8.1)} \qquad \frac{1}{t_k} \int_0^{t_k} X^{N,K}(t)dt \underset{k \to \infty}{\longrightarrow} X(\infty) \quad \text{weakly,}$$

with a random measure $X(\infty)$ on $(E_K)^{\Omega_N^K}$ - it is then easy to show using the Feller property of $X^{N,K}(t)$ that $\mathcal{L}(X(\infty))$ is an invariant measure.

The *uniqueness* of the stationary measure (on the reduced type space) is obtained by modifying the argument of Ethier and Kurtz (1996, Theorem 5.3) for the case in which there is only one site. Their argument is based on two basic steps - the first is the construction of a *successful coupling* for the process without selection, the second the use of *Girsanov's formula* to include selection.

Step 1 We begin with the *coupling* of two processes starting in different initial states. The key observation is that under the given assumptions we can write

$$\text{(8.2)} \qquad M(u, dv) = \bar{m}\rho(dv) + (1 - \bar{m})\bar{M}(u, v)$$

where ρ is counting measure and $\bar{m} > 0$. Consider first the finite neutral K-level model with recombination and mutation, that is, put $s = 0$ so that *no selection* occurs.

Forget for the moment the migration. The idea is to construct a new collection $\widehat{X}_t^{N,K}$ now of *two component* processes, each component representing in every colony one of the two processes to be coupled. In order to carry this out we use as type space $E_K \times E_K$ at each site instead of E_K and simply run the resampling dynamic on that new type space and use a new recombination kernel and new mutation rates so that each of the two marginal processes have the original dynamic. We have to specify this new mutation and recombination term. This is done as follows.

To define the recombination kernel on $E_K \times E_K$ we distinguish the case of points on the diagonal and off the diagonal. (i) On the diagonal there is simultaneous recombination of both components with other types on the diagonal according to the given kernel. (ii) Off the diagonal the two components undergo recombination independently of each other.

Similarly we proceed with the mutation rates. (i) On the diagonal mutation occurs simultaneously in both components. (ii) Off the diagonal we use the decomposition in (2.3) to define two transitions namely we mutate *simultaneously* to the same new type with the $\bar{m}\varrho$ part of the kernel $M(\cdot,\cdot)$ and *independently* in the two components with the $(1-\bar{m})\bar{M}$ part.

We have to now introduce the migration term. For the model with migration and resampling the construction of a coupled process with state space $(E_K \times E_K)^{\Omega_N}$ including migration is obtained in [DGV] subsection 4 c). This carries over to the present case. (Compare section 5(f) step 2 for more details of such a construction).

This process just described is *well-defined* by Theorem 7. We note that the coupled process has the property that when a point on the diagonal migrates it stays on the diagonal.

This coupling is *successful,* i.e. the system ends up with all subpopulations concentrated on the diagonal:

$$P(x^{N,K}_\xi(t)(\{(u,v)|u,v \in E_K,\ u=v\}) \underset{t\to\infty}{\longrightarrow} 1. \tag{8.3}$$

This is because jumps onto the diagonal occur at a *minimal* rate $\bar{m} > 0$ as defined in (2.3), which is independent of the state of the system.

Step 2 We next apply *Girsanov's formula* to processes on $(E_K \times E_K)^{\Omega_N}$ in order to get uniqueness in the *presence of selection.* The Girsanov formula for the finite system allows to find the density of the law of the system with selection with respect to a system without selection. We shall consider two such systems. For that purpose we introduce two fitness functions on $(E_K)^2$:

$$\varphi_i(u_1,u_2) := \varphi(u_i), \quad i=1,2. \tag{8.4}$$

The respective Radon-Nikodym derivatives can be calculated explicitly and are given in (8.6) (the proof of [D], Theorem 7.2.2 can be modified to cover the finite stepping stone model considered here).

Note that the original (i.e. $(E_K)^{\Omega_N}$-valued system) has the property that the component processes $x_\xi(t)$ are semi-martingales and their martingale parts denoted $m_\xi(t)$ can be calculated as

$$\begin{aligned} m_\xi(t) &= x_\xi(t) - x_\xi(0) - \Bigg\{ \int_0^t c \sum_{\xi'} a(\xi,\xi')(x_{\xi'}(s) - x_\xi(s))ds \\ &\quad + m \int_0^t \left[\int_{\mathbb{N}} M(u,\cdot) x_\xi(s)(du) - x_\xi(s) \right] ds \\ &\quad + r \int_0^t \left[\int_{\mathbb{N}\times\mathbb{N}} \alpha(u_1,u_2) R(u_1,u_2;\cdot)(x_\xi(s) \otimes x_\xi(s))(du_1,du_2) - x_\xi(s) \right] ds \Bigg\}. \end{aligned} \tag{8.5}$$

Then if we associate with the coupled process $(\hat{x}_\xi(t))_{\xi\in\Omega_N}$, which is a $(E_K \times E_K)^{\Omega_N}$-valued, the process $\bar{x}_\xi(t) = (\bar{x}^1_\xi(t),\ \bar{x}^2_\xi(t))$ given by $\hat{x}_\xi(t)(\{u\} \times E_K)$ resp. $\hat{x}_\xi(t)(E_K \times \{u\})$, we obtain versions of the two original processes without selection and then we can define for $i = 1,2$:

$$\begin{aligned} \mathcal{R}_i(t) : &= \exp \Bigg\{ \sum_{\xi: d(0,\xi)\le K} \frac{s}{d} \left[\int_0^t \int_{\mathbb{N}\times\mathbb{N}} \varphi_i(u) \bar{x}^i_\xi(s)(du) \right] \bar{m}^i_\xi(ds) \\ &\quad - \frac{s^2}{2d} \int_0^t \int_{\mathbb{N}\times\mathbb{N}} \int_{\mathbb{N}\times\mathbb{N}} \varphi_i(u_1)\varphi_i(u_2) Q_{\bar{x}^i_\xi(s)}\,(du_1,du_2)(ds) \Bigg\}. \end{aligned} \tag{8.6}$$

These functions are according to the Girsanov theorem for measure-valued processes the Radon-Nikodym derivative of the model with selection φ_i with respect to the neutral model. See [D], Theorem 7.2.2.

Note that with $\mathcal{R}_1(t)$ and $\mathcal{R}_2(t)$ we can easily calculate the Radon Nikodym derivative of the $(E_K \times E_K)^{\Omega_N}$-valued process with selection with respect to the process with resampling, mutation and recombination as follows. The fitness of a pair is obtained by simply replacing the expression φ_i in (8.6) by the fitness function $\tilde{\varphi} = \varphi_1 1_D + (\varphi_1 + \varphi_2) 1_{D^c}$ where $D = \{(u,v) \in [0,1]^2 | u = v\}$. We then write out the quadratic expression to check that this density has the form of an exponential again.

This new fitness function $\tilde{\varphi}$ ensures that $(\bar{x}^1_\xi(t))_{\xi \in \Omega_N}$ and $(\bar{x}^2_\xi(t))_{\xi \in \Omega_N}$ are versions of the processes started in the two given initial distributions and *with* selection associated with the fitness function φ. This follows since it is of the form $\varphi_1(u)+$ const for v fixed and u varying.

The proof of the uniqueness of the invariant measure with selection then follows by the analogue of the proof of Ethier and Kurtz Theorem 5.3. By Lemma 5.4. of [EK5] we know that either the original system is ergodic or else there exist two mutually singular invariant measures.

Now assume that the coupled process starts in the product of two measures such that two coordinates start in two extreme invariant (and therefore mutually singular measures). But then we can show that for the dynamics without selection the two marginals after a positive time are not mutually singular using the coupling result for the coupled process without selection derived above in step 1. Then with positive probability there is mass on the diagonal also for the process with selection as can be seen by deriving the density for the fitness function $\tilde{\varphi}$ with respect to the neutral model from (8.6). Then the two extreme invariant measures are not mutually singular, yielding a contradiction. This implies that there is a unique invariant measure and hence completes the proof of Theorem 8 a(i).

(ii) The proof of (ii) by duality is a special case of Part (b) but under the additional assumption that $\bar{m}$ is sufficiently large. The proof under the assumption $\bar{m} > 0$ follows using the ideas in (i) above and will now be outlined. We can use the construction in the proof of (i) to show that in time $[0, T]$, $T > 0$ the event that all components of the coupled process are on the diagonal at time t has positive probability (use the boundedness of the density w.r.t. the neutral law) bounded below independent of the initial condition (recall that φ is bounded above and the number of components finite).

We can then use the "Ney-Nummelin construction" (cf. [EK5], Lemma 5.1 and Prop. 5.2) to build a coupled version with one component starting from an invariant measure and the other an arbitrary measure and prove convergence to the unique equilibrium. The idea is that we can construct a coupled process where the component jumps to the diagonal with positive probability within a time interval of fixed length T, throughout the evolution.

(iii) In case of a one site system this follows first by observing that with no selection this is exactly the case in which the GEM distribution (compare (1.48) for a definition) gives the unique equilibrium measure (cf. [DGV]) (and the equilibrium system is reversible). The invariant measure with added selection is obtained in Overbeck, Röckner, and Schmuland (section 5.3).

Part (b) Now the process we deal with has infinitely many components, so that the argument in (a) does not carry through and we resort to the tool of duality. We shall show that (recall (3.1)) $F_t((\hat{\xi}, f), X^{N,K})$ converges as $t \to \infty$ to a limit which does not depend on $X^{N,K}$, which implies that $\mathcal{L}(X^{N,K}(t))$ converges to a *unique* limit law. The convergence of F_t is now obtained via duality.

We need a *modification* of the *dual process* introduced in section 3(a). This modified dual process can be defined if the mutation rate $M(u,v)$ is bounded below by $\bar{m}\rho > 0$, where ρ is a strictly positive measure. Then we can write $M(u,v)$ in the form

(8.7) $$M(u,v) = \bar{m}\rho(v) + \bar{M}(u,v)$$

and we change the dynamics of the Feynman-Kac dual of section 3 ((3.3) - (3.6)) as follows:

(α) In the mechanism describing the dual process $\mathcal{F}_t$ replace the mutation semigroup M_t by $\bar{M}_t$ (M replaced by $\bar{M}$).

(β) Add to the dynamics of $\widehat{\eta}_t$ the *death* of a partition at rate $\bar{m}$.

(γ) Add to the dynamics of $\mathcal{F}_t$: at the time of such a death of a partition in the driving process $\widehat{\eta}_t$ replace at the same time $\mathcal{F}$ by a new function obtained by integrating the variables corresponding to that partition with respect to the measure ϱ of (8.7) on the type space.

According to [D] chapter 5.5 we still obtain the relation (3.8) for the modified system provided that we establish the exponential tightness (compare (3.10) and the sequel).

Assume for the moment the exponential tightness. It then remains to prove that the system goes to extinction since then the dual process comes to a stop and the limit does not depend on $X^{N,K}(0)$, since now all integrals are with respect to ϱ.

Let us first look at the case of finitely many components. In the finite case we know since $d_0 > 0$ and $\bar{m} > 0$ (and hence the *quadratic part* of the death rate makes the system *subcritical* for large populations per site) that the system $\hat{\eta}_t$ satisfies the condition: there exists a random time τ such that

(8.8) $$|\pi_t| = 0 \quad \text{for} \quad t \geq \tau, \quad \text{Prob}(\tau < \infty) = 1.$$

This immediately implies that with the notation of section 3(a).

(8.9) $$F_t\Big((\hat{\xi}, f), X^{N,K}\Big) \underset{t\to\infty}{\longrightarrow} F_\infty\Big((\hat{\xi}, f), X^{N,K}\Big) = F_\infty\Big((\hat{\xi}, f),\ \varrho^{\otimes K}\Big).$$

In the infinite case extinction will occur provided that $\bar{m}$ is sufficiently large so that the branching particle system (excluding the quadratic death rate) associated with the infinite system is *subcritical* which is the case if $\bar{m}$ is sufficiently large.

All that remains is to show the Feynman-Kac dual exists, that is we have *exponential tightness* for which we need to show that for all $t \in [0, \infty]$:

(8.10) $$E\Bigg(\exp\bigg(s\int_0^t |\pi_s| ds\bigg)\Bigg) < \infty.$$

Note that in contrast to section 3 we need the relation (8.10) even for $t = \infty$.

We proceed to prove (8.10). Define $\lambda(t) = E(\exp(s\int_0^t |\pi_s| ds))$. We first observe that $|\pi_t|$ is stochastically dominated by the branching particle system B_t obtained by ignoring the quadratic death term. Define $\tilde{\lambda}(t) = E\exp(s\int_0^t |B_s| ds)$. Then (cf. (3.13)):

(8.11) $$\begin{aligned} \frac{d\tilde{\lambda}}{dt} &= \Big[(s+r)\tilde{\lambda}^2 - (\bar{m}+s+r)\tilde{\lambda}\Big] + \bar{m} + s\tilde{\lambda} = (s+r)\tilde{\lambda}^2 - (\bar{m}+r)\tilde{\lambda} + \bar{m} \\ \tilde{\lambda}(0) &= 1. \end{aligned}$$

The quadratic expression in (8.11) should have a stable root to the right of 1 in order for $\tilde{\lambda}(t)$ to converge. One checks that this is the case for sufficiently large $\bar{m}$ (specifically $\bar{m} \geq 2s + r$ and in addition $\bar{m}$ must be large enough to satisfy $(\bar{m}+r)^2 > 4(s+r)\bar{m}$). Hence for sufficiently large $\bar{m}$ and fixed selection rate $s > 0$ and recombination rate $r \geq 0$:

(8.12) $$\lim_{t\to\infty} \tilde{\lambda}(t) = \tilde{\lambda}(\infty, s, r, \bar{m}) < \infty, \quad \tilde{\lambda}(\infty, s, r, \bar{m}) \to 1 \text{ as } \bar{m} \to \infty.$$

This completes the argument.

Remark Note that a branching random walk with local quadratic death rate is analogous to the Mueller-Tribe system (or contact process) which has a phase transition between extinction and nonextinction starting with finite initial conditions which suggests that the above argument may fail for large selection rates s. However in the special case of two types without recombination, Shiga and Uchiyama get the result (b) with no additional assumption on $\bar{m}$. (This involves the observation that this system can be obtained by the "normalization" of the linear (that is branching) system. (see 11.23).)

(b) Proof of Theorem 9

We start with part a) of the theorem. We will outline two proofs one using coupling which we give in all detail and then one using duality. The latter generalizes to prove the Corollary to Theorem 9.
Part (a) Coupling proof.
Step 1 We first observe that to prove uniqueness it suffices to prove the uniqueness of the mean curve $m(t) = E(x^*_\xi(t))$ - this is because if we insert this mean value in the generator we obtain the generator of the linear martingale problem and there is a unique solution to this linear martingale problem exactly as in the proof of Theorem 7. The proof of the uniqueness of the mean curve will again be based on a combination of *coupling* techniques for models without selection and *Girsanov's formula* to incorporate selection.

Step 2 Now assume that there are actually two solutions for the mean curve $m_1(t), m_2(t)$ and let $X_1(t),\ X_2(t)$ be the solutions of the linear martingale problems with sources $m_1(t),\ m_2(t)$, respectively. Since by assumption φ is bounded and types are characterized by their fitness we can again embed our type space $\mathbb{N}$ in $[0,1]$. We now work on this type space for convenience. We shall need actually two couplings, one of the driving sources $m_i(t)$ which are elements of $\mathcal{P}([0,1])$ and then based on that coupling of the driving terms, a coupling of the laws of our processes, which are elements of $\mathcal{P}(\mathcal{P}([0,1]))$.

We next construct this coupled version of the two processes. To do this we first couple the $\mathcal{P}([0,1])\times\mathcal{P}([0,1])$-valued pair $m_1(t), m_2(t)$ as in (5.55) and (5.56). The second step is to form a coupled version of the two solutions to the linear *neutral* $(s = r = m = 0)$ martingale problems with sources $m_1(t), m_2(t))$ by simply considering the dynamics of a system on $[0,1]\times[0,1]$ with resampling and immigration and emigration where the immigration source at time t is given by the coupling of $m_1(t), m_2(t)$. The details of this construction are in [DGV], section 4 (c).

We must still incorporate mutation and recombination in the coupling. Here we use the construction in subsection (a) part (i) of this chapter 8 and end up with coupling of the processes leaving out only selection. The law and expectation of the coupled process are denoted by $P_{1,2}$ and $E_{1,2}$ and the component processes are called $X^1(t)$ and $X^2(t)$.

In order to exploit this coupling we want to derive an analogue of Lemma 5.12 in chapter 5, which estimates $||X_1(t) - X_2(t)||$ (here $||\cdot||$ denotes the Vasherstein metric, see (5.55) for a definition) in terms of the corresponding difference in the driving terms $||m_1(t) - m_2(t)||$.

Assume that the number of types is finite. We shall show that there exists a constant $C = C(\alpha, R, M)$ $(\alpha, R, M$ are the recombination and mutation matrices) such that

$$(8.13)\qquad ||X_1(t) - X_2(t)|| \le C\int_0^t ||m_1(s) - m_2(s)||ds.$$

Denote the coupled process by $\widehat{X}(t)$ and the coupled driving terms by $\hat{m}(t)$. (Compare 5(f), step 2). Similarly we denote by $\widehat{M}, \hat{\alpha}, \widehat{R}$ the mutation and recombination matrices of the coupled process.

Furthermore since we have only a finite number of types, say L we can use as type space the interval $[0,1]$ by assigning the points $0, L^{-1}, 2L^{-1}, \ldots, (L-1)L^{-1}$ to the L different types.

Then define

$$h(t) = E \int_{[0,1]^2} |u_1 - u_2| \widehat{X}_t(du_1, du_2) \tag{8.14}$$

$$g(t) = E \int_{[0,1]^2} |u_1 - u_2| \hat{m}_t(du_1, du_2) \tag{8.15}$$

$$f_1(t) = E \int_{[0,1]^2} \left\{ \int_{[0,1]^2} \widehat{M}((u_1, u_2), (dv_1, dv_2))|v_1 - v_2| - |u_1 - u_2| \right\} \widehat{X}_t(du_1, du_2) \tag{8.16}$$

and with the abbreviation $\bar{u} = (u_1, u_2),\ \bar{v} = (v_1, v_2),\ \bar{w} = (w_1, w_2)$:

$$\begin{aligned} f_2(t) \quad = \quad & E\Big[\int_{[0,1]^2} \int_{[0,1]^2} \Big\{ \int_{[0,1]^2} \hat{\alpha}(\bar{v}, \bar{w}) \widehat{R}(\bar{v}, \bar{w}; d\bar{u}) |u_1 - u_2| \widehat{X}_t(d\bar{v}) \widehat{X}_t(d\bar{w}) \\ & - \hat{\alpha}(\bar{v}, \bar{w}) |v_1 - v_2| \widehat{X}_t(d\bar{v}) \widehat{X}_t(d\bar{w}) \Big\} \Big]. \end{aligned} \tag{8.17}$$

We shall show now that

$$h(t) = h(0)e^{-t} + \int_0^t e^{-(t-s)}(g(s) + f_1(s) + f_2(s))ds. \tag{8.18}$$

To see this consider the function $H(\bar{X})$ on $\mathcal{P}([0,1]^2)$ given by:

$$H(\widehat{X}) = \int |u_1 - u_2| \widehat{X}(du_1, du_2). \tag{8.19}$$

Then

$$\frac{\partial H}{\partial \widehat{X}}(u_1, u_2) = |u_1 - u_2|, \qquad \frac{\partial^2 H}{\partial \widehat{X}^2} \equiv 0. \tag{8.20}$$

Inserting the above relation into the defining martingale problem gives the differential equation $\dot{h}(t) = f_1(t) + f_2(t) + g(t) - h(t)$ which has the solution given in (8.18).

We already have

$$g(t) \le ||m_1(t) - m_2(t)|| \tag{8.21}$$

by the definition of the Vasherstein metric. The next step is now to bound $f_i(t)$ $i = 1, 2$ in a similar way. We begin with $f_1(t)$. Note that:

$$E(\widehat{X}(t)) = \hat{m}(t). \tag{8.22}$$

Therefore the bound

$$\left| \int_{[0,1]^2} \widehat{M}(\bar{u}, d\bar{v}) |v_1 - v_2| - |u_1 - u_2| \right| \le C_M |u_1 - u_1| \tag{8.23}$$

implies that

$$f_1(t) \le C_M ||m_1(t) - m_2(t)||. \tag{8.24}$$

However on a finite type space (8.23) follows from the fact that by construction:

$$\widehat{M}(\bar{u}, \{\bar{v} : v_1 \neq v_2\}) = 0 \quad \text{if } u_1 = u_2. \tag{8.25}$$

Next we come to $f_2(t)$. Now the problem is that the second moment measure $E(\widehat{X}_t(d\bar{v})\widehat{X}_t(d\bar{w}))$ appears. We treat first the term

$$\left| \int_{[0,1]^2} \int_{[0,1]^2} \alpha(\bar{v}, \bar{w}) |v_1 - v_2| \widehat{X}(d\bar{v}) \widehat{X}(d\bar{w}) \right| \le C_\alpha \int_{[0,1]} |v_1 - v_2| \widehat{X}(d\bar{v}) \tag{8.26}$$

where

$$C_\alpha = \sup_{\bar{v}} \int_{[0,1]} \alpha(\bar{v}, \bar{w}) d\bar{w}. \tag{8.27}$$

For the other term use that since the recombination kernel satisfies

$$\widehat{R}(\bar{v}, \bar{w};\ \{u | u_1 \neq u_2\}) = 0 \quad \text{if } v_1 = v_2 \text{ and } w_1 = w_2$$

we know that (recall we have finitely many types!)

$$\int_{[0,1]^2} \widehat{R}(\bar{v}, \bar{w}; d\bar{u}) |u_1 - u_2| \le C_R(|v_1 - v_2| + |w_1 - w_2|) \tag{8.28}$$

and hence

$$\int \int \alpha(\bar{v}, \bar{w}) \int \widehat{R}(\bar{v}, \bar{w}; d\bar{u}) |u_1 - u_2| \widehat{X}(d\bar{v}) \widehat{X}(d\bar{w}) \le 2 C_\alpha C_R \int_{[0,1]^2} |v_1 - v_2| \widehat{X}(d\bar{v}). \tag{8.29}$$

Now combine (8.29), (8.26), to get

$$f_2(t) \le (C_\alpha + 2 C_\alpha C_R) ||m_1(t) - m_2(t)||. \tag{8.30}$$

Now define $C = C_M + C_\alpha + 2C_\alpha C_R + 1$ to get from (8.30), (8.23) and (8.21):

$$f(s) + f_1(s) + f_2(s) \le C ||m_1(s) - m_2(s)||. \tag{8.31}$$

Inserting this relation in (8.18) yields the claim (8.13). We remark that the proof immediately generalizes to countably many types if we assume that M, α, R satisfy the conditions:

$$\sup_{X \in \mathcal{P}(\mathbb{N})} \sup_{v} \sum_{w} \alpha(v, w) X(w) < \infty \tag{8.32}$$

(8.33) M and R satisfy (8.23) and (8.28).

Finally it remains to deal with the selection term which we do in the next step.

Step 3 We express the law of the solution of the martingale problem with selection in terms of the martingale problem without selection for each of the component processes of the coupled process from step 2, which were denoted X_1, X_2. This is done using Girsanov's formula (compare (8.6)). Consider a function ϕ on the type space $[0, 1]$, which is Lipschitz with $L(\phi) \le 1$. Note that in particular the fitness of a type is such a functional in our present context.

Let $m_i(t,\phi) := \int \phi(u) m_i(t,du)$. Let $R(t,m_i,X_i)$ $i=1,2$ be the Radon-Nikodym derivatives of the process on $\mathcal{P}[0,1]$ with selection with respect to the one with $s=0$, based on $m_1(t)$ and $m_2(t)$.

Our goal is to apply Gronwall's lemma to $|m_1(t)-m_2(t)|$ by using the coupling in the neutral case combined with Girsanov's formula. Given a function ϕ on $[0,1]$ we write $X(t,\phi)$ for $\int X(t)(du)\phi(u)$. Then we obtain (recall that $E_{1,2}$ is expectation with respect to the coupled process *without selection*, i.e. $s=0$ and E is expectation of the process *with selection*) using the definition of the mean curves $m_i(t)$:

$$(8.34)\qquad \begin{aligned} |m_1(t,\phi)-m_2(t,\phi)| &= |E(X_1(t,\phi)) - E(X_2(t,\phi))| \\ &= |E_{1,2}[\mathcal{R}_1(t,m_1,X_1)X_1(t,\phi)] - E_{1,2}[\mathcal{R}_2(t,m_2,X_2)X_2(t,\phi)]|. \end{aligned}$$

Given the fitness function φ, the mean curve m and path $X \in C([0,T],\mathcal{P}(\mathbb{N}))$ the Radon-Nikodym derivative of the process with selection with respect to the ones without selection are given by the following device.

The process $X(t)$ is (for given mean curve) a semi-martingale and its martingale part denoted $\mathcal{M}(t)$ has the form:

$$(8.35)\qquad \mathcal{M}(t) = X(t) - X(0) - \int_0^t g(X(s))ds - c_0 \int_0^t m(s)ds$$

with $g(\cdot)$ given by

$$(8.36)\qquad \begin{aligned} g(X) &= m_0 \left[\int_{[0,1]} X(du)M(u,\cdot) - X \right] \\ &\quad + r_0 \int_{[0,1]}\int_{[0,1]} X(dv_1)X(dv_2)[\alpha(v_1,v_2)][R(v_1,v_2;\cdot) - X]. \end{aligned}$$

Therefore applying this fact to φ, m_1, X_1 and φ, m_2, X_2 we obtain with $\langle \nu, f\rangle := \int f(u)\nu(du)$ and $M_i(t),\ i=1,2$ denoting the martingale parts of $X_i(t),\ i=1,2$ from the standard Girsanov theorem for measure-valued diffusions ([D], Theorem 7.2.2) that:

$$(8.37)\qquad \begin{aligned} \mathcal{R}(t,m_i,X_i) &= \exp\left\{ \frac{s_0}{d_0} \int_0^t \left(\int_{[0,1]} \varphi(u)(\mathcal{M}_i(ds))(du) \right.\right. \\ &\quad \left.\left. - \frac{s_0^2}{2d_0} \cdot \left[\langle X_i(s), \varphi^2\rangle - (\langle X_i(s),\varphi\rangle)^2 \right] \right) ds \right\}. \end{aligned}$$

We will in the estimate in the sequel view the expressions defined in (8.35) - (8.37) as functionals of X and m and insert also other combinations than m_1, X_1 or m_2, X_2 namely m_1, X_2 or m_2, X_1.

Now return to (8.34). We can estimate as follows:

$$(8.38)\qquad \begin{aligned} |m_1(t,\phi) - m_2(t,\phi)| &\leq |E_{1,2}[\mathcal{R}(t,m_1,X_1)X_1(t,\phi)] - E_{1,2}[\mathcal{R}(t,m_1,X_2)X_2(t,\phi)]| \\ &+ |E_{1,2}[\mathcal{R}(t,m_2,X_2)X_2(t,\phi)] - E_{1,2}[\mathcal{R}(t,m_1,X_2)X_2(t,\phi)]| \end{aligned}$$

We continue by writing (apply the triangle inequality again to first summand)

$$(8.39)\qquad \begin{aligned} |m_1(t,\phi) - m_2(t,\phi)| &\leq E_{1,2}\big\{ |\mathcal{R}(t,m_1,X_1)| |X_1(t,\phi) - X_2(t,\phi)| \\ &\quad + |X_2(t,\phi)| \cdot |\mathcal{R}(t,m_1,X_2) - \mathcal{R}(t,m_1,X_1)| \\ &\quad + |X_2(t,\phi)| \cdot |\mathcal{R}(t,m_1,X_2) - \mathcal{R}(t,m_2,X_2)| \big\}. \end{aligned}$$

We begin by deriving bounds for the differences appearing in the expectation on the r.h.s.

Observe that that $\mathcal{R}$ is bounded as well as $|X_2(t,\phi)|$. Furthermore since ϕ is Lipschitz we have by the neutral coupling that, (with $||\cdot||$ denoting the Vasherstein distance (compare the derivation of (5.65)) and recall (8.13) to get:

$$E_{1,2}|X_1(t,\phi)-X_2(t,\phi)| \le L(\phi)E_{1,2}(||X_1(t)-X_2(t)||) \le CL(\phi)|\int_0^t ||m_1(s)-m_2(s)||ds. \tag{8.40}$$

Furthermore using (8.37) (in combination with (8.35), (8.36)) we see that together with the Lipschitz continuity of φ and the inequality $|\exp a - \exp b| \le C_T|a-b|$ for $a,b \in [0,T]$ the relation (8.40) implies that (recall (8.37) and the text below in connection with (8.35) and (8.36)):

$$E_{1,2}|\mathcal{R}(t,m_1,X_1)-\mathcal{R}(t,m_1,X_2)| \le D_T \int_0^t ||m_1(s)-m_2(s)||ds. \tag{8.41}$$

Finally we estimate $|\mathcal{R}(t,m_1,X_2)-\mathcal{R}(t,m_2,X_2)|$ using (8.35) - (8.37). We see from the formula in (8.37) that since we insert both times X_2 that we need to bound only the term involving the m_i in (8.35). For this purpose we estimate for $0 \le t \le T$:

$$\begin{aligned} |\exp\left(\int_0^t m_1(s,\phi)ds\right) \quad - \quad \exp\left(\int_0^t m_2(s,\phi)ds\right)| &\le C_T|\int_0^t m_1(s,\phi)-m_2(s,\phi)ds| \\ &\le C_T\, L(\varphi)\int_0^t ||m_1(s)-m_2(s)||ds. \end{aligned} \tag{8.42}$$

Therefore we get

$$E_{1,2}|\mathcal{R}(t,m_1,X_2)-\mathcal{R}(t,m_2,X_2)| \le B_T \int_0^t ||m_1(s)-m_2(s)||ds. \tag{8.43}$$

Combining (8.40), (8.41) and (8.43) we obtain from (8.39) for all ϕ with $L(\phi) \le 1$ and $t \in T$ that a.s.:

$$\begin{aligned} |m_1(t,\phi)-m_2(t,\phi)| \quad &\le \quad \Big\{||\mathcal{R}(t,m_1,X_1)||_\infty \int_0^t ||m_1(s)-m_2(s)||ds \\ &+ \quad \text{Const} \cdot \int_0^t ||m_1(s)-m_2(s)||ds\Big\}. \end{aligned} \tag{8.44}$$

Altogether since the constants in the estimate above are independent of ϕ:

$$||m_1(t)-m_2(t)|| \le C\int_0^t ||m_1(s)-m_2(s)||ds, \quad C \in (0,\infty),\ t \le T. \tag{8.45}$$

and the uniqueness follows from Gronwall's lemma using that $m_1(0)=m_2(0)$.

Part (a) Duality proof
We shall introduce a dual process for the process defined via the nonlinear martingale problem. Since we know that the process we defined for fixed N has a dual process (compare section 3 (a)) we simply consider the limiting object

to obtain a candidate for the desired dual process. For that purpose it will be necessary to restrict our attention to exchangeable initial states for the process. The dual process has as set of colonies the set $\mathbb{N}$ (instead of Ω_N as in chapter 3) and will then agree, except the migration part, with the process introduced in section 3 in (3.2) - (3.6). The migration part has the feature that (since we took the limit $N \to \infty$) each migration step leads to a site which has not previously been occupied. This means in particular that coalescence can only occur in the beginning when partitions start initially at the same site or when a new particle is created in the particle process. Formally this migration mechanism can be defined by saying that a jump of a partition elements always leads to a colony of smallest index which has not yet been occupied (recall that the initial state is exchangeable). Then we obtain by a modification of the argument of Theorem 7 the well-posedness of the martingale problem, since all moment measures of the process are determined via the dual process. The new feature here is the fact that we are now working with a non-linear martingale problem and we must now include the case of random initial condition. This is done using the appropriate moment measures (to replace the product measures) for terms involving more than one partition element.

Part (b) Let $\Xi_N(t)$ denote the empirical process with values in $\mathcal{P}(\mathcal{P}(\mathbb{N}))$:

$$(8.46)\qquad \Xi_N(t) := \frac{1}{N}\sum_{\xi=1}^{N} \delta_{x_\xi^N(t)}.$$

The proof follows in two steps, first tightness is shown and then convergence by establishing that every weak limit point solves a well-posed martingale problem.

Step 1 In the first step we check that the processes $\mathcal{L}[\Xi_N((x_\xi^N(t))_{\xi\in\Omega_N})_{t\geq 0}]$ are tight in $D([0,\infty), \quad \mathcal{P}(\mathcal{P}(\mathbb{N})))$ - this is a rather standard argument. First we compactify the type space $\mathbb{N}$ by introducing the type $\{\infty\}$. We can consider on $\mathcal{P}(\mathcal{P}(\mathbb{N}))$ functions of the form

$$(8.47)\qquad H(\Gamma) = \prod_{i=1}^{k}\langle \Gamma, F_i\rangle, \quad F_i(x) = \int x(du) g_i(u) \quad g_i \in \ell_\infty.$$

By considering bounded functions g_i with $g_i(u)$ converging as $u \to \infty$ we define g_i on $\mathbb{N} \cup \{\infty\}$ and we get functions H which can be defined on $\mathcal{P}(\mathcal{P}(\mathbb{N}\cup\{\infty\}))$.

Then we use the fact that $\mathcal{P}(\mathbb{N}\cup\{\infty\})$ is compact to derive with the tightness criterion of Joffe-Metivier [JM] Proposition 3.2.3, which considers sequences of semi-martingales. Consider a function H as above, then we have to show that the semi-martingales $\{(H(\Xi_N(t)))_{t\geq 0}\}$ have local drift and diffusion coefficients $a^N(t,\omega)$, $b^N(t,\omega)$ which satisfy:

$$(8.48)\qquad \sup_N E \sup_{t\leq T}(((b^N(t,\omega)) + (a^N(t,\omega))^2) < \infty.$$

Since H^2 is an eligible function if H is, it suffices to show

$$(8.49)\qquad \sup_N E \sup_{t\leq T}(a^N(t,\omega))^2 < \infty.$$

Let L^N denote the generator of the process $\Xi_N(t)$ (recall that we consider exchangeable initial states). We need to verify that for H in (8.47) and with $\tilde{\mathcal{P}}_N$ denoting probability measures of the form $\frac{1}{N}\sum_{i=1}^{N}\delta_{x_i}$:

$$(8.50)\qquad \sup_N \sup_{\Gamma\in\tilde{\mathcal{P}}_N(\mathcal{P}(\mathbb{N}\cup\{\infty\}))} |L^N(H)(\Gamma)| \leq C_H.$$

We obtain by inserting the special form of H and Γ that is $\Gamma = \frac{1}{N}\sum_{i=1}^{N}\delta_{x_i}$, $x_i \in \mathcal{P}(\mathbb{N})$ that we can express H as a function of $(x_1,\dots,x_N)$ and hence evaluate the generator as follows:

$$(8.51)\qquad (L^N H)(\Gamma) = \frac{1}{N}\,\frac{1}{N^k}\sum_{i_1,\dots,i_k}\sum_{j=1}^{k} G_k^N \tilde{F}_{i_1,\dots,i_k}(x_1,\dots,x_N)$$

with $\tilde{F}_{i_1,\dots,i_k}(x_1,\dots,x_N) = F_1(x_{i_1})\dots F_k(x_{i_k})$ and G_i^N are operators acting on the i-th component of $\tilde{F}$ as the generator of $(x_\xi^N(t))$ does. Now it is elementary to check (8.50) by using the definition of $(x_\xi^N(t))_{\xi=1,\dots,N}$.

It remains now to verify that for a limit law P and every T

$$P(\Xi_\infty(t)(\{\infty\}) > 0 \quad \text{for some} \quad t \le T) = 0. \tag{8.52}$$

To verify this we will check that for every $\varepsilon > 0$:

$$\inf_N P\left[(x_\xi^N(t)(\{1,\dots,K\}) \ge 1-\varepsilon,\ \forall t \le T)\right] \underset{K\to\infty}{\longrightarrow} 1 \tag{8.53}$$

We get this by returning to the martingale problem characterizing the process. Using maximal martingale inequalities and the fact of bounded mutation and immigration rates the derivation of (8.53) is now standard.

Step 2 In the second step we verify that any limit point law is a solution of the nonlinear martingale problem given in (a). The proof of this fact is essentially the same as the proof of Lemma 5.1 in subsection 5 (d) and we refer the reader to this subsection for details. Note that a) has a deterministic solution. Since the original systems were exchangeable, and the limit of the empirical processes is deterministic, this together with the general version of the di Finetti's theorem ([D] section 11.2) immediately yields the propagation of chaos which gives the independence of the $\{x_k^*\}$.

(c) Proof of Theorem 10

(a) Let π_* be a fixed point for the equation (2.18). Now take a one site Fleming-Viot system in equilibrium with selection given by the fitness function φ and with mutation source $\varrho_* = (\alpha\pi_* + (1-\alpha)\varrho)$ and mutation rate $(c_0 + \bar{m}_0)$ where $\alpha = \frac{c_0}{c_0+\bar{m}_0}$. The expression for this equilibrium then follows from Theorem 8(a)(iii). Then by the fixed point property this process coincides with the McKean-Vlasov dynamics and hence gives a stationary state of the latter.

On the other hand given any stationary measure for the McKean-Vlasov system it follows from the martingale problem that the dynamics is that of the Fleming-Viot system described above. Since that system has a unique equilibrium of the form given in (2.20) the mean π_* satisfies the fixed point equation.

(b) The proof is based on a contraction argument. Recall the definition of the map G in (2.18).

Let $\|\cdot\|$ denote the total variation norm. For large enough mutation coefficient $\bar{m}$ we will show that the iteration $\nu_{n+1} = G(\nu_n), \nu_0 = \Theta$ converges in the $||\cdot||$-norm to the the unique fixed point of $\nu = G(\nu)$. To do this, we shall show that the map, G, is a *contraction* for given s if $\bar{m}$ is large enough.

Note that since φ is bounded the exponential term in (2.18) defining G can easily be handled and we are left with two tasks, namely first to compare the difference between the normalizing constants (compare (2.8)) $Z_1^{-1},\ Z_2^{-1}$ if we insert two measures $\nu_1,\ \nu_2$ for ν and second we need to estimate the distance between the respective equilibrias $\Gamma_{\nu_i}^{c,d},\ \ i=1,2$ for the neutral model. We first do the latter.

(i) In order to compare the GEM-distribution (i.e. $\Gamma_\mu^{c,d}$) for the two immigration sources $\mu_i := \alpha\nu_i + (1-\alpha)\varrho \ \ i=1,2$ recall the representation (1.48). We abbreviate the expression involving the V_i as follows:

$$M_i = V_i \prod_{\ell=1}^{i-1}(1-V_\ell). \tag{8.54}$$

Then we can couple $\Gamma_{\mu_i}^{c,d},\ i=1,2$ as follows. First, we use the same weights $\{M_i,\quad i\in\mathbb{N}\}$ in both representations. Secondly we couple the $(U_i^1)_{i\in\mathbb{N}},\ (U_i^2)_{i\in\mathbb{N}}$ by drawing the same points with $\mu_1 \wedge \mu_2$ and independent points with $||\mu_1-\mu_2||^+$ resp. $||\mu_1-\mu_2||^-$.

We can now derive our first estimate. Fix a subset A of the type space and sources μ_1, μ_2 given by $\alpha\nu_1 + (1-\alpha)\varrho$ and $\alpha\nu_2 + (1-\alpha)\varrho$. We can then obtain:

$$(8.55) \qquad |\int_{\mathcal{P}(\mathbb{N})} \lambda(A) \left[\Gamma^{c,d}_{(\alpha\nu_1+(1-\alpha)\varrho)}(d\lambda) - \Gamma^{c,d}_{(\alpha\nu_2+(1-\alpha)\varrho)}(d\lambda)\right]|$$

$$\leq E\left[|\sum_i 1_A(U_i^1)M_i - \sum_i 1_A(U_i^2)M_i|\right] \leq \alpha||\nu_1 - \nu_2||.$$

(ii) A second estimate is now needed to compare Z_1^{-1} and Z_2^{-1}. Namely inserting the definition of Z and using the representation of the coupled $\Gamma^{c,d}_{\mu_i}$ $\quad i = 1, 2$ defined above together with (8.55) we estimate:

$$(8.56) \qquad \begin{aligned} \Big|\frac{1}{Z_1} - \frac{1}{Z_2}\Big| &\leq \Bigg|\int_{\mathcal{P}(\mathbb{N})} \exp(2s \int_{\mathbb{N}} \varphi(u)\lambda(du)) \cdot \left[\Gamma^{c,d}_{(\alpha\nu_1+(1-\alpha)\varrho)}(d\lambda) - \Gamma^{c,d}_{(\alpha\nu_2+(1-\alpha)\varrho)}(d\lambda)\right]\Bigg| \\ &\leq E\Bigg|\left[\exp(2s\sum_i M_i\varphi(U_i^1)) - \exp(2s\sum_i M_i\varphi(U_i^2))\right]\Bigg| \\ &\leq \text{const } E\left[\sum_i M_i|\varphi(U_i^1) - \varphi(U_i^2)|\right] e^{2s} \\ &\leq \text{const } \alpha||\nu_1 - \nu_2||e^{2s}. \end{aligned}$$

Now we return to the task to derive a bound on the difference $G(\nu_1) - G(\nu_2)$. First estimate:

$$(8.57) \qquad \begin{aligned} \|G(\nu_1) - G(\nu_2)\| = \; & sup_A\Bigg|\int \exp(2s\int_{\mathbb{N}} \varphi(u)\lambda(du)) \\ & \cdot \lambda(A)[Z_1^{-1}\Gamma_{(\alpha\nu_1+(1-\alpha)\varrho)}(d\lambda) - Z_2^{-1}\Gamma_{(\alpha\nu_2+(1-\alpha)\varrho)}(d\lambda)]\Bigg|. \end{aligned}$$

Next we use that $|\varphi| \leq 1$ and use the triangle inequality to bring (8.55) and (8.56) into play. We get

$$(8.58) \qquad ||G(\nu_1) - G(\nu_2)|| \leq (C_1 + C_2e^{4s})\alpha||\nu_1 - \nu_2||.$$

Hence G is a contraction mapping if $(C_1 + C_2e^{4s})\alpha < 1$. Here for fixed s and using that by the definition of α by making $\bar{m}$ large, we can make α arbitrarily small we get $(C_1 + C_2e^{4s})\alpha < 1$ for $\bar{m}$ large enough. The result then follows from the contraction mapping principle.

(c) Here we use the duality relation introduced in the duality proof of Theorem 9 (a). The uniform integrability and the convergence follows exactly as in the duality-proof of Theorem 8 (b).

References

[CFG] J. T. Cox, K. Fleischmann, A. Greven (1996), *Comparison of interacting diffusions and an application to their ergodic theory.* Probab. Theory Rel. Fields, **105**, 513-528.

[CGS] J. T. Cox, A. Greven, T. Shiga (1995), *Finite and infinite systems of interacting diffusions.* Probab. Theory Relat. Fields **102**, 165-197.

[D] D. A. Dawson (1993), *Measure-valued Markov Processes.* In: École d'Été de Probabilités de Saint Flour XXI, Lecture Notes in Mathematics **1541**, 1-261, Springer-Verlag.

[DG1] D. A. Dawson and A. Greven (1993a), *Multiple time scale analysis of hierarchically interacting systems.* In: A Festschrift to honor G. Kallianpur, 41-50, Springer-Verlag.

[DG2] D. A. Dawson and A. Greven (1993b), *Multiple time scale analysis of interacting diffusions.* Probab. Theory Rel. Fields **95**, 467-508.

[DG3] D. A. Dawson and A. Greven (1993c), *Hierarchical models of interacting diffusions: multiple time scale phenomena. Phase transition and pattern of cluster-formation.* Probab. Theory Rel. Fields, **96**, 435-473.

[DG4] D. A. Dawson and A. Greven (1996), *Hierarchically interacting Fleming-Viot processes with selection and mutation: Multiple space time scale analysis and quasi equilibria*, Electronic Journal of Probability, Vol. **1**, paper no. 14, pages 1-84.

[DGV] D. A. Dawson, A. Greven, J. Vaillancourt (1995), *Equilibria and Quasiequilibria for Infinite Collections of Interacting Fleming-Viot processes.* Transactions of the American Math. Society, vol **347**, No. 7, 2277-2360.

[DM] D. A. Dawson and P. March (1995), *Resolvent estimates for Fleming-Viot operators and uniqueness of solutions to related Martingale problems*, J. Funct. Anal., Vol. **132**, 417-472.

[E1] S. N. Ethier (1990), *The distribution of frequencies of age-ordered alleles in a diffusion model*, Adv. Appl. Probab. **22**, 519-532.

[E2] S. N. Ethier and T.G. Kurtz (1992), *On the stationary distribution of the neutral one-locus diffusion model in population genetics*, Ann. Appl. Probab. **2**, 24-35.

[E3] S. N. Ethier (1992), *Equivalence of two descriptions of the ages of alleles*, J. Appl. Probab. **29**, 185-189.

[Ed] G. M. Edelmann (1987), *Neural Darwinism*, Basic Books. New York.

[EG] S. N. Ethier and B. Griffiths (1981), *The infinitely-many-sites model as a measure-valued diffusion*, Ann. Probab. **15**, 515-545.

[EK1] S. N. Ethier and T. G. Kurtz (1981), *The infinitely-many-neutral-alleles diffusion model*, Adv. Appl. Probab. **13**, 429-452.

[EK2] S. N. Ethier and T. G. Kurtz (1986/93), *Markov processes, characterization and convergence*, Wiley, New York.

[EK3] S. N. Ethier and T. G. Kurtz (1987), *The infinitely-many-alleles-model with selection as a measure-valued diffusion*, Lecture Notes in Biomath., vol. **70**, Springer-Verlag, 72-86.

[EK4] S. N. Ethier and T. G. Kurtz (1994), *Convergence to Fleming-Viot processes in the weak atomic topology*, Stochastic Process, Appl. **54**, 1-27.

[EK5] S. N. Ethier and T.G. Kurtz (1996), *Coupling and ergodic theorems for Fleming-Viot processes*, Ann. Probab., to appear.

[H] J. Holland (1992), *Adaptation in natural and artificial systems*, MIT Press.

[JM] A. Joffe and M. Métivier (1986), *Weak convergence of sequences of semi-martingales with application to multitype branching processes.* Adv. Appl. Probab. **18**, 20-65.

[KKN] H. C. Kang; S. M. Krone; C. Neuhauser (1995): *Stepping-Stone models with extinction and recolonization.* Ann. Appl. Prob., Vol. **5**, No. 4, 1025-1060

[ORS] L. Overbeck, M. Röckner, J. Schmuland (1995): *An analytic approach to Fleming-Viot processes*, Ann. Prob. 23, 1-36.

[PT] G. P. Patil and C. Taillie (1977), *Diversity as a concept and its implications for random communities*, Bull. Internat. Statist, Inst. **47**, 497-515.

[SF] S. Sawyer and J. Felsenstein (1983). *Isolation by distance in a hierarchically clustered population*, J. Appl. Prob. **20**, 1-10.

[S1] T. Shiga (1982), *Wandering phenomena in infinite allelic diffusion models*, Adv. Appl. Probab. **14**, 457-483.

[S2] T. Shiga (1982), *Continuous time multi-allelic stepping stone models in population genetics*, J. Math. Kyoto Univ. **22**, 1-40.

[SU] T. Shiga, Uchiyama (1986), *Stationary states and the stability of the stepping stone model involving mutation and selection*, Prob. Th. Rel. Fields **73**, 87-117.

[Sm] J. Maynard Smith (1982): *Evolution of the theory of games?* Cambridge University Press

[V] J. Vaillancourt (1990), *Interacting Fleming-Viot processes*, Stochastic Process Appl. **36**, 45-57.

[W] M. J. Wade (1992), *Sewall Wright: gene interaction and the shifting balance theory*, Oxford Surveys Evol. Biol. **8**, 35-62.