EURANDOM PREPRINT SERIES
2012-016
September 10, 2012

# Renormalisation of hierarchically interacting Cannings processes

A. Greven, F. den Hollander, S. Kliem, A. Klimovsky

ISSN 1389-2355

# Renormalisation of hierarchically interacting Cannings processes

A. Greven[1]
F. den Hollander[2]
S. Kliem[3]
A. Klimovsky[2]

September 10, 2012

## Abstract

In order to analyse universal patterns in the large space-time behaviour of interacting multi-type stochastic populations on countable geographic spaces, a key approach has been to carry out a renormalisation analysis in the hierarchical mean-field limit. This has provided considerable insight into the structure of interacting systems of finite-dimensional *diffusions*, such as Fisher-Wright or Feller diffusions, and their infinite-dimensional analogues, such as Fleming-Viot or Dawson-Watanabe superdiffusions.

The present paper brings a new class of interacting *jump processes* into focus. We start from a single-colony $C^{\Lambda}$-process, which arises as the continuum-mass limit of a $\Lambda$-Cannings individual-based population model, where $\Lambda$ is a finite non-negative measure that describes the offspring mechanism, i.e., how individuals in a single colony are replaced via resampling. The key feature of the $\Lambda$-Cannings individual-based population model is that the offspring of a single individual can be a positive fraction of the total population. After that we introduce a system of *hierarchically interacting* $C^{\Lambda}$-processes, where the interaction comes from migration and reshuffling-resampling on *all* hierarchical space-time scales *simultaneously.* More precisely, individuals live in colonies labelled by the hierarchical group $\Omega_N$ of order $N$, and are subject to *migration* based on a sequence of migration coefficients $\underline{c} = (c_k)_{k\in\mathbb{N}_0}$ and to *reshuffling-resampling* based on a sequence of resampling measures $\underline{\Lambda} = (\Lambda_k)_{k\in\mathbb{N}_0}$, both acting in $k$-blocks for all $k \in \mathbb{N}_0$. The reshuffling is linked to the resampling: before resampling in a block takes place all individuals in that block are relocated uniformly, i.e., resampling is done in a locally "panmictic" manner. We refer to this system as the $C_N^{\underline{c},\underline{\Lambda}}$-process. The dual process of the $C^{\Lambda}$-process is the $\Lambda$-coalescent, whereas the dual process of the $C_N^{\underline{c},\underline{\Lambda}}$-process is a spatial coalescent with multi-level block coalescence.

For the above system we carry out a *full renormalisation analysis* in the *hierarchical mean-field limit* $N \to \infty$. Our main result is that, in the limit as $N \to \infty$, on each hierarchical scale $k \in \mathbb{N}_0$ the $k$-block averages of the $C_N^{\underline{c},\underline{\Lambda}}$-process converge to a random process that is a superposition of a $C^{\Lambda_k}$-process and a Fleming-Viot process, the latter with a volatility $d_k$ and with a drift of strength $c_k$ towards the limiting $(k+1)$-block average. It turns out that $d_k$ is a function of $c_l$ and $\Lambda_l$ for all $0 \leq l < k$. Thus, it is through the volatility that the renormalisation manifests itself. We investigate how $d_k$ scales as $k \to \infty$, which requires an analysis of compositions of certain Möbius-transformations, and leads to four different *regimes.*

---

[1] Department Mathematik, Universität Erlangen-Nürnberg, Cauerstraße 11, D-91058 Erlangen, Germany
[2] Mathematical Institute, Leiden University, P.O. Box 9512, NL-2300 RA Leiden, The Netherlands
[3] Fakultät für Mathematik, Universität Duisburg-Essen, Universitätsstraße 2, D-45141 Essen, Germany

We discuss the implications of the scaling of $d_k$ for the behaviour on large space-time scales of the $C_N^{c,\Lambda}$-process. We compare the outcome with what is known from the renormalisation analysis of hierarchically interacting Fleming-Viot diffusions, pointing out several new features. In particular, we obtain a *new classification* for when the process exhibits *clustering* (= develops spatially expanding mono-type regions), respectively, exhibits *local coexistence* (= allows for different types to live next to each other with positive probability). Here, the simple dichotomy of recurrent versus transient migration for hierarchically interacting Fleming-Viot diffusions, namely, $\sum_{k\in\mathbb{N}_0}(1/c_k) = \infty$ versus $< \infty$, is replaced by a dichotomy that expresses a trade-off between migration and reshuffling-resampling, namely, $\sum_{k\in\mathbb{N}_0}(1/c_k)\sum_{l=0}^{k}\Lambda_l([0,1]) = \infty$ versus $< \infty$. Thus, while recurrent migrations still only give rise to clustering, there now are transient migrations that do the same when the block resampling is strong enough, namely, $\sum_{l\in\mathbb{N}_0}\Lambda_l([0,1]) = \infty$. Moreover, in the clustering regime we find a richer scenario for the *cluster formation* than for Fleming-Viot diffusions. In the local-coexistence regime, on the other hand, we find that the types initially present only survive with a positive probability, not with probability one as for Fleming-Viot diffusions. Finally, we show that for finite $N$ the same dichotomy between clustering and local coexistence holds as for $N \to \infty$, even though we lack proper control on the cluster formation, respectively, on the distribution of the types that survive.

**MSC 2010:** Primary 60J25, 60K35; Secondary 60G57, 60J60, 60J75, 82C28, 92D25.

**Key words and phrases:** $C^\Lambda$-process, $\Lambda$-coalescent, hierarchical group, block migration, block reshuffling-resampling, spatial coalescent, hierarchical mean-field limit, renormalisation, McKean-Vlasov process, Möbius-transformation.

**Acknowledgements:** The idea for this paper arose from discussions with P. Pfaffelhuber and A. Wakolbinger during an Oberwolfach meeting on “Random Trees” in January 2009. FdH thanks J. Goodman, R.J. Kooman and E. Verbitskiy for discussions on Möbius-transformations. AG was supported by the Deutsche Forschungsgemeinschaft (grant DFG-GR 876/15-1), FdH by the European Research Council (Advanced Grant VARIS-267356), and AK by the European Commission (project PIEF-GA-2009-251200). SK held a postdoctoral position at EURANDOM from the Summer of 2009 until the Summer of 2011. AK was a guest of the Hausdorff Research Institute for Mathematics in Bonn in the Fall of 2010 (Junior Trimester Program on Stochastics). SK and AK where guests of the Institute for Mathematical Sciences, National University of Singapore, during its 2011 programme on Probability and Discrete Mathematics in Mathematical Biology.

# Contents

# 1 Introduction and main results

## 1.1 Outline

Section 1.2 provides the background for the paper. Section 1.3 defines the single-colony and the multi-colony $C^\Lambda$-process, as well as the so-called McKean-Vlasov $C^\Lambda$-process, a single-colony $C^\Lambda$-process with immigration and emigration from and to a cemetery state arising in the context of the scaling limit of the multi-colony $C^\Lambda$-process with mean-field interaction. Section 1.4 defines a *new* process, the $C_N^{c,\Lambda}$-process, where the countably many colonies are labelled by the hierarchical group $\Omega_N$ of order $N$, and the migration and the reshuffling-resampling on successive hierarchical space-time scales are governed by a sequence $\underline{c} = (c_k)_{k\in\mathbb{N}_0}$ of migration coefficients and a sequence $\underline{\Lambda} = (\Lambda_k)_{k\in\mathbb{N}_0}$ of resampling measures. Section 1.5 introduces *multiple space-time scales* and a collection of *renormalised systems*. It is shown that, in the hierarchical mean-field limit $N \to \infty$, the block averages of the $C_N^{\underline{c},\underline{\Lambda}}$-process on hierarchical space-time scale $k$ converge to a McKean-Vlasov process that is a superposition of a single-colony $C^{\Lambda_k}$-process and a single-colony Fleming-Viot process with a volatility $d_k$ that is a function of $c_l$ and $\Lambda_l$ for all $0 \leq l < k$, and a drift of strength $c_k$ towards the limiting $(k+1)$-st block average. The scaling of $d_k$ as $k \to \infty$ turns out to have several *universality classes*. Section 1.5.5 discusses the implications of this scaling for the behaviour of the $C_N^{\underline{c},\underline{\Lambda}}$-process on large space-time scales, and compares the outcome with what is known for hierarchically interacting Fleming-Viot diffusions.

A key feature of the $C_N^{\underline{c},\underline{\Lambda}}$-process is that it has a spatial $\underline{\Lambda}$-coalescent with block migration and block coalescence as a dual process. This duality, which is of intrinsic interest, and the properties of the dual process are worked out in Section 2. The proofs of the main theorems are given in Sections 3–11. To help the reader, a list of the main symbols used in the paper is added in Section 12.

## 1.2 Background

### 1.2.1 Population dynamics

For the description of spatial populations subject to migration and to neutral stochastic evolution (i.e., resampling without selection, mutation or recombination), it is common to use variants of interacting Fleming-Viot diffusions (Dawson [D93], Donnelly and Kurtz [DK99], Etheridge [E00, E11]). These are processes taking values in $\mathcal{P}(E)^I$, where $I$ is a countable Abelian group playing the role of a *geographic space* labelling the colonies of the population (e.g. $\mathbb{Z}^d$, the $d$-dimensional integer lattice, or $\Omega_N$, the hierarchical group of order $N$), $E$ is a compact Polish space playing the role of a *type space* encoding the possible types of the individuals living in these colonies (e.g. $[0,1]$), and $\mathcal{P}(E)$ is the set of probability measures on $E$. An element in $\mathcal{P}(E)^I$ specifies the frequencies of the types in each of the colonies in $I$.

Let us first consider the (locally finite) populations of individuals from which the above processes arise as continuum-mass limits. Assume that the individuals *migrate* between the colonies according to independent continuous-time random walks on $I$. Inside each colony, the evolution is driven by a change of generation called *resampling*. Resampling, in its simplest form (Moran model), means that after exponential waiting times a pair of individuals ("the parents") is replaced by a new pair of individuals ("the children"), who randomly and independently adopt the type of one of the parents. The process of type *frequencies* in each of the colonies as a result of the migration and the resampling is a jump process taking values in $\mathcal{P}(E)^I$.

If we pass to the *continuum mass limit* of the frequencies by letting the number of individuals per colony tend to infinity, then we obtain a system of *interacting Fleming-Viot diffusions* (Dawson, Greven and Vaillancourt [DGV95]). By picking different resampling mechanisms, occurring at a rate that depends on the state of the colony, we obtain variants of interacting Fleming-Viot diffusions with a state-dependent resampling rate [DM95]. In this context, key questions are: To what extent does the behaviour on large space-time scales depend on the precise form of the resampling mechanism? In particular, to what extent is this behaviour *universal*? For Fleming-Viot models and a small class of state- and type-dependent Fleming-Viot models this question has been answered in [DGV95].

If we consider resampling mechanisms where, instead of a pair of individuals, a positive fraction of the local population is replaced (an idea due to Cannings [C74, C75]), then we enter the world of *jump processes.* In this paper, we will focus on jump processes that are parametrised by a measure $\Lambda$ on $[0,1]$ that models the random proportion of offspring in the population generated by a single individual in a resampling event. It has been argued by many authors that such jump processes are suitable for describing situations with *little biodiversity.* For instance, the jumps may account for selective sweeps, or for extreme reproduction events (occurring on smaller time scales and in a random manner so that an effectively neutral evolution results), such as those observed in certain marine organisms, e.g. Atlantic cod or Pacific oyster (Eldon and Wakeley [EW06]). It is argued in Der, Epstein and Plotkin [DEP11] that mixtures of diffusive dynamics and Cannings dynamics provide a better fit for generation-by-generation empirical data from Drosophila populations. Birkner and Blath [BB08, BB09-MVD] treat the issue of statistical inference on the genealogies corresponding to a one-parameter family of Cannings dynamics.

Our goal is to describe the effect of jumps in a *spatial* setting with a volatile reproduction. To that end we study a system of hierarchically interacting Cannings processes. The interaction is chosen in such a way that the hierarchical lattice mimics the two-dimensional Euclidean space, as will become clear later on. On top of migration and single-colony resampling, we add *multi-colony resampling* by carrying out a Cannings-type resampling in all blocks simultaneously, combined with a *reshuffling* of the individuals inside the block before the resampling is done. The reshuffling mimics the fact that in reproduction the local geographic interaction typically takes place on a smaller time scale, in a random manner, and effectively results in a Cannings jump during a single observation time. We will see that in our model the reshuffling allows us to define the process for all $\Lambda$, and simplifies the analysis by avoiding the need to compensate for small jumps.

The idea to give reproduction a non-local geographic structure, in particular, in two dimensions, was exploited in Barton, Etheridge and Véber [BEV10] and Berestycki, Etheridge and Véber [BEV] also. There, the process on the torus of sidelength $L$ is constructed via its dual, and it is shown that a limiting process on $\mathbb{R}^2$ exists as $L \to \infty$. In [BEV10, BEV] it is assumed that the individual lineages are compound Poisson processes. Freeman [Fr] considers a particular case of the spatially structured Cannings model with a continuum self-similar geographic space, where all individuals in a block are updated upon resampling. This setup does not require compensation for small jumps and allows for their accumulation.

### 1.2.2 Renormalisation

A key approach to understand universality in the behaviour of interacting systems has been a *renormalisation analysis of block averages on successive space-time scales* combined with a *hierarchical mean-field limit.* In this setting, one replaces $I$ by the hierarchical group $\Omega_N$ of order $N$ and passes to the limit $N \to \infty$ ("the hierarchical mean-field limit") [1]. With the limiting dynamics obtained through the hierarchical mean-field limit one associates a (nonlinear) *renormalisation transformation* $\mathcal{F}_c$ (which depends on the migration rate $c$), acting on the resampling rate function $g$ driving the diffusion in single colonies. One studies the *orbit* $(\mathcal{F}^{[k]}(g))_{k\in\mathbb{N}}$, with $\mathcal{F}^{[k]} = \mathcal{F}_{c_{k-1}} \circ \cdots \circ \mathcal{F}_{c_0}$, characterising the behaviour of the system on an

---

[1] Actually, this set-up provides an *approximation* for the geographic space $I = \mathbb{Z}^2$, on which simple random walk migration is critically recurrent (Dawson, Gorostiza and Wakolbinger [DGW]). We will comment on this issue in Section 1.4.2.

increasing sequence of space-time scales, where $(c_k)_{k\in\mathbb{N}}$ represents the sequence of migration coefficients, with the index $k$ labelling the hierarchical distance. The *universality classes* of the system are associated with the fixed points (or the fixed shapes) of $\mathcal{F}_c$, i.e., $g$ with $\mathcal{F}_c(g) = ag$ with $a = 1$ (or $a = a(c) \in (0, \infty)$).

The above *renormalisation program* was developed for various choices of the single-colony state space. Each such choice gives rise to a different universality class with specific features for the large space-time behaviour. For the *stochastic part* of the renormalisation program (i.e., the derivation of the limiting renormalised dynamics), see Dawson and Greven [DG93a], [DG93b], [DG93c], [DG96], [DG99], [DG03], Dawson, Greven and Vaillancourt [DGV95], and Cox, Dawson and Greven [CDG04]. For the *analytic part* (i.e., the study of the renormalisation map $\mathcal{F}$), see Baillon, Clément, Greven and den Hollander [BCGH95], [BCGH97], den Hollander and Swart [HS98], and Dawson, Greven, den Hollander, Sun and Swart [DGHSS08].

So far, two important classes of single-colony processes could not be treated: *Anderson diffusions* [GH07] and *jump processes.* In the present paper, we focus on the second class, in particular, on so-called $C^\Lambda$-processes. In all previously treated models, the renormalisation transformation was a map $\mathcal{F}_c$ acting on the set $M(E)$ of measurable functions on $E$, the single-component state space, while the function $g$ was a branching rate, a resampling rate or other, defining a diffusion function $x \mapsto xg(x)$ on $[0, \infty)$ or $x \mapsto x(1-x)g(x)$ on $[0, 1]$, etc. In the present paper, however, we deal with jump processes that are characterised by a sequence of finite measures $\underline{\Lambda} = (\Lambda_k)_{k\in\mathbb{N}_0}$ on $[0, 1]$, and we obtain a renormalisation map $\mathcal{F}_c$ acting on a pair $(g, \underline{\Lambda})$, where $g \in M(E)$ characterises diffusive behaviour and $\underline{\Lambda}$ characterises resampling behaviour. It turns out that the orbit of this map is of the form

$$(1.1)\quad (d_k g^*, (\Lambda_l)_{l\geq k})_{k\in\mathbb{N}_0},$$

where $g^* \equiv 1$ and $d_k$ depends on $d_{k-1}$, $c_{k-1}$ and the total mass of $\Lambda_{k-1}$. Here, as before, $\underline{c} = (c_k)_{k\in\mathbb{N}_0}$ is the sequence of migration coefficients. The reason behind this reduction is that our single-colony process is a superposition of a $C^\Lambda$-process and a Fleming-Viot process with *state-independent* resampling rates and that both these processes renormalise to a multiple of the latter. It turns out that $d_k$ can be expressed in terms of compositions of certain *Möbius-transformations* with parameters changing from composition to composition. It is through these compositions that the *renormalisation* manifests itself.

If the single-colony process would be a superposition of a $C^\Lambda$-process and a Fleming-Viot process with *state-dependent* resampling rate, i.e., $g$ would *not* be a constant but a function of the state, then the renormalisation transformation would be much more complicated. It remains a challenge to deal with this generalisation.

## 1.3 The Cannings model

The $\Lambda$-Cannings model involves a finite non-negative measure $\Lambda \in \mathcal{M}_f([0, 1])$. We focus on the special case with

$$(1.2)\quad \Lambda(\{0\}) = 0$$

satisfying the so-called *dust-free condition*

$$(1.3)\quad \int_{[0,1]} \frac{\Lambda(\mathrm{d}r)}{r} = \infty.$$

Condition (1.2) excludes the well-studied case of interacting Fleming-Viot diffusions, i.e., we focus on the jumps in the $\Lambda$-Cannings model. Condition (1.3) excludes cases where the jump sizes do not accumulate. Moreover, this condition is needed to have well-defined proportions of the different types in the population in the infinite-population limit (Pitman [P99]), and also to be able to define a genealogical tree for the population (Greven, Pfaffelhuber and Winter [GPW09]) [2].

In Sections 1.3.1–1.3.3, we build up the Cannings model in three steps: single-colony $C^\Lambda$-process, multi-colony $C^\Lambda$-process, and $C^\Lambda$-process with immigration-emigration (McKean-Vlasov limit).

### 1.3.1 Single-colony $C^\Lambda$-process

We recall the definition of the $\Lambda$-Cannings model in its simplest form. This model describes the evolution of allelic types of finitely many individuals living in a single colony. Let $M \in \mathbb{N}$ be the number of individuals, and let $E$ be a compact Polish space encoding the types (a typical choice is $E = [0,1]$). The evolution of the population, whose state space is $E^M$, is as follows.

- The number of individuals stays fixed at $M$ during the evolution.
- Initially, i.i.d. types are assigned to the individuals according to a given distribution

  $$(1.4)\quad \theta \in \mathcal{P}(E).$$

- Let $\Lambda^* \in \mathcal{M}([0,1])$ be the $\sigma$-finite measure defined as

  $$(1.5)\quad \Lambda^*(\mathrm{d}r) = \frac{\Lambda(\mathrm{d}r)}{r^2}.$$

  Consider an inhomogeneous Poisson point process on $[0,\infty) \times [0,1]$ with intensity measure

  $$(1.6)\quad \mathrm{d}t \otimes \Lambda^*(\mathrm{d}r).$$

  For each point $(t, r)$ in this process, we carry out the following transition at time $t$. *Mark* each of the $M$ individuals independently with a 1 or 0 with probability $r$, respectively, $1 - r$. *All* individuals marked by a 1 are killed and are replaced by copies of a *single* individual (= “parent”) that is randomly chosen among all the individuals marked by a 1 (see Fig. 1).

In this way, we obtain a pure-jump Markov process, which is called the $\Lambda$-*Cannings model* with measure $\Lambda$ and population size $M$.

Note that, for a jump to occur, at least two individuals marked by a 1 are needed. Hence, for finite $M$, the rate at which some pair of individuals is marked is

$$(1.7)\quad \int_{[0,1]} \frac{\Lambda(\mathrm{d}r)}{r^2}\tfrac{1}{2}M(M-1)\,r^2 = \tfrac{1}{2}M(M-1)\,\Lambda([0,1]) < \infty,$$

and so only finitely many jumps occur in any finite time interval.

By observing the frequencies of the types, i.e., the number of individuals with a given type divided by $M$, we obtain a measure-valued pure-jump Markov process on $\mathcal{P}(E)$. Letting

---

[2] Condition (1.2) is relevant for some of the questions addressed in this paper, though not for all. We comment on this issue as we go along. Another line of research would be to work with the most general Cannings models that allow for simultaneous multiple resampling events. We do not pursue such a generalisation here.

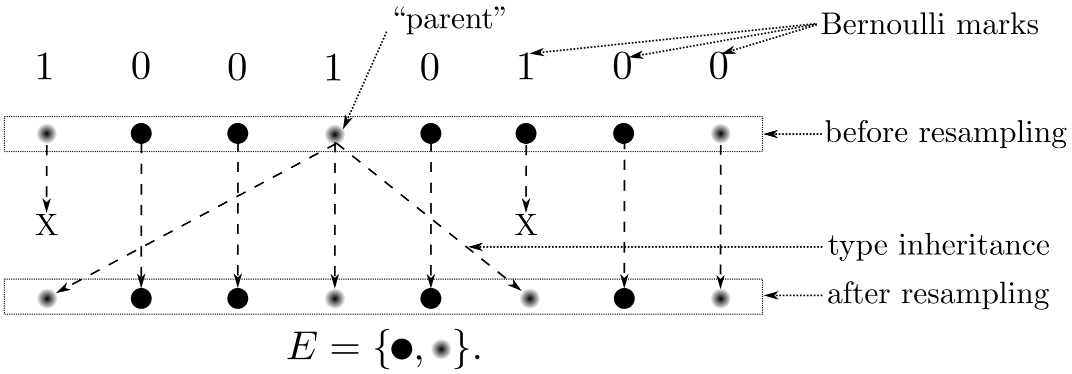

Figure 1: Cannings resampling event in a colony of $M = 8$ individuals of two types. Arrows indicate type inheritance, $X$ indicates death.

$M \to \infty$, we obtain a limiting process $X = (X(t))_{t\geq 0}$, called the *$C^{\Lambda}$-process*, which is a strong Markov jump process with paths in $D([0,\infty),\mathcal{P}(E))$ (the set of càdlàg paths in $\mathcal{P}(E)$ endowed with the Skorokhod $J_1$-topology) and can be characterised as the solution of a well-posed martingale problem (Donnelly and Kurtz [DK99]). This process has countably many jumps in any finite time interval if $\Lambda((0,1]) > 0$ and is the Fleming-Viot diffusion if $\Lambda = \delta_0$. The latter corresponds to Moran resampling.

### 1.3.2 Multi-colony $C^{\Lambda}$-process: mean-field version

Next, we consider the *spatial* $\Lambda$-Cannings model in its standard mean-field version. Consider as geographic space a block of sites $\{0,\dots,N-1\}$ and assign $M$ individuals to each site (= colony). The evolution of the population, whose state space is $(E^M)^N$, is defined as the following pure-jump Markov process.

- The total number of individuals stays fixed at $NM$ during the evolution.
- At the start, each individual is assigned a type that is drawn from $E$ according to some prescribed exchangeable law.
- Individuals *migrate* between colonies at rate $c > 0$, jumping according to the uniform distribution on $\{0,\dots,N-1\}$ (see Fig. 2).
- Individuals *resample* within each colony according to the $\Lambda$-Cannings model with population size corresponding to the current size of the colony.

By considering the frequencies of the types in each of the colonies, we obtain a pure-jump Markov process taking values in $\mathcal{P}(E)^N$.

Letting $M \to \infty$, we pass to the continuum mass limit and we obtain a system of $N$ interacting $C^{\Lambda}$-processes, denoted by

$$(1.8)\quad X^{(N)} = \big(X^{(N)}(t)\big)_{t\geq 0} \quad \text{with} \quad X^{(N)}(t) = \big\{X_i^{(N)}(t)\big\}_{i=0}^{N-1} \in \mathcal{P}(E)^N.$$

The process $X^{(N)}$ can be characterised as the solution of a well-posed martingale problem on $D([0,\infty),\mathcal{P}(E)^N)$ with the product topology on $\mathcal{P}(E)^N$. To this end, we have to consider an algebra $\mathcal{F} \subset C_{\mathrm{b}}(\mathcal{P}(E)^N,\mathbb{R})$ of test functions, and a linear operator $L^{(N)}$ on $C_{\mathrm{b}}(\mathcal{P}(E)^N,\mathbb{R})$

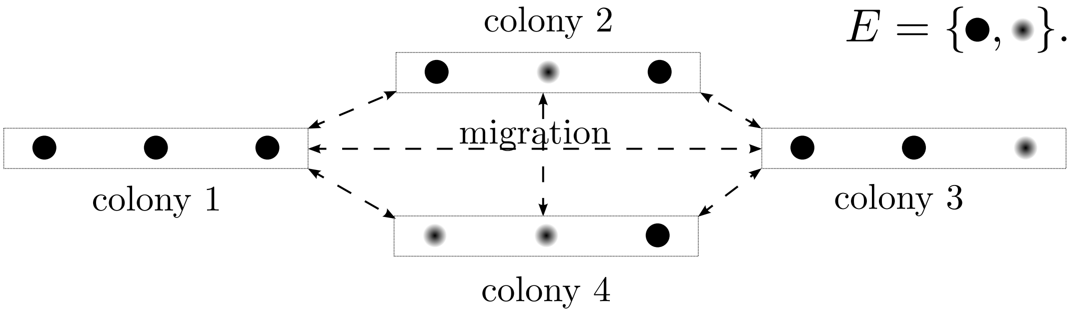

Figure 2: Possible migration paths between $N = 4$ colonies with $M = 3$ individuals of two types in the mean-field version.

with domain $\mathcal{F}$, playing the role of the generator in the martingale problem. Here, we let $\mathcal{F}$ be the algebra of functions $F$ of the form

$$
\begin{aligned}
(1.9)\quad & F(x) = \int_{E^n} \left( \bigotimes_{m=1}^{n} x_{i_m}(\mathrm{d}u^m) \right) \varphi\big(u^1, \dots, u^n\big), \quad x = (x_0, \dots, x_{N-1}) \in \mathcal{P}(E)^N, \\
& n \in \mathbb{N},\ \varphi \in C_{\mathrm{b}}(E^n, \mathbb{R}),\ i_1, \dots, i_n \in \{0, \dots, N-1\}.
\end{aligned}
$$

The generator

$$(1.10)\quad L^{(N)} \colon \mathcal{F} \to C_{\mathrm{b}}\big(\mathcal{P}(E)^N, \mathbb{R}\big)$$

has two parts,

$$(1.11)\quad L^{(N)} = L^{(N)}_{\mathrm{mig}} + L^{(N)}_{\mathrm{res}}.$$

The *migration operator* is given by

$$(1.12)\quad (L^{(N)}_{\mathrm{mig}} F)(x) = \frac{c}{N} \sum_{i,j=0}^{N-1} \int_E (x_j - x_i)(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_i}[\delta_a],$$

where

$$(1.13)\quad \frac{\partial F(x)}{\partial x_i}[\delta_a] = \lim_{h \downarrow 0} \frac{1}{h} \Big[ F(x_0, \dots, x_{i-1}, x_i + h\delta_a, x_{i+1}, \dots, x_{N-1}) - F(x) \Big]$$

is the Gâteaux-derivative of $F$ with respect to $x_i$ in the direction $\delta_a$ (this definition requires that in (1.9) we extend $\mathcal{P}(E)$ to the set of finite signed measure on $E$). Note that the total derivative in the direction $\nu \in \mathcal{P}(E)$ is the integral over $\nu$ of the expression in (1.13), since $\mathcal{P}(E)$ is a Choquet simplex and $F$ is continuously differentiable.

The *resampling operator* is given by (cf. the description of the single-colony $C^\Lambda$-process in Section 1.3.1)

$$
\begin{aligned}
(1.14)\quad (L^{(N)}_{\mathrm{res}} F)(x) = & \sum_{i=0}^{N-1} \int_{[0,1]} \Lambda^*(\mathrm{d}r) \int_E x_i(\mathrm{d}a) \\
& \times \Big[ F\big(x_0, \dots, x_{i-1}, (1-r)x_i + r\delta_a, x_{i+1}, \dots, x_{N-1}\big) - F(x) \Big].
\end{aligned}
$$

Note that, by the law of large numbers, in the limit $M \to \infty$ the evolution in (1.4–1.6) results in the transition $x \to (1-r)x + r\delta_a$ with type $a$ drawn from distribution $x$. This gives rise to (1.14).

**Proposition 1.1. [Multi-colony martingale problem]**
*Without assumption* (1.3)*, for every* $x \in \mathcal{P}(E)^N$*, the martingale problem for* $(L^{(N)}, \mathcal{F}, \delta_x)$ *is well-posed. The unique solution is a strong Markov process with the Feller property.*

### 1.3.3 $C^\Lambda$-process with immigration-emigration: McKean-Vlasov limit

The $N \to \infty$ limit of the $N$-colony model defined in Section 1.3.2 can be described in terms of an independent and identically distributed family of $\mathcal{P}(E)$-valued processes indexed by $\mathbb{N}$. Let us describe the distribution of single member of this family, which can be viewed as a spatial variant of the model in Section 1.3.1 when we add immigration-emigration to/from a cemetery state, with the immigration given by a source that is constant in time. Such processes are of interest in their own right. They are referred to as *McKean-Vlasov processes* for $(c, d, \Lambda, \theta)$, $c, d \in (0, \infty)$, $\Lambda \in \mathcal{M}_f(E)$, $\theta \in \mathcal{P}(E)$, or $C^\Lambda$-processes with immigration-emigration at rate $c$ with source $\theta$ and volatility constant $d$.

Let $\mathcal{F} \subseteq C_\mathrm{b}(\mathcal{P}(E), \mathbb{R})$ be the algebra of functions $F$ of the form

$$(1.15)\quad F(x) = \int_{E^n} x^{\otimes n}(\mathrm{d}u)\, \varphi(u), \qquad x \in \mathcal{P}(E),\, n \in \mathbb{N},\, \varphi \in C_\mathrm{b}(E^n, \mathbb{R}).$$

For $c, d \in (0, \infty)$, $\Lambda \in \mathcal{M}_f([0,1])$ subject to (1.2–1.3) and $\theta \in \mathcal{P}(E)$, let $L_\theta^{c,d,\Lambda}\colon \mathcal{F} \to C_\mathrm{b}(\mathcal{P}(E), \mathbb{R})$ be the linear operator

$$(1.16)\quad L_\theta^{c,d,\Lambda} = L_\theta^c + L^d + L^\Lambda$$

acting on $F \in \mathcal{F}$ as

$$\begin{aligned}
&(L_\theta^c F)(x) = c \int_E (\theta - x)\,(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x}[\delta_a],\\
(1.17)\quad &(L^d F)(x) = d \int_E \int_E Q_x(\mathrm{d}u, \mathrm{d}v)\, \frac{\partial^2 F(x)}{\partial x \partial x}[\delta_u, \delta_v],\\
&(L^\Lambda F)(x) = \int_{[0,1]} \Lambda^*(\mathrm{d}r) \int_E x(\mathrm{d}a)\, \big[F\big((1-r)x + r\delta_a\big) - F(x)\big],
\end{aligned}$$

where

$$(1.18)\quad Q_x(\mathrm{d}u, \mathrm{d}v) = x(\mathrm{d}u)\, \delta_u(\mathrm{d}v) - x(\mathrm{d}u)\, x(\mathrm{d}v)$$

is the Fleming-Viot diffusion function. The three parts of $L_\theta^{c,d,\Lambda}$ correspond to: a *drift* towards $\theta$ of strength $c$ (immigration-emigration), a *Fleming-Viot diffusion* with *volatility* $d$ (Moran resampling), and a $C^\Lambda$-process with resampling measure $\Lambda$ (Cannings resampling). This model arises as the $M \to \infty$ limit of an individual-based model with $M$ individuals at a single site with *immigration* from a constant source with type distribution $\theta \in \mathcal{P}(E)$ and emigration to a cemetery state, both at rate $c$, in addition to the $\Lambda$-resampling.

**Proposition 1.2. [McKean-Vlasov martingale problem]**
*Without assumption* (1.3)*, for every* $x \in \mathcal{P}(E)$*, the martingale problem for* $(L_\theta^{c,d,\Lambda}, \mathcal{F}, \delta_x)$ *is well-posed. The unique solution is a strong Markov process with the Feller property.*

Denote by

$$(1.19)\quad Z_\theta^{c,d,\Lambda} = \big(Z_\theta^{c,d,\Lambda}(t)\big)_{t \geq 0}, \quad Z_\theta^{c,d,\Lambda}(0) = \theta,$$

the solution of the martingale problem in Proposition 1.2 for the special choice $x = \theta$. This is called the *McKean-Vlasov process* with parameters $c, d, \Lambda$ and initial state $\theta$.

## 1.4 The hierarchical Cannings process

The model described in Section 1.3.2 has a *finite* geographical space, an interaction that is mean-field, and a resampling of individuals at the same site. In this section, we introduce two new features into the model:

(1) We consider a *countably infinite* geographic space, namely, the hierarchical group $\Omega_N$ of order $N$, with a migration mechanism that is block-wise exchangeable.

(2) We allow resampling between individuals not only at the same site but also in blocks around a site, which we view as *macro-colonies.*

Both the migration rates and the resampling rates for macro-colonies decay as the distance between the macro-colonies grows. Feature (1) is introduced in Sections 1.4.1–1.4.2, feature (2) in Section 1.4.3. The hierarchical model is defined in Section 1.4.4.

### 1.4.1 Hierarchical group of order $N$

The *hierarchical group* $\Omega_N$ of order $N$ is the set

$$\Omega_N = \Big\{\eta = (\eta^l)_{l\in\mathbb{N}_0} \in \{0,1,\dots,N-1\}^{\mathbb{N}_0}\colon \sum_{l\in\mathbb{N}_0} \eta^l < \infty\Big\}, \qquad N \in \mathbb{N}\backslash\{1\}, \tag{1.20}$$

endowed with the addition operation $+$ defined by $(\eta+\zeta)^l = \eta^l + \zeta^l \pmod N$, $l \in \mathbb{N}_0$. In other words, $\Omega_N$ is the direct sum of the cyclical group of order $N$, a fact that is important for the application of Fourier analysis. The group $\Omega_N$ is equipped with the ultrametric distance $d(\cdot,\cdot)$ defined by

$$d(\eta,\zeta) = d(0,\eta-\zeta) = \min\{k\in\mathbb{N}_0\colon \eta^l = \zeta^l, \text{ for all } l \geq k\}, \qquad \eta,\zeta\in\Omega_N. \tag{1.21}$$

Let

$$B_k(\eta) = \{\zeta\in\Omega_N\colon d(\eta,\zeta)\leq k\}, \qquad \eta\in\Omega_N,\ k\in\mathbb{N}_0, \tag{1.22}$$

denote the $k$-block around $\eta$, which we think of as a *macro-colony.* The geometry of $\Omega_N$ is explained in Fig. 3).

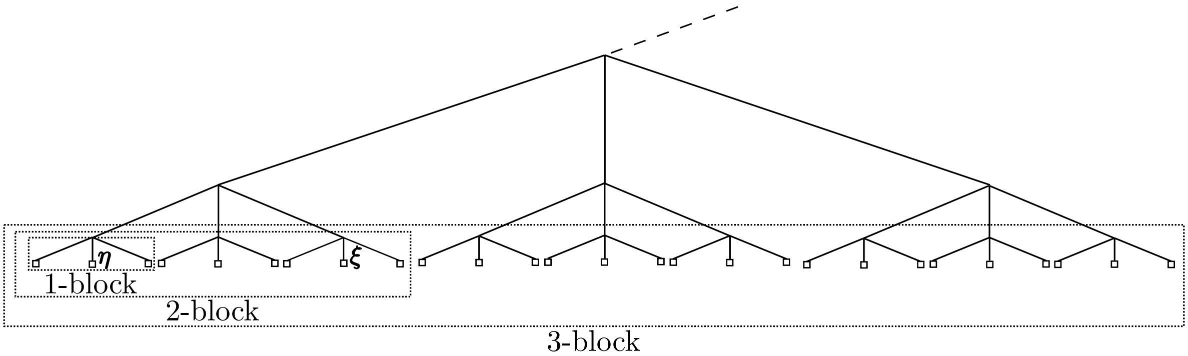

Figure 3: Close-ups of a 1-block, a 2-block and a 3-block in the hierarchical group of order $N = 3$. The elements of the group are the leaves of the tree (□). The hierarchical distance between two elements is the graph distance to the most recent common ancestor: $d(\xi,\eta) = 2$ for $\xi$ and $\eta$ in the picture.

We construct a process

$$(1.23)\quad X^{(\Omega_N)} = \big(X^{(\Omega_N)}(t)\big)_{t\geq 0} \quad \text{with} \quad X^{(\Omega_N)}(t) = \big\{X_\eta^{(\Omega_N)}(t)\big\}_{\eta\in\Omega_N} \in \mathcal{P}(E)^{\Omega_N},$$

by using the same evolution mechanism as for the multi-colony system in Section 1.3.2, except that we replace the migration on $\{0,\dots,N-1\}$ by a migration on $\Omega_N$, and the resampling acting in each colony by a resampling in each of the macro-colonies. On $\mathcal{P}(E)^{\Omega_N}$, we again choose the product of the weak topology on $\mathcal{P}(E)$ as the basic topology.

### 1.4.2 Block migration

We introduce migration on $\Omega_N$ through a random walk kernel. For that purpose, we introduce a sequence of migration rates

$$(1.24)\quad \underline{c} = (c_k)_{k\in\mathbb{N}_0} \in (0,\infty)^{\mathbb{N}_0},$$

and we let the individuals *migrate* as follows:

- Each individual, for every $k\in\mathbb{N}$, chooses at rate $c_{k-1}/N^{k-1}$ the block of radius $k$ around its present location and jumps to a location uniformly chosen at random in that block.

The transition rates of the random walk that is thus performed by each individual are

$$(1.25)\quad a^{(N)}(\eta,\zeta) = \sum_{k\geq d(\eta,\zeta)} \frac{c_{k-1}}{N^{2k-1}}, \qquad \eta,\zeta\in\Omega_N,\ \eta\neq\zeta, \qquad a^{(N)}(\eta,\eta)=0.$$

As shown in Dawson, Gorostiza and Wakolbinger [DGW05], this random walk is recurrent if and only if $\sum_{k\in\mathbb{N}_0}(1/c_k)=\infty$. For the special case where $c_k=c^k$, it is strongly recurrent for $c<1$, critically recurrent for $c=1$, and transient for $c>1$ [3].

Throughout the paper, we assume that [4]

$$(1.26)\quad \limsup_{k\to\infty} \tfrac{1}{k}\log c_k < \infty.$$

This guarantees that the total migration rate per individual is bounded.

### 1.4.3 Block reshuffling-resampling

As we saw in Section 1.3, the idea of the Cannings model is to allow reproduction with an offspring that is of a size comparable to the whole population. Since we have introduced a spatial structure, we now allow, on all hierarchical levels $k$ simultaneously, a reproduction event where each individual treats the $k$-block around its present location as a *macro-colony* and uses it for its resampling. More precisely, we choose a sequence of finite non-negative resampling measures

$$(1.27)\quad \underline{\Lambda} = (\Lambda_k)_{k\in\mathbb{N}_0} \in \mathcal{M}_f([0,1])^{\mathbb{N}_0},$$

---

[3] Loosely speaking, the behaviour is like that of simple random walk on $\mathbb{Z}^d$ with $d<2$, $d=2$ and $d>2$, respectively. More precisely, with the help of potential theory it is possible to associate with the random walk a dimension as a function of $c$ and $N$ that for $N\to\infty$ converges to 2. This shows that in the limit as $N\to\infty$, the potential theory of the hierarchical random walk given by (1.25) choosing $c=1$ is similar to that of simple random walk on $\mathbb{Z}^2$.

[4] In Section 1.6 we will analyse the case $N<\infty$, where (1.26) must be replaced by $\limsup_{k\to\infty}\frac{1}{k}\log c_k < N$.

each subject to (1.2). Assume in addition that

$$(1.28)\quad \int_0^1 \Lambda_k^*(\mathrm{d}r) < \infty, \qquad k \in \mathbb{N},$$

and that $\Lambda_0$ satisfies (1.3). Set

$$(1.29)\quad \lambda_k = \Lambda_k([0,1]), \qquad k \in \mathbb{N}_0.$$

We let individuals *reshuffle-resample* by carrying out the following two steps at once (the formal definition requires the use of a suitable Poisson point process, cf., (2.26), (1.5) and (1.6)):

- For every $\eta \in \Omega_N$ and $k \in \mathbb{N}_0$, choose the block $B_k(\eta)$ at rate $1/N^{2k}$.
- Each individual in $B_k(\eta)$ is first moved to a uniformly chosen random location in $B_k(\eta)$, i.e., a reshuffling takes place (see Fig. 4). After that, $r$ is drawn according to the intensity measure $\Lambda_k^*$ (recall (1.5)), and with probability $r$ each of the individuals in $B_k(\eta)$ is replaced by an individual of type $a$, with $a$ drawn according to the type distribution in $B_k(\eta)$, i.e.,

  $$(1.30)\quad y_{\eta,k} = N^{-k} \sum_{\zeta \in B_k(\eta)} x_\zeta.$$

Note that the reshuffling-resampling affects all the individuals in a macro-colony simultaneously and in the same manner. The reshuffling-resampling occurs at all levels $k \in \mathbb{N}_0$, at a rate that is fastest in single colonies and gets slower as the level $k$ of the macro-colony increases. [5]

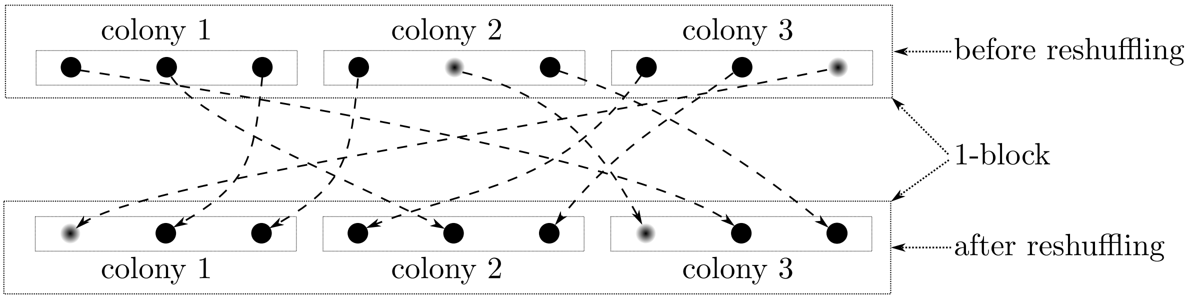

Figure 4: Random reshuffling in a 1-block on the hierarchical lattice of order $N = 3$ with $M = 3$ individuals of two types per colony.

Throughout the paper, we assume that $\underline{\lambda} = (\lambda_k)_{k\in\mathbb{N}_0}$ satisfies [6]

$$(1.31)\quad \limsup_{k\to\infty} \tfrac{1}{k} \log \lambda_k < \infty.$$

Note that each of the $N^k$ colonies in a $k$-block can trigger reshuffling-resampling in that block, and for each colony the block is chosen at rate $N^{-2k}$. Therefore (1.31) guarantees that the total resampling rate per individual is bounded.

---

[5] Because the reshuffling is done first, the resampling always acts on a uniformly distributed state ("panmictic resampling").

[6] In Section 1.6 we will analyse the case $N < \infty$, where (1.31) must be replaced by $\limsup_{k\to\infty} \frac{1}{k} \log \lambda_k < N$.

We note that in the continuum mass limit the reshuffling-resampling operation takes the following form when it acts on the states in the colonies:

$$x_\zeta \text{ is replaced by } (1-r)y_{\eta,k} + r\delta_a \text{ for all } \zeta \in B_k(\eta) \tag{1.32}$$

with $a \in E$ drawn from $y_{\eta,k}$. Note that in the mean-field case and in the single-colony case, $a \in E$ is drawn from $x_\zeta$ (cf. (1.14)) [7].

### 1.4.4 Hierarchical Cannings process

We are now ready to formally define our system of *hierarchically interacting $C^\Lambda$-processes* in terms of a martingale problem. This is the continuum-mass limit ($M \to \infty$) of the individual-based model that we described in Sections 1.4.1–1.4.3. Recall that so far we have considered block migration and block reshuffling-resampling on the hierarchical group of fixed order $N$, starting with $M$ individuals at each site.

We equip the set $\mathcal{P}(E)^{\Omega_N}$ with the product topology to get a state space that is Polish. Let $\mathcal{F} \subset C_\mathrm{b}\big(\mathcal{P}(E)^{\Omega_N}, \mathbb{R}\big)$ be the algebra of functions of the form

$$\begin{aligned} &F(x) = \int_{E^n} \left( \bigotimes_{m=1}^{n} x_{\eta_m}(\mathrm{d}u^m) \right) \varphi\big(u^1, \dots, u^n\big), \quad x = (x_\eta)_{\eta\in\Omega_N} \in \mathcal{P}(E)^{\Omega_N}, \\ &n \in \mathbb{N}, \quad \varphi \in C_\mathrm{b}(E^n, \mathbb{R}), \quad \eta_1, \dots, \eta_m \in \Omega_N. \end{aligned} \tag{1.33}$$

The linear operator for the martingale problem

$$L^{(\Omega_N)} \colon \mathcal{F} \to C_\mathrm{b}\big(\mathcal{P}(E)^{\Omega_N}, \mathbb{R}\big) \tag{1.34}$$

again has two parts,

$$L^{(\Omega_N)} = L^{(\Omega_N)}_{\mathrm{mig}} + L^{(\Omega_N)}_{\mathrm{res}}. \tag{1.35}$$

The *migration operator* is given by

$$(L^{(\Omega_N)}_{\mathrm{mig}} F)(x) = \sum_{\eta,\zeta\in\Omega_N} a^{(N)}(\eta,\zeta) \int_E (x_\zeta - x_\eta)(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\eta}[\delta_a] \tag{1.36}$$

and the *reshuffling-resampling operator* by

$$(L^{(\Omega_N)}_{\mathrm{res}} F)(x) = \sum_{\eta\in\Omega_N} \sum_{k\in\mathbb{N}_0} N^{-2k} \int_{[0,1]} \Lambda^*_k(\mathrm{d}r) \int_E y_{\eta,k}(\mathrm{d}a)\, \left[F\left(\Phi_{r,a,B_k(\eta)}(x)\right) - F(x)\right], \tag{1.37}$$

where $\Phi_{r,a,B_k(\eta)} \colon \mathcal{P}(E)^{\Omega_N} \to \mathcal{P}(E)^{\Omega_N}$ is the *reshuffling-resampling map* acting as

$$\Big[\big(\Phi_{r,a,B_k(\eta)}\big)(x)\Big]_\zeta = \begin{cases} (1-r)y_{\eta,k} + r\delta_a, & \zeta \in B_k(\eta), \\ x_\zeta, & \zeta \notin B_k(\eta), \end{cases} \qquad r \in [0,1],\ a \in E,\ k \in \mathbb{N}_0,\ \eta \in \Omega_N. \tag{1.38}$$

Note that the right-hand side of (1.37) is well defined due to (1.28).

---

[7] Reshuffling is a parallel update affecting all individuals in a macro-colony. Therefore it cannot be seen as a migration of individuals equipped with independent clocks.

**Proposition 1.3. [Hierarchical martingale problem]**
*Without assumption* (1.3)*, for every* $\Theta \in \mathcal{P}(E)^{\Omega_N}$*, the martingale problem for* $(L^{(\Omega_N)}, \mathcal{F}, \delta_\Theta)$ *is well-posed. The unique solution is a strong Markov process with the Feller property.*

The Markov process arising as the solution of the above martingale problem is denoted by $X^{(\Omega_N)} = (X^{(\Omega_N)}(t))_{t\geq 0}$, and is referred to as the $C_N^{\underline{c},\underline{\Lambda}}$-process on $\Omega_N$.

**Remark:** For the analysis of the $C_N^{\underline{c},\underline{\Lambda}}$-process, the following auxiliary models will be important later on. Given $K \in \mathbb{N}_0$, consider the finite geographical space

$$(1.39) \quad G_{N,K} = \{0, \dots, N-1\}^K,$$

which is a truncation of the hierarchical group $\Omega_N$ after $K$ levels. Equip $G_{N,K}$ with coordinate-wise addition modulo $N$, which turns it into a finite Abelian group. By restricting the migration and the resampling to $G_{N,K}$ (i.e., by setting $c_k = 0$ and $\Lambda_k = 0$ for $k \geq K$), we obtain a Markov process with geographic space $G_{N,K}$ that can be characterised by a martingale problem as well. In the limit as $K \to \infty$, this Markov process can be used to approximate the $C_N^{\underline{c},\underline{\Lambda}}$-process.

## 1.5 Main results for $N \to \infty$

Our first set of main results concern a multiscale analysis of $X^{(\Omega_N)}$ in the limit as $N \to \infty$. To that end, we introduce *renormalised* systems with the proper *space-time* scaling.

For each $k \in \mathbb{N}_0$, we look at the *k-block averages* defined by

$$(1.40) \quad Y^{(\Omega_N)}_{\eta,k}(t) = \frac{1}{N^k} \sum_{\zeta \in B_k(\eta)} X^{(\Omega_N)}_\zeta(t), \qquad \eta \in \Omega_N,$$

which constitute a *renormalisation of space* where the component $\eta$ is replaced by the average in $B_k(\eta)$. The corresponding *renormalisation of time* is to replace $t$ by $tN^k$, i.e., $t$ is the associated macroscopic time variable. For each $k \in \mathbb{N}_0$ and $\eta \in \Omega_N$, we can thus introduce a *renormalised* interacting system

$$(1.41) \quad \left( \left( Y^{(\Omega_N)}_{\eta,k}(tN^k) \right)_{\eta \in \Omega_N} \right)_{t \geq 0},$$

which is constant in $B_k(\eta)$ and can be viewed as an interacting system indexed by the set $\Omega_N^{(k)}$ that is obtained from $\Omega_N$ by dropping the first $k$-entries of $\eta \in \Omega_N$. This provides us with a *sequence of renormalised interacting systems*, which for fixed $N$ are however *not* Markov.

Our main results are stated in Sections 1.5.1–1.5.5. In Section 1.5.1, we state the scaling behaviour of the renormalised interacting system in (1.41) as $N \to \infty$ for fixed $k \in \mathbb{N}_0$. In Section 1.5.2, we compare the result with the hierarchical Fleming-Viot process. In Sections 1.5.3–1.5.4, we identify the different regimes for $k \to \infty$. In Section 1.5.5, we look at the interaction chain that captures the scaling behaviour on all scales simultaneously.

### 1.5.1 The hierarchical mean-field limit

Our first main theorem identifies the scaling behaviour of $X^{(\Omega_N)}$ as $N \to \infty$ (the so-called hierarchical mean-field limit) for every fixed block scale $k \in \mathbb{N}_0$. We assume that, for each $N$, the law of $X^{(\Omega_N)}(0)$ is the restriction to $\Omega_N$ of a random field $X$ indexed by $\Omega_\infty = \bigoplus_{\mathbb{N}} \mathbb{N}$ that is taken to be i.i.d. with a single-site mean $\theta$ for some $\theta \in \mathcal{P}(E)$.

Recall (1.29). Let $\underline{d} = (d_k)_{k\in\mathbb{N}_0}$ be the sequence of *volatility constants* defined recursively as

$$(1.42)\quad d_0 = 0, \qquad d_{k+1} = \frac{c_k(\frac{1}{2}\lambda_k + d_k)}{c_k + (\frac{1}{2}\lambda_k + d_k)}, \quad k \in \mathbb{N}_0.$$

Let $\mathcal{L}$ denote law, let $\Longrightarrow$ denote weak convergence on path space, and recall (1.19).

**Theorem 1.4. [Hierarchical mean-field limit and renormalisation]**
*For every $k \in \mathbb{N}_0$, uniformly in $\eta \in \Omega_\infty$,*

$$(1.43)\quad \mathcal{L}\left[\left(Y^{(\Omega_N)}_{\eta,k}(tN^k)\right)_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(Z^{c_k,d_k,\Lambda_k}_{\theta}(t)\right)_{t\geq 0}\right].$$

The limiting process in (1.43) is a McKean-Vlasov process with drift constant $c = c_k$ and resampling measure $d_k\delta_0 + \Lambda_k$. This shows that the class of Cannings models with block resampling is preserved under the renormalisation.

We will see in Section 1.5.5 that the large-scale behaviour of $X^{(\Omega_N)}$ is determined by the sequence $\underline{m} = (m_k)_{k\in\mathbb{N}_0}$ with

$$(1.44)\quad m_k = \frac{\mu_k + d_k}{c_k}, \text{ where } \mu_k = \tfrac{1}{2}\lambda_k.$$

We will argue that the dichotomy

$$(1.45)\quad \sum_{k\in\mathbb{N}_0} m_k = \infty \quad \text{vs.} \quad \sum_{k\in\mathbb{N}_0} m_k < \infty$$

represents qualitatively different situations for the interacting system $X^{(\Omega_N)}$ corresponding to, respectively,

- *clustering* (= formation of large mono-type regions),
- *local coexistence* (= convergence to multi-type equilibria).

(See Section 1.5.5 for more precise definitions.) In the clustering regime the scaling behaviour of $d_k$ is independent of $d_0$, while in the local coexistence regime it depends on $d_0$ (see Section 1.5.5).

For the classical case of hierarchically interacting Fleming-Viot diffusions (i.e., in the absence of block reshuffling-resampling), the dichotomy in (1.45) reduces to

$$(1.46)\quad \sum_{k\in\mathbb{N}_0} (1/c_k) = \infty \quad \text{vs.} \quad \sum_{k\in\mathbb{N}_0} (1/c_k) < \infty,$$

corresponding to the random walk with migration coefficients $\underline{c} = (c_k)_{k\in\mathbb{N}_0}$ being recurrent, respectively, transient. Moreover, it is known that in the clustering regime $\lim_{k\to\infty} \sigma_k d_k = 1$ with $\sigma_k = \sum_{l=0}^{k-1}(1/c_l)$ for all $d_0$.

### 1.5.2 Comparison with the dichotomy for the hierarchical Fleming-Viot process

Our second main theorem provides a comparison of the clustering vs. coexistence dichotomy with the one for the hierarchical Fleming-Viot process. Let

$$(1.47)\quad \underline{d}^* = (d_k^*)_{k\in\mathbb{N}_0}$$

be the sequence of volatility constants when $\mu_0 > 0$ and $\mu_k = 0$ for all $k \in \mathbb{N}$, i.e., there is resampling in single colonies but not in macro-colonies. By (1.42), this sequence has initial value $d_0^* = 0$ and satisfies the recursion relation

$$(1.48)\quad d_1^* = d_1 = \frac{c_0\mu_0}{c_0+\mu_0}, \qquad \frac{1}{d_{k+1}^*} = \frac{1}{c_k} + \frac{1}{d_k^*}, \quad k \in \mathbb{N},$$

whose solution is

$$(1.49)\quad d_k^* = \frac{\mu_0}{1+\mu_0\sigma_k}, \quad k \in \mathbb{N}, \quad \text{with } \sigma_k = \sum_{l=0}^{k-1} \frac{1}{c_l}.$$

**Theorem 1.5. [Comparison with hierarchical Fleming-Viot]**
*The following hold for* $(d_k)_{k\in\mathbb{N}_0}$*:*

- (a) *The maps* $\underline{c} \mapsto \underline{d}$ *and* $\underline{\mu} \mapsto \underline{d}$ *are component-wise non-decreasing.*
- (b) $d_k \geq d_k^*$ *for all* $k \in \mathbb{N}$.
- (c) $\sum_{k\in\mathbb{N}_0} m_k = \infty$ *if and only if* $\sum_{k\in\mathbb{N}_0} (1/c_k) \sum_{l=0}^k \mu_l = \infty$.
- (d) *If* $\lim_{k\to\infty} \sigma_k = \infty$ *and* $\sum_{k\in\mathbb{N}} \sigma_k\mu_k < \infty$*, then* $\lim_{k\to\infty} \sigma_k d_k = 1$.

In words, (a) and (b) say that both migration and reshuffling-resampling increase volatility (recall (1.44–1.45)), (c) says that the dichotomy in (1.46) due to migration is affected by reshuffling-resampling only when the latter is strong enough, i.e., when $\sum_{k\in\mathbb{N}_0} \mu_k = \infty$, while (d) says that the scaling behaviour of $d_k$ in the clustering regime is unaffected by the reshuffling-resampling when the latter is weak enough, i.e., when $\sum_{k\in\mathbb{N}} \sigma_k\mu_k < \infty$. Note that the criterion in (c) shows say that migration tends to inhibit clustering while reshuffling-resampling tends to enhance clustering.

We will see in Section 11.1 that in the *local coexistence regime* $d_k \sim \sum_{l=0}^k \mu_l$ as $k \to \infty$ when this sum diverges and $d_k \to \sum_{l\in\mathbb{N}_0} \mu_l / \prod_{j=l}^\infty (1+m_j) \in (0,\infty)$ when it converges. Thus, in the local coexistence regime the scaling of $d_k$ is determined the resampling-reshuffling.

In the regime where the system *clusters*, i.e., $\sum_{k\in\mathbb{N}_0} m_k = \infty$, it is important to be able to say more about the behaviour of $m_k$ as $k \to \infty$ in order to understand the patterns of cluster formation. For this the key is the behaviour of $d_k$ as $k \to \infty$, which we study in Sections 1.5.3–1.5.4 for polynomial, respectively, exponential growth of the coefficients $c_k$ and $\lambda_k$.

### 1.5.3 Scaling in the clustering regime: polynomial coefficients

Our third main theorem identifies the scaling behaviour of $d_k$ as $k \to \infty$ in four different regimes, defined by the relative size of the migration coefficient $c_k$ versus the block resampling coefficient $\lambda_k$. The necessary *regularity conditions* are stated in (1.55–1.58) below.

Define

$$(1.50)\quad \lim_{k\to\infty} \frac{\mu_k}{c_k} = K \in [0,\infty] \text{ and, if } K = 0, \text{ also } \lim_{k\to\infty} k^2 \frac{\mu_k}{c_k} = L \in [0,\infty].$$

**Theorem 1.6. [Scaling of the volatility in the clustering regime: polynomial coefficients]**
*Assume that the regularity conditions* (1.55–1.58) *hold.*

(a) *If* $K = \infty$, *then*

$$\lim_{k\to\infty} \frac{d_k}{c_k} = 1. \tag{1.51}$$

(b) *If* $K \in (0,\infty)$, *then*

$$\lim_{k\to\infty} \frac{d_k}{c_k} = M \text{ with } M = \tfrac{1}{2}K\left[-1+\sqrt{1+(4/K)}\right] \in (0,1). \tag{1.52}$$

(c) *If* $K = 0$ *and* $L = \infty$, *then*

$$\lim_{k\to\infty} \frac{d_k}{\sqrt{c_k\mu_k}} = 1. \tag{1.53}$$

(d) *If* $K = 0$, $L < \infty$ *and* $a \in (-\infty,1)$, *then*

$$\lim_{k\to\infty} \sigma_k d_k = \bar{N} \text{ with } \bar{N} = \tfrac{1}{2}\left[1+\sqrt{1+4L/(1-a)^2}\right] \in [1,\infty). \tag{1.54}$$

The meaning of these four regimes for the evolution of the population will be explained in Corollary 1.10.

**Regularity conditions.** In Theorem 1.6, we need to impose some mild regularity conditions on $\underline{c}$ and $\underline{\mu}$, which we collect in (1.55–1.58) below. We require that both $c_k$ and $\mu_k$ are *regularly varying at infinity*, i.e.,

$$c_k \sim L_c(k)k^a, \quad a\in\mathbb{R}, \qquad \mu_k \sim L_\mu(k)k^b, \quad b\in\mathbb{R}, \qquad k\to\infty, \tag{1.55}$$

with $L_c, L_\mu$ slowly varying at infinity (Bingham, Goldie and Teugels [BGT87, Section 1.9]). The numbers $a, b$ are referred to as the *indices* of $\underline{c}$ and $\underline{\mu}$ [8].

To handle the boundary cases, where $c_k$, $\mu_k$, $\mu_k/c_k$ and/or $k^2\mu_k/c_k$ are slowly varying, we additionally require that for specific choices of the indices the following functions are *asymptotically monotone*:

$$\begin{aligned} a=0: &\quad k\mapsto \Delta L_c(k)/L_c(k),\ k\mapsto k\Delta L_c(k)/L_c(k),\\ b=0: &\quad k\mapsto \Delta L_\mu(k)/L_\mu(k),\ k\mapsto k\Delta L_\mu(k)/L_\mu(k), \end{aligned} \tag{1.56}$$

and the following functions are *bounded*:

$$\begin{aligned} a=0: &\quad k\mapsto k\Delta L_c(k)/L_c(k),\\ b=0: &\quad k\mapsto k\Delta L_\mu(k)/L_\mu(k), \end{aligned} \tag{1.57}$$

where $\Delta L(k) = L(k+1) - L(k)$. To ensure the *existence* of the limits in (1.50), we also need the following functions to be *asymptotically monotone*:

$$\begin{aligned} a=b: &\quad k\mapsto L_\mu(k)/L_c(k),\\ a=b-2: &\quad k\mapsto k^2L_\mu(k)/L_c(k). \end{aligned} \tag{1.58}$$

---

[8] Regular variation is typically defined with respect to a continuous instead of a discrete variable. However, every regularly varying sequence can be embedded into a regularly varying function.

### 1.5.4 Scaling in the clustering regime: exponential coefficients

We briefly indicate how Theorem 1.6 extends when $c_k$ and $\mu_k$ satisfy

$$\begin{aligned} &c_k = c^k \bar{c}_k,\ \mu_k = \mu^k \bar{\mu}_k \text{ with } c,\mu \in (0,\infty) \text{ and } (\bar{c}_k),\ (\bar{\mu}_k) \text{ regularly varying at infinity,} \\ &\bar{K} = \lim_{k\to\infty} \frac{\bar{\mu}_k}{\bar{c}_k} \in [0,\infty], \end{aligned} \tag{1.59}$$

and the analogues of (1.56–1.58) apply to the regularly varying parts.

**Theorem 1.7. [Scaling of the volatility in the clustering regime: exponential coefficients]**
*Assume that* (1.59) *holds. Then:*

(A) **[like Case (a)]** *$c < \mu$ or $c = \mu$, $\bar{K} = \infty$:* $\lim_{k\to\infty} d_k/c_k = 1/c$.

(B) **[like Case (b)]** *$c = \mu$, $\bar{K} \in (0,\infty)$:* $\lim_{k\to\infty} d_k/c_k = \bar{M}$ *with*

$$\bar{M} = \frac{1}{2c}\left[-(c(\bar{K}+1)-1) + \sqrt{(c(\bar{K}+1)-1)^2 + 4c\bar{K}}\right]. \tag{1.60}$$

(C) *The remainder $c > \mu$ or $c = \mu$, $\bar{K} = 0$ splits into three cases:*

- (C1) **[like Case (d)]** *$1 > c > \mu$ or $1 = c > \mu$,* $\lim_{k\to\infty} \sigma_k = \infty$*:* $\lim_{k\to\infty} \sigma_k d_k = 1$.
- (C2) **[like Case (b)]** *$c = \mu < 1$, $\bar{K} = 0$:* $\lim_{k\to\infty} d_k/c_k = (1-c)/c$.
- (C3) **[like Case (c)]** *$c = \mu > 1$, $\bar{K} = 0$:* $\lim_{k\to\infty} d_k/\mu_k = 1/(\mu - 1)$.

The choices $1 = c > \mu$, $\lim_{k\to\infty} \sigma_k < \infty$ and $c > 1$, $c > \mu$ correspond to local coexistence (and so does $c = \mu > 1$, $\bar{K} = 0$, $\sum_{k\in\mathbb{N}_0} \bar{\mu}_k/\bar{c}_k < \infty$).

### 1.5.5 Multi-scale analysis: the interaction chain

**Multi-scale behaviour.** Our fourth main theorem looks at the implications of the scaling behaviour of $d_k$ as $k \to \infty$ described in Theorems 1.5–1.6, for which we must extend Theorem 1.4 to include *multi-scale renormalisation.* This is done by considering *two* indices $(j,k)$ and introducing an appropriate multi-scale limiting process, called the *interaction chain*

$$M^{(j)} = (M_k^{(j)})_{k=-(j+1),\dots,0}, \qquad j \in \mathbb{N}_0, \tag{1.61}$$

which describes all the block averages of size $N^k$ indexed by $k = -(j+1),\dots,0$ *simultaneously* at time $N^j t$ with $j \in \mathbb{N}_0$ fixed. Formally, the interaction chain is defined as the time-inhomogeneous Markov chain with a prescribed initial state at time $-(j+1)$,

$$M^{(j)}_{-(j+1)} = \theta \in \mathcal{P}(E), \tag{1.62}$$

and with transition kernel

$$K_k(x,\cdot) = \nu_x^{c_k,d_k,\Lambda_k}(\cdot), \quad x \in \mathcal{P}(E),\ k \in \mathbb{N}_0, \tag{1.63}$$

for the transition from time $-(k+1)$ to time $-k$ (for $k = j,\dots,0$). Here, $\nu_x^{c,d,\Lambda}$ is the unique equilibrium of the McKean-Vlasov process $Z_x^{c,d,\Lambda}$ defined in Section 1.3.3 (see Section 4 for details).

**Theorem 1.8. [Multi-scale behaviour]**
*Let* $(t_N)_{N\in\mathbb{N}}$ *be such that*

$$\lim_{N\to\infty} t_N = \infty \text{ and } \lim_{N\to\infty} t_N/N = 0. \tag{1.64}$$

*Then, for every* $j \in \mathbb{N}_0$*, uniformly in* $\eta \in \Omega_\infty$ *and* $u_k \in (0,\infty)$*,* $k = 0,\dots,j$*,*

$$\begin{aligned} &\mathcal{L}\left[\left(Y^{(\Omega_N)}_{\eta,k}(N^j t_N + N^k u_k)\right)_{k=j,\dots,0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(M^{(j)}_{-k}\right)_{k=j,\dots,0}\right], \\ &\mathcal{L}\left[Y^{(\Omega_N)}_{\eta,j+1}(N^j t_N)\right] \underset{N\to\infty}{\Longrightarrow} \delta_\theta. \end{aligned} \tag{1.65}$$

Theorem 1.8 says that, as $N \to \infty$, the system is in a *quasi-equilibrium* $\nu_x^{c_k,d_k,\Lambda_k}$ on time scale $N^j t_N + N^k u$, with $u \in (0,\infty)$ the macroscopic time parameter on level $k$, when $x$ is the average on level $k+1$.

**The basic dichotomy.** Our fifth main theorem lets the index in the multi-scale renormalisation scheme tend to infinity and identifies how the limit depends on the parameters $(\underline{c},\underline{\Lambda})$. Indeed, Theorem 1.8 in combination with Theorems 1.5–1.6 allow us to study the universality properties on large space-time scales when we first let $N \to \infty$ and then $j \to \infty$ [9].

The interaction chain exhibits a *dichotomy*, in the sense that

$$\mathcal{L}\left[M_0^{(j)}\right] \underset{j\to\infty}{\Longrightarrow} \nu_\theta \in \mathcal{P}(\mathcal{P}(E)), \tag{1.66}$$

with $\nu_\theta$ either of the form of a random point measure, i.e.,

$$\nu_\theta = \mathcal{L}[\delta_U], \text{ for some random } U \in E \text{ with } \mathcal{L}[U] = \theta, \tag{1.67}$$

or $\nu_\theta$ spread out, i.e.,

$$\sup_{\psi\in B_1} \mathbb{E}_{\nu_\theta}[\mathrm{Var}_x(\psi)] > 0, \tag{1.68}$$

where $B_1 = C_\mathrm{b}(E,\mathbb{R}) \cap \{\psi\colon\ |\psi| \leq 1\}$ and

$$\mathbb{E}_{\nu_\theta}[\mathrm{Var}_x(\psi)] = \int_{\mathcal{P}(E)} \nu_\theta(\mathrm{d}x)\,\mathrm{Var}_x(\psi) \tag{1.69}$$

with

$$\mathrm{Var}_x(\psi) = \int_{E\times E} [x(\mathrm{d}u)\delta_u(\mathrm{d}v) - x(\mathrm{d}u)x(\mathrm{d}v)]\,\psi(u)\psi(v). \tag{1.70}$$

The first case is called the *clustering regime*, since it indicates the formation of large mono-type regions, while the second case is called the *local coexistence regime*, since it indicates the formation of multi-type local equilibria under which different types can live next to each other with a positive probability. In the local coexistence regime, a remarkable difference occurs with the hierarchical Fleming-Viot process: mono-type regions for $M_0^{(j)}$ as $j \to \infty$ have a probability in the open interval $(0,1)$ rather than probability 0. The latter is referred

---

[9] For several previously investigated systems, the limit as $j \to \infty$ was shown to be interchangeable (Dawson, Greven and Vaillancourt [DGV95], Fleischmann and Greven [FG94]).)

to in [DGV95] by saying that the system is in the *stable regime* (which is stronger than local coexistence). In the present paper, we do not identify the conditions on $\underline{c}$ and $\underline{\lambda}$ that correspond to the stable regime. The dichotomy can be conveniently rephrased as follows: There is either a trivial or a non-trivial *entrance law* for the interaction chain with initial state $\theta \in \mathcal{P}(E)$ at time $-\infty$. [10]

We will show in Section 4.4 that

$$\text{(1.71)}\quad \mathbb{E}_{\mathcal{L}[M_0^{(j)}]}[\mathrm{Var}_x(\psi)] = \left[\prod_{k=0}^{j} \frac{1}{1+m_k}\right] \mathrm{Var}_\theta(\psi), \quad j \in \mathbb{N}_0,\ \psi \in C_{\mathrm{b}}(E,\mathbb{R}),\ \theta \in \mathcal{P}(E).$$

This shows that the entrance law is trivial when $\sum_{k\in\mathbb{N}_0} m_k = \infty$ and non-trivial when $\sum_{k\in\mathbb{N}_0} m_k < \infty$.

**Theorem 1.9. [Dichotomy of the entrance law]**

(a) *The interaction chain converges to an entrance law:*

$$\text{(1.72)}\quad \begin{cases} \mathcal{L}\left[\left(M_k^{(j)}\right)_{k=-(j+1),\dots,0}\right] \underset{j\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(M_k^{(\infty)}\right)_{k=-\infty,\dots,0}\right], \\ M_{-\infty}^{(\infty)} = \theta. \end{cases}$$

(b) **[Clustering]** *If* $\sum_{k\in\mathbb{N}_0} m_k = \infty$, *then* $\mathcal{L}[M_0^{(j)}] \underset{j\to\infty}{\Longrightarrow} \mathcal{L}[\delta_U]$ *with* $\mathcal{L}[U] = \theta$.

(c) **[Local coexistence]** *If* $\sum_{k\in\mathbb{N}_0} m_k < \infty$, *then* $\sup_{\psi\in C_{\mathrm{b}}(E,\mathbb{R})} \mathbb{E}_{\mathcal{L}[M_0^{(\infty)}]}[\mathrm{Var}_x(\psi)] > 0$.

Theorem 1.9 in combination with Theorem 1.5 (c) says that, like for Fleming-Viot diffusions, we have a clear-cut criterion for the two regimes in terms of the migration coefficients and the resampling coefficients.

**Scaling of the variance.** Our sixth main theorem shows what the scaling of $d_k$ in Theorem 1.6 implies for the scaling of $m_k$ and hence of the variance in (1.71) (we will see in Section 11.3 that the conditions for Case (d) imply that $\lim_{k\to\infty} \mu_k\sigma_k = 0$ and $\lim_{k\to\infty} c_k\sigma_k = \infty$).

**Corollary 1.10. [Scaling behaviour of $m_k$]**
*The following asymptotics of $m_k$ for $k \to \infty$ holds in the four cases of Theorem* 1.6*:*

$$\text{(1.73)}\quad \begin{aligned} &\text{(a)}\quad m_k \sim \frac{\mu_k}{c_k} \to \infty, \qquad \text{(b)}\quad m_k \to K + M,\\ &\text{(c)}\quad m_k \sim \sqrt{\frac{\mu_k}{c_k}} \to 0, \qquad \text{(d)}\quad m_k \sim \frac{\bar{N}}{c_k\sigma_k} \to 0. \end{aligned}$$

All four cases fall in the clustering regime. For the variance in (1.71) they imply: (a) superexponential decay; (b) exponential decay, (c–d) subexponential decay.

Note that Case (d) also falls in the *clustering regime* because it assumes that $a \in (-\infty, 1)$, which implies that $\lim_{k\to\infty} \sigma_k = \infty$. Indeed, $1/c_k\sigma_k = (\sigma_{k+1} - \sigma_k)/\sigma_k$, and in Section 11.1 we will see that

$$\text{(1.74)}\quad \lim_{k\to\infty} \sigma_k = \infty \quad \Longleftrightarrow \quad \sum_{k\in\mathbb{N}} \frac{1}{c_k\sigma_k} = \infty.$$

Combining Cases (a–d), we conclude the following:

---

[10] Recall that an entrance law for a sequence of transition kernels $(K_k)_{k=-\infty}^0$ and an entrance state $\theta$ is any law of a Markov chain $(Y_k)_{k=-\infty}^0$ with these transition kernels such that $\lim_{k\to-\infty} Y_k = \theta$.

- The regime of *weak block resampling* (for which the scaling behaviour of $d_k$ is the same as if there were no block resampling) coincides with the choice $K = 0$ and $L < \infty$.
- The regime of *strong block resampling* (for which the scaling behaviour of $d_k$ is different) coincides with $K = 0$ and $L = \infty$ or $K > 0$.

Note that $M \uparrow 1$ as $K \to \infty$, so that Case (b) connects up with Case (a). Further note that $M \sim \sqrt{K}$ as $K \downarrow 0$, so that Case (b) also connects up with Case (c). Finally, note that $\sqrt{c_k \mu_k} \sim \sqrt{L} c_k / k$ as $k \to \infty$ for Case (d) by (1.50), while $c_k \sigma_k \sim k/(1-a)$ as $k \to \infty$ when $a \in (-\infty, 1)$ by (1.55). Hence, Case (d) connects up with Case (c) as well.

**Cluster formation.** In the clustering regime, it is of interest to study the size of the mono-type regions as a function of time, i.e., how do the clusters grow? To that end, we look at the interaction chain $M^{(j)}_{-k}$ for $j \to \infty$ and level scaling $k = k(j)$ for some $k \colon \mathbb{N} \to \mathbb{N}$ with $\lim_{j\to\infty} k(j) = \infty$, suitably chosen such that we obtain a nontrivial limit law. For example, in Dawson and Greven [DG93b] such a result was proved in the case of interacting Fleming-Viot processes when $\underline{c}$ is critically recurrent. Here, different types of limit laws and different types of scaling can occur, corresponding to different clustering regimes. Following Dawson, Greven and Vaillancourt [DGV95] and Dawson and Greven [DG96], it is natural to consider a whole family of scalings $k_\alpha \colon \mathbb{N} \to \mathbb{N}$, $\alpha \in [0,1]$, and single out *fast*, *diffusive* and *slow* clustering regimes, which are defined as follows:

(i) **Fast clustering:** $\lim_{j\to\infty} k_\alpha(j)/j = 1$ for all $\alpha$.

(ii) **Diffusive clustering:** In this regime, $\lim_{j\to\infty} k_\alpha(j)/j = \kappa(\alpha)$ for all $\alpha$, where $\alpha \mapsto \kappa(\alpha)$ is continuous and non-increasing with $\kappa(0) = 1$ and $\kappa(1) = 0$.

(iii) **Slow clustering:** $\lim_{j\to\infty} k_\alpha(j)/j = 0$ for all $\alpha$. This regime borders with the regime of local coexistence.

**Remark:** Diffusive clustering similar to (ii) was previously found for the voter model on $\mathbb{Z}^2$ by Cox and Griffeath [CGr86], where the radii of the clusters of opinion "all 1" or "all 0" scale as $t^{\alpha/2}$ with $\alpha \in [0,1)$, i.e., clusters occur on all scales $\alpha \in [0,1)$. This is different from what happens on $\mathbb{Z}^d$, $d \geq 3$, where clusters occur only on scale $\alpha = 1$. For the model of hierarchically interacting Fleming-Viot diffusions with $c_k \equiv 1$ (= critically recurrent migration), Fleischmann and Greven [FG94]) showed that, for all $N \in \mathbb{N} \setminus \{1\}$ and all $\eta \in \Omega_N$,

$$\mathcal{L}\left[\left(Y^{(\Omega_N)}_{\eta,\lfloor(1-\alpha)t\rfloor}(t)\right)_{\alpha\in[0,1)}\right] \underset{t\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(Y^{(\Omega_N)}\left(\log\left(\frac{1}{1-\alpha}\right)\right)\right)_{\alpha\in[0,1)}\right], \tag{1.75}$$

where $(Y^{(\Omega_N)}(t))_{t\in(0,1]}$ is a time-transformed Fleming-Viot diffusion on $\mathcal{P}(E)$. A similar behaviour occurs for other models, e.g. for branching models (Dawson and Greven [DG96]).

Our next two main theorems show which type of clustering occurs for the various scaling regimes of the coefficients $\underline{c}$ and $\underline{\mu}$ identified in Theorems 1.6–1.7. Polynomial coefficients allow for fast and diffusive clustering only. Exponential coefficients allow for fast, diffusive and slow clustering, with the latter only in a narrow regime.

**Theorem 1.11. [Clustering regimes for polynomial coefficients]**
*Recall the scaling regimes of Theorem* 1.6.

(i) **[Fast clustering]** *In cases* (a-c)*, the system exhibits fast clustering.*

(ii) **[Diffusive clustering]** *In case* (d)*, the system exhibits diffusive clustering, i.e.,*

$$\mathcal{L}\left[\left(M^{(j)}_{-\lfloor(1-\alpha)j\rfloor}\right)_{\alpha\in[0,1)}\right] \underset{j\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(Z^{0,1,0}_{\theta}\left(\log\left(\frac{1}{1-\alpha^{R}}\right)\right)\right)_{\alpha\in[0,1)}\right], \tag{1.76}$$

*where* $R=\bar{N}(1-a)$ *with* $\bar{N}$ *defined in* (1.54) *and* $a$ *the exponent in* (1.55)*.*

**Theorem 1.12. [Clustering regimes for exponential coefficients]**
*Recall the scaling regimes of Theorem* 1.7*.*

(i) **[Fast clustering]** *In cases* (A, B, C1, C2)*, and case* (C3) *with* $\lim_{k\to\infty} k\bar{\mu}_k/\bar{c}_k=\infty$*, the system exhibits fast clustering.*

(ii) **[Diffusive clustering]** *In case* (C3) *with* $\lim_{k\to\infty} k\bar{\mu}_k/\bar{c}_k=C$*, the system exhibits diffusive clustering, i.e.,* (1.76) *holds with* $R=C/(\mu-1)$*.*

(iii) **[Slow clustering]** *In case* (C3) *with* $k\bar{\mu}_k/\bar{c}_k \asymp 1/(\log k)^{\gamma}$*,* $\gamma\in(0,1)$*, the system exhibits slow clustering.*

Note that (1.75) is a statement valid for *all* $N\in\mathbb{N}\setminus\{1\}$. In contrast, Theorems 1.11–1.12 are valid in the hierarchical mean-field limit $N\to\infty$ only. What can we say about the clustering vs. local coexistence dichotomy in our model for finite $N$?

## 1.6 Main results for finite $N$

In this section, we take a look at our system $X^{(\Omega_N)}$ for finite $N$, i.e., without taking the hierarchical mean-field limit. We ask whether this system also exhibits a dichotomy of clustering versus local coexistence, i.e., for fixed $N$ and $t\to\infty$, does $\mathcal{L}[X^{(\Omega_N)}(t)]$ converge to a mono-type state, where the type is distributed according to $\theta$, or to an equilibrium state, where different types live next to each other?

Let $P_t(\cdot,\cdot)$ denote the transition kernel of the random walk on $\Omega_N$ with migration coefficients

$$(c_k+\lambda_{k+1}N^{-(k+1)})_{k\in\mathbb{N}} \tag{1.77}$$

starting at 0. Let

$$\bar{H}_N=\sum_{k\in\mathbb{N}_0}\lambda_k N^{-k}\int_0^\infty P_{2s}(0,B_k(0))\,\mathrm{d}s, \tag{1.78}$$

where $B_k(0)$ is the $k$-block in $\Omega_N$ around 0 (recall (1.22)) and $P_t(0,B_k(0))=\sum_{\zeta\in B_k(0)}P_t(0,\eta)$. We will see in Section 2.4.2 that $\bar{H}_N$ in (1.78) is the *expected hazard* for two partition elements in the spatial $\underline{\Lambda}$-coalescent with block coalescence to coalesce. In particular, the second summand in (1.77) is induced by the reshuffling in the spatial $\underline{\Lambda}$-coalescent with block coalescence.

Our last set of main theorems identify the ergodic behaviour for finite $N$.

**Theorem 1.13. [Dichotomy for finite** $N$**]**
*The following dichotomy holds for every* $N\in\mathbb{N}\backslash\{1\}$*:*

(a) **[Local coexistence]** *If* $\bar{H}_N < \infty$, *then*

$$(1.79)\quad \liminf_{t\to\infty} \sup_{\psi\in B_1} \mathbb{E}_{X_\eta^{(\Omega_N)}(t)}[\mathrm{Var}_x(\psi)] > 0, \quad \textit{for all } \eta \in \Omega_N.$$

(b) **[Clustering]** *If* $\bar{H}_N = \infty$, *then*

$$(1.80)\quad \lim_{t\to\infty} \sup_{\psi\in B_1} \mathbb{E}_{X_\eta^{(\Omega_N)}(t)}[\mathrm{Var}_x(\psi)] = 0, \quad \textit{for all } \eta \in \Omega_N.$$

This dichotomy can be sharpened using duality theory and the complete longtime behaviour of $X^{(\Omega_N)}$ can be identified.

**Theorem 1.14. [Ergodic behaviour for finite** $N$**]**
*The following dichotomy holds:*

(a) **[Local coexistence]** *If* $\bar{H}_N < \infty$, *then for every* $\theta \in \mathcal{P}(E)$ *and every* $X^{(\Omega_N)}(0)$ *whose law is stationary and ergodic w.r.t. translations in* $\Omega_N$ *and has a single-site mean* $\theta$,

$$(1.81)\quad \mathcal{L}\left[X^{(\Omega_N)}(t)\right] \underset{t\to\infty}{\Longrightarrow} \nu_\theta^{(\Omega_N),\underline{c},\underline{\lambda}} \in \mathcal{P}(\mathcal{P}(E)^{\Omega_N})$$

*for some unique law* $\nu_\theta^{(\Omega_N),\underline{c},\underline{\lambda}}$ *that is stationary and ergodic w.r.t. translations in* $\Omega_N$ *and has single-site mean* $\theta$.

(b) **[Clustering]** *If* $\bar{H}_N = \infty$, *then, for every* $\theta \in \mathcal{P}(E)$,

$$(1.82)\quad \mathcal{L}\left[X^{(\Omega_N)}(t)\right] \underset{t\to\infty}{\Longrightarrow} \int_0^1 \theta(\mathrm{d}u)\delta_{(\delta_u)^{\Omega_N}} \in \mathcal{P}(\mathcal{P}(E)^{\Omega_N}).$$

**Theorem 1.15. [Agreement of dichotomy for** $N < \infty$ **and** $N = \infty$**]**
*The dichotomies in Theorems* 1.9 *and* 1.14 *coincide, i.e.,* $\sum_{k\in\mathbb{N}_0} m_k = \infty$ *if and only if* $\bar{H}_N = \infty$.

## 1.7 Discussion

**Summary.** We have constructed the $C_N^{\underline{c},\underline{\Lambda}}$-process, describing hierarchically interacting Cannings processes, and have identified its space-time scaling behaviour in the hierarchical mean field limit $N \to \infty$ (interaction chain). We have fully classified the clustering vs. local coexistence dichotomy in terms of the parameters $\underline{c}, \underline{\Lambda}$ of the model, and found different regimes of cluster formation. Moreover, we have verified the dichotomy also for finite $N$. Our results provide a full generalisation of what was known for hierarchically interacting diffusions, and show that Cannings resampling leads to new phenomena.

**Diverging volatility of the Fleming-Viot part and local coexistence.** The growth of the block resampling rates $(\mu_k)_{k\in\mathbb{N}}$ can lead to a situation, where, as we pass to larger block averages, the volatility of the Fleming-Viot part of the asymptotic limit dynamics diverges, even though on the level of a single component the system exhibits local coexistence. This requires that the migration rates are (barely) transient and the block resampling rate decays very slowly. An example of such a situation is the choice the choice $c_k = k(\log k)^3$ and $\mu_k = 1/k$ which leads to $d_k \sim \log k$ and $m_k \sim 1/k(\log k)^2$ as $k \to \infty$. Thus, the system may be in the local coexistence regime and yet have a diverging volatility on large space-time scales.

**Open problems.** The results of Sections 1.5 and 1.6 suggest that a dichotomy between clustering and local coexistence also holds for a suitably defined Cannings model with non-local resampling on $\mathbb{Z}^d$, $d \geq 3$. In addition, a continuum limit to the geographic space $\mathbb{R}^2$ ought to arise as well. The latter may be easier to investigate in the limit $N \to \infty$, following the approach outlined in Greven [Gre05]. Another open problem concerns the different ways in which cluster formation can occur. Here, the limit $N \to \infty$ could already give a good picture of what is to be expected for finite $N$. A further task is to investigate the genealogical structure of the model, based on the work in Greven, Klimovsky and Winter [GKWpr] for the model without multi-colony Cannings resampling (i.e., $\Lambda_k = \delta_0$ for $k \in \mathbb{N}$).

**Outline.** Section 2 introduces the spatial $\underline{\Lambda}$-coalescent with block coalescence and derives some of its key properties. Sections 3–11 use the results in Section 2 to prove the propositions and the theorems stated in Sections 1.3–1.6. Section 3 handles all issues related to the well-posedness of martingale problems. Section 4 deals with the properties of the McKean-Vlasov process. Section 5 outlines the strategy behind the proofs of the scaling results for the hierarchical Cannings process, which are worked out in Sections 6–9. Section 10 proves the scaling results for the interaction chain. Section 11 derives the scaling results for the volatility constant. Section 12 collects the notation.

## 2 Spatial Λ-coalescent with block-coalescence

In this section, we introduce a new class of spatial $\underline{\Lambda}$-coalescent processes, namely, processes where coalescence of partition elements at distances larger than or equal to zero can occur. This is a generalisation of the spatial coalescent introduced by Limic and Sturm [LS06], which allows for the coalescence of blocks residing at the same location only. Informally, the spatial $\underline{\Lambda}$-coalescent with block coalescence is the process that encodes the *family structure* of a sample from the currently alive population in the $C_N^{c,\underline{\Lambda}}$-process, i.e., it is the process of *coalescing lineages* that occur when the evolution of the spatial $C_N^{c,\underline{\Lambda}}$-Cannings process is traced backwards in time up to a common ancestor. In what follows, we denote this backwards-in-time process by $\mathfrak{C}_N^{c,\underline{\Lambda}}$.

Two Markov processes $X$ and $Y$ with Polish state spaces $\mathcal{E}$ and $\mathcal{E}'$ are called *dual w.r.t. the duality function* $H\colon\ \mathcal{E} \times \mathcal{E}' \to \mathbb{R}$ if

$$(2.1)\quad \mathbb{E}_{X_0}[H(X_t, Y_0)] = \mathbb{E}_{Y_0}[H(X_0, Y_t)] \qquad \text{for all}\,(X_0, Y_0) \in \mathcal{E} \times \mathcal{E}',$$

and if the family $\{H(\cdot, Y_0)\colon\ Y_0 \in \mathcal{E}'\}$ uniquely determines a law on $\mathcal{E}$. Typically, the key point of a duality relation is to translate questions about a complicated process into questions about a simpler process. This translation often allows for an analysis of the long-time behaviour of the process, as well as a proof of existence and uniqueness for associated martingale problems. If $H(\cdot, \cdot) \in C_{\mathrm{b}}(\mathcal{E} \times \mathcal{E}')$, and if $H(\cdot, Y_0)$ and $H(X_0, \cdot)$ are in the domain of the generator of $X$, respectively, $Y$ for all $(X_0, Y_0) \in \mathcal{E} \times \mathcal{E}'$, then it is possible to establish duality by just checking a *generator relation* (see Remark 2.9 below and also Liggett [L85, Section II.3]).

The analysis of the processes on their relevant time scales will lead us to study a number of auxiliary processes on geographic spaces different from $\Omega_N$. The duality will be crucial for the proof of Propositions 1.1–1.3 (martingale well-posedness) in Section 3, and also for statements about the long-time behaviour of the processes and the qualitative properties of their equilibria. In Section 2.1, we define the spatial Λ-coalescent without block coalescence. In Section 2.2, we add block coalescence. In Section 2.3, we formulate and prove the duality relation between the $C_N^{c,\Lambda}$-process and the spatial $\Lambda$-coalescent. In Section 2.4, we look at the long-time behaviour of the $\Lambda$-coalescent.

## 2.1 Spatial $\Lambda$-coalescent without block coalescence

In this section, we briefly recall the definition of the spatial $\Lambda$-coalescent on a countable geographic space $G$ as introduced by Limic and Sturm [LS06]. (For a general discussion of exchangeable coalescents, see Berestycki [B09]). In Section 2.2, we will add *block coalescence*, i.e., coalescence of individuals not necessarily located at the same site.

The following choices of the geographic space $G$ will be needed later on:

(2.2) $G_{N,K} = \{0,\ldots,N-1\}^K,\ K,N \in \mathbb{N}, \qquad G = \Omega_N,\ N \in \mathbb{N}, \qquad G = \{0,*\}, \qquad G = \mathbb{N}.$

The choices in (2.2) correspond to geographic spaces that are needed, respectively, for finite approximations of the hierarchical group, for the hierarchical group, for a single-colony with immigration-emigration, and for the McKean-Vlasov limit. We define the basic transition mechanisms and characterise the process by a martingale problem in order to be able to verify duality and to prove convergence properties. In Section 2.1.1 we define the state space and the evolution rules, in Section 2.1.2 we formulate the martingale problem, while in Section 2.1.3 we introduce coalescents with immigration-emigration.

### 2.1.1 State space, evolution rules, graphical construction and entrance law

**State space.** As with non-spatial exchangeable coalescents, it is convenient to start with finite state spaces and subsequently extend to infinite state spaces via exchangeability. Given $n \in \mathbb{N}$, consider the set

(2.3) $[n] = \{1,\ldots,n\}$

and the set $\Pi_n \subset 2^{[n]}$ of its partitions into partition elements called families:

(2.4) $\Pi_n =$ set of all partitions $\pi = \{\pi_1\}_{i=1}^b$ of $[n]$ into disjoint families $\pi_i \subset [n]$, $i \in [b]$.

Thus, for any $\pi = \{\pi_i\}_{i=1}^b \in \Pi_n$, we have $[n] = \bigcup_{i=1}^b \pi_i$, where $\pi_i \cap \pi_j = \emptyset$ for $i,j \in [b]$ with $i \neq j$. In what follows we denote by

(2.5) $b = b(\pi) \in [n]$

the number of families in $\pi \in \Pi_n$.

**Remark 2.1.** *By a slight abuse of notation, we can associate with $\pi \in \Pi_n$ the mapping $\pi\colon [n] \to [b]$ defined as $\pi(i) = k$, where $k \in [b]$ is such that $i \in \pi_k$. In words, $k$ is the label of the unique family containing $i$.*

The state space of the spatial coalescent is the set of *G-labelled partitions* defined as

(2.6) $\Pi_{G,n} = \Big\{\pi_G = \{(\pi_1,g_1),(\pi_2,g_2),\ldots,(\pi_b,g_b)\}\colon\ \{\pi_1,\ldots,\pi_b\} \in \Pi_n,\ g_1,\ldots,g_b \in G\Big\}.$

For definiteness, we assume that the families of $\pi_G \in \Pi_{G,n}$ are indexed in the increasing order of each family's smallest element, i.e., the enumeration is such that $\min\pi_i < \min\pi_j$ for all $i,j \in [b]$ with $i \neq j$.

Let $S_{G,n} \in \Pi_{G,n}$ denote the labelled partition of $[n]$ into *singletons*, i.e.,

$$(2.7)\quad S_{G,n} = \Big\{(\{1\}, g_1), (\{2\}, g_2), \dots, (\{n\}, g_n)\colon\ g_i \in G, i \in [n]\Big\}.$$

With each $\pi_G \in \Pi_{G,n}$ we can naturally associate the partition $\pi \in \Pi_n$ by *removing the labels*, i.e., with

$$(2.8)\quad \pi_G = \{(\pi_1, g_1), (\pi_2, g_2), \dots, (\pi_b, g_b)\}$$

we associate $\pi = \{\pi_1, \dots, \pi_b\} \in \Pi_n$. With each $\pi_G \in \Pi_{G,n}$ we also associate the set of its *labels*

$$(2.9)\quad L(\pi_G) = \{g_1, \dots, g_b\} \subset G.$$

In addition to the finite-$n$ sets $\Pi_n$ and $\Pi_{G,n}$ considered above, consider their infinite versions

$$(2.10)\quad \Pi = \{\text{partitions of } \mathbb{N}\}, \quad \Pi_G = \{G\text{-labelled partitions of } \mathbb{N}\},$$

and introduce the set of standard initial states

$$(2.11)\quad S_G = \big\{\{(\{i\}, g_i)\}_{i\in\mathbb{N}}\colon\ g_i \in G, i \in \mathbb{N}\big\}.$$

Equip $\Pi_G$ with the following topology. First, equip the set $\Pi_{G,n}$ with the *discrete topology*. In particular, this implies that $\Pi_{G,n}$ is a *Polish space*. We say that the sequence of labelled partitions $\{\pi_G^{(k)} \in \Pi_G\}_{k\in\mathbb{N}}$ converges to the labelled partition $\pi_G \in \Pi_G$ if the sequence $\{\pi_G^{(k)}|_n \in \Pi_{G,n}\}_{k\in\mathbb{N}}$ converges to $\pi_G|_n \in \Pi_{G,n}$ for all $n \in \mathbb{N}$. This topology makes the space $\Pi_G$ Polish too.

**Evolution rules.** Assume that we are given transition rates (= "migration rates") on $G$

$$(2.12)\quad a^*\colon\ G^2 \to \mathbb{R}, \quad a^*(g, f) = a(f, g),$$

where $a(\cdot,\cdot)$ is the migration kernel of the $C^\Lambda$-process with geographic space $G$. The spatial $n$-$\Lambda$-*coalescent* is the continuous-time Markov process $\mathfrak{C}_n^{(G),\mathrm{loc}} = (\mathfrak{C}_n^{(G),\mathrm{loc}}(t) = \pi_G(t) \in \Pi_{G,n})_{t\geq 0}$ with the following dynamics. Given the current state $\pi_G = \mathfrak{C}_n^{(G),\mathrm{loc}}(t-) \in \Pi_{G,n}$, the process $\mathfrak{C}_n^{(G),\mathrm{loc}}$ evolves via:

- *Coalescence.* Independently, at each site $g \in G$, the families of $\pi_G$ with label $g$ *coalesce* according to the mechanism of the non-spatial $n$-$\Lambda$-coalescent. In other words, given that in the current state of the spatial $\Lambda$-coalescent there are $b = b(\pi_G, g) \in [n]$ families with label $g$, among these $i \in [2, b] \cap \mathbb{N}$ *fixed* families coalesce into one family with label $g$ at rate $\lambda_{b,i}^{(\Lambda)}$, where

  $$(2.13)\quad \lambda_{b,i}^{(\Lambda)} = \int_{[0,1]} \Lambda^*(\mathrm{d}r) r^i (1-r)^{b-i}, \quad i \in [2, b] \cap \mathbb{N},$$

  with $\Lambda^*$ given by (1.5).

- *Migration.* Families migrate independently at rate $a^*$, i.e., for any ordered pair of labels $(g, g') \in G^2$, a family of $\pi_G$ with label $g \in G$ changes its label (= "migrates") to $g' \in G$ at rate $a^*(g, g')$.

**Graphical construction.** Next, we recall the explicit construction of the above described spatial $n$-$\Lambda$-coalescent via Poisson point processes (see also Limic and Sturm [LS06]).

Consider the family $\mathfrak{P} = \{\mathfrak{P}_g\}_{g \in G}$ of i.i.d. Poisson point processes on $[0,\infty) \times [0,1] \times \{0,1\}^{\mathbb{N}}$ defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ with intensity measure

$$\text{(2.14)} \quad \mathrm{d}t \otimes \left[\Lambda(\mathrm{d}r)(r\delta_1 + (1-r)\delta_0)^{\otimes \mathbb{N}}\right](\mathrm{d}\omega),$$

where $\omega = (\omega_i)_{i \in \mathbb{N}} \subset \{0,1\}^{\mathbb{N}}$. Note that the second factor of the intensity measure in (2.14) is not a product measure on $[0,1] \times \{0,1\}^{\mathbb{N}}$, in particular, it is not the same as

$$\text{(2.15)} \quad \left[\Lambda^*(\mathrm{d}r)(r\delta_1 + (1-r)\delta_0)\right]^{\otimes \mathbb{N}}(\mathrm{d}\omega).$$

Given $J \subset [n]$ and $g \in G$, define the *labelled coalescence map* $\mathrm{coal}_{J,g}\colon \Pi_{G,n} \to \Pi_{G,n}$, which coalesces the blocks with indices specified by $J$ and locates the new-formed block at $g$, as follows:

$$\text{(2.16)} \quad \mathrm{coal}_{J,g}(\pi_{G,n}) = \left( \bigcup_{i \in J \cap [b(\pi)]} \pi_i, g \right) \cup \left( \pi_{G,n} \setminus \bigcup_{i \in J \cap [b(\pi)]} (\pi_i, g_i) \right), \quad \pi_{G,n} \in \Pi_{G,n}.$$

Using $\mathfrak{P}$, we construct the standard spatial $n$-$\Lambda$-coalescent $\mathfrak{C}_n^{(G),\mathrm{loc}} = (\mathfrak{C}_n^{(G),\mathrm{loc}}(t))_{t \geq 0}$ as a Markov $\Pi_{G,n}$-valued process with the following properties:

- *Initial state.* Assume $\mathfrak{C}_n^{(G),\mathrm{loc}}(0) \in S_{G,n}$.
- *Coalescence.* For each $g \in G$ and each point $(t, r, \omega)$ of the Poisson point process $\mathfrak{P}_g$ satisfying $\sum_{i \in \mathbb{N}} \omega_i \geq 2$, all families $(\pi_i(t-), g_i(t-)) \in \mathfrak{C}_n^{(G),\mathrm{loc}}(t-)$ such that $g_i(t-) = g$ and $\omega_i = 1$ coalesce into a new family labelled by $g$, i.e.,

  $$\text{(2.17)} \quad \mathfrak{C}_n^{(G),\mathrm{loc}}(t) = \mathrm{coal}_{\{i \in [n]\colon\ \omega_i = 1, g_i(t-) = g\}, g}(\mathfrak{C}_n^{(G),\mathrm{loc}}(t-)).$$

- *Migration.* Between the coalescence events, the labels of all partition elements of $\mathfrak{C}_n^{(G),\mathrm{loc}}(t)$ perform independent random walks with transition rates $a^*$ [11].

In what follows, we denote by $\cdot|_n\colon \Pi_{G,m} \to \Pi_{G,n}$ (respectively, $\cdot|_n\colon \Pi_G \to \Pi_{G,n}$) the operation of projection of all families in $[m]$ (respectively, $\mathbb{N}$) onto $[n]$.

**Entrance law.** Note that, by construction, the spatial $n$-$\Lambda$-coalescent satisfies the following consistency property:

$$\text{(2.18)} \quad \mathcal{L}\left[\mathfrak{C}_m^{(G),\mathrm{loc}}|_n\right] = \mathcal{L}\left[\mathfrak{C}_n^{(G),\mathrm{loc}}\right], \qquad n, m \in \mathbb{N},\ n \leq m.$$

Therefore, by the Kolmogorov extension theorem, there exists a process

$$\text{(2.19)} \quad \mathfrak{C}^{(G),\mathrm{loc}} = (\mathfrak{C}^{(G),\mathrm{loc}}(t) \in \Pi_G)_{t \geq 0}$$

such that $\mathfrak{C}^{(G),\mathrm{loc}}|_n = \mathfrak{C}_n^{(G),\mathrm{loc}}$.

**Definition 2.2** ([LS06]). *Call the process $\mathfrak{C}^{(G),\mathrm{loc}}$ the spatial $\Lambda$-coalescent corresponding to the migration rates $a^*$ and the coalescence measure $\Lambda$.*

---

[11] The adjective "between" is well defined because the set of points $(t, r, \omega)$ of $\mathfrak{P}_g$ satisfying the condition $\sum_{i \in \mathbb{N}} \omega_i \geq 2$ is topologically discrete, and hence can be ordered w.r.t. the first coordinate (= time).

### 2.1.2 Martingale problem

In this section, we characterise the spatial $\Lambda$-coalescent as the unique solution of the corresponding well-posed martingale problem.

Let $\mathcal{C}_G$ be the algebra of bounded continuous functions $F\colon \Pi_G \to \mathbb{R}$ such that for all $F \in \mathcal{C}_G$ there exists an $n \in \mathbb{N}$ and a bounded function

$$(2.20)\quad F_n\colon \Pi_{G,n} \to \mathbb{R}$$

with the property that $F(\cdot) = F_n(\cdot|_n)$. In words, $F$ only depends on the family structure of a finite number of individuals. It is easy to check that $\mathcal{C}_G$ separates points on $\Pi_G$. Given $f, g \in G$ and $i \in [n]$, define the *migration map* $\mathrm{mig}_{f\to g,i}\colon \Pi_{G,n} \to \Pi_{G,n}$ as

$$(2.21)\quad \mathrm{mig}_{f\to g,i}(\pi_{G,n}) = \begin{cases} (\pi_i, g) \cup (\pi_{G,n} \setminus (\pi_i, f)), & (\pi_i, f) \in \pi_{G,n}, \\ \pi_{G,n}, & (\pi_i, f) \notin \pi_{G,n}, \end{cases} \qquad \pi_{G,n} \in \Pi_{G,n},$$

describing the jump in which the family labelled $i$ migrates from colony $f$ to colony $g$.

Consider the linear operator $L^*_G$ defined as

$$(2.22)\quad L^*_G = L^*_{\mathrm{mig},G} + L^*_{\mathrm{coal},G},$$

where the operators $L^*_{\mathrm{mig},G}, L^*_{\mathrm{coal},G}\colon \mathcal{C}_G \to \mathcal{C}_G$ are defined for $\pi_G \in \Pi_G$ and $F \in \mathcal{C}_G$ as

$$(2.23)\quad (L^*_{\mathrm{mig},G}F)(\pi_G) = \sum_{i=1}^{b(\pi_G|_n)} \sum_{g,f\in G} a^*(g,f)\big[F_n(\mathrm{mig}_{g\to f,i}(\pi_G|_n)) - F(\pi_G)\big],$$

$$(2.24)\quad (L^*_{\mathrm{coal},G}F)(\pi_G) = \sum_{g\in G} \sum_{\substack{J\subset\{i\in[n]\colon g_i=g\},\\ |J|\geq 2}} \lambda^{(\Lambda)}_{b(\pi_G|_n,g),|J|}\big[F_n(\mathrm{coal}_{J,g}(\pi_G|_n)) - F(\pi_G)\big].$$

**Proposition 2.3. [Martingale problem for the spatial-coalescent without block coalescence]**
*The spatial $\Lambda$-coalescent defined in Section* 2.1.1 *solves the well-posed martingale problem for* $(L^*_G, C_{\mathrm{b}}(\Pi_G), \delta_{S_G})$.

*Proof.* A straightforward inspection of the graphical construction yields the existence. The uniqueness is immediate because we have a duality relation, as we will see in Section 2.3. □

**Remark 2.4.** *Note that, instead of the singleton initial condition in Proposition* 2.3 *(and in the graphical construction of Section* 2.1.1*), we can use any other initial condition in* $\Pi_G$.

### 2.1.3 Mean-field and immigration-emigration $\Lambda$-coalescents

Some special spatial $\Lambda$-coalescents will be needed in the course of our analysis of the hierarchically interacting Cannings process. We define the *mean-field* $\Lambda$*-coalescent* as the spatial $\Lambda$-coalescent with geographic space $G = \{0, \ldots, N-1\}$ and migration kernel $a(i,j) = c/N$ for all $i, j \in G$ with $i \neq j$. Furthermore, we define the $\Lambda$*-coalescent with immigration-emigration* as the spatial $\Lambda$-coalescent with geographic space $G = \{0, *\}$ and migration kernel $a(0,*) = c$, $a(*,0) = 0$. In other words, $*$ is a cemetery migration state.

## 2.2 Spatial Λ-coalescent with block coalescence

In this section, we construct a new type of spatial coalescent process based on a sequence $(\Lambda_k)_{k\in\mathbb{N}}$ of finite measures on $[0,1]$, namely, the *spatial* $\underline{\Lambda}$*-coalescent* on $G=\Omega_N$ *with block coalescence.* For each $k\in\mathbb{N}$, we introduce two additional transition mechanisms: (1) a *block reshuffling* of all partition elements in a ball of radius $k$; (2) a Λ-*block coalescence* with resampling measure $\Lambda_k$ of all partition elements in a ball a radius $k$. In Section 2.2.1 we give definitions, in Section 2.2.2 we formulate the martingale problem.

### 2.2.1 The evolution rules and the Poissonian construction

We start by extending the Poissonian construction from Section 2.1.1 to incorporate the additional transition mechanisms of block reshuffling and block coalescence.

Consider Poisson point processes $\mathfrak{P}^{(\Omega_N)}$ on

$$(2.25)\quad [0,\infty)\times\Omega_N\times\mathbb{N}_0\times[0,1]\times\{0,1\}^{\mathbb{N}}$$

defined on the filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$ with intensity measure

$$(2.26)\quad \mathrm{d}t\otimes\mathrm{d}\eta\otimes\Big(N^{-2k}\mathrm{d}k\,\Big[\Lambda_k(\mathrm{d}r)\big(r\delta_1+(1-r)\delta_0\big)^{\otimes\mathbb{N}}\Big](\mathrm{d}\omega)\Big),$$

where $\omega=(\omega_i)_{i\in\mathbb{N}}\subset\{0,1\}^{\mathbb{N}}$, $(t,\eta,k,r,\omega)\in[0,\infty)\times\Omega_N\times\mathbb{N}_0\times[0,1]\times\{0,1\}^{\mathbb{N}}$, $\mathrm{d}k$ is counting measure on $\mathbb{N}$ and $\mathrm{d}\eta$ is counting measure on $\Omega_N$. Again, note that the third factor in (2.26) is not a product measure (compare (2.15)).

Given $\Sigma\Subset\Omega_N$ (i.e., $\Sigma$ is a finite subset of $\Omega_N$) and $\xi=\{\xi_i\}_{i=1}^{|\Sigma|}\subset\Sigma$, let $\mathrm{resh}_{\Sigma,\xi}\colon\Pi_{\Omega_N}\to\Pi_{\Omega_N}$ be the *reshuffling map* that for all $i$ moves families from $\eta_i\in\Sigma$ to $\xi_i\in\Sigma$:

$$(2.27)\quad \mathrm{resh}_{\Sigma,\xi}(\pi_{\Omega_N})_i=\begin{cases}(\pi_i,\eta_i), & \eta_i\notin\Sigma,\\ (\pi_i,\xi_i), & \eta_i\in\Sigma,\end{cases}\qquad \pi_{\Omega_N}\in\Pi_{\Omega_N},\ i\in[b(\pi_{\Omega_N})].$$

Let

$$(2.28)\quad U_\Sigma=\{U_\Sigma(\xi)\}_{\xi\in\Sigma}$$

be a collection of independent random variables uniformly distributed on $\Sigma$. Using $\mathfrak{C}^{(\Omega_N)}$, we construct the *standard spatial* $n$*-*Λ*-coalescent with block coalescence* $\mathfrak{C}_n^{(\Omega_N)}=(\mathfrak{C}_n^{(\Omega_N)}(t)\in\Pi_{\Omega_N,n})_{t\geq0}$ as the $\Pi_{\Omega_N,n}$-valued Markov process with the following properties:

- *Initial state.* Assume $\mathfrak{C}_n^{(\Omega_N)}(0)\in S_{\Omega_N,n}$.
- *Coalescence with reshuffling.* For each point $(t,\eta,k,r,\omega)$ of the Poisson point process $\mathfrak{P}^{(\Omega_N)}$, all families $(\pi_i,\eta_i)\in\mathfrak{C}_n^{(\Omega_N)}(t-)$ such that $\omega_i=1$ and $\eta_i\in B_k(\eta)$ coalesce into a new family with label $\eta$. Subsequently, all families with labels $\zeta\in B_k(\eta)$ obtain a new label that is drawn independently and uniformly from $B_k(\eta)$. In a formula:

  $$(2.29)\quad \mathfrak{C}_n^{(\Omega_N)}(t)=\mathrm{resh}_{B_k(\eta),U_{B_k(\eta)}}\circ\mathrm{coal}_{\{i\in[n]\colon\ \omega_i=1,\eta_i(t-)\in B_k(\eta)\},\eta}(\mathfrak{C}_n^{(\Omega_N)}(t-)).$$

  Note that, in contrast with the spatial coalescent from Section 2.1, the coalescence mechanism in (2.29) is no longer local: all families whose labels are in $B_k(\eta)$, $k\in\mathbb{N}$, are involved in the coalescence event at site $\eta\in\Omega_N$.

- *Migration.* Independently of the coalescence events, the labels of all partition elements of $\mathfrak{C}_n^{(\Omega_N)}(t)$ perform independent random walks with transition rates $a^{(N)}(\cdot,\cdot)$ (recall (1.25)).

As in Section 2.1, the consistency-between-restrictions property allows us to apply the Kolmogorov extension theorem to the family $\{\mathfrak{C}_n^{(\Omega_N)}\}_{n\in\mathbb{N}}$ to construct the Markov process

$$\mathfrak{C}^{(\Omega_N)} \tag{2.30}$$

taking values in $\Pi_{\Omega_N}$.

**Definition 2.5.** *The process $\mathfrak{C}^{(\Omega_N)}$ is called the spatial $\underline{\Lambda}$-coalescent with block coalescence corresponding to the resampling measures $(\Lambda_k)_{k\in\mathbb{N}_0}$ (recall (1.27)) and the migration coefficients $(c_k)_{k\in\mathbb{N}_0}$ (recall (1.24)).*

**Proposition 2.6. [Feller property]**
*The process $\mathfrak{C}^{(\Omega_N)}$ is a càdlàg strong Markov process with the Feller property.*

*Proof.* This is an immediate consequence of the Poissonian construction. □

### 2.2.2 Martingale problem

In this section, we characterise the spatial Λ-coalescent with block coalescence as the solution of the corresponding martingale problem.

Given $\pi_{\Omega_N,n} \in \Pi_{\Omega_N,n}$ and $B_k(\eta) \subset \Omega_N$, denote the number of families of $\pi_{\Omega_N,n}$ with labels in $B_k(\eta)$ by

$$b(\pi_{\Omega_N,n}, B_k(\eta)) = |\{(\pi_i,\eta_i) \in \pi_{\Omega_N,n}\colon\ \eta_i \in B_k(\eta)\}| \in \mathbb{N}. \tag{2.31}$$

Recall the definition of the algebra of test functions $\mathcal{C}_G$ from Section 2.1.2. Let $\pi_{\Omega_N} = \{(\pi_i,\eta_i)\}_{i\in\mathbb{N}} \in \Pi_{\Omega_N}$, $F \in \mathcal{C}_{\Omega_N}$ and $F(\cdot) = F_n(\cdot|_n)$. Consider the linear operator $L^{(\Omega_N)*}$ defined as

$$L^{(\Omega_N)*} = L^{(\Omega_N)*}_{\text{mig}} + L^{(\Omega_N)*}_{\text{coal}}, \tag{2.32}$$

where the linear operators $L^{(\Omega_N)*}_{\text{mig}}$ and $L^{(\Omega_N)*}_{\text{coal}}$ are defined as follows (recall (2.20)). The migration operator is

$$\left(L^{(\Omega_N)*}_{\text{mig}} F\right)(\pi_{\Omega_N}) = \sum_{i=1}^{b(\pi_{\Omega_N}|_n)} \sum_{\eta,\zeta\in\Omega_N} a^{(N)}(\eta,\zeta)\left[F_n(\text{mig}_{\eta\to\zeta,i}(\pi_{\Omega_N}|_n)) - F(\pi_{\Omega_N})\right], \tag{2.33}$$

and the block-coalescence-reshuffling operator is

$$\begin{aligned} &\left(L^{(\Omega_N)*}_{\text{coal}} F\right)(\pi_{\Omega_N}) = \sum_{\eta\in\Omega_N}\sum_{k\in\mathbb{N}_0} N^{-2k} \sum_{\xi_1\in B_k(\eta)} N^{-k} \sum_{\xi_2\in B_k(\eta)} N^{-k}\cdots \sum_{\xi_{|B_k(\eta)|}\in B_k(\eta)} N^{-k} \\ &\times \sum_{\substack{J\subset[b(\pi_{\Omega_{N,n}},B_k(\eta))],\\ |J|\geq 2}} \lambda^{(\Lambda_k)}_{b(\pi_{G,n},B_k(\eta)),|J|}\left[F_n(\text{resh}_{B_k(\eta),\xi}\circ\text{coal}_{\{i\in J\colon\ \eta_i\in B_k(\eta)\},\eta}(\pi_{\Omega_N}|_n)) - F(\pi_{\Omega_N})\right]. \end{aligned} \tag{2.34}$$

**Proposition 2.7. [Martingale problem: Spatial $\underline{\Lambda}$-coalescent with block coalescence]**
*The spatial $\underline{\Lambda}$-coalescent with block coalescence $\mathfrak{C}^{(\Omega_N)}$ defined in Section* 2.2 *solves the well-posed martingale problem* $(L^{(\Omega_N)*}, \mathcal{C}_{\Omega_N}, \delta_{S_{\Omega_N}})$.

*Proof.* A straightforward inspection of the graphical construction in Section 2.2 yields the existence of a solution. Uniqueness on finite geographic spaces is clear: this follows in the same way as for the single-site case. Once we have well-posedness for finite geographic spaces, we can show uniqueness for $G = \Omega_N$ via approximation. The approximation via finite geographic spaces follows from the fact that the occupation numbers of the sites are stochastically smaller than in the case of pure random walks (see Liggett and Spitzer [LS81]). □

**Remark 2.8.** *Note that, instead of the singleton initial condition in Proposition* 2.7 *(and in the graphical construction of Section* 2.2*), we can use any other initial condition in* $\Pi_{\Omega_N}$.

## 2.3 Duality relations

We next formulate and prove the duality relation between the $C_N^{c,\underline{\Lambda}}$-process and the $\underline{\Lambda}$-coalescent described so far. This follows a general pattern for all choices of the geographic space $G$ in (2.2). We only give the proof for the case $G = \Omega_N$.

Recall (2.1). The construction of the duality function $H(\cdot,\cdot)$ requires some new ingredients. For $n \in \mathbb{N}$ and $\varphi \in C_\mathrm{b}(E^n, \mathbb{R})$, consider the *bivariate function* $H_\varphi^{(n)} \colon \mathcal{P}(E)^G \times \Pi_{G,n} \to \mathbb{R}$ of the form

$$H_\varphi^{(n)}(x, \pi_{G,n}) = \int_{E^{b(\pi)}} \left( \bigotimes_{i=1}^{b(\pi)} x_{\eta_{\pi^{-1}(i)}}(\mathrm{d}u_i) \right) \varphi\big(u_{\pi(1)}, u_{\pi(2)}, \ldots, u_{\pi(n)}\big), \tag{2.35}$$

where $x = (x_\eta)_{\eta\in G} \in \mathcal{P}(E)^G$, $\pi_{G,n} \in \Pi_{G,n}$, $b = b(\pi_{G,n}) = |\pi_{G,n}|$, $(\eta_i)_{i\in[b]} = L(\pi_{G,n})$ are the labels of the partition $\pi_{G,n}$, and (with a slight abuse of notation) $\pi\colon\ [n] \to [b]$ is the map from Remark 2.1. In words, the functions in (2.35) assign the same type to individuals that belong to the same family. Note that these functions form a family of functions on $\mathcal{P}(E)^G$,

$$\left\{ H_\varphi^{(n)}(\cdot, \pi_{G,n}) \colon\ \mathcal{P}(E)^G \to \mathbb{R} \mid \pi_{G,n} \in \Pi_{G,n}, n \in \mathbb{N}, \varphi \in C_\mathrm{b}(E^n, \mathbb{R}) \right\}, \tag{2.36}$$

that separates points. The $C^{\underline{\Lambda}}$-process with block resampling and the spatial $\underline{\Lambda}$-coalescent with block coalescence are mutually *dual* w.r.t. the duality function $H(\cdot, *)$ given by

$$H(x, (\varphi, \pi_{G,n})) = H_\varphi(x, \pi_{G,n}), \qquad x \in \mathcal{E} = \mathcal{P}(E)^G,\ (\varphi, \pi_{G,n}) \in \mathcal{E}', \tag{2.37}$$

with $\mathcal{E}' = \cup_{n\in\mathbb{N}_0} (C_\mathrm{b}(E^n, \mathbb{R}) \times \Pi_{G,n})$.

We proceed with the following observation.

**Remark 2.9.** *Let $X$ and $Y$ be two processes that are dual w.r.t. a continuous and bounded duality function $H(\cdot,\cdot)$. Assume that $X$ and $Y$ are solutions to martingale problems corresponding to operators $L_X$, respectively, $L_Y$. Then the generator relation*

$$[L_X(H(\cdot, Y_0))](X_0) = [L_Y(H(X_0, \cdot))](Y_0), \quad (X_0, Y_0) \in \mathcal{E} \times \mathcal{E}', \tag{2.38}$$

*is equivalent to the duality relation* (2.1) *(see e.g. Ethier and Kurtz* [EK86, Section 4.4]*).*

**Remark 2.10.** *Remark* 2.9 *gives the duality function* $H(\cdot,*)$ *for all* $t \geq 0$ *and* $n \in \mathbb{N}$,

$$\text{(2.39)}\quad \mathbb{E}\left[H_\varphi^{(n)}(X^{(G)}(t), \mathfrak{C}^{(G)}(0)|_n)\right] = \mathbb{E}\left[H_\varphi^{(n)}(X^{(G)}(0), \mathfrak{C}^{(G)}(t)|_n)\right],$$

*as is proved in Proposition* 2.11 *below.*

In our context, we have to verify the following relation for the linear operators in the martingale problem.

**Proposition 2.11. [Operator level duality]**
*For any of the geographic spaces* $G = \Omega_N$, $G = \{0, \dots, N-1\}^K, K \in \mathbb{N}$ *and* $G = \{0, *\}$ *the following holds. For all* $H_\varphi^{(n)}$ *as in* (2.35), *all* $x \in \mathcal{P}(E)^G$, *all* $n \in \mathbb{N}$, *and all* $\pi_G \in \Pi_G$,

$$\text{(2.40)}\quad \left(L^{(G)} H_\varphi^{(n)}(\cdot, \pi_G|_n)\right)(x) = \left(L^{(G)*} H_\varphi^{(n)}(x, \cdot|_n)\right)(\pi_G).$$

*Proof.* We check the statement for $G = \Omega_N$. The proof for the other choices of $G$ is left to the reader.

The claim follows from a straightforward inspection of (1.36–1.37) and (2.33–2.34), respectively. Indeed, duality of the migration operators in (1.36) and (2.33) is evident:

$$\text{(2.41)}\quad \left(L_{\text{mig}}^{(G)} H_\varphi^{(n)}(\cdot, \pi_G|_n)\right)(x) = \left(L_{\text{mig}}^{(G)*} H_\varphi^{(n)}(x, \cdot|_n)\right)(\pi_G).$$

Let us check the duality of the resampling and coalescence operators in (1.37) and (2.34). By a standard approximation argument, it is enough to consider the duality test functions in (2.35) of the product form, i.e., with $\varphi(u) = \prod_{i=1}^n \varphi_i(u_i)$, where $u = (u_i)_{i=1}^n \in E^n$ and $\varphi_i \in C_{\text{b}}(E)$. Using (1.37), (2.13) and simple algebra, for $x \in \mathcal{P}(E)^G$ and $\pi_G \in \Pi_G$ we can rewrite the action of the resampling operator on the duality test function as follows (where for ease of notation we assume that $\pi_G \in S_G$, i.e., $\pi_G$ has the singleton family structure)

$$\begin{aligned}
&\left(L_{\text{res}}^{(\Omega_N)} H_\varphi^{(n)}(\cdot, \pi_G|_n)\right)(x) \\
&= \sum_{\eta \in G} \sum_{k \in \mathbb{N}_0} N^{-2k} \int_{[0,1]} \Lambda_k^*(\mathrm{d}r) N^{-k} \sum_{\rho \in B_k(\eta)} \int_E x_\rho(\mathrm{d}a) \\
&\left( \prod_{i=1}^{b(\pi_{G,n}, B_k(\eta))} \left\langle \left(\Phi_{r,a,B_k(\eta)}(x)\right)_{\eta_{\pi^{-1}(i)}}, \prod_{j:\ \pi(j)=i} \varphi_j \right\rangle \right. \\
&\qquad \left. - \prod_{i=1}^{b(\pi_{G,n}, B_k(\eta))} \left\langle x_{\eta_{\pi^{-1}(i)}}, \prod_{j:\ \pi(j)=i} \varphi_j \right\rangle \right) \\
&= \sum_{\eta \in G} \sum_{k \in \mathbb{N}_0} N^{-2k} \int_{[0,1]} \Lambda_k^*(\mathrm{d}r) N^{-k} \sum_{\rho \in B_k(\eta)} \int_E x_\rho(\mathrm{d}a) \\
&\left( \sum_{\substack{J \subset [b(\pi_{G,n}, B_k(\eta))] \\ |J| \geq 0}} \prod_{i \in [b(\pi_{G,n}, B_k(\eta))] \setminus J} \left\langle (1-r) y_{\eta_{\pi^{-1}(i)},k}, \prod_{j:\ \pi(j)=i} \varphi_j \right\rangle \prod_{i \in J} \left\langle r\delta_a, \prod_{j:\ \pi(j)=i} \varphi_j \right\rangle \right. \\
&\qquad \left. - \prod_{i=1}^{b(\pi_{G,n}, B_k(\eta))} \left\langle x_{\eta_{\pi^{-1}(i)}}, \prod_{j:\ \pi(j)=i} \varphi_j \right\rangle \right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\eta\in G}\sum_{k\in\mathbb{N}_0} N^{-2k} \sum_{\substack{J\subset[b(\pi_{G,n},B_k(\eta))],\\ |J|\geq 2}} \lambda^{(\Lambda_k)}_{b(\pi_{G,n},B_k(\eta)),|J|} \\
&\quad \Bigg( N^{-k}\sum_{\rho\in B_k(\eta)} \prod_{i\in[b(\pi_{G,n},B_k(\eta))]\setminus J} \left\langle N^{-k}\sum_{\xi\in B_k(\eta)} x_\xi, \prod_{j:\pi(j)=i}\varphi_j\right\rangle \prod_{i\in J}\left\langle x_\rho, \prod_{j:\,\pi(j)=i}\varphi_j\right\rangle \\
&\qquad - \prod_{i=1}^{b(\pi_{G,n},B_k(\eta))}\left\langle x_{g_{\pi^{-1}(i)}}, \prod_{j:\,\pi(j)=i}\varphi_j\right\rangle\Bigg).
\end{aligned}
\tag{2.42}
$$

On the other hand, according to (2.34), we have

$$
\begin{aligned}
&\left(L^{(\Omega_N)*}_{\text{coal}} H^{(n)}_{\varphi}(x,\cdot|_n)\right)(\pi_G) = \sum_{\eta\in\Omega_N}\sum_{k\in\mathbb{N}_0} N^{-2k} \sum_{\substack{J\subset[b(\pi_{G,n},B_k(\eta))],\\ |J|\geq 2}} \lambda^{(\Lambda_k)}_{b(\pi_{G,n},B_k(\eta)),|J|} \\
&\times\Bigg( \sum_{\xi_1\in B_k(\eta)} N^{-k} \sum_{\xi_2\in B_k(\eta)} N^{-k}\cdots \sum_{\xi_{b(\pi_{G,n},B_k(\eta))}\in B_k(\eta)} N^{-k} \\
&\quad\times\Bigg( \prod_{i\in[b(\pi_{G,n},B_k(\eta))]\setminus J} \left\langle x_{\xi_i}, \prod_{j:\,\pi(j)=i}\varphi_j\right\rangle \left\langle x_{\xi_{\min\{l:\,l\in J\}}}, \prod_{j:\,\pi(j)\in J}\varphi_j\right\rangle \\
&\qquad - \prod_{i=1}^{b(\pi_{G,n},B_k(\eta))}\left\langle x_{g_{\pi^{-1}(i)}}, \prod_{j:\,\pi(j)=i}\varphi_j\right\rangle\Bigg)\Bigg).
\end{aligned}
\tag{2.43}
$$

Comparing (2.43) with (2.42), we get the claim. □

## 2.4 The long-time behaviour of the spatial $\Lambda$-coalescent with block coalescence

We next investigate the long-time behaviour of the $\underline{\Lambda}$-coalescent. Subsequently, the duality relation allows us to translate results on the long-time behaviour of the $\underline{\Lambda}$-coalescent into results on the long-time behaviour of the $C_N^{\underline{c},\underline{\Lambda}}$-process.

### 2.4.1 The behaviour as $t \to \infty$

In this section, we prove the existence and uniqueness of a limiting state for the $\underline{\Lambda}$-coalescent as $t \to \infty$.

**Proposition 2.12. [Limiting state]**
*Start the $\mathfrak{C}_N^{\underline{c},\underline{\Lambda}}$-process in a labelled partition $\{(\pi_i,\eta_i)\}_{i=1}^n$, where $\{\pi_i\}_{i=1}^n$ form a partition of $\mathbb{N}$ and $\{\eta_i\}_{i=1}^n$ are the corresponding labels. If $x$ is a translation-invariant shift-ergodic random state with mean $\theta\in\mathcal{P}(E)$, then*

$$
\mathcal{L}\left[H^{(n)}_{\varphi}(x,\mathfrak{C}^{(\Omega_N)}_n(t))\right] \underset{t\to\infty}{\Longrightarrow} \mathcal{L}\left[H^{(n)}_{\varphi}(\underline{\theta},\mathfrak{C}^{(\Omega_N)}_n(\infty))\right] \qquad \forall\, n\in\mathbb{N}.
\tag{2.44}
$$

*Proof.* We first observe that $|\mathfrak{C}_n^{(\Omega_N)}(t)|$ is monotone non-increasing, so that there exists a limit for the number of partition elements. This implies that the partition structure converges a.s. to a limit partition, which we call $\mathfrak{C}^{(\Omega_N,n)}(\infty) \in \Pi_{\Omega_N,n}$. We must prove that the locations result in an effective averaging of the configuration $x$, so that we can replace the $|\mathfrak{C}^{(\Omega_N)}(t)|$-locations by any tuple for the (constant) configuration $\underline{\theta}$. This is a standard argument (see e.g. the proof of the ergodic theorem for the voter model in Liggett [L85]). □

**Corollary 2.13. [Limiting state of McKean-Vlasov]** *The convergence in* (2.44) *holds for* $Z_\theta^{c,d,\Lambda}$.

### 2.4.2 The dichotomy: single ancestor versus multiple ancestors

The key question is whether the $\mathfrak{C}_N^{\underline{c},\underline{\Lambda}}$-process converges to a single labelled partition element as $t \to \infty$ with probability one or more than one partition element occurs with positive probability. For that purpose, we have to investigate whether two tagged partition elements coalesce with probability one or not. Recall that, by the projective property of the coalescent, we may focus on the subsystem of just two dual individuals, because this translates into the same dichotomy for any $n$-$\Lambda$-coalescent and hence for the entrance law starting from countably many individuals. However, there is additional reshuffling at all higher levels, which is triggered by a corresponding block-coalescence event. Therefore, we consider two coalescing random walks $(Z_t^1, Z_t^2)_{t\geq 0}$ on $\Omega_N$ with migration coefficients $(c_k + \lambda_{k+1}N^{-(k+1)})_{k\in\mathbb{N}_0}$ and coalescence at rates $(\lambda_k)_{k\in\mathbb{N}_0}$. Consider the time-$t$ accumulated hazard function for coalescence of this pair:

$$(2.45)\quad H_N(t) = \sum_{k\in\mathbb{N}_0} \lambda_k N^{-k} \int_0^t 1\left\{d(Z_s^1, Z_s^2) \leq k\right\} \mathrm{d}s.$$

Here, the rate $N^{-2k}$ to choose a $k$-block is multiplied by $N^k$ because all partition elements in that block can trigger a coalescence event. This yields factor $N^{-k}$ in (2.45). We have coalescence of the random walks (= single common ancestor) with probability one if $\lim_{t\to\infty} H_N(t) = H_N(\infty) \equiv \infty$ a.s., but separation of the random walks (= multiple ancestors) with positive probability if $H_N(\infty) < \infty$ a.s.

**Lemma 2.14. [Zero-one law]** $H_N = H_N(\infty) = \infty$ *a.s. if and only if* $\bar{H}_N = \mathbb{E}[H_N(\infty)] = \infty$.

*Proof.* Write $H_N = \sum_{k\in\mathbb{N}_0} w_k L(k)$ with $w_k = \sum_{l\geq k} \lambda_l N^{-l}$ and the local times $L(k) = \int_0^\infty 1\{d(Z_s^1, Z_s^2) = k\}\mathrm{d}s$. Note that $w_k < \infty$ because of condition (1.31). By the isotropy of the transition kernel of the hierarchical random walk (1.25), we have that $L(k)$, $k \in \mathbb{N}_0$, are independent random variables. Hence, the claim follows from the Borel-Cantelli lemma. □

Let $P_t(\cdot,\cdot)$ denote the time-$t$ transition kernel of the hierarchical random walk on $\Omega_N$ with migration coefficients $(\bar{c}_k^{(N)})_{k\in\mathbb{N}_0}$, where $\bar{c}_k^{(N)} = c_k + \lambda_{k+1}N^{-(k+1)}$. Then

$$(2.46)\quad \bar{H}_N = \sum_{k\in\mathbb{N}_0} \lambda_k N^{-k} \int_0^\infty P_{2s}(0, B_k(0))\,\mathrm{d}s = \sum_{k\in\mathbb{N}_0} \mu_k N^{-k} \int_0^\infty P_t(0, B_k(0))\,\mathrm{d}t.$$

From the formulae in Dawson, Gorostiza and Wakolbinger [DGW05, Section 3.1] we obtain, after a little computation,

$$\begin{aligned}\int_0^\infty P_t(0,\eta)\,\mathrm{d}t &= \frac{1}{D_N}\left[O\left(\frac{1}{N}\right) + \left(1-\frac{1}{N}\right)\sum_{l\geq k}\left(\bar{c}_l^{(N)}\left(\frac{N}{N-1}\right) + \sum_{m\in\mathbb{N}}\frac{\bar{c}_{l+m}^{(N)}}{N^m}\right)^{-1}\right]\\ &\forall\, \eta\in\Omega_N \text{ with } d(0,\eta)=k\in\mathbb{N},\end{aligned} \tag{2.47}$$

where

$$D_N = \sum_{\eta\in\Omega_N} a^{(N)}(0,\eta) = \sum_{k\in\mathbb{N}}(N^k-N^{k-1})\sum_{l\geq k}\frac{\bar{c}_{l-1}^{(N)}}{N^{2l-1}} = \sum_{m\in\mathbb{N}_0}\frac{\bar{c}_m^{(N)}}{N^m}\left(1-\frac{1}{N^{m+1}}\right) \tag{2.48}$$

is the jump rate of our random walk (recall (1.25)). The factor $1/D_N$ in (2.47) is needed because the random walk in [DGW05] has jump rate 1. Note that the sums in (2.47–2.48) are finite because of condition (1.26). (The formula for $\int_0^\infty P_t(0,0)\,\mathrm{d}t$ can be easily deduced from (2.47), but this will not be needed.)

It follows from (2.47–2.48) that

$$\begin{aligned}&\lim_{N\to\infty} D_N = c_0,\\ &\lim_{N\to\infty} |B_k(0)\setminus B_{k-1}(0)|^{-1}\int_0^\infty P_t\big(0,B_k(0)\setminus B_{k-1}(0)\big)\,\mathrm{d}t = \frac{1}{c_0}\left(\sum_{l\geq k} c_l\right)^{-1}, \quad k\in\mathbb{N}.\end{aligned} \tag{2.49}$$

Hence, $\lim_{N\to\infty}\bar{H}_N = \infty$ if and only if $\sum_{k\in\mathbb{N}_0}\mu_k\sum_{l\geq k}(1/c_l) = \infty$, which is the same as $\sum_{k\in\mathbb{N}_0}(1/c_k)\sum_{l=0}^k\mu_l = \infty$, the condition in Theorem 1.5(c). Thus, we see that, in the limit as $N\to\infty$, the condition for clustering in the hierarchical Cannings process is the same as the condition for coalescence in the hierarchical Cannings-coalescent, in line with the duality.

In fact, the above formulae tell us more: the *same* dichotomy holds for finite $N$. This is because, uniformly in $N\in\mathbb{N}\backslash\{1\}$, the quantities in (2.49) are bounded from above and below by a constant times their limit as $N\to\infty$.

# 3 Well-posedness of martingale problems

Our task in this section is to prove Propositions 1.1–1.3, i.e., we have to show that the martingale problem for the single-colony process, the McKean-Vlasov process, the multi-colony process and the hierarchically interacting Cannings process are all well-posed (= have a unique solution). The line of argument is the same for all. In Section 3.1, we make some preparatory observations. In Section 3.2, we give the proofs.

## 3.1 Preparation

We first show that the duality relation and the characterisation of the dual process via a martingale problem allow us to prove the *existence* of a solution to the martingale problem that is strong Markov and has càdlàg paths. To this end, observe that via the dual process we can specify a distribution for every time $t$ and every initial state, since the dual is a unique solution of its martingale problem (being a projective limit of a Markov jump process defined for all times $t\geq 0$). Since the family $\{H(\cdot,Y_0)\colon\, Y_0\in\mathcal{E}'\}$ separates points, this uniquely defines a family of transition kernels $(P_{t,s})_{t\geq s\geq 0}$ satisfying the Kolmogorov equations, and hence defines uniquely a Markov process. By construction, this Markov process solves the martingale problem, provided we can verify the necessary path regularity.

We need to have càdlàg paths to obtain an admissible solution to the martingale problem. For finite geographic space this follows from the theory of Feller semigroups (see Ethier and Kurtz [EK86, Chapter 4]). For $\Omega_N$, we consider the exhausting sequence $(B_j(0))_{j\in\mathbb{N}_0}$ and use the standard tightness criteria for jump processes to obtain a weak limit point solving the martingale problem. The essential step is to control the effect on a single component of the flow of individuals in and out of $B_j(0)$ in finite time as $j\to\infty$.

It is standard to get *uniqueness* of the solution from the existence of the dual process (see e.g. [E00, Section 1.6] or [EK86, Proposition 4.4.7 and Theorem 4.4.11]). Again, this works for all the choices of $G$ in (2.2), with a little extra effort when $G=\Omega_N$.

## 3.2 Proofs of well-posedness

In the section, we prove Propositions 1.1–1.3. We follow the line of argument of Evans [Ev97, Theorem 4.1] and derive existence and uniqueness of the spatial Cannings process from the existence of the corresponding spatial Cannings-coalescent established in Section 2. The main tool is duality (cf. Proposition 2.11). The proofs of Propositions 1.1–1.3 follow the same pattern for $G=\{0,\dots,N-1\}$, $G=\{0,*\}$ and $G=\Omega_N$.

*Proof of Propositions* 1.1–1.3.
• *Well-posedness.* First we show that there exists a Markov transition kernel $Q_t$ on $\mathcal{P}(E)^G$ such that, for all $\varphi\in C_{\mathrm{b}}(E^n,\mathbb{R})$, $\pi\in\Pi_{G,n}$, $X\in\mathcal{P}(E)^G$ and $t\geq 0$,

$$(3.1)\quad \int Q_t(X,\mathrm{d}X')H_\varphi^{(n)}(X',\pi)=\mathbb{E}\left[H_\varphi^{(n)}(X,\mathfrak{C}_n^{(G)}(t))\mid \mathfrak{C}_n^{(G)}(0)=\pi\right].$$

Once (3.1) is established, the general theory of Markov processes implies the existence of a *Hunt-process* with the transition kernel $Q_t$ (see e.g. Blumenthal and Getoor [BG68, Theorem I.9.4]). This càdlàg process is unique and coincides with the process $X^{(G)}$, since (3.1) implies (2.39). There can be at most one process satisfying (2.39), since the family of duality functions $H_\varphi^{(n)}(\cdot,\pi)$ separates points on $\mathcal{P}(E)^G$.

• *Feller property.* To show that $X^{(G)}$ is a Feller process we use duality. It is enough to show that, for any $F\in\mathcal{F}$ and any $t\geq 0$, the map

$$(3.2)\quad \mathcal{P}(E)^G\ni x\mapsto \mathbb{E}\left[F(X^{(G)}(t))\mid X^{(G)}(0)=x\right]\in\mathbb{R}$$

is continuous. In (3.2), instead of the test functions $F(\cdot)\in\mathcal{F}$, it is enough to take the duality test functions $H_\varphi^{(n)}(\cdot,\pi_{G,n})$ from (2.35). The duality in (2.39) implies that

$$(3.3)\quad \mathbb{E}\left[H_\varphi^{(n)}(X^{(G)}(t),\pi_{G,n}|_n)\mid X^{(G)}(0)=x\right]=\mathbb{E}\left[H_\varphi^{(n)}(x,\mathfrak{C}^{(G)}(t)|_n)\right],\quad t\geq 0.$$

Definition (2.35) readily implies that the right-hand side of (3.3) is Lipschitz in $x$. □

# 4 Properties of the McKean-Vlasov process with immigration-emigration

The purpose of this section is to show that the $Z_\theta^{c,d,\Lambda}$-process with immigration-emigration is ergodic (Section 4.1), to identify its equilibrium distribution in terms of the dual (Section 4.3), and to calculate its first and second moment measure (Section 4.4). The characterisation via the dual will allow us to also show that the equilibrium depends continuously on the migration parameter $\theta$ (Section 4.2), a key property that will be needed later on and for which we need that the $\Lambda$-coalescent is *dust-free* (recall (1.3)).

## 4.1 Equilibrium and ergodic theorem

The equilibrium $\nu = \nu_\theta^{c,d,\Lambda} \in \mathcal{P}(\mathcal{P}(E))$ is the solution of the equation

$$(4.1)\quad \left\langle \nu, L_\theta^{c,d,\Lambda} F_\varphi \right\rangle = 0, \quad \varphi \in \mathcal{C}_\mathrm{b}(E^n), \quad n \in \mathbb{N},$$

where we recall (1.15–1.17) for the form of $F_\varphi$.

**Proposition 4.1. [Ergodicity]**
*For every initial state $Z_\theta^{c,d,\Lambda}(0) \in \mathcal{P}(E)$,*

$$(4.2)\quad \mathcal{L}\left[Z_\theta^{c,d,\Lambda}(t)\right] \underset{t\to\infty}{\Longrightarrow} \nu_\theta^{c,d,\Lambda}$$

*and the right-hand side is the unique equilibrium of the process. The convergence holds uniformly in the initial state.*

*Proof.* We use the dual process to show that the expectation in the right-hand side of the duality relation (2.44) converges. Indeed, we showed in (2.44) in Proposition 2.12 and its Corollary 2.13 that the state of the duality function $H(X_0,\cdot)$ applied to the dual process converges in law to a limiting random variable as $t \to \infty$. The duality function viewed as a function of the first argument generates a law-determining family $\{H(\cdot, C_0)\colon\ C_0 \in \Pi\}$ and hence (2.44) proves convergence.

It remains to show that the limit is independent of the initial state. Indeed, this is implied by the fact that if we start with *finitely many* partition elements, then all partition elements eventually jump to the cemetery location $\{*\}$ where all transition rates are zero and the state is $\theta$. The latter implies that the limit is unique. Since $\mathcal{P}(E)$ is compact and the process is Feller, there must exists an equilibrium, and this equilibrium must be equal to the $t \to \infty$ limit. □

## 4.2 Continuity in the centre of the drift

We want to prove that (in the weak topology on the respective spaces)

$$(4.3)\quad \theta \mapsto \nu_\theta^{c,d,\Lambda}$$

is uniformly continuous. To this end, we need to show that, for every $f \in C_\mathrm{b}(\mathcal{P}(E), \mathbb{R}), \theta \mapsto \langle \nu_\theta^{c,d,\Lambda}, f\rangle$ is uniformly continuous. Since the family $\{H(\cdot, C_0), C \in \Pi\}$ is dense in $C_\mathrm{b}(\mathcal{P}(E), \mathbb{R})$, we can approximate $f$ by duality functions in the sup-norm. It is therefore enough to show uniform continuity for the duality function (uniformly on the family). For this purpose, we analyse the limiting random variable for the dual as a function of $\theta$ in the limit as $t \to \infty$. The necessary uniformity follows once we establish the existence of an entrance law starting from $(\{1\}, \{2\}, \cdots)$.

Assume first that $\Lambda$ is such that the single-site $\Lambda$-coalescent comes down from $\infty$. Namely, if we denote by $(C_t^{c,\Lambda})_{t\geq 0}$ the entrance law at time 0 from the state $\{1\}, \{2\}, \cdots$ of the $\Lambda$-coalescent with jumps to the cemetery $\{*\}$ at rate $c$, then $H(\theta, C_\infty^{c,\Lambda})$ determines the McKean-Vlasov limiting law as $t \to \infty$ uniquely. Recall that we associate the value $\theta$ with the cemetery state. The fact that $C^{c,\Lambda}$ is an entrance law follows from the projective property and the fact that

(4.4) the $\Lambda$-coalescent comes down from $\infty$ at the site 0 (not at $\{*\}$!).

Namely, subject to (4.4), on the states, where rates are positive, we only have finitely many partition elements, and it is clear that $C_\infty^{c,\Lambda} = \lim_{t\to\infty} C_t^{c,\Lambda}$ exists. The random variable $C_\infty^{c,\Lambda}$ has partition elements that are all located at the cemetery state, and hence this holds uniformly for all coalescents starting in $n \in \mathbb{N}$ partition elements.

Let

$$(4.5)\quad P_{n,k} = \mathbb{P}\big\{|C_\infty^{c,\Lambda}| = k \mid C_0^{c,\Lambda} = \{\{1\}, \dots, \{n\}\}\big\}.$$

Then, for all $H(\underline{\theta}, (\{1\}, \dots, \{n\})) = \langle \theta, f\rangle^n$, $\theta \in \mathcal{P}(E)$ and $f \in C_\mathrm{b}(E, \mathbb{R})$, we have

$$(4.6)\quad H(\underline{\theta}, C_\infty^{c,\Lambda}) = \sum_{k=1}^{n} P_{n,k} \langle \theta, f\rangle^k.$$

This function is uniformly continuous in $\theta$ for any given parameter $n \in \mathbb{N}$.

The continuity property is now immediate, since the algebra generated by the monomials forms a dense subset in $C_\mathrm{b}(\mathcal{P}(E), \mathbb{R})$.

**Remark 4.2.** *By Limic and Sturm* [LS06, Theorem 12]*, under the dust-free condition* (1.3)*,*

$$(4.7)\quad \lim_{n\to\infty} P_{n,k} = P_{\infty,k}, \quad \sum_{k\in\mathbb{N}} P_{\infty,k} = 1. \quad \square$$

## 4.3 Structure of the McKean-Vlasov equilibrium

In the case of the McKean-Vlasov Fleming-Viot processes, the equilibrium $\nu_\theta^{c,d,\delta_0}$ can be identified as an atomic measure of the form

$$(4.8)\quad \sum_{i\in\mathbb{N}} \Big[W_i \prod_{j=1}^{i-1} (1 - W_j)\Big]\, \delta_{U_i}$$

with $(U_i)_{i\in\mathbb{N}}$ i.i.d. $\theta$-distributed and $(W_i)_{i\in\mathbb{N}}$ i.i.d. $\mathrm{BETA}(1, \frac{c}{d})$-distributed, independently of each other (cf. [DGV95]). What we can say about the equilibrium $\nu_\theta^{c,d,\Lambda}$?

**Proposition 4.3. [Representation of McKean-Vlasov equilibrium]** *Let $\nu_\theta^{c,d,\Lambda}$ be the equilibrium of the process $Z_\theta^{c,d,\Lambda} = (Z_\theta^{c,d,\Lambda}(t))_{t\geq 0}$ with resampling constant $d$ and resampling measure $\Lambda \in \mathcal{M}_f([0,1])$. Assume that $\Lambda$ is dust-free (recall* (1.3)*).*

(a) *The following decomposition holds:*

$$(4.9)\ \nu_\theta^{c,d,\Lambda} = \mathcal{L}\left[\sum_{i\in\mathbb{N}} V_i \delta_{U_i}\right].$$

*Here,* $(V_i)_{i\in\mathbb{N}}$ *and* $(U_i)_{i\in\mathbb{N}}$ *are independent sequences of random variables taking values in* $[0,1]$*, respectively,* $\mathcal{P}(E)$*. Moreover,* $(U_i)_{i\in\mathbb{N}}$ *is i.i.d. with distribution* $\theta$*,* $\sum_{i\in\mathbb{N}} V_i = 1$ *a.s. and*

$$(4.10)\ V_i = W_i \prod_{j=1}^{i-1}(1-W_j),$$

*where*

$$(4.11)\ (W_j)_{j\in\mathbb{N}}$$

*is i.i.d.* $[0,1]$*-valued with some distribution* $\rho$*. This distribution is uniquely determined by the moment measures of* $\nu_\theta^{c,d,\Lambda}$ *(which can be expressed in terms of the dual coalescent process) and depends on* $c, d$ *and* $\Lambda$*.*

(b) *If* $\theta \notin M = \{\delta_u \colon u \in E\}$ *and* $c, d > 0$*, then*

$$(4.12)\ 0 \le \nu_\theta^{c,d,\Lambda}(M) < 1.$$

*Proof.*

(a) The distribution and the independence of $(U_i)_{i\in\mathbb{N}}$ follow from the representation of the state at time $t \in [0,\infty]$ in terms of the entrance law of the $\Lambda$-coalescent starting from the partition $(\{1\},\{2\},\cdots)$. This representation is a consequence of the duality relation in (2.40) and de Finetti's theorem, together with the dust-free condition on $\Lambda$ in (1.3), which guarantees the existence of the frequencies of the partition elements at time $t$. Indeed, every state, including the equilibrium state, can be written as the limit of the empirical distribution of the coalescent entrance law starting from the partition $\{\{1\},\{2\},\cdots\}$ at site 1, where we assign to each dual individual the type of its partition element at time $\infty$, drawn independently from $\theta$, the cemetery state. Here we use the fact that if we condition individuals not to coalesce with a given individual, respectively, its subsequent partition element, then the process is again a coalescent for the smaller (random) subpopulation without that individual, respectively, its subsequent partition element.

The $(V_j)_{j\in\mathbb{N}}$ are the relative frequencies of the partition elements ordered according to their smallest element. By construction, $(V_i)_{i\in\mathbb{N}}$ and $(U_i)_{i\in\mathbb{N}}$ are independent. The i.i.d. property of $(W_j)_{j\in\mathbb{N}}$ follows from the property that, after we exclude the first $j$ partition elements from the population we are left with a $\Lambda$-coalescent entrance law starting from a countable ordered population. More precisely, pick the atom corresponding to $U_1$ and call its weight $W_1$. Let $W_2$ be the fraction of the remaining mass assigned to the atom corresponding to $U_2$, etc. This defines the sequence $(W_j)_{j\in\mathbb{N}}$. From the representation of the state at time $t$ and of the equilibrium state via the coalescent process starting from the partition $(\{1\},\{2\},\dots)$, we conclude that $(W_j)_{j\in\mathbb{N}}$ is i.i.d. It remains to *identify* the law $\rho$ of $W$.

In principle, via the duality we can express the moments in equilibrium

$$(4.13)\ E_{\nu_\theta^{c,d,\Lambda}}[\langle X, f\rangle^n]$$

in terms of $\langle \theta, f\rangle^k$, $k = 1, \dots, n$, and the coalescence probabilities before the migration jumps into the cemetery state. The latter in turn can be calculated in terms of

$$(4.14)\quad c,\ d,\ r^k(1-r)^{n-k}\Lambda(\mathrm{d}r).$$

These relations uniquely determine the statistics of the atom sizes, which in turn uniquely determine the marginal distribution of the $W_i$'s via (4.10).

**Remark 4.4.** *In the case where $\Lambda = \delta_0$ (the McKean-Vlasov Fleming-Viot process) it is possible to identify the law of $W$ as the* BETA$(1, \frac{c}{d})$*-distribution. It remains an open problem to identify the law of $W$ in general as function of the ingredients in* (4.14)*. This is more complex task because of the presence of the measure $\Lambda$.*

(b) First consider the case $\Lambda = \delta_0$. Let us verify that, for $c > 0$ and $\theta \notin M$, there can be no mass in $M$. Indeed, if there would be an atom somewhere in $M$, then there would also be an atom in $M$ after we merge types into a finite type set. However, in the latter situation the $W_i$'s are BETA-distributed, hence do not have an atom at 0 or 1, and so also the law of the $V_i$'s has no atom at 0 or 1. This immediately gives the claim, because it means that $\nu_\theta^{c,d,\Lambda}(M) = 0$.

Next consider the case $\Lambda \neq \delta_0$. Then new types keep on coming in. We need to prove that the event that $\mathfrak{C}_\infty^{(\{0,*\})}$ contains more than one partition element has a positive probability. But this is obviously true when $c, d > 0$. □

## 4.4 First and second moment measure

We can identify the first and second moments of the equilibrium explicitly, and we can use the outcome to calculate the variance of $M_k^{(j)}$ for $k = 0, \dots, j$, the interaction chain defined in Section 1.5.5). Recall the definition of $\mathbb{E}_{\nu_\theta}[\mathrm{Var}_x(\psi)]$ from (1.69).

**Proposition 4.5.** **[Variance]** *For every $\psi \in \mathcal{C}_\mathrm{b}(E)$,*

$$(4.15)\quad \mathbb{E}_{\nu_\theta^{c,d,\Lambda}}\left[\mathrm{Var}_x(\psi)\right] = \int_{\mathcal{P}(E)} \nu_\theta^{c,d,\Lambda}(\mathrm{d}x)\left(\langle \psi^2, x\rangle - \langle \psi, x\rangle^2\right) = \frac{2c}{2c+\lambda+2d}\,\mathrm{Var}_\theta(\psi).$$

*Proof.* We calculate the expectation of $\langle \varphi, x\rangle$, $\varphi \in \mathcal{C}_\mathrm{b}(E)$, and $\langle \varphi, x^{\otimes 2}\rangle$, $\varphi \in \mathcal{C}_\mathrm{b}(E^2)$, in equilibrium.

It follows from (4.1) with $\nu = \nu_\theta^{c,d,\Lambda}$ that

$$(4.16)\quad n=1,\ \varphi \in \mathcal{C}_\mathrm{b}(E):\qquad 0 = c\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\,\langle \varphi, (\theta - x)\rangle,$$

i.e., $\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\,\langle \varphi, x\rangle = \langle \varphi, \theta\rangle$. It further follows that

$$(4.17)\quad \begin{aligned} n=2,\ \varphi \in \mathcal{C}_\mathrm{b}(E^2):\qquad 0 &= -2c\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\left\langle \varphi, x^{\otimes 2}\right\rangle \\ &+ c\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\left[\langle \varphi, \theta\otimes x\rangle + \langle \varphi, x\otimes\theta\rangle\right] \\ &+ 2d\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\left(\int_E x(\mathrm{d}a)\left\langle \varphi, \delta_a^{\otimes 2}\right\rangle - \left\langle \varphi, x^{\otimes 2}\right\rangle\right) \\ &+ \lambda\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\int_E x(\mathrm{d}a)\left\langle \varphi, (\delta_a - x)^{\otimes 2}\right\rangle. \end{aligned}$$

We can rewrite this relation as

$$
\begin{aligned}
&\int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \int_E x(\mathrm{d}a) \left\langle \varphi, (\delta_a - x)^{\otimes 2} \right\rangle \\
&\quad = \int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \left( \int_E x(\mathrm{d}a) \left\langle \varphi, \delta_a^{\otimes 2} \right\rangle - \left\langle \varphi, x^{\otimes 2} \right\rangle \right) \\
&\quad = \frac{c}{\lambda + 2d} \left( 2 \int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \left\langle \varphi, x^{\otimes 2} \right\rangle - \int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \left[ \langle \varphi, \theta \otimes x \rangle + \langle \varphi, x \otimes \theta \rangle \right] \right).
\end{aligned}
\tag{4.18}
$$

From this, we see that

$$
\begin{aligned}
&\int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \left\langle \varphi, x^{\otimes 2} \right\rangle \\
&\quad = \frac{\lambda + 2d}{2c + \lambda + 2d} \left( \frac{c}{\lambda + 2d} \int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \left[ \langle \varphi, \theta \otimes x \rangle \right. \right. \\
&\qquad\qquad \left. \left. + \ \langle \varphi, x \otimes \theta \rangle \right] + \int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \int_E x(\mathrm{d}a) \left\langle \varphi, \delta_a^{\otimes 2} \right\rangle \right) \\
&\quad = \frac{\lambda + 2d}{2c + \lambda + 2d} \left( \frac{2c}{\lambda + 2d} \left\langle \varphi, \theta^{\otimes 2} \right\rangle + \int_E \theta(\mathrm{d}a) \left\langle \varphi, \delta_a^{\otimes 2} \right\rangle \right),
\end{aligned}
\tag{4.19}
$$

where we use (4.16) in the last line. Substituting this back into (4.18) and using (4.16) once more, we get

$$
\begin{aligned}
&\int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \int_E \int_E Q_x(\mathrm{d}u, \mathrm{d}v)\, \varphi(u, v) \\
&\quad = \int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \left( \int_E x(\mathrm{d}a) \left\langle \varphi, \delta_a^{\otimes 2} \right\rangle - \left\langle \varphi, x^{\otimes 2} \right\rangle \right) \\
&\quad = \frac{2c}{\lambda + 2d} \left( \int_{\mathcal{P}(E)} \nu(\mathrm{d}x) \left\langle \varphi, x^{\otimes 2} \right\rangle - \left\langle \varphi, \theta^{\otimes 2} \right\rangle \right) \\
&\quad = \frac{2c}{2c + \lambda + 2d} \left( \int_E \theta(\mathrm{d}a) \left\langle \varphi, \delta_a^{\otimes 2} \right\rangle - \left\langle \varphi, \theta^{\otimes 2} \right\rangle \right) \\
&\quad = \frac{2c}{2c + \lambda + 2d} \int_E \int_E Q_\theta(\mathrm{d}u, \mathrm{d}v)\, \varphi(u, v).
\end{aligned}
\tag{4.20}
$$

Pick $\varphi = \psi \times \psi$ in (4.20) to get the claim. □

For $\lambda = \Lambda([0,1]) = 0$, (4.15) is the same as Dawson, Greven and Vaillancourt [DGV95, Eq. (2.5)].

**Corollary 4.6. [Asymptotic variance of entrance law]**
*For $\varphi \in C_{\mathrm{b}}(E, \mathbb{R})$, the interaction chain satisfies*

$$
\lim_{j \to \infty} \mathbb{E}_{\mathcal{L}(M_0^{(j)})} [\mathrm{Var}_x(\varphi)] = 0 \quad (\textit{respectively, } > 0),
\tag{4.21}
$$

*if $\sum_{k \in \mathbb{N}} m_k = \infty$ (respectively, $\sum_{k \in \mathbb{N}} m_k < \infty$) with $m_k$ defined in* (1.44)*.*

*Proof.* From (4.15), we have the formula

$$\mathbb{E}_{\nu_\theta^{c,d,\Lambda}}[\mathrm{Var}_x(\varphi)] = \frac{2c}{2c+\lambda+2d}\,\mathrm{Var}_\theta(\varphi). \tag{4.22}$$

Hence, we have the relation (recall (1.63) for the definition of $K_k(\theta, \mathrm{d}x)$)

$$\int_{\mathcal{P}(E)} K_k(\theta, \mathrm{d}x)\,\mathrm{Var}_x(\varphi) = \frac{2c_k}{2c_k+\lambda_k+2d_k}\,\mathrm{Var}_\theta(\varphi), \tag{4.23}$$

which says that in one step of the interaction chain the variance is modified by the factor

$$n_k = \frac{2c_k}{2c_k+\lambda_k+2d_k} = \frac{1}{1+m_k}. \tag{4.24}$$

Iteration gives

$$\mathbb{E}_{\mathcal{L}(M_0^{(j)})}[\mathrm{Var}_x(\varphi)] = \left(\prod_{k=0}^{j} n_k\right)\mathrm{Var}_\theta(\varphi) = \left(\prod_{k=0}^{j}\left(\frac{1}{1+m_k}\right)\right)\mathrm{Var}_\theta(\varphi). \tag{4.25}$$

Therefore, taking logarithms, we see that (4.21) is equivalent to

$$\sum_{k\in\mathbb{N}_0} m_k = \infty \text{ (respectively, } <\infty). \tag{4.26}$$

□

We next prove a result that is similar to, but more involved than, [DGV95], Eq. (6.12). This result is necessary for the proof of Theorem 1.11 on diffusive clustering.

**Proposition 4.7. [Variance of the integral against a test function]** *For every $\psi \in \mathcal{C}_\mathrm{b}(E)$, $j\in\mathbb{N}$ and $0\le k\le j+1$,*

$$\begin{aligned}\mathrm{Var}_{\mathcal{L}(M_{-k}^{(j)})}(\langle x,\psi\rangle) &= \mathbb{E}_{\mathcal{L}(M_{-k}^{(j)})}[\langle x,\psi\rangle^2] - \left(\mathbb{E}_{\mathcal{L}(M_{-k}^{(j)})}[\langle x,\psi\rangle]\right)^2\\ &= \left(\sum_{i=k}^{j}\left(\frac{d_{i+1}}{c_i}\prod_{l=i+1}^{j}\frac{1}{1+m_l}\right)\right)\mathrm{Var}_\theta(\psi).\end{aligned} \tag{4.27}$$

*Proof.* The proof uses the following two ingredients. Combining (4.15) and (4.24), we have

$$\mathbb{E}_{\nu_\theta^{c_k,d_k,\Lambda_k}}[\mathrm{Var}_x(\psi)] = \frac{1}{1+m_k}\,\mathrm{Var}_\theta(\psi). \tag{4.28}$$

The first and the third line of (4.20) yield

$$\mathrm{Var}_\nu(\langle x,\psi\rangle) = \frac{\lambda+2d}{2c}\,\mathbb{E}_\nu[\mathrm{Var}_x(\psi)]. \tag{4.29}$$

Together with (4.15) and (1.42), we therefore obtain

$$\mathrm{Var}_{\nu_\theta^{c_k,d_k,\Lambda_k}}(\langle x,\psi\rangle) = \frac{\lambda_k+2d_k}{2c_k+\lambda_k+2d_k}\,\mathrm{Var}_\theta(\psi) = \frac{d_{k+1}}{c_k}\,\mathrm{Var}_\theta(\psi). \tag{4.30}$$

Fix $j \in \mathbb{N}$. The proof follows by downward induction over $0 \leq k \leq j+1$. The initial case $k = j+1$ is obvious because $M^{(j)}_{-(j+1)} = \theta$ by (1.62). Let us therefore assume that the claim holds for $k+1$. By (1.62–1.63),

$$\begin{aligned} \operatorname{Var}_{\mathcal{L}(M^{(j)}_{-k})}(\langle x,\psi\rangle) &= \mathbb{E}_{\mathcal{L}(M^{(j)}_{-k})}[\langle x,\psi\rangle^2] - \Big(\mathbb{E}_{\mathcal{L}(M^{(j)}_{-k})}[\langle x,\psi\rangle]\Big)^2 \\ &= \int_{\mathcal{P}(E)} \nu_{\theta}^{c_j,d_j,\Lambda_j}(\mathrm{d}\theta_j) \int_{\mathcal{P}(E)} \nu_{\theta_j}^{c_{j-1},d_{j-1},\Lambda_{j-1}}(\mathrm{d}\theta_{j-1}) \ldots \int_{\mathcal{P}(E)} \nu_{\theta_{k+1}}^{c_k,d_k,\Lambda_k}(\mathrm{d}\theta_k)\langle\theta_k,\psi\rangle^2 - \langle\theta,\psi\rangle^2. \end{aligned} \tag{4.31}$$

Next, use (4.30) to rewrite the inside integral as

$$\int_{\mathcal{P}(E)} \nu_{\theta_{k+1}}^{c_k,d_k,\Lambda_k}(\mathrm{d}\theta_k)\langle\theta_k,\psi\rangle^2 = \mathbb{E}_{\nu_{\theta_{k+1}}^{c_k,d_k,\Lambda_k}}(\langle x,\psi\rangle^2) = \langle\theta_{k+1},\psi\rangle^2 + \frac{d_{k+1}}{c_k}\operatorname{Var}_{\theta_{k+1}}(\psi). \tag{4.32}$$

Substitute this back into (4.31), to obtain

$$\begin{aligned} &\operatorname{Var}_{\mathcal{L}(M^{(j)}_{-k})}(\langle x,\psi\rangle) = \operatorname{Var}_{\mathcal{L}(M^{(j)}_{-(k+1)})}(\langle x,\psi\rangle) \\ &\quad + \frac{d_{k+1}}{c_k}\int_{\mathcal{P}(E)} \nu_{\theta}^{c_j,d_j,\Lambda_j}(\mathrm{d}\theta_j)\int_{\mathcal{P}(E)} \nu_{\theta_j}^{c_{j-1},d_{j-1},\Lambda_{j-1}}(\mathrm{d}\theta_{j-1}) \\ &\qquad \cdots \int_{\mathcal{P}(E)} \nu_{\theta_{k+2}}^{c_{k+1},d_{k+1},\Lambda_{k+1}}(\mathrm{d}\theta_{k+1})\operatorname{Var}_{\theta_{k+1}}(\psi). \end{aligned} \tag{4.33}$$

The first term is given by the induction hypothesis. For the second term we use (4.28), to see that the inside integral equals

$$\int_{\mathcal{P}(E)} \nu_{\theta_{k+2}}^{c_{k+1},d_{k+1},\Lambda_{k+1}}(\mathrm{d}\theta_{k+1})\operatorname{Var}_{\theta_{k+1}}(\psi) = \mathbb{E}_{\nu_{\theta_{k+2}}^{c_{k+1},d_{k+1},\Lambda_{k+1}}}(\operatorname{Var}_x(\psi)) = \frac{1}{1+m_{k+1}}\operatorname{Var}_{\theta_{k+2}}(\psi). \tag{4.34}$$

Iteration of this reasoning for the second term in (4.33) leads to

$$\begin{aligned} \operatorname{Var}_{\mathcal{L}(M^{(j)}_{-k})}(\langle x,\psi\rangle) &= \operatorname{Var}_{\mathcal{L}(M^{(j)}_{-(k+1)})}(\langle x,\psi\rangle) + \frac{d_{k+1}}{c_k}\prod_{l=k+1}^{j}\frac{1}{1+m_l}\operatorname{Var}_\theta(\psi) \\ &= \left(\sum_{i=k+1}^{j}\left(\frac{d_{i+1}}{c_i}\prod_{l=i+1}^{j}\frac{1}{1+m_l}\right)\right)\operatorname{Var}_\theta(\psi) + \frac{d_{k+1}}{c_k}\prod_{l=k+1}^{j}\frac{1}{1+m_l}\operatorname{Var}_\theta(\psi), \end{aligned} \tag{4.35}$$

which proves the claim. □

If $\lambda_k = \Lambda_k([0,1]) = 0$, $k \in \mathbb{N}_0$, then (4.27) reduces to [DGV95], Eq. (6.12). Indeed, in that case $d_{i+1}\prod_{l=i+1}^{j}\frac{1}{1+m_l}$ is equal to $d_{j+1}$. (Note the typo in [DGV95], Eq. (6.12): $d_k$ should be replaced by $d_{k+1}$.)

**Remark 4.8.** *The results in this section can alternatively be inferred from the long-time behaviour of the spatial* $\Lambda$*-coalescent with* $G = \{0,*\}$.

# 5 Strategy of the proof of the main scaling theorem

The proof of Theorem 1.4 will be carried out in Sections 6–8. In this section we explain the main line of the argument.

## 5.1 General scheme and three main steps

In Dawson, Greven and Vaillancourt [DGV95], a general scheme was developed to derive the scaling behaviour of space-time block averages for hierarchically interacting Fleming-Viot processes, with the interaction coming from migration, i.e., a system similar to ours but *without* $\Lambda$-Cannings block resampling (so for $\Lambda = \delta_0$, which results in diffusion processes rather than jump processes). Nevertheless, this scheme is widely applicable and indicates what estimates have to be established in a concrete model (with methods that may be specific to that model).

For our model, the difficulty sits in the fact that *diffusions* are replaced by *jump processes*, even in the many-individuals-per-site limit. Below we explain how we can use the special properties of the dual process derived in Section 2 to deal with this difficulty. In Sections 6–8 the various steps will be carried out in detail to prove our scaling result in Theorem 1.4. In these sections we will mainly focus on the new features coming from the $\Lambda$-Cannings block resampling. The refined multi-scale result in Theorem 1.8 will be proved in Section 9. This can be largely based on earlier work ([DGV95, Section 4]), where the line of argument was developed in detail for Fleming-Viot and needs no new ideas for Cannings process: it only requires carrying out a new moment calculation.

The analysis in Sections 6-8 proceeds in three main steps:

- Show that for the mean-field system, i.e., $G = G_{N,1} = \{0, 1, \ldots, N-1\}$, in the limit as $N \to \infty$ we obtain for single sites on time scale $t$ independent McKean-Vlasov processes, and for block averages on time scale $Nt$ Fleming-Viot processes with a resampling constant $d_1$ corresponding to $\Lambda_0$ and $c_0$. With an additional $\Lambda_1$-block resampling at rate $N^{-2}$ there is no effect on time scale $t$, and so on time scale $Nt$ we obtain a $C^{\widetilde{\Lambda}}$-process with $\widetilde{\Lambda} = d_1\delta_0 + \Lambda_1$. This is done in Section 6.

- Consider the $C_N^{\underline{c},\underline{\Lambda}}$-process restricted to $G_{N,K}$. More precisely, study its components and its $k$-block averages for $1 \leq k \leq j < K$ on time scales $N^j + tN^k$. This is done in Section 7.

- Treat the $(j,k)$ renormalised systems for $1 \leq k \leq j < K$, approximating the $C_N^{\underline{c},\underline{\Lambda}}$-process on $\Omega_N$, in the limit as $N \to \infty$ and on time scales at most $N^K t$ for a fixed but otherwise arbitrary $K \in \mathbb{N}$, by the process on $G_{N,K}$ from the previous step. This is done in Section 8.

The three steps above are carried out following the scheme of proof developed in [DGV95]. What is new for *jump* processes? The key difference is that now semi-martingales (arising from functionals of the process) are no longer controlled just by the *compensator* and the *increasing process* of the linear functional $\langle X_t, f\rangle$, where $X = (X_t)_{t\geq 0}$ is the process in question and $f \in C_{\mathrm{b}}(E)$, as in the case of diffusions where linear and quadratic functions $\langle X_t, f\rangle, \langle X_t, f\rangle^2$ in $\mathcal{F}$ suffice to establish both *tightness in path space* and *convergence of finite-dimensional distributions.* The new ingredients are the analysis of the linear operators of the martingale problem acting on all of $\mathcal{F}$, and the extension of the tightness arguments to handle the jumps. This relies heavily on the duality relation to the $\underline{\Lambda}$-coalescent with block coalescence.

### 5.2 Embedding

In the proofs, we view the $C_N^{\underline{c},\underline{\Lambda}}$-process the $C^\Lambda$-processes with $G = \{0,1,\dots,N-1\}$, $G = G_{N,K} = \{0,1,\dots,N-1\}^K$ and $G = \Omega_N$, as embedded in a process for the choices $G = \mathbb{N}$, $G = \mathbb{N}^K$ and $G = \Omega_\infty$, where

$$(5.1)\ \ \Omega_\infty = \bigcup_{M\in\mathbb{N}} \Omega_M \subseteq \mathbb{N}^{\mathbb{N}}.$$

Note that $\Omega_\infty$ is *countable*, but that the $\Omega_M$'s are *not* subgroups of $\Omega_\infty$. The embedding requires us to embed the test functions and the generators on $\Omega_M$ into those on $\Omega_\infty$. In the calculations in Sections 6–8, we use this embedding without writing it out formally.

Proving weak convergence in path space for solutions of martingale problems with operators acting only on functions of the current state reduces to showing convergence for generators, combined with the compact containment condition for the path. Often, for processes with values in *compact* state spaces, the latter follows from bounds on the generator as well. More precisely, we have to choose a dense subset $\mathcal{A}$ of $C_{\mathrm{b}}((\mathcal{P}(E))^{\mathbb{N}},\mathbb{R})$ and show that the compensator terms satisfy, for all $F \in \mathcal{A}$

$$(5.2)\ \ \mathcal{L}\left[\left(\int_0^t L^{(G_N)}F((X_s^N))\mathrm{d}s\right)_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(\int_0^t (L^{(G)}F)(X_s)\mathrm{d}s\right)_{t\geq 0}\right].$$

We can then conclude that the limit points satisfy the desired martingale problem, after which verification of the well-posedness of the latter gives convergence of the finite-dimensional distributions. Once we can guarantee tightness in path space via the compensator convergence, we are done.

The above procedure works well in the case of Fleming-Viot processes. In our case, since we have jump processes, an additional argument is needed for the tightness in path space. This argument will be based on Jakubowski's criterion, which reduces the tightness of measure-valued processes to collections of real-valued processes (semi-martingales), whose tightness in turn can be based on the characteristics of the semi-martingales. The latter requires an additional argument to cope with the jumps, but is still essentially based on generator calculations.

In summary, the role of Sections 6–8 is to carry out first some generator calculations and then an asymptotic evaluation of the resulting expressions, leading to a limiting form. There we will use an *averaging principle* for local variables based on the local equilibrium dictated by the macroscopic slowly changing variable based on the idea of *local equilibria.*

## 6 The mean-field limit of $C^\Lambda$-processes

This section deals with the case $G = \{0,1,\dots,N-1\}$ for a model that includes mean-field migration and Cannings reproduction at rate 1 with resampling measure $\Lambda_0$ in single colonies. We analyse the single components and the block averages on time scales $t$, $Nt$ and $Nt+u$ with $u \in \mathbb{R}$. The key results are formulated in Propositions 6.1 and 6.3 below. We will see that we can also incorporate block resampling at rate $N^{-2}\Lambda_1$ and still get the same results.

The analysis for mean-field interacting Fleming-Viot processes with drift is given in detail in [DGV95, Section 4]. The reader unfamiliar with the arguments involved is referred to this paper (see, in particular, the outline of the abstract scheme in [DGV95, Section 4(b)(i), pp. 2314–2315]). In what follows, we provide the main ideas again, and focus on the changes arising from the replacement of the Fleming-Viot process by the $\Lambda$-Cannings resampling process, i.e., the change from continuous to càdlàg semi-martingales.

We always start the process in an

$$\text{(6.1)}\quad \text{i.i.d. random state with mean-measure } \mathbb{E}[X_0^{(N)}(0)] = \theta \in \mathcal{P}(E).$$

The system will be analysed in the limit as $N \to \infty$ in two steps: (1) component-wise on time scale $t$ (Section 6.1); (2) block-wise on time scale $Nt$ and component-wise on time scale $Nt + u$ with $u \in \mathbb{R}$ (Section 6.2).

## 6.1 Propagation of chaos: Single colonies and the McKean-Vlasov process

In this section, we consider the $C^\Lambda$-mean-field model in Section 1.3.2 with $G = \{0, 1, \ldots, N-1\}$. We prove propagation of chaos for the collection $(\{X_0^{(N)}(t), \ldots, X_{N-1}^{(N)}(t)\})_{t\geq 0}$ in the limit as $N \to \infty$, i.e., we prove asymptotic independence of the components via duality and component-wise convergence to the McKean-Vlasov process with parameters $d_0 = 0, c_0, \Lambda_0, \theta$.

**Proposition 6.1. [McKean-Vlasov limit, propagation of chaos]**
*Under assumption* (6.1)*, for any* $L \in \mathbb{N}$ *fixed,*

$$\text{(6.2)}\quad \mathcal{L}\left[(X_0^{(N)}(t), \ldots, X_L^{(N)}(t))_{t\geq 0}\right] \underset{t\to\infty}{\Longrightarrow} \left(\mathcal{L}\left[Z_\theta^{c_0,d_0,\Lambda_0}\right]\right)^{\otimes(L+1)}$$

*with* $Z_\theta^{c_0,d_0,\Lambda_0}$ *as in* (1.19)*.*

**Corollary 6.2. [McKean-Vlasov limit with block resampling]**
*Consider the system above with an additional rate* $N^{-2}\Lambda_1$ *of block resampling. Then (6.2) continues to hold.*

In order to prove (6.2), we will argue that the laws $\mathcal{L}[(\{X_\xi^{(N)}(t), \xi = 0, \ldots, L\})_{t\geq 0}]$, $N \in \mathbb{N}$, are tight by showing this first for components (Section 6.1.1) and characterise the weak limit points. In order to carry this out, we verify asymptotic independence (Section 6.1.2), calculate explicitly the action of the generator on the test functions in the martingale problem of $X^{(N)}$ (Section 6.1.3), and show, for functions depending on one component, uniform convergence to the generator of the McKean-Vlasov operator with parameter $\theta = \mathbb{E}[X_0^{(N)}(0)]$ (Section 6.1.4).

### 6.1.1 Tightness on path space in $N$

We focus first on one component $(X_\xi(t))_{t\geq 0}$. We use Jakubowski's criterion for measure-valued processes (see [D93, Theorem 3.6.4]). This requires us to show: (1) a compact containment condition for the path, i.e., for all $\epsilon, T > 0$ there exists a $K_{T,\epsilon}$ compact such that

$$\text{(6.3)}\quad \mathbb{P}(\{X_\xi^{(N)}(t) \in K_{T,\epsilon} \text{ for all } t \in [0,T]\}) \geq 1 - \varepsilon;$$

(2) tightness of certain evaluation processes $(F(X^{(N)}(t)))_{t\geq 0}$ in path space. In our case the compact containment condition in (1) is immediate because we have a compact type space. Condition (2) can be verified by using a criterion for tightness of Kurtz (see Dawson [D93, Corollary 3.6.3]). Here, we use test functions as in (1.9) that only depend on the first $L$ coordinates. We further make use of the boundedness of the characteristics of the generator

as a function of $N$ when acting on a test function (recall (1.38)). Namely, we will see in Sections 6.1.3 that

$$\sup_{N\in\mathbb{N}} \|L^{(N)}F\|_\infty < \infty, \quad \text{for all } F \in \mathcal{F}. \tag{6.4}$$

This allows us to verify that linear combinations of

$$\{\mathcal{L}[(\langle X_\xi^{(N)}(t), f\rangle^n)_{t\geq 0}]\colon\ n\in\mathbb{N}, f\in C_{\mathrm{b}}(E,\mathbb{R})\} \tag{6.5}$$

are tight, which verifies the tightness criterion in (2).

In order to verify (6.5), we use a criterion for the tightness of semi-martingales based on the local characterisation via the Joffe-Métivier criterion ([JM86], see also [D93, Theorem 3.6.6.]).

### 6.1.2 Asymptotic independence

In this section, we use duality to prove the factorisation of spatial mixed moments (including the case with block coalescence at rate $N^{-2}\Lambda_1$):

$$\limsup_{N\to\infty} \left| \mathbb{E}\left[\prod_{\xi=0}^{L} \Big(\langle X_\xi^{(N)}(t), f_\xi\rangle\Big)^{k_\xi}\right] - \prod_{\xi=0}^{L} \mathbb{E}\left[\Big(\langle X_\xi^{(N)}(t), f_\xi\rangle\Big)^{k_\xi}\right]\right| = 0, \quad \text{for all } t\geq 0. \tag{6.6}$$

A similar result holds for mixed moments over different time points.

*Proof of* (6.6). Obviously, no block coalescence takes place in the time interval $[0,T]$ in the limit as $N\to\infty$. We verify the remaining claim by showing that any two partition elements of the dual process never meet, so that for $n$ partition elements none of the possible pairs will ever meet. Indeed, the probability for two random walks to meet is the waiting time for the rate-$2c_0$ random walk to hit 2 starting from 1. This waiting time is the sum of a geometrically distributed number of jumps with parameter $N^{-1}$, each occurring after an $\exp(2c_0)$-distributed waiting time. By explicit calculation, the probability for this event to occur before time $t$ is $O(N^{-1})$, which gives the claim. □

### 6.1.3 Generator convergence

In order to show the convergence of $L^{(N)}F$, we investigate the migration and the resampling part separately.

• **Migration part.** Recall from (1.12) that the migration operator for the geographic space $G = G_{N,1} = \{0,1,\dots,N-1\}$ is

$$(L_{\mathrm{mig}}^{(N)}F)(x) = \frac{c_0}{N}\sum_{\xi,\zeta\in G_{N,1}} \int_E (x_\zeta - x_\xi)(\mathrm{d}a)\,\frac{\partial F(x)}{\partial x_\xi}[\delta_a], \tag{6.7}$$

where $F\in\mathcal{F}\subset C_{\mathrm{b}}(\mathcal{P}(E)^N,\mathbb{R})$, with $\mathcal{F}$ the algebra of functions of the form (1.9). We rewrite (6.7) as

$$\begin{aligned}(L_{\mathrm{mig}}^{(N)}F)(x) &= c_0 \sum_{\xi\in G_{N,1}} \int_E \frac{1}{N}\sum_{\zeta\in G_{N,1}} (x_\zeta - x_\xi)\,(\mathrm{d}a)\,\frac{\partial F(x)}{\partial x_\xi}[\delta_a]\\ &= c_0\sum_{\xi\in G_{N,1}} \int_E (y - x_\xi)\,(\mathrm{d}a)\,\frac{\partial F(x)}{\partial x_\xi}[\delta_a],\end{aligned} \tag{6.8}$$

where $y = N^{-1}\sum_{\zeta=0}^{N-1} x_\zeta = N^{-1}\sum_{\zeta\in G_{N,1}} x_\zeta$ denotes the block average. We will show that, in the limit $N\to\infty$, $(L^{(N)}_{\mathrm{mig}}F)(x)$ only depends on the mean type measure $\theta$ of the initial state, i.e., it converges to

$$(L^{c_0}_\theta F)(x) = c_0 \sum_{\xi\in\mathbb{N}_0} \int (\theta - x_\xi)(\mathrm{d}a) \frac{\partial F(x)}{\partial x_\xi}[\delta_a], \tag{6.9}$$

where we use for this generator acting on $C_{\mathrm{b}}(\mathcal{P}(E))^{\mathbb{N}},\mathbb{R})$ the same notation we used for the McKean-Vlasov process with immigration-emigration on $\mathcal{P}(E)$. Furthermore, we show that

(6.10) $\theta\mapsto L^{c_0}_\theta F \in C_{\mathrm{b}}(\mathcal{P}(E),\mathbb{R})$ is continuous for all $\theta\in\mathcal{P}(E)$.

To show the convergence, define

$$\mathbb{B}_\theta = \Big\{x\in(\mathcal{P}(E))^{\mathbb{N}}\colon\ N^{-1}\sum_{\xi\in G_{N,1}} x_\xi \underset{N\to\infty}{\longrightarrow} \theta\Big\} \subseteq (\mathcal{P}(E))^{\mathbb{N}}, \tag{6.11}$$

and

$$\mathbb{B} = \bigcup_{\theta\in\mathcal{P}(E)} \mathbb{B}_\theta. \tag{6.12}$$

If we have an i.i.d. initial law (respectively, an exchangeable law) with mean measure $\theta$, then the process $X^{(N)}$ satisfies

$$\mathcal{L}[X^{(N)}(t)](\mathbb{B}) = 1 \quad (\text{respectively, } \mathcal{L}[X^{(N)}(t)(\mathbb{B}_\theta)] = 1). \tag{6.13}$$

Indeed, as we will see in Section 6.2, $Y^{(N)}_{\xi,1}$ (recall (1.40)) evolves on time scale $Nt$. More precisely, $(Y^{(N)}_{\xi,1}(tN))_{t\geq 0}$ is tight in path space and therefore converges over a finite time horizon to the mean type measure $\theta$ of the initial state. In a formula (the right-hand side means a constant path):

$$\mathcal{L}[(Y^{(N)}_{\xi,1}(t))_{t\in[0,T]}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(\underline{\theta})_{t\in[0,T]}]. \tag{6.14}$$

Therefore, on $\mathbb{B}_\theta$ we have

$$|(L^{(N)}_{\mathrm{mig}}F)(x) - (L^{c_0}_\theta F)(x)| \underset{N\to\infty}{\longrightarrow} 0, \quad \text{for all } x\in\mathbb{B}_\theta, \tag{6.15}$$

Hence, on the path space, we have

$$\mathcal{L}\left[\left(\left|\int_0^t (L^{(N)}_{\mathrm{mig}}F)(X^N(s))\mathrm{d}s - \int_0^t (L^{c_0}_{Y^{(N)}_\xi(s)}F)(X^N(s))\mathrm{d}s\right|\right)_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \delta_{\underline{0}}. \tag{6.16}$$

• **Resampling part.** The action of the resampling term on each component (recall (1.14)) does not depend on $N$ and hence we obtain, by the law of large numbers for the marking operation (recall that $F$ depends on finitely many coordinates only)

$$|(L^{(N)}_{\mathrm{res}}F)(x) - (L^{\Lambda_0}F)(x)| \underset{N\to\infty}{\longrightarrow} 0, \quad \text{for all } x\in(\mathcal{P}(E)^{\mathbb{N}}), \tag{6.17}$$

where

$$\begin{aligned}&(L^{\Lambda_0}F)(x)\\ &= \sum_{\xi\in\mathbb{N}_0}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\Big[F\big(x_0,\dots,x_{\xi-1},(1-r)x_\xi + r\delta_a, x_{\xi+1},\dots,x_{N-1}\big) - F(x)\Big].\end{aligned} \tag{6.18}$$

### 6.1.4 Convergence to the McKean-Vlasov process

In what follows, we fix $\xi \in \mathbb{N}_0$ and let

$$(6.19)\quad G(x_\xi) = \int_{E^n} x_\xi^{\otimes n}(\mathrm{d}u)\,\varphi(u) = \left\langle \varphi, x_\xi^{\otimes n} \right\rangle, \qquad n \in \mathbb{N}, \varphi \in C_\mathrm{b}(E^n, \mathbb{R}).$$

We know that $(X_\xi^{(N)}(t))_{\xi\in\mathbb{N}_0}$ is tight and that all weak limit points are systems of independent random processes ("propagation of chaos"). It remains to identify the unique marginal law.

Let the initial condition $(X_\xi^{(\infty)}(0))_{\xi\in\mathbb{N}_0}$ be i.i.d. $\mathcal{P}(E)$-valued random variables with intensity measure $\theta$. Then each single component converges and the limiting coordinate process has generator

$$(6.20)\quad \begin{aligned} (L_\theta^{c_0,0,\Lambda_0}G)(x_\xi) =& c_0 \int_E (\theta - x_\xi)(\mathrm{d}a)\,\frac{\partial G(x_\xi)}{\partial x_\xi}[\delta_a] \\ &+ \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\,\big[G\big((1-r)x_\xi + r\delta_a\big) - G(x_\xi)\big], \end{aligned}$$

where $\theta \in \mathcal{P}(E)$ is the initial mean measure. Indeed, we may now reason as in [D93, the second part of Section 2.9]. Tightness of the processes $(X^{(N)}(t))_{t\geq 0}$ was shown in Section 6.1.1. Fix $\xi \in \mathbb{N}_0$ and consider a convergent subsequence $(X_\xi^{(N_k)}(t))_{t\geq 0}$, $k \in \mathbb{N}$. We claim that the limiting process is the unique solution to the well-posed martingale problem with corresponding generator $L_\theta^{c_0,0,\Lambda_0}$ and initial condition $\theta$. Recall from Section 6.1.3 that, for all test functions $F \in \mathcal{F}$,

$$(6.21)\quad \mathcal{L}\left[\left(\int_0^t (L_{\mathrm{mig}}^{(N)} + L_{\mathrm{Res}}^{(N)})(F)(X^N(s))\mathrm{d}s\right)_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(\int_0^t L_\theta^{c_0,d_0,\Lambda_0}(X^\infty(s))\mathrm{d}s\right)_{t\geq 0}\right]$$

in the sense of processes. Hence all weak limit points of $X^{(N)}$ solve the $L_\theta^{c_0,d_0,\Lambda_0}$-martingale problem of Section 1.3.3. The right-hand side is the compensator of a well-posed martingale problem, and hence we have convergence.

## 6.2 The mean-field finite-system scheme

In this section we verify the mean-field "finite system scheme" for the $C^\Lambda$-process, i.e., we consider $L+1$ tagged sites $\{X_0^{(N)}(t), \ldots, X_L^{(N)}(t)\}$ and the block average $Y^{(N)}(t) = N^{-1}\sum_{\xi\in G_{N,1}} X_\xi^{(N)}(t)$ and we prove:

- convergence of $(Y^{(N)}(Nt))_{t\geq 0}$ to the Fleming-Viot diffusion $Y(t) = Z_\theta^{0,d_1,0}(t)$ with parameter $d_1 = \frac{c_0\lambda_0}{2c_0+\lambda_0}$ and initial state $\theta$;
- convergence of the components $(\{X_\xi^{(N)}(Nt+u), \xi = 0, \ldots, L\})_{u\geq 0}$ to the equilibrium McKean-Vlasov process with immigration-emigration $(Z_{\theta(t)}^{c_0,d_0,\Lambda_0}(u))_{u\geq 0}$ starting from distribution $\nu_{\theta(t)}^{c_0,d_0,\Lambda_0}$ (recall (4.1)) with $\theta(t) = Y(t)$ (recall that $d_0 = 0$).

**Proposition 6.3. [Mean-field finite system scheme]**
*For initial laws with i.i.d. initial configuration and mean measure $\theta$,*

$$(6.22)\quad \mathcal{L}[(Y^{(N)}(Nt))_{t\geq 0}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(Z_\theta^{0,d_1,\delta_0}(t))_{t\geq 0}]$$

*with* $d_1 = \frac{c_0\lambda_0}{2c_0+\lambda_0}$. *Moreover, for every* $u \in \mathbb{R}$ *and* $L \in \mathbb{N}$,

$$(6.23)\quad \mathcal{L}[(X_\xi^{(N)}(Nt+u))_{\xi=0,\dots,L}] \underset{N\to\infty}{\Longrightarrow} \int_{\mathcal{P}(E)} P_t(\mathrm{d}\theta')\left(\nu_{\theta'}^{c_0,d_0,\Lambda_0}\right)^{\otimes L+1} \qquad \text{with } P_t = \mathcal{L}[Z_\theta^{0,d_1,\delta_0}(t)].$$

**Corollary 6.4. [Mean-field finite system scheme with $\Lambda_1$-block resampling]** *Consider the model above with additional block resampling at rate* $N^{-2}\Lambda_1$. *Then, in the right-hand side of* (6.22), $Z_\theta^{0,d_1,\delta_0}$ *must be replaced by* $Z_\theta^{0,d_1,\Lambda_1}$, *and similarly in the definition of* $P_t$ *in* (6.23).

The proof of the mean-field finite system scheme follows the abstract argument developed in [DGV95]. Namely, we first establish tightness of the sequence of processes $(Y^{(N)}(Nt))_{t\geq 0}$, $N \in \mathbb{N}$, which can be done as in Section 6.1.1 for $(X_0^{(N)}(t),\dots,X_L^{(N)}(t))_{t\geq 0}$, $N\in\mathbb{N}$, once we have calculated the generators. A representation for the generator of the process is found in Sections 6.2.1–6.2.2 below. This can be pursued, with the help of the idea of local equilibria based on the ergodic theorems of Section 4, to obtain first (6.23) and then (6.22).

In Sections 6.2.1-6.2.2, we calculate the action of the generator of the martingale problem on the test functions induced by the functions necessary to arrive at the action of the generator of the limiting process. In Section 6.2.4, we pass to the limit $N\to\infty$, where as in Section 6.1, we have to use an averaging principle. However, instead of a simple law of large numbers, this now is a *dynamical averaging principle* with *local equilibria* for the single components necessary to obtain the expression for the limiting block-average process.

By the definition of the generator of a process, $M^{x,F} = (M_t^{x,F})_{t\geq 0}$,

$$(6.24)\quad M_t^{x,F} = F(x_t) - F(x_0) - \int_0^t \mathrm{d}s\left(L_{\mathrm{mig}}^{(N)}F + L_{\mathrm{res}}^{(N)}F\right)(x_s)$$

is a martingale for all $F$, as in (6.19). The same holds with $x$ replaced by the block averages $y$ (by the definition of $y$). Once again, we will investigate the migration and the resampling operator separately, this time for the block average.

### 6.2.1 Migration

In this section, we consider functions $F\circ y$ with $F$ as in (6.19) and $y$ a block average (which is equivalent to choosing $F$ as in (1.9) with $N=1$). We will show below that $L_{\mathrm{mig}}^{(N)}(F\circ y)=0$, so that migration has no effect.

Recall $(L_{\mathrm{mig}}^{(N)}F)(x)$ as rewritten in (6.8). For the block averages $y$ the migration operator can be calculated as follows. Since $y=y(x)$ and $F(y)=(F\circ y)(x)$ can be seen as functions of $x$ in the algebra $\mathcal{F}$ of functions in $x$ of the form (6.19), we have

$$(6.25)\quad (L_{\mathrm{mig}}^{(N)}F)(y) = \left(L_{\mathrm{mig}}^{(N)}(F\circ y)\right)(x) = \sum_{\xi\in G_{N,1}} c_0\int_E (y-x_\xi)(\mathrm{d}a)\frac{\partial(F\circ y)(x)}{\partial x_\xi}[\delta_a].$$

For $y = N^{-1}\sum_{\xi\in G_{N,1}} x_\xi$ this yields

$$(6.26)\quad \frac{\partial(F\circ y)(x)}{\partial x_\xi}[\delta_a] = \frac{\partial F(y)}{\partial y}\left[\frac{\delta_a}{N}\right]$$

and hence

$$(6.27)\quad (L_{\mathrm{mig}}^{(N)}F)(y) = \sum_{\xi\in G_{N,1}} c_0\int_E (y-x_\xi)(\mathrm{d}a)\frac{\partial F(y)}{\partial y}\left[\frac{\delta_a}{N}\right] = 0.$$

### 6.2.2 From Λ-Cannings to Fleming-Viot

Next, we evaluate the moment measures of the average in the limit as $N \to \infty$ and show convergence of the terms to the Fleming-Viot second order term. We denote by $L_{\mathrm{res}}^{(N)[1]}$ the generator $L_{\mathrm{res}}^{(N)}$ on time scale $Nt$.

**Lemma 6.5. [Generator convergence: resampling]**
*On time scale $Nt$, in the limit as $N \to \infty$,*

$$\begin{aligned}&(L_{\mathrm{res}}^{(N)[1]}F)(y)\\ &= \frac{1}{N}\sum_{\xi\in G_{N,1}}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\frac{1}{2}\frac{\partial^2 F(y)}{\partial y\partial y}[r(-x_\xi+\delta_a), r(-x_\xi+\delta_a)] + O\big(N^{-1}\big)\,.\end{aligned} \tag{6.28}$$

*Proof of Lemma 6.5.* We first rewrite $F(y_t)$ in terms of $x_t$:

$$\begin{aligned}F(y_t) = \langle\varphi, y_t^{\otimes n}\rangle &= \left\langle\varphi, \left(\frac{1}{N}\sum_{\xi\in G_{N,1}} x_\xi(t)\right)^{\otimes n}\right\rangle\\ &= \frac{1}{N^n}\sum_{\xi_1\in G_{N,1}}\cdots\sum_{\xi_n\in G_{N,1}}\langle\varphi, x_{\xi_1}(t)\otimes\ldots\otimes x_{\xi_n}(t)\rangle\\ &= \frac{1}{N^n}\left(\bigotimes_{i=1}^n\sum_{\xi_i\in G_{N,1}}\right)\langle\varphi, x_{\xi_1}(t)\otimes\cdots\otimes x_{\xi_n}(t)\rangle\,.\end{aligned} \tag{6.29}$$

Abbreviate

$$F^{(\xi_1,\ldots,\xi_n)}(x) = \int_{E^n}\left(\bigotimes_{i=1}^n x_{\xi_i}\Big(\mathrm{d}u^{(i)}\Big)\right)\varphi\Big(u^{(1)},\ldots,u^{(n)}\Big) = \left\langle\varphi, \bigotimes_{i=1}^n x_{\xi_i}\right\rangle. \tag{6.30}$$

Note that, in this notation, $\xi_i = \xi_j$ for $i \neq j$ is possible. Recall that $(x_t)_{t\geq0}$ has generator $L^{(N)}$ and is the unique solution of the martingale problem (6.24). If we use (6.29) in (6.24) with $x$ replaced by $y$, then we obtain that $(y_t)_{t\geq0}$ solves the martingale problem with generator

$$(L_{\mathrm{res}}^{(N)}F)(y) = \frac{1}{N^n}\left(\bigotimes_{i=1}^n\sum_{\xi_i\in G_{N,1}}\right)L_{\mathrm{res}}\Big(F^{(\xi_1,\ldots,\xi_n)}\Big)(x) \tag{6.31}$$

for the resampling part. Together with (1.14) this yields the expression

$$\begin{aligned}(L_{\mathrm{res}}^{(N)}F)(y) &= \frac{1}{N^n}\left(\bigotimes_{i=1}^n\sum_{\xi_i\in G_{N,1}}\right)\sum_{\xi\in G_{N,1}}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\\ &\times\Big[F^{(\xi_1,\ldots,\xi_n)}\big(x_0,\ldots,x_{\xi-1},(1-r)x_\xi+r\delta_a,x_{\xi+1},\ldots,x_{N-1}\big) - F^{(\xi_1,\ldots,\xi_n)}(x)\Big]\,.\end{aligned} \tag{6.32}$$

We must analyse this expression in the limit as $N \to \infty$. To do so, we collect the leading order terms. The key quantity is the cardinality of the set $\{\xi_1,\ldots,\xi_n\}$, for which we distinguish three cases.

**Case 1:** $|\{\xi_1, \dots, \xi_n\}| = n$, i.e., all $\xi_i, 1 \le i \le n$ are distinct.

The contribution to (6.32) is zero. For $\xi \notin \{\xi_1, \dots, \xi_n\}$ this is obvious by the definition of $F^{(\xi_1,\dots,\xi_n)}(x)$ in (6.30). Otherwise, we have

$$
\begin{aligned}
&\int_E x_\xi(\mathrm{d}a)\, \Big[F^{(\xi_1,\dots,\xi_n)}\big(x_0, \dots, x_{\xi-1}, (1-r)x_\xi + r\delta_a, x_{\xi+1}, \dots, x_{N-1}\big) - F^{(\xi_1,\dots,\xi_n)}(x)\Big] \\
&= \int_E x_\xi(\mathrm{d}a) \\
&\quad \times \Bigg[ \Bigg\langle \varphi, x_{\xi_1} \otimes \dots \otimes \underbrace{\big((1-r)x_\xi + r\delta_a\big)}_{\substack{\text{only change (unique)} \\ \text{position with } \xi_i = \xi}} \otimes \dots \otimes x_{\xi_n} \Bigg\rangle - \langle \varphi, x_{\xi_1} \otimes \dots \otimes x_{\xi_n} \rangle \Bigg] \\
&= 0,
\end{aligned}
\tag{6.33}
$$

where in the last line we use that $\langle x_\xi, 1\rangle = 1$.

**Case 2:** $|\{\xi_1, \dots, \xi_n\}| \le n-2$.

The contribution to (6.32) is of order $N^{-2}$. Indeed, the contribution is bounded from above by

$$
\frac{1}{N^n} \left( \bigotimes_{i=1}^{n} \sum_{\xi_i \in G_{N,1}} \right) 1_{\{|\{\xi_1,\dots,\xi_n\}| \le n-2\}} \lambda_0 C_F = N^{-2} \lambda_0 C_F,
\tag{6.34}
$$

where $C_F$ denotes a generic constant that depends on $F$ only (as in (6.19)), and thereby on $\varphi$ and $n$. Here we use (1.38) and the fact that the sum $\sum_{\xi \in G_{N,1}}$ yields at most $n$ non-zero summands by the definition of $F^{(\xi_1,\dots,\xi_n)}(x)$ in (6.30).

**Case 3:** $|\{\xi_1, \dots, \xi_n\}| = n-1$.

There exist $1 \le m_1 < m_2 \le n$ such that $\xi_{m_1} = \xi_{m_2}$ while all other $\xi_i, 1 \le i \le n$, are different. By the reasoning as in (6.33), we see that the only non-zero contribution of the sum $\sum_{\xi \in G_{N,1}}$ to the generator in (6.32) comes from the case where $\xi = \xi_{m_1} = \xi_{m_2}$. We therefore obtain

$$
\begin{aligned}
&(L_{\mathrm{res}}^{(N)} F)(y) \\
&= \frac{1}{N^n} \left( \bigotimes_{i=1}^{n} \sum_{\xi_i \in G_{N,1}} \right) 1_{\{|\{\xi_1,\dots,\xi_n\}| = n-1\}} \sum_{1 \le m_1 < m_2 \le n} 1_{\{\xi_{m_1} = \xi_{m_2} = \xi\}} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a) \\
&\quad \times \Big[F^{(\xi_1,\dots,\xi_n)}\big(x_0, \dots, x_{\xi-1}, (1-r)x_\xi + r\delta_a, x_{\xi+1}, \dots, x_{N-1}\big) - F^{(\xi_1,\dots,\xi_n)}(x)\Big] + O\big(N^{-2}\big).
\end{aligned}
\tag{6.35}
$$

Reasoning similarly to (6.34), we see that extending

$$
\left( \bigotimes_{i=1}^{n} \sum_{\xi_i \in G_{N,1}} \right) 1_{\{|\{\xi_1,\dots,\xi_n\}| = n-1\}} \sum_{1 \le m_1 < m_2 \le n} 1_{\{\xi_{m_1} = \xi_{m_2}\}}
\tag{6.36}
$$

in (6.35) to

$$
\sum_{1 \le m_1 < m_2 \le n} \sum_{\xi_{m_1} \in G_{N,1}} 1_{\{\xi_{m_1} = \xi_{m_2}\}} \left( \bigotimes_{i \in \{1,\dots,n\} \setminus \{m_1, m_2\}} \sum_{\xi_i \in G_{N,1}} \right)
\tag{6.37}
$$

only produces an additional error of order $N^{-2}$. Using this observation in (6.35), we get

$$
\begin{aligned}
&(L_{\mathrm{res}}^{(N)}F)(y)\\
&=\frac{1}{N^2}\sum_{1\le m_1<m_2\le n}\sum_{\xi\in G_{N,1}}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\\
&\times\Bigg[\Big\langle \varphi, y_{\xi_1}\otimes\cdots\otimes \underbrace{((1-r)x_\xi+r\delta_a)}_{\text{only change position }\xi_{m_1}}\otimes\cdots\otimes\underbrace{((1-r)x_\xi+r\delta_a)}_{\text{and position }\xi_{m_2}}\otimes\cdots\otimes y_{\xi_n}\Big\rangle\\
&-\Big\langle \varphi, y_{\xi_1}\otimes\cdots\otimes \underbrace{x_\xi}_{\text{only change position }\xi_{m_1}}\otimes\cdots\otimes\underbrace{x_\xi}_{\text{and position }\xi_{m_2}}\otimes\cdots\otimes y_{\xi_n}\Big\rangle\Bigg]\\
&+O\big(N^{-2}\big)\,.
\end{aligned}
\tag{6.38}
$$

Now use that

$$
\int_E x_\xi(\mathrm{d}a)\Big\langle \varphi, y_{\xi_1}\otimes\cdots\otimes \underbrace{(x_\xi)}_{\substack{\text{only change}\\ \text{position }\xi_{m_1}}}\otimes\cdots\otimes\underbrace{(-rx_\xi+r\delta_a)}_{\substack{\text{and position }\xi_{m_2}\\ \text{for }m_1,m_2\text{ fixed}}}\otimes\cdots\otimes y_{\xi_n}\Big\rangle=0
\tag{6.39}
$$

to obtain from (6.38) that

$$
\begin{aligned}
&(L_{\mathrm{res}}^{(N)}F)(y)\\
&=\frac{1}{N^2}\sum_{1\le m_1<m_2\le n}\sum_{\xi\in G_{N,1}}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\\
&\quad\times\Big\langle \varphi, y_{\xi_1}\otimes\cdots\otimes \underbrace{(r(-x_\xi+\delta_a))}_{\text{only change position }\xi_{m_1}}\otimes\cdots\otimes\underbrace{(r(-x_\xi+\delta_a))}_{\text{and position }\xi_{m_2}}\otimes\cdots\otimes y_{\xi_n}\Big\rangle\\
&\quad+O\big(N^{-2}\big)\\
&=\frac{1}{N^2}\sum_{\xi\in G_{N,1}}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\frac{1}{2}\frac{\partial^2F(y)}{\partial y\partial y}[r(-x_\xi+\delta_a),r(-x_\xi+\delta_a)]+O\big(N^{-2}\big)\,.
\end{aligned}
\tag{6.40}
$$

Comparing Cases (1)–(3), we see that only the latter contributes to the leading term. Changing to time scale $Nt$ in (6.40), i.e., multiplying $L_{\mathrm{res}}^{(N)}$ by $N$, we complete the proof. □

### 6.2.3 A comment on coupling and duality

The techniques of coupling and duality are of major importance. One application can be found in [DGV95, Section 4], namely, to prove Equation (4.17) therein. The key point is to obtain control on the difference between $\mathcal{L}[Z_t]$ and $\mathcal{L}[Z_t']$ for two Markov processes with *identical dynamics* but *different initial states.* Such estimates can be derived via coupling of the two dynamics, or alternatively, via dual processes that are based on finite particle systems with non-increasing particle numbers, allowing for an entrance law starting from a countably infinite number of particles. Both these properties hold in our model. This fact is used to argue that the configuration locally converges on time scale $Nt$ to an equilibrium by the following restart argument.

At times $Nt$ and $Nt - t_N$, with $\lim_{N\to\infty} t_N = \infty$ and $\lim_{N\to\infty} t_N/N = 0$, the empirical mean remains constant. Hence, we can argue that, in the limit as $N \to \infty$, a system started at time $Nt - t_N$ converges over time $t_N$ to the equilibrium dictated by the current mean. Two facts are needed to make this rigorous: (1) the map $\theta \mapsto \nu_\theta^{c,d,\Lambda}$ must be continuous; (2) the ergodic theorem must hold uniformly in the initial state. Both coupling and duality do the job. This is why we can work with both in [DGV95].

### 6.2.4 McKean-Vlasov process of the 1-block averages on time scale $Nt$

Recall the definition of the Fleming-Viot diffusion operator $Q$ in (1.18) and the equilibrium $\nu$ of the McKean-Vlasov process in the line preceding (4.1). In what follows we keep to denote by $L_{\text{res}}^{(N)[1]}$ the generator $L_{\text{res}}^{(N)}$ on time scale $Nt$. Observe that the compensators of $M^{x,F}$, see (6.24) are functionals of the empirical measure of the configuration. The set of configurations on which $X^{(N)}$ concentrates in the limit as $N \to \infty$ turns out to be

$$\text{(6.41)}\quad \mathbb{B}_\theta^* = \mathbb{B}_\theta \cap \left\{ \underline{x} \in (\mathcal{P}(E))^{\mathbb{N}} \colon \frac{1}{N} \sum_{\xi=1}^{N} \delta_{(x_\xi)} \underset{N\to\infty}{\Longrightarrow} \nu_\theta^{c_0,0,\Lambda_0} \right\},$$

where $\theta$ is called the intensity of the configuration and

$$\text{(6.42)}\quad \mathbb{B}^* = \bigcup_{\theta \in \mathcal{P}(E)} \mathbb{B}_\theta^*.$$

**Lemma 6.6. [Local equilibrium]**

(a) *The block resampling term satisfies, with $y$ the intensity of the configuration $\underline{x}$ for $\underline{x} \in B^*$,*

$$\text{(6.43)}\quad \begin{aligned} \lim_{N\to\infty} (L_{\text{res}}^{(N)[1]} F)(y) &= \frac{\lambda_0}{2} \int_{\mathcal{P}(E)} \nu_y^{c_0,0,\Lambda_0}(\mathrm{d}\widetilde{x}) \int_E \int_E Q_{\widetilde{x}}(\mathrm{d}u, \mathrm{d}v)\, \frac{\partial^2 F(y)}{\partial y \partial y}[\delta_u, \delta_v] \\ &= \frac{c_0 \lambda_0}{2c_0 + \lambda_0} \int_E \int_E Q_y(\mathrm{d}u, \mathrm{d}v)\, \frac{\partial^2 F(y)}{\partial y \partial y}[\delta_u, \delta_v]. \end{aligned}$$

(b) *If the system starts i.i.d. with some finite intensity measure, then every weak limit point of $\mathcal{L}[(X^{(N)}(Nt+u))_{u\in\mathbb{R}}]$ as $N \to \infty$ has paths that satisfy*

$$\text{(6.44)}\quad \mathbb{P}(X^{(\infty)}(t,u) \in \mathbb{B}^*) = 1 \qquad \forall\, t \in [0,\infty),\ u \in \mathbb{R}.$$

*Proof.* (a) The proof uses the line of argument in [DGV95, Section 4(d)] (recall the comment in Section 6.2.3), together with (4.20) and the definition of $Q$. In what follows two observations are important:

(i) We use the results on the existence and uniqueness of a stationary distribution to (6.20) on the time scale $t$ with $N \to \infty$, including the convergence to the stationary distribution uniformly in the initial state, combined with the Feller property of the limiting dynamics (see Section 4). Note, in particular, that with (4.20) we get the second assertion in (6.43) from the first assertion.

(ii) We use the property that the laws of the processes $(Y^{(N)}(Nt))_{t\geq 0}$, $N \in \mathbb{N}$, are tight in path space.

The combination of (i) and (ii) will allow us to derive the claim.

To verify (ii), use (6.40) to establish boundedness of the characteristics, which gives the tightness through a standard criterion (Dawson [D93, Theorem 3.6.6]). To verify (i), we want to show that the weak limit points satisfy the $(\delta_\theta, L^{0,\mathrm{d}_1,\delta_0})$-martingale problem. For that we have to show that

$$\begin{aligned}
&\mathcal{L}\left[\left(F\big(Y^{(N)}(tN)\big) - F\big(Y^{(N),[1]}(0)\big) - \int_0^t (L^{(N),[1]}F)\big(Y^{(N)}(sN)\big)\mathrm{d}s\right)_{t\geq 0}\right] \\
&\qquad \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(F\big(Z^{0,\mathrm{d}_1,0)}(t)\big) - F(\theta) - \int_0^t (L^{0,\mathrm{d}_1,0}F)\big(Z^{0,\mathrm{d}_1,0}(s)\big)\mathrm{d}s\right)_{t\geq 0}\right].
\end{aligned} \tag{6.45}$$

In order to do so, we first need some information on $L^{(N),[1]}$. Since we are on time scale $Nt$ with $N \to \infty$, we get

$$\begin{aligned}
&\lim_{N\to\infty} (L_{\mathrm{res}}^{(N)[1]}F)(y) \\
&= \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_{\mathcal{P}(E)} \nu_y^{c_0,0,\Lambda_0}(\mathrm{d}x) \int_E x(\mathrm{d}a) \frac{1}{2}\frac{\partial^2 F(y)}{\partial y \partial y}[r(-x+\delta_a), r(-x+\delta_a)] \\
&= \frac{\lambda_0}{2} \int_{\mathcal{P}(E)} \nu_y^{c_0,d_0,\Lambda_0}(\mathrm{d}x) \int_E x(\mathrm{d}a) \frac{\partial^2 F(y)}{\partial y \partial y}[-x+\delta_a, -x+\delta_a] \qquad \forall\, \underline{x} \in \mathbb{B}_y^*,\ y \in \mathcal{P}(E).
\end{aligned} \tag{6.46}$$

Use the definition of the Fleming-Viot diffusion operator $Q$ from (1.18) to obtain the claim in (6.43).

(b) To show that the relevant configurations (under the limiting laws) are in $\mathbb{B}^*$, we use a restart argument in combination with the ergodic theorem for the McKean-Vlasov process. Namely, to study the process at time $Nt+u$ we consider the time $Nt+u-t_N$ with $\lim_{N\to\infty} t_N = \infty$ and $\lim_{N\to\infty} t_N/N = 0$. We know that the density process $Y^{(N)}$ at times $Nt+u-t_N$ and $Nt+u$ is the same in the limit $N\to\infty$, say equal to $\theta$, and so over the time stretch $t_N$ the process converges to the equilibrium $(\nu_\theta^{c_0,0,\Lambda_0})^{\otimes\mathbb{N}}$. By the law of large numbers this gives the claim. Therefore all possible limiting dynamics allow for an averaging principle with the local equilibrium. □

**Conclusion of the proof of Proposition 6.3**

Recall from (6.27) that migration has no effect. Lemma 6.6 shows the effect of the block resampling term on time scale $Nt$ for $N\to\infty$. Adding both effects together, we have that all weak limit points of $\mathcal{L}[(Y^{(N)}(Nt))_{t\geq 0}]$, $N\in\mathbb{N}$, satisfy

(6.47) the $(\delta_\theta, L_\theta^{0,\mathrm{d}_1,\delta_0})$-martingale problem with $d_1 = \dfrac{c_0\lambda_0}{2c_0+\lambda_0}$.

# 7 Hierarchical $C^\Lambda$-process

The next step in our construction is to consider finite spatial systems with a hierarchical structure of $K$ levels and to study the $k$-block averages with $k = 0, 1, \ldots, K$ on their natural time scales $N^k t$ and $N^k t + u$. This section therefore deals with the geographic space

$$G = G_{N,K} = \{0, 1, \dots, N-1\}^K, \qquad N, K \in \mathbb{N}. \tag{7.1}$$

Define the Cannings process on $G_{N,K}$ by restricting the Cannings process $X^{(\Omega_N)}$ to $B_K(0)$ and putting

$$c_k, \lambda_k = 0, \quad \text{for all } k \geq K. \tag{7.2}$$

The corresponding process will be denoted by $X^{(N,K)}$ and its generator by $L^{(N,K)}$, etc. It is straightforward to also include a block resampling at rate $N^{-2K}$ with resampling measure $\Lambda_K$ (compare Corollary 6.2).

In this section our principal goal is to understand how we move up $0 \leq k \leq K$ levels when starting from level 0. However, in order to also understand a system with $k$ levels starting from level, say, $L$ and moving up to level $L + k$, we will add a Fleming-Viot term to the generator of $X^{(N)}$, i.e., we consider the case $d_0 > 0$. We do not need to add Fleming-Viot terms acting on higher blocks. In addition, we include a Fleming-Viot term with volatility $d_0 > 0$. As we saw in Lemma 6.6, a resampling term can result, on a higher time scale and in the limit as $N \to \infty$, in a Fleming-Viot term. For instance, if we choose $d_0 = 0$ in the beginning, then we obtain $d_1 = \frac{c_0 \lambda_0}{2c_0 + \lambda_0} > 0$ on time scale $Nt$ for the 1-block average (recall (6.47)).

We look at the block averages on space scales $N^k$ and time scales $N^k t$ with $k = 1, \dots, K$. In Section 7.1 we will focus on the case $K = 2$, where most of the difficulties for general $K$ are already present. For $K > 2$ lower order perturbations arise, which we will discuss only briefly in Section 7.2 because they can be treated similarly as in [DGV95]. In Section 8 we will take the limit $K \to \infty$ and show how this approximates the model with $G = \Omega_N$ on all the time scales we are interested in for our main theorem.

## 7.1 Two-level systems

The geographic space is $G_{N,2} = \{0, 1, \dots, N-1\}^2$, we pick $d_0, c_0, c_1, \lambda_0, \lambda_1 > 0$ and put $c_k, \lambda_k$ to zero for $k \geq 2$. We will prove the following: (1) On time scales $t$ and $Nt$ we obtain the same limiting objects as described in Section 6, but with an additional Fleming-Viot term ($d_0 > 0$) and with block resampling via $\Lambda_1$; (2) On time scale $Nt$ for 1-block averages (each belonging to an address $\eta \in \{0, 1, \cdots, N-1\}$) we introduce

$$Y_\eta^{(N)}(t) = N^{-1} \sum_{\xi \in G_{N,1}} X_{\xi,\eta}^{(N)}(t). \tag{7.3}$$

Next on time scale $N^2 t$ we consider the total average

$$Z^{(N)}(t) = N^{-2} \sum_{\zeta \in G_{N,2}} X_\zeta^{(N)}(t), \tag{7.4}$$

we get a similar structure.

Namely, we can replace the system $(Z^{(N)}, X^{(N)})$ for $N \to \infty$ by a system of the type in Section 6, with the role of components on time scale $t$ taken over by 1-block averages on time scale $Nt$ and the role of the total (1-block) average on time scale $Nt$ taken over by the 2-block average on time scale $N^2 t$. Once again, we only focus on the new features arising in our model. The general scheme of the proof for the two-level system can be found in [DGV95, Section 5(a), pp. 2328–2337]. The calculations in Sections 7.1.1–7.1.3 correspond to Steps 4–5 in [DGV95, Section 5(a)], with the focus now shifted from the characteristics of diffusions to the *full* generator because we are dealing with jump processes.

**Proposition 7.1. [Two-level rescaling]**
*Under the assumptions made above,*

$$(7.5)\quad \mathcal{L}[(X_\zeta^{(N)}(t))_{t\geq 0}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(Z_\theta^{c_0,d_0,\Lambda_0}(t))_{t\geq 0}] \qquad \forall\, \zeta \in G_{N,2},$$

*and*

$$(7.6)\quad \mathcal{L}[(Y_\xi^{(N)}(Nt))_{t\geq 0}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(Z_\theta^{c_1,d_1,\Lambda_1}(t))_{t\geq 0}] \quad \textit{with} \quad d_1 = \frac{c_0(\lambda_0+2d_0)}{2c_0+\lambda_0+2d_0}, \quad \xi \in G_{N,1}$$

*and (with $\Lambda = \delta_0$)*

$$(7.7)\quad \mathcal{L}[(Z^{(N)}(N^2t))_{t\geq 0}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(Z_\theta^{0,d_2,\delta_0}(t))_{t\geq 0}] \quad \textit{with } d_2 = \frac{c_1(\lambda_1+2d_1)}{2c_1+\lambda_1+2d_1}.$$

The proof of (7.5–7.7) is carried out in Sections 7.1.1–7.1.3.

### 7.1.1 The single components on time scale $t$

In this section, our main goal is to argue that the components of $X^{(N)}$ change on time scale $t$ as before, and that the same holds on time scales $Nt+u$ and $N^2t+u$ with $u \in \mathbb{R}$, provided we use the appropriate value for the 1-block average as the centre of drift.

We first look at the components on time scale $t$. Due to the Markov property and the continuity in $\theta$ of the law of the McKean-Vlasov process, the behaviour of the components on time scales $Nt+u$ and $N^2t+u$ with $u \in \mathbb{R}$ is immediate once we have the tightness of $Y^{(N)}$ and $Z^{(N)}$ on these scales. Again, our convergence results are obtained by: (1) establishing *tightness* in path space; (2) verifying convergence of the finite-dimensional distributions via establishing asymptotic independence and the *generator calculation* for the martingale problem. Since the latter is key also for the tightness arguments, we give the analysis of the generator terms first. In fact, the rest of the argument is the same as in Section 6.1.

**Migration part.** Consider the migration operator in (1.36) and (1.25) applied to functions $F \in \mathcal{F}$, the algebra of functions in (1.33). The migration operator can be rewritten as

$$(7.8)\quad \begin{aligned}
(L_{\mathrm{mig}}^{(N,2)}F)(x) &= \sum_{\xi,\zeta\in G_{N,2}} a_{\xi,\zeta}^{(N)} \int_E (x_\zeta - x_\xi)\,(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\xi}[\delta_a] \\
&= \sum_{\xi,\zeta\in G_{N,2}} \sum_{d(\xi,\zeta)\leq k\leq 2} c_{k-1}N^{1-2k} \int_E (x_\zeta - x_\xi)\,(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\xi}[\delta_a] \\
&= \sum_{\xi\in G_{N,2}} \sum_{k\leq 2} c_{k-1}N^{1-2k} \sum_{\zeta\in B_k(\xi)} \int_E (x_\zeta - x_\xi)\,(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\xi}[\delta_a] \\
&= \sum_{\xi\in G_{N,2}} \sum_{k\leq 2} c_{k-1}N^{1-k} \int_E (y_{\xi,k} - x_\xi)\,(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\xi}[\delta_a],
\end{aligned}$$

where we use (1.30) in the last line. Thus, for $F$ as in (1.33), we obtain

$$(7.9)\quad (L^{(N,2)}_{\mathrm{mig}}F)(x) = \sum_{\xi\in G_{N,2}} c_0 \int_E (y_{\xi,1}-x_\xi)\,(\mathrm{d}a)\,\frac{\partial F(x)}{\partial x_\xi}[\delta_a] + E^{(N)},$$

where

$$(7.10)\quad |E^{(N)}| \le N^{-1}c_1C_F = O\big(N^{-1}\big)$$

with $C_F$ a generic constant depending on the choice of $F$ only. Here we use that, by the definition of $F$ in (1.33), the sum over $\xi \in G_{N,2}$ is a sum over finitely many coordinates only, with the number depending on $F$ only.

**Resampling part.** Consider the resampling operator $(L^{(N,2)}_{\mathrm{res}}F)(x)$ in (1.37). We have

$$(7.11)\quad (L^{(N,2)}_{\mathrm{res}}F)(x) = \sum_{\xi\in G_{N,2}} \int_{[0,1]} \Lambda^*_0(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\,\big[F\left(\Phi_{r,a,B_0(\xi)}(x)\right) - F(x)\big] + E^{(N)}$$

with

$$(7.12)\quad |E^{(N)}| \le N^{-2}\int_{[0,1]} \Lambda^*_1(\mathrm{d}r)C_Fr^2N = C_FN^{-1}\lambda_1 = O\big(N^{-1}\big)\,.$$

Here we use (1.38) in the first inequality, together with the fact that $F(\Phi_{r,a,B_1(\xi)}(x)) - F(x)$ is non-zero for at most $C_FN$ different values of $\xi \in G_{N,2}$.

**Additional Fleming-Viot part.** Recall that in this section we consider the case $d_0 > 0$, i.e., we add the Fleming-Viot generator

$$(7.13)\quad (L^{(N,2)}_{\mathrm{FV}}F)(x) = d_0 \sum_{\xi\in G_{N,2}} \int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(x)}{\partial x_\xi \partial x_\xi}[\delta_u,\delta_v].$$

Contrary to the migration and the resampling operator, the Fleming-Viot operator does not act on higher block levels.

**The resulting generator.** Combining the migration parts (7.9) and (7.10), the resampling parts (7.11) and (7.12), and the Fleming-Viot part (7.13), we obtain

$$(7.14)\quad \begin{aligned}(L^{(N,2)}F)(x) = &\sum_{\xi\in G_{N,2}} c_0 \int_E (y_{\xi,1}-x_\xi)\,(\mathrm{d}a)\,\frac{\partial F(x)}{\partial x_\xi}[\delta_a]\\ &+ \sum_{\xi\in G_{N,2}} \int_{[0,1]} \Lambda^*_0(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\,\big[F\left(\Phi_{r,a,B_0(\xi)}(x)\right) - F(x)\big]\\ &+ d_0 \sum_{\xi\in G_{N,2}} \int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(x)}{\partial x_\xi \partial x_\xi}[\delta_u,\delta_v] + O\big(N^{-1}\big)\,.\end{aligned}$$

The order term is independent of $x$.

**Convergence to McKean-Vlasov process.** We can use (7.14) to argue that

$$\|L^{(L,2)}F - L^{c_0,d_0,\Lambda_0}_{y_{\xi,1}}F\|_\infty \leq C_F N^{-1}. \tag{7.15}$$

Next, following again the line of argument in Section 6.1, we obtain that $X^{(N)}$ converges as a process to the McKean-Vlasov limit, which is an i.i.d. collection of single components indexed by $\mathbb{N}_0$ with generator

$$\begin{aligned}(L^{c_0,d_0,\Lambda_0}_\theta G)(x_\xi) =& c_0 \int_E (\theta - x_\xi)(\mathrm{d}a)\, \frac{\partial G(x_\xi)}{\partial x_\xi}[\delta_a] \\ &+ \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\, \big[G\big((1-r)x_\xi + r\delta_a\big) - G(x_\xi)\,\big] \\ &+ d_0 \int_E \int_E Q_{x_\xi}(\mathrm{d}u, \mathrm{d}v)\, \frac{\partial^2 G(x)}{\partial x_\xi \partial x_\xi}[\delta_u, \delta_v],\end{aligned} \tag{7.16}$$

where $\theta \in \mathcal{P}(E)$ is the initial mean measure. This completes the proof of (7.5).

### 7.1.2 The 1-block averages on time scale $Nt$

Again, we need to prove: (1) tightness in path space of $(Y^{(N)}_\xi(Nt))_{t\geq 0}$; (2) convergence of finite-dimensional distributions via asymptotic independence and generator convergence. As we saw in Section 6, the latter is also the key to tightness. Therefore we proceed by first calculating the generator of 1-block averages on time scale $Nt$ and then using this generator to show convergence of the process. At that point we need that the average over the full space remains $\theta$, in the sense of a constant path on time scale $Nt$. The latter property will be proved in Section 7.1.3.

**Basic generator formula.** We proceed as in Section 6.2.1. Since $G = G_{N,2}$, the 1-block averages are now indexed too. We use the following notation for the indexing of 1-block averages. Recall the notation $y_{\zeta,1} = N^{-1}\sum_{\xi\in B_1(\zeta)} x_\xi$ from (1.30), which is the 1-block around $\zeta$. This 1-block coincides with the 1-block around $\xi$ if and only if $d(\zeta,\xi) \leq 1$. To endow every 1-block with a unique label, we proceed as follows. Let $\phi$ be the shift-operator

$$\phi\colon\ G_{N,K} \to G_{N,K-1}, (\phi\xi)_i = \xi_{i+1}, \qquad 0 \leq i \leq K-1,\ K \in \mathbb{N}. \tag{7.17}$$

We consider the evolution in time of the 1-block averages indexed block-wise, i.e.,

$$y^{[1]}_\eta = N^{-1} \sum_{\xi\in G_{N,2}, \phi\xi=\eta} x_\xi, \tag{7.18}$$

where once again we suppress the dependence of $y^{[1]}_\eta$ on $N$. Note. in particular, that

$$y_{\xi,1} = y^{[1]}_\eta \text{ for all } \xi \text{ such that } \phi\xi = \eta. \tag{7.19}$$

We will often drop the superscript [1], especially when the context is clear.

This time, we consider functions $F \in \mathcal{F}$ (see (1.33)) applied to $y^{[1]} = y^{[1]}(x)$, where $y^{[1]} = (y^{[1]}_\eta)_{\eta\in G_{N,1}}$. By explicit calculation of the different terms below, we will obtain the following expression:

$$
\begin{aligned}
(L^{(N,2)[1]}F)(y) &= \left(L^{(N,2)[1]}_{\mathrm{mig}} + L^{(N,2)[1]}_{\mathrm{res},0} + L^{(N,2)[1]}_{\mathrm{res},1} + L^{(N,2)[1]}_{\mathrm{FV}}\right)(F)(y) \\
&= \sum_{\eta\in G_{N,1}} c_1 \int_E \left(y^{[1]}_{\phi\eta} - y_\eta\right)(\mathrm{d}a)\,\frac{\partial F(y)}{\partial y_\eta}[\delta_a] \\
&\quad + \sum_{m=1}^{q} \frac{1}{N} \sum_{\xi:\ \phi\xi=\eta^{(m)}} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\,\frac{1}{2}\,\frac{\partial^2 F(y)}{\partial y_{\eta^{(m)}}\partial y_{\eta^{(m)}}}[r(-x_\xi+\delta_a), r(-x_\xi+\delta_a)] \\
&\quad + \sum_{\eta\in G_{N,1}} \int_{[0,1]} \Lambda_1^*(\mathrm{d}r) \int_E y_\eta(\mathrm{d}a)\,[F(\Phi_{r,a,\eta}(y)) - F(y)] \\
&\quad + d_0 \sum_{\eta\in G_{N,1}} \frac{1}{N} \sum_{\xi:\ \phi\xi=\eta} \int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(y)}{\partial y_\eta\partial y_\eta}[\delta_u,\delta_v] + O\big(N^{-1}\big)\,.
\end{aligned}
\tag{7.20}
$$

**Convergence to McKean-Vlasov process.** We first argue how to conclude the argument, and then further below we carry out the necessary generator calculations.

We have to argue first that the $N$ different 1-blocks satisfy the propagation of chaos property (recall (6.6)), where we had this for components. The proof again uses duality, namely, dual particles from different 1-blocks need a time of order $N^2$ to meet and hence do not meet on time scale $Nt$. We do not repeat the details here.

Once we have the propagation of chaos property, it suffices to consider *single blocks*, which we do next. We have to verify tightness in path space and convergence of the finite-dimensional distributions. As we saw before, this reduces to convergence of the generator on $\mathcal{F}$ by the same tightness argument used in Section 6.1.1, but now based on (7.20). Consider the resampling and Fleming-Viot parts of the generator separately.

Reason as in the proof of Lemma 6.6 to see that

$$
\begin{aligned}
&\lim_{N\to\infty} (L^{(N,2)[1]}_{\mathrm{res},0}F)(y) \\
&= \lim_{N\to\infty} \sum_{m=1}^{q} \frac{1}{N} \sum_{\xi:\ \phi\xi=\eta^{(m)}} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\,\frac{1}{2}\,\frac{\partial^2 F(y)}{\partial y_{\eta^{(m)}}\partial y_{\eta^{(m)}}}[r(-x_\xi+\delta_a), r(-x_\xi+\delta_a)] \\
&= \frac{\lambda_0}{2} \sum_{\eta\in\mathbb{N}_0} \int_{\mathcal{P}(E)} \nu^{c_0,d_0,\Lambda_0}_{y_\eta}(\mathrm{d}x) \int_E\int_E Q_x(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(y)}{\partial y_\eta\partial y_\eta}[\delta_u,\delta_v] \\
&= \frac{c_0\lambda_0}{2c_0+\lambda_0+2d_0} \sum_{\eta\in\mathbb{N}_0} \int_E\int_E Q_{y_\eta}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(y)}{\partial y_\eta\partial y_\eta}[\delta_u,\delta_v],
\end{aligned}
\tag{7.21}
$$

where by (4.20) the second assertion follows from the first. Similarly, we have

$$
\lim_{N\to\infty} (L^{(N,2)[1]}_{\mathrm{FV}}F)(y) = d_0 \sum_{\eta\in\mathbb{N}_0} \int_{\mathcal{P}(E)} \nu^{c_0,d_0,\Lambda_0}_{y_\eta}(\mathrm{d}x) \int_E\int_E Q_x(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(y)}{\partial y_\eta\partial y_\eta}[\delta_u,\delta_v]. \tag{7.22}
$$

Using (4.20) once more, we get

$$
\text{r.h.s. (7.22)} = \frac{2c_0d_0}{2c_0+\lambda_0+2d_0} \sum_{\eta\in\mathbb{N}_0} \int_E\int_E Q_{y_\eta}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(y)}{\partial y_\eta\partial y_\eta}[\delta_u,\delta_v]. \tag{7.23}
$$

Combine (7.21) with (7.23) and argue as in Section 6.1.4, to see that each single component of the 1-block averages $y = y^{[1]} = (y^{[1]}_\eta)_{\eta \in G_{N,1}}$ converges and the limiting coordinate process has generator

$$
\begin{aligned}
(L_\theta^{c_1,d_1,\Lambda_1} G)(y_\eta) &= c_1 \int_E (\theta - y_\eta)(\mathrm{d}a)\, \frac{\partial G(y_\eta)}{\partial y_\eta}[\delta_a] \\
&\quad + d_1 \int_E \int_E Q_{y_\eta}(\mathrm{d}u, \mathrm{d}v)\, \frac{\partial^2 G(y)}{\partial y_\eta \partial y_\eta}[\delta_u, \delta_v] \\
&\quad + \int_{[0,1]} \Lambda_1^*(\mathrm{d}r) \int_E y_\eta(\mathrm{d}a)\, \big[G\big((1-r)y_\eta + r\delta_a\big) - G(y_\eta)\big],
\end{aligned}
\tag{7.24}
$$

where $\theta \in \mathcal{P}(E)$ is the initial mean measure of a component and $d_1 = \frac{c_0(\lambda_0 + 2d_0)}{2c_0 + \lambda_0 + 2d_0}$. At this point we use that the average over the complete population remains the path that stands still at $\theta$ on time scale $Nt$.

**Generator calculation: proof of (7.20).** We next verify the expression given in (7.20). We calculate separately the action of the various terms in the generator on the function $F$. In what follows a change to time scale $N^k t$ is denoted by an additional superscript $[k]$.

**Migration part.** Recall $(L_{\mathrm{mig}}^{(N,2)} F)(x)$ from (7.8). Proceeding along the lines of (6.25–6.27), we get

$$
\begin{aligned}
(L_{\mathrm{mig}}^{(N,2)} F)(y) &= \sum_{\xi \in G_{N,2}} \sum_{k \le 2} c_{k-1} N^{1-k} \int_E (y_{\xi,k} - x_\xi)(\mathrm{d}a) \frac{\partial (F \circ y)(x)}{\partial x_\xi}[\delta_a] \\
&= \sum_{\xi \in G_{N,2}} \sum_{k \le 2} c_{k-1} N^{1-k} \int_E \Big(y^{[1]}_{\phi^k \xi} - x_\xi\Big)(\mathrm{d}a) \frac{\partial F(y)}{\partial y_{\phi\xi}}\left[\frac{\delta_a}{N}\right] \\
&= N \sum_{\eta \in G_{N,1}} \sum_{k \le 2} c_{k-1} N^{1-k} \int_E \Big(y^{[1]}_{\phi^{k-1}\eta} - y_\eta\Big)(\mathrm{d}a) \frac{\partial F(y)}{\partial y_\eta}\left[\frac{\delta_a}{N}\right] \\
&= \sum_{\eta \in G_{N,1}} \sum_{k \le 1} c_k N^{1-k} \int_E \Big(y^{[1]}_{\phi^k \eta} - y_\eta\Big)(\mathrm{d}a) \frac{\partial F(y)}{\partial y_\eta}\left[\frac{\delta_a}{N}\right].
\end{aligned}
\tag{7.25}
$$

Next, for functions $F$ that are linear combinations of functions in (1.33), we have

$$
N \frac{\partial F(y)}{\partial y_\eta}\left[\frac{\delta_a}{N}\right] = \frac{\partial F(y)}{\partial y_\eta}[\delta_a].
\tag{7.26}
$$

On the time scale $Nt$, we have

$$
(L_{\mathrm{mig}}^{(N,2)[1]} F)(y) = \sum_{\eta \in G_{N,1}} c_1 \int_E \Big(y^{[1]}_{\phi\eta} - y_\eta\Big)(\mathrm{d}a) \frac{\partial F(y)}{\partial y_\eta}[\delta_a].
\tag{7.27}
$$

**Resampling part.** The calculations proceed along the same lines as in Section 6.2.2. Apart from an additional higher-order term, the main extension is that we consider $F(y_t) = F(y_t^{[1]}) = \langle \varphi, \bigotimes_{l=1}^q y_{\eta^{(l)}}^{\otimes n_l} \rangle$ with $y = y^{[1]} = (y^{[1]}_\eta)_{\eta \in G_{N,1}}$, $\eta^{(l)} \in G_{N,1}$, $q \in \{1, \dots, N\}$ and $n_l \in \mathbb{N}, 1 \le l \le q$, instead of (6.29) (which corresponds to the case $q = 1$). We will restrict ourselves to functions $F$ of the form

$$\begin{aligned} &F(y) = \int_{E^{n_1+\dots+n_q}} \left( \bigotimes_{l=1}^{q} y_{\eta^{(l)}}^{\otimes n_l}(\mathrm{d}u^{(l)}) \right) \varphi(u^{(1)}, \dots, u^{(q)}), \quad y = (y_\eta)_{\eta \in G_{N,1}} \in \mathcal{P}(E)^N, \\ &q \in \{1, \dots, N\},\ n_l \in \mathbb{N}, \eta^{(l)} \in G_{N,1},\ l \in \{1, \dots, q\}, \\ &\eta^{(l)} \neq \eta^{(l')}, \text{ for all } l \neq l',\ u^{(l)} \in E^{n_l},\ \varphi \in C_{\mathrm{b}}(E^{n_1+\dots+n_q}, \mathbb{R}). \end{aligned} \tag{7.28}$$

The only difference with (1.33) is the restriction of the ordering of the entries. This facilitates the notation in the computation below, but is no loss of generality because the set of functions in (7.28) generates the same algebra $\mathcal{F}$. We will now show that

$$\begin{aligned} &(L_{\mathrm{res}}^{(N,2)[1]} F)(y) \\ &= \sum_{m=1}^{q} \frac{1}{N} \sum_{\xi:\ \phi\xi = \eta^{(m)}} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a) \frac{1}{2} \frac{\partial^2 F(y)}{\partial y_{\eta^{(m)}} \partial y_{\eta^{(m)}}} [r(-x_\xi + \delta_a), r(-x_\xi + \delta_a)] \\ &\quad + \sum_{\eta \in G_{N,1}} \int_{[0,1]} \Lambda_1^*(\mathrm{d}r) \int_E y_\eta(\mathrm{d}a) \left[ F\left(\Phi_{r,a,\eta}(y)\right) - F(y) \right] + O\big(N^{-1}\big). \end{aligned} \tag{7.29}$$

Recall the notation in (6.30) and set

$$L = \sum_{l=1}^{q} n_l. \tag{7.30}$$

Proceeding as in (6.29-6.31), we obtain

$$(L_{\mathrm{res}}^{(N,2)} F)(y) = \frac{1}{N^L} \left( \bigotimes_{l=1}^{q} \bigotimes_{i=1}^{n_l} \sum_{\xi_i^l:\ \phi\xi_i^l = \eta^{(l)}} \right) L_{\mathrm{res}} \Big( F^{\left(\xi_1^1, \dots, \xi_{n_q}^q\right)} \Big) (x). \tag{7.31}$$

As in Section 6.2, we distinguish between the different cases for the structure of the set $\{\xi_1^1, \cdots, \xi_{n_q}^1\}$ and we obtain, using the definition of the resampling operator in (1.37),

$$\begin{aligned} &(L_{\mathrm{res}}^{(N,2)} F)(y) \\ &= \frac{1}{N^L} \left( \bigotimes_{l=1}^{q} \bigotimes_{i=1}^{n_l} \sum_{\xi_i^l:\ \phi\xi_i^l = \eta^{(l)}} \right) \sum_{\xi \in G_{N,2}} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a) \\ &\quad \times \Big[ F^{\left(\xi_1^1, \dots, \xi_{n_q}^q\right)} \big(\Phi_{r,a,B_0(\xi)}(x)\big) - F^{\left(\xi_1^1, \dots, \xi_{n_q}^q\right)}(x) \Big] \\ &\quad + \frac{1}{N^L} \left( \bigotimes_{l=1}^{q} \bigotimes_{i=1}^{n_l} \sum_{\xi_i^l:\ \phi\xi_i^l = \eta^{(l)}} \right) \sum_{\xi \in G_{N,2}} N^{-2} \int_{[0,1]} \Lambda_1^*(\mathrm{d}r) \int_E y_{\xi,1}(\mathrm{d}a) \\ &\quad \times \Big[ F^{\left(\xi_1^1, \dots, \xi_{n_q}^q\right)} \big(\Phi_{r,a,B_1(\xi)}(x)\big) - F^{\left(\xi_1^1, \dots, \xi_{n_q}^q\right)}(x) \Big] \\ &= I_0 + I_1. \end{aligned} \tag{7.32}$$

For the first term $I_0$ in (7.32) we proceed along the lines of (6.33–6.34) to conclude that the only non-negligible contribution to the sum in $I_0$ comes from terms with $|\{\xi_i^l, 1 \leq l \leq q, 1 \leq$

$i \leq n_l\}| = L-1$. It remains to investigate the terms with $|\{\xi_i^l, 1 \leq l \leq q, 1 \leq i \leq n_l\}| = L-1$. Since $\phi\xi_i^l = \eta^{(l)}$, this implies that there exist $1 \leq m \leq q$ and $1 \leq m_1 < m_2 \leq n_m$ such that $\xi_{m_1}^m = \xi_{m_2}^m$ and all other $\xi_i^l$ different. By the same reasoning as in (6.33), we see that the only non-zero contribution of the sum $\sum_{\xi \in G_{N,2}}$ comes from $\xi = \xi_{m_1}^m = \xi_{m_2}^m$. We therefore obtain

(7.33)
$$\begin{aligned} I_0 =& \frac{1}{N^L}\left(\bigotimes_{l=1}^{q}\bigotimes_{i=1}^{n_l}\sum_{\xi_i^l:\ \phi\xi_i^l=\eta^{(l)}}\right) 1_{\{|\{\xi_i^l, 1\leq l\leq q, 1\leq i\leq n_l\}|=L-1\}} \sum_{m=1}^{q}\sum_{1\leq m_1<m_2\leq n_m} 1_{\{\xi_{m_1}^m=\xi_{m_2}^m=\xi\}} \\ & \times \int_{[0,1]} \Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\left[F^{(\xi_1^1,\dots,\xi_{n_q}^q)}\big(\Phi_{r,a,B_0(\xi)}(x)\big) - F^{(\xi_1^1,\dots,\xi_{n_q}^q)}(x)\right] + O\big(N^{-2}\big)\,. \end{aligned}$$

Now follow the reasoning from (6.35) to (6.40), to get

$$\begin{aligned} I_0 =& \frac{1}{N^2}\sum_{m=1}^{q}\sum_{\xi:\ \phi\xi=\eta^{(m)}}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\frac{1}{2}\frac{\partial^2 F(y)}{\partial y_{\eta^{(m)}}\partial y_{\eta^{(m)}}}[r(-x_\xi+\delta_a), r(-x_\xi+\delta_a)] \\ & + O\big(N^{-2}\big)\,. \end{aligned} \tag{7.34}$$

For the second term $I_1$ in (7.32), we obtain, by the definition of $\Phi_{r,a,B_1(\xi)}(x)$ in (1.38) and using (7.19),

(7.35)
$$\begin{aligned} I_1 =& \frac{1}{N^L}\left(\bigotimes_{l=1}^{q}\bigotimes_{i=1}^{n_l}\sum_{\xi_i^l:\ \phi\xi_i^l=\eta^{(l)}}\right)\sum_{\xi\in G_{N,2}} N^{-2}\int_{[0,1]}\Lambda_1^*(\mathrm{d}r)\int_E y_{\xi,1}(\mathrm{d}a) \\ & \times \left[F^{(\xi_1^1,\dots,\xi_{n_q}^q)}\big(\Phi_{r,a,B_1(\xi)}(x)\big) - F^{(\xi_1^1,\dots,\xi_{n_q}^q)}(x)\right] \\ =& \frac{1}{N^L}\left(\bigotimes_{l=1}^{q}\bigotimes_{i=1}^{n_l}\sum_{\xi_i^l:\ \phi\xi_i^l=\eta^{(l)}}\right)\sum_{\eta\in G_{N,1}} N^{-1}\int_{[0,1]}\Lambda_1^*(\mathrm{d}r)\int_E y_\eta(\mathrm{d}a) \\ & \times \left[F^{(\xi_1^1,\dots,\xi_{n_q}^q)}\Big(\Phi_{r,a,\eta}^{[1]}(x)\Big) - F^{(\xi_1^1,\dots,\xi_{n_q}^q)}(x)\right] \end{aligned}$$

with

$$\left[\Phi_{r,a,\eta}^{[1]}(x)\right]_\xi = \begin{cases} (1-r)y_\eta + r\delta_a, & \phi\xi = \eta, \\ x_\xi, & \text{otherwise.} \end{cases} \tag{7.36}$$

Now observe that the sum $\sum_{\eta\in G_{N,1}}$ in (7.35) yields non-zero contributions only for $\eta \in$

$\{\eta^{(1)},\dots,\eta^{(q)}\}$, and so we can rewrite $I_1$ as

$$\begin{aligned}
I_1 =&\frac{1}{N^L}\left(\bigotimes_{l=1}^{q}\bigotimes_{i=1}^{n_l}\sum_{\xi_i^l\colon\ \phi\xi_i^l=\eta^{(l)}}\right)\sum_{l=1}^{q}N^{-1}\int_{[0,1]}\Lambda_1^*(\mathrm{d}r)\int_E y_{\eta^{(l)}}(\mathrm{d}a)\\
&\times\Bigg[\Big\langle\varphi, x_{\xi_1^1}\otimes\cdots\otimes x_{\xi_{n_{l-1}}^{l-1}}\otimes\underbrace{\Big((1-r)y_{\eta^{(l)}}+r\delta_a\Big)}_{\text{change from position }\xi_1^l}\\
&\qquad\otimes\cdots\otimes\underbrace{\Big((1-r)y_{\eta^{(l)}}+r\delta_a\Big)}_{\text{to position }\xi_{n_l}^l}\otimes x_{\xi_1^{l+1}}\otimes\cdots\otimes x_{\xi_{n_q}^q}\Big\rangle-\Big\langle\varphi,x_{\xi_1^1}\otimes\cdots\otimes x_{\xi_{n_q}^q}\Big\rangle\Bigg]\\
=&\sum_{l=1}^{q}N^{-1}\int_{[0,1]}\Lambda_1^*(\mathrm{d}r)\int_E y_{\eta^{(l)}}(\mathrm{d}a)\\
&\times\Bigg[\Big\langle\varphi,y_{\eta^{(1)}}^{\otimes n_1}\otimes\cdots\otimes y_{\eta^{(l-1)}}^{\otimes n_{l-1}}\otimes\Big((1-r)y_{\eta^{(l)}}+r\delta_a\Big)^{\otimes n_l}\otimes y_{\eta^{(l+1)}}^{\otimes n_{l+1}}\\
&\qquad\otimes\cdots\otimes y_{\eta^{(q)}}^{\otimes n_q}\Big\rangle-\Big\langle\varphi,\bigotimes_{l=1}^{q}y_{\eta^{(l)}}^{\otimes n_l}\Big\rangle\Bigg]\\
=&\sum_{\eta\in G_{N,1}}N^{-1}\int_{[0,1]}\Lambda_1^*(\mathrm{d}r)\int_E y_\eta(\mathrm{d}a)\left[F\left(\Phi_{r,a,\eta}(y)\right)-F(y)\right].
\end{aligned}\tag{7.37}$$

Combining (7.32), (7.34) and (7.37), we obtain (7.29) on time scale $Nt$.

**Additional Fleming-Viot part.** We proceed as for the migration operator and write

$$\begin{aligned}
(L_{\mathrm{FV}}^{(N,2)}F)(y)&=\Big(L_{\mathrm{FV}}^{(N,2)}(F\circ y)\Big)(x)\\
&=d_0\sum_{\xi\in G_{N,2}}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2\left(F\circ y\right)(x)}{\partial x_\xi\partial x_\xi}[\delta_u,\delta_v],
\end{aligned}\tag{7.38}$$

where the definition of $y=y^{[1]}$ in (7.18) yields

$$\frac{\partial^2\left(F\circ y\right)(x)}{\partial x_\xi\partial x_\xi}[\delta_u,\delta_v]=\frac{\partial^2F(y)}{\partial y_{\phi\xi}\partial y_{\phi\xi}}\left[\frac{\delta_u}{N},\frac{\delta_v}{N}\right].\tag{7.39}$$

Hence, on time scale $Nt$,

$$\begin{aligned}
(L_{\mathrm{FV}}^{(N,2)[1]}F)(y)&=d_0N\sum_{\eta\in G_{N,1}}\sum_{\xi\colon\ \phi\xi=\eta}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(y)}{\partial y_\eta\partial y_\eta}\left[\frac{\delta_u}{N},\frac{\delta_v}{N}\right]\\
&=d_0\sum_{\eta\in G_{N,1}}\frac{1}{N}\sum_{\xi\colon\ \phi\xi=\eta}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(y)}{\partial y_\eta\partial y_\eta}[\delta_u,\delta_v],
\end{aligned}\tag{7.40}$$

where in the last line we use that, for $F$ a linear combination of the functions in (1.33),

$$N^2\frac{\partial^2F(y)}{\partial y_\eta\partial y_\eta}\left[\frac{\delta_u}{N},\frac{\delta_v}{N}\right]=\frac{\partial^2F(y)}{\partial y_\eta\partial y_\eta}[\delta_u,\delta_v].\tag{7.41}$$

**The resulting generator.** Combining the migration (7.27), resampling (7.29) and Fleming-Viot (7.40) parts for the 1-block averages on time scale $Nt$, we obtain (7.20). This completes the proof of (7.6).

### 7.1.3 The total average on time scale $N^2t$

Denote the total average by by

$$z = N^{-1} \sum_{\eta \in G_{N,1}} y_\eta^{[1]} = N^{-2} \sum_{\xi \in G_{N,2}} x_\xi. \tag{7.42}$$

(This is a 2-block average because we are considering the case $K = 2$.) We must prove: (1) $\{\mathcal{L}[(Z^{(N)}(tN^2))_{t\geq 0},\ N \in \mathbb{N}\}$ is *tight* in path space; (2) the weak limit points of this sequence are solutions of the martingale problem (recall Section 6.1). From the uniqueness of the solution of the martingale problem we get the claim.

To verify (1), we first observe that we have a martingale (compare with (7.43)) for every $N$. Hence, we want to show convergence to a continuous martingale, i.e., we have to apply tightness criteria for martingales. The necessary information is collected below by going through the various mechanisms step by step, which is the key to (2) as well, the convergence of the finite-dimensional distributions.

**Migration part.** For the total average the migration operator can be obtained from (7.27) by writing $z = z(y)$ and using the analogue to (6.26),

$$(L_{\mathrm{mig}}^{(N,2)[1]}F)(z) = (L_{\mathrm{mig}}^{(N,2)[1]}(F \circ z))(y) = \sum_{\eta \in G_{N,1}} c_1 \int_E \left(y_{\phi\eta}^{[1]} - y_\eta\right)(\mathrm{d}a) \frac{\partial F(z)}{\partial z}\left[\frac{\delta_a}{N}\right]. \tag{7.43}$$

Using that $z = y_{\phi\eta}^{[1]} = N^{-1}\sum_{\eta \in G_{N,1}} y_\eta^{[1]}$ for all $\eta \in G_{N,1}$, we get

$$(L_{\mathrm{mig}}^{(N,2)[1]}F)(z) = (L_{\mathrm{mig}}^{(N,2)[2]}F)(z) = 0. \tag{7.44}$$

**Resampling part.** Consider $F(z) = \langle \varphi, z^{\otimes n}\rangle$. Follow the derivation of (6.31) to obtain

$$(L_{\mathrm{res}}^{(N,2)}F)(z) = \frac{1}{N^n}\left(\bigotimes_{i=1}^{n} \sum_{\eta_i \in G_{N,1}}\right) L_{\mathrm{res}}\left(F^{(\eta_1,\dots,\eta_n)}\right)(y) = I_0' + I_1' \tag{7.45}$$

with $F^{(\eta_1,\dots,\eta_n)}(y) = \langle \varphi, \bigotimes_{i=1}^n y_{\eta_i}\rangle$ as in (6.30), where we recall from (7.32) that

$$\begin{aligned}
&(L_{\mathrm{res}}^{(N,2)}F^{(\eta_1,\dots,\eta_n)})(y)\\
&= \frac{1}{N^n}\left(\bigotimes_{l=1}^{n} \sum_{\xi_l\colon\ \phi\xi_l=\eta_l}\right) \sum_{\xi \in G_{N,2}} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\\
&\quad \times \left[F^{(\xi_1,\dots,\xi_n)}\left(\Phi_{r,a,B_0(\xi)}(x)\right) - F^{(\xi_1,\dots,\xi_n)}(x)\right]\\
&\quad + \frac{1}{N^n}\left(\bigotimes_{l=1}^{n} \sum_{\xi_l\colon\ \phi\xi_l=\eta_l}\right) \sum_{\xi \in G_{N,2}} N^{-2} \int_{[0,1]} \Lambda_1^*(\mathrm{d}r) \int_E y_{\xi,1}(\mathrm{d}a)\\
&\quad \times \left[F^{(\xi_1,\dots,\xi_n)}\left(\Phi_{r,a,B_1(\xi)}(x)\right) - F^{\left(\xi_1^1,\dots,\xi_n\right)}(x)\right]\\
&= I_0'' + I_1''.
\end{aligned} \tag{7.46}$$

Let us begin with the second term $I_1''$ in (7.46), which corresponds to $I_1$ in (7.32) and was rewritten in (7.35–7.37) as

$$(7.47)\quad I_1'' = \sum_{\eta\in G_{N,1}} N^{-1}\int_{[0,1]}\Lambda_1^*(\mathrm{d}r)\int_E y_\eta(\mathrm{d}a)\left[F^{(\eta_1,\ldots,\eta_n)}\left(\Phi_{r,a,\eta}(y)\right) - F^{(\eta_1,\ldots,\eta_n)}(y)\right].$$

Combine (7.45) and (7.47), change to timescale $Nt$ and compare the result to (6.32). We obtain that $I_1'$ on time scale $Nt$ behaves analogously to (6.32) on time scale $t$. By moving one time scale upwards, we obtain as in (6.43) (respectively, (7.21) with $d_1 = \frac{c_0(\lambda_0+2d_0)}{2c_0+\lambda_0+2d_0} > 0$) that

$$(7.48)\quad \lim_{N\to\infty}(I_1')^{[2]} = \frac{c_1\lambda_1}{2c_1+\lambda_1+2d_1}\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(z)}{\partial z\partial z}[\delta_u,\delta_v].$$

The term $I_0'$ can be handled in the same spirit as $I_0$ in (7.32). To obtain non-zero contributions in $I_0''$, we need to have $|\{\xi_l, \phi\xi_l = \eta_l, 1\le l\le n\}| < n$ (recall (6.33)). This is possible only if $|\eta_1,\ldots,\eta_n| < n$. Reasoning similarly as in (6.34), we obtain negligible terms if $|\{\xi_l,\phi\xi_l = \eta_l, 1\le l\le n\}| < n-1$. Indeed, two equivalent sites already result in a factor of $O(N^{-2})$ (on time scale $t$): first a common block has to be chosen ($|\eta_1,\ldots,\eta_n| = n-1$), which contributes a factor $N^{-2}\sum_{\eta\in G_{N,1}}$, and subsequently a common site has to be chosen, which contributes a factor $N^{-2}\sum_{\xi:\phi\xi=\eta}$. Any additional choice results in terms that vanish for $N\to\infty$ on time scale $N^2t$. Consequently, we can reason as in (6.35–6.40) to obtain

$$(7.49)\quad \begin{aligned}(I_0')^{[0]} &= \frac{1}{N^2}\sum_{\eta\in G_{N,1}}\frac{1}{N^2}\sum_{\xi:\ \phi\xi=\eta}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\\ &\quad\times\int_E x_\xi(\mathrm{d}a)\frac{1}{2}\frac{\partial^2F(z)}{\partial z\partial z}[r(-x_\xi+\delta_a), r(-x_\xi+\delta_a)] + O\big(N^{-3}\big).\end{aligned}$$

**Additional Fleming-Viot part.** We proceed as for the migration operator. Recall (7.40), to get

$$(7.50)\quad (L_{\mathrm{FV}}^{(N,2)[1]}F)(z) = d_0\sum_{\eta\in G_{N,1}}\frac{1}{N}\sum_{\xi:\ \phi\xi=\eta}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2(F\circ z)(y)}{\partial y_\eta\partial y_\eta}[\delta_u,\delta_v].$$

Now use the analogue to (7.39), to obtain

$$(7.51)\quad (L_{\mathrm{FV}}^{(N,2)[1]}F)(z) = d_0\sum_{\eta\in G_{N,1}}\frac{1}{N}\sum_{\xi:\ \phi\xi=\eta}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}\left[\frac{\delta_u}{N},\frac{\delta_v}{N}\right].$$

After changing to time scale $N^2t$, we have

$$(7.52)\quad (L_{\mathrm{FV}}^{(N,2)[2]}F)(z) = d_0\frac{1}{N}\sum_{\eta\in G_{N,1}}\frac{1}{N}\sum_{\xi:\ \phi\xi=\eta}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v].$$

**Tightness.** We have to bound the generator and apply the tightness criterion, as explained in Section 6.1.1. We omit the details.

**Convergence to McKean-Vlasov process.** We have to identify the limiting generators. One approach would be to try and make the following heuristics rigorous.

*Begin heuristics.* On time scale $N^2t$, we obtain, by reasoning as in (7.21),

$$
\begin{aligned}
\lim_{N\to\infty}(I_0')^{[2]} &= \frac{\lambda_0}{2}\lim_{N\to\infty}\frac{1}{N}\sum_{\eta\in G_{N,1}}\int_{\mathcal{P}(E)}\nu_{y_\eta}^{c_0,d_0,\Lambda_0}(\mathrm{d}x)\int_E\int_E Q_x(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v]\\
&= \frac{c_0\lambda_0}{2c_0+\lambda_0+2d_0}\lim_{N\to\infty}\frac{1}{N}\sum_{\eta\in G_{N,1}}\int_E\int_E Q_{y_\eta}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v]\\
&= \frac{c_0\lambda_0}{2c_0+\lambda_0+2d_0}\int_{\mathcal{P}(E)}\nu_z^{c_1,d_1,\Lambda_1}(\mathrm{d}y)\int_E\int_E Q_y(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v]\\
&= \frac{2c_1}{2c_1+\lambda_1+2d_1}\,\frac{c_0\lambda_0}{2c_0+\lambda_0+2d_0}\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v].
\end{aligned}
\tag{7.53}
$$

Combine (7.48) with (7.53), to get

$$
\begin{aligned}
&\lim_{N\to\infty}(L_{\mathrm{res}}^{(N,2)[2]}F)(z)\\
&= \frac{2c_1}{2c_1+\lambda_1+2d_1}\left(\frac{\lambda_1}{2}+\frac{c_0\lambda_0}{2c_0+\lambda_0+2d_0}\right)\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v].
\end{aligned}
\tag{7.54}
$$

For the Fleming-Viot part, we obtain, by reasoning once more as in (7.21),

$$
\begin{aligned}
&\lim_{N\to\infty}(L_{\mathrm{FV}}^{(N,2)[2]}F)(z)\\
&= d_0\lim_{N\to\infty}\frac{1}{N}\sum_{\eta\in G_{N,1}}\int_{\mathcal{P}(E)}\nu_{y_\eta}^{c_0,d_0,\Lambda_0}(\mathrm{d}x)\int_E\int_E Q_x(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v]\\
&= \frac{2c_0d_0}{2c_0+\lambda_0+2d_0}\lim_{N\to\infty}\frac{1}{N}\sum_{\eta\in G_{N,1}}\int_E\int_E Q_{y_\eta}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v]\\
&= \frac{2c_0d_0}{2c_0+\lambda_0+2d_0}\int_{\mathcal{P}(E)}\nu_z^{c_1,d_1,\Lambda_1}(\mathrm{d}y)\int_E\int_E Q_y(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v]\\
&= \frac{2c_1}{2c_1+\lambda_1+2d_1}\,\frac{2c_0d_0}{2c_0+\lambda_0+2d_0}\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v].
\end{aligned}
\tag{7.55}
$$

Collecting the limiting terms as $N\to\infty$ on time scale $N^2t$ for migration (7.44), resampling (7.54) and Fleming-Viot (7.55), we obtain

$$
\begin{aligned}
&\lim_{N\to\infty}(L^{(N,2)[2]}F)(z)\\
&= \frac{2c_1}{2c_1+\lambda_1+2d_1}\left(\frac{\lambda_1}{2}+\frac{c_0\lambda_0+2c_0d_0}{2c_0+\lambda_0+2d_0}\right)\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z\partial z}[\delta_u,\delta_v].
\end{aligned}
\tag{7.56}
$$

In order to obtain the convergence in (7.53–7.55), we would need to restrict the set of configurations, argue that the law of the process lives on that set of configurations, and show that therefore the compensators of the martingale problems converge to the compensator of the limit process. However, it is technically easier to follow a different route, as we do below. *End heuristics.*

We want to view the expression for $(L^{(N,2),[2]}F)(z)$ as an average over $N$ different 1-block averages. If we replace the 1-block averages by a system of $N$ exchangeable Fleming-Viot diffusions with resampling constant $d_1$ (for which we have a formula in terms of $c_0, d_0$ and $\lambda_0$), which on time scale $Nt$ lead to the generator

$$(7.57)\quad L^{(N,2),[1]}_{\mathrm{mig}}(G)(y) + \frac{c_0(\lambda_0+2d_0)}{2c_0+\lambda_0+2d_0}\int\limits_E\int\limits_E Q_y(\mathrm{d}u,\mathrm{d}v)\frac{\partial^2 G(y)}{\partial y\partial y}[\delta_u,\delta_v] + (L^{(N,2),[1]}_{\mathrm{res},1}G)(y),$$

then we can apply the analysis of Section 6 to this new collection of processes, denoted by

$$(7.58)\quad \left\{\widetilde{Y}_i^{(N)}(tN),\, i=1,\dots,N\right\},$$

to conclude that the block average $\widetilde{Z}^{(N)}(tN) = N^{-1}\sum_{i=1}^{N}\widetilde{Y}_i^{N}(Nt)$ satisfies

$$(7.59)\quad \mathcal{L}[(\widetilde{Z}^{(N)}(tN^2))_{t\geq 0}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(\widetilde{Z}(t))_{t\geq 0}],$$

where $\widetilde{Z}$ is a Fleming-Viot diffusion with resampling constant

$$(7.60)\quad \frac{c_1}{2c_1+\lambda_1+2d_1}(\lambda_1+2d_1), \text{ where } d_1 = \frac{c_0(\lambda_0+2d_0)}{2c_0+\lambda_0+2d_0}.$$

Hence we obtain a limit process with a generator acting on $F$ as

$$(7.61)\quad \frac{c_1(\lambda_1+2d_1)}{2c_1+\lambda_1+2d_1}\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(z)}{\partial z\partial z}[\delta_u,\delta_v].$$

Hence, the weak limit points of the laws $\{\mathcal{L}[(\widetilde{Z}^{(N)}(tN^2))_{t\geq 0}],\, N\in\mathbb{N}\}$ satisfy the martingale problem with generator $(L^{0,d_2,\delta_0}_\theta G)(z)$ with $d_2 = \frac{c_1(\lambda_1+2d_1)}{2c_1+\lambda_1+2d_1}$.

Since we know that the martingale problem for the generator $L^{0,d,\Lambda}$ and for the test functions given by $C^2_{\mathrm{b}}(\mathcal{P}(E),\mathbb{R})$ is well-posed, we have the claimed convergence in (7.7) on path space if $Z$ (a weak limit point for the original problem) and $\widetilde{Z}$ agree. Thus, we have to argue that it is legitimate to

$$(7.62)\quad \text{replace } \{((Y_i^{(N)}(Nt))_{i=1,\dots,N})_{t\geq 0}\} \text{ by } \{(\widetilde{Y}_i^{(N)}(Nt)_{i=1,\dots,N})_{t\geq 0}\}.$$

For that purpose, observe that we know from Section 6 that, for a suitable subsequence along which $\mathcal{L}[(Z^{(N)}(sN^2))_{s\geq 0}]$ converges to $Z(s)$,

$$(7.63)\quad \mathcal{L}[((Y_i^{(N)}(N^2 s+Nt))_{i=1,\dots,N})_{t\geq 0}] \underset{N\to\infty}{\longrightarrow} \mathcal{L}[(Y_i^{(\infty)}(s,t))_{t\geq 0}],$$

where the right-hand side is the McKean-Vlasov process with Fleming-Viot part at rate $d_1$, Cannings part $\Lambda_1$, and immigration-emigration at rate $c_1$ from the random source $Z(s)$. We need to argue that the latter implies that $Z$ and $\widetilde{Z}$ agree.

For $F\in C^2_{\mathrm{b}}(\mathcal{P}(E),\mathbb{R})$, define $G_N\in C^2_{\mathrm{b}}((\mathcal{P}(E))^N,\mathbb{R})$ and $H_N\in C^2_{\mathrm{b}}((\mathcal{P}(E))^{N^2},\mathbb{R})$ by

$$(7.64)\quad F(z) = G_N(y) = H_N(x), \quad x\in(\mathcal{P}(E))^{N^2}, \quad y\in(\mathcal{P}(E))^N, \quad z\in\mathcal{P}(E),$$

with

$$(7.65)\quad z = \frac{1}{N}\sum_{i\in\{1,\dots,N\}} y_i, \quad y_i = \frac{1}{N}\sum_{j\in\{1,\dots,N\}} x_{j,i}.$$

In order to verify that $Z$ and $\widetilde{Z}$ agree, it suffices to show that the compensator processes for $\widetilde{Z}$ and $Z$ agree for a measure-determining family of functions $F \in C_{\mathrm{b}}^2(\mathcal{P}(E),\mathbb{R})$, namely,

$$\begin{aligned}
&\mathcal{L}\Big[\Big(\int_0^{tN^2} \mathrm{d}s \,\Big[\int_{E\times E} d_1 \sum_{i=1}^N Q_{y_i(s)}(\mathrm{d}u,\mathrm{d}v)\frac{\partial^2 G_N(y(s))}{\partial y_i \partial y_i}[\delta_u,\delta_v] + L_{\mathrm{res},1}^{(N,2)[1]} G_N(y(s))\Big]\Big)_{t\geq 0}\Big] \\
&- \mathcal{L}\Big[\Big(\int\limits_0^{tN^2} \mathrm{d}s \,\Big[L_{\mathrm{res},1}^{(N,2)[1]} G_N(y_j(s)) + \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N L_{\mathrm{res},0}^{(N,2)} H_N(x_{j,i}(s))\Big]\Big)_{t\geq 0}\Big] \\
&\underset{N\to\infty}{\Longrightarrow} \text{Zero measure.}
\end{aligned} \tag{7.66}$$

To that end, first note that the two terms with $L_{\mathrm{res},1}^{(N,2),[1]}$ cancel each other out. Regarding the remaining terms, after we transform $s$ to $sN^2$, we must show that for each $s \in [0,t]$ the term in the second line converges weakly to the term in the first line (the joint law of the density and the empirical measure converges). When worked out in detail, this requires a somewhat subtle argument. However, nothing is specific to our model: a detailed argument along these lines can be found in [DGV95], pp. 2322-2339.

### 7.2 Finite-level systems

The next step is to consider general $K \geq 3$ (recall the beginning of Section 7). We can copy the arguments used for $K=2$, and argue recursively, namely, we can view the $(j-1), j, (j+1)$-block averages as a *two-level system* on time scales $tN^{j-1}, N(tN^{j-1}), N^2(tN^{j-1})$.The limit as $N\to\infty$ is a two-level system with migration rates $c_{j-1}, c_j, c_{j+1}$ instead of $c_0, c_1, c_2$, resampling measures $\Lambda_{j-1}, \Lambda_j, \Lambda_{j+1}$ instead of $\Lambda_0, \Lambda_1, \Lambda_2$, and volatility $\mathrm{d}_{j-1}$ instead of $\mathrm{d}_0$. If we would have $c_0 = c_1 = \cdots = c_{j-2} = 0$ and $\lambda_0 = \cdots = \lambda_{j-2} = 0$, then this would be literally the case. Hence, the key point is to show that the lower-order perturbation terms play no role in the renormalised dynamics after they have played their role in determining the coefficients $d_{j-1}, d_j, d_{j+1}$.

The argument has again a tightness part, which is the same as before and which we do not discuss, and a finite-dimensional distributions part. Since the solution of the martingale problem is uniquely determined by the marginal distributions (see [EK86, Section 4.4.2]), this part is best based on duality, which determines the transition kernel of the process as follows.

We have to verify that the dual of the $(j+1)$-level system on the time scales $N^{j-1}t, N^j t$ behaves like the dual process of a two-level system. This means that the dual process can be replaced by the system where the locations up to level $j-2$ are uniformly distributed and all partition elements originally within that distance have coalesced. This can be obtained by showing that the dual system with the lower-order terms is instantaneously uniformly distributed in small balls, and that within that distance coalescence is instantaneous, since we are working with times at least $tN^{j-1}$. Therefore, the dynamics as $N\to\infty$ results *effectively* in a coalescent corresponding to a two-level system.

## 8 Proof of the hierarchical mean-field scaling limit

We are finally ready to prove Theorem 1.4. In this section we approximate our infinite spatial system by finite spatial systems of the type studied in Section 7. As before, we denote the finite system with geographic space $G_{N,K}$ by $X^{(N,K)}$ and the one with $G=\Omega_N$ by $X^{(\Omega_N)}$.

**Proposition 8.1.** **[$K$-level approximation]**
*For $t \in (0,\infty)$ and $s_N \in (0,\infty)$ with $\lim_{N\to\infty} s_N = \infty$ and $\lim_{N\to\infty} s_N/N = 0$, consider $Y^{(\Omega_N)}_{\xi,k}$ and $Y^{(N,K)}_{\xi,k}$ on time scale $tN^j + s_N N^k$ for $0 \leq k \leq j < K$. Then*

$$(8.1)\quad d_{\text{Prokh}} \left(\mathcal{L}\left[\left(Y^{(\Omega_N)}_{\xi,k}(tN^j + s_N N^k)\right)\right], \mathcal{L}\left[\left(Y^{(N,K)}_{\xi,k}(tN^j + s_N N^k)\right]\right)\right) \underset{N\to\infty}{\Longrightarrow} 0,$$

*where $d_{\text{Prokh}}$ is the Prokhorov metric.*

Once we have proved this proposition, we obtain Theorem 1.4 by observing that (8.1) allows us to replace our system on $\Omega_N$ by the one on $G^{N,K}$ when we are interested only in block averages of order $\leq K$ on time scales of order $< N^K$. In that case we can use the result of Section 7 to obtain the claim of the theorem for $(j,k)$ with $k \leq j < K$. Thus, it remains only to prove Proposition 8.1. We give the proof for $K = 1$, and later indicate how to extend to $K \in \mathbb{N}$.

The main idea is the following. We want to compare the laws of the solution of two martingale problems at a fixed time and show that their difference goes to zero in the weak topology. To this end, it suffices to show that the difference of the action of the two generators in the martingale problems on the functions in the algebra $\mathcal{F}$ tends to zero. Indeed, we then easily get the claim with the help of the formula of partial integration for two semigroups $(V_t)_{t\geq 0}$ and $(U_t)_{t\geq 0}$ (see e.g. Ethier and Kurtz [EK86, Section 1, (5.19)])):

$$(8.2)\quad V_t = U_t + \int_0^t U_{t-s}(L_V - L_U)V_s \mathrm{d}s.$$

In Sections 8.1–8.2, we calculate and asymptotically evaluate the difference of the generator acting on $\mathcal{F}$ on the two spatial and temporal scales.

## 8.1 The single components on time scale $t$

For an $F$ that depends only on $x_\xi$, $\xi \in B_1(0)$, we have, as we will see below,

$$(8.3)\quad (L^{(\Omega_N)}F)(x) = (L^{(N,2)}F)(x) + O(N^{-1})$$

where the error term is uniform in $x$ and only depends on the choice of $F$. By the formula of partial integration for semigroups, it follows that

$$(8.4)\quad \left|\mathbb{E}\left[F(X^{(\Omega_N)}(t))\right] - \mathbb{E}\left[F(X^{(N,K)}(t))\right]\right| \leq tO(N^{-1}).$$

Since our test functions are measure-determining, the claim follows for any finite time horizon. To prove (8.3), we discuss the different parts of the generators separately.

Consider the migration operator in (1.36) applied to functions $F \in \mathcal{F}$, the algebra of functions of the form in (1.33). The migration operator can be rewritten, similarly as in (7.8),

$$(8.5)\quad (L^{(\Omega_N)}_{\text{mig}}F)(x) = \sum_{\xi\in\Omega_N}\sum_{k\in\mathbb{N}} c_{k-1}N^{1-k}\int_E (y_{\xi,k} - x_\xi)(\mathrm{d}a)\,\frac{\partial F(x)}{\partial x_\xi}[\delta_a].$$

We obtain

$$(8.6)\quad (L^{(\Omega_N)}_{\text{mig}}F)(x) = \sum_{\xi\in\Omega_N} c_0 \int_E (y_{\xi,1} - x_\xi)(\mathrm{d}a)\,\frac{\partial F(x)}{\partial x_\xi}[\delta_a] + E^{(N)},$$

where

$$(8.7)\ \ |E^{(N)}| \le N^{-1} C_F \sum_{k\in\mathbb{N}\setminus\{1\}} c_{k-1} N^{2-k},$$

with $C_F$ a generic constant depending on the choice of $F$ only. Here we use that, by the definition of $F$ in (1.33), the sum over $\xi \in \Omega_N$ is a sum over finitely many coordinates only, with the number depending on $F$ only. By (1.26) we get

$$(8.8)\ \ |E^{(N)}| \le O\big(N^{-1}\big)\,.$$

For the resampling operator we obtain, applying first (1.38) and then (1.31), similarly as in (1.37),

$$(8.9)\ \ (L^{(\Omega_N)}_{\mathrm{res}} F)(x) = \sum_{\xi\in\Omega_N} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a) \left[F\left(\Phi_{r,a,B_0(\xi)}(x)\right) - F(x)\right] + E^{(N)}$$

with

$$(8.10)\ \ |E^{(N)}| \le \sum_{k\in\mathbb{N}} N^{-2k} \int_{[0,1]} \Lambda_k^*(\mathrm{d}r) C_F N^k r^2 = C_F \sum_{k\in\mathbb{N}} N^{-k}\lambda_k = O\big(N^{-1}\big)\,.$$

Finally, the Fleming-Viot operator reads as in (7.13):

$$(8.11)\ \ (L^{(\Omega_N)}_{\mathrm{FV}} F)(x) = d_0 \sum_{\xi\in\Omega_N} \int_E \int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\, \frac{\partial^2 F(x)}{\partial x_\xi \partial x_\xi}[\delta_u,\delta_v].$$

Combining the migration parts in (8.6) and (8.8), the resampling parts in (8.9) and (8.10), and the Fleming-Viot part in (8.11), we obtain

$$\begin{aligned}
(L^{(\Omega_N)}F)(x) &= \sum_{\xi\in\Omega_N} c_0 \int_E (y_{\xi,1} - x_\xi)(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\xi}[\delta_a] + O(N^{-1}) \\
&\quad + \sum_{\xi\in\Omega_N} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a) \left[F\left(\Phi_{r,a,B_0(\xi)}(x)\right) - F(x)\right] + O\big(N^{-1}\big) \qquad (8.12)\\
&\quad + d_0 \sum_{\xi\in\Omega_N} \int_E \int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\, \frac{\partial^2 F(x)}{\partial x_\xi \partial x_\xi}[\delta_u,\delta_v].
\end{aligned}$$

Combining (8.12) with (8.5—8.11) and (7.14) (also recall the discussion on embeddings from Section 5.2), we get the claim.

### 8.2 The 1-block averages on time scale $Nt$

As before we prove, for $F$ depending on $\xi \in B_1(0)$ only,

$$(8.13)\ \ (L^{(\Omega_N)[1]})(y) = (L^{(N,2)[1]}F)(y) + O(N^{-1})$$

after which the claim follows in the limit as $N \to \infty$ by the same argument as in Section!8.1. We prove (8.13) by considering separately the different parts of the generator.

For the 1-block averages $y = y^{[1]}$ the migration operator can be calculated as in (7.25). Using (7.26), we get

$$(L_{\mathrm{mig}}^{(\Omega_N)}F)(y) = \frac{1}{N}\sum_{\eta\in\Omega_N}\sum_{k\in\mathbb{N}} c_k N^{1-k}\int_E \left(y^{[1]}_{\phi^k\eta} - y_\eta\right)(\mathrm{d}a)\frac{\partial F(y)}{\partial y_\eta}[\delta_a]. \tag{8.14}$$

We obtain on the time scale $Nt$

$$(L_{\mathrm{mig}}^{(\Omega_N)[1]}F)(y) = \sum_{\eta\in\Omega_N} c_1 \int_E \left(y^{[1]}_{\phi\eta} - y_\eta\right)(\mathrm{d}a)\frac{\partial F(y)}{\partial y_\eta}[\delta_a] + E^{(N)}, \tag{8.15}$$

where

$$\left|E^{(N)}\right| \leq C_F \sum_{k\in\mathbb{N}\setminus\{1\}} c_k N^{1-k} = O\big(N^{-1}\big)\,. \tag{8.16}$$

Note that, by (7.27),

$$(L_{\mathrm{mig}}^{(\Omega_N)[1]}F)(y) = (L_{\mathrm{mig}}^{(N,2)[1]}F)(y) + O\big(N^{-1}\big)\,. \tag{8.17}$$

For the resampling operator, the only change to (7.31) is that (7.32) gets replaced by

$$(L_{\mathrm{res}}^{(\Omega_N)}F)(y) = I_0 + I_1 + E^{(N)} \tag{8.18}$$

with $I_0, I_1$ as in (7.32) (with $G_{N,2}$ replaced by $\Omega_N$) and

$$\begin{aligned}\left|E^{(N)}\right| &\leq \frac{1}{N^L}\left(\bigotimes_{l=1}^{q}\bigotimes_{i=1}^{n_l}\sum_{\xi_i^l:\ \phi\xi_i^l=\eta^{(l)}}\right)\sum_{k\in\mathbb{N}\setminus\{1\}} N^{-2k}\int_{[0,1]}\Lambda_k^*(\mathrm{d}r)LN^kC_Fr^2\\ &= C_F\sum_{k\in\mathbb{N}\setminus\{1\}} N^{-k}\lambda_k = O\big(N^{-2}\big)\,.\end{aligned} \tag{8.19}$$

After a change to time scale $Nt$, we therefore have

$$(L_{\mathrm{res}}^{(\Omega_N)[1]}F)(y) = (L_{\mathrm{res}}^{(N,2)[1]}F)(y) + O\big(N^{-1}\big) \tag{8.20}$$

with $(L_{\mathrm{res}}^{(N,2)}F)(y)$ as in (7.31).

The Fleming-Viot operator on time scale $t$ reads as in (7.38), respectively, on time scale $Nt$ as in (7.40),

$$(L_{\mathrm{FV}}^{(\Omega_N)[1]}F)(y) = (L_{\mathrm{FV}}^{(N,2)[1]}F)(y). \tag{8.21}$$

## 8.3 Arbitrary truncation level

For every $K \in \mathbb{N}$, consider the block averages up to level $K-1$ on time scales up to $N^Kt$, estimate the generator difference, bound this by an $O(N^{-1})$-term and get the same conclusion as above. There are more indices involved in the notation, but the argument is the same. The details are left to the reader.

# 9 Multiscale analysis

## 9.1 The interaction chain

In this section, we prove Theorem 1.8. In addition to Theorem 1.4, what is needed is the convergence of the *joint law* of the collection of $k$-level block averages for $k = 0, \dots, j+1$ on the corresponding time scales $N^j t_N + N^k t$, with $\lim_{N\to\infty} t_N = \infty$ and $\lim_{N\to\infty} t_N/N = 0$. We already know that the $\ell$-block averages for $\ell > k$ do not change on time scale $tN^k$ and that this holds in path space as well. Hence, in particular, the $(j+1)$-block average converges to a constant path at times $N^j t_N + N^k t$ for all $0 \leq k \leq j$. We also have the convergence of the marginal distributions for each $k = 0, \dots, j+1$, namely, we know that the process on level $k$ solves a martingale problem on time scale $tN^k$, which we have identified and where only the block average on the next level appears as a parameter. Therefore, arguing downward from level $j+1$ to level $j$, we see that the *Markov property* holds for the limiting law. It therefore only remains to identify the transition probability.

We saw in Section 7 that when going from level $k+1$ to level $k$, we get the corresponding equilibrium law of the level-$k$ limiting dynamics as a McKean-Vlasov process with parameters $(c_k, \theta, d_k, \Lambda_k)$ with $\theta$ equal to the limiting state on level $k+1$. Note here that, instead of $N^{k+1}s + N^k t$, we can write $N^{k+1}s + N^k t_N$ with $\lim_{N\to\infty} t_N = \infty$ and $\lim_{N\to\infty} t_N/N = 0$, since an $o(1)$ perturbation of $s$ has no effect as $N \to \infty$. For more details, consult [DGV95, Section 5(f)].

In the remainder of this section, we prove the implications of the scaling results of $(d_k)_{k\in\mathbb{N}}$ for the hierarchical multiscale analysis of the process $X^{(\Omega_N)}$, involving clustering versus coexistence (Section 9.2), related phase transitions (Section 9.1), as well as a more detailed description of the properties of the different regimes (Section 10), as discussed in Section 1.5.5.

## 9.2 Dichotomy for the interaction chain

In this section, we prove Theorem 1.9. Fix $j \in \mathbb{N}_0$. The first observation is that the interaction chain $(M_k^{(j)})_{k=-(j+1),\dots,0}$ is a $\mathcal{P}(E)$-valued Markov chain such that, for any $\varphi \in C_{\mathrm{b}}(E)$,

(9.1) $\big(\langle M_k^{(j)}, \varphi\rangle\big)_{k=-(j+1),\dots,0}$ is a *square-integrable martingale*

(because it is bounded). For the analysis of the interaction chain for Fleming-Viot diffusions, carried out in [DGV95, Section 6], this fact was central in combination with the formula for the variance of evaluations analogous to Proposition 4.5. We argue as follows.

Since the map $\theta \mapsto \nu_\theta^{c,d,\Lambda}$ is continuous, the convergence as $j \to \infty$ in the *local coexistence* regime is a standard argument (see [DGV95, Section 6a]). In the *clustering* regime, the convergence to the mono-type state follows by showing, with the help of the variance formula, that $\lim_{j\to\infty} \mathbb{E}_{\mathcal{L}(M_0^{(j)})}[\mathrm{Var}_x(\varphi)] = 0$ for all $\varphi \in C_{\mathrm{b}}(E)$, so that all limit points of $\mathcal{L}[M^{(j)}]$ are concentrated on $\delta$-measures on $E$ (recall that $\mathcal{P}(E)$ is compact). This argument is identical to the one in [DGV95, Section 6a]. The mixing measure for the value of the mono-type state can be identified via the martingale property.

It remains to show that in the case where $\mathbb{E}_{\mathcal{L}(M_0^{(j)})}[\mathrm{Var}_x(\varphi)]$ is bounded away from zero, the limit points allow for the coexistence of types. The argument in [DGV95, Section 6a] shows that for $\Lambda = \delta_0$,

(9.2) $\nu_\theta^{c,d,\Lambda}(M) = 0$ if $d > 0$, $\quad M = \{\delta_u\colon\ u \in E\}$.

This is no longer true for $\Lambda \neq \delta_0$. Instead, we have $\nu_\theta^{c,d,\Lambda}(M) \in [0,1)$, as proven in Section 4.3 (see (4.12)), and hence the variance is $> 0$.

## 9.3 Scaling for the interaction chain

In this Section we prove Theorems 1.11 and 1.12

The proof of the scaling result in the regime of diffusive clustering in [DGV95, Section 6(b), Steps 1–3] uses two ingredients:

(I) The following processes is a square-integrable martingale:

$$(9.3)\quad \Big(\langle M_k^{(j)}, f\rangle\Big)_{k=-(j+1),\dots,0}, \qquad \forall\, f \in C_{\mathrm{b}}(E,\mathbb{R}).$$

(II) For $c_k \to c \in (0,\infty)$ as $k \to \infty$, by [DGV95, Eq. (6.12)],

$$(9.4)\quad \mathrm{Var}\left(\langle M_{k_2}^{(j)}, f\rangle \mid M_{k_1}^{(j)} = \theta\right) = \frac{(-k_1)-(-k_2)+1}{c+(-k_1)}\,\mathrm{Var}_\theta(f), \qquad \forall\, f \in C_{\mathrm{b}}(E,\mathbb{R}).$$

In [DGV95, Section 6(b)], (I–II) led to the conclusion that if $\lim_{j\to\infty}(-k_i)/j = \bar\beta_i \in [0,1]$, $i=1,2$, with $\bar\beta_1 > \bar\beta_2$, then

$$(9.5)\quad \lim_{j\to\infty} \mathrm{Var}\left(\langle M_{k_2}^{(j)}, f\rangle \mid M_{k_1}^{(j)} = \theta\right) = \frac{\bar\beta_1 - \bar\beta_2}{\bar\beta_1}\,\mathrm{Var}_\theta(f).$$

Thus, as soon as we have these formulae, we get the claim by repeating the argument in [DGV95, Section 6(b)], which includes the time transformation $\bar\beta = e^{-s}$ in Step 3 to obtain a time-homogeneous expression from (9.5).

We know the necessary first and second moment formulae from Section 4.4. Replace [DGV95, Eq. (6.12)] by (4.27), to see that we must make sure that

$$(9.6)\quad \lim_{j\to\infty} \sum_{i=\lfloor\bar\beta_2 j\rfloor}^{\lfloor\bar\beta_1 j\rfloor} \left(\frac{d_{i+1}}{c_i} \prod_{l=i+1}^{\lfloor\bar\beta_1 j\rfloor} \frac{1}{1+m_l}\right) = 1 - \left(\frac{\bar\beta_2}{\bar\beta_1}\right)^R.$$

Note that (9.6) remains valid also for $\bar\beta_2 = 0$.

Moreover, by following the reasoning in [DGV95, Section 6(b), Step 4], we obtain by using 4.27 instead of [DGV95, (6.34)] that

$$(9.7)\quad \left\{\begin{array}{c}\text{fast growing clusters}\\ \text{slowly growing clusters}\end{array}\right\} \quad \text{if} \quad \sum_{i=n}^{m}\left(\frac{d_{i+1}}{c_i}\prod_{l=i+1}^{m}\frac{1}{1+m_l}\right)\left\{\begin{array}{c}\to 0\\ \to 1\end{array}\right\}$$

when $m,n\to\infty$ such that $n/m \to \alpha$, for all $\alpha \in (0,1)$.

*Proof of Theorem 1.11.* The proof follows by inserting the asymptotics of $c_k$, $d_k$ and $m_k$ obtained in Theorem 1.6 and Corollary 1.10 into (9.6) or (9.7).

(i) In Cases (a) and (b), the asymptotics in (1.51–1.52) and (1.73) imply

$$(9.8)\quad \sum_{i=\lfloor\alpha m\rfloor}^{m}\left(\frac{d_{i+1}}{c_i}\prod_{l=i+1}^{m}\frac{1}{1+m_l}\right) = O\left(\mathrm{e}^{-Cn}\right), \quad C>0.$$

In Case (c), using the fact that $d_{i+1}/c_i \sim m_i \to 0$ and $\sum_{l\in\mathbb{N}_0} m_l = \infty$, we obtain

$$(9.9)\quad \sum_{i=\lfloor \alpha m\rfloor}^{m} \left( \frac{d_{i+1}}{c_i} \prod_{l=i+1}^{m} \frac{1}{1+m_l} \right) \to 0.$$

(ii) In Case (d), for any $\varepsilon > 0$ and $l$ large enough we have $|m_l - R/l| \leq \varepsilon R/l$. This implies

$$(9.10)\quad \prod_{l=i+1}^{\lfloor \bar\beta_2 j\rfloor} \frac{1}{1+m_l} = \exp\left[ - \sum_{l=i+1}^{\bar\beta_1 j} \left( \frac{R}{l} + O(m_l^2) \right) \right].$$

Since $d_{i+1}/c_i \sim R/i$ and $m_l = O(1/l)$, it follows that

$$(9.11)\quad \sum_{i=\lfloor \bar\beta_2 j\rfloor}^{\lfloor \bar\beta_1 j\rfloor} \left( \frac{d_{i+1}}{c_i} \prod_{l=i+1}^{\lfloor \bar\beta_2 j\rfloor} \frac{1}{1+m_l} \right) \sim \sum_{i=\lfloor \bar\beta_2 j\rfloor}^{\lfloor \bar\beta_1 j\rfloor} \frac{R}{i} \left( \frac{\bar\beta_1 j}{i} \right)^{-R} \to 1 - \left( \frac{\bar\beta_2}{\bar\beta_1} \right)^{R}.$$

□

*Proof of Theorem 1.12.* In Case (A), $m_k \to \infty$, which implies fast clustering. In Case (B), $m_k \to \bar K + \bar M > 0$, which implies fast clustering. In Case (C1), $m_k \sim (c_k \sigma_k)^{-1} \to C > 0$, which implies fast clustering. In Case (C2), $d_k/c_k \sim m_k \sim (1-c)/c > 0$, which implies fast clustering. In Case (C3), $d_k/c_k \sim m_k \sim \mu_k/(c_k(\mu-1))$, which implies fast, diffusive and slow clustering depending on the asymptotic behaviour of $k\mu_k/c_k$. □

## 10 Dichotomy between clustering and coexistence for finite $N$

In this section, we prove Theorems 1.13–1.14.

*Proof of Theorem* 1.13. The key is the spatial version of the formulae for the first and second moments in terms of the coalescent process. The variance tends to zero for all evaluations if and only if the coalescent started from two individuals at a single site coalesces into one partition element. Therefore, all we have to show is that the hazard function for the time to coalesce is $H_N$, and then show that $\lim_{N\to\infty} H_N = \infty$ a.s. if and only if $\lim_{N\to\infty} \bar H_N = \infty$. The latter was already □

*Proof of Theorem* 1.14. We first note that the set of functions

$$(10.1)\quad \left\{ H_\varphi^{(n)}(\cdot, \pi_{G,n})\colon\ n \in \mathbb{N},\ \varphi \in C_\mathrm{b}(E^n, \mathbb{R}),\ \pi_{G,n} \in \Pi_{G,n} \right\},$$

is a distribution-determining subset of the set of bounded continuous functions on $\mathcal{P}(\mathcal{P}(E))^G$. It therefore suffices to establish the following:

(1) For all initial laws $\mathcal{L}[X^{(\Omega_N)}(0)]$ satisfying our assumptions for a given parameter $\theta \in \mathcal{P}(E)$ and all admissible $n, \varphi, \pi_{G,n}$,

$$(10.2)\quad \mathbb{E}\left[ H_\varphi^{(n)}(X^{(\Omega_N)}(t), \pi_{G,n}) \right] \xrightarrow[t\to\infty]{} F((\varphi, n, \pi_{G,n}), \theta),$$

which implies that $\mathcal{L}[X^{(\Omega_N)}(t)]$ converges to a limit law as $t \to \infty$ that depends on the initial law only through the parameter $\theta$.

(2) Depending on whether $\bar{H}_N < \infty$ or $\bar{H}_N = \infty$, the quantity in the right-hand side of (10.2) corresponds to the form of the limit claimed in (1.79–1.80).

Item (2) follows from Theorem 1.13 once we have proved the convergence result in (10.2), since (1.80) implies that the marginal law of the limiting state is $\delta_\theta$, and we will see in (10.5) below that recurrence of the transition kernel $\widehat{a}$ implies that

$$(10.3)\quad \mathbb{E}^{(\Omega_N)}_{\nu_\theta,\underline{c},\underline{\Lambda}}\left[\langle\varphi,\bigotimes_{i=1}^{n} x_{\eta_i}\rangle\right] = \langle f^n(u),\theta\rangle, \quad \text{for } \varphi(u_1,\cdots,u_n)=\prod_{i=1}^{n} f(u_i),$$

which in turn implies

$$(10.4)\quad \nu^{(\Omega_N)}_{\theta,\underline{c},\underline{\Lambda}} = \int_K (\delta_u)^{\otimes\Omega_N}\theta(\mathrm{d}u).$$

In order to prove item (1), we use duality and express the expectation in the left-hand side of (10.2) as an expectation over a coalescent $\mathfrak{C}^{(\Omega_N)}_t$ starting with $n$ partition elements. We therefore know that the number of partition elements, which is nonincreasing in $t$, converges to a limit as $t \to \infty$, which is 1 for $\bar{H}_N = \infty$ and a random number in $\{1,\ldots,n\}$ for $\bar{H}_N < \infty$. This means that there exists a finite random time after which the partition elements never meet again, and keep on moving by migration only. For such a scenario, it was proven in [DGV95], Lemma 3.2, that the positions of the partition elements are given, asymptotically, by $k = 1,\ldots,n$ random walks, all starting at the origin. Using that the initial state is ergodic, we can then calculate, for $\varphi(u_1,\cdots,u_n) = \Pi_{k=1}^n f(u_k)$,

$$(10.5)\quad \lim_{t\to\infty} E\left[H^{(n)}_\varphi\left(X^{(\Omega_N)}(0),\mathfrak{C}^{(\Omega_N)}_t\right)\right] = \sum_{k=1}^{n}\langle f,\theta\rangle^k q_k^{(\pi_{G,n})},$$

with $q_k^{(\pi_{G,n})}$ the probability that the coalescent starting in $\pi_{G,n}$ in the limit has $k$ remaining partition elements. Furthermore, if the initial positions of a sequence $(\pi^{(m)}_{G,n})_{m\in\mathbb{N}}$ of initial states satisfies $\lim_{m\to\infty} d(\eta_i^{(m)},\eta_j^{(m)}) = \infty$ for $i \neq j$, then for transient $\widehat{a}$ we know that

$$(10.6)\quad \lim_{m\to\infty} q_k^{(\pi^{(m)}_{G,n})} = 0, \quad \forall\, k = 1,\ldots,n-1 \text{ and } \lim_{m\to\infty} q_n^{(\Pi^{(m)}_{G,n})} = 1.$$

In view of (10.5), this proves that the law on $(\mathcal{P}(E))^G$ defined by the right-hand side of (10.2) is a translation-invariant and ergodic probability measure, with mean measure $\theta$ (see [DGV95], p. 2310, for details). □

## 11 Scaling of the volatility in the clustering regime

In Section 11.1, we prove Theorems 1.5 and 1.15, in Section 11.3 we prove Theorem 1.6.

### 11.1 Comparison with the hierarchical Fleming-Viot process

(a) Rewrite the recursion relation in (1.42) as

$$(11.1)\quad d_0 = 0, \qquad \frac{1}{d_{k+1}} = \frac{1}{c_k} + \frac{1}{\mu_k + d_k}, \qquad k \in \mathbb{N}_0.$$

From (11.1), it is immediate that $\underline{c} \mapsto \underline{d}$ and $\underline{\mu} \mapsto \underline{d}$ are component-wise non-decreasing.

(b) To compare $\underline{d}$ with $\underline{d}^*$, the solution of the recursion relation in (1.48) when $\mu_0 > 0$ and $\mu_k = 0$ for all $k \in \mathbb{N}$, simply note that $d_1 = d_1^* = c_0\mu_0/(c_0 + \mu_0)$. This gives

$$(11.2) \quad d_k \geq d_k^*, \quad k \in \mathbb{N},$$

with $d_k^*$ given by (1.49).

(c) Inserting the definition $m_k = (\mu_k + d_k)/c_k$ into (11.1), we get the recursion relation

$$(11.3) \quad c_0 m_0 = \mu_0, \qquad c_{k+1} m_{k+1} = \mu_{k+1} + \frac{c_k m_k}{1 + m_k}, \quad k \in \mathbb{N}_0.$$

Iterating (11.3), we get

$$(11.4) \quad c_k m_k = \sum_{l=0}^{k} \frac{\mu_l}{\prod_{j=l}^{k}(1 + m_j)}.$$

Ignoring the terms in the denominator, we get

$$(11.5) \quad m_k \leq \frac{1}{c_k} \sum_{l=0}^{k} \mu_l,$$

which proves that $\sum_{k \in \mathbb{N}_0} (1/c_k) \sum_{l=0}^{k} \mu_l < \infty$ implies $\sum_{k \in \mathbb{N}_0} m_k < \infty$. To prove the reverse, suppose that $\sum_{k \in \mathbb{N}_0} m_k < \infty$. Then $\prod_{j \in \mathbb{N}_0}(1 + m_j) = C < \infty$. Hence (11.4) gives

$$(11.6) \quad m_k \geq \frac{1}{C} \frac{1}{c_k} \sum_{l=0}^{k} \mu_l,$$

which after summation over $k \in \mathbb{N}_0$ completes the proof.

(d) We know from (1.49) that $d_k \geq d_k^* = \mu_0/(1 + \mu_0 \sigma_k)$ for $k \in \mathbb{N}$. Hence, if $\lim_{k\to\infty} \sigma_k = \infty$, then $\liminf_{k\to\infty} \sigma_k d_k \geq 1$. To get the reverse, note that iteration of (11.1) gives

$$(11.7) \quad \begin{aligned} \frac{1}{d_k} &= \sum_{l=0}^{k-1} \frac{1}{c_l \prod_{j=l+1}^{k-1}(1 + \frac{\mu_j}{d_j})} \geq \sum_{l=0}^{k-1} \frac{1}{c_l \prod_{j=l+1}^{k-1}(1 + \frac{\mu_j}{d_j^*})} \\ &\geq \sum_{l=0}^{k-1} \frac{1}{c_l \prod_{j=l+1}^{\infty}(1 + \frac{\mu_j}{\mu_0}[1 + \mu_0 \sigma_j])}. \end{aligned}$$

If $\sum_{j \in \mathbb{N}} \sigma_j \mu_j < \infty$, then the product in the last line tends to 1 as $l \to \infty$. Hence, if also $\lim_{k\to\infty} \sigma_k = \infty$, then it follows that $\liminf_{k\to\infty}(1/\sigma_k d_k) \geq 1$.

Note from the proof of (c) and (d) that in the *local coexistence regime* $d_k \sim \sum_{l=0}^{k} \mu_l$ as $k \to \infty$ when this sum diverges and $d_k \to \sum_{l \in \mathbb{N}_0} \mu_l / \prod_{j=l}^{\infty}(1 + m_j) \in (0, \infty)$ when it converges.

We close with the following observation. Since $1/c_k\sigma_k = (\sigma_{k+1} - \sigma_k)/\sigma_k$, $k \in \mathbb{N}$, and

$$(11.8) \quad \frac{\sigma_{k+1} - \sigma_k}{\sigma_1} \geq \frac{\sigma_{k+1} - \sigma_k}{\sigma_k} \geq \int_{\sigma_k}^{\sigma_{k+1}} \frac{\mathrm{d}x}{x}, \quad k \in \mathbb{N},$$

we have

$$(11.9) \quad \lim_{k\to\infty} \sigma_k = \infty \quad \Longleftrightarrow \quad \sum_{k \in \mathbb{N}} \frac{1}{c_k \sigma_k} = \infty.$$

## 11.2 Preparation: Möbius-transformations

To draw the scaling behaviour of $d_k$ as $k \to \infty$ from (11.1), we need to analyse the recursion relation

$$x_0 = 0, \qquad x_{k+1} = f_k(x_k), \quad k \in \mathbb{N}_0, \tag{11.10}$$

where

$$f_k(x) = \frac{c_k x + c_k \mu_k}{x + (c_k + \mu_k)}, \qquad x \neq -(c_k + \mu_k). \tag{11.11}$$

The map $x \mapsto f_k(x)$ is a Möbius-transformation on $\mathbb{R}^*$, the one-point compactification of $\mathbb{R}$. It has determinant $c_k(c_k+\mu_k) - c_k\mu_k = c_k^2 > 0$ and therefore is hyperbolic (see Kooman [K98]; a Möbius-transformation $f$ on $\mathbb{R}^*$ is called *hyperbolic* when it has two distinct fixed points at which the derivatives are not equal to $-1$ or $+1$.) Since

$$f_k'(x) = \left(\frac{c_k}{x + (c_k + \mu_k)}\right)^2, \qquad x \neq -(c_k + \mu_k), \tag{11.12}$$

it is strictly increasing except at $x = -(c_k + \mu_k)$, is strictly convex for $x < -(c_k + \mu_k)$ and strictly concave for $x > -(c_k + \mu_k)$, has horizontal asymptotes at height $c_k$ at $x = \pm\infty$ and vertical asymptotes at $x = -(c_k + \mu_k)$, and has two fixed points

$$x_k^+ = \tfrac{1}{2}\mu_k[-1 + \sqrt{1 + 4c_k/\mu_k}] \in (0, \infty), \qquad x_k^- = \tfrac{1}{2}\mu_k[-1 - \sqrt{1 + 4c_k/\mu_k}] \in (-\infty, 0), \tag{11.13}$$

of which the first is attractive ($f_k'(x_k^+) < 1$) and the second is repulsive ($f_k'(x_k^-) > 1$). For us only $x_k^+$ is relevant because, as is clear from (11.10), our iterations take place on $(0, \infty)$. See Fig. 5 for a picture of $f_k$.

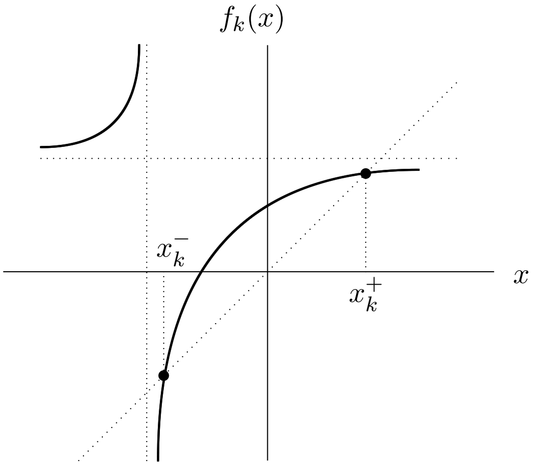

Figure 5: The Möbius-transformation $x \mapsto f_k(x)$.

In what follows, we will use the following two theorems of Kooman [K98]. We state the version of these theorems for $\mathbb{R}$, although they apply for $\mathbb{C}$ as well.

**Theorem 11.1.** [Kooman [K98], Corollary 6.5]
*Given a sequence of Möbius-transformations* $(f_k)_{k\in\mathbb{N}_0}$ *on* $\mathbb{R}^*$ *that converges pointwise to a Möbius-transformation* $f$ *that is hyperbolic. Then, for one choice of* $x_0 \in \mathbb{R}^*$ *the solution of the recursion relation* $x_{k+1} = f_k(x_k)$, $k \in \mathbb{N}_0$, *converges to the repulsive fixed point* $x^-$ *of* $f$, *while for all other choices of* $x_0$ *it converges to the attractive fixed point* $x^+$ *of* $f$.

**Theorem 11.2.** [Kooman [K98], Theorem 7.1]
*Given a sequence of Möbius-transformations* $(f_k)_{k\in\mathbb{N}_0}$ *on* $\mathbb{R}^*$ *whose fixed points are of bounded variation and converge to (necessarily finite) distinct limits, i.e.,*

$$\begin{aligned} &\sum_{k\in\mathbb{N}_0} |x^+_{k+1} - x^+_k| < \infty, \qquad \sum_{k\in\mathbb{N}_0} |x^-_{k+1} - x^-_k| < \infty, \\ &x^+ = \lim_{k\to\infty} x^+_k \in \mathbb{R}^*, \quad x^- = \lim_{k\to\infty} x^-_k \in \mathbb{R}^*, \quad x^+ \neq x^-. \end{aligned} \tag{11.14}$$

*If*

$$\prod_{k\in\mathbb{N}_0} |f'_k(x^+_k)| = 0, \tag{11.15}$$

*then, for one choice of* $x_0 \in \mathbb{R}^*$, *the solution of the recursion relation* $x_{k+1} = f_k(x_k)$, $k \in \mathbb{N}_0$, *converges to* $x^-$, *while for all other choices of* $x_0$ *it converges to* $x^+$. *If, on the other hand,*

$$\prod_{k\in\mathbb{N}_0} |f'_k(x^+_k)| > 0, \tag{11.16}$$

*then all choices of* $x_0 \in \mathbb{R}^*$ *lead to different limits.*

Theorem 11.1 deals with the situation in which there is a limiting hyperbolic Möbius-transformation, while Theorem 11.2 deals with the more general situation in which the limiting Möbius-transformation may not exist or may not be hyperbolic, but the fixed points do converge to distinct finite limits and they do so in a summable manner. (In Theorem 11.1, it is automatic that the fixed points of $f_k$ converge to the fixed points of $f$.) The conditions in (11.14–11.15) are necessary to ensure that the solutions of the recursion relation can reach the limits of the fixed points. Indeed, condition (11.16) prevents precisely that. As is evident from Fig. 5, the single value of $x_0$ for which the solution converges to the limit of the repulsive fixed point must satisfy $x_0 < 0$, which is excluded in our case because $x_0 = 0$.

## 11.3 Scaling of the volatility for polynomial coefficients

Theorem 1.6 shows *four regimes.* Our key assumptions are (1.55–1.58). For the scaling behaviour as $k \to \infty$ of the attractive fixed point $x^+_k$ given in (11.13) there are three regimes depending on the value of $K$:

$$x^+_k \sim \begin{cases} c_k, & \text{if } K = \infty, \\ M^+ c_k, & \text{if } K \in (0,\infty) \text{ with } M^+ = \frac{1}{2}K[-1 + \sqrt{1 + (4/K)}], \\ \sqrt{c_k \mu_k}, & \text{if } K = 0. \end{cases} \tag{11.17}$$

Our target will be to show that (recall $x_k$ from (11.10))

$$x_k \sim x^+_k \quad \text{as} \quad k \to \infty, \tag{11.18}$$

which is the scaling we are after in Theorems 1.6(a–c). We will see that (11.18) holds for $K \in (0,\infty]$, and also for $K = 0$ when $L = \infty$. A different situation arises for $K = 0$ when $L < \infty$, namely, $x_k \sim 1/\sigma_k$, which is the scaling we are after in Theorem 1.6(d).

For the proofs given in Sections 11.3.1–11.3.4, below we make use of Theorems 11.1–11.2 after doing the appropriate change of variables. Along the way we need the following elementary facts:

(I) If $(a_k)$ and $(b_k)$ have bounded variation, then both $(a_k + b_k)$ and $(a_k b_k)$ have bounded variation.

(II) If $(a_k)$ has bounded variation and $h\colon \mathbb{R} \to \mathbb{R}$ is globally Lipschitz on a compact interval containing the tail of $(a_k)$, then $(h(a_k))$ has bounded variation.

(III) If $(a_k)$ is bounded and is asymptotically monotone, then it has bounded variation.

Moreover, the following notion will turn out to be useful. According to Bingham, Goldie and Teugels [BGT87, Section 1.8], a strictly positive sequence $(a_k)$ is said to be *smoothly varying* with index $\rho \in \mathbb{R}$ if

$$(11.19)\quad \lim_{k\to\infty} k^n a_k^{[n]}/a_k = \rho(\rho-1)\times\cdots\times(\rho-n+1), \qquad n \in \mathbb{N},$$

where $a_k^{[n]}$ is the $n$-th order discrete derivative, i.e., $a_k^{[0]} = a_k$ and $a_k^{[n+1]} = a_{k+1}^{[n]} - a_k^{[n]}$, $k, n \in \mathbb{N}_0$.

(IV) If $(a_k)$ is smoothly varying with index $\rho \notin \mathbb{N}_0$, then $(a_k^{[n]})$ is asymptotically monotone for all $n \in \mathbb{N}$, while if $\rho \in \mathbb{N}$, then the same is true for all $n \in \mathbb{N}$ with $n \leq \rho$.

This observation will be useful in combination with (I–III).

According to [BGT87, Theorem 1.8.2], if $(a_k)$ is regularly varying with index $\rho \in \mathbb{R}$, then there exist smoothly varying $(a'_k)$ and $(a''_k)$ with index $\rho$ such that $a'_k \leq a_k \leq a''_k$ and $a'_k \sim a''_k$. In words, any regularly varying function can be sandwiched between two smoothly varying functions with the same asymptotic behaviour. In view of the monotonicity property in Theorem 1.5(a), it therefore suffices to prove Theorem 1.6 under the following assumption, which is stronger than (1.55):

$$(11.20)\quad \begin{array}{l} (c_k),\ (\mu_k),\ (\mu_k/c_k),\ (k^2\mu_k/c_k) \text{ are smoothly varying} \\ (\text{with index } a,\ b,\ a-b, \text{ respectively, } 2+a-b). \end{array}$$

### 11.3.1 Case (b)

Let $K \in (0,\infty)$. Put $y_k = x_k/c_k$. Then the recursion relation in (11.10) becomes

$$(11.21)\quad y_0 = 0, \qquad y_{k+1} = g_k(y_k), \quad k \in \mathbb{N}_0,$$

where

$$(11.22)\quad g_k(y) = \frac{A_k y + B_k}{C_k y + D_k}, \qquad y \in \mathbb{R}^*,$$

with coefficients

$$(11.23)\quad A_k = \frac{c_k^2}{c_{k+1}}, \quad B_k = \frac{c_k\mu_k}{c_{k+1}}, \quad C_k = c_k, \quad D_k = c_k + \mu_k.$$

By (1.55), we have $c_k/c_{k+1} \sim 1$, and hence $A_k \sim C_k \sim c_k$, $B_k \sim Kc_k$, $D_k \sim (K+1)c_k$. Therefore, (11.22) yields

$$\text{(11.24)} \quad \lim_{k\to\infty} g_k(y) = g(y) = \frac{y+K}{y+(K+1)}, \qquad y \in \mathbb{R}^*.$$

Since $g$ is hyperbolic with fixed points $y^{\pm} = M^{\pm} = \frac{1}{2}K[-1 \pm \sqrt{1+(4/K)}]$, we can apply Theorem 11.1 and conclude that

$$\text{(11.25)} \quad \lim_{k\to\infty} y_k = M^{+}.$$

### 11.3.2 Case (a)

Let $K = \infty$. Again put $y_k = x_k/c_k$. Then the same recursion relation as in (11.21–11.22) holds with the same coefficients as in (11.23), but this time $c_k/c_{k+1} \sim 1$ gives $A_k \sim C_k \sim c_k$, $B_k \sim D_k \sim \mu_k$, and

$$\text{(11.26)} \quad \lim_{k\to\infty} g_k(y) = g(y) = 1, \qquad y \in \mathbb{R}^*.$$

Since $g$ is not hyperbolic, we cannot apply Theorem 11.1. To compute $y^{\pm} = \lim_{k\to\infty} y_k^{\pm}$, we note that $g_k$ has fixed points

$$\text{(11.27)} \quad y_k^{\pm} = \frac{1}{a_k} h^{\pm}(b_k/a_k^2) \text{ with } h^{\pm}(x) = \frac{1}{2x}\big(1 \mp \sqrt{1+4x}\big),\ a_k = \frac{A_k - D_k}{B_k},\ b_k = \frac{C_k}{B_k}$$

(use that $a_k < 0$ for $k$ large enough). Since $c_k/\mu_k \to 0$, we have $a_k \to -1$ and $b_k \to 0$. It follows that $y_k^{+} \to y^{+} = 1$ and $y_k^{-} \to y^{-} = -\infty$, so that we can apply Theorem 11.2. To prove that $y_k \to y^{+} = 1$, we need to check that (recall (11.14–11.15))

(1) $(y_k^{+})_{k\in\mathbb{N}_0}$ has bounded variation.

(2) $\prod_{k\in\mathbb{N}_0} g_k'(y_k^{+}) = 0$.

(What happens near $y_k^{-}$ is irrelevant because $x_k > 0$ for all $k$.)

To prove (1), note that $h^{+}$ is globally Lipschitz near zero. Since, by (11.23) and (11.27),

$$\text{(11.28)} \quad a_k = \frac{c_k}{\mu_k}\left(1 - \frac{c_{k+1}}{c_k}\right) - \frac{c_{k+1}}{c_k}, \quad b_k = \frac{c_k}{\mu_k}\frac{c_{k+1}}{c_k},$$

it follows from (1.56), (I), (III–IV) and (11.20) that $(a_k)$ and $(b_k)$ have bounded variation. Since $a_k \to -1$ and $b_k \to 0$, it in turn follows from (I–II) that $(1/a_k)$ and $(b_k/a_k^2)$ have bounded variation. Via (I–II) this settles (1).

To prove (2), note that

$$\text{(11.29)} \quad g_k'(y_k^{+}) = \frac{\Delta_k}{(C_k y_k^{+} + D_k)^2} \text{ with } \Delta_k = A_k D_k - B_k C_k.$$

Since $y_k^{+} > 0$ and $D_k > \mu_k$, we have

$$\text{(11.30)} \quad \prod_{k\in\mathbb{N}_0} g_k'(y_k^{+}) \le \prod_{k\in\mathbb{N}_0} \frac{\Delta_k}{\mu_k^2}.$$

But $\Delta_k = c_k^3/c_{k+1}$ and so, because $c_k/c_{k+1} \sim 1$, we have $\Delta_k/\mu_k^2 = c_k^3/c_{k+1}\mu_k^2 \sim (c_k/\mu_k)^2 \to 0$. Hence (2) indeed holds.

### 11.3.3 Case (c)

Let $K = 0$ and $L = \infty$. Put $y_k = x_k/\sqrt{c_k\mu_k}$. Then the same recursion relation as in (11.21–11.22) holds with coefficients

$$(11.31)\quad A_k = c_k\sqrt{\frac{c_k\mu_k}{c_{k+1}\mu_{k+1}}}, \quad B_k = c_k\mu_k\sqrt{\frac{1}{c_{k+1}\mu_{k+1}}}, \quad C_k = \sqrt{c_k\mu_k}, \quad D_k = c_k + \mu_k.$$

By (1.55), $c_{k+1}/c_k \sim 1$ and $\mu_{k+1}/\mu_k \sim 1$, and hence $A_k \sim D_k \sim c_k$, $B_k \sim C_k \sim \sqrt{c_k\mu_k}$. Therefore (11.22) yields

$$(11.32)\quad \lim_{k\to\infty} g_k(y) = g(y) = y, \qquad y \in \mathbb{R}^*.$$

Since $g$ is not hyperbolic, we cannot apply Theorem 11.1. To compute $y^\pm = \lim_{k\to\infty} y_k^\pm$ from (11.27), we abbreviate

$$(11.33)\quad \alpha_k = \frac{c_{k+1}}{c_k} - 1, \quad \beta_k = \frac{\mu_{k+1}}{\mu_k} - 1, \quad \gamma_k = \frac{\mu_k}{c_k},$$

and write

$$(11.34)\quad a_k = \frac{1}{\sqrt{\gamma_k}}\left[1 - (1+\gamma_k)\sqrt{(1+\alpha_k)(1+\beta_k)}\right], \quad b_k = \sqrt{(1+\alpha_k)(1+\beta_k)}.$$

We have $\alpha_k \to 0$, $\beta_k \to 0$, $\gamma_k \to 0$. Moreover, (1.56–1.58), (IV) and (11.20) imply that $(k\alpha_k)$ and $(k\beta_k)$ are asymptotically monotone and bounded. Together with $\lim_{k\to\infty} k^2\gamma_k = \infty$ this in turn implies that $\alpha_k/\sqrt{\gamma_k} \to 0$ and $\beta_k/\sqrt{\gamma_k} \to 0$. Hence $a_k \to 0$ and $b_k \to 1$, and therefore (11.27) yields $y^\pm = \pm 1$, so that we can apply Theorem 11.2.

To prove (1), note that (1.56–1.58), (IV) and (11.20) also imply that $(\sqrt{\gamma_k})$ and $(1/\sqrt{k^2\gamma_k})$, are asymptotically monotone and bounded. By (11.34) and (I–III), this in turn implies that $(a_k)$ and $(b_k)$ have bounded variation. Indeed, the first equality in (11.34) can be rewritten as

$$(11.35)\quad a_k = \frac{1}{\sqrt{\gamma_k}}\,\frac{1 - (1+\gamma_k)^2(1+\alpha_k)(1+\beta_k)}{1 + (1+\gamma_k)\sqrt{(1+\alpha_k)(1+\beta_k)}}.$$

The denominator tends to 2, is Lipschitz near 2, and has bounded variation because $(\alpha_k)$, $(\beta_k)$, $(\gamma_k)$ have bounded variation. The numerator equals $-\alpha_k - \beta_k - 2\gamma_k$ plus terms that are products of $\alpha_k$, $\beta_k$ and $\gamma_k$. Writing $\alpha_k/\sqrt{\gamma_k} = k\alpha_k/\sqrt{k^2\gamma_k}$ and $\beta_k/\sqrt{\gamma_k} = k\beta_k/\sqrt{k^2\gamma_k}$ and using that $\sqrt{k^2\gamma_k} \to \infty$, we therefore easily get the claim.

To prove (2), note that

$$(11.36)\quad \Delta_k = c_k^2\sqrt{\frac{c_k\mu_k}{c_{k+1}\mu_{k+1}}} = c_k^2/\sqrt{(1+\alpha_k)(1+\beta_k)}, \quad C_k y_k^+ + D_k = c_k(1 + y_k^+\sqrt{\gamma_k} + \gamma_k),$$

and hence

$$(11.37)\quad \prod_{k\in\mathbb{N}_0} g_k'(y_k^+) \leq \prod_{k\in\mathbb{N}_0} \frac{1}{\sqrt{(1+\alpha_k)(1+\beta_k)}(1 + y_k^+\sqrt{\gamma_k})^2}.$$

The term under the product equals

$$(11.38)\quad 1 - 2y^+\sqrt{\gamma_k}\,[1 + o(1)],$$

which yields (2) because $\sqrt{k^2\gamma_k} \to \infty$.

### 11.3.4 Case (d)

Let $K = 0$ and $L < \infty$. Put $y_k = \sigma_k x_k$. Then the same recursion relation as in (11.21–11.22) holds with coefficients

$$A_k = c_k \frac{\sigma_{k+1}}{\sigma_k}, \quad B_k = c_k \mu_k \sigma_{k+1}, \quad C_k = \frac{1}{\sigma_k}, \quad D_k = c_k + \mu_k. \tag{11.39}$$

Abbreviate

$$\delta_k = \frac{\sigma_{k+1}}{\sigma_k} - 1 = \frac{1}{c_k \sigma_k}. \tag{11.40}$$

We have $k\mu_k/c_k \to 0$ and, by (1.55), $c_{k+1}/c_k \sim 1$, $\sigma_{k+1}/\sigma_k \sim 1$ and $k\delta_k \to 1-a$ with $a \in (-\infty, 1)$ the exponent in (1.55). It therefore follows that

$$\frac{A_k}{D_k} \to 1, \quad \frac{B_k}{D_k} \sim \mu_k \sigma_k = \frac{k\mu_k}{c_k} \frac{1}{k\delta_k} \to 0, \quad \frac{C_k}{D_k} \sim \frac{1}{c_k \sigma_k} = \delta_k \to 0. \tag{11.41}$$

Hence (11.22) yields

$$\lim_{k\to\infty} g_k(y) = g(y) = y, \qquad y \in \mathbb{R}^*. \tag{11.42}$$

Since $g$ is not hyperbolic, we cannot apply Theorem 11.1. To compute $y^\pm = \lim_{k\to\infty} y_k^\pm$, we rewrite (11.27) as

$$y_k^\pm = \tfrac{1}{2}\left(\bar{a}_k \pm \sqrt{\bar{a}_k^2 + 4\bar{b}_k}\right) \qquad \text{with} \qquad \bar{a}_k = \frac{A_k - D_k}{C_k}, \quad \bar{b}_k = \frac{B_k}{C_k}, \tag{11.43}$$

and note that

$$\begin{aligned} \bar{a}_k &= \frac{c_k}{c_{k+1}} - \mu_k \sigma_k = \frac{c_k}{c_{k+1}} - \frac{k\mu_k}{c_k} \frac{1}{k\delta_k}, \\ \bar{b}_k &= c_k \mu_k \sigma_k \sigma_{k+1} = \frac{k^2 \mu_k}{c_k} \frac{\sigma_{k+1}}{\sigma_k} \frac{1}{(k\delta_k)^2}. \end{aligned} \tag{11.44}$$

Since $k^2\mu_k/c_k \to L < \infty$ and $k\delta_k \to 1-a$ with $a \in (-\infty,1)$ the exponent in (1.55), it follows that $\bar{a}_k \to 1$ and $\bar{b}_k \to L/(1-a)^2$. Hence $y_k^\pm \to y^\pm = \frac{1}{2}(1 \pm \sqrt{1 + 4L/(1-a)^2})$, so that we can apply Theorem 11.2.

To prove (1), note that (1.56–1.58), (I–IV) and (11.20) imply that $(\bar{a}_k)$ and $(\bar{b}_k)$ have bounded variation. This yields the claim via (11.43).

To prove (2), note that

$$\begin{aligned} \Delta_k &= c_k^2 \frac{\sigma_{k+1}}{\sigma_k} = c_k^2(1+\delta_k), \\ C_k y_k^+ + D_k &= \frac{y_k^+}{\sigma_k} + c_k + \mu_k = c_k \left(1 + \delta_k y_k^+ + \frac{\mu_k}{c_k}\right), \end{aligned} \tag{11.45}$$

and, hence,

$$\prod_{k\in\mathbb{N}_0} g_k'(y_k^+) \leq \prod_{k\in\mathbb{N}_0} \frac{1+\delta_k}{(1+\delta_k y_k^+)^2}. \tag{11.46}$$

The term under the product equals

$$1 - (2y^+ - 1)\delta_k\,[1 + o(1)], \tag{11.47}$$

Since $y^+ \geq 1$, it follows that (2) holds if and only if $\sum_{k\in\mathbb{N}_0} \delta_k = \infty$, which by (11.9) and (11.40) holds if and only $\lim_{k\to\infty} \sigma_k = \infty$. Theorem 11.2 shows that failure of (2) implies that $y_k$ converges to a limit different from 1.

### 11.4 Scaling of the volatility for exponential coefficients

In this section, we briefly comment on how to extend the proof of Theorem 1.6 to cover the case of Theorem 1.7.

The claims made for Cases (A) and (B) follow from minor adaptations of the arguments for Cases (a) and (b) in Sections 11.3.2 and 11.3.1. The claim made for Case (C1) follows from Theorem 1.5(d). The claims made for Cases (C2) and (C3) follow from minor adaptations of the arguments for Cases (b) and (c) in Sections 11.3.1 and 11.3.3. The details are left to the reader.

## 12 Notation index

### 12.1 General notation

- $E \rightsquigarrow$ compact Polish space of types.
- $\mathcal{P}(E) \rightsquigarrow$ set of probability measures on $E$.
- $M(E) \rightsquigarrow$ set of measurable functions on $E$.
- $\mathcal{M}([0,1]) \rightsquigarrow$ set of non-negative measures on $[0,1]$.
- $\mathcal{M}_f([0,1]) \rightsquigarrow$ set of finite non-negative measures on $[0,1]$.
- $\mathcal{L} \rightsquigarrow$ law.
- $\Longrightarrow \rightsquigarrow$ weak convergence on path space.
- $\Lambda^* \in \mathcal{M}([0,1]) \rightsquigarrow$ (cf. (1.5)).
- $\Lambda \in \mathcal{M}_f([0,1]) \rightsquigarrow$ (cf. Section 1.3).
- $\frac{\partial F(x)}{\partial x_i}[\delta_a] \rightsquigarrow$ Gâteaux-derivative of $F$ with respect to $x_i$ in the direction $\delta_a$ (cf. (1.13)).
- $D(T, \mathcal{E}) \rightsquigarrow$ set of càdlàg paths in $\mathcal{E}$ indexed by the elements of $T \subset \mathbb{R}$ and equipped with the Skorokhod $J_1$-topology.
- $C_{\mathrm{b}}(\mathcal{E}, \mathcal{E}') \rightsquigarrow$ set of continuous bounded mappings from $\mathcal{E}$ to $\mathcal{E}'$.

### 12.2 Interacting Λ-Cannings processes

- $\Omega_N \rightsquigarrow$ hierarchical group of order $N$ (cf. (1.20)).
- $\underline{c} = (c_k)_{k \in \mathbb{N}_0} \in (0,\infty)^{\mathbb{N}_0} \rightsquigarrow$ migration coefficients (cf. (1.24)).
- $\underline{\Lambda} = (\Lambda_k)_{k \in \mathbb{N}_0} \in \mathcal{M}_f([0,1])^{\mathbb{N}_0} \rightsquigarrow$ offspring measures (cf. (1.27)).
- $\lambda_k = \Lambda_k([0,1]) \rightsquigarrow$ resampling rates (cf. (1.29)).
- $\underline{d} = (d_k)_{k \in \mathbb{N}_0} \rightsquigarrow$ volatility constants (cf. (1.42)).
- $\underline{m} = (m_k)_{k \in \mathbb{N}_0} \rightsquigarrow$ (cf. (1.44)).
- $\mu_k = \frac{1}{2}\lambda_k \rightsquigarrow$ (cf. (1.44)).

- $\sigma_k \rightsquigarrow$ (cf. (1.49)).
- $B_k(\eta) \rightsquigarrow k$-block around $\eta$ (cf. (1.22)).
- $y_{\eta,k} \rightsquigarrow$ type distribution in $B_k(\eta)$ (cf. (1.30)).
- $C^\Lambda$-process $\rightsquigarrow$ non-spatial continuum-mass $\Lambda$-Cannings process (cf. Section 1.3.1).
- $a^{(N)}(\cdot,\cdot) \rightsquigarrow$ hierarchical random walk kernel on $\Omega_N$ (cf. (1.25)).
- $C_N^{c,\Lambda}$-process $\rightsquigarrow$ hierarchically interacting Cannings process on $\Omega_N$ (cf. Section 1.4.4).
- $L^{(N)}$, $L^{(N)}_{\mathrm{mig}}$, $L^{(N)}_{\mathrm{res}} \rightsquigarrow$ generators of the mean-field Cannings process (cf. (1.11)).
- $L^{(\Omega_N)}$, $L^{(\Omega_N)}_{\mathrm{mig}}$, $L^{(\Omega_N)}_{\mathrm{res}} \rightsquigarrow$ generators of the hierarchical Cannings process (cf. (1.35)).
- $\Phi_{r,a,B_k(\eta)} \rightsquigarrow$ reshuffling-resampling map (cf. (1.38)).
- $X^{(\Omega_N)} \rightsquigarrow C_N^{c,\Lambda}$-process (cf. Section 1.4.4).
- $Y^{(\Omega_N)}_{\eta,k}(\cdot) \rightsquigarrow$ macroscopic observables (= block averages) of $X^{(\Omega_N)}$ (cf. (1.40)).
- $y^{[1]}_\eta \rightsquigarrow$ 1-block averages indexed block-wise (cf. (7.18)).
- $G_{N,K} \rightsquigarrow K$-level truncation of $\Omega_N$ (cf. (1.39)).
- $X^{(N)} \rightsquigarrow$ mean-field interacting Cannings process (cf. Section 1.3.2).
- $Q_x(\mathrm{d}u,\mathrm{d}v) \rightsquigarrow$ Fleming-Viot diffusion function (cf. (1.18)).
- $L^{c,d,\Lambda}_\theta$, $L^c_\theta$, $L^d$, $L^\Lambda \rightsquigarrow$ generators of the McKean-Vlasov process (cf. (1.16)).
- $Z^{c,d,\Lambda}_\theta \rightsquigarrow$ McKean-Vlasov process with immigration-emigration (cf. Section 1.3.3).
- $\nu^{c,d,\Lambda}_\theta \rightsquigarrow$ unique equilibrium of $Z$ (cf. (4.1)).
- $Y^{(\Omega_N)}_{\eta,k}(\cdot) \rightsquigarrow$ macroscopic observables (= block averages) of $X^{(\Omega_N)}$ (cf. (1.40)).
- $(M^{(j)}_k)_{k=-(j+1),\dots,0} \rightsquigarrow$ interaction chain (cf. Section 1.5.5).

### 12.3 Spatial Λ-coalescents

- $[n] = \{1,\dots,n\}$.
- $\Pi_n \rightsquigarrow$ set of all partitions of $[n]$ into disjoint families (cf. (2.4)).
- $\Pi_{G,n} \rightsquigarrow$ set of $G$-labelled partitions of $[n]$ (cf. (2.6)).
- $S_{G,n} \in \Pi_{G,n} \rightsquigarrow G$-labelled partition into singletons (cf. (2.7)).
- $\Pi$, $\Pi_G \rightsquigarrow$ partitions of $\mathbb{N}$, $G$-labelled partitions of $\mathbb{N}$ (cf. (2.10)).
- $L(\pi_G) \rightsquigarrow$ set of labels of partition $\pi_G$ (cf. (2.9)).
- $\lambda^{(\Lambda)}_{b,i} \rightsquigarrow$ coalescence-rates (cf. (2.13)).

- $\cdot|_n$ ⇝ operation of projection from $[m]$ (respectively, $\mathbb{N}$) onto $[n]$.
- $L^{(\Omega_N)*}$, $L^{(\Omega_N)*}_{\text{mig}}$, $L^{(\Omega_N)*}_{\text{coal}}$ ⇝ generators of the hierarchical Cannings-coalescent (cf. (2.32)).
- $\mathfrak{P}$ ⇝ field of Poisson point processes driving the spatial $\Lambda$-coalescent (cf. (2.14)).
- $\mathfrak{P}^{(\Omega_N)}$ ⇝ driving Poisson point process for the spatial $n$-$\Lambda$-coalescent with block coalescence (cf. (2.26)).
- $\mathfrak{C}^{(G)}_n$ ⇝ spatial $n$-$\Lambda$-coalescent on $G$ (cf. (2.17)).
- $\mathfrak{C}^{(G)}$ ⇝ spatial $\Lambda$-coalescent (cf. (2.19)).
- $\mathfrak{C}^{(\Omega_N)}$ ⇝ spatial $\underline{\Lambda}$-coalescent with block coalescence (cf. (2.30)).

## References

[BCGH95] J.B. Baillon, Ph. Clément, A. Greven and F. den Hollander, On the attracting orbit of a nonlinear transformation arising from renormalization of hierarchically interacting diffusions, Part 1: The compact case, Can. J. Math. 47 (1995) 3–27.

[BCGH97] J.B. Baillon, Ph. Clément, A. Greven and F. den Hollander, On the attracting orbit of a nonlinear transformation arising from renormalization of hierarchically interacting diffusions, Part 2: The non-compact case, J. Funct. Anal. 147 (1997) 236–298.

[BEV10] N.H. Barton, A.M. Etheridge and A. Véber, A new model for evolution in a spatial continuum, Electr. J. Prob. 15 (2010) 162–216.

[B09] N. Berestycki, Recent progress in coalescent theory, Ensaios Matematicos 16 (2009) 1–193.

[BEV] N. Berestycki, A.M. Etheridge and A. Véber, Large scale behaviour of the spatial Lambda-Fleming-Viot process, to appear in Ann. Inst. H. Poincaré: Probab. Statist.

[BL05] J. Bertoin and J.-F. Le Gall, Stochastic flows associated to coalescent processes II: Stochastic differential equations, Ann. Inst. H. Poincaré: Probab. Statist. 41 (2005) 307–333.

[BGT87] N.H. Bingham, C.M. Goldie and J.L. Teugels, *Regular Variation*, Encyclopedia of Mathematics and its Applications, Vol. 27, Cambridge University Press, Cambridge, 1987.

[BB08] M. Birkner and J. Blath, Computing likelihoods for coalescents with multiple collisions in the infinitely many sites model, J. Math. Biol. 57 (2008) 435–465.

[BB09-MVD] M. Birkner and J. Blath, Measure-valued diffusions, general coalescents and population genetic inference, Trends in stochastic analysis, London Math. Soc. Lecture Note Ser. 353, Cambridge University Press, Cambridge, 2009, pp. 329–363.

[BG68] R.M. Blumenthal and R.K. Getoor, *Markov Processes and Potential Theory*, Academic Press, New York, 1968.

[C74] C. Cannings, The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models, Adv. Appl. Probab. 6 (1974) 260–290.

[C75] C. Cannings, The latent roots of certain Markov chains arising in genetics: a new approach, II. Further haploid models, Adv. Appl. Probab. 7 (1975) 264–282.

[CDG04] J.T. Cox, D.A. Dawson and A. Greven, *Mutually Catalytic Super Branching Random Walks: Large Finite Systems and Renormalization Analysis*, Memoirs of the American Mathematical Society, Volume 171, Number 809, Amer. Math. Soc., Providence, RI, 2004.

[CGr86] J.T. Cox and D. Griffeath, Diffusive clustering in the two-dimensional voter model, Ann. Probab. 14 (1986) 347–370.

[D93] D.A. Dawson, *Measure-valued Markov processes*, in: École d'Été de Probabilités de Saint-Flour XXI-1991, Lecture Notes in Mathematics 1541, Springer, Berlin, 1993, pp. 1–260.

[DGW05] D.A. Dawson, L.G. Gorostiza and A. Wakolbinger, Degrees of transience and recurrence and hierarchical random walks, Potential Anal. 22 (2005) 305–350.

[DG93a] D.A. Dawson and A. Greven, Multiple time scale analysis of hierarchically interacting diffusions, in: *A Festschrift in Honour of Gopinath Kallianpur* (eds. S. Cambanis, J.K. Gosh, R.L. Karandikar and P.K. Sen), Springer, 1993, pp. 50–51.

[DG93b] D.A. Dawson and A. Greven, Multiple scale analysis of interacting diffusions, Probab. Theory Relat. Fields 95 (1993) 467–508.

[DG93c] D.A. Dawson and A. Greven, Hierarchical models of interacting diffusions: multiple time scales, phase transitions and cluster-formation, Probab. Theory Relat. Fields 96 (1993) 435–473.

[DG96] D.A. Dawson and A. Greven, Multiple space-time scale analysis for interacting branching models, Electr. J. Probab. 1 (1996) 1–84.

[DG99] D.A. Dawson and A. Greven, Hierarchically interacting Fleming-Viot processes with selection and mutation: Multiple space-time scale analysis and quasi-equilibria, Electr. J. Probab. 4 (1999) 1–81.

[DG03] D.A. Dawson and A. Greven, State dependent multitype spatial branching processes and their longtime behavior, Electr. J. Probab. 8 (2003) 1–93.

[DGHSS08] D.A. Dawson, A. Greven, F. den Hollander, R. Sun and J. Swart, The renormalization transformation for two-type branching models, Ann. I. Henri Poincaré: Probab. Statist. 44 (2008) 1038–1077.

[DGV95] D.A. Dawson, A. Greven and J. Vaillancourt, Equilibria and quasi-equilibria for infinite systems of Fleming-Viot processes, Trans. Amer. Math. Soc. 347 (1995) 2277–2360.

[DGW] D.A. Dawson, L.G. Gorostiza, and A. Wakolbinger, Hierarchical random walks, in: *Asymptotic Methods in Stochastics: Festschrift for Miklós Csörgő* (eds. L. Horváth, B. Szyszkowicz), Fields Inst. Commun. 44, Amer. Math. Soc., Providence, RI, 2004, pp. 173–193.

[DM95] D.A. Dawson and P. March, Resolvent estimates for Fleming-Viot operators and uniqueness of solutions to related martingale problems, J. Funct. Anal. 132 (1995) 417–472.

[DEP11] R. Der, Ch. Epstein and J.B. Plotkin, Generalized population models and the nature of genetic drift, Theor. Popul. Biol. 80 (2011) 80–99.

[DK99] P. Donnelly and T. Kurtz, Particle representations for measure-valued population models, Ann. Probab. 1 (1999) 166–205.

[EW06] B. Eldon and J. Wakeley, Coalescent processes when the distribution of offspring number among individuals is highly skewed, Genetics 172 (2006) 2621–2633.

[E00] A.M. Etheridge, *An Introduction to Superprocesses*, Amer. Math. Soc., Providence, RI, 2000.

[E11] A.M. Etheridge, *Some Mathematical Models from Population Genetics*, in: École d'Été de Probabilités de Saint-Flour XXXIX-2009, Lecture Notes in Mathematics 2012, Springer, Berlin, 2011.

[EK86] S.N. Ethier and T. Kurtz, *Markov Processes. Characterization and Convergence*, John Wiley, New York, 1986.

[Ev97] S.N. Evans. Coalescing Markov labelled partitions and a continuous sites genetics model with infinitely many types, Ann. Inst. H. Poincaré Probab. Statist. 33 (1997) 339–358.

[FG94] K. Fleischmann and A. Greven, Diffusive clustering in an infinite system of hierarchically interacting Fisher-Wright diffusions, Probab. Theory Relat. Fields 98 (1994) 517–566.

[Fr] N. Freeman, Phase transitions in a spatial coalescent, preprint 2012, arXiv:1109.4363v2 [math.PR].

[Gre05] A. Greven, Renormalization and universality for multitype population models, in: *Interacting Stochastic Systems* (eds. J.-D. Deuschel, and A. Greven), Springer, Berlin, 2005, pp. 209–244.

[GH07] A. Greven and F. den Hollander, Phase transitions for the long-time behavior of interacting diffusions, Ann. Probab. 35 (2007) 1250–1306.

[GKWpr] A. Greven, A. Klimovsky and A. Winter, Tree-valued Cannings dynamics, in preparation.

[GLW10] A. Greven, V. Limic and A. Winter, Coalescent processes arising in a study of diffusive clustering, to appear in Electr. J. Probab.

[GPW09] A. Greven, P. Pfaffelhuber and A. Winter, Convergence in distribution of random metric measure spaces: Lambda-coalescent measure trees, Probab. Theory Relat. Fields 145 (2009) 285–322.

[HS98] F. den Hollander and J. Swart, Renormalization of hierarchically interacting isotropic diffusions, J. Stat. Phys. 93 (1998) 243–291.

[JM86] A. Joffe and M. Métivier, Weak convergence of sequences of semimartingales with applications to multitype bra/nching processes, Adv. Appl. Probab. 18 (1986) 20–65.

[K98] R.J. Kooman, Asymptotic behaviour of solutions of linear recurrences and sequences of Möbius-transformations, J. Approx. Theory 93 (1998) 1–58.

[L85] T.M. Liggett, *Interacting Particle Systems*, Springer, Berlin, 1985.

[LS81] T.M. Liggett and F. Spitzer, Ergodic theorems for coupled random walks and other systems with locally interacting components, Z. Wahrsch. Verw. Gebiete 56 (1981) 443–468.

[LS06] V. Limic and A. Sturm, The spatial Λ-coalescent, Electr. J. Probab. 11 (2006) 363–393.

[P99] J. Pitman, Coalescents with multiple collisions, Ann. Probab. 27 (1999) 1870–1902.