ALEA, Lat. Am. J. Probab. Math. Stat. 11 (1), 43–140 (2014)

Renormalisation of hierarchically interacting Cannings processes

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

Department Mathematik, Universität Erlangen-Nürnberg
Cauerstraße 11, D-91058 Erlangen, Germany
E-mail address: greven@mi.uni-erlangen.de
URL: http://www.mathematik.uni-erlangen.de/~greven

Mathematical Institute, Leiden University
P.O. Box 9512, NL-2300 RA Leiden, The Netherlands
E-mail address: denholla@math.leidenuniv.nl
URL: http://www.math.leidenuniv.nl/~denholla

AG Wahrscheinlichkeitstheorie, Fakultät für Mathematik
Universität Duisburg-Essen, D-45117 Essen, Germany
E-mail address: sandra.kliem@uni-due.de
URL: https://www.uni-due.de/~hm0135

AG Wahrscheinlichkeitstheorie, Fakultät für Mathematik
Universität Duisburg-Essen, D-45117 Essen, Germany
E-mail address: ak@aklimovsky.net
URL: http://www.aklimovsky.net/

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^Λ-process, which arises as the continuum-mass limit of a Λ-Cannings individual-based population model, where Λ 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 Λ-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^Λ-processes, where the interaction comes from migration and

Received by the editors December 11, 2012; accepted february 7, 2014.

2010 *Mathematics Subject Classification.* Primary 60J25, 60K35; Secondary 60G57, 60J60, 60J75, 82C28, 92D25.

Key words and phrases. C^Λ-process, Λ-coalescent, hierarchical group, migration, multi-scale reshuffling-resampling, spatial coalescent, hierarchical mean-field limit, renormalisation, McKean-Vlasov process, Möbius-transformation.

reshuffling-resampling on *all* hierarchical space-time scales *simultaneously*. More precisely, individuals live in colonies labelled by the hierarchical group Ω_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-macro-colonies*, for all $k \in \mathbb{N}_0$. The reshuffling is linked to the resampling: before resampling in a macro-colony takes place all individuals in that macro-colony 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^Λ-process is the Λ-coalescent, whereas the dual process of the $C_N^{\underline{c},\underline{\Lambda}}$-process is a spatial coalescent with multi-scale non-local 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-macro-colony averages of the $C_N^{\underline{c},\underline{\Lambda}}$-process at the macroscopic time scale N^k (= the volume of the k-macrocolony) converge to a random process that is a superposition of a C^{Λ_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 $(k+1)$-macro-colony average. It turns out that d_k is a function of c_l and Λ_l for all $0 \le l < k$. 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*.

We discuss the implications of the scaling of d_k for the behaviour on large space-time scales of the $C_N^{\underline{c},\underline{\Lambda}}$-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 non-local 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.

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^Λ-process, as well as the so-called McKean-Vlasov C^Λ-process, a single-colony C^Λ-process with immigration and emigration from and to a cemetery state arising in the context of the scaling limit of the multi-colony C^Λ-process with mean-field interaction. Section 1.4 defines a *new* process, the $C_N^{\underline{c},\underline{\Lambda}}$-process, where the countably many colonies are labelled by the hierarchical group Ω_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^{Λ_k}-process and a single-colony Fleming-Viot process with a volatility d_k that is a function of c_l and Λ_l for all $0 \le l < k$, 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*. The implications of this scaling for the behaviour of the $C_N^{\underline{c},\underline{\Lambda}}$-process on large space-time scales is discussed in detail, and the outcome is compared 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 multi-scale non-local 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 (1993); Donnelly and Kurtz (1999); Etheridge (2000, 2011)). 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 Ω_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 et al. (1995)). 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 Dawson and March (1995). 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 Dawson et al. (1995).

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 (1974, 1975)), then we enter the world of *jump processes*. In this paper, we will focus on jump processes that are parametrised by a measure Λ 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 (2006)). It is argued in Der et al. (2011) that mixtures of diffusive dynamics and Cannings dynamics provide a better fit to generation-by-generation empirical data from *Drosophila* populations. Birkner and Blath (2008, 2009) treat the issue of statistical inference on the genealogies corresponding to a one-parameter family of Cannings dynamics. None of these models includes the effect of geography.

Our goal is to describe the effect of jumps in a *spatial* setting with a volatile reproduction. To that end, we add two ingredients: (1) a geographic space with a migration mechanism; (2) a spatially structured reproduction mechanism. As a result, we obtain a system of *interacting Cannings processes*.

As geographic space, we choose a hierarchically structured lattice: the hierarchical group, i.e., we study a system of hierarchically interacting Cannings processes. The interaction is chosen in such a way that the geographic space mimics the *two*-dimensional Euclidean space, with the migration of individuals given by independent random walks.

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. This is a *first attempt* to account for the fact that the volatility the Cannings model tries to capture results from *catastrophic events* on a smaller time scale (with a geographic structure). In this view, 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 and in a complete geographic redistribution of individuals during a single observation time. To carry out this idea fully, the mechanism should actually be modelled by specifying a random environment. In this work, however, we concentrate on the case of spatially homogeneous parameters. The case of spatially inhomogeneous parameters (modelled via a random environment) is left for future work. On a technical level, we will see that in our model the reshuffling substantially simplifies the analysis.

The idea to give reproduction a non-local geographic structure, in particular, in two dimensions, was exploited by Barton et al. (2010) and by Berestycki et al. (2013) also[1]. There, the process lives on the torus of sidelength L and is constructed via its dual, and it is shown that a limiting process on $\mathbb{R}^2$ exists as $L \to \infty$. In Barton et al. (2010); Berestycki et al. (2013), it is assumed that the individual lineages are compound Poisson processes. Freeman (2013) 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. The latter set-up 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 Ω_N of order N and passes to the limit $N \to \infty$ ("the hierarchical mean-field limit")[2]. 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 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 (1993c,a,b, 1996, 1999, 2003); Dawson et al. (1995), and Cox et al. (2004). For the *analytic part* (i.e., the study of the renormalisation map $\mathcal{F}$), see Baillon et al. (1995, 1997); den Hollander and Swart (1998), and Dawson et al. (2008).

So far, two important classes of single-colony processes could not be treated: *Anderson diffusions* Greven and den Hollander (2007) and *jump processes*. In the present paper, we focus on the second class, in particular, on so-called C^Λ-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

[1] In the literature, there is an alternative terminology – "generalised Λ-Fleming-Viot process" or "jump-type Fleming-Viot process" – which refers to the continuum-mass limit of the original discrete individual-based Cannings model. In this paper, we stick to the name "Cannings process" also for the continuum-mass limit.

[2] 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 et al. (2004)). We will comment on this issue in Section 1.4.2.

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

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

where $g^* \equiv 1$ and d_k depends on d_{k-1}, c_{k-1} and the total mass of Λ_{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^Λ-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^Λ-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 Λ-Cannings model involves a finite non-negative measure $\Lambda \in \mathcal{M}_f([0,1])$. Below, we often assume that

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

and Λ satisfying the so-called *dust-free condition*

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

Condition (1.2) excludes the well-studied case of *Fleming-Viot diffusions*. In this paper, we are primarily interested in the new effects brought by the *pure jump* case in the Λ-Cannings model. These effects were not studied using renormalisation techniques previously. Besides the pure jump case, later on, we allow for superpositions of Fleming-Viot diffusion and pure-jump Λ-Cannings models (cf. Sections 1.3.3 and 1.4.4). 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 (1999, Theorem 8)), and also to be able to define a genealogical tree for the population (Greven et al. (2009))[3].

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

[3]Condition (1.3) is relevant for some of the questions addressed in this paper, though not for all. We comment on this issue 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.

1.3.1. *Single-colony C^Λ-process.* We recall the definition of the Λ-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
$$\theta \in \mathcal{P}(E). \tag{1.4}$$
- Let $\Lambda^* \in \mathcal{M}([0,1])$ be the σ-finite non-negative measure defined as
$$\Lambda^*(\{0\}) = 0, \qquad \Lambda^*(\mathrm{d}r) = \frac{\Lambda(\mathrm{d}r)}{r^2}, \quad r \in (0,1]. \tag{1.5}$$
Consider an inhomogeneous Poisson point process on $[0,\infty) \times [0,1]$ with intensity measure
$$\mathrm{d}t \otimes \Lambda^*(\mathrm{d}r). \tag{1.6}$$
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 uniformly chosen at random among all the individuals marked by a 1 (see Fig. 1.1).

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

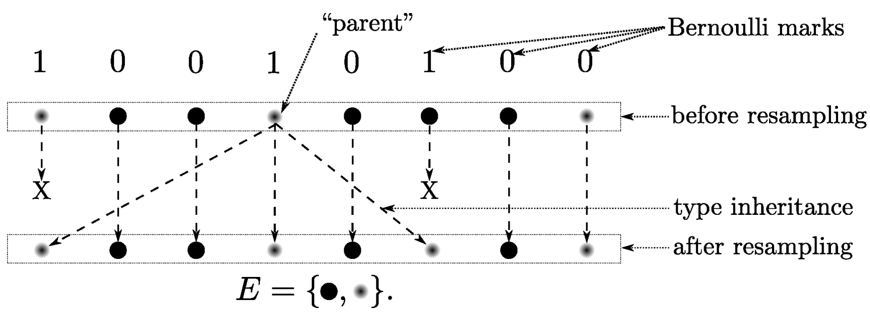

FIGURE 1.1. Cannings resampling event in a colony of $M = 8$ individuals of two types. Arrows indicate type inheritance, X indicates death.

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
$$\int_{(0,1]} \frac{\Lambda(\mathrm{d}r)}{r^2} \tfrac{1}{2} M(M-1)\, r^2 = \tfrac{1}{2} M(M-1)\, \Lambda((0,1]) < \infty, \tag{1.7}$$
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 $\mathcal{P}(E)$. Equip $\mathcal{P}(E)$ with the topology of weak convergence of probability measures.

Letting $M \to \infty$, we obtain a limiting process $X = (X(t))_{t\geq 0}$, called the C^Λ-*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 (1999)). This process has countably many jumps in any finite time interval when $\Lambda((0,1]) > 0$.

Note that the limiting case $\Lambda = \delta_0$ is the Fleming-Viot diffusion (cf. Section 1.3.3). It is well known that this limiting case is obtained as a scaling limit of the Moran model.

1.3.2. *Multi-colony C^Λ-process: mean-field version.* Next, we consider the *spatial* Λ-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. 1.2).
- Individuals *resample* within each colony according to the Λ-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 $\mathcal{P}(E)^N$.

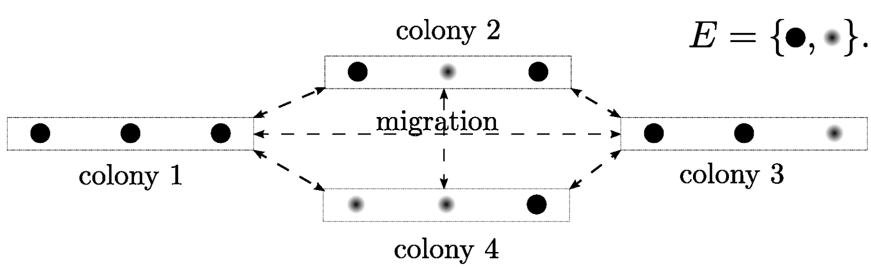

FIGURE 1.2. Possible one-step migration paths between $N = 4$ colonies with $M = 3$ individuals of two types in the mean-field version.

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

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

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 $C_\mathrm{b}(\mathcal{P}(E)^N,\mathbb{R})$ 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

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

The generator

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

has two parts,

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

The *migration operator* is given by

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

where

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

is the Gâteaux-derivative of F with respect to x_i in the direction δ_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 ν 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 verbal description of the single-colony C^Λ-process in Section 1.3.1)

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

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.

The proof of Proposition 1.1 is given in Section 3.2.

1.3.3. *C^Λ-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 a 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, Λ, θ), $c, d \in (0, \infty)$,

$\Lambda \in \mathcal{M}_f([0,1])$, $\theta \in \mathcal{P}(E)$, or C^Λ-processes with immigration-emigration at rate c with source θ 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

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

Define the second Gâteaux-derivative of F with respect to x as

$$\frac{\partial^2 F(x)}{\partial x^2}[\delta_u,\delta_v] = \frac{\partial}{\partial x}\left(\frac{\partial F(x)}{\partial x}[\delta_u]\right)[\delta_v], \quad u,v \in E. \tag{1.16}$$

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

$$L_\theta^{c,d,\Lambda} = L_\theta^c + L^d + L^\Lambda \tag{1.17}$$

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

$$\begin{aligned}
(L_\theta^c F)(x) &= c\int_E (\theta - x)\,(\mathrm{d}a)\,\frac{\partial F(x)}{\partial x}[\delta_a],\\
(L^d F)(x) &= d\int_E\int_E Q_x(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(x)}{\partial x^2}[\delta_u,\delta_v],\\
(L^\Lambda F)(x) &= \int_{(0,1]} \Lambda^*(\mathrm{d}r)\int_E x(\mathrm{d}a)\,\big[F\big((1-r)x + r\delta_a\big) - F(x)\big],
\end{aligned} \tag{1.18}$$

where

$$Q_x(\mathrm{d}u,\mathrm{d}v) = x(\mathrm{d}u)\,\delta_u(\mathrm{d}v) - x(\mathrm{d}u)\,x(\mathrm{d}v) \tag{1.19}$$

is the Fleming-Viot diffusion coefficient. The three parts of $L_\theta^{c,d,\Lambda}$ correspond to: a *drift* towards θ of strength c (immigration-emigration), a *Fleming-Viot diffusion* with *volatility* d (Moran resampling), and a C^Λ-process with resampling measure Λ (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 Λ-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.

The proof of Proposition 1.2 is given in Section 3.2.

Denote by

$$Z_\theta^{c,d,\Lambda} = \big(Z_\theta^{c,d,\Lambda}(t)\big)_{t\geq 0}, \quad Z_\theta^{c,d,\Lambda}(0) = \theta, \tag{1.20}$$

the solution of the martingale problem in Proposition 1.2 for the special choice $x = \theta$. This is called the *McKean-Vlasov process*[4] with parameters c,d,Λ and initial state θ.

[4]The terminology stems from the fact that this process describes the limiting behaviour of an interacting particle system for which propagation of chaos holds. The physics terminology is related to the fact that the system of independent components is more random (= more chaotic) than the one with dependent components. In our context, in the mean-field limit ($N \to \infty$), the components of the system become independent of each other. Therefore, “chaos propagates”.

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 Ω_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* Ω_N of order N is the set

$$\Omega_N = \Big\{\eta = (\eta^l)_{l\in\mathbb{N}_0} \in \{0,1,\dots,N-1\}^{\mathbb{N}_0}\colon\ \sum_{l\in\mathbb{N}_0} \eta^l < \infty\Big\}, \qquad N \in \mathbb{N}\backslash\{1\}, \tag{1.21}$$

endowed with the addition operation $+$ defined by $(\eta+\zeta)^l = \eta^l + \zeta^l \pmod N$, $l \in \mathbb{N}_0$ (see Fig. 1.3 for the case $N = 3$). In other words, Ω_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 Ω_N is equipped with the ultrametric distance $d(\cdot,\cdot)$ defined by

$$d(\eta,\zeta) = d(0,\eta-\zeta) = \min\{k\in\mathbb{N}_0\colon\ \eta^l = \zeta^l,\ \text{for all } l \geq k\}, \qquad \eta,\zeta\in\Omega_N. \tag{1.22}$$

Let

$$B_k(\eta) = \{\zeta\in\Omega_N\colon\ d(\eta,\zeta)\leq k\}, \qquad \eta\in\Omega_N,\ k\in\mathbb{N}_0, \tag{1.23}$$

denote the k-block around η, which we think of as a *macro-colony*. The geometry of Ω_N is explained in Fig. 1.3).

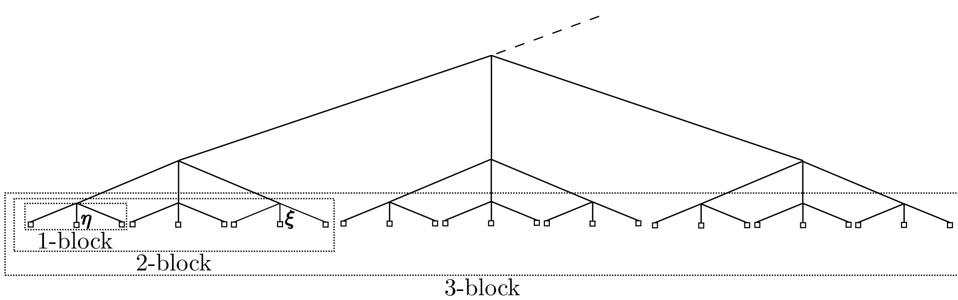

FIGURE 1.3. Close-ups of a 1-block, a 2-block and a 3-block in the hierarchical group of order $N = 3$. The elements of the group are the leaves of the tree ($\Box$). The hierarchical distance between two elements is the graph distance to the most recent common ancestor: $d(\xi,\eta) = 2$ for ξ and η in the picture.

We construct a process

$$X^{(\Omega_N)} = \big(X^{(\Omega_N)}(t)\big)_{t\geq 0} \quad \text{with} \quad X^{(\Omega_N)}(t) = \big\{X^{(\Omega_N)}_\eta(t)\big\}_{\eta\in\Omega_N} \in \mathcal{P}(E)^{\Omega_N}, \tag{1.24}$$

by using the same evolution mechanism as for the multi-colony system in Section 1.3.2, except that we replace the migration on $\{0,\ldots,N-1\}$ by a migration on Ω_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 Ω_N through a random walk kernel. For that purpose, we introduce a sequence of migration rates

$$\underline{c} = (c_k)_{k\in\mathbb{N}_0} \in (0,\infty)^{\mathbb{N}_0}, \tag{1.25}$$

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 kernel of the random walk that is thus performed by each individual are

$$a^{(N)}(\eta,\zeta) = \sum_{k\geq d(\eta,\zeta)} \frac{c_{k-1}}{N^{2k-1}}, \qquad \eta,\zeta\in\Omega_N,\ \eta\neq\zeta, \qquad a^{(N)}(\eta,\eta)=0. \tag{1.26}$$

As shown in Dawson et al. (2005), 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 $c>1$[5].

Throughout the paper, we assume that[6]

$$\limsup_{k\to\infty} \tfrac{1}{k}\log c_k < \infty. \tag{1.27}$$

This guarantees that the total migration rate per individual is bounded (at least for sufficiently large N).

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

$$\underline{\Lambda} = (\Lambda_k)_{k\in\mathbb{N}_0} \in \mathcal{M}_f([0,1])^{\mathbb{N}_0}, \tag{1.28}$$

each subject to (1.2). Assume in addition that

$$\int_{(0,1]} \Lambda_k^*(\mathrm{d}r) < \infty, \qquad k\in\mathbb{N}, \tag{1.29}$$

and that Λ_0 satisfies (1.3). The condition in (1.29) is needed to guarantee that in finite time a colony is affected by finitely many reshuffling-resampling events

[5] 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 (1.26) with $c=1$ is similar to that of simple random walk on $\mathbb{Z}^2$.

[6] In Section 1.5.3, we will analyse the case $N<\infty$, where (1.27) must be replaced by $\limsup_{k\to\infty}\frac{1}{k}\log c_k < \log N$.

only, since otherwise this transition cannot be defined (see Remark 1.3 at the end of Section 1.4). The condition in (1.3) guarantees that the population has a well-defined genealogy and most of the population at a site goes back to a finite number of ancestors after a positive finite time.

Set

$$\lambda_k = \Lambda_k([0,1]), \qquad \lambda_k^* = \Lambda_k^*([0,1]), \qquad k \in \mathbb{N}_0. \tag{1.30}$$

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. (1.5–1.6) and (2.28)):

- 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. 1.4). After that, r is drawn according to the intensity measure Λ_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.,

$$y_{\eta,k} \equiv N^{-k} \sum_{\zeta \in B_k(\eta)} x_\zeta. \tag{1.31}$$

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.[7]

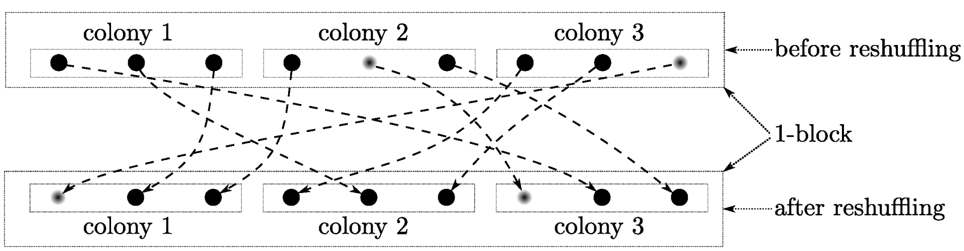

FIGURE 1.4. Random reshuffling in a 1-block on the hierarchical lattice of order $N = 3$ with $M = 3$ individuals of two types per colony.

Throughout the paper, we assume that $\underline{\lambda}^* = (\lambda_k^*)_{k\in\mathbb{N}_0}$ (recall the definition of λ_k^* from (1.30)) satisfies[8]

$$\limsup_{k\to\infty} \tfrac{1}{k} \log \lambda_k^* < \infty. \tag{1.32}$$

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 (1.32) guarantees that the total resampling rate per individual is bounded.

In the continuum-mass limit, the reshuffling-resampling operation, when it acts on the states in the colonies, takes the form

$$x_\zeta \text{ is replaced by } (1-r)y_{\eta,k} + r\delta_a \text{ for all } \zeta \in B_k(\eta) \tag{1.33}$$

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

[8] In Section 1.5.3, we will analyse the case $N < \infty$, where (1.32) must be replaced by $\limsup_{k\to\infty} \frac{1}{k} \log \lambda_k^* < \log N$.

with $a \in E$ drawn from $y_{\eta,k}$ (the type distribution in $B_k(\eta)$ (cf. (1.31)). Note that in the mean-field case and in the single-colony case of Section 1.3.1, $a \in E$ is drawn from x_ζ (cf. (1.14) and the comment following it)[9].

1.4.4. *Hierarchical Cannings process.* We are now ready to formally define our system of *hierarchically interacting C^Λ-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 non-local 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

$$\begin{aligned} &F(x) = \int_{E^n} \left(\bigotimes_{m=1}^{n} x_{\eta_m}(\mathrm{d}u^m) \right) \varphi\big(u^1, \ldots, u^n\big), \quad x = (x_\eta)_{\eta \in \Omega_N} \in \mathcal{P}(E)^{\Omega_N}, \\ &n \in \mathbb{N}, \quad \varphi \in C_{\mathrm{b}}(E^n, \mathbb{R}), \quad \eta_1, \ldots, \eta_n \in \Omega_N. \end{aligned} \tag{1.34}$$

The linear operator for the martingale problem

$$L^{(\Omega_N)} \colon \mathcal{F} \to C_{\mathrm{b}}\big(\mathcal{P}(E)^{\Omega_N}, \mathbb{R}\big) \tag{1.35}$$

again has two parts,

$$L^{(\Omega_N)} = L^{(\Omega_N)}_{\mathrm{mig}} + L^{(\Omega_N)}_{\mathrm{res}}. \tag{1.36}$$

The *migration operator* is given by

$$(L^{(\Omega_N)}_{\mathrm{mig}} F)(x) = \sum_{\eta,\zeta \in \Omega_N} a^{(N)}(\eta,\zeta) \int_E (x_\zeta - x_\eta)(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\eta}[\delta_a] \tag{1.37}$$

and the *reshuffling-resampling operator* by

$$\begin{aligned} (L^{(\Omega_N)}_{\mathrm{res}} F)(x) = \sum_{\eta \in \Omega_N} \Bigg(&(L^{d_0}_\eta F)(x) + \int_{(0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\eta(\mathrm{d}a) \\ &\qquad \times \big[F\big(\Phi_{r,a,\{\eta\}}(x)\big) - F(x)\big] \\ &+ \sum_{k \in \mathbb{N}} N^{-2k} \int_{(0,1]} \Lambda_k^*(\mathrm{d}r) \int_E y_{\eta,k}(\mathrm{d}a) \\ &\qquad \times \big[F\big(\Phi_{r,a,B_k(\eta)}(x)\big) - F(x)\big] \Bigg), \end{aligned} \tag{1.38}$$

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

$$\Big[\big(\Phi_{r,a,B_k(\eta)}\big)(x)\Big]_\zeta = \begin{cases} (1-r) y_{\eta,k} + r\delta_a, & \zeta \in B_k(\eta), \\ x_\zeta, & \zeta \notin B_k(\eta), \end{cases} \tag{1.39}$$

where $r \in [0,1]$, $a \in E$, $k \in \mathbb{N}_0$, $\eta \in \Omega_N$, and $L^{d_0}_\eta$ is the Fleming-Viot diffusion operator with volatility d_0 (see (1.18)) acting on the colony x_η with

$$d_0 \geq 0. \tag{1.40}$$

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

Remark 1.3. (1) If $d_0 = 0$, then the operator in (1.38) is pure-jump.
(2) The right-hand side of (1.38) is well-defined because of (1.29). Indeed, by Taylor-expanding the inner integral in (1.38) in powers of r, we get

$$\int_E y_{\eta,k}(\mathrm{d}a)\left[F\left(\Phi_{r,a,B_k(\eta)}(x)\right) - F(x)\right] = F(y_{\eta,k}) - F(x) + O(r^2), \quad \text{as } r \downarrow 0. \tag{1.41}$$

To have a well-defined resampling operator (1.38), the expression in (1.41) must be integrable with respect to $\Lambda_k^*(\mathrm{d}r)$, which is equivalent to assumption (1.29).

Proposition 1.4. [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*[10]*. The unique solution is a strong Markov process with the Feller property.*

The proof of Proposition 1.4 is given in Section 3.2.

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 Ω_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

$$G_{N,K} = \{0, \dots, N-1\}^K, \tag{1.42}$$

which is a truncation of the hierarchical group Ω_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. This approximation of $X^{(\Omega_N)}$ by $X^{(G_{N,K})}$ is made rigorous in Proposition 8.1

Remark: Similarly to the mean-field Cannings process $X^{(N)}$ from Section 1.3.2, the hierarchical Cannings process $X^{(\Omega_N)}$ can be obtained as a $M \to \infty$ limit of the finite M individual-based models.

1.5. *Main results.* Our main results concern a multiscale analysis of the $C_N^{\underline{c},\underline{\Lambda}}$-process on Ω_N, $X^{(\Omega_N)}$ (cf. below Proposition 1.4) 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

$$Y_{\eta,k}^{(\Omega_N)}(t) = \frac{1}{N^k} \sum_{\zeta \in B_k(\eta)} X_\zeta^{(\Omega_N)}(t), \qquad \eta \in \Omega_N, \tag{1.43}$$

which constitute a *renormalisation of space* where the component η 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

$$\left(\left(Y_{\eta,k}^{(\Omega_N)}(tN^k)\right)_{\eta\in\Omega_N}\right)_{t\geq 0}, \tag{1.44}$$

[10] As a part of the definition of the martingale problem, we always require that the solution has càdlàg paths and is adapted to a natural filtration.

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 Ω_N by dropping the first k-entries of $\eta \in \Omega_N$ (recall (1.21)). 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 1.5.1–1.5.2. In Section 1.5.1, we state the scaling behaviour of the renormalised interacting system in (1.44) as $N \to \infty$ for fixed $k \in \mathbb{N}_0$. In Section 1.5.2, we look at the interaction chain that captures the scaling behaviour on all scales simultaneously. In Section 1.5.3, we take a look at our system $X^{(\Omega_N)}$ for finite N. In Section 1.5.4, we compare the result with the hierarchical Fleming-Viot process. In Sections 1.5.5–1.5.6, we identify the different regimes for $k \to \infty$ and in Section 1.5.7 we investigate cluster formation.

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 Ω_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 θ for some $\theta \in \mathcal{P}(E)$.

Recall (1.30) and (1.40). Let $\underline{d} = (d_k)_{k\in\mathbb{N}_0}$ be the sequence of *volatility constants* defined recursively as

$$d_{k+1} = \frac{c_k(\frac{1}{2}\lambda_k + d_k)}{c_k + (\frac{1}{2}\lambda_k + d_k)}, \quad k \in \mathbb{N}_0. \tag{1.45}$$

Let $\mathcal{L}$ denote law, let $\Longrightarrow$ denote weak convergence on path space, and recall (1.20).

Theorem 1.5. [Hierarchical mean-field limit and renormalisation]
For every $k \in \mathbb{N}$, uniformly in $\eta \in \Omega_\infty$,

$$\mathcal{L}\left[\left(Y_{\eta,k}^{(\Omega_N)}(tN^k)\right)_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(Z_\theta^{c_k,d_k,\Lambda_k}(t)\right)_{t\geq 0}\right]. \tag{1.46}$$

For $k = 0$, (1.46) is still true, but the McKean-Vlasov process must be started from $Z(0) = X_\eta^{(\Omega_N)}(0)$ instead of $Z(0) = \theta$ (cf. (1.20)).

The proof of Theorem 1.5 is given in Section 8. The limiting process in (1.46) is a *McKean-Vlasov process* with drift constant $c = c_k$ and resampling measure $d_k\delta_0 + \Lambda_k$ (cf. (1.18)). This shows that the class of Cannings models with block resampling is preserved under the renormalisation.

Heuristics. In order to understand the origin of the recursion relation in (1.45), let us start by explaining where $d_1 = c_0\lambda_0/(2c_0 + \lambda_0)$ comes from. Consider two *lineages*[11] drawn at random from a macro-colony of order 1, say $B_1(\eta)$ for some $\eta \in \Omega_N$. Due to migration, both lineages are uniformly distributed over the macro-colony after the first migration step. For each lineage, marking the migration steps that result in being in the same colony, we get a Poisson process with rate $2c_0$ on timescale Nt. For every such mark, the rate to coalesce is $\lambda_0 N$ (on time scale Nt), while the rate to migration away is $2c_0N$. Hence, the probability that the two lineages coalesce before they migrate away is $\lambda_0/(2c_0 + \lambda_0)$. Therefore, thinning the Poisson process with rate $2c_0$, we see that the two lineages coalesce at rate

[11] The fact that we consider coalescing lineages as opposed to type distributions is actually the essence of the *duality* approach to the study of the dynamics of interacting particle systems. In the present context, duality is developed in Section 2.

$2c_0\lambda_0/(2c_0+\lambda_0)$. Since the coalescence rate is twice the diffusion coefficient (cf., Section 4.4), this gives a heuristic explanation for d_1. Note that three lineages are within the same colony only after a time of order N^2, so three lineages do not coalesce on time scale Nt.

To understand the generic step of the recursion relation, i.e., d_{k+1}, consider a macro-colony of order $k+1$, say $B_{k+1}(\eta)$ for some $\eta \in \Omega_N$, and two lineages drawn at random from this macro-colony. Consider only migration on level k, i.e., migration events between the macro-colonies of order k, which occur at rate $2c_k N^{-k}$. For every such event, the rate of coalescence is $2d_k+\lambda_k$, while migration of one of them occurs at rate $2c_k$. Hence, the probability that the two lineages coalesce before one of them migrates is $(2d_k+\lambda_k)/(2c_k+2d_k+\lambda_k)$. After speeding up time by a factor N, we see that the coalescence rate is $2c_k(2d_k+\lambda_k)/(2c_k+2d_k+\lambda_k)$. Since the coalescence rate is twice the diffusion coefficient, this gives a heuristic explanation for d_k. Again, three or more lineages do not coalesce on the same time scale.

1.5.2. *Multi-scale analysis: the interaction chain.* **Multi-scale behaviour.** Our second main theorem looks at the implications of the scaling behaviour of d_k as $k \to \infty$, to be described in Theorems 1.11–1.12 in Section 1.5.4–1.5.5, for which we must extend Theorem 1.5 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*

$$M^{(j)} = (M_k^{(j)})_{k=-(j+1),\dots,0}, \qquad j \in \mathbb{N}_0, \tag{1.47}$$

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 $-(j+1)$,

$$M_{-(j+1)}^{(j)} = \theta \in \mathcal{P}(E), \tag{1.48}$$

and with transition kernel

$$K_k(x,\cdot) = \nu_x^{c_k,d_k,\Lambda_k}(\cdot), \quad x \in \mathcal{P}(E),\, k \in \mathbb{N}_0, \tag{1.49}$$

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 (1.18) of Section 1.3.3 (see Section 4 for details).

Theorem 1.6. **[Multi-scale behaviour]**
Let $(t_N)_{N\in\mathbb{N}}$ be such that

$$\lim_{N\to\infty} t_N = \infty \text{ and } \lim_{N\to\infty} t_N/N = 0. \tag{1.50}$$

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

$$\begin{aligned} &\mathcal{L}\left[\left(Y_{\eta,k}^{(\Omega_N)}(N^j t_N + N^k u_k)\right)_{k=j,\dots,0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(M_{-k}^{(j)}\right)_{k=j,\dots,0}\right],\\ &\mathcal{L}\left[Y_{\eta,j+1}^{(\Omega_N)}(N^j t_N)\right] \underset{N\to\infty}{\Longrightarrow} \delta_\theta, \end{aligned} \tag{1.51}$$

where $\theta \in \mathcal{P}(E)$ is the single-site mean of the initial distribution $X^{(\Omega_N)}(0)$, cf. Section 1.5.1.

The proof of Theorem 1.6 is given in Section 9.

Theorem 1.6 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 $k+1$.

Heuristics. The effect described in Theorem 1.6 results from the fact that on the smaller time scale uN^k a k-block average evolves effectively like a single component of the $N-1$ other k-block averages with a mean-field migration mechanism. This leads to *propagation of chaos*, i.e., convergence to a system of independently evolving components that interact only because they feel the overall type density in the $(k+1)$-block. Since we look at the system at a late time $N^j t_N$, we see that the dynamics at scale $N^k u$, which is $o(N^j t_N)$, has already reached equilibrium, as is clear from a restart argument that absorbs an order-N^k term into $N^j t_N$.

The basic dichotomy. We next let the index in the multi-scale renormalisation scheme tend to infinity and identify how the limit depends on the parameters $(\underline{c}, \underline{\Lambda})$. Indeed, Theorem 1.6, in combination with Theorems 1.11–1.12 in Sections 1.5.4–1.5.5, allows us to study the universality properties on large space-time scales when we first let $N \to \infty$ and then $j \to \infty$[12].

The interaction chain exhibits a *dichotomy*, as will be seen in Theorem 1.7 below, in the sense that

$$\mathcal{L}\left[M_0^{(j)}\right] \underset{j\to\infty}{\Longrightarrow} \nu_\theta \in \mathcal{P}(\mathcal{P}(E)), \tag{1.52}$$

with ν_θ either (I) of the form of a random single-atom measure, i.e.,

$$\nu_\theta = \mathcal{L}[\delta_U], \text{ for some random } U \in E \text{ with } \mathcal{L}[U] = \theta, \tag{1.53}$$

or (II) ν_θ spread out. To be more specific, define

$$\mathrm{Var}_x(\psi) = \int_{E\times E} [x(\mathrm{d}u)\delta_u(\mathrm{d}v) - x(\mathrm{d}u)x(\mathrm{d}v)]\, \psi(u)\psi(v). \tag{1.54}$$

Then, ν_θ is spread out iff

$$\sup_{\psi\in B_1} \mathbb{E}_{\nu_\theta}[\mathrm{Var}_\cdot(\psi)] > 0, \tag{1.55}$$

where $B_1 \equiv C_\mathrm{b}(E,\mathbb{R}) \cap \{\psi\colon\ |\psi| \le 1\}$ and the expectation is taken with respect to the parameter x in (1.54), i.e.,

$$\mathbb{E}_{\nu_\theta}[\mathrm{Var}_\cdot(\psi)] = \int_{\mathcal{P}(E)} \nu_\theta(\mathrm{d}x)\, \mathrm{Var}_x(\psi). \tag{1.56}$$

Case (I) is called the *clustering regime*, since it indicates the formation of large mono-type regions, while case (II) 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 comparing 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 (see Proposition 4.2(b) below). The latter is referred to in Dawson et al. (1995) 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

[12] For several previously investigated systems, the limit as $j \to \infty$ was shown to be interchangeable (Dawson et al. (1995); Fleischmann and Greven (1994))

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 $-\infty$[13].

Explicit dichotomy criterion. The large-scale behaviour of $X^{(\Omega_N)}$ is determined by the sequence $\underline{m} = (m_k)_{k\in\mathbb{N}_0}$ with

$$m_k = \frac{\mu_k + d_k}{c_k}, \text{ where } \mu_k = \tfrac{1}{2}\lambda_k \tag{1.57}$$

(recall c_k from (1.25), λ_k from (1.30) and d_k from (1.45)). We will argue that the dichotomy

$$\sum_{k\in\mathbb{N}_0} m_k = \infty \quad \text{vs.} \quad \sum_{k\in\mathbb{N}_0} m_k < \infty \tag{1.58}$$

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).

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. In (4.26) of Section 4.4, we will show that

$$\mathbb{E}_{\mathcal{L}[M_0^{(j)}]}\left[\mathrm{Var.}(\psi)\right] = \left[\prod_{k=0}^{j} \frac{1}{1+m_k}\right] \mathrm{Var}_\theta(\psi), \quad j \in \mathbb{N}_0,\, \psi \in C_\mathrm{b}(E,\mathbb{R}),\, \theta \in \mathcal{P}(E). \tag{1.59}$$

This implies that the entrance law is trivial when $\sum_{k\in\mathbb{N}_0} m_k = \infty$ and non-trivial when $\sum_{k\in\mathbb{N}_0} m_k < \infty$. Our third main theorem identifies the dichotomy.

Theorem 1.7. [Dichotomy of the entrance law]

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

$$\begin{cases} \mathcal{L}\left[\left(M_k^{(j)}\right)_{k=-(j+1),\dots,0}\right] \underset{j\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(M_k^{(\infty)}\right)_{k=-\infty,\dots,0}\right], \\ M_{-\infty}^{(\infty)} = \theta. \end{cases} \tag{1.60}$$

(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*

$$\sup_{\psi\in C_\mathrm{b}(E,\mathbb{R})} \mathbb{E}_{\mathcal{L}[M_0^{(\infty)}]}[\mathrm{Var.}(\psi)] > 0. \tag{1.61}$$

The proof of Theorem 1.7 is given in Section 9.2.

Theorem 1.7, in combination with Theorem 1.11(c) in Section 1.5.4, 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.

Heuristics. If the resampling happens only locally, i.e., $\lambda_k = 0$, for $k \in \mathbb{N}$, we simply obtain the two regimes depending on whether two ancestral lines coalesce with probability 1 or < 1, giving after a long time monotype or coexistence, if and only if they meet with probability 1 or < 1. Now, the ancestral lines can coalesce

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

due to the reshuffling-resampling in a k-ball and hence the occupation time of two ancestral lines in the distances k weighted by the λ_k is the relevant quantity.

1.5.3. *Main results for finite N.* In this section, we take a look at our system $X^{(\Omega_N)}$ ($C_N^{\underline{c},\Lambda}$-process on Ω_N, cf. below Proposition 1.4) 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 θ, or to an equilibrium state, where different types live next to each other?

As it will turn out below, in the finite-N case there is the dichotomy and, moreover, the quantitative criterion is the same as in the $N \to \infty$ limit.

Concretely, let $P_t(\cdot,\cdot)$ denote the transition kernel of the random walk on Ω_N with migration coefficients

$$\bar{c}_k(N) = c_k + N^{-1}\lambda_{k+1}, \quad k \in \mathbb{N}_0 \tag{1.62}$$

starting at 0 (cf. Section 1.4.2). Let

$$\bar{H}_N = \sum_{k\in\mathbb{N}_0} \lambda_k N^{-k} \int_0^\infty P_{2s}(0, B_k(0))\,\mathrm{d}s, \tag{1.63}$$

where $B_k(0)$ is the k-block in Ω_N around 0 (recall (1.23)) and $P_t(0,B_k(0)) \equiv \sum_{\zeta\in B_k(0)} P_t(0,\eta)$. We will see in Section 2.4.2 that $\bar{H}_N$ in (1.63) is the *expected hazard* for two partition elements in the spatial $\underline{\Lambda}$-coalescent with non-local coalescence to coalesce. Note in particular that the second summand in (1.62) is induced by the reshuffling in the spatial $\underline{\Lambda}$-coalescent with non-local coalescence.

Our next three main theorems identify the ergodic behaviour for finite N.

Theorem 1.8. **[Dichotomy for finite N]**
The following dichotomy holds for every $N \in \mathbb{N}\backslash\{1\}$ fixed:

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

$$\liminf_{t\to\infty} \sup_{\psi\in B_1} \mathbb{E}_{X_\eta^{(\Omega_N)}(t)}[\mathrm{Var.}(\psi)] > 0, \quad \textit{for all } \eta \in \Omega_N. \tag{1.64}$$

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

$$\lim_{t\to\infty} \sup_{\psi\in B_1} \mathbb{E}_{X_\eta^{(\Omega_N)}(t)}[\mathrm{Var.}(\psi)] = 0, \quad \textit{for all } \eta \in \Omega_N. \tag{1.65}$$

The proof of Theorem 1.8 is given in Section 10.

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

Theorem 1.9. **[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 Ω_N and has a single-site mean θ,*

$$\mathcal{L}\left[X^{(\Omega_N)}(t)\right] \underset{t\to\infty}{\Longrightarrow} \nu_\theta^{(\Omega_N),\underline{c},\underline{\lambda}} \in \mathcal{P}(\mathcal{P}(E)^{\Omega_N}) \tag{1.66}$$

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

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

$$\mathcal{L}\left[X^{(\Omega_N)}(t)\right] \underset{t\to\infty}{\Longrightarrow} \int_0^1 \theta(\mathrm{d}u)\delta_{(\delta_u)^{\Omega_N}} \in \mathcal{P}(\mathcal{P}(E)^{\Omega_N}). \tag{1.67}$$

The proof of Theorem 1.9 is given in Section 10.

Theorem 1.10. [Agreement of dichotomy for $N < \infty$ and $N = \infty$]
Under the weak regularity condition

$$\textit{either} \quad \limsup_{k\to\infty} \frac{\lambda_{k+1}}{c_k} < \infty \quad \textit{or} \quad \liminf_{k\to\infty} \left(\frac{\lambda_{k+1}}{c_k} \wedge \frac{\lambda_k}{\lambda_{k+1}}\right) > 0, \tag{1.68}$$

the dichotomies in Theorems 1.7 and 1.9 coincide i.e., $\sum_{k\in\mathbb{N}_0} m_k = \infty$ if and only if $\bar{H}_N = \infty$.

The proof of Theorem 1.10 is given in Section 11.1.

1.5.4. *Comparison with the dichotomy for the hierarchical Fleming-Viot process.* We return to the case $N = \infty$. For the classical case of hierarchically interacting Fleming-Viot diffusions (i.e., in the absence of non-local reshuffling-resampling), the dichotomy was analysed in Dawson et al. (1995). It was shown there that the dichotomy in (1.58) reduces to

$$\sum_{k\in\mathbb{N}_0} (1/c_k) = \infty \quad \text{vs.} \quad \sum_{k\in\mathbb{N}_0} (1/c_k) < \infty, \tag{1.69}$$

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.

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

$$\underline{d}^* = (d_k^*)_{k\in\mathbb{N}_0} \tag{1.70}$$

be the sequence of volatility constants when $\mu_0 > 0$ and $\mu_k = 0$ for all $k \in \mathbb{N}$ ($\mu_k = \frac{1}{2}\lambda_k$, see (1.57)), i.e., there is resampling in single colonies but not in macro-colonies. By (1.45), this sequence has initial value $d_0^* = 0$ and satisfies the recursion relation

$$d_1^* = d_1 = \frac{c_0\mu_0}{c_0 + \mu_0}, \qquad \frac{1}{d_{k+1}^*} = \frac{1}{c_k} + \frac{1}{d_k^*}, \quad k \in \mathbb{N}, \tag{1.71}$$

whose solution is

$$d_k^* = \frac{\mu_0}{1 + \mu_0\sigma_k}, \quad k \in \mathbb{N}, \quad \text{with } \sigma_k = \sum_{l=0}^{k-1} \frac{1}{c_l}. \tag{1.72}$$

Theorem 1.11. [Comparison with hierarchical Fleming-Viot]
The following hold for $(d_k)_{k\in\mathbb{N}_0}$ as in (1.45) (also recall (1.57)):

(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 $\lim_{k\to\infty} \sigma_k d_k = 1$.*

The proof of Theorem 1.11 is given in Section 11.1.

In words, (a) and (b) say that both migration and reshuffling-resampling increase volatility (recall ((1.57)–1.58)), (c) says that the dichotomy in (1.69) 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 the last paragraph of 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 1.5.5–1.5.6 for polynomial, respectively, exponential growth of the coefficients c_k and λ_k.

Heuristics. The recursion relation in (1.45) has the shape $d_{k+1}=f_k(d_k)$ with $f_k\colon\mathbb{R}\to\mathbb{R}$ a Möbius-transformation (see Section 11.2). Thus, to obtain the asymptotics of d_k as $k\to\infty$ we must study *inhomogeneous iterates of Möbius-transformations.* For each $k\in\mathbb{N}$, f_k is hyperbolic with two fixed points: a repulsive fixed point $x_k^-<0$ and an attractive fixed point $x_k^+>0$. Depending on the scaling of the coefficients c_k and λ_k, the scaling of x_k^+ exhibits four regimes. For three of the regimes, it turns out that $d_k\sim x_k^+$ as $k\to\infty$, i.e., the iterates of the Möbius-transformations attract towards the fixed point of the last one. The fourth regime is different. In Section 1.5.5 we deal with polynomial coefficients, in Section 1.5.6 with exponential coefficients. In order to obtains sharp results, the coefficients c_k and λ_k must satisfy certain *regularity conditions.*

1.5.5. *Scaling in the clustering regime: polynomial coefficients.* The following 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 λ_k. The necessary *regularity conditions* are stated in (1.78–1.81) below.

Define

$$\lim_{k\to\infty}\frac{\mu_k}{c_k}=K\in[0,\infty]\text{ and, if } K=0,\text{ also }\lim_{k\to\infty}k^2\frac{\mu_k}{c_k}=L\in[0,\infty]. \tag{1.73}$$

Theorem 1.12. [Scaling of the volatility in the clustering regime: polynomial coefficients]
Assume that the regularity conditions (1.78–1.81) *hold.*

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

$$\lim_{k\to\infty}\frac{d_k}{c_k}=1. \tag{1.74}$$

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

$$\lim_{k\to\infty}\frac{d_k}{c_k}=M\ \text{ with } M=\tfrac12 K\left[-1+\sqrt{1+(4/K)}\right]\in(0,1). \tag{1.75}$$

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

$$\lim_{k\to\infty} \frac{d_k}{\sqrt{c_k \mu_k}} = 1. \tag{1.76}$$

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

$$\lim_{k\to\infty} \sigma_k d_k = M^* \text{ with } M^* = \tfrac{1}{2}\left[1 + \sqrt{1 + 4L/(1-a)^2}\right] \in [1, \infty). \tag{1.77}$$

The proof of Theorem 1.12 in given Section 11.3. The meaning of the four regimes for the evolution of the population will be explained in Corollary 1.13. Case (a) can be termed "reshuffling-resampling dominated", cases (c) and (d) "migration dominated", and case (b) "balanced".

Regularity conditions. In Theorem 1.12, we need to impose some mild regularity conditions on $\underline{c}$ and $\underline{\mu}$, which we collect in (1.78–1.81) below. We require that both c_k and μ_k are *regularly varying at infinity*, i.e., there exist $a, b \in \mathbb{R}$ such that

$$c_k \sim L_c(k) k^a, \qquad \mu_k \sim L_\mu(k) k^b, \qquad k \to \infty, \tag{1.78}$$

with L_c, L_μ slowly varying at infinity (Bingham et al. (1987, Section 1.9)). The numbers a, b are referred to as the *indices* of $\underline{c}$ and $\underline{\mu}$[14]. Note that (1.68) is satisfied.

To handle the boundary cases, where c_k, μ_k, μ_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*:

$$\begin{aligned} a = 0: &\quad k \mapsto \Delta L_c(k)/L_c(k),\ k \mapsto k\Delta L_c(k)/L_c(k), \\ b = 0: &\quad k \mapsto \Delta L_\mu(k)/L_\mu(k),\ k \mapsto k\Delta L_\mu(k)/L_\mu(k), \end{aligned} \tag{1.79}$$

and the following functions are *bounded*:

$$\begin{aligned} a = 0: &\quad k \mapsto k\Delta L_c(k)/L_c(k), \\ b = 0: &\quad k \mapsto k\Delta L_\mu(k)/L_\mu(k), \end{aligned} \tag{1.80}$$

where $\Delta L(k) = L(k+1) - L(k)$. To ensure the *existence* of the limits in (1.73), we also need the following functions to be *asymptotically monotone*:

$$\begin{aligned} a = b: &\quad k \mapsto L_\mu(k)/L_c(k), \\ a = b - 2: &\quad k \mapsto k^2 L_\mu(k)/L_c(k). \end{aligned} \tag{1.81}$$

Scaling of the variance. The next corollary shows what the scaling of d_k in Theorem 1.12 implies for the scaling of m_k and hence of the variance in (1.59) (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 1.13. **[Scaling behaviour of m_k]** *The following asymptotics of* m_k *for* $k \to \infty$ *holds in the four cases of* Theorem 1.12*:*

$$\begin{aligned} &\text{(a)} \quad m_k \sim \frac{\mu_k}{c_k} \to \infty, \quad \text{(b)} \quad m_k \to K + M, \\ &\text{(c)} \quad m_k \sim \sqrt{\frac{\mu_k}{c_k}} \to 0, \quad \text{(d)} \quad m_k \sim \frac{M^*}{c_k \sigma_k} \to 0. \end{aligned} \tag{1.82}$$

[14] 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.

All four cases fall in the clustering regime. For the variance in (1.59) 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

$$\lim_{k\to\infty} \sigma_k = \infty \quad \Longleftrightarrow \quad \sum_{k\in\mathbb{N}} \frac{1}{c_k\sigma_k} = \infty. \tag{1.83}$$

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

- 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 (1.73), while $c_k\sigma_k \sim k/(1-a)$ as $k \to \infty$ when $a \in (-\infty, 1)$ by (1.78). Hence, Case (d) connects up with Case (c) as well.

1.5.6. *Scaling in the clustering regime: exponential coefficients.* We briefly indicate how Theorem 1.12 extends when c_k and μ_k satisfy

$$\begin{aligned} &c_k = c^k\bar{c}_k,\ \mu_k = \mu^k\bar{\mu}_k \text{ with } c,\mu \in (0,\infty) \text{ and} \\ &\qquad\qquad\qquad (\bar{c}_k),\ (\bar{\mu}_k) \text{ regularly varying at infinity,} \\ &\bar{K} = \lim_{k\to\infty} \frac{\bar{\mu}_k}{\bar{c}_k} \in [0,\infty], \end{aligned} \tag{1.84}$$

and the analogues of (1.79–1.81) apply to the regularly varying parts. Again, note that (1.68) is satisfied.

Theorem 1.14. [Scaling of the volatility in the clustering regime: exponential coefficients]
Assume that (1.84) *holds. Recall the cases* (a–d) *from* Theorem *1.12. Then:*

(A) **[scaling like Case (a)]** *$c < \mu$ or $c = \mu$, $\bar{K} = \infty$: $\lim_{k\to\infty} d_k/c_k = 1/c$.*
(B) **[scaling like Case (b)]** *$c = \mu$, $\bar{K} \in (0,\infty)$: $\lim_{k\to\infty} d_k/c_k = \bar{M}$ with*

$$\bar{M} = \frac{1}{2c}\left[-(c(\bar{K}+1)-1) + \sqrt{(c(\bar{K}+1)-1)^2 + 4c\bar{K}}\right]. \tag{1.85}$$

(C) *The remainder $c > \mu$ or $c = \mu$, $\bar{K} = 0$ splits into three cases:*
 (C1) **[scaling 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) **[scaling like Case (b)]** *$c = \mu < 1$, $\bar{K} = 0$: $\lim_{k\to\infty} d_k/c_k = (1-c)/c$.*
 (C3) **[scaling like Case (c)]** *$c = \mu > 1$, $\bar{K} = 0$: $\lim_{k\to\infty} d_k/\mu_k = 1/(\mu-1)$.*

Remark 1.15. The analogue of L (cf., (1.73) and Theorem 1.12) no longer plays a role for exponential coefficients (cf., Theorem 1.14).

The proof of Theorem 1.14 is given in Section 11.4. The choices $1 = c > \mu$, $\lim_{k\to\infty} \sigma_k < \infty$ and $c > 1$, $c > \mu$ correspond to local coexistence (and so does $c = \mu > 1$, $\bar{K} = 0$, $\sum_{k\in\mathbb{N}_0} \bar{\mu}_k/\bar{c}_k < \infty$).

1.5.7. *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 fast do the clusters grow? To that end, we look at the interaction chain $M^{(j)}_{-k(j)}$ for $j \to \infty$, where the *level scaling* function $k\colon \mathbb{N} \to \mathbb{N}$ with $\lim_{j\to\infty} k(j) = \infty$ is suitably chosen such that we obtain a *nontrivial clustering limiting law*, i.e.,

$$\lim_{j\to\infty} \mathcal{L}\left[M^{(j)}_{-k(j)}\right] = \mathcal{L}\left[\hat{\theta}\right], \tag{1.86}$$

where the limiting random measure $\hat{\theta}$ satisfies

$$\mathbb{P}\{\hat{\theta} = \delta_U,\ \text{for some } U \in E\} < 1. \tag{1.87}$$

For example, in Dawson and Greven (1993a) such a question was answered in the case of the interacting Fleming-Viot processes with critically recurrent migration $\underline{c}$. There, different types of limit laws and different types of scaling can occur, corresponding to different clustering regimes. Following Dawson et al. (1995) and Dawson and Greven (1996), it is natural to consider a whole family of scalings $k_\alpha\colon \mathbb{N} \to \mathbb{N}$, $\alpha \in [0,1)$ satisfying (1.86). We single out *fast*, *diffusive* and *slow* clustering regimes:

(i) **Fast clustering:** $\lim_{j\to\infty} k_\alpha(j)/j = 1$ for all α.
(ii) **Diffusive clustering:** In this regime, $\lim_{j\to\infty} k_\alpha(j)/j = \kappa(\alpha)$ for all α, 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 α. 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 (1986), 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 $\mathbb{Z}^1$, where the clusters occur only on scales $\chi \cdot t^{1/2}$, where χ is random, see Arratia (1979). For the model of hierarchically interacting Fleming-Viot diffusions with $c_k \equiv 1$ (= critically recurrent migration), Fleischmann and Greven (1994) showed that, for all $N \in \mathbb{N} \setminus \{1\}$ and all $\eta \in \Omega_N$,

$$\mathcal{L}\left[\left(Y^{(\Omega_N)}_{\eta,\lfloor(1-\alpha)t\rfloor}(N^t)\right)_{\alpha\in[0,1)}\right] \underset{t\to\infty}{\overset{\text{f.d.d.}}{\Longrightarrow}} \mathcal{L}\left[\left(Y\left(\log\left(\frac{1}{1-\alpha}\right)\right)\right)_{\alpha\in[0,1)}\right], \tag{1.88}$$

where $(Y(t))_{t\in[0,\infty)}$ is the standard Fleming-Viot diffusion on $\mathcal{P}(E)$. A similar behaviour occurs for other models, e.g., for branching models (Dawson and Greven (1996)).

Our last 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 1.12–1.14. 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 1.16. [Clustering regimes for polynomial coefficients]
Recall the scaling regimes of Theorem 1.12.

(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.,*

$$\mathcal{L}\left[\left(M^{(j)}_{-\lfloor(1-\alpha)j\rfloor}\right)_{\alpha\in[0,1)}\right] \underset{j\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(Z^{0,1,0}_{\theta}\left(\log\left(\frac{1}{1-\alpha^R}\right)\right)\right)_{\alpha\in[0,1)}\right], \qquad (1.89)$$

where $R = M^*(1-a)$ *with* M^* *defined in* (1.77) *and* a *the exponent in* (1.78)*.*

Theorem 1.17. [Clustering regimes for exponential coefficients]
Recall the scaling regimes of Theorem 1.14*.*

(i) **[Fast clustering]** *In cases* (A, B, C1, C2)*, and case* (C3) *with*

$$\lim_{k\to\infty} k\bar{\mu}_k/\bar{c}_k = \infty, \qquad (1.90)$$

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.,* (1.89) *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.*

The proofs of Theorems 1.16–1.17 are given in Section 9.3. Note that (1.88) is a statement valid for *all* $N \in \mathbb{N}\setminus\{1\}$. In contrast, Theorems 1.16–1.17 are valid in the hierarchical mean-field limit $N\to\infty$ only.

1.6. *Discussion.* **Summary.** We have constructed the $C_N^{\underline{c},\underline{\Lambda}}$-process in Section 1.4.4, describing hierarchically interacting Cannings processes, and have identified its space-time scaling behaviour in the hierarchical mean field limit $N\to\infty$ (interaction chain, cf. Theorem 1.6). We have fully classified the clustering vs. local coexistence dichotomy in terms of the parameters $\underline{c},\underline{\Lambda}$ of the model (cf. Theorem 1.7), and found different regimes of cluster formation (cf. Theorems 1.16, 1.17). Moreover, we have verified the dichotomy also for finite N (cf. Theorems 1.8–1.10). Our results provide a full generalisation of what was known for hierarchically interacting diffusions, and show that Cannings resampling leads to new phenomena (cf. Theorem 1.11 and comment following it).

Diverging volatility of the Fleming-Viot part and local coexistence. The growth of the block resampling rates $(\lambda_k)_{k\in\mathbb{N}}$ 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 (recall Theorem 1.7(c)). 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 $c_k = k(\log k)^3$ 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 Section 1.5 and suggest that a dichotomy between clustering and local coexistence also holds for a suitably defined Cannings model with non-local resampling on $\mathbb{Z}^d$, $d\ge 3$. In addition, a continuum limit to the geographic space $\mathbb{R}^2$ ought to arise as well, cf. Barton et al. (2010). The latter may be easier to investigate in the limit $N\to\infty$, following the approach outlined in

Greven (2005). 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, Greven et al. (2014) for the model without multi-colony Cannings resampling (i.e., $\Lambda_k = \delta_0$ for $k \in \mathbb{N}$).

Outline of the remainder of the paper. 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 1.3–1.5. Here is a *roadmap*:

- Section 3 handles all issues related to the well-posedness of martingale problems. The proofs of Propositions 1.1–1.4 are in Section 3.2.
- Section 4 deals with the properties of the McKean-Vlasov process, including its equilibrium distribution.
- 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 as follows: Theorem 1.5 is proved in Section 8 with preparatory work being done in Sections 6–7, Theorem 1.6 is proved in Section 9.1, Theorem 1.7 in Section 9.2, and Theorems 1.16–1.17 in Section 9.3.
- Section 10 proves the scaling results for the interaction chain stated in Theorems 1.8 and 1.9.
- Section 11 derives the scaling results for the volatility constant: Theorems 1.10 and 1.11 are proved in Section 11.1, Möbius-transformations are introduced in Section 11.2, Theorem 1.12 is proved in Section 11.3, and Theorem 1.14 in Section 11.4.
- Section 12 collects the notation.

2. Spatial Λ-coalescent with non-local 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 (2006), which allows for the coalescence of the partition elements (= families = lineages) residing at the same location only. Informally, the spatial $\underline{\Lambda}$-coalescent with non-local coalescence is the process that encodes the *family structure* of a sample from the currently alive population in the $C_N^{\underline{c},\underline{\Lambda}}$-process, i.e., it is the process of *coalescing lineages* that occur when the evolution of the spatial $C_N^{\underline{c},\underline{\Lambda}}$-Cannings process is traced backwards in time up to a common ancestor. In what follows, we denote this backwards-in-time process by $\mathfrak{C}_N^{\underline{c},\underline{\Lambda}}$.

Recall that 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

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

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 (1985, 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 Ω_N. The duality will be crucial for the proof of Propositions 1.1–1.4 (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 with local coalescence. In Section 2.2, we add non-local coalescence. In Section 2.3, we formulate and prove the duality relation between the $C_N^{\underline{c},\underline{\Lambda}}$-process and the spatial Λ-coalescent with non-local coalescence. In Section 2.4, we look at the long-time behaviour of the spatial Λ-coalescent with non-local coalescence.

2.1. *Spatial Λ-coalescent with local coalescence.* In this section, we briefly recall the definition of the spatial Λ-coalescent on a countable geographic space G as introduced by Limic and Sturm (2006). (For a general discussion of exchangeable coalescents, see Berestycki (2009).) Here, we do not need assumption (1.2) on measure Λ. In Section 2.2, we will add *non-local 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:

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

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

$$[n] = \{1, \dots, n\} \tag{2.3}$$

and the set Π_n of its partitions into *families*:

$$\begin{aligned} \Pi_n = {} & \text{set of all partitions } \pi = \{\pi_i \subset [n]\}_{i=1}^b \\ & \text{of set } [n] \text{ into disjoint families } \pi_i,\ i \in [b]. \end{aligned} \tag{2.4}$$

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

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

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

Remark 2.1 (Notation). 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.

Abbreviate

$$\pi^{-1}(k) = \min\{i \in [n] : \pi(i) = k\}, \quad k \in [b]. \tag{2.6}$$

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

$$\begin{aligned} \Pi_{G,n} = \Big\{ &\pi_G = \{(\pi_1, g_1), (\pi_2, g_2), \ldots, (\pi_b, g_b)\} \\ &: \{\pi_1, \ldots, \pi_b\} \in \Pi_n,\, g_1, \ldots, g_b \in G, b \in [n] \Big\}. \end{aligned} \tag{2.7}$$

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.,

$$S_{G,n} = \Big\{ (\{1\}, g_1), (\{2\}, g_2), \ldots, (\{n\}, g_n) : g_i \in G, i \in [n] \Big\}. \tag{2.8}$$

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

$$\pi_G = \{(\pi_1, g_1), (\pi_2, g_2), \ldots, (\pi_b, g_b)\} \tag{2.9}$$

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

$$L(\pi_G) = \{g_1, \ldots, g_b\} \subset G. \tag{2.10}$$

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

$$\Pi = \{\text{partitions of } \mathbb{N}\}, \quad \Pi_G = \{G\text{-labelled partitions of } \mathbb{N}\}, \tag{2.11}$$

and introduce the set of standard initial states

$$S_G = \big\{ \{(\{i\}, g_i)\}_{i \in \mathbb{N}} : g_i \in G, i \in \mathbb{N} \big\}. \tag{2.12}$$

Equip Π_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 Π_G Polish, too.

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

$$a^* : G^2 \to \mathbb{R}, \quad a^*(g, f) = a(f, g), \tag{2.13}$$

where $a(\cdot, \cdot)$ is the migration kernel of the corresponding C^Λ-process with geographic space G as in (2.2). The spatial n-Λ-*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 π_G with label g *coalesce* according to the mechanism of the non-spatial n-Λ-coalescent. In other words, given that in the current state of the spatial Λ-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

$$\lambda_{b,i}^{(\Lambda)} = \int_{[0,1]} \Lambda^*(\mathrm{d}r) r^i (1-r)^{b-i}, \quad i \in [2, b] \cap \mathbb{N}, \tag{2.14}$$

with Λ^* 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 π_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-Λ-coalescent via Poisson point processes (see also Limic and Sturm (2006)).

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

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

where $\omega = (\omega_i)_{i\in\mathbb{N}} \subset \{0,1\}^{\mathbb{N}}$. We assume that point processes $\mathfrak{P}$ are adapted to filtration $(\mathcal{F}_t)_{t\geq 0}$.

Note that the second factor of the intensity measure in (2.15) is not a product measure on $[0,1] \times \{0,1\}^{\mathbb{N}}$, in particular, it is not the same as

$$\left[\Lambda^*(\mathrm{d}r)(r\delta_1 + (1-r)\delta_0)\right]^{\otimes\mathbb{N}}(\mathrm{d}\omega). \tag{2.16}$$

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:

$$\mathrm{coal}_{J,g}(\pi_{G,n}) = \left(\bigcup_{i\in J\cap[b(\pi)]} \pi_i, g\right) \cup \left(\pi_{G,n} \setminus \bigcup_{i\in J\cap[b(\pi)]} (\pi_i, g_i)\right), \quad \pi_{G,n} \in \Pi_{G,n}. \tag{2.17}$$

Using $\mathfrak{P}$, we construct the standard spatial n-Λ-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, ω) 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.,
$$\mathfrak{C}_n^{(G),\mathrm{loc}}(t) = \mathrm{coal}_{\{i\in[n]\colon\, \omega_i=1, g_i(t-)=g\},g}(\mathfrak{C}_n^{(G),\mathrm{loc}}(t-)). \tag{2.18}$$
- *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^*[15].

In what follows, we denote by $\cdot|_n\colon \Pi_{G,m} \to \Pi_{G,n}$, for $m \geq n$, (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-Λ-coalescent satisfies the following consistency property:

$$\mathcal{L}\left[\mathfrak{C}_m^{(G),\mathrm{loc}}|_n\right] = \mathcal{L}\left[\mathfrak{C}_n^{(G),\mathrm{loc}}\right], \qquad n, m \in \mathbb{N},\ n \leq m. \tag{2.19}$$

[15] The adjective "between" is well defined because the set of points (t, r, ω) 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).

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

$$\mathfrak{C}^{(G),\mathrm{loc}} = (\mathfrak{C}^{(G),\mathrm{loc}}(t) \in \Pi_G)_{t \geq 0} \tag{2.20}$$

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

Definition 2.2 (Limic and Sturm (2006))**.** Call the process $\mathfrak{C}^{(G),\mathrm{loc}}$ the spatial Λ-coalescent corresponding to the migration rates a^* and the coalescence measure Λ.

2.1.2. *Martingale problem.* In this section, we characterise the spatial Λ-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

$$F_n\colon \Pi_{G,n} \to \mathbb{R} \tag{2.21}$$

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 Π_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

$$\mathrm{mig}_{f\to g,i}(\pi_{G,n}) = \begin{cases} (\pi_i, g) \cup (\pi_{G,n} \setminus (\pi_i, f)), & (\pi_i, f) \in \pi_{G,n}, \\ \pi_{G,n}, & (\pi_i, f) \notin \pi_{G,n}, \end{cases} \quad \pi_{G,n} \in \Pi_{G,n}, \tag{2.22}$$

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

Consider the linear operator $L^{(G)*}$ defined as

$$L^{(G)*} = L^{(G)*}_{\mathrm{mig}} + L^{(G)*}_{\mathrm{coal}}, \tag{2.23}$$

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

$$(L^{(G)*}_{\mathrm{mig}} F)(\pi_G) = \sum_{i=1}^{b(\pi_G|_n)} \sum_{g,f\in G} a^*(g,f)\big[F_n(\mathrm{mig}_{g\to f,i}(\pi_G|_n)) - F(\pi_G)\big], \tag{2.24}$$

$$(L^{(G)*}_{\mathrm{coal}} F)(\pi_G) = \sum_{g\in G} \sum_{\substack{J\subset\{i\in[n]\colon g_i = g\},\\ |J|\geq 2}} \lambda^{(\Lambda)}_{b(\pi_G|_n,g),|J|}\big[F_n(\mathrm{coal}_{J,g}(\pi_G|_n)) - F(\pi_G)\big] \tag{2.25}$$

(recall definitions (2.5), (2.13), (2.14) and (2.17)).

Proposition 2.3. [Martingale problem for the spatial Λ-coalescent with local coalescence] *The spatial Λ-coalescent with local coalescence defined in Section 2.1.1 solves the well-posed martingale problem for $(L^{(G)*}, C_{\mathrm{b}}(\Pi_G), \delta_{S_G})$ with S_G as in (2.12).*

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 Π_G.

2.1.3. *Mean-field and immigration-emigration* Λ*-coalescents.* Some special spatial Λ-coalescents will be needed in the course of our analysis of the hierarchically interacting Cannings process. We define the *mean-field* Λ*-coalescent* as the spatial Λ-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 Λ*-coalescent with immigration-emigration* as the spatial Λ-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 non-local coalescence.* In this section, we construct a new type of spatial coalescent process based on a sequence $\underline{\Lambda} = (\Lambda_k)_{k\in\mathbb{N}_0}$ of finite measures on $[0, 1]$ as in (1.28), namely, the *spatial* $\underline{\Lambda}$*-coalescent* on $G = \Omega_N$ *with non-local 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 *non-local* Λ*-coalescence* of partition elements in a ball of radius k.

In this section, we assume that, for all $k \in \mathbb{N}$, measure Λ_k satisfy (1.2). But we do not assume that measure Λ_0 satisfies (1.2). Denote

$$d_0 = \Lambda_0\{0\}. \tag{2.26}$$

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.* In what follows, we consider $G = \Omega_N$. We start by extending the graphical construction from Section 2.1.1 to incorporate the additional transition mechanisms of non-local reshuffling and coalescence.

Given the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq0}, \mathbb{P})$, consider Poisson point processes $\mathfrak{P}^{(\Omega_N)}$ on

$$[0, \infty) \times \Omega_N \times \mathbb{N}_0 \times [0, 1] \times \{0, 1\}^{\mathbb{N}} \tag{2.27}$$

having intensity measure

$$\mathrm{d}t \otimes \mathrm{d}\eta \otimes \Big(N^{-2k}\mathrm{d}k \left[\Lambda_k^*(\mathrm{d}r)\big(r\delta_1 + (1-r)\delta_0\big)^{\otimes\mathbb{N}}\right](\mathrm{d}\omega)\Big), \tag{2.28}$$

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 Ω_N. Again, note that the third factor in (2.28) is not a product measure (compare (2.16)).

Given $\Sigma \Subset \Omega_N$ (i.e., Σ is a finite subset of Ω_N) and $\xi = \{\xi_i\}_{i=1}^{|\Sigma|}, \xi_i \in \Sigma$, 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 $\xi_i \in \Sigma$:

$$\mathrm{resh}_{\Sigma,\xi}(\pi_{\Omega_N})_i = \begin{cases} (\pi_i, \eta_i), & \eta_i \notin \Sigma, \\ (\pi_i, \xi_i), & \eta_i \in \Sigma, \end{cases} \qquad \pi_{\Omega_N} \in \Pi_{\Omega_N},\, i \in [b(\pi_{\Omega_N})]. \tag{2.29}$$

Let

$$U_\Sigma = \{U_\Sigma(\xi)\}_{\xi\in\Sigma} \tag{2.30}$$

be a collection of independent Σ-valued random variables uniformly distributed on Σ. We construct the *standard spatial* n*-*Λ*-coalescent with non-local 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 $\mathfrak{C}_n^{(\Omega_N)}(0) \in S_{\Omega_N,n}$ (recall (2.8)).
- *Coalescence with reshuffling.* For each point (t, η, k, r, ω) of the Poisson point process $\mathfrak{P}^{(\Omega_N)}$ (cf. (2.27)–(2.28)), 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 η. 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 (recall (2.17), (2.29)–(2.30)):

$$\mathfrak{C}_n^{(\Omega_N)}(t) = \mathrm{resh}_{B_k(\eta), U_{B_k(\eta)}} \circ \mathrm{coal}_{\{i\in[n]:\ \omega_i=1, \eta_i(t-)\in B_k(\eta)\},\eta}(\mathfrak{C}_n^{(\Omega_N)}(t-)). \tag{2.31}$$

  Note that, in contrast with the spatial coalescent with local coalescence from Section 2.1, the coalescence mechanism in (2.31) 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 (1.26) and (2.13)).

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

$$\mathfrak{C}^{(\Omega_N)} \tag{2.32}$$

taking values in Π_{Ω_N}.

Definition 2.5. The process $\mathfrak{C}^{(\Omega_N)}$ is called the spatial $\underline{\Lambda}$-coalescent with non-local coalescence corresponding to the resampling measures $(\Lambda_k)_{k\in\mathbb{N}_0}$ (recall (1.28)) and the migration coefficients $(c_k)_{k\in\mathbb{N}_0}$ (recall (1.25)).

Proposition 2.6. [Feller property] *The process $\mathfrak{C}^{(\Omega_N)}$ is a cádlág 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 $\underline{\Lambda}$-coalescent with non-local coalescence as the solution of the corresponding martingale problem.

Given $\pi_{\Omega_N,n} \in \Pi_{\Omega_N,n}$ and $\eta \in \Omega_N$, denote the number of families of $\pi_{\Omega_N,n}$ with labels in $B_k(\eta)$ (recall (1.23)) by

$$b(\eta) = b(\pi_{\Omega_N,n}, B_k(\eta)) = |\{(\pi_i, \eta_i) \in \pi_{\Omega_N,n}:\ \eta_i \in B_k(\eta)\}| \in \mathbb{N}. \tag{2.33}$$

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 $F(\cdot) = F_n(\cdot|_n)$ (recall (2.21)). Consider the linear operator $L^{(\Omega_N)*}$ defined as

$$L^{(\Omega_N)*} = L_{\mathrm{mig}}^{(\Omega_N)*} + L_{\mathrm{coal}}^{(\Omega_N)*}, \tag{2.34}$$

where the linear operators $L_{\mathrm{mig}}^{(\Omega_N)*}$ and $L_{\mathrm{coal}}^{(\Omega_N)*}$ are defined as follows (recall (2.21)). The migration operator is[16]

$$\left(L_{\mathrm{mig}}^{(\Omega_N)*} F\right)(\pi_{\Omega_N}) = \sum_{i=1}^{b(\pi_{\Omega_N}|_n)} \sum_{\eta,\zeta\in\Omega_N} a^{(N)*}(\eta,\zeta) \left[F_n(\mathrm{mig}_{\eta\to\zeta,i}(\pi_{\Omega_N}|_n)) - F(\pi_{\Omega_N})\right], \tag{2.35}$$

[16]Note that $a^{(N)} = a^{(N)*}$ for the hierarchical random walk (cf. (2.24)).

and the block-coalescence-reshuffling operator is (recall (2.14), (2.17), (2.29) and (2.33))

$$\Big(L_{\mathrm{coal}}^{(\Omega_N)*}F\Big)(\pi_{\Omega_N}) = \sum_{\eta\in\Omega_N}\sum_{k\in\mathbb{N}_0} N^{-2k} \sum_{\xi_1\in B_k(\eta)} N^{-k} \sum_{\xi_2\in B_k(\eta)} N^{-k}\cdots \sum_{\xi_{|B_k(\eta)|}\in B_k(\eta)} N^{-k}$$
$$\times \sum_{\substack{J\subset[b(\eta)],\\ |J|\geq 2}} \lambda_{b(\eta),|J|}^{(\Lambda_k)} \left[F_n(\mathrm{resh}_{B_k(\eta),\xi}\circ \mathrm{coal}_{\{i\in J:\ \eta_i\in B_k(\eta)\},\eta}(\pi_{\Omega_N}|_n)) - F(\pi_{\Omega_N})\right]. \tag{2.36}$$

Proposition 2.7. [Martingale problem: Spatial $\underline{\Lambda}$-coalescent with non-local coalescence] *The spatial $\underline{\Lambda}$-coalescent with non-local coalescence $\mathfrak{C}^{(\Omega_N)}$ defined in Section 2.2.1 solves the well-posed martingale problem $(L^{(\Omega_N)*},\mathcal{C}_{\Omega_N},\delta_{S_{\Omega_N}})$ with S_{Ω_N} as in (2.12).*

Proof: A straightforward inspection of the graphical construction in Section 2.2.1 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 (1981)). □

Remark 2.8. Note that, instead of the singleton initial condition in Proposition 2.7 (and in the graphical construction of Section 2.2.1), we can use any other initial condition in Π_{Ω_N}.

2.3. *Duality relations.* We next formulate and prove the duality relation between the $C_N^{\underline{c},\underline{\Lambda}}$-process from Section 1.4.4 and the spatial $\underline{\Lambda}$-coalescent with non-local coalescence $\mathfrak{C}^{(\Omega_N)}$ 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

$$H_\varphi^{(n)}(x,\pi_{G,n}) = \int_{E^b}\left(\bigotimes_{i=1}^{b} x_{\eta_{\pi^{-1}(i)}}(\mathrm{d}u_i)\right)\varphi\big(u_{\pi(1)},u_{\pi(2)},\ldots,u_{\pi(n)}\big), \tag{2.37}$$

where $x=(x_\eta)_{\eta\in G}\in\mathcal{P}(E)^G$, $\pi_{G,n}\in\Pi_{G,n}$, $b=b(\pi_{G,n})=|\pi_{G,n}|$ (cf. (2.5)), $(\eta_i)_{i\in[b]}=L(\pi_{G,n})$ (cf. (2.10)) 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 (2.37) assign the same type to individuals that belong to the same family. Note that these functions form a family of functions on $\mathcal{P}(E)^G$,

$$\left\{H_\varphi^{(n)}(\cdot,\pi_{G,n})\colon\ \mathcal{P}(E)^G\to\mathbb{R}\mid \pi_{G,n}\in\Pi_{G,n}, n\in\mathbb{N},\varphi\in C_{\mathrm{b}}(E^n,\mathbb{R})\right\}, \tag{2.38}$$

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

$$H(x,(\varphi,\pi_{G,n})) = H_\varphi^{(n)}(x,\pi_{G,n}),\qquad x\in\mathcal{E}=\mathcal{P}(E)^G,\ (\varphi,\pi_{G,n})\in\mathcal{E}', \tag{2.39}$$

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

We proceed with the following observation. Recall the definition of duality in the paragraph including (2.1).

Remark 2.9. (a) 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

$$[L_X(H(\cdot,Y_0))](X_0) = [L_Y(H(X_0,\cdot))](Y_0), \quad \text{for all } (X_0,Y_0)\in\mathcal{E}\times\mathcal{E}', \tag{2.40}$$

is equivalent to the duality relation (2.1) (see, e.g., Ethier and Kurtz (1986, Section 4.4)).

(b) Item (a) gives the duality function $H(\cdot,\cdot)$ for all $t\geq 0$ and $n\in\mathbb{N}$, as is proved in Proposition 2.10 below. In particular, the following holds

$$\mathbb{E}\left[H_\varphi^{(n)}(X^{(G)}(t),\mathfrak{C}^{(G)}(0)|_n)\right] = \mathbb{E}\left[H_\varphi^{(n)}(X^{(G)}(0),\mathfrak{C}^{(G)}(t)|_n)\right], \tag{2.41}$$

with $X^{(G)}$ as below Proposition 1.4 and $\mathfrak{C}^{(G)}$ as in Definition 2.5.

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

Proposition 2.10. [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 $n\in\mathbb{N}$, for all $H_\varphi^{(n)}$ as in (2.37), all $x\in\mathcal{P}(E)^G$, and all $\pi_G\in\Pi_G$,*

$$\left(L^{(G)}H_\varphi^{(n)}(\cdot,\pi_G|_n)\right)(x) = \left(L^{(G)*}H_\varphi^{(n)}(x,\cdot|_n)\right)(\pi_G). \tag{2.42}$$

Proof: We check the statement for $G=\Omega_N$. In this case, $L^{(G)}$ is as in (1.35) and $L^{(G)*}$ is as in (2.34). The proof for the other choices of G is left to the reader.

The claim follows from a straightforward inspection of (1.37–1.38) and (2.35–2.36), respectively. Indeed, duality of the migration operators in (1.37) and (2.35) is evident:

$$\left(L_{\text{mig}}^{(G)}H_\varphi^{(n)}(\cdot,\pi_G|_n)\right)(x) = \left(L_{\text{mig}}^{(G)*}H_\varphi^{(n)}(x,\cdot|_n)\right)(\pi_G). \tag{2.43}$$

Let us check the duality of the resampling and coalescence operators in (1.38) and (2.36). It is enough to assume that $d_0=0$, since it is well-known that Fleming-Viot operator L^d (cf. (1.18)) is dual with the generator of the Kingman coalescent which is the special case of $L_{\text{coal}}^{(G)*}$ (cf. (2.25)) with $\Lambda_0 = d_0\delta_0$.

By a standard approximation argument, it is enough to consider the duality test functions in (2.37) 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 $\varphi_i\in C_b(E)$. Using (1.38)–(1.39), (2.14), (2.33) 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 $\pi_G \in S_G$ (cf. (2.12)), i.e., π_G has the singleton family structure)

$$
\begin{aligned}
&\left(L_{\mathrm{res}}^{(\Omega_N)} H_\varphi^{(n)}(\cdot, \pi_G|_n)\right)(x) \\
&= \sum_{\eta\in G}\sum_{k\in\mathbb{N}_0} N^{-2k}\int_{[0,1]}\Lambda_k^*(\mathrm{d}r)N^{-k}\sum_{\rho\in B_k(\eta)}\int_E x_\rho(\mathrm{d}a) \\
&\quad\times\left(\prod_{i=1}^{b(\eta)}\left\langle\left(\Phi_{r,a,B_k(\eta)}(x)\right)_{\eta_{\pi^{-1}(i)}}, \prod_{j:\ \pi(j)=i}\varphi_j\right\rangle - \prod_{i=1}^{b(\eta)}\left\langle x_{\eta_{\pi^{-1}(i)}}, \prod_{j:\ \pi(j)=i}\varphi_j\right\rangle\right) \\
&= \sum_{\eta\in G}\sum_{k\in\mathbb{N}_0} N^{-2k}\int_{[0,1]}\Lambda_k^*(\mathrm{d}r)N^{-k}\sum_{\rho\in B_k(\eta)}\int_E x_\rho(\mathrm{d}a) \\
&\qquad\times\Bigg(\sum_{\substack{J\subset[b(\eta)]\\|J|\geq 0}}\prod_{i\in[b(\eta)]\setminus J}\left\langle(1-r)y_{\eta_{\pi^{-1}(i)},k}, \prod_{j:\ \pi(j)=i}\varphi_j\right\rangle\prod_{i\in J}\left\langle r\delta_a, \prod_{j:\ \pi(j)=i}\varphi_j\right\rangle \\
&\qquad\quad-\prod_{i=1}^{b(\eta)}\left\langle x_{\eta_{\pi^{-1}(i)}}, \prod_{j:\ \pi(j)=i}\varphi_j\right\rangle\Bigg) \\
&= \sum_{\eta\in G}\sum_{k\in\mathbb{N}_0} N^{-2k}\sum_{\substack{J\subset[b(\eta)],\\|J|\geq 2}}\lambda_{b(\eta),|J|}^{(\Lambda_k)} \\
&\qquad\times\Bigg(N^{-k}\sum_{\rho\in B_k(\eta)}\prod_{i\in[b(\eta)]\setminus J}\left\langle N^{-k}\sum_{\xi\in B_k(\eta)}x_\xi, \prod_{j:\pi(j)=i}\varphi_j\right\rangle\prod_{i\in J}\left\langle x_\rho, \prod_{j:\ \pi(j)=i}\varphi_j\right\rangle \\
&\qquad\quad-\prod_{i=1}^{b(\eta)}\left\langle x_{g_{\pi^{-1}(i)}}, \prod_{j:\ \pi(j)=i}\varphi_j\right\rangle\Bigg).
\end{aligned}
\tag{2.44}
$$

On the other hand, according to (2.36) (also recall (2.17), (2.29)), we have

$$
\begin{aligned}
&\left(L_{\mathrm{coal}}^{(\Omega_N)*} H_\varphi^{(n)}(x, \cdot|_n)\right)(\pi_G) = \sum_{\eta\in\Omega_N}\sum_{k\in\mathbb{N}_0} N^{-2k}\sum_{\substack{J\subset[b(\eta)],\\|J|\geq 2}}\lambda_{b(\eta),|J|}^{(\Lambda_k)} \\
&\times\Bigg(\sum_{\xi_1\in B_k(\eta)}N^{-k}\sum_{\xi_2\in B_k(\eta)}N^{-k}\cdots\sum_{\xi_{b(\eta)}\in B_k(\eta)}N^{-k} \\
&\quad\times\Bigg(\prod_{i\in[b(\eta)]\setminus J}\left\langle x_{\xi_i}, \prod_{j:\ \pi(j)=i}\varphi_j\right\rangle\left\langle x_{\xi_{\min\{l:\ l\in J\}}}, \prod_{j:\ \pi(j)\in J}\varphi_j\right\rangle \\
&\qquad-\prod_{i=1}^{b(\eta)}\left\langle x_{g_{\pi^{-1}(i)}}, \prod_{j:\ \pi(j)=i}\varphi_j\right\rangle\Bigg)\Bigg).
\end{aligned}
\tag{2.45}
$$

Comparing (2.45) with (2.44), we get the claim. □

2.4. *The long-time behaviour of the spatial* Λ*-coalescent with non-local coalescence.* We next investigate the long-time behaviour of the spatial $\underline{\Lambda}$-coalescent with non-local coalescence. Subsequently, the duality relation allows us to translate results on the long-time behaviour of the spatial $\underline{\Lambda}$-coalescent with non-local coalescence 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 spatial $\underline{\Lambda}$-coalescent with non-local coalescence as $t \to \infty$.

Proposition 2.11. [Limiting state] *Start the* $\mathfrak{C}^{(\Omega_N)}$*-process from* (2.32) *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*

$$\mathcal{L}\left[H_\varphi^{(n)}(x, \mathfrak{C}_n^{(\Omega_N)}(t))\right] \underset{t\to\infty}{\Longrightarrow} \mathcal{L}\left[H_\varphi^{(n)}(\underline{\theta}, \mathfrak{C}_n^{(\Omega_N)}(\infty))\right] \qquad \forall\, n \in \mathbb{N}, \tag{2.46}$$

where $\mathfrak{C}_n^{(\Omega_N)}$ *is as in Section* 2.2.1 *and* $H_\varphi^{(n)}$ *as in* (2.37)*.*

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 $\mathfrak{C}^{(\Omega_N,n)}(\infty) \in \Pi_{\Omega_N,n}$ (cf. (2.7)). We must prove that the locations result in an effective averaging of the configuration x, so that we can replace the $|\mathfrak{C}_n^{(\Omega_N)}(t)|$-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 (1985)). $\square$

Recall the definition of the spatial Λ-coalescent with immigration-emigration introduced in Section 2.1.3.

Corollary 2.12. [Limiting state of the Λ**-coalescent with immigration-emigration]** *The analogous to* (2.46) *statement holds if we substitute* $\mathfrak{C}_n^{(\Omega_N)}$ *with the the* Λ*-coalescent with immigration-emigration (see Section* 2.1.3*), i.e., the spatial* Λ*-coalescent* $\mathfrak{C}_n^{(G),\mathrm{loc}}$ *with geographic space* $G = \{0, *\}$ *and migration kernel* $a(0,*) = c$*,* $a(*,0) = 0$*.*

$$\mathcal{L}\left[H_\varphi^{(n)}(x, \mathfrak{C}_n^{(\{0,*\}),\mathrm{loc}}(t))\right] \underset{t\to\infty}{\Longrightarrow} \mathcal{L}\left[H_\varphi^{(n)}((y, x_*), \mathfrak{C}_n^{(\{0,*\}),\mathrm{loc}}(\infty))\right] \forall y \in \mathcal{P}(E), n \in \mathbb{N}, \tag{2.47}$$

where $H_\varphi^{(n)}$ *as in* (2.37) *and* $x = (x_0, x_*) \in \mathcal{P}(E)^2$*. Note that the right hand side of* (2.47) *does not depend on* y*.*

2.4.2. *The dichotomy: single ancestor versus multiple ancestors.* A key question is whether the $\mathfrak{C}^{(\Omega_N)}$-process from (2.32) converges to a single labelled partition element as $t \to \infty$ with probability one. To answer this question, 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 $\mathfrak{C}_n^{(\Omega_N)}$-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 Ω_N with migration coefficients $(\bar{c}_k(N))_{k\in\mathbb{N}_0}$ (cf. (1.62)) and coalescence at rates $(\lambda_k)_{k\in\mathbb{N}_0}$. Consider the *time-t accumulated hazard for coalescence* of this pair:

$$H_N(t) = \sum_{k\in\mathbb{N}_0} \lambda_k N^{-k} \int_0^t 1\left\{d(Z_s^1, Z_s^2) \leq k\right\} \mathrm{d}s. \tag{2.48}$$

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 explains the factor N^{-k} in (2.48). Let

$$H_N \equiv \lim_{t\to\infty} H_N(t). \tag{2.49}$$

We have coalescence of the random walks (= common ancestor) with probability one, when $H_N = \infty$ a.s., but separation of the random walks (= different ancestors) with positive probability, when $H_N < \infty$ a.s.

Lemma 2.13. [Zero-one law] *$H_N = \infty$ a.s. if and only if $\bar{H}_N = \mathbb{E}[H_N] = \infty$. Moreover, under the weak regularity condition in (1.68) the latter is equivalent to*

$$\sum_{k\in\mathbb{N}_0} \frac{1}{c_k} \sum_{l=0}^{k} \lambda_l = \infty. \tag{2.50}$$

Proof: Write $H_N = \sum_{k\in\mathbb{N}_0} w_k L(k)$ with

$$w_k(N) = \sum_{j\geq k} \lambda_j N^{-j}, \qquad L(k) = \int_0^\infty 1\{d(Z_s^1, Z_s^2) = k\}\mathrm{d}s. \tag{2.51}$$

Note that $w_k(N) < \infty$ because of condition (1.32). We want to show that $\bar{H}_N = \infty$ implies $H_N = \infty$ (the reverse is immediate). Recall from Section 1.5.3 that $P_t(\cdot,\cdot)$ denotes the time-t transition kernel of the hierarchical random walk on Ω_N with migration coefficients $(\bar{c}_k(N))_{k\in\mathbb{N}_0}$ given by (1.62). In the computations below, we pretend that the coefficients are $(c_k)_{k\in\mathbb{N}_0}$. Afterwards, we can replace c_k by $\bar{c}_k(N)$.

Note that $(Z_s^1 - Z_s^2)_{s\geq 0}$ has the same law as a single copy $(Z_s)_{s\geq 0}$ of the hierarchical random walk but moving at twice the speed. Thus, in law, we may replace $L(k)$ by $L(k) = \int_0^\infty 1\{|Z_{2s}| = k\}\mathrm{d}s$.

Step 1. As shown in Dawson et al. (2005, Eq. (3.1.5)), for the hierarchical random walk with jump rate 1,

$$P_t(0,\eta) = \sum_{j\geq k} K_{jk}(N)\, \frac{\exp\left[-h_j(N)t\right]}{N^j}, \qquad t\geq 0,\, \eta\in\Omega_N\colon\ |\eta| = k\in\mathbb{N}_0, \tag{2.52}$$

where

$$K_{jk}(N) = \begin{cases} 0, & j=k=0,\\ -1, & j=k>0,\\ N-1, & \text{otherwise,}\end{cases} \qquad j,k\in\mathbb{N}_0, \tag{2.53}$$

and

$$h_j(N) = \frac{N}{N-1} r_j(N) + \sum_{i>j}^{\infty} r_i(N), \quad j\in\mathbb{N}, \tag{2.54}$$

where, for the hierarchical random walk defined in Section 1.4.2,

$$r_j(N) = \frac{1}{D(N)} \sum_{i\geq j} \frac{c_{i-1}}{N^{2i-j-1}}, \qquad j\in\mathbb{N}, \tag{2.55}$$

with $D(N)$ the normalising constant such that $\sum_{j\in\mathbb{N}} r_j(N) = 1$.

The random walk in Dawson et al. (2005) has jump rate 1, while our hierarchical random walk has jump rate

$$\begin{aligned} D^*(N) &= \sum_{\eta\in\Omega_N} a^{(N)}(0,\eta) = \sum_{k\in\mathbb{N}} (N^k - N^{k-1}) \sum_{j\geq k} \frac{c_{j-1}}{N^{2j-1}} \\ &= \sum_{m\in\mathbb{N}_0} \frac{c_m}{N^m}\left(1 - \frac{1}{N^{m+1}}\right). \end{aligned} \tag{2.56}$$

Therefore, after computing H_N with the help of the above formulas, we must divide H_N by $D^*(N)$ to get the correct expression.

Note that (2.54–2.55) simplify considerably when $N \to \infty$, namely,

$$h_j(N) \sim r_j(N) \sim \frac{c_{j-1}}{D(N)N^{j-1}}, \qquad D(N) \sim c_0, \tag{2.57}$$

while also (2.51) and (2.56) simplify to

$$w_k(N) \sim \frac{\lambda_k}{N^k}, \qquad D^*(N) \sim c_0. \tag{2.58}$$

Moreover, because $\limsup_{k\to\infty} \frac{1}{k}\log c_k < \log N$ and $\limsup_{k\to\infty} \frac{1}{k}\log\lambda_k < \log N$ (see the footnotes in Sections 1.4.2–1.4.3), the following holds:

For every $N \in \mathbb{N}\backslash\{1\}$ the quantities $h_j(N)$, $r_j(N)$, $D(N)$, $w_k(N)$ and $D^*(N)$ are bounded from above and below by positive finite constants times their $N \to \infty$ asymptotics uniformly in the indices j, k. (2.59)

Step 2. For $M \in \mathbb{N}_0$, define the truncated hazard

$$H_N^{(M)} = \sum_{k=0}^{M} w_k(N) L(k). \tag{2.60}$$

For a non-negative random variable V with a finite second moment, Cauchy-Schwarz gives

$$\mathbb{P}\{V > 0\} \geq (\mathbb{E}[V])^2/\mathbb{E}[V^2]. \tag{2.61}$$

Therefore,

$$\mathbb{P}\left\{H_N^{(M)} \Big/ \mathbb{E}\left[H_N^{(M)}\right] > 0\right\} \geq \left(\mathbb{E}\left[H_N^{(M)}\right]\right)^2 \Big/ \mathbb{E}\left[(H_N^{(M)})^2\right]. \tag{2.62}$$

To compute the quotient in the right-hand side of (2.62), we write

$$\begin{aligned} \mathbb{E}\left[H_N^{(M)}\right] &= \sum_{k=0}^{M} w_k(N) \int_0^\infty \mathrm{d}s\, P\{|Z_{2s}| = k\} \\ &= \tfrac{1}{2} \sum_{k=0}^{M} w_k(N) \sum_{\eta\in\partial B_k(0)} G(0,\eta) \end{aligned} \tag{2.63}$$

and

$$\begin{aligned}
\mathbb{E}\left[\left(H_N^{(M)}\right)^2\right] &= \sum_{k,l=0}^{M} w_k(N) w_l(N) \int_0^\infty \mathrm{d}s \int_0^\infty \mathrm{d}t\, P\{|Z_{2s}| = k\} P\{|Z_{2t}| = l\} \\
&= \tfrac{1}{2} \sum_{k,l=0}^{M} w_k(N) w_l(N) \sum_{\substack{\eta \in \partial B_k(0) \\ \eta' \in \partial B_l(0)}} G(0,\eta) G(0,\eta'-\eta).
\end{aligned} \tag{2.64}$$

Here, G is the Green function of the hierarchical random walk, which by (2.52) equals

$$G(0,\eta) = G_k(N), \qquad \eta \in \Omega_N \colon\, |\eta| = k \in \mathbb{N}_0, \qquad G_k(N) = \sum_{j \geq k} K_{jk}(N)\, \frac{1}{h_j(N) N^j}. \tag{2.65}$$

Let

$$N[k] = \begin{cases} 1, & k = 0, \\ N^k - N^{k-1}, & k > 0, \end{cases} \qquad \bar{N}[k] = \begin{cases} 1, & k = 0, \\ N^k - 2N^{k-1}, & k > 0, \end{cases} \tag{2.66}$$

denote the number of sites at distance k from the origin, respectively, at distance k from both the origin and a given site itself at distance k from the origin. A straightforward counting argument shows that

$$\begin{aligned}
\text{r.h.s.}(2.63) &= \tfrac{1}{2} \sum_{k=0}^{M} w_k(N)\, N[k]\, G_k(N), \\
\text{r.h.s.}(2.64) &= \tfrac{1}{2} \sum_{k,l=0}^{M} w_k(N)\, w_l(N)\, N[k]\, N[l]\, G_{k \vee l}^2(N) \\
&\quad + \tfrac{1}{2} \sum_{k=0}^{M} w_k^2(N)\, N[k]\, G_k(N) \Big\{ \big(\bar{N}[k] - N[k]\big)\, G_k(N) \\
&\qquad\qquad + \sum_{m=0}^{k-1} N[m]\, G_m(N) \Big\}.
\end{aligned} \tag{2.67}$$

For $N \to \infty$, substituting (2.53) and (2.57) into (2.65) and the resulting expression into (2.67), we get

$$\mathbb{E}\left[H_N^{(M)}\right] \sim \sum_{k=0}^{M} \mu_k \sum_{m \geq k} \frac{1}{c_m} \tag{2.68}$$

and

$$\mathbb{E}\left[\left(H_N^{(M)}\right)^2\right] \sim 2 \sum_{k,l=0}^{M} \mu_k \mu_l \left(\sum_{m \geq k \vee l} \frac{1}{c_m} \right)^2, \tag{2.69}$$

where we use that the dominant term in the sum defining $G_k(N)$ in (2.65) is the one with $j = k+1$, and we also use that $\mu_k = \frac{1}{2}\lambda_k$ as in (1.57). Thus, for every M, the right-hand side of (2.62) is bounded from below by a number that tends to $\frac{1}{2}$ as $N \to \infty$. Together with the observation made below (2.57–2.58), it therefore follows that there exists a $\delta > 0$ independent of M and N such that

$$\mathbb{P}\left\{ H_N^{(M)} \Big/ \mathbb{E}\left[H_N^{(M)}\right] > 0 \right\} \geq \delta. \tag{2.70}$$

Step 3.

Since $H_N^{(M)} \leq H_N$ and $H_N = \lim_{M\to\infty} H_N^{(M)}$, it follows from (2.70) that

$$\mathbb{P}\left\{ H_N / \mathbb{E}[H_N] > 0 \right\} \geq \delta. \tag{2.71}$$

Thus, $\mathbb{E}[H_N] = \infty$ implies $\mathbb{P}\{H_N = \infty\} \geq \delta$. But the event $\{H_N = \infty\}$ lies in the tail-sigma-algebra of the hierarchical random walk, which is trivial, and therefore this event has probability 0 or 1. Consequently, $P\{H_N = \infty\} = 1$.

Step 4. Finally, replacing c_k by $\bar{c}_k(N) = c_k + N^{-1}\lambda_{k+1}$ (recall (1.62)), noting that (2.59) continues to apply, and using (2.68) with $M = \infty$, we get that $P\{H_N = \infty\} = 1$ if and only if

$$\sum_{k\in\mathbb{N}_0} \lambda_k \sum_{m\geq k} \frac{1}{c_m + N^{-1}\lambda_{m+1}} = \infty, \tag{2.72}$$

which is the same as

$$\sum_{k\in\mathbb{N}_0} \frac{1}{c_k + N^{-1}\lambda_{k+1}} \sum_{l=0}^{k} \lambda_l = \infty. \tag{2.73}$$

Under the weak regularity condition in (1.68) the latter is equivalent to (2.50). □

3. Well-posedness of martingale problems

Our task in this section is to prove Propositions 1.1–1.4, 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 $\{H(\cdot, Y_0)\colon\, Y_0 \in \mathcal{E}'\}$ (cf. (2.39)) 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 (1986, Chapter 4)). For Ω_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., Etheridge (2000, Section 1.6) or Ethier and Kurtz (1986, 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 this section, we prove Propositions 1.1–1.4. We follow the line of argument of Evans (1997, 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 2.10 respectively (2.41)). The proofs of Propositions 1.1–1.4 follow the same pattern for $G = \{0, \dots, N-1\}$, $G = \{0, *\}$ and $G = \Omega_N$.

Proof of Propositions 1.1–1.4:

- *Well-posedness.* First we show that there exists a Markov transition kernel Q_t on $\mathcal{P}(E)^G$ such that, for all $\varphi \in C_{\mathrm{b}}(E^n, \mathbb{R})$, $\pi \in \Pi_{G,n}$ (cf. (2.7)), $X \in \mathcal{P}(E)^G$ and $t \geq 0$,

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

where $H_\varphi^{(n)}$ as in (2.37) and $\mathfrak{C}_n^{(G)}$ as in (2.20) resp. (2.32) depending on the choice of G. 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 (1968, Theorem I.9.4)). This cádlág process is unique and coincides with the process $X^{(G)}$ from (1.8) resp. (1.20) resp. from below Proposition 1.4, since (3.1) implies (2.41). There can be at most one process satisfying (2.41), since the family of duality functions $H_\varphi^{(n)}(\cdot, \pi)$ separates points on $\mathcal{P}(E)^G$.

Finally, the transition kernel Q_t satisfying (3.1) exists as a solution of the Hausdorff moment problem (3.1) and is Markov due to the Markov property of the spatial coalescent on the right-hand side of (3.1) (see Evans (1997, Theorem 4.1) for details).

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

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

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 (2.37). The duality in (2.41) implies that

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

Recall that we equip $\mathcal{P}(E)^G$ with the topology of weak convergence. Definition (2.37) readily implies that the right-hand side of (3.3) is continuous 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 (cf. Section 1.3.3) 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 θ (Section 4.2), a key property that will be needed later on and for which we need that the Λ-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

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

where we recall (1.15–1.18) for the form of F_φ and $L_\theta^{c,d,\Lambda}$.

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

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

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, namely, the Λ-coalescent with immigration-emigration (see Section 2.1.3), to show that the expectation in the right-hand side of the duality relation (2.41) converges. Indeed, we showed in (2.46) in Proposition 2.11 and its Corollary 2.12 that the state of the duality function $H(X_0, \cdot)$, cf., (2.37), 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 $\{H(\cdot, C_0)\colon\ C_0 \in \mathcal{E}'\}$ ($\mathcal{E}'$ as below (2.39)) and hence (2.46) 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 θ. The latter implies that the limit is unique. Since $\mathcal{P}(E)$ is compact and the process is Feller, there must exist 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

$$\mathcal{P}(E) \ni \theta \mapsto \nu_\theta^{c,d,\Lambda} \in \mathcal{P}(\mathcal{P}(E)) \tag{4.3}$$

is uniformly continuous for suitably chosen metrics (in the weak topology on the respective metrisable spaces). We will choose the metrics in (4.7–4.8) below. Recall the definition of the duality functions H from (2.38–2.39). Since the family $\{H(\,\cdot\,, C_0)\colon\ C_0 \in \mathcal{E}'\}$ is dense in $C_\mathrm{b}(\mathcal{P}(E), \mathbb{R})$, we can approximate any function in $C_\mathrm{b}(\mathcal{P}(E), \mathbb{R})$ by duality functions in the supremum norm. In fact, even the smaller family $\{H_\varphi(\,\cdot\,, \{\{1\}, \dots, \{n\}\})\colon\ n \in \mathbb{N}, \varphi \in C_\mathrm{b}(E)\}$ is dense in $C_\mathrm{b}(\mathcal{P}(E), \mathbb{R})$. It is enough to prove uniform continuity for the duality function uniformly in the family, even with the additional restriction $\|\varphi\|_\infty < 1$. For this purpose, we analyse the limiting random variable for the corresponding dual as a function of θ in the limit as $t \to \infty$.

If $(C_t^{c,\Lambda})_{t\geq 0}$ denotes the spatial Λ-coalescent with immigration-emigration starting from $\{(\{1\},0),\dots,(\{n\},0)\}$ and jumping to the cemetery state $\{*\}$ at rate c, then $H(\theta, C_\infty^{c,\Lambda})$ uniquely determines the McKean-Vlasov limit law $\nu_\theta^{c,d,\Lambda}$ for $t \to \infty$. Recall that we associate the distribution of types θ with the cemetery state. 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.

Let

$$P_{n,k} = \mathbb{P}\big\{|C_\infty^{c,\Lambda}| = k \mid C_0^{c,\Lambda} = \{\{1\},\dots,\{n\}\}\big\}. \tag{4.4}$$

For all $\theta \in \mathcal{P}(E)$ and all $\varphi \in C_{\mathrm{b}}(E)$ with $\|\varphi\|_\infty < 1$, taking $H_\varphi(\underline{\theta}, (\{1\},\dots,\{n\}\}) = \langle\theta,\varphi\rangle^n$ we have

$$\mathbb{E}\left[H(\underline{\theta}, C_\infty^{c,\Lambda}) \mid C_0^{c,\Lambda} = \{\{1\},\dots,\{n\}\}\right] = \sum_{k=1}^n P_{n,k}\langle\theta,\varphi\rangle^k. \tag{4.5}$$

From the right-hand side of (4.5), we read off that the family of functions

$$\left\{\mathbb{E}\left[H_\varphi(\underline{\theta}, C_\infty^{c,\Lambda}) \mid C_0^{c,\Lambda} = \{\{1\},\dots,\{n\}\}\right] : \, n\in\mathbb{N}\right\} \text{ is uniformly continuous in } \theta. \tag{4.6}$$

On $\mathcal{P}(E)$ we choose the metric

$$\rho_{\mathcal{P}(E)}(\theta,\theta') \equiv \sum_{k\in\mathbb{N}} 2^{-k}|\langle\theta-\theta',\varphi_k\rangle|, \quad \theta,\theta'\in\mathcal{P}(E), \tag{4.7}$$

where $\{\varphi_k \in C(E)\colon\, k\in\mathbb{N}\}$ with $\sup_{k\in\mathbb{N}}\|\varphi_k\|_\infty < 1$ separates points and therefore generates the topology. On $\mathcal{P}(\mathcal{P}(E))$, we choose the metric

$$\begin{aligned}\rho_{\mathcal{P}(\mathcal{P}(E))}(X,X') \equiv \sum_{n\in\mathbb{N}}\sum_{k\in\mathbb{N}} 2^{-k-n}\mathbb{E}\left[|H_{\varphi_k}(X-X',\{\{1\},\dots,\{n\}\})|\right],\\ X,X'\in\mathcal{P}(\mathcal{P}(E)).\end{aligned} \tag{4.8}$$

Combining (4.6–4.8), we get the uniform continuity of (4.3).

4.3. *Structure of the McKean-Vlasov equilibrium.* In the case of the McKean-Vlasov Fleming-Viot processes, the equilibrium $\nu_\theta^{c,d,0}$ can be identified as an atomic measure of the form

$$\sum_{i\in\mathbb{N}}\left[W_i\prod_{j=1}^{i-1}(1-W_j)\right]\delta_{U_i} \tag{4.9}$$

with $(U_i)_{i\in\mathbb{N}}$ i.i.d. θ-distributed and $(W_i)_{i\in\mathbb{N}}$ i.i.d. $\mathrm{BETA}(1,\frac{c}{d})$-distributed, independently of each other (cf. Dawson et al. (1995)). What we can say about the equilibrium $\nu_\theta^{c,d,\Lambda}$?

Proposition 4.2. [Towards a representation for 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 Λ has the dust-free property (recall (1.3)).*

(a) *The following decomposition holds:*

$$\nu_\theta^{c,d,\Lambda} = \mathcal{L}\left[\sum_{i\in\mathbb{N}} V_i\delta_{U_i}\right]. \tag{4.10}$$

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* θ*,* $\sum_{i\in\mathbb{N}} V_i = 1$ *a.s., and*

$$V_i = W_i \prod_{j=1}^{i-1}(1-W_j), \tag{4.11}$$

where

$$(W_j)_{j\in\mathbb{N}} \tag{4.12}$$

is a sequence of $[0,1]$*-valued random variables whose joint 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* Λ*. (See* Remark 4.3 *below.)*

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

$$0 \le \nu_\theta^{c,d,\Lambda}(M) < 1. \tag{4.13}$$

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 Λ-coalescent starting from the partition into singletons: $\{\{1\},\{2\},\ldots\}$. This representation is a consequence of the duality relation in (2.41) and de Finetti's theorem, together with the dust-free condition on Λ 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 $\{\{1\},\{2\},\ldots\}$ at site 1, where we assign to each dual individual the type of its partition element at time ∞, drawn independently from θ, 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.

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

$$\mathbb{E}_{\nu_\theta^{c,d,\Lambda}}[\langle X, f\rangle^n] \tag{4.14}$$

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

$$c,\ d,\ r^k(1-r)^{n-k}\Lambda(\mathrm{d}r). \tag{4.15}$$

These relations uniquely determine the distribution of the atom sizes, which in turn uniquely determines the marginal distribution of the W_i's via (4.11).

(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 $\mathfrak{C}_\infty^{(\{0,*\})}$ (the limit of the dual coalescent) contains more than one partition element has a positive probability. But this is obviously true when $c, d > 0$. $\square$

Remark 4.3. It is well known (cf. Dawson et al. (1995)) that if $\Lambda = \delta_0$ (the McKean-Vlasov Fleming-Viot process), then the W_i's are i.i.d. with distribution BETA$(1, \frac{c}{d})$. It remains an open problem to identify the law of the W_i's for the general Cannings resampling as function of the ingredients in (4.15). We note that if the W_i's happen to be independent, then W_i has distribution BETA$(1-\alpha, i\alpha+\beta)$ for some $\alpha \in [0,1]$ and $\beta \in [0,\infty)$ (see Pitman (2006, Theorem 3.4)).

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 1.5.2. Recall the definition of $\mathbb{E}_{\nu_\theta}[\mathrm{Var}.(\psi)]$ from (1.56) and of $\mathrm{Var}_x(\psi)$ from (1.54). Recall $\lambda = \Lambda([0;1])$.

Proposition 4.4. [Variance] *For every* $\psi \in \mathcal{C}_\mathrm{b}(E)$,

$$\mathbb{E}_{\nu_\theta^{c,d,\Lambda}}\left[\mathrm{Var}.\left(\psi\right)\right] = \int_{\mathcal{P}(E)} \nu_\theta^{c,d,\Lambda}(\mathrm{d}x)\left(\left\langle \psi^2, x\right\rangle - \left\langle \psi, x\right\rangle^2\right) = \frac{2c}{2c+\lambda+2d}\,\mathrm{Var}_\theta\left(\psi\right). \tag{4.16}$$

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

$$n = 1,\ \varphi \in \mathcal{C}_\mathrm{b}(E){:} \qquad 0 = c\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\,\langle \varphi, (\theta - x)\rangle\,, \tag{4.17}$$

i.e., $\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\,\langle \varphi, x\rangle = \langle \varphi, \theta\rangle$. It further follows that, for $n = 2$, $\varphi \in \mathcal{C}_\mathrm{b}(E^2)$,

$$\begin{aligned} 0 = {} & -2c\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\left\langle \varphi, x^{\otimes 2}\right\rangle \\ & + c\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\,\left[\langle \varphi, \theta\otimes x\rangle + \langle \varphi, x\otimes\theta\rangle\right] \\ & + 2d\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\left(\int_E x(\mathrm{d}a)\,\langle \varphi, \delta_a^{\otimes 2}\rangle - \left\langle \varphi, x^{\otimes 2}\right\rangle\right) \\ & + \lambda\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\int_E x(\mathrm{d}a)\left\langle \varphi, (\delta_a - x)^{\otimes 2}\right\rangle. \end{aligned} \tag{4.18}$$

We can rewrite (4.18) as

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

From this, we see that

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

where we use (4.17) in the last line. Substituting this back into (4.19) and using (4.17) once more, we get

$$\begin{aligned}
&\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\int_E\int_E Q_x(\mathrm{d}u,\mathrm{d}v)\,\varphi(u,v)\\
&\quad = \int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\left(\int_E x(\mathrm{d}a)\,\left\langle \varphi, \delta_a^{\otimes 2}\right\rangle - \left\langle \varphi, x^{\otimes 2}\right\rangle\right)\\
&\quad = \frac{2c}{\lambda+2d}\left(\int_{\mathcal{P}(E)} \nu(\mathrm{d}x)\,\left\langle \varphi, x^{\otimes 2}\right\rangle - \left\langle \varphi, \theta^{\otimes 2}\right\rangle\right)\\
&\quad = \frac{2c}{2c+\lambda+2d}\left(\int_E \theta(\mathrm{d}a)\,\left\langle \varphi, \delta_a^{\otimes 2}\right\rangle - \left\langle \varphi, \theta^{\otimes 2}\right\rangle\right)\\
&\quad = \frac{2c}{2c+\lambda+2d}\int_E\int_E Q_\theta(\mathrm{d}u,\mathrm{d}v)\,\varphi(u,v).
\end{aligned} \tag{4.21}$$

Pick $\varphi = \psi\times\psi$ in (4.21) to get the claim. □

For $\lambda = \Lambda([0,1]) = 0$, (4.16) is the same as Dawson et al. (1995, Eq. (2.5)).

Corollary 4.5. [Asymptotic variance of entrance law] *For $\varphi \in C_{\mathrm{b}}(E,\mathbb{R})$, the interaction chain (cf., Section 1.5.2) satisfies*

$$\lim_{j\to\infty} \mathbb{E}_{\mathcal{L}(M_0^{(j)})}[\mathrm{Var}_{\cdot}(\varphi)] = 0 \quad (\textit{respectively}, \ > 0), \tag{4.22}$$

if $\sum_{k\in\mathbb{N}} m_k = \infty$ (respectively, $\sum_{k\in\mathbb{N}} m_k < \infty$) with m_k defined in (1.57) and d_k in (1.45).

Proof: From (4.16), we have the formula

$$\mathbb{E}_{\nu_\theta^{c,d,\Lambda}}[\mathrm{Var}_{\cdot}(\varphi)] = \frac{2c}{2c+\lambda+2d}\,\mathrm{Var}_\theta(\varphi). \tag{4.23}$$

Hence, we have the relation (recall (1.49) for the definition of $K_k(\theta,\mathrm{d}x)$)

$$\int_{\mathcal{P}(E)} K_k(\theta,\mathrm{d}x)\,\mathrm{Var}_x(\varphi) = \frac{2c_k}{2c_k+\lambda_k+2d_k}\,\mathrm{Var}_\theta(\varphi), \tag{4.24}$$

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

$$n_k \equiv \frac{2c_k}{2c_k+\lambda_k+2d_k} = \frac{1}{1+m_k}. \tag{4.25}$$

Iteration gives

$$\mathbb{E}_{\mathcal{L}(M_0^{(j)})}[\mathrm{Var}_{\cdot}(\varphi)] = \left(\prod_{k=0}^{j} n_k\right) \mathrm{Var}_\theta(\varphi) = \left(\prod_{k=0}^{j}\left(\frac{1}{1+m_k}\right)\right) \mathrm{Var}_\theta(\varphi). \tag{4.26}$$

Therefore, taking logarithms, we see that (4.22) is equivalent to

$$\sum_{k\in\mathbb{N}_0} m_k = \infty \text{ (respectively, } < \infty). \tag{4.27}$$

□

We next prove a result that is similar to, but more involved than, Dawson et al. (1995), Eq. (6.12). This result is necessary for the proof of Theorem 1.16 on diffusive clustering.

Proposition 4.6. [Variance of the integral against a test function] *For every $\psi \in \mathcal{C}_\mathrm{b}(E)$, $j \in \mathbb{N}$ and $0 \le k \le j+1$,*

$$\begin{aligned}\mathrm{Var}_{\mathcal{L}(M_{-k}^{(j)})}(\langle\cdot,\psi\rangle) &= \mathbb{E}_{\mathcal{L}(M_{-k}^{(j)})}[\langle\cdot,\psi\rangle^2] - \left(\mathbb{E}_{\mathcal{L}(M_{-k}^{(j)})}[\langle\cdot,\psi\rangle]\right)^2 \\ &= \left(\sum_{i=k}^{j}\left(\frac{d_{i+1}}{c_i}\prod_{l=i+1}^{j}\frac{1}{1+m_l}\right)\right)\mathrm{Var}_\theta(\psi).\end{aligned} \tag{4.28}$$

Proof: The proof uses the following two ingredients. Combining (4.16) and (4.25), we have

$$\mathbb{E}_{\nu_\theta^{c_k,d_k,\Lambda_k}}[\mathrm{Var}_{\cdot}(\psi)] = \frac{1}{1+m_k}\mathrm{Var}_\theta(\psi). \tag{4.29}$$

The first and the third line of (4.21) yield

$$\mathrm{Var}_{\nu_\theta^{c_k,d_k,\Lambda_k}}(\langle\cdot,\psi\rangle) = \frac{\lambda+2d}{2c}\,\mathbb{E}_{\nu_\theta^{c_k,d_k,\Lambda_k}}[\mathrm{Var}_{\cdot}(\psi)]. \tag{4.30}$$

Together with (4.16) and (1.45), we therefore obtain

$$\mathrm{Var}_{\nu_\theta^{c_k,d_k,\Lambda_k}}(\langle\cdot,\psi\rangle) = \frac{\lambda_k+2d_k}{2c_k+\lambda_k+2d_k}\,\mathrm{Var}_\theta(\psi) = \frac{d_{k+1}}{c_k}\,\mathrm{Var}_\theta(\psi). \tag{4.31}$$

Fix $j \in \mathbb{N}$. The proof follows by downward induction over $0 \le k \le j+1$. The initial case $k = j+1$ is obvious because $M_{-(j+1)}^{(j)} = \theta$ by (1.48). Let us therefore assume that the claim holds for $k+1$. By (1.48–1.49),

$$\begin{aligned}\mathrm{Var}_{\mathcal{L}(M_{-k}^{(j)})}(\langle\cdot,\psi\rangle) &= \mathbb{E}_{\mathcal{L}(M_{-k}^{(j)})}[\langle\cdot,\psi\rangle^2] - \left(\mathbb{E}_{\mathcal{L}(M_{-k}^{(j)})}[\langle\cdot,\psi\rangle]\right)^2 \\ &= \int_{\mathcal{P}(E)}\nu_\theta^{c_j,d_j,\Lambda_j}(\mathrm{d}\theta_j)\int_{\mathcal{P}(E)}\nu_{\theta_j}^{c_{j-1},d_{j-1},\Lambda_{j-1}}(\mathrm{d}\theta_{j-1}) \\ &\quad \dots\int_{\mathcal{P}(E)}\nu_{\theta_{k+1}}^{c_k,d_k,\Lambda_k}(\mathrm{d}\theta_k)\langle\theta_k,\psi\rangle^2 - \langle\theta,\psi\rangle^2.\end{aligned} \tag{4.32}$$

Next, use (4.31) to rewrite the inside integral as

$$\int_{\mathcal{P}(E)}\nu_{\theta_{k+1}}^{c_k,d_k,\Lambda_k}(\mathrm{d}\theta_k)\langle\theta_k,\psi\rangle^2 = \mathbb{E}_{\nu_{\theta_{k+1}}^{c_k,d_k,\Lambda_k}}(\langle\cdot,\psi\rangle^2) = \langle\theta_{k+1},\psi\rangle^2 + \frac{d_{k+1}}{c_k}\mathrm{Var}_{\theta_{k+1}}(\psi). \tag{4.33}$$

Substitute this back into (4.32), to obtain

$$\begin{aligned} \operatorname{Var}_{\mathcal{L}(M^{(j)}_{-k})}(\langle\cdot,\psi\rangle) &= \operatorname{Var}_{\mathcal{L}(M^{(j)}_{-(k+1)})}(\langle\cdot,\psi\rangle) \\ &\quad + \frac{d_{k+1}}{c_k}\int_{\mathcal{P}(E)} \nu_\theta^{c_j,d_j,\Lambda_j}(\mathrm{d}\theta_j)\int_{\mathcal{P}(E)} \nu_{\theta_j}^{c_{j-1},d_{j-1},\Lambda_{j-1}}(\mathrm{d}\theta_{j-1}) \\ &\quad\quad \cdots\int_{\mathcal{P}(E)} \nu_{\theta_{k+2}}^{c_{k+1},d_{k+1},\Lambda_{k+1}}(\mathrm{d}\theta_{k+1})\operatorname{Var}_{\theta_{k+1}}(\psi). \end{aligned} \tag{4.34}$$

The first term is given by the induction hypothesis. For the second term we use (4.29), to see that the inside integral equals

$$\begin{aligned} \int_{\mathcal{P}(E)} \nu_{\theta_{k+2}}^{c_{k+1},d_{k+1},\Lambda_{k+1}}(\mathrm{d}\theta_{k+1})\operatorname{Var}_{\theta_{k+1}}(\psi) &= \mathbb{E}_{\nu_{\theta_{k+2}}^{c_{k+1},d_{k+1},\Lambda_{k+1}}}(\operatorname{Var}_{\cdot}(\psi)) \\ &= \frac{1}{1+m_{k+1}}\operatorname{Var}_{\theta_{k+2}}(\psi). \end{aligned} \tag{4.35}$$

Iteration of this reasoning for the second term in (4.34) leads to

$$\begin{aligned} \operatorname{Var}_{\mathcal{L}(M^{(j)}_{-k})}(\langle\cdot,\psi\rangle) &= \operatorname{Var}_{\mathcal{L}(M^{(j)}_{-(k+1)})}(\langle\cdot,\psi\rangle) + \frac{d_{k+1}}{c_k}\prod_{l=k+1}^{j}\frac{1}{1+m_l}\operatorname{Var}_\theta(\psi) \\ &= \left(\sum_{i=k+1}^{j}\left(\frac{d_{i+1}}{c_i}\prod_{l=i+1}^{j}\frac{1}{1+m_l}\right)\right)\operatorname{Var}_\theta(\psi) \\ &\quad + \frac{d_{k+1}}{c_k}\prod_{l=k+1}^{j}\frac{1}{1+m_l}\operatorname{Var}_\theta(\psi), \end{aligned} \tag{4.36}$$

which proves the claim. □

If $\lambda_k = \Lambda_k([0,1]) = 0$, $k \in \mathbb{N}_0$, then (4.28) reduces to Dawson et al. (1995, Eq. (6.12)). Indeed, in that case $d_{i+1}\prod_{l=i+1}^{j}\frac{1}{1+m_l}$ is equal to d_{i+1}. (Note the typo in Dawson et al. (1995, Eq. (6.12)): d_k should be replaced by d_{k+1}.)

Remark 4.7. The results in this section can alternatively be inferred from the long-time behaviour of the spatial Λ-coalescent with $G = \{0,*\}$.

5. Strategy of the proof of the main scaling theorem

The proof of Theorem 1.5 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 et al. (1995), a general scheme was developed to derive the scaling behaviour of space-time block averages as in (1.44) for hierarchically interacting Fleming-Viot processes, with the interaction coming from migration, i.e., a system similar to ours but *without* Λ-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 1.5. In these sections, we focus on the new features coming from the Λ-Cannings block resampling. The refined multi-scale result in Theorem 1.6 will be proved in Section 9. The line of argument can be largely based on the work in Dawson et al. (1995, Section 4), where it was developed in detail for Fleming-Viot. No new ideas are needed for the Cannings process: only a new moment calculation is required.

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

- Show that for the mean-field system from Section 1.3.2, 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 (recall Section 1.3.3), and for block averages on time scale Nt Fleming-Viot processes with a resampling constant d_1 corresponding to Λ_0 and c_0. With an additional Λ_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 from Section 1.4.4 restricted to $G_{N,K}$ as in (1.42). More precisely, study its components and its k-block averages (1.43) 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 $1 \leq k \leq j < K$ via an approximation of the $C_N^{\underline{c},\underline{\Lambda}}$-process on Ω_N by the process on $G_{N,K}$ from the previous step, in the limit as $N \to \infty$ and on time scales at most $N^K t$ for a fixed but otherwise arbitrary $K \in \mathbb{N}$. This is done in Section 8.

The three steps above are carried out following the scheme of proof developed in Dawson et al. (1995). What is new for *jump* processes? We are dealing with sequences of measure-valued processes $X = (X_t)_{t\geq 0}$, and the key difference is that now semi-martingales arising from functionals of the process of the form $\langle X_t, f\rangle^n$ with $f \in C_\mathrm{b}(E)$ are no longer controlled just by the *compensator* and the *increasing process* of the linear functional $\langle X_t, f\rangle$. This is different from the case of diffusions, where linear and quadratic functions $\langle X_t, f\rangle$ and $\langle X_t, f\rangle^2$ in a set $\mathcal{F}$ of test-functions suffice to establish both *tightness in path space* and *convergence of finite-dimensional distributions (f.d.d.s)*.

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 necessary to handle the jumps. We explain the basic structure of the argument in the next section.

5.2. *Convergence criteria.* In the proofs, we view the process with $G = \{0,1,\ldots,N-1\}$, $G = G_{N,K} = \{0,1,\ldots,N-1\}^K$ and $G = \Omega_N$ (cf. (1.21)) as embedded in the process with $G = \mathbb{N}$, $G = \mathbb{N}^K$ and $G = \Omega_\infty$, where

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

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

The claims we have to prove require us to show that certain sequences of probability measures $(P_n)_{n\in\mathbb{N}}$ on $D([0,\infty),E)$ converge to a specified limit P. Therefore we have to show

- tightness on $D([0,\infty),E)$,
- convergence of the f.d.d.'s to the ones of the claimed limit.

What we will use to establish tightness (and later also f.d.d.-convergence) is that the P_n's and P are solutions to martingale problems for measure-valued processes. We write $X^{(N)}, X$ to denote realisations of these processes.

The states of our processes are *probability* measures on the type space (recall (1.43) and (1.46)). We use Jakubowski's criterion for measure-valued processes (see Dawson (1993, Theorem 3.6.4)). This requires us to prove: (1) a compact containment condition for the path, i.e., for all $\epsilon, T > 0$ there exists a $K_{T,\epsilon}$ compact such that

$$\mathbb{P}(\{X^{(N)}(t) \in K_{T,\epsilon} \text{ for all } t \in [0,T]\}) \geq 1-\varepsilon; \tag{5.2}$$

(2) tightness of evaluation processes $(F(X^{(N)}(t)))_{t\geq 0}$ in path space for all $F \in \mathcal{D}$, with $\mathcal{D}$ a dense subspace of continuous functions on type space. We will use for $\mathcal{D}$ the set

$$\mathcal{D} = \{\langle X, f\rangle^n | f \in C_b(E,\mathbb{R}), \quad n \in \mathbb{N}\} \subseteq C_b(\mathcal{P}(E),\mathbb{R}). \tag{5.3}$$

In our setting, the compact containment condition in (1) is immediate, because we have a compact type space and the probability measures on it form a compact set in the weak topology. Condition (2) can be verified by using a criterion for tightness by Kurtz (see Dawson (1993, Corollary 3.6.3)). (Alternatively, we could use a tightness criterion by Joffe-Métivier Dawson (1993, Theorem 3.6.6 and Corollary 3.6.7).) In particular, we get that (2) follows from

$$\sup_{N\in\mathbb{N}} \|L^{(N)}F\|_\infty < \infty, \quad \forall\, F \in \mathcal{D}. \tag{5.4}$$

Thus, to conclude tightness, we have to calculate $L^{(N)}F$, for $F \in \mathcal{D}$, and bound it in the supremum norm.

In order to show f.d.d.-convergence of $X^{(N)}$ to the claimed limit X, we use that these measure-valued processes arise as the solution to the $(L^{(N)}, \mathcal{D}, \delta_{X_0^N})$-martingale problem, respectively, the $(L, \mathcal{D}, \delta_{X_0})$-martingale problem, where the latter is well-posed. It then suffices to show that, for a dense subset $\mathcal{A}$ of $C_{\mathrm{b}}((\mathcal{P}(E))^{\mathbb{N}},\mathbb{R})$ and all all $F \in \mathcal{A}$, the compensator terms satisfy:

$$\mathcal{L}\left[\left(\int_0^t L^{(G_N)}F((X_s^N))\mathrm{d}s\right)_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(\int_0^t (L^{(G)}F)(X_s)\mathrm{d}s\right)_{t\geq 0}\right] \tag{5.5}$$

and the initial laws satisfy

$$\mathcal{L}[X_0^{(N)}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[X_0]. \tag{5.6}$$

This allows us to conclude that $X^{(N)}$ converges in f.d.d. to X, so that we get (2).

Thus, to prove the convergence as claimed, we have to verify (5.4) and (5.5) for each of the three processes mentioned in Section 5.1. For the proof of (5.5), it is necessary to use the *duality relation*, in order to establish certain properties of the process $X^{(N)}$ in the limit as $N \to \infty$ that allow us to draw more information from the generator calculation. This includes a proof that certain higher-order terms can be discounted, or an argument that establishes independence over sufficiently large distances.

The averaging arguments we will use in the following sections are close in spirit to those in Kurtz (1992). In our case, however, the latter work does not apply immediately, in particular, because we deal with N-dependent state space.

In summary, the role of Sections 6–8 is to first carry out some generator calculations, leading to the bound in (5.4), and then an asymptotic evaluation of the resulting generator expressions, leading to a limiting form that uniquely determines the limiting process in (5.5). The latter will be based on a direct calculation. In view of the large time scales involved, we can use an *averaging principle* for local variables, based on the *local equilibria* dictated by the macroscopic slowly changing variables. The properties of the limiting process are established in Section 4.

6. The mean-field limit of C^Λ-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 Λ_0 in single colonies (cf. Section 1.3.2). 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 Dawson et al. (1995, Section 4). The reader unfamiliar with the arguments involved is referred to this paper (see, in particular, the outline of the abstract scheme in Dawson et al. (1995, 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 Λ-Cannings resampling process, i.e., the change from continuous to cádlág semi-martingales.

We always start the process in a product state with law $\chi^{\otimes N}$ with $\chi \in \mathcal{P}(\mathcal{P}(E))$ satisfying

$$\int_{\mathcal{P}(E)} x\chi(\mathrm{d}x) = \theta \in \mathcal{P}(E). \tag{6.1}$$

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^Λ-mean-field model from Section 1.3.2 with $G = \{0, 1, \dots, N-1\}$. We prove propagation of chaos for the collection

$$(\{X_0^{(N)}(t), \dots, X_{N-1}^{(N)}(t)\})_{t \geq 0} \tag{6.2}$$

in the limit as $N \to \infty$, i.e., we prove asymptotic independence of the components via duality as well as component-wise convergence to the McKean-Vlasov process with parameters $d_0 = 0, c_0, \Lambda_0, \theta$ (cf. (1.18)).

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

$$\mathcal{L}\left[(X_0^{(N)}(t), \dots, X_L^{(N)}(t))_{t \geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \bigotimes_{i=0}^{L} \mathcal{L}\left[Z_{i,\theta}^{c_0,d_0,\Lambda_0}\right], \tag{6.3}$$

where $Z_{i,\theta}^{c_0,d_0,\Lambda_0}$ solves the martingale problem for $(L_\theta^{c_0,d_0,\Lambda_0}, \mathcal{F}, \chi)$.

Corollary 6.2. **[McKean-Vlasov limit with block resampling]** *Consider the system above with an additional rate $N^{-2}\Lambda_1$ of block resampling per site. Then* (6.3) *continues to hold.*

In order to prove (6.3), 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. We show this first for components (Section 6.1.1). Then, 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.* Since we have a state in $(\mathcal{P}(E))^{\mathbb{N}}$ equipped with the product topology, it suffices to establish tightness for L-tuples of components. We focus first on one component $(X_\xi(t))_{t\geq 0}$ and conclude later the result for tuples of L-components.

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 (1.7), (1.12) and (1.14)). Namely, we will see in Section 6.1.3 (in (6.6), (6.16) and (6.17) below) that the generator $L^{(N)}F$ satisfies

$$\sup_{N\in\mathbb{N}} \|L^{(N)}F\|_\infty < \infty, \quad \text{for all } F \in C_b^2(\mathcal{P}(E), \mathbb{R}). \tag{6.4}$$

As we outlined in Section 5.2, this guarantees tightness.

6.1.2. *Asymptotic independence.* In this section, we use duality to prove the factorisation of spatial mixed moments (including the case with non-local coalescence at rate $N^{-2}\Lambda_1$). Namely, we show that for any $L \in \mathbb{N}$, any $k_\xi \in \mathbb{N}$, $\xi \in [L]$,

$$\limsup_{N\to\infty} \left| \mathbb{E}\left[\prod_{\xi=0}^{L} \left(\langle X_\xi^{(N)}(t), f_\xi \rangle \right)^{k_\xi} \right] - \prod_{\xi=0}^{L} \mathbb{E}\left[\left(\langle X_\xi^{(N)}(t), f_\xi \rangle \right)^{k_\xi} \right] \right| = 0, \text{for all } t \geq 0. \tag{6.5}$$

Similar to (6.5) decorrelation holds also for mixed moments at different time points.

Proof of (6.5)*:* Obviously, no non-local 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 starting at different sites 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 $G = G_{N,1} = \{0, 1, \ldots, N-1\}$ is

$$(L_{\rm mig}^{(N)} F)(x) = \frac{c_0}{N} \sum_{\xi,\zeta \in G_{N,1}} \int_E (x_\zeta - x_\xi)(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\xi}[\delta_a], \tag{6.6}$$

where $F \in \mathcal{F} \subset C_{\rm b}(\mathcal{P}(E)^N, \mathbb{R})$, with $\mathcal{F}$ the algebra of functions of the form (1.9). We rewrite (6.6) as

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

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 $N \to \infty$, $(L_{\rm mig}^{(N)} F)(x)$ only depends on the mean type measure θ of the initial state, i.e., it converges to

$$(L_\theta^{c_0} F)(x) \equiv c_0 \sum_{\xi \in \mathbb{N}_0} \int (\theta - x_\xi)(\mathrm{d}a) \frac{\partial F(x)}{\partial x_\xi}[\delta_a], \tag{6.8}$$

where we use for this generator acting on $C_{\rm b}(\mathcal{P}(E))^{\mathbb{N}}, \mathbb{R})$ the same notation we used for the McKean-Vlasov process with immigration-emigration on $\mathcal{P}(E)$ (cf. (1.17)). Furthermore, we show that

$$\theta \mapsto L_\theta^{c_0} F \in C_{\rm b}(\mathcal{P}(E), \mathbb{R}) \text{ is continuous for all } \theta \in \mathcal{P}(E). \tag{6.9}$$

To show the convergence, define

$$\mathbb{B}_\theta = \Big\{ x \in (\mathcal{P}(E))^{\mathbb{N}_0} \colon\ N^{-1} \sum_{\xi \in G_{N,1}} x_\xi \underset{N \to \infty}{\longrightarrow} \theta \Big\} \subseteq (\mathcal{P}(E))^{\mathbb{N}}, \tag{6.10}$$

and

$$\mathbb{B} = \bigcup_{\theta \in \mathcal{P}(E)} \mathbb{B}_\theta. \tag{6.11}$$

For $x \in \mathcal{P}(E)^{\mathbb{N}_0}$ and $n \in \mathbb{N}$, denote $x|_n = (x_0, x_1, \ldots, x_{n-1})$.

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

$$\mathcal{L}[X^{(N)}(t)](\mathbb{B}|_N) = 1 \quad \text{(respectively, } \mathcal{L}[X^{(N)}(t)(\mathbb{B}_\theta|_N)] = 1). \tag{6.12}$$

Indeed, as we will see in Section 6.2, the 1-block average $Y_{\xi,1}^{(N)}$ (recall (1.43)) evolves on time scale Nt. More precisely, $(Y_{\xi,1}^{(N)}(tN))_{t \geq 0}$ is tight in path space and therefore converges over a finite time horizon to the mean type measure θ of the initial state. In a formula (the right-hand side means a constant path):

$$\mathcal{L}[(Y_{\xi,1}^{(N)}(t))_{t \in [0,T]}] \underset{N \to \infty}{\Longrightarrow} \mathcal{L}[(\underline{\theta})_{t \in [0,T]}]. \tag{6.13}$$

Therefore, we have

$$|(L_{\rm mig}^{(N)} F)(x|_N) - (L_\theta^{c_0} F)(x)| \underset{N \to \infty}{\longrightarrow} 0, \quad \text{for all } x \in \mathbb{B}_\theta, \tag{6.14}$$

Hence, on the path space, by dominated convergence, we have

$$\mathcal{L}\left[\left(\left|\int_0^t (L^{(N)}_{\mathrm{mig}}F)(X^{(N)}(s))\mathrm{d}s - \int_0^t (L^{c_0}_{Y^{(N)}_{\xi,1}(s)}F)(X^{(N)}(s))\mathrm{d}s\right|\right)_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \delta_{\underline{0}}. \tag{6.15}$$

• **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 as in (1.9) depends on finitely many coordinates only)

$$|(L^{(N)}_{\mathrm{res}}F)(x|_N) - (L^{\Lambda_0}F)(x)| \underset{N\to\infty}{\longrightarrow} 0, \quad \text{for all } x \in (\mathcal{P}(E)^{\mathbb{N}}), \tag{6.16}$$

where

$$\begin{aligned}(L^{\Lambda_0}F)(x) \equiv \sum_{\xi\in\mathbb{N}_0} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\Big[F\big(x_0,\dots,x_{\xi-1},(1-r)x_\xi + r\delta_a,\\ x_{\xi+1},\dots,x_{N-1}\big) - F(x)\Big].\end{aligned} \tag{6.17}$$

Again, 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 $\mathcal{P}(E)$ (cf. (1.17)).

6.1.4. *Convergence to the McKean-Vlasov process.* In this section, we finally show the convergence of the mean-field C$^{\Lambda}$-process (see Section 1.3.2) to the McKean-Vlasov process (see Section 1.3.3) which was claimed in Proposition 6.1.

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

$$G(x_\xi) = \int_{E^n} x_\xi^{\otimes n}(\mathrm{d}u)\,\varphi(u) = \left\langle \varphi, x_\xi^{\otimes n}\right\rangle, \qquad n\in\mathbb{N}, \varphi\in C_{\mathrm{b}}(E^n,\mathbb{R}). \tag{6.18}$$

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

Let the initial condition $(X^{(\infty)}_\xi(0))_{\xi\in\mathbb{N}_0}$ be i.i.d. $\mathcal{P}(E)$-valued random variables with mean θ. Then each single component converges and the limiting coordinate process has generator (recall (1.17))

$$\begin{aligned}(L_\theta^{c_0,0,\Lambda_0}G)(x_\xi) =& c_0 \int_E (\theta - x_\xi)(\mathrm{d}a)\,\frac{\partial G(x_\xi)}{\partial x_\xi}[\delta_a]\\ &+ \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\left[G\big((1-r)x_\xi + r\delta_a\big) - G(x_\xi)\right],\end{aligned} \tag{6.19}$$

where $\theta \in \mathcal{P}(E)$ is the initial mean measure. Indeed, we may now reason as in Dawson (1993, 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 $(X^{(N_k)}_\xi(t))_{t\geq 0}$, $k\in\mathbb{N}$. 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 distribution $\mathcal{L}[X_\xi(0)]$. Recall from Section 6.1.3 that, for all test functions $F \in \mathcal{F}$,

$$\mathcal{L}\left[\left(\int_0^t (L_{\text{mig}}^{(N)} + L_{\text{res}}^{(N)})(F)(X^N(s))\mathrm{d}s\right)_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[\left(\int_0^t L_\theta^{c_0,d_0,\Lambda_0}(X^\infty(s))\mathrm{d}s\right)_{t\geq 0}\right] \tag{6.20}$$

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 of (6.20) is the compensator of a well-posed martingale problem (recall Proposition 1.2), and hence we have convergence (6.3).

6.2. *The mean-field finite-system scheme.* In this section, we verify the mean-field "finite system scheme" for the C^Λ-process, i.e., we consider $L+1$ tagged sites $\{X_0^{(N)}(t), \dots, X_L^{(N)}(t)\}$ evolving as in Section 1.3.2 and the corresponding block average $Y^{(N)}(t) = N^{-1}\sum_{\xi\in G_{N,1}} X_\xi^{(N)}(t)$. 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 θ (cf. Section 1.3.3 and recall (1.45) with $d_0 = 0$);
- convergence of the components $(\{X_\xi^{(N)}(Nt+u), \xi = 0, \dots, L\})_{u\geq 0}$ 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 θ,*

$$\mathcal{L}[(Y^{(N)}(Nt))_{t\geq 0}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(Z_\theta^{0,d_1,0}(t))_{t\geq 0}] \tag{6.21}$$

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

$$\mathcal{L}[(X_\xi^{(N)}(Nt+u))_{\xi=0,\dots,L}] \underset{N\to\infty}{\Longrightarrow} \int_{\mathcal{P}(E)} P_t(\mathrm{d}\theta')\left(\nu_{\theta'}^{c_0,d_0,\Lambda_0}\right)^{\otimes(L+1)} \quad \text{with } P_t = \mathcal{L}[Z_\theta^{0,d_1,0}(t)]. \tag{6.22}$$

Corollary 6.4. **[Mean-field finite system scheme with Λ_1-block resampling]** *Consider the model above with additional block resampling at rate $N^{-2}\Lambda_1$. Then, in the right-hand side of (6.21), $Z_\theta^{0,d_1,0}$ must be replaced by $Z_\theta^{0,d_1,\Lambda_1}$, and similarly in the definition of P_t in (6.22).*

The proof of the mean-field finite system scheme follows the abstract argument developed in Dawson et al. (1995). 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. With the help of the idea of local equilibria based on the ergodic theorems of Section 4, we obtain first (6.22) and then (6.21) in Section 6.2.4.

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, $M^{x,F} = (M_t^{x,F})_{t\geq 0}$,

$$M_t^{x,F} = F(x_t) - F(x_0) - \int_0^t \mathrm{d}s \left(L_{\text{mig}}^{(N)} F + L_{\text{res}}^{(N)} F \right)(x_s) \tag{6.23}$$

is a martingale for all F, as in (6.18). 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 (6.18) and

$$y = N^{-1} \sum_{\xi \in G_{N,1}} x_\xi \tag{6.24}$$

a block average (with $G_{N,1} = \{0, 1, \dots, N-1\}$). We will show below that $L_{\text{mig}}^{(N)}(F \circ y) = 0$, so that migration has no effect.

Recall $(L_{\text{mig}}^{(N)} F)(x)$ as rewritten in (6.7). 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 (6.18), we have

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

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

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

and hence

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

6.2.2. *From Λ-Cannings to Fleming-Viot.* Next, we evaluate the moment measures of the average (6.24) in the limit as $N \to \infty$ and show convergence of the terms to the Fleming-Viot second order term.

Remark 6.5 (Notation for the rescaled generators)*.* Given a generator L of a Markov process, we denote by $L^{[k]}$ (for $k \in \mathbb{N}$) the generator of the Markov process on time scale $N^k t$. Evidently, this time speed-up simply amounts to multiplication of the original generator L by N^k.

We are interested in the action of the rescaled generator $L_{\text{res}}^{(N)[1]}$ on the functions of the corresponding 1-block averages (6.24).

Lemma 6.6. [Generator convergence: resampling]
On time scale Nt, in the limit as $N \to \infty$,

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

Proof of Lemma 6.6: 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

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

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 (6.23). If we use (6.29) in (6.23) with x replaced by y, then we obtain that $(y_t)_{t\geq0}$ solves the martingale problem with generator

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

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

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

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: $|\{\xi_1,\ldots,\xi_n\}|=n$, i.e., all $\xi_i, 1\leq i\leq n$ 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

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

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 (as in (6.18)) only, and thereby on φ and n. Here we use (1.39) 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

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

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

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

Now use that

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

to obtain from (6.38), for $F(y)=\langle\varphi,y^{\otimes n}\rangle$, that

$$\begin{aligned}
&(L^{(N)}_{\mathrm{res}}F)(y)\\
&=\frac{1}{N^2}\sum_{1\le m_1<m_2\le n}\ \sum_{\xi\in G_{N,1}}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\\
&\quad\times\Bigg\langle\varphi, y_{\xi_1}\otimes\cdots\otimes\underbrace{(r(-x_\xi+\delta_a))}_{\text{only change position }\xi_{m_1}}\otimes\cdots\otimes\underbrace{(r(-x_\xi+\delta_a))}_{\text{and position }\xi_{m_2}}\otimes\cdots\otimes y_{\xi_n}\Bigg\rangle\\
&\quad+O\big(N^{-2}\big)\\
&=\frac{1}{N^2}\sum_{\xi\in G_{N,1}}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\\
&\quad\times\frac12\frac{\partial^2F(y)}{\partial y^2}[r(-x_\xi+\delta_a),r(-x_\xi+\delta_a)]+O\big(N^{-2}\big)\,.
\end{aligned} \tag{6.40}$$

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 $L^{(N)}_{\mathrm{res}}$ 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 Dawson et al. (1995, 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 (recall Section 4.2); (2) the ergodic theorem must hold uniformly in the initial state. Both coupling and duality do the job, which is why both work in Dawson et al. (1995).

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 (1.19) and the equilibrium ν of the McKean-Vlasov process in the line preceding (4.1). Observe that the compensators of $M^{x,F}$, see (6.23) 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

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

where θ is called the *intensity of the configuration* and

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

Lemma 6.7. [Local equilibrium]

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

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

(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*

$$\mathbb{P}(X^{(\infty)}(t,u) \in \mathbb{B}^*) = 1, \qquad \textit{for all } t \in [0,\infty),\ u \in \mathbb{R}. \tag{6.44}$$

Proof: (a) The proof uses the line of argument in Dawson et al. (1995, Section 4(d)) (recall the comment in Section 6.2.3), together with (4.21) 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 (6.19) 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 (4.21) 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) together with (6.27) to establish that $\|L_{\text{res}}^{(N)[1]}(F)\|_\infty$ is bounded in N, which gives the tightness (recall Section 5.2). To verify (i), we want to show that the weak limit points satisfy the $(\delta_\theta, L_\theta^{0,d_1,0})$-martingale problem. For that, we have to show that

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

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

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

Use the definition of the Fleming-Viot diffusion operator Q from (1.19) to obtain the first line of the claim in (6.43). The second line follows with the help of (4.21) (recall $d_0 = 0$ in this section).

(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 θ, 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 6.7 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

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

7. Hierarchical C^Λ-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 $k = 0, 1, \dots, K$ 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 $X^{(\Omega_N)}$ from Section 1.4.4 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 include also a block resampling at rate N^{-2K} with resampling measure Λ_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. As we saw in Lemma 6.7, 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. Many features from Section 6.2 reappear here, but we have to be aware that level-one averages are forming only asymptotically a mean-field system of the type we had in Section 6 and we have to prove that we can in fact ignore this perturbation. 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 Dawson et al. (1995). 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, λ_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 Λ_1; (2) For 1-block averages (each belonging to an address $\eta \in \{0, 1, \cdots, N-1\}$) we introduce the notation

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

Next, 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 to the one in Section 6. Namely, we can replace the system $(Y^{(N)}, Z^{(N)})$ for $N \to \infty$ by a system of the type in Section 6, where the role of components on time scale t is 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^2t. 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 Dawson et al. (1995, Section 5(a), pp. 2328–2337). The calculations in Sections 7.1.1–7.1.3 correspond to Steps 4–5 in Dawson et al. (1995, 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,*

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

and

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

and

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

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 θ of the law of the McKean-Vlasov process (cf., Section 4.2), 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 by means of establishing asymptotic independence and the *generator calculation* for the martingale problem. Since the latter is key also for the tightness arguments (recall (5.4)), 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 (1.37) with (1.26) applied to functions $F \in \mathcal{F}$, the algebra of functions in (1.34). The migration operator can be rewritten as (recall that the upper index 2 in $L^{(N,2)}$ indicates that we consider

$K = 2$ levels)

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

where we use (1.31) in the last line. Thus, for F as in (1.34), we obtain

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

where

$$|E^{(N)}| \le N^{-1}c_1 C_F = O\big(N^{-1}\big) \tag{7.10}$$

with C_F a generic constant depending on the choice of F only. Here we use that, by the definition of F in (1.34), 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. Recall (1.34). For $F \in \mathcal{F}$, consider the resampling operator $(L_{\mathrm{res}}^{(N,2)}F)(x)$ in (1.38)–(1.39). We have

$$(L_{\mathrm{res}}^{(N,2)}F)(x) = \sum_{\xi\in G_{N,2}} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a)\,\big[F\big(\Phi_{r,a,B_0(\xi)}(x)\big) - F(x)\big] + E^{(N)} \tag{7.11}$$

with

$$|E^{(N)}| \le N^{-2} \int_{[0,1]} \Lambda_1^*(\mathrm{d}r) C_F r^2 N = C_F N^{-1}\lambda_1 = O\big(N^{-1}\big)\,. \tag{7.12}$$

Here we use (1.39) 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 $C_F N$ 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

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

with Q_{x_ξ} as in (1.19). 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

$$\begin{aligned}(L^{(N,2)}F)(x) = &\sum_{\xi\in G_{N,2}} c_0 \int_E (y_{\xi,1}-x_\xi)\,(\mathrm{d}a)\,\frac{\partial F(x)}{\partial x_\xi}[\delta_a]\\ &+\sum_{\xi\in G_{N,2}}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\int_E x_\xi(\mathrm{d}a)\left[F\left(\Phi_{r,a,B_0(\xi)}(x)\right)-F(x)\right]\\ &+d_0\sum_{\xi\in G_{N,2}}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(x)}{\partial x_\xi^2}[\delta_u,\delta_v]+O\big(N^{-1}\big)\,,\end{aligned} \tag{7.14}$$

where $O\big(N^{-1}\big)$ is uniform in x.

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

$$\|L^{(N,2)}F-L^{c_0,d_0,\Lambda_0}_{y_{\xi,1}}F\|_\infty\leq C_FN^{-1},\quad \|L^{c_0,d_0,\Lambda_0}_{y_{\xi,1}}F\|\leq C(F),\quad n\in\mathbb{N},\, F\in\mathcal{F} \tag{7.15}$$

with $\mathcal{F}$ as in (1.34). Next, following again the line of argument in Section 5.2, we see that $\mathcal{L}[X^{(N)}]$ is tight in path space and, following the argument as 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

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

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) uniform boundedness (in N) of the generator in the supremum norm for test-functions in $\mathcal{F}$ to get tightness in path space of $(Y^{(N)}_\xi(Nt))_{t\geq0}$ (cf. (5.4)); (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 total average over the full space (cf. (7.4)) remains θ on time scale Nt, 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. 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 (1.31), which is the 1-block around ζ. This 1-block coincides with the 1-block around ξ if and only if $d(\zeta,\xi)\leq1$. To endow every 1-block with a unique label, we proceed as follows. Let ϕ 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.,

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

where we suppress the dependence of $y_\eta^{[1]}$ on N. Note in particular that

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

We will often drop the superscript [1] to lighten the notation.

This time, we consider functions $F \in \mathcal{F}$ (see (1.34)) applied to $y^{[1]} \equiv y^{[1]}(x)$, where $y^{[1]} = (y_\eta^{[1]})_{\eta \in G_{N,1}}$. Recall the $^{[k]}$-notation for the rescaled generators from Section 6.2.2. By explicit calculation of the different terms below, we will obtain the following expression (recall $\Phi_{r,a,\eta}$ from (1.39) and Q_{x_ξ} from (1.19)):

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

Here, we assumed that F can be written as follows: $F(y_t) = F(y_t^{[1]}) = \langle \varphi, \bigotimes_{l=1}^q y_{\eta^{(l)}}^{\otimes n_l} \rangle$ with $y = y^{[1]} = (y_\eta^{[1]})_{\eta \in G_{N,1}}$, $\eta^{(l)} \in G_{N,1}$, $q \in \{1, \ldots, N\}$ and $n_l \in \mathbb{N}, 1 \leq l \leq q$. We give more detail in (7.28) below.

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 (6.5), 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 showing that the action of the generators is uniformly bounded in N in the sup-norm on $\mathcal{F}$, so that we have convergence of the generator on $\mathcal{F}$ by the same tightness argument as used in Section 6.2.4, but now based on (7.20). Consider the resampling and Fleming-Viot parts of the generator in (7.20) separately.

Reason as in the proof of Lemma 6.7 to see that (recall the definition of $\nu^{c_0,d_0,\Lambda_0}_{y_\eta}$ from (4.1))

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

where by (4.21) the second assertion follows from the first. Recall (7.13). Similarly, we have

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

Using (4.21) once more, we get

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

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

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

for test-functions G of the form (6.18). Note that $\theta \in \mathcal{P}(E)$ 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 θ 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 $(L^{(N,2)}_{\mathrm{mig}} F)(x)$ from (7.8) and that the upper index 2 in $L^{(N,2)}$ indicates that we consider $K = 2$ levels. Let F be as in (1.34). Denote

$\phi^k \equiv \underbrace{\phi \circ \phi \circ \ldots \circ \phi}_{k \text{ times}}$. Proceeding along the lines of (6.25–6.27), we get

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

Next, for functions F that are linear combinations of functions in (1.34), 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 (recall that the upper index [1] indicates time scale $N^1 t$)

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

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 $F(y_t) = F(y^{[1]}_t) = \langle \varphi, \bigotimes_{l=1}^q y^{\otimes n_l}_{\eta^{(l)}} \rangle$ with $y = y^{[1]} = (y^{[1]}_\eta)_{\eta \in G_{N,1}}$, $\eta^{(l)} \in G_{N,1}$, $q \in \{1, \ldots, N\}$ and $n_l \in \mathbb{N}, 1 \le l \le q$, instead of restricting ourselves to test-functions of the form (6.29) (which corresponds to the case $q = 1$). We will now use functions F of the form

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

The only difference with (1.34) 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

$$\begin{aligned}&(L^{(N,2)[1]}_{\mathrm{res}} F)(y) \\ &= \sum_{m=1}^q \frac{1}{N} \sum_{\xi:\ \phi\xi = \eta^{(m)}} \int_{[0,1]} \Lambda^*_0(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a) \frac{1}{2} \frac{\partial^2 F(y)}{\partial y^2_{\eta^{(m)}}}[r(-x_\xi + \delta_a), r(-x_\xi + \delta_a)] \\ &\quad + \sum_{\eta \in G_{N,1}} \int_{[0,1]} \Lambda^*_1(\mathrm{d}r) \int_E y_\eta(\mathrm{d}a) \left[F\left(\Phi_{r,a,\eta}(y)\right) - F(y)\right] + O\big(N^{-1}\big)\end{aligned} \tag{7.29}$$

with $\Phi_{r,a,\eta}$ as in (1.39).

Recall the notation in (7.28) and set

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

Proceeding as in (6.29-6.31), we obtain

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

with $F^{\left(\xi_1^1,\dots,\xi_{n_q}^q\right)}$ as in (6.30). As in Section 6.2.2, 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 (1.38)–(1.39),

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

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 $|\{\xi_i^l, 1 \le l \le q, 1 \le i \le n_l\}| = L-1$. It remains to investigate the terms with $|\{\xi_i^l, 1 \le l \le q, 1 \le i \le n_l\}| = L-1$. Since $\phi\xi_i^l = \eta^{(l)}$, this implies that there exist $1 \le m \le q$ and $1 \le m_1 < m_2 \le n_m$ such that $\xi_{m_1}^m = \xi_{m_2}^m$ and all other ξ_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

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

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

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

For the second term I_1 in (7.32), we obtain, by the definition of $\Phi_{r,a,B_1(\xi)}(x)$ in (1.39) and using (7.19),

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

with

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

Now observe that the sum $\sum_{\eta\in G_{N,1}}$ in (7.35) yields non-zero contributions only for $\eta\in\left\{\eta^{(1)},\dots,\eta^{(q)}\right\}$, and so we can rewrite I_1 as

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

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

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

Additional Fleming-Viot part. We proceed as with the migration operator (recall that in the present Section 7 we added a Fleming-Viot term to the generator, i.e., we consider the case $d_0 > 0$) and write

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

with Q_{x_ξ} as in (1.19) and where the definition of $y = y^{[1]}$ in (7.18) yields

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

Hence, on time scale Nt,

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

where in the last line we use that, for F a linear combination of the functions in (1.34),

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

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 (recall $y_\eta^{[1]}$ from (7.18))

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

(This is a 2-block average because we are considering the case $K = 2$.) Recall notation (7.4). We must prove: (1) the sequence of laws $\{\mathcal{L}[(Z^{(N)}(tN^2))_{t\geq 0}, N \in \mathbb{N}\}$ is *tight* in path space; (2) the weak limit points of this sequence are solutions of the martingale problem for $Z_\theta^{0,d_2,0}$ (cf. (7.7)) by showing (5.5) (recall Section 5.2). From the uniqueness of the solution to the martingale problem, we get the claim.

We now verify these points by calculating the generator. Recall the $^{[k]}$-notation from Section 6.2.2 for the rescaled generators.

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), (cf., (7.17) for the definition of ϕ)

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

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

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

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

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

with $\Phi_{r,a,B_k(\xi)}$ as in (1.39).

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

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

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

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

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 $|\eta_1,\dots,\eta_n| < n$. 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 sites residing in a common 1-block already result in a factor of $O(N^{-2})$ (on time scale t): first a common block has to be chosen ($|\eta_1,\dots,\eta_n| = n-1$), 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 on time scale t

$$\begin{aligned}(I_0')^{[0]} = \frac{1}{N^2}\sum_{\eta\in G_{N,1}}\frac{1}{N^2}\sum_{\xi:\ \phi\xi=\eta}\int_{[0,1]}\Lambda_0^*(\mathrm{d}r)\\ \times\int_E x_\xi(\mathrm{d}a)\frac{1}{2}\frac{\partial^2 F(z)}{\partial z^2}[r(-x_\xi+\delta_a), r(-x_\xi+\delta_a)] + O\big(N^{-3}\big)\,.\end{aligned} \tag{7.49}$$

Additional Fleming-Viot part. We proceed as for the migration operator. Recall (7.40), to get

$$(L_{\mathrm{FV}}^{(N,2)[1]}F)(z) = d_0\sum_{\eta\in G_{N,1}}\frac{1}{N}\sum_{\xi:\ \phi\xi=\eta}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2(F\circ z)(y)}{\partial y_\eta^2}[\delta_u,\delta_v]. \tag{7.50}$$

Now use the analogue to (7.39), to obtain

$$(L_{\mathrm{FV}}^{(N,2)[1]}F)(z) = d_0\sum_{\eta\in G_{N,1}}\frac{1}{N}\sum_{\xi:\ \phi\xi=\eta}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(z)}{\partial z^2}\left[\frac{\delta_u}{N},\frac{\delta_v}{N}\right]. \tag{7.51}$$

After changing to time scale N^2t, we have

$$(L_{\mathrm{FV}}^{(N,2)[2]}F)(z) = d_0\frac{1}{N}\sum_{\eta\in G_{N,1}}\frac{1}{N}\sum_{\xi:\ \phi\xi=\eta}\int_E\int_E Q_{x_\xi}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2 F(z)}{\partial z^2}[\delta_u,\delta_v]. \tag{7.52}$$

Tightness. We have to bound the generator, i.e., show that $\sup_N\|L^{(N,2)[1]}(F)\|<\infty$, in order to apply the tightness criterion, as explained in Section 5.2. (Recall that the upper index [1] indicates time scale N^1t and that the upper index 2 indicates that we consider $K=2$ levels.) This we read off from (7.44), (7.46), (7.47), (7.49) and (7.52).

Convergence to McKean-Vlasov process. We have to identify the limiting generator. 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), using (7.49), now on time scale tN^2, together with (4.21) in the second and fourth equation,

$$
\begin{aligned}
\lim_{N\to\infty}(I_0')^{[2]} &= \frac{\lambda_0}{2}\lim_{N\to\infty}\frac{1}{N}\\
&\quad\times\sum_{\eta\in G_{N,1}}\int_{\mathcal{P}(E)}\nu_{y_\eta}^{c_0,d_0,\Lambda_0}(\mathrm{d}x)\int_E\int_E Q_x(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z^2}[\delta_u,\delta_v]\\
&= \frac{c_0\lambda_0}{2c_0+\lambda_0+2d_0}\lim_{N\to\infty}\frac{1}{N}\sum_{\eta\in G_{N,1}}\int_E\int_E Q_{y_\eta}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z^2}[\delta_u,\delta_v]\\
&= \frac{c_0\lambda_0}{2c_0+\lambda_0+2d_0}\int_{\mathcal{P}(E)}\nu_z^{c_1,d_1,\Lambda_1}(\mathrm{d}y)\int_E\int_E Q_y(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z^2}[\delta_u,\delta_v]\\
&= \frac{2c_1}{2c_1+\lambda_1+2d_1}\,\frac{c_0\lambda_0}{2c_0+\lambda_0+2d_0}\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z^2}[\delta_u,\delta_v].
\end{aligned}
\tag{7.53}
$$

Combine (7.48) with (7.53), to get from (7.45)

$$
\begin{aligned}
&\lim_{N\to\infty}(L_{\mathrm{res}}^{(N,2)[2]}F)(z)\\
&= \frac{2c_1}{2c_1+\lambda_1+2d_1}\left(\frac{\lambda_1}{2}+\frac{c_0\lambda_0}{2c_0+\lambda_0+2d_0}\right)\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial z^2}[\delta_u,\delta_v].
\end{aligned}
\tag{7.54}
$$

For the Fleming-Viot part in (7.52), we obtain, by reasoning once more as in (7.21), using (4.21),

$$
\begin{aligned}
&\lim_{N\to\infty}(L_{\mathrm{FV}}^{(N,2)[2]}F)(z)\\
&= d_0\lim_{N\to\infty}\frac{1}{N}\sum_{\eta\in G_{N,1}}\int_{\mathcal{P}(E)}\nu_{y_\eta}^{c_0,d_0,\Lambda_0}(\mathrm{d}x)\int_E\int_E Q_x(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial_z^2}[\delta_u,\delta_v]\\
&= \frac{2c_0d_0}{2c_0+\lambda_0+2d_0}\lim_{N\to\infty}\frac{1}{N}\sum_{\eta\in G_{N,1}}\int_E\int_E Q_{y_\eta}(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial_z^2}[\delta_u,\delta_v]\\
&= \frac{2c_0d_0}{2c_0+\lambda_0+2d_0}\int_{\mathcal{P}(E)}\nu_z^{c_1,d_1,\Lambda_1}(\mathrm{d}y)\int_E\int_E Q_y(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial_z^2}[\delta_u,\delta_v]\\
&= \frac{2c_1}{2c_1+\lambda_1+2d_1}\,\frac{2c_0d_0}{2c_0+\lambda_0+2d_0}\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial_z^2}[\delta_u,\delta_v].
\end{aligned}
\tag{7.55}
$$

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

$$
\begin{aligned}
&\lim_{N\to\infty}(L^{(N,2)[2]}F)(z)\\
&= \frac{2c_1}{2c_1+\lambda_1+2d_1}\left(\frac{\lambda_1}{2}+\frac{c_0\lambda_0+2c_0d_0}{2c_0+\lambda_0+2d_0}\right)\int_E\int_E Q_z(\mathrm{d}u,\mathrm{d}v)\,\frac{\partial^2F(z)}{\partial_z^2}[\delta_u,\delta_v].
\end{aligned}
\tag{7.56}
$$

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 the generator of the total average on time scale tN^2 with $K = 2$ levels, $(L^{(N,2),[2]}F)(z)$, as an average over N different 1-block averages. If we replace the $\big(L^{(N,2)[1]}_{\mathrm{res},0} + L^{(N,2)[1]}_{\mathrm{FV}}\big)$-part of the 1-block averages (cf. (7.20)) 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 λ_0, cf. (7.6)), which on time scale Nt lead to the generator

$$L^{(N,2),[1]}_{\mathrm{mig}}(F)(y) + \frac{c_0(\lambda_0 + 2d_0)}{2c_0 + \lambda_0 + 2d_0} \int\limits_E \int\limits_E Q_y(\mathrm{d}u, \mathrm{d}v) \frac{\partial^2 F(y)}{\partial y^2}[\delta_u, \delta_v] + (L^{(N,2),[1]}_{\mathrm{res},1}F)(y), \tag{7.57}$$

then we can apply the analysis of Section 6 to this new collection of processes, denoted by

$$\left\{ \widetilde{Y}^{(N)}_i(tN) \colon i = 1, \ldots, N \right\}, \tag{7.58}$$

to conclude that on time scale tN^2 the block average $\widetilde{Z}^{(N)}(tN) = N^{-1} \sum\limits_{i=1}^{N} \widetilde{Y}^N_i(Nt)$ satisfies,

$$\mathcal{L}[(\widetilde{Z}^{(N)}(tN^2))_{t \geq 0}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(\widetilde{Z}(t))_{t\geq 0}], \tag{7.59}$$

where $\widetilde{Z}$ is a Fleming-Viot diffusion with resampling constant

$$\frac{c_1}{2c_1 + \lambda_1 + 2d_1}(\lambda_1 + 2d_1), \text{ where } d_1 = \frac{c_0(\lambda_0 + 2d_0)}{2c_0 + \lambda_0 + 2d_0}. \tag{7.60}$$

Hence, we obtain a limit process with a generator acting on F as

$$\frac{c_1(\lambda_1 + 2d_1)}{2c_1 + \lambda_1 + 2d_1} \int_E \int_E Q_z(\mathrm{d}u, \mathrm{d}v) \, \frac{\partial^2 F(z)}{\partial^2_z}[\delta_u, \delta_v]. \tag{7.61}$$

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 $(L^{0,d_2,0}_\theta G)(z)$ 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 $L^{0,d_2,0}$ and for the test functions given in (1.34) is well-posed (recall Proposition 1.2), 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

$$\text{replace } \{((Y^{(N)}_i(Nt))_{i=1,\ldots,N})_{t\geq 0}\} \text{ by } \{(\widetilde{Y}^{(N)}_i(Nt)_{i=1,\ldots,N})_{t\geq 0}\}. \tag{7.62}$$

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 $Z(s)$,

$$\mathcal{L}[((Y^{(N)}_i(N^2 s + Nt))_{i=1,\ldots,N})_{t\geq 0}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[((Y^{(\infty)}_i(s,t))_{i\in\mathbb{N}})_{t\geq 0}], \tag{7.63}$$

where the right-hand side is the McKean-Vlasov process with Fleming-Viot part at rate d_1, Cannings part Λ_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

$$F(z) = G_N(y) = H_N(x), \quad x \in (\mathcal{P}(E))^{N^2}, \quad y \in (\mathcal{P}(E))^N, \quad z \in \mathcal{P}(E), \tag{7.64}$$

with

$$z = \frac{1}{N} \sum_{i\in\{1,\ldots,N\}} y_i, \quad y_i = \frac{1}{N} \sum_{j\in\{1,\ldots,N\}} x_{j,i}. \tag{7.65}$$

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,

$$\begin{aligned}
&\mathcal{L}\Big[\Big(\int_0^{tN^2} \mathrm{d}s\,\Big[\int_{E\times E} d_1 \sum_{i=1}^N Q_{y_i(s)}(\mathrm{d}u,\mathrm{d}v)\frac{\partial^2 G_N(y(s))}{\partial y_i^2}[\delta_u,\delta_v] \\
&\qquad\qquad + L_{\mathrm{res},1}^{(N,2)[1]} G_N(y(s))\Big]\Big)_{t\geq 0}\Big] \\
&\quad - \mathcal{L}\Big[\Big(\int_0^{tN^2} \mathrm{d}s\,\Big[L_{\mathrm{res},1}^{(N,2)[1]} G_N(y_j(s)) \\
&\qquad\qquad + \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \Big(L_{\mathrm{res},0}^{(N,2)} + L_{\mathrm{FV}}^{(N,2)}\Big) H_N(x_{j,i}(s))\Big]\Big)_{t\geq 0}\Big] \\
&\underset{N\to\infty}{\Longrightarrow} \text{Zero measure.}
\end{aligned} \tag{7.66}$$

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 Dawson et al. (1995), 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 then 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 d_{j-1} instead of 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 Ethier and Kurtz (1986, Theorem 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 1.5. Recall the $C_N^{\underline{c},\underline{\Lambda}}$-process on Ω_N, denoted $X^{(\Omega_N)}$ from Section 1.4.4 and (1.43). Also recall the discussion on convergence criteria from Section 5.2. We establish the tightness by checking the bound on the generator action. Having Section 7, all we need is to show that the higher-order term action on monomials is bounded in N in the considered time scale. This is readily checked from the explicit form of the terms. In order to show convergence of the finite dimensional distribution, 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 the k-block averages $Y_{\xi,k}^{(\Omega_N)}$ and $Y_{\xi,k}^{(N,K)}$ on time scale $tN^j + s_N N^k$ for $0 \le k \le j < K$. Then*

$$d_{\mathrm{Prokh}}\left(\mathcal{L}\left[\left(Y_{\xi,k}^{(\Omega_N)}(tN^j + s_N N^k)\right)\right], \mathcal{L}\left[\left(Y_{\xi,k}^{(N,K)}(tN^j + s_N N^k)\right]\right)\right) \underset{N\to\infty}{\Longrightarrow} 0, \quad (8.1)$$

where d_{Prokh} is the Prokhorov metric.

Once we have proved this proposition, we obtain Theorem 1.5 by observing that (8.1) allows us to replace our system on Ω_N by the one on $G^{N,K}$ when we are interested only in block averages of order $\le K$ 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 $k \le j < K$. Thus, it remains only to prove Proposition 8.1. We give the proof for $K = 2$, and later indicate how to extend it 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 $(V_t)_{t\ge 0}$ and $(U_t)_{t\ge 0}$ (see, e.g., Ethier and Kurtz (1986, Section 1, (5.19))):

$$V_t = U_t + \int_0^t U_{t-s}(L_V - L_U)V_s \mathrm{d}s. \quad (8.2)$$

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 $F \in \mathcal{F}$ (cf. (1.34)) that depends only on $\{x_\xi,\, \xi \in B_1(0)\}$ (cf., (1.23)), we have (as we will see below)

$$(L^{(\Omega_N)}F)(x) = (L^{(N,2)}F)(x) + (L^{\mathrm{err}}F)(x), \quad (8.3)$$

where $\|L^{\mathrm{err}}\| = O(N^{-1})$ ($\|\cdot\|$ is the operator norm generated by the sup-norm). By the formula of partial integration for semigroups, it follows that

$$\left|\mathbb{E}\left[F(X^{(\Omega_N)}(t))\right] - \mathbb{E}\left[F(X^{(N,2)}(t))\right]\right| \le tO(N^{-1}). \quad (8.4)$$

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 (1.37) applied to functions $F \in \mathcal{F}$. The migration operator can be rewritten, similarly as in (7.8),

$$(L_{\mathrm{mig}}^{(\Omega_N)} F)(x) = \sum_{\xi \in \Omega_N} \sum_{k \in \mathbb{N}} c_{k-1} N^{1-k} \int_E (y_{\xi,k} - x_\xi)\,(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\xi}[\delta_a]. \tag{8.5}$$

We obtain

$$(L_{\mathrm{mig}}^{(\Omega_N)} F)(x) = \sum_{\xi \in \Omega_N} c_0 \int_E (y_{\xi,1} - x_\xi)\,(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\xi}[\delta_a] + E^{(N)}, \tag{8.6}$$

where

$$|E^{(N)}| \le N^{-1} C_F \sum_{k \in \mathbb{N}\setminus\{1\}} c_{k-1} N^{2-k}, \tag{8.7}$$

with C_F a generic constant depending on the choice of F only. Here we use that, by the definition of F in (1.34), the sum over $\xi \in \Omega_N$ is a sum over finitely many coordinates only, with the number depending on F only. By (1.27) we get

$$|E^{(N)}| \le O\big(N^{-1}\big)\,. \tag{8.8}$$

For the resampling operator in (1.38), applying first (1.39) and then (1.32), we obtain,

$$(L_{\mathrm{res}}^{(\Omega_N)} F)(x) = \sum_{\xi \in \Omega_N} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a) \left[F\left(\Phi_{r,a,B_0(\xi)}(x)\right) - F(x)\right] + E^{(N)} \tag{8.9}$$

with

$$|E^{(N)}| \le \sum_{k \in \mathbb{N}} N^{-2k} \int_{[0,1]} \Lambda_k^*(\mathrm{d}r) C_F N^k r^2 = C_F \sum_{k \in \mathbb{N}} N^{-k} \lambda_k = O\big(N^{-1}\big)\,. \tag{8.10}$$

Finally, the Fleming-Viot operator reads as in (7.13):

$$(L_{\mathrm{FV}}^{(\Omega_N)} F)(x) = d_0 \sum_{\xi \in \Omega_N} \int_E \int_E Q_{x_\xi}(\mathrm{d}u, \mathrm{d}v)\, \frac{\partial^2 F(x)}{\partial x_\xi^2}[\delta_u, \delta_v]. \tag{8.11}$$

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

$$\begin{aligned}
(L^{(\Omega_N)} F)(x) = &\sum_{\xi \in \Omega_N} c_0 \int_E (y_{\xi,1} - x_\xi)\,(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\xi}[\delta_a] + O(N^{-1}) \\
&+ \sum_{\xi \in \Omega_N} \int_{[0,1]} \Lambda_0^*(\mathrm{d}r) \int_E x_\xi(\mathrm{d}a) \left[F\left(\Phi_{r,a,B_0(\xi)}(x)\right) - F(x)\right] + O\big(N^{-1}\big) \\
&+ d_0 \sum_{\xi \in \Omega_N} \int_E \int_E Q_{x_\xi}(\mathrm{d}u, \mathrm{d}v)\, \frac{\partial^2 F(x)}{\partial x_\xi^2}[\delta_u, \delta_v].
\end{aligned} \tag{8.12}$$

Combining (8.12) with (8.5–8.11) and (7.14) (also recall the discussion on embeddings from Section 5.2), we get (8.3).

8.2. *The* 1*-block averages on time scale* Nt*.* As before, we prove, for $F \in \mathcal{F}$ depending on $\{x_\xi,\ \xi \in B_1(0)\}$ only (recall that the upper index [1] indicates time scale $N^1 t$ and that the upper index 2 indicates that we consider $K=2$ levels),

$$(L^{(\Omega_N)[1]})(y) = (L^{(N,2)[1]}F)(y) + O(N^{-1}) \tag{8.13}$$

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

$$(L^{(\Omega_N)}_{\mathrm{mig}}F)(y) = \frac{1}{N}\sum_{\eta\in\Omega_N}\sum_{k\in\mathbb{N}} c_k N^{1-k}\int_E \Big(y^{[1]}_{\phi^k\eta} - y_\eta\Big)(\mathrm{d}a)\frac{\partial F(y)}{\partial y_\eta}[\delta_a]. \tag{8.14}$$

We obtain on the time scale Nt

$$(L^{(\Omega_N)[1]}_{\mathrm{mig}}F)(y) = \sum_{\eta\in\Omega_N} c_1 \int_E \Big(y^{[1]}_{\phi\eta} - y_\eta\Big)(\mathrm{d}a)\frac{\partial F(y)}{\partial y_\eta}[\delta_a] + E^{(N)}, \tag{8.15}$$

where

$$\Big|E^{(N)}\Big| \leq C_F \sum_{k\in\mathbb{N}\backslash\{1\}} c_k N^{1-k} = O\big(N^{-1}\big)\,. \tag{8.16}$$

Note that, by (7.27),

$$(L^{(\Omega_N)[1]}_{\mathrm{mig}}F)(y) = (L^{(N,2)[1]}_{\mathrm{mig}}F)(y) + O\big(N^{-1}\big)\,. \tag{8.17}$$

For the resampling operator, the only change to (7.31) is that (7.32) gets replaced by

$$(L^{(\Omega_N)}_{\mathrm{res}}F)(y) = I_0 + I_1 + E^{(N)} \tag{8.18}$$

with I_0, I_1 as in (7.32) (with $G_{N,2}$ replaced by Ω_N) and

$$\begin{aligned}\Big|E^{(N)}\Big| &\leq \frac{1}{N^L}\left(\bigotimes_{l=1}^{q}\bigotimes_{i=1}^{n_l}\sum_{\xi^l_i:\ \phi\xi^l_i=\eta^{(l)}}\right)\sum_{k\in\mathbb{N}\backslash\{1\}} N^{-2k}\int_{[0,1]}\Lambda^*_k(\mathrm{d}r)LN^kC_Fr^2\\ &= C_F\sum_{k\in\mathbb{N}\backslash\{1\}} N^{-k}\lambda_k = O\big(N^{-2}\big)\,.\end{aligned} \tag{8.19}$$

After a change to time scale Nt, we therefore have

$$(L^{(\Omega_N)[1]}_{\mathrm{res}}F)(y) = (L^{(N,2)[1]}_{\mathrm{res}}F)(y) + O\big(N^{-1}\big) \tag{8.20}$$

with $(L^{(N,2)}_{\mathrm{res}}F)(y)$ 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),

$$(L^{(\Omega_N)[1]}_{\mathrm{FV}}F)(y) = (L^{(N,2)[1]}_{\mathrm{FV}}F)(y). \tag{8.21}$$

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 $N^K t$, 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 interested reader.

9. Multiscale analysis

9.1. *The interaction chain.* In this section, we prove Theorem 1.6. In addition to Theorem 1.5, what is needed is the convergence of the *joint law* of the collection of k-level block averages for $k = 0, \ldots, j+1$ 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 ℓ-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 $0 \le k \le j$. We also have the convergence of the marginal distributions for each $k = 0, \ldots, j+1$, 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 θ 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 Dawson et al. (1995, 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 9.3), as well as a more detailed description of the properties of the different regimes (Section 10), as discussed in Section 1.5.2.

9.2. *Dichotomy for the interaction chain.* In this section, we prove Theorem 1.7.

Proof of Theorem 1.7. Fix $j \in \mathbb{N}_0$. The first observation is that the interaction chain $(M_k^{(j)})_{k=-(j+1),\ldots,0}$ from Section 1.5.2 is a $\mathcal{P}(E)$-valued Markov chain such that

$$\big(\langle M_k^{(j)}, \varphi\rangle\big)_{k=-(j+1),\ldots,0} \text{ is a } \textit{square-integrable martingale}, \text{ for any } \varphi \in C_\mathrm{b}(E) \tag{9.1}$$

(because it is bounded). For the analysis of the interaction chain for Fleming-Viot diffusions, carried out in Dawson et al. (1995, Section 6), this fact was central in combination with the formula for the variance of evaluations analogous to Proposition 4.4. We argue as follows.

Since the map $\theta \mapsto \nu_\theta^{c,d,\Lambda}$ is continuous (cf. Section 4.2), the convergence as $j \to \infty$ in the *local coexistence* regime is a standard argument (see Dawson et al. (1995, Section 6a)). In the *clustering* regime, the convergence to the mono-type state follows by showing, with the help of the variance formula (4.26), that $\lim_{j\to\infty} \mathbb{E}_{\mathcal{L}(M_0^{(j)})}[\mathrm{Var.}(\varphi)] = 0$ for all $\varphi \in C_\mathrm{b}(E)$ (cf., Corollary 4.5), so that all limit points of $\mathcal{L}[M^{(j)}]$ are concentrated on δ-measures on E (recall that $\mathcal{P}(E)$ is compact). This argument is identical to the one in Dawson et al. (1995, 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 $\mathbb{E}_{\mathcal{L}(M_0^{(j)})}[\mathrm{Var.}(\varphi)]$ is bounded away from zero, the limit points allow for the coexistence of types. The argument in

Dawson et al. (1995, Section 6a) shows that for $\Lambda = 0$,

$$\nu_\theta^{c,d,\Lambda}(M) = 0 \text{ if } d > 0, \quad M = \{\delta_u \colon u \in E\}. \tag{9.2}$$

This is no longer true for $\Lambda \neq 0$. Instead, we have $\nu_\theta^{c,d,\Lambda}(M) \in [0,1)$, as proven in Section 4.3 (see (4.13)), and hence the variance is > 0.

9.3. *Scaling for the interaction chain.* In this section, we prove Theorems 1.16 and 1.17

The proof of the scaling result in the regime of diffusive clustering in Dawson et al. (1995, Section 6(b), Steps 1–3) uses two ingredients:

(I) Assertion (9.1).

(II) For $c_k \to c \in (0,\infty)$ as $k \to \infty$, by Dawson et al. (1995, Eq. (6.12)),

$$\mathrm{Var}\left(\langle M_{k_2}^{(j)}, f\rangle \mid M_{k_1}^{(j)} = \theta\right) = \frac{(-k_1) - (-k_2) + 1}{c + (-k_1)}\, \mathrm{Var}_\theta(f), \qquad \forall\, f \in C_\mathrm{b}(E,\mathbb{R}). \tag{9.3}$$

In Dawson et al. (1995, Section 6(b)), (I–II) led to the conclusion that if $\lim_{j\to\infty}(-k_j)/j = \bar\beta_i \in [0,1]$, $i = 1,2$, with $\bar\beta_1 > \bar\beta_2$, then

$$\lim_{j\to\infty} \mathrm{Var}\left(\langle M_{k_2}^{(j)}, f\rangle \mid M_{k_1}^{(j)} = \theta\right) = \frac{\bar\beta_1 - \bar\beta_2}{\bar\beta_1}\, \mathrm{Var}_\theta(f). \tag{9.4}$$

Thus, as soon as we have these formulae, we get the claim by repeating the argument in Dawson et al. (1995, Section 6(b)), which includes the time transformation $\bar\beta = e^{-s}$ in Step 3 to obtain a time-homogeneous expression from (9.4).

We know the necessary first and second moment formulae from Section 4.4. Replace Dawson et al. (1995, Eq. (6.12)) by (4.28), to see that we must make sure that

$$\lim_{j\to\infty} \sum_{i=\lfloor\bar\beta_2 j\rfloor}^{\lfloor\bar\beta_1 j\rfloor} \left(\frac{d_{i+1}}{c_i} \prod_{l=i+1}^{\lfloor\bar\beta_1 j\rfloor} \frac{1}{1+m_l} \right) = 1 - \left(\frac{\bar\beta_2}{\bar\beta_1}\right)^R \tag{9.5}$$

(recall (1.45) and (1.57) for the definition of d_k and m_k). Note that (9.5) remains valid also for $\bar\beta_2 = 0$.

Moreover, by following the reasoning in Dawson et al. (1995, Section 6(b), Step 4), we obtain by using (4.28) instead of Dawson et al. (1995, (6.34)) that

$$\left\{ \begin{array}{c} \text{fast growing clusters} \\ \text{slowly growing clusters} \end{array} \right\} \quad \text{if} \quad \sum_{i=n}^{m} \left(\frac{d_{i+1}}{c_i} \prod_{l=i+1}^{m} \frac{1}{1+m_l} \right) \left\{ \begin{array}{c} \to 0 \\ \to 1 \end{array} \right\} \tag{9.6}$$

when $m, n \to \infty$ such that $n/m \to \alpha$, for all $\alpha \in (0,1)$.

Proof of Theorem 1.16: The proof follows by inserting the asymptotics of c_k, d_k and m_k obtained in Theorem 1.12 and Corollary 1.13 into (9.5) or (9.6).

(i) In Cases (a) and (b), the asymptotics in (1.74–1.75) and (1.82) imply

$$\sum_{i=\lfloor\alpha m\rfloor}^{m} \left(\frac{d_{i+1}}{c_i} \prod_{l=i+1}^{m} \frac{1}{1+m_l} \right) = O\left(\mathrm{e}^{-Cm}\right), \quad C > 0. \tag{9.7}$$

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

$$\sum_{i=\lfloor \alpha m\rfloor}^{m} \left(\frac{d_{i+1}}{c_i} \prod_{l=i+1}^{m} \frac{1}{1+m_l} \right) \to 0. \tag{9.8}$$

(ii) In Case (d), for any $\varepsilon > 0$ and l large enough we have $|m_l - R/l| \le \varepsilon R/l$. This implies

$$\prod_{l=i+1}^{\lfloor \bar\beta_2 j\rfloor} \frac{1}{1+m_l} = \exp\left[- \sum_{l=i+1}^{\bar\beta_1 j} \left(\frac{R}{l} + O(m_l^2) \right) \right]. \tag{9.9}$$

Since $d_{i+1}/c_i \sim R/i$ and $m_l = O(1/l)$, it follows that

$$\sum_{i=\lfloor \bar\beta_2 j\rfloor}^{\lfloor \bar\beta_1 j\rfloor} \left(\frac{d_{i+1}}{c_i} \prod_{l=i+1}^{\lfloor \bar\beta_2 j\rfloor} \frac{1}{1+m_l} \right) \sim \sum_{i=\lfloor \bar\beta_2 j\rfloor}^{\lfloor \bar\beta_1 j\rfloor} \frac{R}{i} \left(\frac{\bar\beta_1 j}{i} \right)^{-R} \to 1 - \left(\frac{\bar\beta_2}{\bar\beta_1} \right)^{R}. \tag{9.10}$$

□

Proof of Theorem 1.17: In Case (A), $m_k \to \infty$, which by (9.6) implies fast clustering. In Case (B), $m_k \to \bar K + \bar M > 0$, which also 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 1.8–1.9.

Proof of Theorem 1.8: 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 carried out in Section 2.4.2. □

Proof of Theorem 1.9: We first note that the set of functions

$$\left\{ H_\varphi^{(n)}(\cdot, \pi_{G,n})\colon\ n\in\mathbb{N},\, \varphi \in C_{\mathrm{b}}(E^n,\mathbb{R}),\, \pi_{G,n} \in \Pi_{G,n} \right\}, \tag{10.1}$$

(recall the definition of $H_\varphi^{(n)}$ from (2.37) and of $\Pi_{G,n}$ from (2.7)) 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 $\mathcal{L}[X^{(\Omega_N)}(0)]$, where $X^{(\Omega_N)}$ is the $C_N^{\underline{c},\underline{\Lambda}}$-process on Ω_N satisfying our assumptions for a given parameter $\theta \in \mathcal{P}(E)$ (see below Proposition 1.4), and all admissible $n, \varphi, \pi_{G,n}$, we have

$$\mathbb{E}\left[H_\varphi^{(n)}(X^{(\Omega_N)}(t), \pi_{G,n}) \right] \underset{t\to\infty}{\longrightarrow} F((\varphi, n, \pi_{G,n}), \theta), \tag{10.2}$$

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 θ.

(2) Depending on whether $\bar{H}_N < \infty$ or $\bar{H}_N = \infty$, with $\bar{H}_N$ as in Section 2.4.2, the quantity in the right-hand side of (10.2) corresponds to the form of the limit claimed in (1.66–1.67).

Item (2) follows from Theorem 1.8 once we have proved the convergence result in (10.2), since (1.65) implies that the marginal law of the limiting state is δ_θ, and we will see in (10.5) below that recurrence of the migration mechanism a (recall (1.26)) implies that

$$\mathbb{E}^{(\Omega_N)}_{\nu_\theta,\underline{c},\underline{\Lambda}}\left[\langle\varphi,\bigotimes_{i=1}^n x_{\eta_i}\rangle\right] = \langle f^n(u),\theta\rangle, \quad \text{for } \varphi(u_1,\cdots,u_n) = \prod_{i=1}^n f(u_i), \tag{10.3}$$

which in turn implies

$$\nu^{(\Omega_N)}_{\theta,\underline{c},\underline{\Lambda}} = \int_K (\delta_u)^{\otimes\Omega_N}\theta(\mathrm{d}u). \tag{10.4}$$

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 $\mathfrak{C}^{(\Omega_N)}_t$ as in (2.32) 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 $\{1,\dots,n\}$ 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 Dawson et al. (1995), Lemma 3.2, that the positions of the partition elements are given, asymptotically, by $k = 1,\dots,n$ random walks, all starting at the origin. Using that the initial state is ergodic, we can then calculate, for $\varphi(u_1,\cdots,u_n) = \Pi^n_{k=1} f(u_k)$,

$$\lim_{t\to\infty} E\left[H^{(n)}_\varphi\left(X^{(\Omega_N)}(0),\mathfrak{C}^{(\Omega_N)}_t\right)\right] = \sum_{k=1}^n \langle f,\theta\rangle^k q_k^{(\pi_{G,n})}, \tag{10.5}$$

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 a we know that

$$\lim_{m\to\infty} q_k^{(\pi^{(m)}_{G,n})} = 0, \quad \forall\, k = 1,\dots,n-1 \text{ and } \lim_{m\to\infty} q_n^{(\Pi^{(m)}_{G,n})} = 1. \tag{10.6}$$

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 θ (see Dawson et al. (1995), p. 2310, for details). □

11. Scaling of the volatility in the clustering regime

In Section 11.1, we prove Theorems 1.10 and 1.11, in Section 11.3 we prove Theorem 1.12.

11.1. *Comparison with the hierarchical Fleming-Viot process.*

Proof of Theorem 1.11: (a) Rewrite the recursion relation in (1.45) as

$$d_0 = 0, \qquad \frac{1}{d_{k+1}} = \frac{1}{c_k} + \frac{1}{\mu_k + d_k}, \quad k \in \mathbb{N}_0. \tag{11.1}$$

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 (1.71) 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

$$d_k \geq d_k^*, \quad k \in \mathbb{N}, \tag{11.2}$$

with d_k^* given by (1.72).

(c) Inserting the definition $m_k = (\mu_k + d_k)/c_k$ into (11.1), we get the recursion relation

$$c_0 m_0 = \mu_0, \qquad c_{k+1} m_{k+1} = \mu_{k+1} + \frac{c_k m_k}{1 + m_k}, \quad k \in \mathbb{N}_0. \tag{11.3}$$

Iterating (11.3), we get

$$c_k m_k = \sum_{l=0}^{k} \frac{\mu_l}{\prod_{j=l}^{k} (1 + m_j)}. \tag{11.4}$$

Ignoring the terms in the denominator, we get

$$m_k \leq \frac{1}{c_k} \sum_{l=0}^{k} \mu_l, \tag{11.5}$$

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

$$m_k \geq \frac{1}{C} \frac{1}{c_k} \sum_{l=0}^{k} \mu_l, \tag{11.6}$$

which after summation over $k \in \mathbb{N}_0$ proves the claim.

(d) We know from (1.72) 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

$$\begin{aligned} \frac{1}{d_k} &\geq \sum_{l=0}^{k-1} \frac{1}{c_l \prod_{j=l+1}^{k-1} (1 + \frac{\mu_j}{d_j})} \geq \sum_{l=0}^{k-1} \frac{1}{c_l \prod_{j=l+1}^{k-1} (1 + \frac{\mu_j}{d_j^*})} \\ &\geq \sum_{l=0}^{k-1} \frac{1}{c_l \prod_{j=l+1}^{\infty} (1 + \frac{\mu_j}{\mu_0}[1 + \mu_0\sigma_j])}. \end{aligned} \tag{11.7}$$

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

$$\frac{\sigma_{k+1} - \sigma_k}{\sigma_1} \geq \frac{\sigma_{k+1} - \sigma_k}{\sigma_k} \geq \int_{\sigma_k}^{\sigma_{k+1}} \frac{\mathrm{d}x}{x}, \quad k \in \mathbb{N}, \tag{11.8}$$

we have

$$\lim_{k\to\infty} \sigma_k = \infty \quad\Longleftrightarrow\quad \sum_{k\in\mathbb{N}} \frac{1}{c_k\sigma_k} = \infty. \tag{11.9}$$

□

Proof of Theorem 1.10. Combining Lemma 2.13 with Theorem 1.11(c), we get the claim. □

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 (1998); 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

$$x_k^+ = \tfrac{1}{2}\mu_k[-1+\sqrt{1+4c_k/\mu_k}] \in (0,\infty), \quad x_k^- = \tfrac{1}{2}\mu_k[-1-\sqrt{1+4c_k/\mu_k}] \in (-\infty,0), \tag{11.13}$$

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. 11.5 for a picture of f_k.

In what follows, we will use the following two theorems of Kooman (1998). We state the version of these theorems for $\mathbb{R}$, although they apply for $\mathbb{C}$ as well.

Theorem 11.1. [Kooman (1998), 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 (1998), 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.,

$$\begin{gathered}\sum_{k\in\mathbb{N}_0} |x_{k+1}^+ - x_k^+| < \infty, \quad \sum_{k\in\mathbb{N}_0} |x_{k+1}^- - x_k^-| < \infty, \\ x^+ = \lim_{k\to\infty} x_k^+ \in \mathbb{R}^*, \quad x^- = \lim_{k\to\infty} x_k^- \in \mathbb{R}^*, \quad x^+ \neq x^-.\end{gathered} \tag{11.14}$$

If

$$\prod_{k\in\mathbb{N}_0} |f_k'(x_k^+)| = 0, \tag{11.15}$$

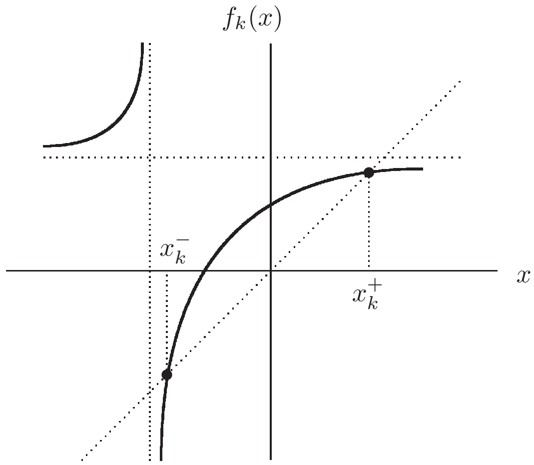

FIGURE 11.5. The Möbius-transformation $x \mapsto f_k(x)$.

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. 11.5, 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 $x_0 = 0$. We therefore also do not need the bounded variation condition in the second part of the first line of (11.14).

11.3. *Scaling of the volatility for polynomial coefficients. Proof of Theorem 1.12.* Theorem 1.12 shows *four regimes*. Our key assumptions are (1.78–1.81). 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 1.12(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 1.12(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 et al. (1987, Section 1.8), a strictly positive sequence (a_k) is said to be *smoothly varying* with index $\rho \in \mathbb{R}$ if

$$\lim_{k\to\infty} k^n a_k^{[n]}/a_k = \rho(\rho-1)\times\cdots\times(\rho-n+1), \qquad n \in \mathbb{N}, \tag{11.19}$$

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 $n \le \rho$.

This observation will be useful in combination with (I–III).

According to Bingham et al. (1987, 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 ρ such that $a_k' \le a_k \le a_k''$ 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 1.11(a), it therefore suffices to prove Theorem 1.12 under the following assumption, which is stronger than (1.78):

$$\begin{aligned}&(c_k),\ (\mu_k),\ (\mu_k/c_k),\ (k^2\mu_k/c_k) \text{ are smoothly varying}\\ &(\text{with index } a,\ b,\ a-b, \text{ respectively, } 2+a-b).\end{aligned} \tag{11.20}$$

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

$$y_0 = 0, \qquad y_{k+1} = g_k(y_k), \quad k \in \mathbb{N}_0, \tag{11.21}$$

where

$$g_k(y) = \frac{A_k y + B_k}{C_k y + D_k}, \qquad y \in \mathbb{R}^*, \tag{11.22}$$

with coefficients

$$A_k = \frac{c_k^2}{c_{k+1}}, \quad B_k = \frac{c_k \mu_k}{c_{k+1}}, \quad C_k = c_k, \quad D_k = c_k + \mu_k. \tag{11.23}$$

By (1.78), we have $c_k/c_{k+1} \sim 1$, and hence $A_k \sim C_k \sim c_k$, $B_k \sim K c_k$, $D_k \sim (K+1)c_k$. Therefore, (11.22) yields

$$\lim_{k\to\infty} g_k(y) = g(y) = \frac{y+K}{y+(K+1)}, \qquad y \in \mathbb{R}^*. \tag{11.24}$$

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

$$\lim_{k\to\infty} y_k = M^+. \tag{11.25}$$

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

$$\lim_{k\to\infty} g_k(y) = g(y) = 1, \qquad y \in \mathbb{R}^*. \tag{11.26}$$

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

$$y_k^\pm = \frac{1}{a_k} h^\pm(b_k/a_k^2) \ \text{ with } \ h^\pm(x) = \frac{1}{2x}\big(1 \mp \sqrt{1+4x}\,\big), \ a_k = \frac{A_k - D_k}{B_k}, \ b_k = \frac{C_k}{B_k} \tag{11.27}$$

(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),

$$a_k = \frac{c_k}{\mu_k}\left(1 - \frac{c_{k+1}}{c_k}\right) - \frac{c_{k+1}}{c_k}, \quad b_k = \frac{c_k}{\mu_k}\,\frac{c_{k+1}}{c_k}, \tag{11.28}$$

it follows from (1.79), (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

$$g_k'(y_k^+) = \frac{\Delta_k}{(C_k y_k^+ + D_k)^2} \ \text{ with } \ \Delta_k = A_k D_k - B_k C_k. \tag{11.29}$$

Since $y_k^+ > 0$ and $D_k > \mu_k$, we have

$$\prod_{k\in\mathbb{N}_0} g_k'(y_k^+) \le \prod_{k\in\mathbb{N}_0} \frac{\Delta_k}{\mu_k^2}. \tag{11.30}$$

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

$$A_k = c_k\sqrt{\frac{c_k\mu_k}{c_{k+1}\mu_{k+1}}}, \quad B_k = c_k\mu_k\sqrt{\frac{1}{c_{k+1}\mu_{k+1}}}, \quad C_k = \sqrt{c_k\mu_k}, \quad D_k = c_k + \mu_k. \tag{11.31}$$

By (1.78), $c_{k+1}/c_k \sim 1$ 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

$$\lim_{k\to\infty} g_k(y) = g(y) = y, \qquad y \in \mathbb{R}^*. \tag{11.32}$$

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

$$\alpha_k = \frac{c_{k+1}}{c_k} - 1, \quad \beta_k = \frac{\mu_{k+1}}{\mu_k} - 1, \quad \gamma_k = \frac{\mu_k}{c_k}, \tag{11.33}$$

and write

$$a_k = \frac{1}{\sqrt{\gamma_k}}\left[1 - (1+\gamma_k)\sqrt{(1+\alpha_k)(1+\beta_k)}\right], \quad b_k = \sqrt{(1+\alpha_k)(1+\beta_k)}. \tag{11.34}$$

We have $\alpha_k \to 0$, $\beta_k \to 0$, $\gamma_k \to 0$. Moreover, (1.79–1.81), (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 (1.79–1.81), (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

$$a_k = \frac{1}{\sqrt{\gamma_k}}\,\frac{1-(1+\gamma_k)^2(1+\alpha_k)(1+\beta_k)}{1+(1+\gamma_k)\sqrt{(1+\alpha_k)(1+\beta_k)}}. \tag{11.35}$$

The denominator tends to 2, is Lipschitz near 2, and has bounded variation because (α_k), (β_k), (γ_k) have bounded variation. The numerator equals $-\alpha_k - \beta_k - 2\gamma_k$ plus terms that are products of α_k, β_k and γ_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

$$\Delta_k = c_k^2\sqrt{\frac{c_k\mu_k}{c_{k+1}\mu_{k+1}}} = c_k^2/\sqrt{(1+\alpha_k)(1+\beta_k)}, \quad C_k y_k^+ + D_k = c_k(1 + y_k^+\sqrt{\gamma_k} + \gamma_k), \tag{11.36}$$

and hence

$$\prod_{k\in\mathbb{N}_0} g_k'(y_k^+) \le \prod_{k\in\mathbb{N}_0} \frac{1}{\sqrt{(1+\alpha_k)(1+\beta_k)}(1+y_k^+\sqrt{\gamma_k})^2}. \tag{11.37}$$

The term under the product equals

$$1 - 2y^+\sqrt{\gamma_k}\,[1+o(1)], \tag{11.38}$$

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 (1.78), $c_{k+1}/c_k \sim 1$, $\sigma_{k+1}/\sigma_k \sim 1$ and $k\delta_k \to 1 - a$ with $a \in (-\infty, 1)$ the exponent in (1.78). 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

$$y_k^\pm = \tfrac{1}{2}\left(\bar{a}_k \pm \sqrt{\bar{a}_k^2 + 4\bar{b}_k}\right) \quad \text{ with } \quad \bar{a}_k = \frac{A_k - D_k}{C_k}, \quad \bar{b}_k = \frac{B_k}{C_k}, \tag{11.43}$$

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 (1.78), 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 (1.79–1.81), (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,

$$\prod_{k\in\mathbb{N}_0} g_k'(y_k^+) \le \prod_{k\in\mathbb{N}_0} \frac{1 + \delta_k}{(1 + \delta_k y_k^+)^2}. \tag{11.46}$$

The term under the product equals

$$1 - (2y^+ - 1)\delta_k\,[1 + o(1)], \tag{11.47}$$

Since $y^+ \ge 1$, 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. Proof of Theorem 1.14.* In this section, we briefly comment on how to extend the proof of Theorem 1.12 to cover the case of Theorem 1.14.

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 1.11(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 δ_a (cf. (1.13)).
- $\frac{\partial^2 F(x)}{\partial x^2}[\delta_u,\delta_v] \rightsquigarrow$ second Gâteaux-derivative of F with respect to x in the directions δ_u, δ_v (cf. (1.16)).
- $D(T,\mathcal{E}) \rightsquigarrow$ set of cádlág paths in $\mathcal{E}$ indexed by the elements of $T \subset \mathbb{R}$ and equipped with the Skorokhod J_1-topology.
- $C_{\mathrm{b}}\big(\mathcal{E},\mathcal{E}'\big) \rightsquigarrow$ 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. (1.21)).
- $\underline{c} = (c_k)_{k\in\mathbb{N}_0} \in (0,\infty)^{\mathbb{N}_0} \rightsquigarrow$ migration coefficients (cf. (1.25)).
- $\underline{\Lambda} = \big(\Lambda_k\big)_{k\in\mathbb{N}_0} \in \mathcal{M}_f([0,1])^{\mathbb{N}_0} \rightsquigarrow$ offspring measures (cf. (1.28)).
- $\lambda_k = \Lambda_k([0,1]) \rightsquigarrow$ resampling rates (cf. (1.30)).
- $\underline{d} = (d_k)_{k\in\mathbb{N}_0} \rightsquigarrow$ volatility constants (cf. (1.45)).
- $\underline{m} = (m_k)_{k\in\mathbb{N}_0} \rightsquigarrow$ (cf. (1.57)).
- $\mu_k = \frac{1}{2}\lambda_k \rightsquigarrow$ (cf. (1.57)).
- $\sigma_k \rightsquigarrow$ (cf. (1.72)).
- $B_k(\eta) \rightsquigarrow$ k-macro-colony around η (cf. (1.23)).
- $y_{\eta,k} \rightsquigarrow$ type distribution in $B_k(\eta)$ (cf. (1.31)).
- C^{Λ}-process $\rightsquigarrow$ non-spatial continuum-mass Λ-Cannings process (cf. Section 1.3.1).
- $a^{(N)}(\cdot,\cdot) \rightsquigarrow$ hierarchical random walk kernel on Ω_N (cf. (1.26)).
- $C_N^{\underline{c},\underline{\Lambda}}$-process $\rightsquigarrow$ hierarchically interacting Cannings process on Ω_N (cf. Section 1.4.4).
- $\mathcal{F} \rightsquigarrow$ algebra of test functions on $\mathcal{P}(E)^{\Omega_N}$ (cf., (1.34)).
- $L^{(N)}$, $L^{(N)}_{\mathrm{mig}}$, $L^{(N)}_{\mathrm{res}} \rightsquigarrow$ generators of the mean-field Cannings process (cf. (1.11)).

- $L^{(\Omega_N)}$, $L^{(\Omega_N)}_{\text{mig}}$, $L^{(\Omega_N)}_{\text{res}}$ $\rightsquigarrow$ generators of the hierarchical Cannings process (cf. (1.36)).
- $\Phi_{r,a,B_k(\eta)}$ $\rightsquigarrow$ reshuffling-resampling map (cf. (1.39)).
- $X^{(\Omega_N)}$ $\rightsquigarrow$ $C^{\underline{c},\underline{\Lambda}}_N$-process (cf. Section 1.4.4).
- $Y^{(\Omega_N)}_{\eta,k}(\cdot)$ $\rightsquigarrow$ macroscopic observables (= block averages) of $X^{(\Omega_N)}$ (cf. (1.43)).
- $y^{[1]}_\eta$ $\rightsquigarrow$ 1-block averages indexed block-wise (cf. (7.18)).
- $L^{(N)[k]}_{\text{res}}$, $L^{(N)[k]}_{\text{mig}}$ $\rightsquigarrow$ generators of the k-block averaged hierarchically interacting Cannings process at the time scale $t^k N$ (cf. 6.2.2).
- $G_{N,K}$ $\rightsquigarrow$ K-level truncation of Ω_N (cf. (1.42)).
- $X^{(N)}$ $\rightsquigarrow$ mean-field interacting Cannings process (cf. Section 1.3.2).
- $Q_x(\mathrm{d}u, \mathrm{d}v)$ $\rightsquigarrow$ Fleming-Viot diffusion function (cf. (1.19)).
- $L^{c,d,\Lambda}_\theta$, L^c_θ, L^d, L^Λ $\rightsquigarrow$ generators of the McKean-Vlasov process (cf. (1.17)).
- $Z^{c,d,\Lambda}_\theta$ $\rightsquigarrow$ McKean-Vlasov process with immigration-emigration (cf. Section 1.3.3).
- $\nu^{c,d,\Lambda}_\theta$ $\rightsquigarrow$ unique equilibrium of Z (cf. (4.1)).
- $(M^{(j)}_k)_{k=-(j+1),\dots,0}$ $\rightsquigarrow$ interaction chain (cf. Section 1.5.2).

12.3. *Spatial Λ-coalescents.*

- $[n] = \{1, \dots, n\}$.
- Π_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. (2.7)).
- $S_{G,n} \in \Pi_{G,n}$ $\rightsquigarrow$ G-labelled partition into singletons (cf. (2.8)).
- Π, Π_G $\rightsquigarrow$ partitions of $\mathbb{N}$, G-labelled partitions of $\mathbb{N}$ (cf. (2.11)).
- $L(\pi_G)$ $\rightsquigarrow$ set of labels of partition π_G (cf. (2.10)).
- $\lambda^{(\Lambda)}_{b,i}$ $\rightsquigarrow$ coalescence-rates (cf. (2.14)).
- $\cdot|_n$ $\rightsquigarrow$ operation of projection from $[m]$ (respectively, $\mathbb{N}$) onto $[n]$.
- $L^{(G)*}$, $L^{(G)*}_{\text{mig}}$, $L^{(G)*}_{\text{coal}}$ generators of the spatial coalescent on G (cf., (2.23)).
- $L^{(\Omega_N)*}$, $L^{(\Omega_N)*}_{\text{mig}}$, $L^{(\Omega_N)*}_{\text{coal}}$ $\rightsquigarrow$ generators of the spatial Λ-coalescent with non-local coalescence (cf. (2.34)).
- $\mathfrak{P}$ $\rightsquigarrow$ field of Poisson point processes driving the spatial Λ-coalescent (cf. (2.15)).
- $\mathfrak{P}^{(\Omega_N)}$ $\rightsquigarrow$ driving Poisson point process for the spatial n-Λ-coalescent with non-local coalescence (cf. (2.28)).
- $\mathfrak{C}^{(G)}_n$ $\rightsquigarrow$ spatial finite n-Λ-coalescent on G (cf. (2.18)).
- $\mathfrak{C}^{(G)}$ $\rightsquigarrow$ spatial Λ-coalescent on G (cf. (2.20)).
- $\mathfrak{C}^{(\Omega_N)}$ $\rightsquigarrow$ spatial $\underline{\Lambda}$-coalescent with non-local coalescence (cf. (2.32)).

Acknowledgments

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) and by NWO (grant 613.000.913). SK held a postdoctoral position at EURANDOM from the Summer of 2009 until the Summer of 2011. AK was Postdoctoral Marie Curie Fellow at EURANDOM from December 2009 until July 2012, and held a postdoctoral position in Leiden from August 2012 until May 2013. 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 were guests of the Institute for Mathematical Sciences, National University of Singapore, during its 2011 programme on Probability and Discrete Mathematics in Mathematical Biology. The authors thank the anonymous referees for a careful reading of the manuscript and for helpful suggestions on the presentation.

References

R.A. Arratia. *Coalescing Brownian motions on the line*. Ph.D. thesis, University of Wisconsin-Madison (1979). MR2630231.

J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander. On the attracting orbit of a non-linear transformation arising from renormalization of hierarchically interacting diffusions. I. The compact case. *Canad. J. Math.* **47** (1), 3–27 (1995). MR1319687.

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. II. The non-compact case. *J. Funct. Anal.* **146** (1), 236–298 (1997). MR1446381.

N.H. Barton, A.M. Etheridge and A. Véber. A new model for evolution in a spatial continuum. *Electron. J. Probab.* **15**, no. 7, 162–216 (2010). MR2594876.

N. Berestycki. *Recent progress in coalescent theory*, volume 16 of *Ensaios Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, Rio de Janeiro (2009). ISBN 978-85-85818-40-1. MR2574323.

N. Berestycki, A.M. Etheridge and A. Véber. Large scale behaviour of the spatial Λ-Fleming–Viot process. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** (2), 374–401 (2013). MR3088374.

N.H. Bingham, C.M. Goldie and J.L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge (1987). ISBN 0-521-30787-2. MR898871.

M. Birkner and J. Blath. Computing likelihoods for coalescents with multiple collisions in the infinitely many sites model. *J. Math. Biol.* **57** (3), 435–465 (2008). MR2411228.

M. Birkner and J. Blath. Measure-valued diffusions, general coalescents and population genetic inference. In *Trends in stochastic analysis*, volume 353 of *London Math. Soc. Lecture Note Ser.*, pages 329–363. Cambridge Univ. Press, Cambridge (2009). MR2562160.

R.M. Blumenthal and R.K. Getoor. *Markov processes and potential theory*. Pure and Applied Mathematics, Vol. 29. Academic Press, New York (1968). MR0264757.

C. Cannings. The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. *Advances in Appl. Probability* **6**, 260–290 (1974). MR0343949.

C. Cannings. The latent roots of certain Markov chains arising in genetics: a new approach. II. Further haploid models. *Advances in Appl. Probability* **7**, 264–282 (1975). MR0371430.

J.T. Cox, D.A. Dawson and A. Greven. Mutually catalytic super branching random walks: large finite systems and renormalization analysis. *Mem. Amer. Math. Soc.* **171** (809), viii+97 (2004). MR2074427.

J.T. Cox and D. Griffeath. Diffusive clustering in the two-dimensional voter model. *Ann. Probab.* **14** (2), 347–370 (1986). MR832014.

D. Dawson and A. Greven. Multiple time scale analysis of interacting diffusions. *Probab. Theory Related Fields* **95** (4), 467–508 (1993a). MR1217447.

D.A. Dawson. Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI—1991*, volume 1541 of *Lecture Notes in Math.*, pages 1–260. Springer, Berlin (1993). MR1242575.

D.A. Dawson, L.G. Gorostiza and A. Wakolbinger. Hierarchical random walks. In *Asymptotic methods in stochastics*, volume 44 of *Fields Inst. Commun.*, pages 173–193. Amer. Math. Soc., Providence, RI (2004). MR2106854.

D.A. Dawson, L.G. Gorostiza and A. Wakolbinger. Degrees of transience and recurrence and hierarchical random walks. *Potential Anal.* **22** (4), 305–350 (2005). MR2135263.

D.A. Dawson and A. Greven. Hierarchical models of interacting diffusions: multiple time scale phenomena, phase transition and pattern of cluster-formation. *Probab. Theory Related Fields* **96** (4), 435–473 (1993b). MR1234619.

D.A. Dawson and A. Greven. Multiple time scale analysis of hierarchically interacting systems. In *Stochastic processes*, pages 41–50. Springer, New York (1993c). MR1427299.

D.A. Dawson and A. Greven. Multiple space-time scale analysis for interacting branching models. *Electron. J. Probab.* **1**, no. 14, 84 pp. (electronic) (1996). MR1423467.

D.A. Dawson and A. Greven. Hierarchically interacting Fleming-Viot processes with selection and mutation: multiple space time scale analysis and quasi-equilibria. *Electron. J. Probab.* **4**, no. 4, 81 pp. (electronic) (1999). MR1670873.

D.A. Dawson and A. Greven. State dependent multitype spatial branching processes and their longtime behavior. *Electron. J. Probab.* **8**, no. 4, 93 pp. (electronic) (2003). MR1961166.

D.A. Dawson, A. Greven, F. den Hollander, Rongfeng Sun and J.M. Swart. The renormalization transformation of two-type branching models. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** (6), 1038–1077 (2008). MR2469334.

D.A. Dawson, A. Greven and J. Vaillancourt. Equilibria and quasiequilibria for infinite collections of interacting Fleming-Viot processes. *Trans. Amer. Math. Soc.* **347** (7), 2277–2360 (1995). MR1297523.

D.A. Dawson and P. March. Resolvent estimates for Fleming-Viot operators and uniqueness of solutions to related martingale problems. *J. Funct. Anal.* **132** (2), 417–472 (1995). MR1347357.

R. Der, C.L. Epstein and J.B. Plotkin. Generalized population models and the nature of genetic drift. *Theoretical Population Biology* **80** (2), 80–99 (2011). DOI: 10.1016/j.tpb.2011.06.004.

P. Donnelly and T.G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.* **27** (1), 166–205 (1999). MR1681126.

B. Eldon and J. Wakeley. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **172** (4), 2621–2633 (2006). DOI: 10.1534/genetics.105.052175.

A. Etheridge. *An introduction to superprocesses*, volume 20 of *University Lecture Series*. American Mathematical Society, Providence, RI (2000). ISBN 0-8218-2706-5. MR1779100.

A. Etheridge. *Some mathematical models from population genetics*, volume 2012 of *Lecture Notes in Mathematics*. Springer, Heidelberg (2011). ISBN 978-3-642-16631-0. Lectures from the 39th Probability Summer School held in Saint-Flour, 2009. MR2759587.

S.N. Ethier and T.G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York (1986). ISBN 0-471-08186-8. Characterization and convergence. MR838085.

S.N. Evans. Coalescing Markov labelled partitions and a continuous sites genetics model with infinitely many types. *Ann. Inst. H. Poincaré Probab. Statist.* **33** (3), 339–358 (1997). MR1457055.

K. Fleischmann and A. Greven. Diffusive clustering in an infinite system of hierarchically interacting diffusions. *Probab. Theory Related Fields* **98** (4), 517–566 (1994). MR1271108.

N. Freeman. The segregated Lambda-coalescent. *ArXiv Mathematics e-prints* (2013). arXiv:1109.4363 [math.PR].

A. Greven. Renormalization and universality for multitype population models. In *Interacting stochastic systems*, pages 209–246. Springer, Berlin (2005). MR2118576.

A. Greven and F. den Hollander. Phase transitions for the long-time behavior of interacting diffusions. *Ann. Probab.* **35** (4), 1250–1306 (2007). MR2330971.

A. Greven, A. Klimovsky and A. Winter. Tree-valued Cannings dynamics (2014). In preparation.

A. Greven, P. Pfaffelhuber and A. Winter. Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Related Fields* **145** (1-2), 285–322 (2009). MR2520129.

F. den Hollander and J.M. Swart. Renormalization of hierarchically interacting isotropic diffusions. *J. Statist. Phys.* **93** (1-2), 243–291 (1998). MR1656371.

R.J. Kooman. Asymptotic behaviour of solutions of linear recurrences and sequences of Möbius-transformations. *J. Approx. Theory* **93** (1), 1–58 (1998). MR1612782.

T.G. Kurtz. Averaging for martingale problems and stochastic approximation. In *Applied stochastic analysis (New Brunswick, NJ, 1991)*, volume 177 of *Lecture Notes in Control and Inform. Sci.*, pages 186–209. Springer, Berlin (1992). MR1169928.

T.M. Liggett. *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York (1985). ISBN 0-387-96069-4. MR776231.

T.M. Liggett and F. Spitzer. Ergodic theorems for coupled random walks and other systems with locally interacting components. *Z. Wahrsch. Verw. Gebiete* **56** (4), 443–468 (1981). MR621659.

V. Limic and A. Sturm. The spatial Λ-coalescent. *Electron. J. Probab.* **11**, no. 15, 363–393 (electronic) (2006). MR2223040.

J. Pitman. Coalescents with multiple collisions. *Ann. Probab.* **27** (4), 1870–1902 (1999). MR1742892.

J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin (2006). ISBN 978-3-540-30990-1; 3-540-30990-X. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard. MR2245368.