ALEA, Lat. Am. J. Probab. Math. Stat. **15**, 295–351 (2018)
DOI: 10.30757/ALEA.v15-14

The hierarchical Cannings process in random environment

Andreas Greven, Frank den Hollander and Anton Klimovsky

Department Mathematik, Universität Erlangen-Nürnberg
Cauerstrasse 11
D-91058 Erlangen
Germany
E-mail address: greven@mi.uni-erlangen.de

Mathematisch Instituut, Universiteit Leiden
P.O. Box 9512
NL-2300RA Leiden
The Netherlands
E-mail address: denholla@math.leidenuniv.nl

Fakultät für Mathematik, Universität Duisburg-Essen
D-45117 Essen
Germany
E-mail address: anton.klymovskiy@uni-due.de

Received by the editors March 7th, 2017; accepted February 16th, 2018.

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

Key words and phrases. Hierarchical Cannings process, random environment, migration, block reshuffling, block resampling, block coalescence, hierarchical mean field limit, random Möbius transformations.

AG was supported by the Deutsche Forschungsgemeinschaft (grant DFG-GR 876/15-2), FdH was supported by the European Research Council (Advanced Grant VARIS-267356) and by the Netherlands Organization for Scientific Research (Gravitation Grant NETWORKS-024.002.003), AK was supported by the Netherlands Organization for Scientific Research (grant 613.000.913).

Abstract. In an earlier paper, we introduced and studied a system of hierarchically interacting measure-valued random processes that arises as the continuum limit of a large population of individuals carrying different types. Individuals live in colonies labelled by the hierarchical group of order N, and are subject to *migration* and *resampling* on all hierarchical scales simultaneously. The resampling mechanism is such that a random positive fraction of the population in a block of colonies inherits the type of a random single individual in that block, which is why we refer to our system as the hierarchical Cannings process. Before resampling in a block takes place, all individuals in that block are relocated uniformly, which we call *reshuffling*. The evolution of the system seen backwards in time leads to a dual process of coalescing random walks (representing the lineages) in random environment. The space-time scaling behaviour of the dual determines that of the system forward in time.

In the present paper, we study a version of the hierarchical Cannings process in *random environment*, namely, the resampling measures controlling the change of type of individuals in different blocks are chosen randomly with a given mean and are kept fixed in time, i.e., we work in the *quenched* setting. We give a necessary and sufficient condition under which a *multi-type* equilibrium is approached (= coexistence) as opposed to a *mono-type* equilibrium (= clustering). Moreover, in the hierarchical mean-field limit $N \to \infty$, with the help of a *renormalization analysis* we obtain a full picture of the space-time scaling behaviour of block averages on all hierarchical scales simultaneously. We show that the k-block averages are distributed as the superposition of a Fleming-Viot diffusion with a deterministic volatility constant d_k and a Cannings process with a random jump rate, both depending on k. In the random environment d_k turns out to be smaller than in the homogeneous environment of the same mean. We investigate how d_k scales with k. This leads to five *universality classes of cluster formation* in the mono-type regime. We find that if clustering occurs, then the random environment slows down the growth of the clusters, i.e., enhances the diversity of types. In some universality classes, the growth of the clusters depends on the realisation of the random environment, in others it does not.

Contents

1. Introduction 297
1.1. Motivation and goal 297
1.2. Summary of the main results 299
1.3. Outline 301
2. The model 301
2.1. The hierarchical Cannings process 301
2.2. The hierarchical Cannings process in random environment 307
3. Main theorems 309
3.1. Results for fixed N 309
3.2. Results for $N \to \infty$ 311
3.3. Summary of the effects of the random environment 321
4. Existence, uniqueness, duality and equilibrium 321
4.1. The spatial coalescent in random environment 322
4.2. Dualities 324
4.3. Well-posedness of the martingale problems and equilibria 324
4.4. Consequences for the Cannings process 325
5. Dichotomy: coexistence versus clustering 327
5.1. Mean hazard 328
5.2. Zero-one law 332
6. Multi-scale analysis 336
6.1. The mean-field finite-system scheme 336
6.2. The hierarchical mean-field limit 338
6.3. Dichotomy in the hierarchical mean-field limit 339
7. The orbit of the renormalization transformations 340
7.1. Random environment lowers the volatility 340
7.2. Scaling of the volatility: polynomial coefficients 341
7.3. Scaling of the volatility: exponential coefficients 344
8. Identification of the universality classes of cluster formation 344
8.1. Random cluster size 345
8.2. Random cluster order 347
Acknowledgements 350
References 350

1. Introduction

1.1. *Motivation and goal.* Two models play a central role in the world of stochastic multi-type population dynamics:

(1) The Moran model and its limit for large populations, the Fleming-Viot measure-valued *diffusion*.
(2) The Cannings model and its limit for large populations, the Cannings measure-valued *jump process* (also called the generalized Fleming-Viot process).

The Cannings model accounts for situations in which *resampling* is such that a random positive fraction of the population in the next generation inherits the type

of a random single individual in the current generation, even in the infinite population limit (see Cannings, 1974, 1975). In order to describe a setting where this effect has a geographical structure, i.e., where *migration* of individuals is allowed as well, different models have been proposed in Limic and Sturm (2006), Blath et al. (2007), Barton et al. (2010), Berestycki et al. (2013), and Greven et al. (2014). The behaviour of these models has been studied in detail and its dependence on the geographic space is fairly well understood.

The type space is typically chosen to be a compact Polish space E. In Greven et al. (2014), we focused on the case where the geographic space is the hierarchical group Ω_N of order N, since this allowed us to carry out a full *renormalization analysis*. In the *hierarchical mean-field limit* $N \to \infty$, the migration can be chosen in such a way that it approximates migration on the geographic space $\mathbb{Z}^2$, a possibility that was exploited by Sawyer and Felsenstein (1983) (see also Dawson et al., 2004).

We analyze the model introduced in Greven et al. (2014), but add the effect that the Cannings resampling mechanism is controlled by *catastrophic events* on a small time scale, for which it is appropriate to assume that the rate of occurrence has a spatially inhomogeneous structure. This leads us to consider spatial Cannings models with block resampling in *random environment*, i.e., both the form and the overall rate of the block resampling mechanism depend on the geographic location.

Remark 1.1. In a catastrophic event, a part of the population is killed in a large spatial area and is subsequently replenished via a rapid recolonization, resulting in a bottleneck effect consisting of compression and subsequent expansion of the descendants of a single ancestor. The mechanisms behind such events are functions of the background environment, which is inhomogeneous in space but constant in time. It would be interesting to derive our continuum model (defined in Section 2.2) from an individual-based model with *two time scales*: the catastrophic events happen on a fast time scale, while the migration and resampling happen on a slow time scale. Moreover, in our individual-based model we do *reshuffling* before resampling, which must be motivated likewise. Carrying out the details of such a derivation would merit a paper in its own right. □

Remark 1.2. In the present paper, we *only* work with continuum models. However, we motivate these models by viewing them as the large-population limit of individual-based models. □

The *goal* of the present paper is three-fold:

(1) *Construction* of the hierarchical Cannings process in random environment via a well-posed martingale problem and derivation of a *duality relation* with a hierarchical spatial coalescent in random environment.
(2) Analysis of the longtime behaviour, in particular, the *dichotomy* between a multi-type equilibrium and a mono-type equilibrium.
(3) Scaling analysis of a collection of *renormalised processes* obtained by looking at the evolution of blocks averages on successive space-time scales in the hierarchical mean-field limit and the consequences for universality classes of the mono-type cluster formation.

We are particularly interested in *new effects caused by the random environment*. To capture these effects, we need to deal with spatially inhomogeneous processes. Below, we list what is different and what is the same compared to the model in

homogeneous environment treated in Greven et al. (2014) (see Section 3.3 for more details), and what new techniques are employed.

▶ Properties that are *different*:

- The volatility of the block averages is *lower* compared to that in the homogeneous (= average) environment. The random environment causes fluctuations in the resampling, which in turn reduce the clustering.
- The growing of mono-type clusters exhibits both environment-dependent *and* environment-independent space-time scales.
- The equilibrium is environment-dependent.
- Formulas involving averages of random Möbius transformations are obtained for the parameters controlling the renormalized system on successive space-time scales in the hierarchical mean-field limit.

▶ Properties that are the *same*:

- The criterion for the dichotomy of co-existence versus clustering.
- The form of the duality (backwards in time with block-coalescing random walks).
- The space-time scaling of the system in the hierarchical mean-field limit can be computed explicitly.

▶ New techniques are:

- The harmonic analysis that was used to study the dual process is not applicable anymore and has to be replaced by potential-theoretic tools.
- The dual process for the second moments involves two random walks instead of one. The reason is that in random environment the difference of two ancestral paths is no longer equal in distribution to that of one ancestral path running at double the speed.
- To analyze the orbits of the renormalization transformation, which arise from iterations of averages of random Möbius transformations, comparison metholds are needed.

Finally, compared with Greven et al. (2014), we have made further progress in finding the appropriate *conceptual form* for the study of the growth of mono-type clusters, in particular, in classifying the different clustering regimes.

Our proofs focus on the new ingredients compared to Greven et al. (2014). However, in order to be self-contained, we everywhere expose the full line of argument, sometimes referring to Greven et al. (2014) to cut the presentation short. The key mathematical tools that we will exploit are the duality of the hierarchical Cannings process in random environment with a *hierarchical spatial coalescent in random environment*, and the scaling of the block averages towards a mean-field process in random environment called the *McKean-Vlasov process*. This in turn will lead us to study two independent hierarchical random walks in the same random environment, and to analyze the orbit of iterations of non-linear transformations arising from random Möbius transformations that link the behaviour on successive hierarchical scales.

1.2. *Summary of the main results.* In an earlier paper, we introduced and studied a system of hierarchically interacting measure-valued random processes that arises as the continuum limit of a large population of individuals subject to migration,

reshuffling and resampling Greven et al. (2014). More precisely, individuals live in colonies labelled by Ω_N, the hierarchical group 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 *resampling* based on a sequence of resampling measures $\underline{\Lambda} = (\Lambda_k)_{k\in\mathbb{N}_0}$, both acting on blocks of colonies (= macro-colonies) on all hierarchical scales $k \in \mathbb{N}_0$ simultaneously. The resampling mechanism is such that a random positive fraction of the population in a block of colonies inherits the type of a random single individual in that block, even in the infinite population limit, which is why we refer to our system as the hierarchical Cannings process. Before resampling in a block takes place, all individuals in that block are relocated uniformly. This relocation is called *reshuffling* and means that resampling is done in a locally "panmictic" manner. The *justification* of this mechanism via a model of catastrophic events on a slow time scale would require a detailed further study and therefore is not carried out here (recall Remark 1.1).

In the present paper, we study a version of the hierarchical Cannings process in *random environment*, namely, the resampling measures in different blocks are chosen randomly with mean $\underline{\Lambda}$ and are kept fixed in time, i.e., we consider the *quenched* version of the system. We construct the hierarchical Cannings process in random environment via a well-posed martingale problem, and establish duality with a system of coalescing hierarchical random walks with block coalescence in random environment. We study the long-time behaviour of the process, in particular, we give a necessary and sufficient condition on $\underline{c}$ and $\underline{\Lambda}$ under which almost sure convergence to a *multi-type equilibrium* occurs (= coexistence), as opposed to a *mono-type equilibrium* (= clustering). The equilibrium depends on the environment, but the condition on $\underline{c}$ and $\underline{\Lambda}$ for its occurrence does not.

To obtain more detailed information on the evolution of the system, we consider the *hierarchical mean-field limit* $N \to \infty$. In this limit, with the help of a *renormalization analysis*, we obtain a full picture of the space-time scaling behaviour on all hierarchical scales simultaneously. Our main result is that, on each hierarchical scale $k \in \mathbb{N}_0$, the k-block averages on time scale N^k converge to a random process that is a superposition of a Cannings process with a resampling measure equal to the associated k-block resampling measure (which depends on the environment) and an additional Fleming-Viot process with volatility d_k, reflecting the macroscopic impact of the lower-order resampling and of the drift of strength c_k towards the limiting $(k+1)$-block average (which is constant on the limiting time scale). It turns out that d_k is a function of c_l and Λ_l for all $0 \le l < k$, and of the law of the random environment. Thus, *it is through the volatility that the renormalization manifests itself.*

We show that *the random environment makes the system less volatile*, i.e., d_k is strictly smaller than its corresponding value for the homogenous system where the resampling measures are replaced by their mean. We investigate how d_k scales as $k \to \infty$, which leads to *various different cases* depending on the choice of $\underline{c}$ and $\underline{\Lambda}$. We find that if migration and resampling occur with comparable rates on all hierarchical scales, then the lower volatility persists in the limit as $k \to \infty$. The renormalization transformation connecting d_{k+1} to d_k turns out to be a nonlinear transformation arising from a *random Möbius transformation*. The scaling behaviour of the iterates of these transformations is studied in detail. We find that if clustering occurs, then the random environment slows down the growth of

the clusters, i.e., enhances the diversity of types. We find five *universality classes of cluster formation* in the regime of clustering. These are linked to the different cases of scaling behaviour of d_k. We find that if the growth of the clusters is *rapid*, then the rate of growth depends on the realization of the environment, while if the growth is *slow*, then the effect of the environment averages out. The latter happens e.g. in the *critical regime* where the system is barely clustering.

The key to an understanding of the hierarchical Cannings process is its duality to finite systems of coalescing random walks in random environment on the hierarchical group. Here, the coalescence is a Λ_0-coalescent for single components and, for each $k \in \mathbb{N}$, a Λ_k-coalescent for k-blocks in Ω_N followed by a relocation uniformly in the block. It is clear that the behaviour of such systems depends on the behaviour of random walks in random environment viewed on successively larger space-time scales. We can no longer employ harmonic analysis as for homogenous random walks. Instead, we need potential theory to control the effect of the random environment.

1.3. *Outline.* Sections 2–5 deal with the model for finite N, while Sections 6–8 deal with the hierarchical mean-field limit $N \to \infty$. In Section 2, we define the hierarchical Cannings process in random environment (the presentation necessarily has considerable overlap with Greven et al., 2014). In Section 3, we state our main theorems (organized in a manner similar to Greven et al., 2014) and summarize the effects of the random environment. Section 4 provides the proof of existence and uniqueness of the hierarchical Cannings process and its dual, and establishes convergence to an equilibrium. Section 5 proves the dichotomy between coexistence (multi-type equilibrium) versus clustering (mono-type equilibrium), and provides the parameter range for both. Section 6 describes the multi-scale analysis for the evolution of block averages on successive space-time scales in the hierarchical mean-field limit, proves the dichotomy in that limit, and identifies the renormalization transformations connecting the successive scales. Section 7 analyzes the orbit of the iterations of these transformations and identifies various different cases for the scaling of the volatility of the block averages. Section 8 links these cases to the universality classes of cluster formation.

Sections 2–3 make up one third of the paper and contain all the main theorems and their background. Sections 4–8 provide their proofs.

2. The model

In this section, we define the hierarchical Cannings process in random environment and construct its dual: a spatial coalescent in random environment. We begin in Section 2.1 by recalling the process without random environment introduced in Greven et al. (2014). In Section 2.2, we explain how the random environment is added.

2.1. *The hierarchical Cannings process.* In Sections 2.1.1–2.1.3, we recall the definition of the hierarchical Canning process given in Greven et al. (2014). In Section 2.1.4, we add the random environment and indicate how the definition needs to be modified.

2.1.1. *The hierarchical group of order N.* The *hierarchical group* Ω_N of order N is the set

$$\Omega_N = \Big\{\eta = (\eta^l)_{l\in\mathbb{N}_0} \in \{0,1,\dots,N-1\}^{\mathbb{N}_0} \colon \sum_{l\in\mathbb{N}_0} \eta^l < \infty\Big\}, \qquad N \in \mathbb{N}\backslash\{1\}, \quad (2.1)$$

endowed with the addition operation $+$ defined by $(\eta+\zeta)^l = \eta^l + \zeta^l \pmod N$, $l \in \mathbb{N}_0$. 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 $d_{\Omega_N}(\cdot,\cdot)$ defined by

$$d_{\Omega_N}(\eta,\zeta) = d_{\Omega_N}(0,\eta-\zeta) = \min\{k \in \mathbb{N}_0 \colon \eta^l = \zeta^l \ \forall\, l \geq k\}, \qquad \eta,\zeta \in \Omega_N. \quad (2.2)$$

Let

$$B_k(\eta) = \{\zeta \in \Omega_N \colon d_{\Omega_N}(\eta,\zeta) \leq k\}, \qquad \eta \in \Omega_N,\ k \in \mathbb{N}_0 \quad (2.3)$$

denote the k-block around η (i.e., the ball of hierarchical radius k around η), which we think of as a *macro-colony*. The geometry of Ω_N is explained in Fig. 2.1.

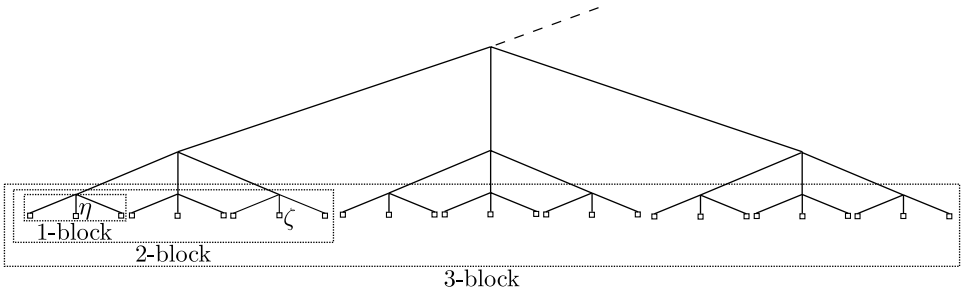

FIGURE 2.1. Close-ups of a 1-block, a 2-block and a 3-block in the hierarchical group of order $N = 3$. The elements of the group are the leaves of the tree (indicated by □'s). The hierarchical distance between two elements is the graph distance to the most recent common ancestor: $d_{\Omega_3}(\eta,\zeta) = 2$ for η and ζ in the picture.

In what follows, we consider a system of individuals organized in colonies labelled by Ω_N. Initially, each colony has M individuals, each carrying a type drawn from

$$E = \text{a Polish type space that is compact.} \quad (2.4)$$

Subsequently, individuals are subject to block migration (Section 2.1.2) and block reshuffling-resampling (Section 2.1.3). In the continuum-mass limit $M \to \infty$, the evolution converges to the hierarchical Cannings process (Section 2.1.4).

2.1.2. *Block migration.* We introduce migration on Ω_N through a random walk kernel. For that purpose, we introduce a sequence of migration rates

$$\underline{c} = (c_k)_{k\in\mathbb{N}_0} \in (0,\infty)^{\mathbb{N}_0}, \quad (2.5)$$

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 chosen uniformly at random in that block.

The transition kernel of the random walk thus performed by the individuals is

$$a^{(N)}(\eta,\zeta) = \sum_{k \geq d_{\Omega_N}(\eta,\zeta)} \frac{c_{k-1}}{N^{2k-1}}, \qquad \eta,\zeta \in \Omega_N,\ \eta \neq \zeta, \qquad a^{(N)}(\eta,\eta) = 0. \tag{2.6}$$

Remark 2.1. The behaviour of the random walk in (2.6) is known in great detail. Dawson, Gorostiza and Wakolbinger Dawson et al. (2005) showed that it is recurrent if and only if $\sum_{k\in\mathbb{N}_0}(1/c_k) = \infty$. They introduced the concept of *degree of recurrence/transience* γ_N (Dawson et al., 2005, Definition 2.1.1), which for the special case where $c_k = c^k$, $c \in (0,\infty)$, equals $\gamma(N) = \log c/\log(N/c)$. Note that

$$\gamma(N) \begin{cases} < 0, & c < 1 \text{ (strongly recurrent)}, \\ = 0, & c = 1 \text{ (critically recurrent)}, \\ > 0, & c > 1 \text{ (transient)}. \end{cases} \tag{2.7}$$

This is the same as for simple random walk on $\mathbb{Z}^d$ with dimension

$$d = d(N,c) = (2\log N)/\log(N/c), \tag{2.8}$$

where N and c are such that $d \in \mathbb{N}$ (see Dawson et al., 2005, Eq. (3.2.2)). In particular, $d = d(N,1) = 2$ for all $N \in \mathbb{N}$. (In (2.8), we may in principle allow for non-integer dimensions $d \in (0,\infty)$ as well.)

Throughout the paper, we assume that

$$\limsup_{k\to\infty} \tfrac{1}{k} \log c_k < \log N. \tag{2.9}$$

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

2.1.3. *Block reshuffling-resampling.* The idea of the Cannings resampling mechanism 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 resampling measures

$$\underline{\Lambda} = (\Lambda_k)_{k\in\mathbb{N}_0} \in \mathcal{M}_f([0,1])^{\mathbb{N}_0}, \tag{2.10}$$

where $\mathcal{M}_f([0,1])$ denotes the set of finite non-negative measures on $[0,1]$, satisfying

$$\Lambda_0(\{0\}) = 0, \qquad \int_{(0,1]} \frac{\Lambda_0(\mathrm{d}r)}{r} = \infty, \tag{2.11}$$

and

$$\Lambda_k(\{0\}) = 0, \qquad \int_{(0,1]} \frac{\Lambda_k(\mathrm{d}r)}{r^2} < \infty. \qquad k \in \mathbb{N}. \tag{2.12}$$

The assumption in (2.12) guarantees that the local dynamics is a pure jump process (finitely many jumps in finite time intervals). A general Λ-Cannings process can have countably many jumps. We assume (2.12) to avoid explicit construction of the Cannings process via a limiting procedure: cutting off the support of Λ_k in the ϵ-vicinity of 0 and letting $\epsilon \downarrow 0$ afterwards, (see Remark 2.2 below).

Let $\Lambda_k^*(\mathrm{d}r) = \Lambda_k(\mathrm{d}r)/r^2$, $r \in (0,1]$. Set

$$\lambda_k = \Lambda_k((0,1]), \qquad \lambda_k^* = \Lambda_k^*((0,1]), \qquad k \in \mathbb{N}_0, \tag{2.13}$$

and assume that

$$\underline{\lambda} = (\lambda_k)_{k\in\mathbb{N}_0} \in (0,\infty)^{\mathbb{N}_0}. \tag{2.14}$$

We let individuals *reshuffle-resample* by carrying out the following two steps at once:

- For every $\eta \in \Omega_N$ and $k \in \mathbb{N}_0$, choose the block $B_k(\eta)$ at rate $1/N^{2k}$.
- First, each individual in $B_k(\eta)$ independently is moved to a uniformly random location in $B_k(\eta)$, i.e., a reshuffling takes place (see Fig. 2.2). After that, $r \in (0,1]$ is chosen according to the intensity measure Λ_k^* and $a \in E$ is drawn according to the current type distribution in $B_k(\eta)$, and each of the individuals in $B_k(\eta)$ independently is replaced by an individual of type a with probability r.

Note that the reshuffling-resampling affects all the individuals in a macro-colony simultaneously and in the same manner[1]. 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[2]

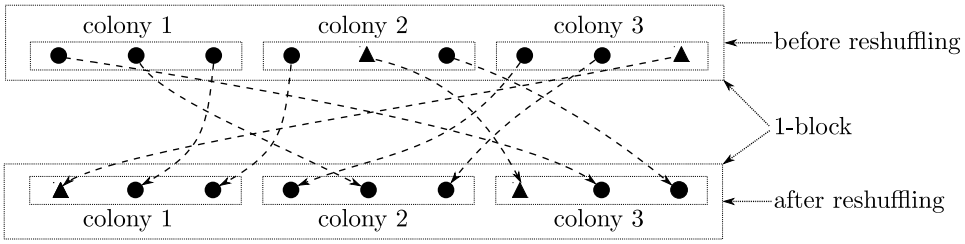

FIGURE 2.2. Random reshuffling in a 1-block on the hierarchical lattice of order $N = 3$, with $M = 3$ individuals of two types (• and ▲) per colony. *Note*: Typically, a random reshuffling does not preserve the number of individuals per colony, but in the example drawn here it does.

The first conditions in (2.11) and (2.12) make the resampling a *jump process*. Later we will add in diffusion by hand. The second condition in (2.11) guarantees that the population has a well-defined genealogy, in the sense that after a positive finite time most of the population in a colony descends from a finite number of ancestors: this is the so-called dust-free case (see, for example, Pitman, 1999). The second condition in (2.12) is needed to guarantee that in finite time a macro-colony is affected by finitely many reshuffling-resampling events, otherwise the resampling cannot be properly defined.

Throughout the paper, we assume that

$$\limsup_{k\to\infty} \tfrac{1}{k} \log \lambda_k^* < \log N. \tag{2.15}$$

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

[1]The average number of individuals in a colony does not change over time, provided it is the same for all colonies initially.

[2]Because the reshuffling is done first, the resampling always acts on a uniformly distributed state ("panmictic resampling"). 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. Reshuffling is biologically less compelling, but mathematically convenient.

2.1.4. *The generator and the martingale problem.* We are now ready to formally define the hierarchical Cannings process in terms of a martingale problem. The process arises as the continuum-mass limit of the individual-based model described in Sections 2.1.1–2.1.3. Namely, in each colony of size M, instead of recording the *numbers* of individuals of a given type we record the *empirical distribution* of the types and pass to the limit $M \to \infty$. For our understanding it is helpful to keep the individual-based model in mind, even though we *only* deal with the continuum process itself.

Let $\mathcal{P}(E)$ denote the set of probability measures on E, equipped with the topology of weak convergence [3]. 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

$$F(x) = \int_{E^n} \left(\bigotimes_{m=1}^{n} x_{\eta_m}\big(\mathrm{d}u^m\big) \right) f\big(u^1, \dots, u^n\big), \quad x = (x_\eta)_{\eta \in \Omega_N} \in \mathcal{P}(E)^{\Omega_N}, \tag{2.16}$$
$$n \in \mathbb{N}, \quad f \in C_{\mathrm{b}}(E^n, \mathbb{R}), \quad \eta_1, \dots, \eta_n \in \Omega_N.$$

The linear operator for the martingale problem

$$L^{(\Omega_N)} \colon \mathcal{F} \to C_{\mathrm{b}}\big(\mathcal{P}(E)^{\Omega_N}, \mathbb{R}\big) \tag{2.17}$$

has two parts,

$$L^{(\Omega_N)} = L^{(\Omega_N)}_{\mathrm{mig}} + L^{(\Omega_N)}_{\mathrm{res}}. \tag{2.18}$$

The *migration operator* is given by

$$(L^{(\Omega_N)}_{\mathrm{mig}} F)(x) = \sum_{\eta, \zeta \in \Omega_N} a^{(N)}(\eta, \zeta) \int_E (x_\zeta - x_\eta)(\mathrm{d}a)\, \frac{\partial F(x)}{\partial x_\eta}[\delta_a] \tag{2.19}$$

and the *reshuffling-resampling operator* by (well-defined by Remark 2.2 below)

$$\begin{aligned}(L^{(\Omega_N)}_{\mathrm{res}} F)(x) &= \sum_{\eta \in \Omega_N} \sum_{k \in \mathbb{N}_0} N^{-2k} \int_{(0,1]} \Lambda^*_k(\mathrm{d}r) \int_E y_{\eta,k}(\mathrm{d}a) \left[F\left(\Phi_{r,a,B_k(\eta)}(x)\right) - F(x) \right] \\ &\qquad + \sum_{\eta \in \Omega_N} (L^{d_0}_\eta F)(x),\end{aligned} \tag{2.20}$$

where

$$y_{\eta,k} = N^{-k} \sum_{\zeta \in B_k(\eta)} x_\zeta \tag{2.21}$$

is the k-block average of the components of x in $B_k(\eta)$, $\Phi_{r,a,B_k(\eta)} \colon \mathcal{P}(E)^{\Omega_N} \to \mathcal{P}(E)^{\Omega_N}$ is the *reshuffling-resampling map* acting as

$$\Big[\big(\Phi_{r,a,B_k(\eta)}\big)(x)\Big]_\zeta = \begin{cases} (1-r) y_{\eta,k} + r\delta_a, & \zeta \in B_k(\eta), \\ x_\zeta, & \zeta \in \Omega_N \backslash B_k(\eta), \end{cases} \tag{2.22}$$

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 $d_0 \geq 0$, acting on the colony x_η, given by

$$(L^{d_0}_\eta F)(x) = d_0 \int_E \int_E Q_{x_\eta}(\mathrm{d}u, \mathrm{d}v)\, \frac{\partial^2 F(x)}{\partial x_\eta^2}[\delta_u, \delta_v] \tag{2.23}$$

[3] For countable E, in $\mathcal{P}(E)$ weak convergence implies convergence in total variation norm.

with

$$Q_y(\mathrm{d}u,\mathrm{d}v) = y(\mathrm{d}u)\,\delta_u(\mathrm{d}v) - y(\mathrm{d}u)\,y(\mathrm{d}v), \qquad y \in \mathcal{P}(E), \tag{2.24}$$

the Fleming-Viot diffusion coefficient, and

$$\frac{\partial^2 F(x)}{\partial x_\eta^2}[\delta_u,\delta_v] = \frac{\partial}{\partial x_\eta}\left(\frac{\partial F(x)}{\partial x_\eta}[\delta_u]\right)[\delta_v], \quad u,v \in E. \tag{2.25}$$

Remark 2.2.

(a) Note that the right-hand side of (2.20) is well-defined because of assumption (2.12). Indeed, using (2.16) to Taylor-expand the inner integral in (2.20) in powers of r, and noting that the first order term in r cancels because of (2.21)–(2.22), we get

$$\int_E y_{\eta,k}(\mathrm{d}a)\left[F\left(\Phi_{r,a,B_k(\eta)}(x)\right) - F(x)\right] = O(r^2), \qquad r \downarrow 0, \tag{2.26}$$

with the error term uniform in x. Consequently,

$$\int_{(0,1]} \Lambda_k^*(\mathrm{d}r) \int_E y_{k,\eta}(\mathrm{d}a)\left[F(\Phi_{r,a,B_k(\eta)}) - F(x)\right] = O(1) \tag{2.27}$$

under (2.12).

(b) To remove assumption (2.12), we consider $(\Lambda_k)_{k\in\mathbb{N}_0}$ that are truncated below level $\delta > 0$. The corresponding process is constructed as in Section 4, with a well-posed dual process (see Theorems 3.1 and 4.2 below). It is easy to show that this dual processes has a limit as $\delta \downarrow 0$ that is a pure jump process (see Greven et al., 2014, Eq. (1.7)). Having a dual process for a duality function that is measure-determining in the dual variable, we can construct the forward *transition probabilities* for any two time-points, and define a process $X^{(\Omega_N)}$ with càdlàg paths that is the limit as $\delta \downarrow 0$ of the truncated process, which is of the kind we can deal with under assumption (2.12). Except for Proposition 2.3 and Theorem 3.1 below, which deal with the martingale problem, nothing changes. To characterize the limit process, we cannot work with the operator $L^{(\Omega_N)}$, as pointed out above. Instead, to characterize the martingale problem, we have to work with the *compensator*,

$$\left(\int_s^t L^{(\Omega_N)}(X^{(\Omega_N)}(u))\,\mathrm{d}u\right)_{t\geq s\geq 0}, \tag{2.28}$$

and show that this object is well-defined by using the fact that the dual process is. Consequently, all the arguments in Greven et al. (2014) based on $L^{(\Omega_N)}$ have to be adapted, basing them on the compensator. We refrain from giving the details of the construction.

□

The following proposition was proved in Greven et al. (2014).

Proposition 2.3 (Hierarchical martingale problem).
For every $x \in \mathcal{P}(E)^{\Omega_N}$, the martingale problem for $(L^{(\Omega_N)}, \mathcal{F}, \delta_x)$ is well-posed [4]*. The unique solution is a strong Markov process with the Feller property.* □

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

The Markov process arising as the solution of this martingale problem is denoted by

$$X^{(\Omega_N)} = (X^{(\Omega_N)}(t))_{t\geq 0}, \tag{2.29}$$

and is referred to as the $C_N^{\underline{c},\underline{\Lambda}}$-process on Ω_N. Proposition 2.3 does not actually need the second condition in (2.11).

This condition will be needed only later.

2.2. *The hierarchical Cannings process in random environment.* Our task in this section is to modify the first term in the right-hand side of (2.20) so as to include the effect of a random environment on the Cannings resampling mechanism. Section 2.2.1 defines the random environment, Section 2.2.2 introduces the generator in random environment.

2.2.1. *The random environment on the full tree.* Recall that Ω_N is the set of *leaves* of the tree in Fig. 2.1. To introduce the random environment, we need to consider the *full tree*, i.e.,

$$\Omega_N^{\mathbb{T}} = \bigcup_{k\in\mathbb{N}_0} \Omega_N^{(k)} \quad \text{with} \quad \Omega_N^{(k)} = \Omega_N / B_k(0), \tag{2.30}$$

where $\Omega_N/B_k(0)$ denotes the quotient group of Ω_N modulo $B_k(0)$, which can be identified with the layer of the tree situated at height k above the leaves. Indeed, because d_{Ω_N} is an ultrametric distance (recall (2.2)), for each $k \in \mathbb{N}_0$ the set Ω_N decomposes into disjoint balls of radius k, which can be labelled by the set $\Omega_N^{(k)}$. For $\xi \in \Omega_N^{\mathbb{T}}$, we write

$$|\xi| = \text{ the height of } \xi \text{ (counting from the leaves)}, \tag{2.31}$$

i.e., $|\xi| = k$ when $\xi \in \Omega_N^{(k)}$ for $k \in \mathbb{N}_0$, and we define

$$B_{|\xi|} \subset \Omega_N \tag{2.32}$$

to be the set of sites in Ω_N that lie below ξ (see Fig. 2.3). We can define the distance on the layer $\Omega_N^{(k)}$ as the graph distance to the most recent common ancestor, and the distance on the full tree $\Omega_N^{\mathbb{T}}$ as the *largest* of the two graph distances to the most recent common ancestor (recall Fig. 2.1). The latter will be denoted by $d_{\Omega_N^{\mathbb{T}}}$. We write

$$\mathrm{MC}_k(\eta) \tag{2.33}$$

to denote the *vertex in* $\Omega_N^{\mathbb{T}}$ *at height* $k \in \mathbb{N}_0$ *above* $\eta \in \Omega_N$ (which we can think of as "ancestor k of macro-colony η"). This site carries the rate for the random walk on Ω_N to become uniformly distributed on the $k-$ball around η. Since $\Omega_N^{\mathbb{T}}$ is isomorphic to $\Omega_N \times \mathbb{N}_0$, we sometimes write $\mathrm{MC}_k(\eta) = \xi = (\eta, k)$.

We want to make the reshuffling-resampling *spatially random*. To that end, we let

$$\underline{\Lambda}(\omega) = \left\{\Lambda^{\xi}(\omega)\colon\ \xi \in \Omega_N^{\mathbb{T}}\right\} \tag{2.34}$$

be a random field of $\mathcal{M}_f([0,1])$-valued resampling measures indexed by the tree.

- Throughout the paper, we *use the symbol* ω *to denote the random environment and the symbol* $\mathbb{P}$ *to denote the law of* ω.

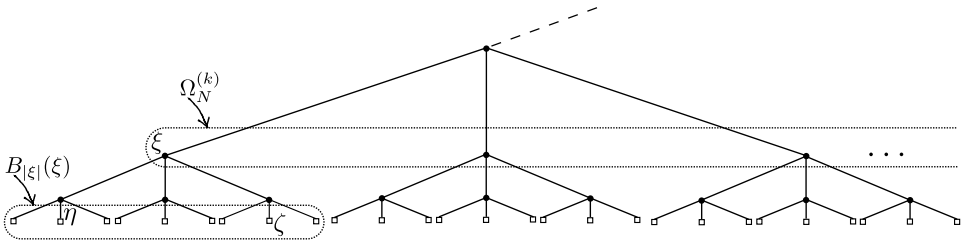

FIGURE 2.3. $\Omega_N^{\mathbb{T}}$ with $N = 3$, $\xi \in \Omega_N^{\mathbb{T}}$ with $|\xi| = k = 2$, and $\eta, \zeta \in B_{|\xi|}$. The elements of $\Omega_N^{\mathbb{T}}$ are the vertices of the tree (indicated by •'s and □'s). The elements of Ω_N are the leaves of the tree (indicated by □'s).

In what follows, we assume that $\Lambda^\xi(\omega)$ is of the form

$$\Lambda^\xi(\omega) = \lambda_{|\xi|}\chi^\xi(\omega), \tag{2.35}$$

where $\underline{\lambda} = (\lambda_k)_{k\in\mathbb{N}_0}$ is a deterministic sequence in $(0,\infty)$ (playing the role of modulation coefficients) and

$$\{\chi^\xi(\omega)\colon\ \xi \in \Omega_N^{\mathbb{T}}\} \tag{2.36}$$

is a random field of $\mathcal{M}_f([0,1])$-valued resampling measures that is *stationary under upward and sideway translations* in $\Omega_N^{\mathbb{T}}$ (see Fig. 2.3), and satisfies (2.11) when $\xi = 0$ and (2.12) when $\xi \neq 0$ [5].

Define

$$\lambda_k^*(\omega) = \int_{(0,1]} \frac{\Lambda^{\mathrm{MC}_k(0)}(\mathrm{d}r)(\omega)}{r^2}. \tag{2.37}$$

Throughout the paper, we assume the following analogue of (2.15):

$$\limsup_{k\to\infty} \frac{1}{k}\log\lambda_k^*(\omega) < \log N \quad \mathbb{P}\text{-a.e.} \tag{2.38}$$

Abbreviate

$$\rho^\xi(\omega) = \chi^\xi(\omega)((0,1]), \tag{2.39}$$

which is the total mass of $\chi^\xi(\omega)$. Clearly,

$$\{\rho^\xi(\omega)\colon\ \xi \in \Omega_N^{\mathbb{T}}\} \tag{2.40}$$

is a random field of $(0,\infty)$-valued total masses that is also stationary under translations in $\Omega_N^{\mathbb{T}}$. Throughout the paper, we assume that

$$\mathbb{E}[\rho^\xi(\omega)] = 1, \qquad \mathbb{E}[(\rho^\xi(\omega))^2] = C \in (1,\infty), \tag{2.41}$$

and that the sigma-algebra at infinity associated with (2.39), defined by

$$\mathcal{T} = \bigcap_{L\in\mathbb{N}_0} \mathcal{F}_L, \qquad \mathcal{F}_L = \sigma\Big(\rho^\xi(\cdot)\ \Big|\ \xi \in \Omega_N^{\mathbb{T}}\colon\ d_{\Omega_N^{\mathbb{T}}}(0,\xi) \geq L\Big), \tag{2.42}$$

is trivial, i.e., all its events have probability 0 or 1 under the law $\mathbb{P}$. For one of the theorems below we need to strengthen (2.41) to

$$\mathbb{E}[\rho^\xi(\omega)] = 1, \qquad \exists\,\delta > 0\colon\ \delta \leq \rho^\xi(\omega) \leq \delta^{-1}\ \forall\,\xi \in \Omega_N^{\mathbb{T}} \text{ for } \mathbb{P}\text{-a.e. } \omega. \tag{2.43}$$

[5] We will only need translation invariance in the upward and sideway direction, starting from some level in $\mathbb{N}$. So there is no contradiction with the fact that (2.11) and (2.12) differ.

2.2.2. *The generator in random environment.* Throughout the sequel, we use the symbols η, ζ to denote elements of Ω_N and the symbol ξ to denote elements of $\Omega_N^{\mathbb{T}}$.

In random environment, we keep the definitions in Section 2.1.4, but we replace the *reshuffling-resampling operator* $L_{\mathrm{res}}^{(\Omega_N)}$ in (2.20) by

$$\begin{aligned}(L_{\mathrm{res}}^{(\Omega_N)}(\omega)F)(x) \\ = \sum_{\xi\in\Omega_N^{\mathbb{T}}} N^{-2|\xi|} \int_{(0,1]} (\Lambda^\xi(\omega))^*(\mathrm{d}r) \int_E y_\xi(\mathrm{d}a) \left[F\left(\Phi_{r,a,B_{|\xi|}}(x)\right) - F(x)\right] \\ + \sum_{\eta\in\Omega_N} (L_\eta^{d_0}F)(x) \end{aligned} \tag{2.44}$$

with $(\Lambda^\xi(\omega))^*(\mathrm{d}r) = \Lambda^\xi(\omega)(\mathrm{d}r)/r^2$, $r \in (0,1]$, where $y_\xi \in \mathcal{P}(E)$ is given by

$$y_\xi = N^{-|\xi|} \sum_{\zeta\in B_{|\xi|}} x_\zeta, \quad \xi \in \Omega_N^{\mathbb{T}}, \tag{2.45}$$

and $\Phi_{r,a,B_{|\xi|}} \colon \mathcal{P}(E)^{\Omega_N} \to \mathcal{P}(E)^{\Omega_N}$ is the *reshuffling-resampling map* acting as

$$\left[\left(\Phi_{r,a,B_{|\xi|}}\right)(x)\right]_\zeta = \begin{cases} (1-r)y_\xi + r\delta_a, & \zeta \in B_{|\xi|}, \\ x_\zeta, & \zeta \in \Omega_N \backslash B_{|\xi|}, \end{cases} \qquad r \in [0,1],\ a \in E,\ \xi \in \Omega_N^{\mathbb{T}}. \tag{2.46}$$

The difference between (2.20) and (2.44) is that the resampling in blocks occurs according to the resampling measure associated with the center of the block, labelled by $\Omega_N^{\mathbb{T}}$. The full generator is

$$L^{(\Omega_N)}(\omega) = L_{\mathrm{mig}}^{(\Omega_N)} + L_{\mathrm{res}}^{(\Omega_N)}(\omega) \tag{2.47}$$

with $L_{\mathrm{mig}}^{(\Omega_N)}$ the migration operator in (2.19).

3. Main theorems

In Section 3.1, we present results for fixed N, in Section 3.2, for $N \to \infty$, the hierarchical mean-field limit. In Section 3.3, we summarize the effects of the random environment. Throughout the paper, the environment ω is fixed and we use the symbol $\mathcal{L}[W]$ to denote the law of a random variable W.

3.1. *Results for fixed N.* Section 3.1.1 establishes the well-posedness of the martingale problem, Section 3.1.2 the convergence to an equilibrium that depends on ω.

3.1.1. *Well-posedness of the martingale problem.* We begin by establishing that the martingale problem characterizes the process uniquely and specifies a strong Markov process.

Theorem 3.1 (Well-posedness of the martingale problem). *Fix $N \in \mathbb{N}\backslash\{1\}$. For $\mathbb{P}$-a.e. ω and every $x \in \mathcal{P}(E)^{\Omega_N}$, the $(L^{(\Omega_N)}(\omega), \mathcal{F}, \delta_x)$-martingale problem is well-posed. The unique solution is a strong Markov process with the Feller property.* □

The Markov process arising as the solution of the martingale problem is denoted by

$$X^{(\Omega_N)}(\omega) = (X^{(\Omega_N)}(\omega;t))_{t\geq 0} = ((X_\eta(\omega,t))_{\eta\in\Omega_N})_{t\geq 0}, \tag{3.1}$$

and is referred to as the hierarchical Cannings process on Ω_N in the environment ω. Theorem 3.1 does not actually need the second condition in (2.11). This condition will be needed only later.

3.1.2. *Dichotomy: coexistence versus clustering.* We next show that the law of our process converges to a limit law that depends on ω. Since the law $\mathbb{P}$ of ω is assumed to be stationary and ergodic under translations in $\Omega_N^{\mathbb{T}}$ (recall Section 2.2.1), the limit law is stationary and ergodic under translations in Ω_N, provided its initial state is.

Theorem 3.2 (Equilibrium). *Fix $N \in \mathbb{N}\backslash\{1\}$. Suppose that, under the law $\mathbb{P}$ of ω, the law of the initial state $X^{(\Omega_N)}(\omega;0)$ is stationary and ergodic under translations in Ω_N, with mean single-coordinate measure*

$$\theta = \mathbb{E}\big[X_0^{(\Omega_N)}(\omega;0)\big] \in \mathcal{P}(E). \tag{3.2}$$

Then, for $\mathbb{P}$-a.e. ω, there exists an equilibrium measure $\nu_\theta^N(\omega) \in \mathcal{P}(\mathcal{P}(E)^{\Omega_N})$, arising as

$$\lim_{t\to\infty} \mathcal{L}\big[X^{(\Omega_N)}(\omega;t)\big] = \nu_\theta^N(\omega), \tag{3.3}$$

satisfying

$$\int_{\mathcal{P}(E)^{\Omega_N}} x_0\, \nu_\theta^N(\omega)(\mathrm{d}x) = \theta. \tag{3.4}$$

Moreover, under the law $\mathbb{P}$, $\nu_\theta^N(\omega)$ is stationary and ergodic under translations in Ω_N. □

Note that $\nu_\theta^N(\omega)$ depends on ω even though its mean single-coordinate measure θ (which is determined by the initial state) does not. The proof of Theorem 3.2 is based on a computation with the dual hierarchical Cannings process, which allows us to control second moments. As we will see in Section 5, in random environment this computation is delicate because it involves two random walks in the same environment, and the *difference* of these two random walks is *not* a random walk itself, like in the average environment.

Using the stationarity and ergodicity of ν_θ^N, we next identify the parameter regime for which $\nu_\theta^N(\omega)$ is a *multi-type equilibrium* (= coexistence given ω), i.e.,

$$\sup_{f\in C_{\mathrm{b}}(E)} \int_{\mathcal{P}(E)^{\Omega_N}} \nu_\theta^N(\omega)(\mathrm{d}x) \int_E [f(u) - \langle\theta, f\rangle]^2\, x_0(\mathrm{d}u) > 0 \tag{3.5}$$

with $\langle\theta, f\rangle = \int_E \theta(\mathrm{d}u) f(u)$, respectively, a *mono-type equilibrium* (= clustering given ω), i.e.,

$$\nu_\theta^N(\omega) = \int_E \delta_{(\delta_u)^{\Omega_N}}\, \theta(\mathrm{d}u). \tag{3.6}$$

The two regimes are complementary. In the latter regime the system grows mono-type clusters that eventually cover any finite subset of Ω_N, in the sense that the probability to see only one type in the set at time t tends to 1 as $t \to \infty$. (Types may or may not change infinitely often in the course of time.)

Theorem 3.3 (Dichotomy for finite N). *Fix $N \in \mathbb{N}\backslash\{1\}$ and assume (2.43).*

(a) *Let* $\mathcal{C}_N = \{\omega\colon$ *in* ω *coexistence occurs*$\}$. *Then* $\mathbb{P}(\mathcal{C}_N) \in \{0,1\}$.
(b) $\mathbb{P}(\mathcal{C}_N) = 1$ *if and only if (recall* (2.34)*):*

$$\sum_{k\in\mathbb{N}_0} \frac{1}{c_k + N^{-1}\lambda_{k+1}} \sum_{l=0}^{k} \lambda_l < \infty. \tag{3.7}$$

□

Cox and Klenke (2000) give a criterion in the clustering regime for when the type at a given site changes infinitely often. For interacting Fleming-Viot processes they show that this happens as soon as θ is not a δ-measure. Because of the reasoning in Section 5, we therefore get the following.

Corollary 3.4 (Change of types). *In the clustering regime, if* $\theta \neq \delta_u$ *for some* $u \in E$*, then at every site the type changes infinitely often.* □

3.2. *Results for* $N \to \infty$. Our remaining theorems capture the space-time scaling behaviour of our process in the hierarchical mean-field limit $N \to \infty$. In this limit, the degree of recurrence/transience $\gamma(N)$ tends to 0, while the Hausdorff dimension $d(N)$ tends to 2 (recall Remark 2.1), so that our process is *near-critical* [6].

In Section 3.2.1, we introduce a key process, called the *McKean-Vlasov process*, which naturally arises in this limit [7] In Section 3.2.2, we define the random environment for $N = \infty$. In Section 3.2.3 we look at the block averages on successive space-time scales and show that as $N \to \infty$ these converge to a sequence of McKean-Vlasov processes with *renormalized volatilities*. In Section 3.2.5, we identify the scaling behaviour of the volatility on hierarchical scale k in the limit as $k \to \infty$, which leads to various different cases as a function of $\underline{c}$ and $\underline{\Lambda}$. In Section 3.2.4, we identify the parameter regimes that correspond to coexistence, respectively, clustering. In Section 3.2.6, we link the different cases of scaling to five universality classes of cluster formation.

3.2.1. *McKean-Vlasov process.* We need some definitions and basic facts about the McKean-Vlasov process from Greven et al. (2014).

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

$$F(y) = \int_{E^n} y^{\otimes n}(\mathrm{d}u)\, f(u), \qquad y \in \mathcal{P}(E),\ n \in \mathbb{N},\ f \in C_{\mathrm{b}}(E^n, \mathbb{R}). \tag{3.8}$$

For $c, d \in [0,\infty)$, $\Lambda \in \mathcal{M}_f([0,1])$ subject to (2.11) 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

$$L_\theta^{c,d,\Lambda} = L_\theta^c + L^d + L^\Lambda \tag{3.9}$$

[6] View Ω_N as embedded in its continuum version $\{(\eta^l)_{l\in\mathbb{Z}} \in \{0,1,\dots,N-1\}^{\mathbb{Z}}\colon \sum_{l\in\mathbb{N}_0} \eta^l < \infty\}$ (recall (2.1)), and determine the growth of the volume of a ball as a function of its radius to find the dimension $d(N)$ in the metric $\eta \mapsto e^{|\eta|}$.

[7] For the special initial state of the mean-field system considered in Proposition 3.5 below, it is not visible that θ is to be viewed as a functional of the law of the state. For general initial states (i.e., exchangeable random states of the mean-field system), θ is a functional of the law of the state, namely, it is de Finetti's measure, which specifies the limit of the averages of all the components. Consequently, the system is a *non-linear* Markov process, and the name McKean-Vlasov process rather than "process with immigration-emigration" is appropriate.

acting on $F \in \mathcal{F}$ as (recall (2.24))

$$\begin{aligned}(L_\theta^c F)(y) &= c \int_E (\theta - y)\,(\mathrm{d}a)\,\frac{\partial F(y)}{\partial y}[\delta_a],\\ (L^d F)(y) &= d \int_E \int_E Q_y(\mathrm{d}u, \mathrm{d}v)\,\frac{\partial^2 F(y)}{\partial y^2}[\delta_u, \delta_v],\\ (L^\Lambda F)(y) &= \int_{(0,1]} \Lambda^*(\mathrm{d}r) \int_E y(\mathrm{d}a)\,\big[F\big((1-r)y + r\delta_a\big) - F(y)\big].\end{aligned} \tag{3.10}$$

The above three parts of $L_\theta^{c,d,\Lambda}$ correspond to:

(1) a *drift* towards θ of strength c ("immigration-emigration");
(2) a *Fleming-Viot diffusion* with *volatility* d ("Moran resampling");
(3) a *Cannings 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* at rate c from a constant source with type distribution $\theta \in \mathcal{P}(E)$, *emigration* at rate c to a cemetery state, diffusive resampling at rate d, and Λ-resampling.

The following proposition was proved in Greven et al. (2014).

Proposition 3.5 (McKean-Vlasov martingale problem).

(a) *For every $y \in \mathcal{P}(E)$, the martingale problem for $(L_\theta^{c,d,\Lambda}, \mathcal{F}, \delta_y)$ is well-posed. The unique solution is a strong Markov process with the Feller property.*
(b) *For every $c \in (0,\infty)$, the solution from* (a) *is ergodic in time: as $t \to \infty$ its law converges to the unique equilibrium measure $\nu_\theta^{c,d,\Lambda}$, irrespective of the initial state. For $c = 0$, the solution from* (a) *is not ergodic in time, and $\nu_\theta^{0,d,\Lambda}$ is defined as the unique equilibrium measure, as $t \to \infty$, when the initial state is chosen to be $y = \theta$, i.e., $\nu_\theta^{0,d,\Lambda} = \int_E \theta(\mathrm{d}a)\,\delta_a$.*
(c) *For $c > 0$,*

$$\nu_\theta^{c,d,\Lambda} = \nu_\theta^{1,d/c,\Lambda/c}. \tag{3.11}$$

□

Denote by

$$Z_\theta^{c,d,\Lambda} = \big(Z_\theta^{c,d,\Lambda}(t)\big)_{t\geq 0}, \quad Z_\theta^{c,d,\Lambda}(0) = \theta, \tag{3.12}$$

the solution of the martingale problem in Proposition 3.5 for the special choice $y = \theta$. This is called the *McKean-Vlasov process* with parameters c, d, Λ and initial state θ.

3.2.2. *Random environment for $N = \infty$.* In order to be able to pass to the limit $N \to \infty$, we need to define a random environment for $N = \infty$ in which all the random environments for finite N are embedded. To that end, define $\Omega_\infty = \oplus_{\mathbb{N}} \mathbb{N}_0$, and let (recall (2.30))

$$\Omega_\infty^{\mathbb{T}} = \bigcup_{k\in\mathbb{N}_0} \Omega_\infty^{(k)} \quad \text{with} \quad \Omega_\infty^{(k)} = \Omega_\infty / B_k(0). \tag{3.13}$$

Note that for any $N \in \mathbb{N}$ there is a natural embedding of $\Omega_N^{\mathbb{T}}$ into $\Omega_\infty^{\mathbb{T}}$. Similarly as in Section 2.2.1, we let

$$\underline{\Lambda}(\omega) = \big\{\Lambda^\xi(\omega)\colon\ \xi \in \Omega_\infty^{\mathbb{T}}\big\} \tag{3.14}$$

be a random field of $\mathcal{M}_f([0,1])$-valued resampling measures index by the full tree, where ω again denotes the random environment. We retain the symbol $\mathbb{P}$ for the law of of ω. As in (2.35)–(2.42), we assume that $\Lambda^\xi(\omega) = \lambda_{|\xi|}\chi^\xi(\omega)$ where, under the law $\mathbb{P}$, $\{\chi^\xi(\omega)\colon\ \xi \in \Omega_\infty^{\mathbb{T}}\}$ is stationary under translations in $\Omega_\infty^{\mathbb{T}}$, and is such that the total masses $\rho^\xi(\omega) = \chi^\xi(\omega)((0,1])$ have first moment equal to 1, second moment finite, and a trivial sigma-algebra at infinity. For any $N \in \mathbb{N}$, the natural restriction of the random field in (3.14) equals the random field in (2.34).

3.2.3. *Renormalization via block averages.* For each $k \in \mathbb{N}_0$, we look at the *k-block averages* defined by (recall Fig. 2.1)

$$Y_{\eta,k}^{(\Omega_N)}(\omega;t) = \frac{1}{N^k} \sum_{\zeta \in B_k(\eta)} X_\zeta^{(\Omega_N)}(\omega;t), \qquad \eta \in \Omega_N, \tag{3.15}$$

which constitute a *renormalization of space* where the component η is replaced by the average of the components in $B_k(\eta)$. After a corresponding *renormalization of time* where t is replaced by tN^k, i.e., t is the associated macroscopic time variable, we obtain a *renormalized* interacting system

$$\left(\left(Y_{\eta,k}^{(\Omega_N)}(\omega;tN^k)\right)_{\eta\in\Omega_N}\right)_{t\geq 0}, \qquad k \in \mathbb{N}_0,\ \eta \in \Omega_N, \tag{3.16}$$

which is constant in $B_k(\eta)$ and can be viewed as an interacting system indexed by the set $\Omega_N^{(k)}$ (see Fig. 2.1). This provides us with a *sequence of renormalized interacting systems*, which for fixed N are *not* Markov.

The key ingredient to study the $N \to \infty$ limit of (3.16) is the following. Let $\underline{d} = (d_k)_{k\in\mathbb{N}_0}$ be the sequence of *volatility constants* defined recursively as (compare with Greven et al., 2014, Eq. (1.45))

$$d_{k+1} = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{c_k(\mu_k\rho + d_k)}{c_k + (\mu_k\rho + d_k)}\right], \quad k \in \mathbb{N}_0, \tag{3.17}$$

where

$$\mu_k = \tfrac{1}{2}\lambda_k, \quad k \in \mathbb{N}_0, \tag{3.18}$$

ρ is the $(0,\infty)$-valued random variable whose law $\mathcal{L}_\rho$ is the same as that of $\rho^0(\omega)$ under $\mathbb{P}$ (recall (2.41–2.42)), and $\mathbb{E}_{\mathcal{L}_\rho}$ is expectation w.r.t. $\mathcal{L}_\rho$. For fixed $\underline{c}$, $\underline{\Lambda}$ and d_0, the recursion in (3.17) determines $\underline{d}$. The right-hand side is the *average of a random Möbius transformation that depends on* ρ. Recall that ρ has mean 1 and strictly positive variance.

Heuristics behind the recursion formula for the volatilities. In order to understand the recursion formula in (3.17), we consider the 1-block around the origin 0 on time scale Nt and let $N \to \infty$. Note that, in this limit, the time scales for the jumps to different levels separate (recall (2.6)), so that we can focus on each of the time scales separately.

If we randomly draw two lineages from the 1-block and ask whether they have a common ancestor some time back (so that they are of the same type), then we get exactly the event that generates the variance of the 1-block average (otherwise the lineages and their types would be independent and would have an asymptotically vanishing contribution to the variance). The fact that the lineages behave like a spatial coalescent follows from the duality formally introduced in Section 4. The lineages have to meet in order to have a common ancestor, which takes them a time

of order Nt. Note that triples of lineages have a negligible probability to meet at times of order Nt in the limit of $N \to \infty$.

If the lineages meet, then they may coalesce. This happens at rate

$$\lambda^{(\eta,0)}(\omega) = \Lambda^{(\eta,0)}((0,1])(\omega) \tag{3.19}$$

when they both sit at $\eta \in \Omega$. However, they may also move before they coalesce, i.e., make a migration jump away, which happens with probability $2c_0/(2c_0+\lambda^{(\eta,0)}(\omega))$. Hence the effective coalescence rate is $\lambda^{(\eta,0)}(\omega)[2c_0/(2c_0+\lambda^{(\eta,0)}(\omega))]$. Since the vertex where the lineages meet is uniformly distributed over the 1-block, the average rate is given by

$$\mathbb{E}\left[\frac{2c_0\lambda_0\rho(\omega)}{2c_0+\lambda_0\rho(\omega)}\right], \tag{3.20}$$

where we use that $\lambda^{(\eta,0)}(\omega)$ has the same distribution as $\lambda_0\rho(\omega)$ (recall (2.35)–(2.39)). If we would have a diffusive part as well, at constant rate $2d_0$, then the lineages would coalesce at the same rate but with $\lambda(\omega)$ replaced by $2d_0+\lambda(\omega)$. Since the volatility turns out to be equal to this rate, we get the recursion formula

$$2d_1 = \mathbb{E}\left[\frac{2c_0(2d_0+\lambda_0\varrho(\omega))}{2c_0+(2d_0+\lambda_0\varrho(\omega))}\right]. \tag{3.21}$$

By the same reasoning for k-blocks on time scale tN^k, we get a heuristic explanation for the recursion formula in (3.17).

Our next theorem states that for each $k \in \mathbb{N}_0$ the k-block averages in the limit as $N \to \infty$ evolve according to the McKean-Vlasov process defined in Section 3.2.1 with certain k-dependent parameters.

Theorem 3.6 (**Hierarchical mean-field limit and renormalization**). *Suppose that for each N the random field $X^{(\Omega_N)}(\omega;0)$ is the restriction to Ω_N of a random field $X(\omega)$ indexed by $\Omega_\infty = \bigoplus_{\mathbb{N}} \mathbb{N}_0$ that is i.i.d. with single-component mean $\theta \in \mathcal{P}(E)$. Then, for $\mathbb{P}$-a.e. ω and every $k \in \mathbb{N}$ and $\eta \in \Omega_\infty$,*

$$\lim_{N\to\infty} \mathcal{L}\left[\left(Y^{(\Omega_N)}_{\eta,k}(\omega;tN^k)\right)_{t\geq 0}\right] = \mathcal{L}\left[\left(Z_\theta^{c_k,d_k,\Lambda^{\mathrm{MC}_k(\eta)}(\omega)}(t)\right)_{t\geq 0}\right], \tag{3.22}$$

where (recall (2.33), (2.35) and (3.12))

$$\mathrm{MC}_k(\eta) = \text{ unique site in } \Omega_\infty^{\mathbb{T}} \text{ at height } k \text{ above } \eta \in \Omega_\infty, \tag{3.23}$$

i.e., the label of the block (= macro-colony) of radius k in Ω_∞ around $\eta \in \Omega_\infty$ (see Fig. 2.1*). The same is true for $k=0$ when the initial condition for the McKean-Vlasov process in the right-hand side of (3.22) is $Z_\theta^{c_0,d_0,\Lambda^\eta(\omega)}(0) = X^{(\Omega_N)}(\omega;0)$ instead of $Z_\theta^{c_0,d_0,\Lambda^\eta(\omega)}(0)=\theta$.* □

Note that among the parameters $c_k, d_k, \Lambda^{\mathrm{MC}_k(\eta)}(\omega)$ of the limiting McKean-Vlasov process, the volatility d_k is the result of a *self-averaging* with respect to the random environment up to and including level k, as exemplified by (3.17). *It is through this recursion relation that the renormalization manifests itself.*

Our next theorem looks at successive block averages simultaneously.

Theorem 3.7 (**Multi-scale analysis and the interaction chain**). *Let $(t_N)_{N\in\mathbb{N}}$ be such that*

$$\lim_{N\to\infty} t_N = \infty \text{ and } \lim_{N\to\infty} t_N/N = 0. \tag{3.24}$$

Then, for $\mathbb{P}$-a.e. ω, every $j \in \mathbb{N}$ and every $\eta \in \Omega_\infty$,

$$\lim_{N\to\infty} \mathcal{L}\left[\left(Y^{(\Omega_N)}_{\eta,k}(\omega; t_N N^k)\right)_{k=j+1,j,\dots,0}\right] = \mathcal{L}\left[\left(M^{(j)}_{\eta,k}(\omega)\right)_{k=-(j+1),-j,\dots,0}\right], \tag{3.25}$$

where $M^{(j)}_\eta(\omega) = (M^{(j)}_{\eta,k}(\omega))_{k=-(j+1),-j,\dots,0}$ is the time-inhomogeneous Markov chain with initial state

$$M^{(j)}_{\eta,-(j+1)}(\omega) = \theta, \tag{3.26}$$

and transition kernel from time $-(k+1)$ to $-k$ given by

$$K_{\eta,k}(\omega;\theta,\cdot) = \nu_\theta^{c_k,d_k,\Lambda^{\mathrm{MC}_k(\eta)}(\omega)}(\cdot). \tag{3.27}$$

□

The right-hand side of (3.24) describes the large space-time scaling behaviour of our hierarchical Cannings process.

Definition 3.8 (Interaction chain). $M^{(j)}_\eta(\omega)$ is called the interaction chain at level j at location $\eta \in \Omega_\infty$ given ω. □

Remark 3.9. Theorem 3.7 only specifies the limiting distribution of the one-dimensional spatial marginals, i.e., the single interaction chains. Similarly as in Dawson et al. (1995, Section 0e), it is possible to also specify the joint distribution of the interaction chains, which can be viewed as a *field of Markov chains indexed by* $\Omega^{\mathbb{T}}_\infty$. □

An important characteristic of $M^{(j)}_\eta$ is the variance of $M^{(j)}_{\eta,0}$, calculated as

$$\mathrm{Var}\langle M^{(j)}_{\eta,0}, f\rangle = \prod_{k=0}^{j} \frac{2c_k}{2c_k + \lambda_k\rho_k + 2d_k}\,\mathrm{Var}_\theta(f). \tag{3.28}$$

This shows that a key ingredient for $M^{(j)}_\eta$ is the sequence of volatilities $\underline{d} = (d_k)_{k\in\mathbb{N}_0}$ and the way this sequence grows or decays. How is this affected by the randomness of the environment?

Our next theorem shows that the volatility d_k in the random environment can be sandwiched between the volatility d^0_k in the zero environment ($\mathcal{L}_\rho = \delta_0$, i.e., the system without resampling) and the volatility d^1_k in the average environment ($\mathcal{L}_\rho = \delta_1$, i.e., the system with average resampling).

Theorem 3.10 (Randomness lowers volatility). *If $d^0_0 = d_0 = d^1_0$, then $d^0_k < d_k < d^1_k$ for all $k \in \mathbb{N}$.* □

3.2.4. *Dichotomy for the interaction chain.* How are the qualitative properties of the Cannings process for large N reflected in the interaction chain? What about the dichotomy clustering versus coexistence? Before answering these questions we need to first establish the existence of the entrance law of the interaction chain from level ∞, which we will obtain from the level j interaction chain as limit $j \to \infty$. With this object, we can address the question of coexistence versus clustering.

Proposition 3.11 (Entrance law of interaction chain exists).
The limit as $j \to \infty$ of $M^{(j)}_\eta$ exists. □

The object corresponding to the equilibrium of the stochastic system for finite N in the hierarchical mean-field limit $N \to \infty$ is the *field of entrance laws* of the interaction chain from level ∞ (recall Remark 3.9), in particular, its marginal law $\Pi_\eta \nu_\theta(\omega)$ at level 0 in η, which is element of $\mathcal{P}(\mathcal{P}(E))$.

Definition 3.12 (**Entrance law of interaction chain**). For $\mathbb{P}$-a.e. ω and all $\eta \in \Omega_\infty$,

$$\lim_{j\to\infty} \mathcal{L}\left[M^{(j)}_{\eta,0}(\omega)\right] = \Pi_\eta \nu_\theta(\omega), \tag{3.29}$$

where $\nu_\theta(\omega) \in \mathcal{P}(\mathcal{P}(E))^{\Omega_\infty}$ is the entrance law from level ∞ of the (tree-indexed) interaction chain at level 0, and $\Pi_\eta \nu_\theta(\omega)$ denotes the projection of $\nu_\theta(\omega)$ on η. □

Our next theorem is indeed the analogue of Theorem 3.3 for $N \to \infty$. In this limit, coexistence and clustering in ω are defined for $(M^{(\infty)}_{\eta,0})_{\eta\in\Omega_N}$ in the same way as in (3.5)–(3.6).

Theorem 3.13 (**Dichotomy for $N = \infty$**).

(a) *Let $\mathcal{C} = \{\omega\colon$ in ω coexistence occurs$\}$. Then $\mathbb{P}(\mathcal{C}) \in \{0, 1\}$.*

(b) *$\mathbb{P}(\mathcal{C}) = 1$ if and only if*

$$\sum_{k\in\mathbb{N}_0} \frac{1}{c_k} \sum_{l=0}^{k} \lambda_l < \infty. \tag{3.30}$$

□

Note that condition (3.30) is the limit of condition (3.7) as $N \to \infty$. In fact, the two conditions are equivalent when the following *weak regularity condition* holds:

$$\text{either} \quad \limsup_{k\to\infty} \frac{\lambda_{k+1}}{c_k} < \infty \quad \text{or} \quad \liminf_{k\to\infty} \left(\frac{\lambda_{k+1}}{c_k} \wedge \frac{\lambda_k}{\lambda_{k+1}}\right) > 0. \tag{3.31}$$

An important question is whether the equilibrium measure $\nu_\theta(\omega)$ is the limit as $N \to \infty$ of the equilibrium measure $\nu^N_\theta(\omega)$ (recall (3.3)). The answer is yes. We only prove the following.

Corollary 3.14 (**Hierarchical mean field limit of equilibrium**).
For $\mathbb{P}$-a.e. all ω and all $\eta \in \Omega_\infty$,

$$\lim_{N\to\infty} \Pi_\eta \nu^N_\theta = \Pi_\eta \nu_\theta. \tag{3.32}$$

□

3.2.5. *Scaling of the volatility.* We are interested in the behaviour of the variance of the interaction chain $M^{(j)}$ as $j \to \infty$, since this allows us to identify *universality classes* for the scaling behaviour of our stochastic system. From the variance formula, we see that $(d_k)_{k\in\mathbb{N}_0}$ is the key input, and so we study this sequence first. Note that the variance formula only depends on the ratios $d_k/c_k, \mu_k/d_k$, which we encounter and analyze below.

Our next two theorems identify the scaling behaviour of d_k as $k \to \infty$ *in the regime of clustering*. The first theorem considers the case of *polynomial coefficients*, i.e.,

$$c_k \sim L_c(k)\, k^a, \qquad \mu_k \sim L_\mu(k)\, k^b, \qquad k \to \infty, \tag{3.33}$$

with $a, b \in \mathbb{R}$ and L_c, L_μ slowly varying at infinity. In what follows, we assume that

$$K = \lim_{k\to\infty} \frac{\mu_k}{c_k} \in [0,\infty], \qquad L = \lim_{k\to\infty} \frac{k^2\mu_k}{c_k} \in [0,\infty], \tag{3.34}$$

exist and write K_k and L_k for the respective sequences. There are *five cases* according to the values of K and L. Four of these, labelled (a)–(d), we can analyze in detail. For the remaining case, labelled (e), see Remark (3.16). For cases (c)–(d), we need extra regularity conditions on L_c, L_μ in (3.33), for which we refer the reader to Greven et al. (2014, Eqs. (1.79)–(1.81)).

Theorem 3.15 (**Scaling of the Fleming-Viot volatility: polynomial coefficients**).
Under the polynomial scaling assumptions (3.33)–(3.34)*, the following cases apply:*

(a) *If* $K = \infty$*, then* $\lim_{k\to\infty} d_k/c_k = 1$.
(b) *If* $K \in (0,\infty)$*, then* $\lim_{k\to\infty} d_k/c_k = M$ *with* $M \in (0,1)$ *the unique solution of the equation*

$$M = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{(K\rho + M)}{1 + (K\rho + M)}\right]. \tag{3.35}$$

(c) *If* $K = 0$ *and* $L = \infty$*, then* $\lim_{k\to\infty} d_k/\sqrt{c_k\mu_k} = 1$.
(d) *If* $K = 0$*,* $L \in [0,\infty)$ *and* $a \in (-\infty, 1)$*, then* $\lim_{k\to\infty} \sigma_k d_k = M$ *with* $\sigma_k = \sum_{l=0}^{k-1}(1/c_l)$ *and* $M \in [1,\infty)$ *given by*

$$M = \tfrac{1}{2}\left[1 + \sqrt{1 + 4L/(1-a)^2}\right]. \tag{3.36}$$

□

Remark 3.16. It is straightforward to check with the help of (3.33)–(3.34) that all four cases (a)-(d) correspond to choices of $\underline{c}$ and $\underline{\lambda}$ for which clustering holds, i.e., the sum in (3.30) diverges (note that $\lim_{k\to\infty} \sigma_k = \infty$ in case (d)). However, they are not exhaustive: there is a fifth case (e), corresponding to $K = 0$, $L \in [0,\infty)$, $a = 1$ and $\lim_{k\to\infty} \sigma_k = \infty$, for which we have no scaling result. This case lies at the border of the clustering regime. An example is $c_k \sim k(\log k)^\gamma$, $\gamma \in (-\infty, 1]$, and $\mu_k = k^{-2}c_k$, which we were able to handle in the deterministic model in Greven et al. (2014), but cannot handle in the random model treated here. □

The second theorem considers the case of *exponential coefficients*, i.e.,

$$c_k = c^k \bar{c}_k, \qquad \mu_k = \mu^k \bar{\mu}_k \tag{3.37}$$

with $c, \mu \in (0,\infty)$ and $\bar{c}_k, \bar{\mu}_k$ satisfying (3.33) with exponents a, b. We further assume that

$$\bar{K} = \lim_{k\to\infty} \frac{\bar{\mu}_k}{\bar{c}_k} \in [0,\infty], \text{ we write } \bar{K}_k = \frac{\bar{\mu}_k}{\bar{c}_k}. \tag{3.38}$$

exists.

Theorem 3.17 (**Scaling of the Fleming-Viot volatility: exponential coefficients**).
Under the exponential scaling assumptions in (3.37)–(3.38)*, the following cases apply (cf.* Theorem 3.15*):*

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

(B) [Like Case (b)] $c = \mu$ *and* $\bar{K} \in (0,\infty)$*:* $\lim_{k\to\infty} d_k/c_k = \bar{M}/c$ *with* $\bar{M} \in (0,1)$ *the unique solution of the equation*

$$\bar{M} = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{(cK\rho + \bar{M})}{c + (cK\rho + \bar{M})}\right]. \tag{3.39}$$

(C) *The case* $\bar{K} = 0$*, with* $c = \mu$ *or* $c > \mu$*, splits into three sub-cases:*
- (C1) [Like Case (b)] $c = \mu < 1$, $\bar{K} = 0$*:* $\lim_{k\to\infty} d_k/c_k = (1-c)/c$.
- (C2) [Like Case (c)] $c = \mu > 1$, $\bar{K} = 0$, $\sum_{k\in\mathbb{N}_0} \bar{K}_k = \infty$*:* [8] $\lim_{k\to\infty} d_k/\mu_k = 1/(\mu - 1)$.
- (C3) [Like Case (d)] $1 > c > \mu$ *and* $\bar{K} = 0$*, or* $1 = c > \mu$*,* $\bar{K} = 0$ *and* $a \in (-\infty, 1)$*:* $\lim_{k\to\infty} \sigma_k d_k = 1$.

□

The same observation as in Remark 3.16 applies. Again, the critical case $a = 1$ is missing in (C3).

3.2.6. *Cluster formation.* Within 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 the clusters where one type prevails grow.

This question has been addressed for other population models. For the voter model on $\mathbb{Z}^2$, Cox and Griffeath (1986) showed that the radii of the clusters with opinion "all 1" or "all 0" scale as $t^{\alpha/2}$ with $\alpha \in [0,1)$, i.e., clusters occur on all scales $\alpha \in [0,1)$. 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 $\eta \in \Omega_N$,

$$\lim_{t\to\infty} \mathcal{L}\left[\left(Y^{(\Omega_N)}_{\eta,\lfloor(1-\alpha)t\rfloor}(N^t)\right)_{\alpha\in[0,1)}\right] = \mathcal{L}\left[\left(Y\left(\log\left(\frac{1}{1-\alpha}\right)\right)\right)_{\alpha\in[0,1)}\right] \tag{3.40}$$

in the sense of finite-dimensional distributions, 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., branching models as shown in Dawson and Greven (1996).

The advantage of the hierarchical group is that we can analyze the cluster formation as a function of N and let $N \to \infty$ to approach the critically recurrent case (recall Remark 2.1). We can do this by using the *interaction chain*. In Greven et al. (2014), we analysed the Cannings model in the limit as $N \to \infty$, namely, we proved that for some *level scaling function* $k\colon \mathbb{N}_0 \to \mathbb{N}_0$, satisfying $0 \le k(j) \le j+1$ and $\lim_{j\to\infty} k(j) = \infty$, we obtained a *non-trivial clustering limiting law* (henceforth we pick $\eta = 0$ and drop it from the notation)

$$\lim_{j\to\infty} \mathcal{L}\left[M^{(j)}_{-k(j)}\right] = \mathcal{L}\big[M^\infty\big] \tag{3.41}$$

for some $M^\infty \in \mathcal{P}(E)$ satisfying $E[M^\infty] = E[X_0^{(\Omega_N)}(0)] = \theta \in \mathcal{P}(E)$ that is not of the form $M^\infty = \delta_U$ for some possibly random $U \in E$. We will do the same in the random environment ω, namely, our aim is to show that for $\mathbb{P}$-a.e. ω

$$\lim_{j\to\infty} \mathcal{L}\left[M^{(j)}_{-k(j)}(\omega)\right] = \mathcal{L}\big[M^\infty(\omega)\big] \tag{3.42}$$

[8] In Greven et al. (2014), the condition $\sum_{k\in\mathbb{N}_0} \bar{K}_k = \infty$ was mistakenly omitted.

for some $M^\infty(\omega) \in \mathcal{P}(E)$ satisfying $\mathbb{E}[M^\infty(\omega)] = \theta$ that is not of the form $M^\infty(\omega) = \delta_{U(\omega)}$ for some possibly random $U(\omega) \in E$.

As in Dawson and Greven (1993); Dawson et al. (1995); Dawson and Greven (1996), and similarly as in (3.40), in order to obtain the *profile of cluster formation* it is necessary to consider a whole *family of scalings* $k_\alpha \colon \mathbb{N}_0 \to \mathbb{N}_0$, $\alpha \in I$, with $I = \mathbb{N}_0$, $I = [0, \infty)$ or $I = [0, 1)$, and with $j \mapsto k_\alpha(j)$ non-decreasing, $0 \le k_\alpha(j) \le j + 1$ and $\lim_{j\to\infty} k_\alpha(j) = \infty$, such that

$$\lim_{j\to\infty} \mathcal{L}\left[M^j_{-k_\alpha(j)}(\omega)\right] = \mathcal{L}\left[M^*_\alpha(\omega)\right] \quad \text{for } \mathbb{P}\text{-a.e. } \omega \text{ and all } \alpha \in I, \tag{3.43}$$

for some non-constant Markov process $M^* = (M^*_\alpha(\omega))_{\alpha\in I}$ on $\mathcal{P}(E)$ that preserves the mean θ. The convergence in (3.43) is in the weak topology on the product space of $\mathcal{P}(E)$ and the space of the environment.

There are five *universality classes* of clustering behaviour (see Dawson and Greven, 1996):

Definition 3.18 (Clustering classes).

(I) **Concentrated clustering** (M^* is a Markov chain):
  (I1) $k_\alpha(j) = 0 \vee (j + 1 - \alpha)$, $\alpha \in \mathbb{N}_0$, M^* is trapped after one step.
  (I2) $k_\alpha(j) = 0 \vee (j + 1 - \alpha)$, $\alpha \in \mathbb{N}_0$, M^* is not trapped.

(II) **Diffusive clustering** (M^* is a diffusion process):
  (II1) *Fast clustering*: $k_\alpha(j) = 0 \vee \lfloor j+1-\alpha h(j) \rfloor$, $\alpha \in [0, \infty)$, where $h \colon \mathbb{N}_0 \to [0, \infty)$ is such that $\lim_{j\to\infty} h(j) = \infty$ and $\lim_{j\to\infty} h(j)/j = 0$.
  (II2) *Moderate clustering*: $k_\alpha(j) = \lfloor (1 - \alpha)(j + 1) \rfloor$, $\alpha \in [0, 1)$.
  (II3) *Slow clustering*: $\lim_{j\to\infty} k_\alpha(j)/j = 0$, $\alpha \in [0, 1)$.

(The terminology is slightly different from Greven et al., 2014.) The volume of the clusters at time t in these five universality classes (arranged in decreasing order of magnitude) equals, respectively, N^t, ZN^t, $N^{t-o(t)}$, N^{Zt}, $N^{o(t)}$, with $Z \in (0, 1)$ some random variable. Note that slow clustering borders with the regime of coexistence (= no clustering).

Recall (a)-(d) in Theorem 3.15 and (A)-(C) in Theorem 3.17. Recall that, under the law $\mathbb{P}$, the law of the initial state $X^{(\Omega_N)}(\omega; 0)$ is stationary and ergodic under translations in $\Omega_N^{\mathbb{T}}$, with mean single-coordinate measure

$$\theta = \mathbb{E}[X_0^{(\Omega_N)}(\omega; 0)] \in \mathcal{P}(E). \tag{3.44}$$

The interaction chain on level j, arising in the scaling limit $N \to \infty$, starts in θ. We need this for the scaling limit $j \to \infty$.

Theorem 3.19 (Cluster formation). *Fix $N \in \mathbb{N}\backslash\{1\}$. The five universality classes in the regime of clustering, linked to the different cases of scaling behaviour of $(d_k)_{k\in\mathbb{N}}$, are as follows:*

- **(a), (A):** *The scaling in regime* (I1) *yields* (3.43) *with $I = \mathbb{N}_0$. The scaling limit M^* is the time-homogeneous Markov chain on $\mathcal{P}(E)$ starting in θ with transition kernel $K(\theta, \cdot)$ given by*

$$K(\theta, \cdot) = \int_E \theta(\mathrm{d}u)\delta_{\delta_u}(\cdot), \tag{3.45}$$

  which satisfies $K_\alpha = K$, for all $\alpha \in \mathbb{N}$.

- **(b), (B), (C1), (C3)[first subcase]:** *The scaling in regime* (I2) *yields* (3.43) *with* $I = \mathbb{N}_0$. *The scaling limit* M^* *is the time-inhomogeneous Markov chain on* $\mathcal{P}(E)$ *in random environment* $(\chi_\alpha)_{\alpha\in\mathbb{N}_0}$ *starting in* θ *with transition kernels* $\{K_\alpha(\theta,\cdot)(\omega)\}_{\alpha\in\mathbb{N}_0}$ *given by (recall* (3.11)*)*

$$K_\alpha(\theta,\cdot)(\omega) = \nu_\theta^{1,\widetilde{M},2\widetilde{K}\chi_\alpha(\omega)}(\cdot) \tag{3.46}$$

with

$$(\widetilde{M},\widetilde{K}) = \begin{cases} (M,K), & \text{(b)},\\ (\bar{M}/c,\bar{K}), & \text{(B)},\\ ((1-c)/c,0), & \text{(C1), (C3)[first subcase]}. \end{cases} \tag{3.47}$$

In the last two cases, the random environment does not affect the scaling limit, and the scaling is the same as for the homogeneous environment with the same mean. In the first two cases, the measure-valued process $(\chi_\alpha(\omega))_{\alpha\in\mathbb{N}_0}$ *in* (3.46) *is constructed by extending the one-sided stationary random environment* $(\chi^{(\eta,k)}(\omega))_{\eta\in\Omega_N,\,k\in\mathbb{N}_0}$ *introduced in* (2.36) *to a two-sided stationary random environment* $(\chi^{(\eta,k)}(\omega))_{\eta\in\Omega_N,\,k\in\mathbb{Z}}$, *and defining*

$$\mathcal{L}[(\chi_\alpha(\omega))_{\alpha\in\mathbb{N}_0}] = \mathcal{L}[(\chi^{(\eta,-\alpha)}(\omega))_{\alpha\in\mathbb{N}_0}], \quad \eta\in\Omega_N, \tag{3.48}$$

which by stationarity does not depend on η. *Furthermore,* $\chi_\alpha(\omega)$ *is an* $\mathcal{M}_f([0,1])$*-valued resampling measure with* $\mathbb{E}[\chi_\alpha(\omega)] = \bar{\chi}_\alpha$ *satisfying*

$$\bar{\chi}_\alpha((0,1]) = 1. \tag{3.49}$$

- **(c), (C2)[subcase** $\lim_{k\to\infty} k\bar{K}_k = \infty$**]:** *The scaling in regime* (II1) *yields* (3.43) *with* $I = [0,\infty)$. *The scaling limit* M^* *is the time-changed standard Fleming-Viot process*

$$M^*_\alpha = Z_\theta^{0,1,0}(\ell(\alpha)), \quad \alpha\in[0,\infty) \tag{3.50}$$

with
  - **(c):** $\ell(\alpha)=\alpha$, $h(j) = 1/\sqrt{K_j}$.
  - **(C2)**$[\lim_{k\to\infty} k\bar{K}_k = \infty]$**:** $\ell(\alpha) = \frac{\mu}{\mu-1}\alpha$, $h(j) = 1/K_j$.
- **(d), (C2)[subcase** $\lim_{k\to\infty} k\bar{K}_k = \bar{N}$**], (C3)[second subcase]:** *The scaling in regime* (II2) *yields* (3.43) *with* $I=[0,1)$. *The scaling limit* M^* *is the time-changed standard Fleming-Viot process* [9]

$$M^*_\alpha = Z_\theta^{0,1,0}\left(\log\left(\frac{1}{(1-\alpha)^R}\right)\right), \quad \alpha\in[0,1) \tag{3.51}$$

with
  - **(d):** $R = M(1-a)$.
  - **(C2)[subcase** $\lim_{k\to\infty} k\bar{K}_k = \bar{N}$**]:** $R = \bar{N}\frac{\mu}{\mu-1}$.
  - **(C3)[second subcase]:** $R = 1-a$.

For reasons explained in Remark 8.2, *in cases* **(c), (d), (C2), (C3)[second subcase]** *only convergence in* $\mathbb{P}$*-probability and not* $\mathbb{P}$*-a.s. is obtained.* □

Remark 3.20. We expect that also *slow* clustering occurs, namely, in the case (e) that was not treated in Theorems 3.15 and 3.17 (recall Remark 3.16). □

[9] Greven et al. (2014) contains a typo: there the time scaling $1/(1-\alpha)^R$ was wrongly written as $1/(1-\alpha^R)$.

3.3. *Summary of the effects of the random environment.*
1. Theorem 3.1 says that the hierarchical Cannings process in random environment is well-defined for $\mathbb{P}$-a.e. ω, while Theorem 3.2 shows that it converges to an ω-dependent equilibrium that preserves the single-component mean.
2. Theorem 3.6 (mono-scale) and Theorem 3.7 (multi-scale) identify the behaviour of the k-block averages in the limit as $N \to \infty$ in terms of the McKean-Vlasov process with parameters that depend on ω and k. The volatility d_k depends on the parameters c_l, λ_l, $0 \leq l < k$, and on the law $\mathcal{L}_\rho$ via the recursion relation in (3.17), which is a *randomized* version of the recursion relation in Greven et al. (2014).
3. Theorems 3.3 and 3.13 show that the dichotomy "coexistence versus clustering" is *not* affected by the random environment: the *same* conditions apply to the homogeneous hierarchical Cannings process studied in Greven et al. (2014). Apparently, for the nature of the equilibrium only the large-scale properties of the random environment matter. Since the resampling measures are stationary under translations with total masses whose sigma-algebra at infinity is trivial, only the average medium behaviour is relevant. The proof of the dichotomy in Theorem 3.3 requires assumption (2.43) rather than assumption (2.41). We believe this strengthening to be redundant, but a proof would require considerable extra work.
4. Theorem 3.10 shows that *the effect of the random environment is to lower the volatility parameter d_k on every hierarchical scale k* compared to the average environment. The intuition behind this is that the random environment causes fluctuations in the resampling, which in turn reduce the clustering. The sandwich between the volatilities for the zero environment and the average environment is useful to control the scaling.
5. Theorem 3.15 (polynomial coefficients) and Theorem 3.17 (exponential coefficients) show that for Cases (b) and (B), where *migration and resampling occur at comparable rates*, the phenomenon of lower volatility d_k in random environment persists in the limit as $k \to \infty$: even though the scaling of d_k as $k \to \infty$ is the same as for the average environment, it has a *different prefactor* (e.g. M solving (3.35) is strictly smaller than M^* solving (3.35) with $\mathcal{L}_\rho$ replaced by δ_1, as is easily shown by applying Jensen's inequality). For all other cases both the scaling and the prefactor are the same as for the average environment.
6. Theorem 3.19 shows that for Cases (b), (B), (C1), (C3)[first subcase] the scaling of the clusters in the random environment changes compared to that in the average environment: the random environment is visible even in the scaling limit. *The effect of the random environment is to slow down the growth of the clusters*, i.e., to enhance the diversity of types. For all other cases the scaling of the clusters is the same as for the average environment.

4. Existence, uniqueness, duality and equilibrium

In this section, we prove Theorems 3.1–3.2. In Section 4.1, we construct the dual process with the help of a graphical representation based on Poisson random measures. In Section 4.2, we exhibit the duality. In Section 4.3, we establish the existence and uniqueness of the dual process and show the existence of its equilibrium. In Section 4.4, we use these results to prove Theorems 3.1–3.2. Theorems 4.1–4.4 below do not need a separate proof: this is verbatim the same as the proof for the homogeneous environment given in Greven et al. (2014).

4.1. *The spatial coalescent in random environment.* In this section, we introduce a hierarchical coalescent process in random environment that will serve as a dual to the hierarchical Cannings process in random environment.

The coalescent is a Markov process taking values in the set of partitions of $\mathbb{N}$ labelled by the points of a geographical space. We recall the basic objects and notations, and refer to Greven et al. (2014, Section 2) for details.

Let G be a discrete geographical space. Our target geographical space is $G = \Omega_N$. This will be approximated by a sequence of geographical spaces

$$G_{N,K} = \{0, \dots, N-1\}^K, \quad K \in \mathbb{N}, \tag{4.1}$$

which are to be thought of as a sequence of blocks filling Ω_N. We will also need to consider the mean-field geographical space

$$G = \{0, *\}, \tag{4.2}$$

where $\{*\}$ is a cemetery location. The state space of the spatial coalescent is the set of *G-labelled partitions* defined as

$$\Pi_{G,n} = \Big\{\pi_G = \{(\pi_1, g_1), (\pi_2, g_2), \dots, (\pi_b, g_b)\}: \tag{4.3}$$

$$\{\pi_1, \dots, \pi_b\} \in \Pi_n,\ g_1, \dots, g_b \in G, b \in [n]\Big\}, \tag{4.4}$$

where $n \in \mathbb{N}$ and

$$\Pi_n = \text{ set of partitions } \pi = \{\pi_i \subset [n]\}_{i=1}^b \text{ of } [n] \text{ into disjoint families } \pi_i,\ i \in [b]. \tag{4.5}$$

We equip the set $\Pi_{G,n}$ with the *discrete topology*. Let a be a continuous-time random walk transition (rate) kernel on G. When $G = \Omega_N$ we use the hierarchical random walk with kernel $a = a^{(N)}$ in (2.6), when $G = G_{N,K}$ we use the same hierarchical random walk kernel but with $c_k = 0$ for $k > K$, and when $G = \{0, *\}$ we use the random walk with kernel $a(0, *) = c$, $a(*, 0) = 0$.

Given the random environment ω (recall Section 2.2.1), the spatial coalescent in random environment is the Markov process on state space $\Pi_{G,n}$ with the following dynamics:

- **[Migration]** Each partition block performs an independent random walk on G with random walk kernel a^*, where $a^*(g_1, g_2) = a(g_2, g_1)$, $g_1, g_2 \in G$, is the conjugate random walk kernel.
- **[Local coalescence]** Independently at each location $g \in G$, the l-tuples of the partition elements at g coalesce into a single partition element at g at rate

$$\lambda_{b,l}(\omega) = \int_{(0,1]} r^l (1-r)^{b-l} \frac{\Lambda^{[g]}(\omega)(\mathrm{d}r)}{r^2}, \tag{4.6}$$

  where b is the current total number of partition elements and $\Lambda^{[g]}(\omega)$ is the resampling measure at g in environment ω.
- **[Non-local coalescence with reshuffling]** In the case $G = \Omega_N$, independently at each location $g \in B_{|\xi|}$, $\xi \in \Omega_N^{\mathbb{T}}$, the l out of b tuples of the partition elements in $B_{|\xi|}$ coalesce into a single partition element at g at rate

$$N^{-2k} \lambda_{b,l}^{(\xi)}(\omega), \qquad \lambda_{b,l}^{(\xi)}(\omega) = \int_{(0,1]} r^l (1-r)^{b-l} \frac{\Lambda^{\xi}(\omega)(\mathrm{d}r)}{r^2}. \tag{4.7}$$

Subsequently, all the partition elements located in $B_{|\xi|}$ are uniformly *reshuffled*, i.e., all the partition elements in $B_{|\xi|}$ get a new location that is drawn uniformly from $B_{|\xi|}$. In what follows, denote by

$$b(\xi) = b(\pi_{\Omega_N,n}, B_{|\xi|}) = |\{(\pi_i, \eta_i) \in \pi_{\Omega_N,n} \colon \eta_i \in B_{|\xi|}\}| \in \mathbb{N} \tag{4.8}$$

the total number of partition blocks with labels in $B_{|\xi|}$.

Note that in the case $G = \Omega_N$ the partition elements of the coalescent perform a hierarchical random walk on Ω_N in the environment ω with migration coefficients given by (recall (2.6))

$$c_k(\omega)(N,\eta) = c_k + N^{-1}\,\lambda^{\mathrm{MC}_{k+1}(\eta)}(\omega), \qquad \eta \in \Omega_N, k \in \mathbb{N}_0, \tag{4.9}$$

where $\mathrm{MC}_k(\eta)$ is the unique site at height k above $\eta \in \Omega_N$ and $\lambda^{\mathrm{MC}_{k+1}(\eta)}(\omega) = \Lambda^{\mathrm{MC}_{k+1}(\eta)}(\omega)\,((0,1])$ (recall the notation introduced in Section 2.2 and see Fig. 2.3). The extra term in the right-hand side of (4.9) comes from the reshuffling that takes place prior to the resampling.

The *coalescence* rate of two partition elements in $B_{|\xi|}$ equals (recall (2.39))

$$N^{-2|\xi|}\lambda^{\xi}(\omega), \qquad \lambda^{\xi}(\omega) = \lambda_{|\xi|}\rho^{\xi}(\omega), \qquad \xi \in \Omega_N^{\mathbb{T}}. \tag{4.10}$$

We specify the spatial coalescent as a Markov process on $\Pi_G = \cup_{n\in\mathbb{N}}\Pi_{G,n}$ by providing its generator. To that end, we need a space of test functions on Π_G. Namely, 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 $n = n(F) \in \mathbb{N}$ and a bounded function

$$F_n\colon\ \Pi_{G,n} \to \mathbb{R} \tag{4.11}$$

with the property that $F(\cdot) = F_n(\cdot|_n)$. Consider the linear operator $L^{(G)*}(\omega)\colon \mathcal{C}_G \to \mathcal{C}_G$ defined as

$$L^{(G)*}(\omega) = L^{(G)*}_{\mathrm{mig}} + L^{(G)*}_{\mathrm{coal}}(\omega), \tag{4.12}$$

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

$$(L^{(G)*}_{\mathrm{mig}}F)(\pi_G) = \sum_{i=1}^{b(\pi_G|_n)} \sum_{g,f\in\Omega_N} a^*(g,f)\big[F_n\big(\mathrm{mig}_{g\to f,i}(\pi_{G,n})\big) - F(\pi_G)\big] \tag{4.13}$$

and

$$\begin{aligned}(L^{(G)*}_{\mathrm{coal}}(\omega)F)(\pi_G) &= \sum_{k\in\mathbb{N}_0}\sum_{\xi\in\Omega_N^{(k)}} N^{-2k} \\ &\times \sum_{\substack{J\subset[b(\xi)],\\ |J|\geq 2}} \Big(\lambda^{(\xi)}_{b(\xi),|J|}(\omega)\,\big[F_n\left(\mathrm{resh}_{B_{|\xi|}}\circ\mathrm{coal}_{J,\xi}(\pi_{\Omega_N,n})\right) - F(\pi_{\Omega_N})\big]\Big)\end{aligned} \tag{4.14}$$

Here, the *migration map* $\mathrm{mig}_{g\to f,i}(\pi_G|_n)$ changes the spatial coordinate of the i-th partition block from g to f (if such a partition element exists), the *coalescence map* $\mathrm{coal}_{J,g}(\pi_{G,n})$ coalesces the partition blocks with indices in J and location g (if any) into one block, while the *reshuffling map* $\mathrm{resh}_{B_{|\xi|}}$ independently relocates each partition element located in $B_{|\xi|}$ to a new location in $B_{|\xi|}$ that is randomly chosen.

Theorem 4.1 (Existence and uniqueness). *For every $\pi \in \Pi_G$, the $(L^{(G)*}(\omega), \mathcal{C}_G, \delta_\pi)$-martingale problem is well-posed.* □

We denote the solution of the $(L^{(G)*}(\omega), \mathcal{C}^{(G)}, \delta_\pi)$-martingale problem by

$$\mathfrak{C}^{(G)}(\omega) = \big(\mathfrak{C}^{(G)}(\omega;t)\big)_{t\geq 0}, \qquad \mathfrak{C}^{(G)}(\omega;0) = \delta_\pi. \tag{4.15}$$

For every $n \in \mathbb{N}$, when restricted to $\Pi_{G,n}$, $\mathfrak{C}^{(G)}(\omega)$ becomes a strong Markov process $\mathfrak{C}_n^{(G)}(\omega)$ with the Feller property.

4.2. *Dualities.* By a slight abuse of notation, we 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.

Consider the map

$$H_\varphi^{(n)}(x, \pi_{G,n}) = \int_{E^b} \left(\bigotimes_{i=1}^{b} x_{g_{\pi^{-1}(i)}}(\mathrm{d}u_i) \right) \varphi\big(u_{\pi(1)}, \dots, u_{\pi(n)}\big), \tag{4.16}$$

where $n \in \mathbb{N}$, $\phi \in C_{\mathrm{b}}(\mathcal{P}(E)^n)$, $x = (x_\eta)_{\eta\in G} \in \mathcal{P}(E)^G$,

$$\pi_{G,n} = \{(\pi_1, g_1), \dots, (\pi_b, g_b)\} \in \Pi_{G,n}, \quad \{\pi_1, \dots, \pi_b\} \in \Pi_n, \quad g_1, \dots, g_b \in G, \tag{4.17}$$

$b = b(\pi_{G,n}) = |\pi_{G,n}|$ and

$$\pi^{-1}(k) = \min\{i \in [n]\colon\ \pi(i) = k\}, \quad k \in [b]. \tag{4.18}$$

Theorem 4.2 (Duality). *Fix $N \in \mathbb{N}\backslash\{1\}$. For each of the choices G in (4.1) and (4.2),*

$$E[H_\varphi^{(n)}(X^{(G)}(\omega,t), \mathfrak{C}^{(G)}(\omega,0)|_n)] = E[H_\varphi^{(n)}(X^{(G)}(\omega,0), \mathfrak{C}^{(G)}(\omega,t)|_n)] \tag{4.19}$$

for all $n \in \mathbb{N}$ and $\phi \in C_{\mathrm{b}}(\mathcal{P}(E)^n)$, where the same ω is used on both sides. □

This theorem is a consequence of the generator relation

$$\Big(L^{(G)}(\omega) H_\varphi^{(n)}(\cdot, \pi_{G,n})\Big)(x) = \Big(L^{(G)*}(\omega) H_\varphi^{(n)}(x, \cdot)\Big)(\pi_G) \quad \text{for } \mathbb{P}\text{-a.e. } \omega. \tag{4.20}$$

This relation has been verified for the homogeneous model in Greven et al. (2014), but here works the same.

4.3. *Well-posedness of the martingale problems and equilibria.* Theorem 3.1 can be formulated for geographic spaces that are countable Abelian groups, in particular, the hierarchical group and the Euclidean lattice. For us the following generalization of Theorem 4.1 suffices.

Theorem 4.3 (Well-posedness). *For each of the choices G in (4.1) and (4.2), the following holds: For $\mathbb{P}$-a.e. ω and every $\pi \in \Pi_G$, the $(L^{(G)}(\omega), \mathcal{C}^{(G)}, \delta_\pi)$-martingale problem is well-posed.* □

Theorem 4.4 (Equilibrium). *Fix $N \in \mathbb{N}\backslash\{1\}$. Fix $n \in \mathbb{N}$ and start the $\mathfrak{C}^{(\Omega_N)}(\omega)$-process in a labelled partition $\{(\pi_i, \eta_i)\}_{i=1}^n$, where $\{\pi_i\}_{i=1}^n$ form a partition of $\mathbb{N}$ and $\{\eta_i\}_{i=1}^n$ represent the labels. If x is a random state with mean $\theta \in \mathcal{P}(E)$ whose law is invariant and ergodic under translations, then*

$$\lim_{t\to\infty} \mathcal{L}\left[H_\varphi^{(n)}\left(x, \mathfrak{C}_n^{(\Omega_N)}(\omega;t)\right)\right] = \mathcal{L}\left[H_\varphi^{(n)}\left(\underline{\theta}, \mathfrak{C}_n^{(\Omega_N)}(\infty)\right)\right] \quad \text{for } \mathbb{P}\text{-a.e. } \omega \tag{4.21}$$

for all $\phi \in C_{\mathrm{b}}(\mathcal{P}(E)^n)$. □

In order to prove Theorem 4.4, we follow the argument in Dawson et al. (1995, Section 3)). The partition-valued process converges to a limiting partition. If the locations of the partition elements would follow a homogeneous random walk, then the key to the argument would be the averaging property one can prove via Fourier analysis

$$\lim_{t\to\infty} \sum_{\zeta\in\Omega_N} p_t(\eta,\zeta) f(\zeta) = \int_{\Omega_N} f(\xi)\nu(\mathrm{d}\xi) \qquad \forall\, \eta\in\Omega_N,\ f\in C_b(\Omega_N), \tag{4.22}$$

where $p_t(\cdot,\cdot)$ is the time-t transition kernel of the random walk on Ω_N and $\nu(\cdot)$ is the Haar measure on Ω_N (see Evans and Fleischmann, 1996). We need to show that the same holds for our random walk in random environment ω, which goes as follows.

Place a Poisson clock at every $\xi\in\Omega_N^{\mathbb{T}}\backslash\Omega_N$. Let the clock at ξ ring at rate

$$N^{-(|\xi|-1)}\big[c_{|\xi|-1} + N^{-1}\lambda^{\xi}(\omega)\big]. \tag{4.23}$$

At any moment of time, let the random walk look at the ancestral line above its current position (see Fig. 2.3) and redistribute itself uniformly over the block around its current position whose hierarchical label corresponds to the height of the first clock on that ancestral line that rings. The resulting random walk is the same as the hierarchical random walk in environment ω with migration coefficients given by (4.9).

Next, let $K_t(\eta)$ be the highest hierarchical level at which prior to time t a Poisson clock that lies on the ancestral line above η has rang. Then at time t the random walk starting from η is uniformly distributed on the $K_t(\eta)$-block around η. Hence we have

$$\sum_{\zeta\in\Omega_N} p_t^{\omega}(\eta,\zeta) f(\zeta) = \sum_{k\in\mathbb{N}_0} P(K_t(\eta)=k) \times \left[N^{-k}\sum_{\zeta\in B_k(\eta)} f(\zeta) + \sum_{\zeta\in\Omega_N\backslash B_k(\eta)} p_t^{\omega}(\eta,\zeta) f(\zeta)\right], \tag{4.24}$$

where $p_t^{\omega}(\cdot,\cdot)$ is the time-t transition kernel of the random walk in ω. Fix $\eta\in\Omega_N$ and $f\in C_b(\Omega_N)$. The first term between the square brackets in (4.24) tends to $\int_{\Omega_N} f(\xi)\nu(\mathrm{d}\xi)$ as $k\to\infty$. The second term is bounded from above by $\|f\|_\infty\, p_t^{\omega}(\eta,\Omega_N\backslash B_k(\eta))$, which tends to zero as $k\to\infty$. Finally, since all Poisson clocks ring at a strictly positive rate, we have

$$P\left(\lim_{t\to\infty} K_t(\eta)=\infty\right)=1. \tag{4.25}$$

It therefore follows that the right-hand side of (4.24) tends to the right-hand side of (4.22) as $t\to\infty$.

4.4. *Consequences for the Cannings process.* The claims in Theorems 3.1–3.2 follow from Theorems 4.3–4.4. As argued in Greven et al. (2014), the proof follows the strategy for the two-type case given in Evans (1997, Theorem 4.1), which says that for spatial coalescent processes well-posedness and existence carry over from the dual process to the original process.

We next prove Corollary 3.14.

Proof: We analyze both ν_θ^N and ν_θ with the help of duality relations and show that the dual representation of the former converges to the dual representation of the latter.

Step 1: ν_θ. We have to construct a dual process for the entrance law of a Markov chain, namely, the interaction chain running from level ∞ down to level 0. We consider the process that is dual to the interaction chain at level j. This dual process is a discrete-time Markov chain whose transition kernel we can determine, for fixed j and in the limit as $j \to \infty$, via an explicit construction. This dual Markov chain is a spatial coalescent on $\{0, 1, \dots, j\}$, or on $\mathbb{N}$ when we consider all j simultaneously and are interested in its limit state as $j \to \infty$.

We first focus on the dual transition kernel at one particular level. In the interaction chain this is defined via the equilibrium of the McKean-Vlasov process. How did this equilibrium arise? We consider a *mean-field* system of size N^k with parameters c_k, d_k, Λ_k and take the mean-field dual started in n individuals at mutual distance k. This dual is shown to *converge*, in the limit as $N \to \infty$ and on time scale $t_N N^k$ with $t_N \to \infty$ and $t_N = o(N)$, to a limiting process that is a coalescent on the geographic space $k \cup \{\triangle\}$, with $\triangle$ a cemetery state, such that the process jumps from k to $\triangle$ at rate c_k and does Kingman coalescence at rate d_k and Λ-coalescence according to $\Lambda_k(\mathrm{MC}_k(0))(\omega)$. This limiting process is run for infinite time to obtain the dual transition kernel at the k-th step. This partition at $\triangle$ is used as input for the next step of the dual with label $k+1$. Altogether this procedure defines the full Markov chain, i.e., the new site and the new partition element. We denote the path of the dual Markov chain by

$$(\Pi_k^\infty)_{k\in\mathbb{N}_0}, \text{ with } \Pi_\infty^\infty \text{ its limiting state as } k \to \infty. \tag{4.26}$$

(Recall that partitions are ordered and hence the limiting state exists.) The line of argument is the same for each k. A detailed argument can be found in Greven et al. (2014, Corollary 2.12).

We need an explicit description as an $\mathbb{N}$-marked partition-valued process, namely, the above mentioned random walk, moving one step to the right on $\mathbb{N}$, doing Kingman coalescence at rate d_k and Λ-coalescence according to $\Lambda_k(\mathrm{MC}_k(\eta))$ in state k, provided the rate-c_k clock does not ring first.

The dual chain after j steps gives the expression $E_\theta[\langle M_0^{(j)}, f\rangle^n](\omega)$, which in the limit as $j \to \infty$ equals $\int_0^1 \Pi_\eta \nu_\theta(\omega)(\mathrm{d}x)\, \langle x_\eta, f\rangle^n$ by the definition of $\nu_\theta(\omega)$. The dual expectation is the expression $E[\langle \theta, f\rangle^{|\Pi_\infty^\infty|}](\omega)$. It therefore suffices to show that the latter is obtained from the dual representation of $\nu_\theta^N(\omega)$ as $N \to \infty$.

Remark 4.5. What is the dual counterpart of Theorem 3.6? The connection between the renormalized system and the interaction chain on the level of the dual is as follows. Consider the dual process for the j-level *hierarchical* system for finite N, starting with n partition elements at one site and letting $t \to \infty$ and $N \to \infty$ in the following way. Consider time scales $(t_N^k)_{k\in\mathbb{N}_0}$ with $t_N^k/N^{k+1} \to 0$ and $t_N^k/N^k \to \infty$ as $N \to \infty$. Then the coalescent reaches a partition Π_∞^{k+1}, with the remaining partition elements in B_k uniformly distributed. After that move to the next time scale. Finally, first take $N \to \infty$ and then take

$$\Pi_\infty^\infty \text{ as the limiting partition element for } j \to \infty. \tag{4.27}$$

By our scaling result in (3.25), this object gives us the dual process of the interaction chain at level j. □

Step 2: ν_θ^N. We return to the representation of ν_θ^N, respectively, its marginal law at level 0. The convergence of the dual chain for the Cannings process on Ω_N, and its limit as $t \to \infty$ followed by $N \to \infty$ to the dual chain of the interaction chain, will follow from the fact that the partitions become successively finer and hence converge to a limit partition, and the fact that the time scales for the random walk to reach distance k separate as $N \to \infty$. Since the monomials are convergence determining, this will yield the claim.

We have to show that the coalescent on Ω_N, starting with n individuals at site 0, converges to a limit process as $t \to \infty$, which we can investigate in the limit as $N \to \infty$. We need to show that this process has the property that, when we consider the times where the coalescent makes jumps to the next larger block, we get an embedded Markov chain with index in $\mathbb{N}_0$, describing the successive maximal jump sizes and values in partitions. The corresponding partition converges to Π_∞^N as the index k tends to infinity. The claim is that, as $N \to \infty$, this Markov chain converges to a birth process in the first component, which moves one step to the right, with Kingman-coalescence at rate d_k and Λ-coalescence according to $\Lambda_k(\mathrm{MC}_k(0))(\omega)$, provided the rate-$c_k$ clock does not ring first. This gives Π_k^∞. As $k \to \infty$ we get (4.27). This is done in Dawson and Greven (1996) for the Kingman coalescent, but the necessary modification is straightforward. □

5. Dichotomy: coexistence versus clustering

In this section, we prove Theorem 3.3. The question is whether $\mathfrak{C}^{(G)}(\omega)$, the hierarchical coalescent in the environment ω defined in (4.15), 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 $\mathfrak{C}_n^{(G)}(\omega)$ for any $n \in \mathbb{N}$, and hence for the entrance law starting from n partition elements. However, there is additional reshuffling at all higher levels, which is triggered by a corresponding block-coalescence event. Therefore, we need to consider two coalescing random walks with slightly *adapted* migration coefficients, lacking in particular the random walk property.

Recall the notation introduced in Sections 2.1.1–2.1.3 and 2.2.1. Recall that $\mathrm{MC}_k(\eta)$ is the unique site at height k above $\eta \in \Omega_N$ (see Fig. 2.3). Consider two independent copies

$$Y(\omega) = (Y_t(\omega))_{t\geq 0}, \qquad Y'(\omega) = (Y'_t(\omega))_{t\geq 0}, \tag{5.1}$$

of the hierarchical random walk on Ω_N in the environment ω with migration coefficients given by (4.9) and coalescence rates given by (4.10). Write $P^\omega, P^{\omega,\prime}$ for the marginal laws of $Y(\omega), Y'(\omega)$ and $\bar{P}^\omega = P^\omega \times P^{\omega,\prime}$ for the joint law of the pair $\bar{Y}(\omega) = (Y(\omega), Y'(\omega))$. Consider the time-$t$ accumulated hazard for coalescence:

$$H_N(\omega; t) = \sum_{k\in\mathbb{N}_0} N^{-k} \sum_{\substack{\eta,\eta'\in\Omega_N \\ d_{\Omega_N}(\eta,\eta')\leq k}} \lambda^{\mathrm{MC}_k(\eta)}(\omega) \int_0^t 1_{\{Y_s(\omega)=\eta, Y'_s(\omega)=\eta'\}}\, \mathrm{d}s, \tag{5.2}$$

where we use that $\mathrm{MC}_k(\eta) = \mathrm{MC}_k(\eta')$ when $d_{\Omega_N}(\eta,\eta') \leq k$. The rate N^{-2k} to choose a k-block for the coalescence is multiplied by N^k because all partition

elements in that block can trigger a coalescence event, which explains the factor N^{-k} in (5.2).

Let $\lim_{t\to\infty} H_N(\omega;t) = H_N(\omega;\infty)$. We have coalescence of the two random walks ("common ancestor") with probability 1 when $H_N(\omega;\infty) = \infty$ $\bar{P}^\omega$-a.s., but separation of the two random walks ("different ancestors") with positive probability when $H_N(\omega;\infty) < \infty$ $\bar{P}^\omega$-a.s. In Section 5.1 we identify the dichotomy for the mean hazard $\bar{E}^\omega[H_N(\omega;\infty)]$ combining Fourier analysis with potential theory of reversible Markov chains to handle the fact that our migration is no longer a random walk. In Section 5.2 we use a zero-one law to show that the same dichotomy holds for the hazard $H_N(\omega;\infty)$.

5.1. *Mean hazard.*

Lemma 5.1. *For every $N \in \mathbb{N}\backslash\{1\}$ and $\mathbb{P}$-a.e. ω,*

$$\bar{E}^\omega[H_N(\omega;\infty)] = \infty \quad \Longleftrightarrow \quad \sum_{k\in\mathbb{N}_0} \frac{1}{c_k + N^{-1}\lambda_{k+1}} \sum_{l=0}^{k} \lambda_l = \infty. \tag{5.3}$$

□

Proof: Write

$$p_t^\omega(\eta,\zeta) = P^\omega\{Y_t(\omega) = \zeta \mid Y_0(\omega) = \eta\}, \quad \eta,\zeta \in \Omega_N, \tag{5.4}$$

to denote the time-t transition kernel. Let

$$G^\omega((0,0),(\eta,\eta')) = \int_0^\infty p_t^\omega(0,\eta) p_t^\omega(0,\eta')\,\mathrm{d}t, \qquad (\eta,\eta') \in \Omega_N \times \Omega_N, \tag{5.5}$$

denote the Green function for $\bar{Y}(\omega)$. Then (5.2) gives

$$\begin{aligned}
\bar{E}^\omega[H_N(\omega;\infty)] &= \sum_{k\in\mathbb{N}_0} N^{-k} \sum_{\substack{\eta,\eta'\in\Omega_N \\ d_{\Omega_N}(\eta,\eta')\le k}} \lambda^{\mathrm{MC}_k(\eta)}(\omega)\, G^\omega((0,0),(\eta,\eta')) \\
&= \sum_{k\in\mathbb{N}_0} N^{-k} \sum_{\xi\in\Omega_N^{(k)}} \lambda^\xi(\omega) \sum_{\substack{\eta\in\Omega_N \\ \mathrm{MC}_k(\eta)=\xi}} \sum_{\eta'\in B_k(\eta)} G^\omega((0,0),(\eta,\eta')) \\
&= \sum_{k\in\mathbb{N}_0} N^{-k} \sum_{\xi\in\Omega_N^{(k)}} \lambda^\xi(\omega) \sum_{\eta,\eta'\in B_{|\xi|}} G^\omega((0,0),(\eta,\eta'))
\end{aligned} \tag{5.6}$$

(recall (2.32)). The proof comes in two steps. In Step 1, we pretend that the ω-dependent term in the right-hand side of (4.9) is replaced by its mean, i.e., the two hierarchical random walks are homogenous with migration coefficients $\bar{c}_k$ given by

$$\bar{c}_k(N) = \mathbb{E}[c_k(\omega)(N,\eta)] = c_k + N^{-1}\,\lambda_{k+1}, \tag{5.7}$$

and show that the same dichotomy as in (5.3) holds. In Step 2, we explain why this replacement does not affect the dichotomy. The Green function of the two homogeneous hierarchical random walks will be denoted by $G((0,0),(\eta,\eta'))$.

Step 1. In what follows, we use the explicit form of the transition kernel $p_t(\eta,\zeta)$, $\eta,\zeta \in \Omega_N$, for the *homogeneous* hierarchical random walk computed in Dawson et al. (2005) with the help of Fourier analysis. Namely,

$$p_t(0,\eta) = \sum_{j\geq k} K_{jk}(N)\,\frac{\exp[-h_j(N)t]}{N^j}, \qquad t\geq 0,\ \eta\in\Omega_N\colon\ d_{\Omega_N}(0,\eta)=k\in\mathbb{N}_0, \tag{5.8}$$

where

$$K_{jk}(N) = \begin{cases} 0, & j=k=0,\\ -1, & j=k>0,\\ N-1, & \text{otherwise}, \end{cases} \qquad j,k\in\mathbb{N}_0, \tag{5.9}$$

and

$$h_j(N) = \frac{N-1}{N}\,r_j(N) + \sum_{i>j} r_i(N), \qquad j\in\mathbb{N}, \tag{5.10}$$

with

$$r_j(N) = \frac{1}{D(N)}\,\frac{N-1}{N}\sum_{i\geq j}\frac{\bar c_{i-1}(N)}{N^{2i-j-1}}, \qquad j\in\mathbb{N}, \tag{5.11}$$

where $D(N)$ is the normalizing constant such that $\sum_{j\in\mathbb{N}} r_j(N)=1$. Note that the expressions in (5.10)–(5.11) simplify considerably in the limit as $N\to\infty$, namely, the term with $i=j$ dominates and

$$h_j(N)\sim r_j(N)\sim \frac{\bar c_{j-1}(N)}{D(N)N^{j-1}}, \quad j\in\mathbb{N}, \qquad D(N)\sim \bar c_0(N). \tag{5.12}$$

Also note that, because of (2.9) and (2.15), the following holds:

> For $N\in\mathbb{N}\backslash\{1\}$, the quantities $h_j(N), r_j(N), D(N)$ are bounded from above and below by positive finite constants times the right-hand side of (5.12) uniformly in the index j.

(5.13)

To compute the sum in (5.6), we need to distinguish two cases: (1) $\xi=0^k\in\Omega_N^{(k)}$, the unique site in $\Omega_N^{\mathbb{T}}$ at height k above $0\in\Omega_N$; (2) $\xi\in\Omega_N^{(k)}\backslash\{0^k\}$.

(1). $\underline{\xi=0^k}$. Write

$$\sum_{\eta,\eta'\in B_{|\xi|}} G((0,0),(\eta,\eta')) = \sum_{0\leq p,q\leq k} N[p]N[q]\,G((0,0),(\eta^{(p)},\eta^{(q)})), \tag{5.14}$$

where $\eta^{(p)}$ is any site in Ω_N such that $d_{\Omega_N}(0,\eta^{(p)})=p$, and

$$N[p] = |B_p(0)\backslash B_{p-1}(0)| = \begin{cases} N^p - N^{p-1}, & p>0,\\ 1, & p=0. \end{cases} \tag{5.15}$$

With the help of (5.8) we obtain

$$G((0,0),(\eta^{(p)},\eta^{(q)})) = \sum_{m\geq p}\sum_{n\geq q} K_{mp}(N)K_{nq}(N)\,N^{-m-n}\,\frac{1}{h_m(N)+h_n(N)}. \tag{5.16}$$

Inserting (5.9) and (5.12), we get

$$\text{r.h.s. (5.16)} \sim \frac{\bar c_0(N)}{(1+1_{\{p=q\}})\bar c_{p\wedge q}(N)N^{p\vee q}}, \qquad N\to\infty, \tag{5.17}$$

where the asymptotics comes from the terms with $m = p+1$ and $n = q+1$. Combining (5.14–5.17), we obtain

$$\begin{aligned} &N^{-k} \sum_{\eta,\eta' \in B_k(0^k)} G((0,0),(\eta,\eta')) \\ &\quad \sim N^{-k} \left(\sum_{0 \le p \le k} N^{2p} \frac{\bar{c}_0(N)}{2\bar{c}_p(N) N^p} + 2 \sum_{0 \le p < q \le k} N^{p+q} \frac{\bar{c}_0(N)}{\bar{c}_p(N) N^q} \right) \sim \frac{\bar{c}_0(N)}{2\bar{c}_k(N)}, \end{aligned} \tag{5.18}$$

where the asymptotics comes from the term with $p = k$.

(2). $\underline{\xi \in \Omega_N^{(k)} \backslash \{0^k\}}$. Now, $p_t(0,\eta)$ is the same for all $\eta \in B_{|\xi|}$, and so we have

$$\begin{aligned} &N^{-k} \sum_{\xi \in \Omega_N^{(k)} \backslash \{0^k\}} \lambda^\xi(\omega) \sum_{\eta,\eta' \in B_{|\xi|}} G((0,0),(\eta,\eta')) \\ &\quad = N^k \sum_{\xi \in \Omega_N^{(k)} \backslash \{0^k\}} \lambda^\xi(\omega)\, G((0,0),(\eta^{(k)},\eta^{(k)})) \\ &\quad \sim N^k \sum_{d \in \mathbb{N}} \sum_{\substack{\xi \in \Omega_N^{(k)} \\ d_{\Omega_N^{(k)}}(0^k,\xi) = d}} \lambda^\xi(\omega)\, \frac{\bar{c}_0(N)}{2N^{k+d}\bar{c}_{k+d}(N)} \\ &\quad = \frac{1}{2} \sum_{d \in \mathbb{N}} \frac{\bar{c}_0(N)}{\bar{c}_{k+d}(N)} \left(\frac{1}{N^d} \sum_{\substack{\xi \in \Omega_N^{(k)} \\ d_{\Omega_N^{(k)}}(0^k,\xi) = d}} \lambda^\xi(\omega) \right), \end{aligned} \tag{5.19}$$

where we use (5.16–5.17) with $p = q = k$, and $d_{\Omega_N^{(k)}}$ denotes the distance within $\Omega_N^{(k)}$.

Combining (4.10), (5.6), (5.18)–(5.19), we arrive at

$$\bar{E}[H_N(\omega;\infty)] \sim \tfrac{1}{2}\,\bar{c}_0(N) \sum_{k \in \mathbb{N}_0} \frac{1}{\bar{c}_k(N)} \sum_{l=0}^{k} \lambda_l \Big\{ \tfrac{1}{2}\, 1_{\{l=k\}}\, \Theta(\omega;0,k) + 1_{\{l<k\}}\, \Theta(\omega;k-l,l) \Big\}, \tag{5.20}$$

where we abbreviate

$$\Theta(\omega;a,b) = \frac{1}{N[a]} \sum_{\substack{\xi \in \Omega_N^{(b)} \\ d_{\Omega_N^{(b)}}(0^b,\xi) = a}} \rho^\xi(\omega), \qquad a,b \in \mathbb{N}_0. \tag{5.21}$$

Now, by (2.41) we have, for some $C < \infty$,

$$\mathbb{E}\big[\Theta(\omega;a,b)\big] = 1, \quad \mathbb{E}\big[\Theta(\omega;a,b)\Theta(\omega;a',b')\big] \le C \quad \forall\, a,b,a',b' \in \mathbb{N}_0. \tag{5.22}$$

Because $\{\rho^\xi(\omega)\colon\ \xi \in \Omega_N^{\mathbb{T}}\}$ is stationary, ergodic and tail trivial (recall (2.42)), it follows from a standard second-moment estimate that the sum in the right-hand of (5.20) is infinite if and only if its expectation w.r.t. $\mathbb{P}$ is infinite. Since

$$\mathbb{E}(\text{r.h.s.}(5.20)) = \sum_{k \in \mathbb{N}_0} \frac{1}{\bar{c}_k(N)} \sum_{l=0}^{k} \lambda_l \big\{ \tfrac{1}{2}\, 1_{\{l=k\}} + 1_{\{l<k\}} \big\}, \tag{5.23}$$

we get the claim in (5.3) for the hierarchical random walk with homogeneous migration coefficients $\bar{c}_k(N)$ defined in (5.7) (the factor $\frac{1}{2}$ is harmless for the convergence or divergence of the right-hand side of (5.23)).

Step 2. It remains to show that the same dichotomy holds for the coefficients in (4.9) rather than (5.7). We start with the observation that the hierarchical random walk in random environment is *symmetric* and therefore is *reversible* with respect to the *Haar measure* on Ω_N. We have the representation (see Bovier and den Hollander, 2015, Chapter 7)

$$\begin{aligned} G^\omega((0,0),(\eta,\eta')) &= \int_0^\infty p_t^\omega(0,\eta)p_t^\omega(0,\eta')\,\mathrm{d}t \\ &= \frac{P^\omega_{(0,0)}(\tau_{(\eta,\eta')}<\infty)}{a^\omega((\eta,\eta'))\,P^\omega_{(\eta,\eta')}(\hat\tau_{(\eta,\eta')}=\infty)}, \quad (\eta,\eta')\in\Omega_N\times\Omega_N, \end{aligned} \tag{5.24}$$

where $a^\omega((a,b)) = \sum_{(c,d)} a^\omega((a,b),(c,d))$ is the total rate at which the random walk jumps out of (a,b), and

$$\begin{aligned} \tau_{(a,b)} &= \inf\big\{t\geq 0\colon\ Y_t(\omega)=(a,b)\big\}, \\ \hat\tau_{(a,b)} &= \inf\big\{t\geq 0\colon\ Y_t(\omega)=(a,b),\ \exists\, 0<s<t\colon\ Y_s(\omega)\neq(a,b)\big\}, \end{aligned} \tag{5.25}$$

are the first hitting time, respectively, the first return time of (a,b). The point of (5.24) is that both the numerator and the denominator can be controlled with the help of the *Dirichlet Principle*, as follows.

Let

$$\mathcal{E}^\omega(f,f) = \sum_{(a,b),(c,d)} \frac{a^\omega((a,b),(c,d))}{a^\omega((a,b))}\,[f((a,b))-f((c,d))]^2 \tag{5.26}$$

be the *Dirichlet form* associated with the two random walks in random environment. By classical potential theory, the escape probability in the denominator of (5.24) is given by the capacity of the pair (η,η') and ∞,

$$P^\omega_{(\eta,\eta')}(\hat\tau_{(\eta,\eta')}=\infty) = \mathrm{cap}^\omega((\eta,\eta'),\infty) = \inf_{\substack{f\colon\ \Omega_N\to[0,1]\\ f((\eta,\eta'))=1,\,f(\infty)=0}} \mathcal{E}^\omega(f,f), \tag{5.27}$$

where $f(\infty)=0$ stands for $\lim_{(\eta,\eta')\to\infty} f((\eta,\eta'))=0$ with $(\eta,\eta')\to\infty$ short hand for $d_{\Omega_N}(0,\eta)+d_{\Omega_N}(0,\eta')\to\infty$ (recall (2.2)). The hitting probability in the numerator of (5.24) can also be expressed in terms of capacities after we use a renewal argument. Write

$$\begin{aligned} P^\omega_{(0,0)}(\tau_{(\eta,\eta')}<\infty) &= \frac{P^\omega_{(0,0)}(\tau_{(\eta,\eta')}<\hat\tau_{(0,0)})}{1-P^\omega_{(0,0)}(\hat\tau_{(0,0)}<\tau_{(\eta,\eta')})} \\ &= \frac{P^\omega_{(0,0)}(\tau_{(\eta,\eta')}<\hat\tau_{(0,0)})}{P^\omega_{(0,0)}(\tau_{(\eta,\eta')}<\hat\tau_{(0,0)})+P^\omega_{(0,0)}(\tau_{(\eta,\eta')}=\hat\tau_{(0,0)}=\infty)}. \end{aligned} \tag{5.28}$$

We have

$$P^\omega_{(0,0)}(\tau_{(\eta,\eta')}<\hat\tau_{(0,0)}) = \mathrm{cap}^\omega((0,0),(\eta,\eta')) = \inf_{\substack{f\colon\ \Omega_N\to[0,1]\\ f((\eta,\eta'))=1,\,f((0,0))=0}} \mathcal{E}^\omega(f,f). \tag{5.29}$$

Moreover,

$$P^\omega_{(0,0)}(\tau_{(\eta,\eta')}=\hat\tau_{(0,0)}=\infty) = P^\omega_{(0,0)}(\hat\tau_{(0,0)}=\infty) - P^\omega_{(0,0)}(\tau_{(\eta,\eta')}<\infty,\hat\tau_{(0,0)}=\infty). \tag{5.30}$$

The first term equals $\mathrm{cap}^{\omega}((0,0),\infty)$, while the second term is bounded from above by $P^{\omega}_{(0,0)}(\tau_{(\eta,\eta')}<\infty)$, which tends to zero as $(\eta,\eta')\to\infty$ when $G^{\omega}<\infty$, i.e., when the random walk in random environment is transient. Below we will show that, under assumption (2.43), $G^{\omega}<\infty$ *if and only if* $G<\infty$.

We are now ready to explain why the estimates in Step 1 carry over. The transition rates of the random walk in random environment are given by

$$a^{\omega}((a,b)(c,d))=\begin{cases} a^{\omega,(N)}(a,c), & b=d,\\ a^{\omega,(N)}(b,d), & a=c,\\ 0, & \text{else.}\end{cases} \tag{5.31}$$

where $a^{\omega,(N)}$ is the transition kernel in (2.6), but with c_k replaced by $c_k(\omega)(N,\eta)$ in (4.9):

$$a^{\omega,(N)}(\eta,\zeta)=\begin{cases}\sum_{k\geq d_{\Omega_N}(\eta,\zeta)}\left(c_{k-1}+N^{-1}\lambda^{\mathrm{MC}_k(\eta)}(\omega)\right)/N^{2k-1}, & \eta\neq\zeta,\\ 0, & \eta=\zeta,\end{cases}\qquad \eta,\zeta\in\Omega_N. \tag{5.32}$$

By (4.10), we have $\lambda^{\mathrm{MC}_k(\eta)}(\omega)=\lambda_k\rho^{\mathrm{MC}_k(\eta)}(\omega)$. Assumption (2.43) implies $\delta\leq\lambda^{\mathrm{MC}_k(\eta)}(\omega)/\lambda_k\leq\delta^{-1}$ for all $k\in\mathbb{N}_0$, $\eta\in\Omega_N$ and $\mathbb{P}$-a.e. ω, which in turn implies

$$\delta\leq\frac{a^{\omega}((a,b)(c,d))}{a((a,b)(c,d))}\leq\delta^{-1}\quad\forall\,a,b,c,d\in\Omega_N\text{ for }\mathbb{P}\text{-a.e. }\omega, \tag{5.33}$$

where a is the transition kernel in (2.6), but with c_k replaced by $\bar{c}_k(N)$ in (5.7). Inserting these bounds into the formulas for the capacities in (5.27) and (5.29), and recalling (5.24), we see that

$$\exists\,\delta'>0\colon\quad \delta'\leq\frac{G^{\omega}((0,0),(\eta,\eta'))}{G((0,0),(\eta,\eta'))}\leq\delta'^{-1}\quad\forall\,\eta,\eta'\in\Omega_N\text{ for }\mathbb{P}\text{-a.e. }\omega. \tag{5.34}$$

This shows that the Green function for the random walk in random environment is comparable to the Green function of the homogeneous random walk. Hence, the argument in Step 1 carries over.

Note that $P^{\omega}_{(0,0)}(\hat{\tau}_{(0,0)}=\infty)$ in (5.30) is comparable to $P_{(0,0)}(\hat{\tau}_{(0,0)}=\infty)$, which is a strictly positive constant when $G<\infty$. Consequently, by the observation made below (5.30), also $P^{\omega}_{(0,0)}(\tau_{(\eta,\eta')}=\hat{\tau}_{(0,0)}=\infty)$ in (5.30) is comparable to $P_{(0,0)}(\tau_{(\eta,\eta')}=\hat{\tau}_{(0,0)}=\infty)$ when $G,G^{\omega}<\infty$.

It remains to show that, under assumption (2.43), $G^{\omega}<\infty$ if and only if $G<\infty$. This is easy. Indeed, if $G<\infty$, then $P_{(0,0)}(\hat{\tau}_{(0,0)}=\infty)>0$, hence $P^{\omega}_{(0,0)}(\hat{\tau}_{(0,0)}=\infty)>0$, and hence $G^{\omega}<\infty$ by (5.24) because $P^{\omega}_{(0,0)}(\tau_{(\eta,\eta')}<\infty)\leq 1$. Conversely, if $G=\infty$, then $P_{(0,0)}(\hat{\tau}_{(0,0)}=\infty)=0$, hence $P^{\omega}_{(0,0)}(\hat{\tau}_{(0,0)}=\infty)=0$, and hence $G^{\omega}=\infty$ by (5.24) because $P^{\omega}_{(0,0)}(\tau_{(\eta,\eta')}<\infty)>0$. □ □

5.2. *Zero-one law.* To conclude the proof of the dichotomy in Theorem 3.3, we use the following zero-one law.

Lemma 5.2 (Zero-one law). *For every* $N\in\mathbb{N}\backslash\{1\}$ *and* $\mathbb{P}$*-a.e.* ω, $H_N(\omega;\infty)=\infty$ *if and only if* $\bar{E}^{\omega}[H_N(\omega;\infty)]=\infty$. □

Proof: The proof comes in five steps.

Step 1. For $M, N \in \mathbb{N}$, let $H_N^{(M)}(\omega;\infty)$ denote the truncation of $H_N(\omega;\infty)$ obtained by setting $\lambda_k = 0$ for $k > M$ (no resampling in blocks of hierarchical size larger than M). The key to the proof is the following second-moment estimate:

$$\exists\, C < \infty\colon \qquad \bar{E}^\omega\big[\big(H_N^{(M)}(\omega;\infty)\big)^2\big] \le C\big(\bar{E}^\omega\big[H_N^{(M)}(\omega;\infty)\big]\big)^2 \qquad \forall\, M, N \in \mathbb{N}. \tag{5.35}$$

Before proving (5.35), we complete the proof of Theorem 3.3. By Cauchy-Schwarz, for any non-negative random variable V we have

$$\bar{P}^\omega(V > 0) \ge (\bar{E}^\omega[V])^2 / \bar{E}^\omega[V^2]. \tag{5.36}$$

Picking $V = H_N^{(M)}(\omega;\infty)/\bar{E}^\omega[H_N^{(M)}(\omega;\infty)]$ in (5.36) and using (5.35), we obtain

$$\bar{P}^\omega\Big(H_N^{(M)}(\omega;\infty)/\bar{E}^\omega\big[H_N^{(M)}(\omega;\infty)\big] > 0\Big) \ge \frac{1}{C} \qquad \forall\, M, N \in \mathbb{N}. \tag{5.37}$$

Since $H_N^{(M)}(\omega;\infty) \le H_N(\omega;\infty)$ and the lower bound in (5.37) is uniform in M and N, it follows that if $\bar{E}^\omega[H_N(\omega;\infty)] = \lim_{M\to\infty} \bar{E}^\omega[H_N^{(M)}(\omega;\infty)] = \infty$, then $\bar{P}^\omega(H_N(\omega;\infty) = \infty) \ge 1/C$. By (5.2), $\{\omega\colon H_N(\omega;\infty) = \infty\}$ is an element of the sigma-algebra at infinity defined in (2.42), which is trivial. The latter event therefore has probability either 0 or 1, and since it has positive probability we get the claim.

Step 2. Write out (recall (5.2))

$$\begin{aligned}
\bar{E}^\omega\big[\big(H_N^{(M)}(\omega;\infty)\big)^2\big] &= \sum_{k,l=0}^{M} N^{-k-l} \sum_{\substack{\eta,\eta'\in\Omega_N\\ d_{\Omega_N}(\eta,\eta')\le k}} \sum_{\substack{\zeta,\zeta'\in\Omega_N\\ d_{\Omega_N}(\zeta,\zeta')\le l}} \lambda^{\mathrm{MC}_k(\eta)}(\omega)\lambda^{\mathrm{MC}_l(\zeta)}(\omega) \\
&\quad \times \bar{E}^\omega\left[\int_0^\infty \mathrm{d}s\, 1_{\{Y_s(\omega)=\eta, Y'_s(\omega)=\eta'\}} \int_0^\infty \mathrm{d}u\, 1_{\{Y_u(\omega)=\zeta, Y'_u(\omega)=\zeta'\}}\right] \\
&= \sum_{k,l=0}^{M} N^{-k-l} \sum_{\substack{\eta,\eta'\in\Omega_N\\ d_{\Omega_N}(\eta,\eta')\le k}} \sum_{\substack{\zeta,\zeta'\in\Omega_N\\ d_{\Omega_N}(\zeta,\zeta')\le l}} \lambda^{\mathrm{MC}_k(\eta)}(\omega)\lambda^{\mathrm{MC}_l(\zeta)}(\omega) \\
&\quad \times 2\, G^\omega((0,0),(\eta,\eta'))\, G^\omega((\eta,\eta'),(\zeta,\zeta')) \\
&= 2\sum_{k,l=0}^{M} N^{-k-l} \sum_{\xi\in\Omega_N^{(k)}} \lambda^{\xi}(\omega) \sum_{\xi'\in\Omega_N^{(l)}} \lambda^{\xi'}(\omega) \\
&\quad \times \sum_{\eta,\eta'\in B_{|\xi|}} G^\omega((0,0),(\eta,\eta')) \sum_{\zeta,\zeta'\in B_{|\xi'|}} G^\omega((\eta,\eta'),(\zeta,\zeta')).
\end{aligned} \tag{5.38}$$

In what follows, we consider the hierarchical random walk with homogeneous migration coefficients $\bar{c}_k$ defined in (5.7). In Step 4 we incorporate the ω-dependence.

Use symmetry to replace $\sum_{k,l=0}^M$ by $2\sum_{k,l=0}^M 1_{\{k<l\}} + \sum_{k,l=0}^M 1_{\{k=l\}}$. Due to the ultrametricity of the hierarchical distance and the isotropy of the hierarchical random walk, we have $G((\eta,\eta'),(\zeta,\zeta')) = G((0,0),(\zeta,\zeta'))$ for all $\eta,\eta' \in B_{|\xi|}$ in the following three cases (where $\xi < \xi'$ means that ξ' is an ancestor of ξ):

(1) $k < l$, $\xi \not< \xi'$ and $\zeta, \zeta' \in B_{|\xi'|}$.
(2) $k < l$ and $\xi < \xi'$ and $\zeta, \zeta' \in B_{|\xi'|} \backslash B_{|\xi|}$.
(3) $k = l$, $\xi \ne \xi'$ and $\zeta, \zeta' \in B_{|\xi'|}$.

Therefore we have

$$\bar{E}\big[\big(H_N^{(M)}(\omega;\infty)\big)^2\big] = 2\bar{E}\big[H_N^{(M)}(\omega;\infty)\big]^2 + R \tag{5.39}$$

with R a correction term given by

$$\begin{aligned} R = 4 \sum_{0\le k<l\le M} N^{-k-l} \sum_{\xi\in\Omega_N^{(k)}} \lambda^{\xi}(\omega) \sum_{\xi'\in\Omega_N^{(l)}} \lambda^{\xi'}(\omega)\, 1_{\{\xi<\xi'\}} \sum_{\eta,\eta'\in B_{|\xi|}} G((0,0),(\eta,\eta')) \\ \times \sum_{\substack{\zeta,\zeta'\in B_{|\xi'|} \\ \zeta\in B_{|\xi|} \text{ and/or } \zeta'\in B_{|\xi|}}} \big[G((\eta,\eta'),(\zeta,\zeta')) - G((0,0),(\zeta,\zeta'))\big] \\ + 2 \sum_{0\le k\le M} N^{-2k} \sum_{\xi\in\Omega_N^{(k)}} [\lambda^{\xi}(\omega)]^2 \sum_{\eta,\eta'\in B_{|\xi|}} G((0,0),(\eta,\eta')) \\ \times \sum_{\zeta,\zeta'\in B_{|\xi|}} \big[G((\eta,\eta'),(\zeta,\zeta')) - G((0,0),(\zeta,\zeta'))\big]. \end{aligned} \tag{5.40}$$

If R would be absent from (5.39), then we would have proved (5.35) with $C = 2$. Thus, it remains to show that R can only raise the constant. We will do this by showing that $R \le O(N^{-2})\, \bar{E}[H_N^{(M)}(\omega;\infty)]^2$ as $N\to\infty$, uniformly in M, and by appealing to the observation made in (5.13).

Step 3. By translation invariance, $G((\eta,\eta'),(\zeta,\zeta')) = G((0,0),(\zeta-\eta,\zeta'-\eta'))$. By isotropy, $\sum_{\zeta,\zeta'\in B_{|\xi|}} G((0,0),(\zeta-\eta,\zeta'-\eta')) = \sum_{\zeta,\zeta'\in B_{|\xi|}} G((0,0),(\zeta,\zeta'))$ for all $\eta,\eta'\in B_{|\xi|}$. Hence, in the first sum in (5.40) the term with $\zeta,\zeta'\in B_{|\xi|}$ vanishes, while the second sum in (5.40) vanishes altogether, and so R simplifies to

$$\begin{aligned} R = 8 \sum_{0\le k<l\le M} N^{-k-l} \sum_{\xi\in\Omega_N^{(k)}} \lambda^{\xi}(\omega) \sum_{\xi'\in\Omega_N^{(l)}} \lambda^{\xi'}(\omega)\, 1_{\{\xi<\xi'\}} \sum_{\eta,\eta'\in B_{|\xi|}} G((0,0),(\eta,\eta')) \\ \times \sum_{\substack{\zeta\in B_{|\xi|} \\ \zeta'\in B_{|\xi'|}\setminus B_{|\xi|}}} \big[G((0,0),(\zeta-\eta,\zeta'-\eta')) - G((0,0),(\zeta,\zeta'))\big]. \end{aligned} \tag{5.41}$$

By isotropy, $\sum_{\zeta\in B_{|\xi|}} G((0,0),(\zeta-\eta,\zeta'-\eta')) = \sum_{\zeta\in B_{|\xi|}} G((0,0),(\zeta,\zeta'-\zeta))$ for all $\eta,\eta'\in B_{|\xi|}$ when $\zeta'\in B_{|\xi'|}\setminus B_{|\xi|}$, and so R simplifies further to

$$\begin{aligned} R = 8 \sum_{0\le k<l\le M} N^{-k-l} \sum_{\xi\in\Omega_N^{(k)}} \lambda^{\xi}(\omega) \sum_{\xi'\in\Omega_N^{(l)}} \lambda^{\xi'}(\omega)\, 1_{\{\xi<\xi'\}} \sum_{\eta,\eta'\in B_{|\xi|}} G((0,0),(\eta,\eta')) \\ \times \sum_{\zeta\in B_{|\xi|}} \left[\sum_{\zeta''\in B_l(0)\setminus B_k(0)} G((0,0),(\zeta,\zeta'')) - \sum_{\zeta'\in B_{|\xi'|}\setminus B_{|\xi|}} G((0,0),(\zeta,\zeta')) \right]. \end{aligned} \tag{5.42}$$

If $0\in B_{|\xi'|}$, then $B_l(0) = B_{|\xi'|}$, in which case the term between brackets equals

$$\sum_{\zeta'\in B_{|\xi|}} G((0,0),(\zeta,\zeta')) - \sum_{\zeta''\in B_k(0)} G((0,0),(\zeta,\zeta'')). \tag{5.43}$$

If also $0\in B_{|\xi|}\subset B_{|\xi'|}$, then also $B_k(0) = B_{|\xi|}$, in which case the latter difference vanishes. Hence we obtain the bound

$$\begin{aligned} R \le 8 \sum_{0\le k<l\le M} N^{-k-l} \sum_{\xi\in\Omega_N^{(k)}} \lambda^{\xi}(\omega) \sum_{\xi'\in\Omega_N^{(l)}} \lambda^{\xi'}(\omega)\, 1_{\{\xi<\xi'\}} \sum_{\eta,\eta'\in B_{|\xi|}} G((0,0),(\eta,\eta')) \end{aligned}$$

$$\times \left[1_{\{0 \not\leq \xi,\, 0 \leq \xi'\}} \sum_{\zeta, \bar{\zeta} \in B_{|\xi|}} G((0,0),(\zeta,\bar{\zeta})) + 1_{\{0 \not\leq \xi'\}} \sum_{\substack{\zeta \in B_{|\xi|} \\ \bar{\zeta} \in B_l(0) \setminus B_k(0)}} G((0,0),(\zeta,\bar{\zeta})) \right]. \tag{5.44}$$

The sums over η, η' and $\zeta, \bar{\zeta}$ can be computed with the help of (5.16). Recalling (5.17)–(5.19), we obtain

$$\begin{aligned} \sum_{\eta,\eta' \in B_{|\xi|}} G((0,0),(\eta,\eta')) &= \sum_{\zeta,\bar{\zeta} \in B_{|\xi|}} G((0,0),(\eta,\eta')) \\ &\sim N^{2k} \frac{\bar{c}_0(N)}{2N^{k+d(\xi)} \bar{c}_{k+d(\xi)}(N)} = N^k \frac{\bar{c}_0(N)}{2N^{d(\xi)} \bar{c}_{k+d(\xi)}(N)} \end{aligned} \tag{5.45}$$

with $d(\xi) = d_{\Omega_N^{(k)}}(0^k, \xi)$ and

$$\begin{aligned} \sum_{\substack{\zeta \in B_{|\xi|} \\ \bar{\zeta} \in B_l(0) \setminus B_k(0)}} G((0,0),(\zeta,\bar{\zeta})) &\sim \sum_{\substack{\zeta \in B_{|\xi|} \\ \bar{\zeta} \in B_l(0) \setminus B_k(0)}} \frac{\bar{c}_0(N)}{N^{l+d'(\xi')} \bar{c}_{d''(\bar{\zeta})}(N)} \\ &= N^{-l+k} \sum_{\bar{\zeta} \in B_l(0) \setminus B_k(0)} \frac{\bar{c}_0(N)}{N^{d'(\xi')} \bar{c}_{d''(\bar{\zeta})}(N)} \sim N^k \frac{\bar{c}_0(N)}{N^{d'(\xi')} \bar{c}_l(N)} \end{aligned} \tag{5.46}$$

with $d'(\xi') = d_{\Omega_N^{(l)}}(0^l, \xi')$ and $d''(\bar{\zeta}) = d_{\Omega_N}(0, \bar{\zeta})$. Here we use that $\xi \neq 0^k$ when $0 \not\leq \xi$ and $\xi' \neq 0^l$ when $0 \not\leq \xi'$, and also that $l + d'(\xi') > d''(\bar{\zeta})$ for all $\bar{\zeta} \in B_l(0)$. Inserting (5.45)–(5.46) into (5.44), we get

$$\begin{aligned} R \leq 8\,[1+o(1)] \sum_{0 \leq k < l \leq M} N^{k-l} \sum_{\xi \in \Omega_N^{(k)}} \lambda^{\xi}(\omega) \sum_{\xi' \in \Omega_N^{(l)}} \lambda^{\xi'}(\omega)\, 1_{\{\xi < \xi',\, 0 \not\leq \xi'\}} \\ \times \left[1_{\{0 \not\leq \xi,\, 0 \leq \xi'\}} \left(\frac{\bar{c}_0(N)}{2N^{d(\xi)} \bar{c}_{k+d(\xi)}(N)} \right)^2 + 1_{\{0 \not\leq \xi'\}} \frac{\bar{c}_0(N)}{2N^{d(\xi)} \bar{c}_{k+d(\xi)}(N)} \frac{\bar{c}_0(N)}{N^{d'(\xi')} \bar{c}_l(N)} \right]. \end{aligned} \tag{5.47}$$

Step 4. If $0 \not\leq \xi$, then $d(\xi) \in \mathbb{N}$. Hence the first part of (5.47) equals $8\,[1+o(1)]$ times

$$\sum_{0 \leq k < l \leq M} N^{k-l} \lambda_k \lambda_l \sum_{d=1}^{l} \left(\frac{\bar{c}_0(N)}{2N^d \bar{c}_{k+d}(N)} \right)^2 \sum_{\substack{\xi \in \Omega_N^{(k)} \\ d(\xi) = d}} \rho^{\xi}(\omega) \rho^{\xi^{l-k}}(\omega), \tag{5.48}$$

where we recall (4.10) and write ξ^{l-k} to denote the ancestor of ξ at height l. Because $\{\rho^{\xi}(\omega)\colon\ \xi \in \Omega_N^{\mathbb{T}}\}$ is stationary, ergodic and tail trivial (recall (2.42)), the last sum scales as $\sim N^d \mathbb{E}[\rho^{0^k}(\omega) \rho^{0^l}(\omega)]$, where the expectation is finite because of (2.41). Hence, (5.48) is

$$\leq C[1+o(1)]\, \tfrac{1}{4} \bar{c}_0(N)^2 \sum_{0 \leq k < l \leq M} N^{k-l} \lambda_k \lambda_l \sum_{d=1}^{l} \frac{1}{N^d \bar{c}_{k+d}(N)^2}. \tag{5.49}$$

The last sum scales as $\sim 1/N\bar{c}_{k+1}(N)^2$, and so (5.48) is

$$\leq C[1+o(1)]\,\tfrac{1}{4}\bar{c}_0(N)\sum_{k=0}^{M}\frac{\lambda_k}{\bar{c}_{k+1}(N)^2}\sum_{l=k+1}^{M}\lambda_l N^{k-l-1}=O(N^{-2})\,\bar{E}[H_N^{(M)}(\omega;\infty)]^2, \tag{5.50}$$

where the equality follows from (5.20) with k,l truncated at M.

If $0\not\leq\xi'$, then $d'(\xi')\in\mathbb{N}$ and $d(\xi)=l-k+d'(\xi')$. Hence, the second part of (5.47) equals $8\,[1+o(1)]$ times

$$\sum_{0\leq k<l\leq M}N^{k-l}\lambda_k\lambda_l\sum_{d'\in\mathbb{N}}\frac{\bar{c}_0(N)}{2N^{l-k+d'}\bar{c}_{l+d'}(N)}\frac{\bar{c}_0(N)}{N^{d'}\bar{c}_l(N)}\sum_{\substack{\xi\in\Omega_N^{(k)}\\ d(\xi)=l-k+d'}}\rho^{\xi}(\omega)\rho^{\xi^{l-k}}(\omega). \tag{5.51}$$

The last sum is $\leq C[1+o(1)]N^{l-k+d'}$. Hence (5.51) is

$$\leq C[1+o(1)]\,\tfrac{1}{2}\bar{c}_0(N)^2\sum_{0\leq k<l\leq M}N^{k-l}\lambda_k\lambda_l\sum_{d'\in\mathbb{N}}\frac{1}{N^{d'}\bar{c}_{l+d'}(N)\bar{c}_l(N)}. \tag{5.52}$$

The last sum scales as $\sim 1/N\bar{c}_{l+1}(N)\bar{c}_l(N)$, and so (5.51) is

$$\leq C[1+o(1)]\,\tfrac{1}{2}\bar{c}_0(N)^2\sum_{k=0}^{M}\lambda_k\sum_{l=k+1}^{M}\frac{\lambda_l}{\bar{c}_{l+1}(N)\bar{c}_l(N)}\,N^{k-l-1}=O(N^{-2})\,\bar{E}[H_N^{(M)}(\omega;\infty)]^2. \tag{5.53}$$

Step 5. We can again use (5.34) to show that the proof carries over to the random walk in random environment. □

Lemmas 5.1–5.2 combine to yield Theorem 3.3 (recall the discussion at the beginning of this section).

6. Multi-scale analysis

In this section, we prove Theorem 3.6. We first consider a *mean-field system*, i.e., the geographic space is $G=\{1,\dots,N\}$ with $N\to\infty$. In Section 6.1, we look at this system on time scale t (on which the single components evolve) and on time scale Nt (on which the block average evolves). In Section 6.2, we use the results to analyze the system on Ω_N as $N\to\infty$. Our general strategy runs parallel to that in Greven et al. (2014) for the homogeneous model. We only point out which *new issues* arise. Thus, this section is *not autonomous*, the principal steps of the arguments are given but not all formulas are repeated, and for an understanding of the fine details the reader must check the relevant passages in Greven et al. (2014).

6.1. *The mean-field finite-system scheme.* As geographic space and transition kernel, we take

$$\Omega=\{1,\dots,N\},\qquad a(i,j)=\frac{1}{N},\quad i,j\in\Omega. \tag{6.1}$$

As migration rate, we take c_0, and as resampling measures

$$\Lambda^i=\lambda_0\chi^i,\quad i\in\Omega, \tag{6.2}$$

with total masses $\rho^i=\chi^i((0,1])$. We assume that $(\chi^i)_{i\in\mathbb{N}}$ is stationary and ergodic such that ϱ^i has mean 1. We also allow a component with Fleming-Viot resampling

at rate d_0. The corresponding stochastic system is denoted by $(Z^{(N)}(t))_{t\geq 0}$ with $Z^{(N)}(t) = (Z_1^{(N)}(t), \ldots, Z_N^{(N)}(t))$.

We consider time scales t and Nt for the components, respectively, the block average:

$$\begin{aligned} &(Z^{(N)}(t))_{t\geq 0}, \\ &(\bar{Z}^{(N)}(t))_{t\geq 0} \text{ with } \bar{Z}^{(N)}(t) = \frac{1}{N}\sum_{i=1}^{N} Z_i^{(N)}(Nt). \end{aligned} \tag{6.3}$$

Theorem 6.1 ([**Mean-field finite-system scheme).** *Suppose that the initial state is i.i.d. with mean measure $\theta \in \mathcal{P}(E)$. Then*

$$\lim_{N\to\infty} \mathcal{L}\left[(Z^{(N)}(t))_{t\geq 0}\right] = \bigotimes_{i\in\mathbb{N}} \mathcal{L}\left[(Z_\theta^{c_0,d_0,\Lambda^i}(t))_{t\geq 0}\right] \tag{6.4}$$

and

$$\lim_{N\to\infty} \mathcal{L}\left[(\bar{Z}^{(N)}(t))_{t\geq 0}\right] = \mathcal{L}\left[(Z_\theta^{0,d_1,0}(t))_{t\geq 0}\right], \tag{6.5}$$

where $(Z_\theta^{c,d,\Lambda}(t))_{t\geq 0}$ is the McKean-Vlasov process defined in Section 3.2.1. □

Proof: We follow Greven et al. (2014, Section 6). The proof of (6.4) carries over in a straightforward way. In the proof of (6.5) a new issue arises: the increasing process of the limit process incorporates an additional averaging over the random environment controlling the resampling for the single components. This is handled as follows.

Calculate the generator for a polynomial of $\bar{Z}^{(N)}(t)$, namely, a function F of the form

$$F(z) = \langle f, z^{\otimes n}\rangle, \quad f \in C_b(E^n, \mathbb{R}), \tag{6.6}$$

applied to a $z \in E$ of the form $z = \frac{1}{N}\sum_{i=1}^{N} z_i$. This expression can be expanded in terms of sums of products of monomials of single components. The action of the generator was calculated and analysed in Greven et al. (2014, Section 6). We can argue in the same way with the following changes. In the action of the generator, integrals are taken with respect to the random sequence of resampling measures $(\Lambda^i)_{i\in\Omega}$ rather than a fixed resampling measure Λ. This entails that for the block average we get a sum of terms where the random sequence $(\rho^i)_{i\in\Omega}$ appears as weights. This in turn requires us to change the definition of the set of configurations on which the generator converges in the limit as $N \to \infty$ (see Greven et al., 2014, Eq. (6.41)–(6.42)) as follows.

Let $\mathbb{B}^*$ be the set of $\underline{x} = (x_i)_{i\in\Omega} \in \mathcal{P}(E)^{\mathbb{N}}$ with

$$\lim_{N\to\infty} \mathcal{L}\left[\frac{1}{N}\sum_{i=1}^{N} \delta_{(\chi^i(\omega),x_i)}\right] = \Gamma \in \mathcal{P}\big(\mathcal{M}_f([0,1]) \times \mathcal{P}(E)\big), \tag{6.7}$$

where

$$\Gamma(\,\cdot\,, \mathcal{P}(E)) = \mathcal{L}[\chi^1], \quad \Gamma(\mathrm{d}x_1 \mid \chi^1) = \nu^{c_0,0,\chi^1}(\mathrm{d}x_1), \quad x_1 \in \mathcal{P}(E). \tag{6.8}$$

In order to calculate the sum of the resampling operators as in Greven et al. (2014, Eq. (6.46)), we have to account for the presence of χ^i, $i \in \Omega$, and invoke the law of large numbers for the expression in the variance formula, namely, $2c_0/(2c_0 + \lambda_0\rho^i + 2d_0)$, $i \in \Omega$. We write the latter as

$$\frac{c_0}{c_0 + \mu_0\rho^i + d_0}, \qquad i \in \Omega. \tag{6.9}$$

The expressions appearing in the generator, which are averages of local functions of the configuration and their shifts to any of the N locations, result in the same expression as the one we obtain by using (6.9) averaged over $i \in \Omega$. In the limit as $N \to \infty$, this leads to the recursion formula in (3.17) for $k = 0$. With these changes, the argument runs as in the case of the homogeneous environment. □

6.2. *The hierarchical mean-field limit.* In this section, we prove the results claimed in Section 3.2.3. The strategy of the proof is to approximate our system with *infinitely* many hierarchies of components and time scales by systems with *finitely* many hierarchies of components and time scales, uniformly in N. The latter are analyzed by using the multiscale analysis of the mean-field system. In Section 6.2.1, we consider 2-level systems with N^2 components, in Section 6.2.2 k-level systems with N^k components, and in Section 6.2.3 we pass to the limit $k \to \infty$ of infinitely many hierarchies. Along the way, we make frequent reference to Dawson, Greven and Vaillancourt (Dawson et al., 1995) and the work on the homogeneous version of the model in Greven et al. (2014).

6.2.1. *The 2-level system on 3 time scales.* The geographic space is

$$G_{N,2} = \{0, 1, \dots, N-1\}^2 = G_{N,1}^2. \tag{6.10}$$

We pick $d_0 > 0$, $c_0, c_1, \mu_0, \mu_1 > 0$ and $c_k, \mu_k = 0$ for $k \geq 2$. We choose the random environment that is obtained by restricting the random environment of Section 2 to the subtree corresponding to the 2-block around 0. We show that, on time scales t and Nt, we obtain the same limiting objects as described in Section 6.1, but with additional volatility and block resampling.

For the 1-block averages, we use the notation

$$Y_\eta^{(N)}(t) = N^{-1} \sum_{\sigma \in G_{N,1}} X_{(\sigma,\eta)}^{(N)}(t), \qquad \eta \in G_{N,1}, \tag{6.11}$$

and for the 2-block average (= total average)

$$Z^{(N)}(t) = N^{-2} \sum_{(\sigma,\eta) \in G_{N,2}} X_{(\sigma,\eta)}^{(N)}(t) = N^{-1} \sum_{\eta \in G_{N,1}} Y_\eta^{(N)}(t). \tag{6.12}$$

Proposition 6.2 ([Two-level rescaling). *Under the above assumptions,*

$$\begin{aligned}
&\lim_{N\to\infty} \mathcal{L}\left[\left(X_{(\sigma,\eta)}^{(N)}(t)\right)_{t\geq 0}\right] = \mathcal{L}\left[\left(Z_\theta^{c_0,d_0,\Lambda^{\mathrm{MC}_1(\cdot,\eta)}(\omega)}(t)\right)_{t\geq 0}\right] \qquad \forall\, (\sigma,\eta) \in G_{N,2},\\
&\lim_{N\to\infty} \mathcal{L}\left[\left(Y_\eta^{(N)}(Nt)\right)_{t\geq 0}\right] = \mathcal{L}\left[\left(Z_\theta^{c_1,d_1,\Lambda^{\mathrm{MC}_2(\cdot,\cdot)}(\omega)}(t)\right)_{t\geq 0}\right] \qquad \forall\, \eta \in G_{N,1},\\
&\lim_{N\to\infty} \mathcal{L}\left[\left(Z^{(N)}(N^2t)\right)_{t\geq 0}\right] = \mathcal{L}\left[\left(Z_\theta^{0,d_2,0}(t)\right)_{t\geq 0}\right],
\end{aligned} \tag{6.13}$$

with

$$d_1 = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{c_0(\mu_0\rho(\omega)+d_0)}{c_0+(\mu_0\rho(\omega)+d_0)}\right], \qquad d_2 = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{c_1(\mu_1\rho(\omega)+d_1)}{c_1+(\mu_1\rho(\omega)+d_1)}\right]. \tag{6.14}$$

□

To prove the above results in the homogeneous environment, we used *uniform estimates* for higher-order perturbations of generators. These no longer hold in the random environment, due to the unboundedness of the random resampling rates $\rho^{(\cdot,\eta)}(\omega)$. (There is no problem under assumption (2.43), and the proof carries over from Greven et al. (2014).)

To handle this problem, we first consider the system where the coefficients $\lambda^{\mathrm{MC}_k(\cdot,\eta)}(\omega)$, $k = 1, 2$, are truncated at level $M < \infty$. For this system, we show, with the help of a coupling argument, that on time scale $N^k t$, $k = 1, 2$, and averaged over the random environment and the dynamics, the effect of the truncation goes to zero as $M \to \infty$. The same holds for the limiting objects, so that we get the claim by using the existence of the expectation in combination with the stationarity of ω.

To get tightness of the approximating sequence of processes, as in Greven et al. (2014, Eq. (7.52), p. 117), we use the fact that the laws conditioned on the environment ω of the averages in (6.11)–(6.12) are tight. To prove the latter, we use the criterion of Joffe and Metivier in the form as given in Dawson (1993, p. 55), observing that $\chi^{\mathrm{MC}_k(\cdot,\eta)}(\omega)$, $\eta \in G_{N,1}$, $k = 1, 2$, are integrable uniformly in N. To check the criterion, we observe that we can code the information on the random environment into the initial condition of the process. With this observation, the proof works as for the homogeneous environment.

6.2.2. *The k-level system on $k+1$ time scales.* The reasoning addresses the same points raised above and runs otherwise exactly as in Greven et al. (2014, Section 7.2).

6.2.3. *The infinite-level system on infinitely many time scales.* The problem is again the extension of the uniform perturbation arguments, which have to be adapted to guarantee that cutting off higher hierarchical levels leads to an approximation by finite systems, for which we can apply the reasoning in the previous section, on the relevant time scales. To get the necessary arguments and estimates, we refer the reader to the material in Greven et al. (2014, Sections 8.1–8.2).

The argument used for the homogeneous environment to obtain uniforms bounds does not apply because the perturbation of the migration and the resampling coming from the hierarchical levels $\geq k+1$ is unbounded. However, the perturbation terms can be stochastically bounded by a random variable that has a *finite expectation* over the random environment. Again, it suffices to show with the help of a coupling argument that the stochastic dynamics with k hierarchical levels approximates the infinite stochastic dynamics on time scales tN^l with $0 \leq l \leq k$. Apart from that, the argument is the same.

6.3. *Dichotomy in the hierarchical mean-field limit.* In this section, we prove Theorem 3.13. First, we argue that the entrance law exists, a fact that was established in Dawson et al. (1995)[Section 6(a), Proposition 6.2] for the Fleming-Viot model, based on a variance estimate and the convergence of the sum in the coexistence criterion. The argument from that paper carries over despite the ω-dependence of the transition kernels of the interaction chain (read this of from (6.16) and (6.18) below).

Next, we argue that the dichotomy holds. Here, we again follow the strategy for the homogeneous environment by calculating the variance of $\langle M_{\eta,0}^{(j)}, f\rangle$ for every $\eta \in \Omega_\infty$ and $f \in C_b(E,\mathbb{R})$ and showing that as $j \to \infty$ this variance converges to zero, respectively, remains positive, depending on whether the sum in (3.30) is infinite or finite.

The variance formula reads

$$\mathrm{Var}_{\nu_\theta^{c,d,\Lambda}}(\langle\,\cdot\,, f\rangle) = \frac{2c}{2c+\lambda\rho(\omega)+2d}\,\mathrm{Var}_\theta(f). \tag{6.15}$$

Consequently, by iteration,

$$\mathrm{Var}\langle M_{\eta,0}^{(j)}, f\rangle = \left[\prod_{k=0}^{j} \frac{2c_k}{2c_k+\lambda_k\rho_k(\omega)+2d_k}\right]\mathrm{Var}_\theta(f), \tag{6.16}$$

where $\underline{d} = (d_k)_{k\in\mathbb{N}_0}$ is determined by the recursion relation in (3.17). Taking logarithms, we see that the product tends to a positive limit as $j \to \infty$ if and only if

$$\sum_{k\in\mathbb{N}_0} \frac{1}{c_k}(\mu_k\rho_k(\omega)+d_k) < \infty. \tag{6.17}$$

By assumptions (2.41)–(2.42), the sum converges ω-a.s. if and only if

$$\sum_{k\in\mathbb{N}_0} \frac{1}{c_k}(\mu_k+d_k) < \infty. \tag{6.18}$$

Indeed, the variance of the sum in (6.17) equals the variance of the ρ-field times $\sum_{k\in\mathbb{N}_0}(\frac{\mu_k}{c_k})^2$, and the latter is bounded from above by the square of the average of the sum. As shown in Greven et al. (2014, Theorem 3.7(c)), the criterion in (6.18) is the same as the criterion in (3.30).

7. The orbit of the renormalization transformations

In Section 7.1, we show the ordering in Theorem 3.10. In Sections 7.2–7.3, we derive the scaling behaviour in Theorems 3.15–3.17.

7.1. *Random environment lowers the volatility.*

Proof of Theorem 3.10: Recall the notation introduced in Section 2.2. Fix $\underline{c}$ and $\underline{\lambda}$. Let $\underline{d}$ be the solution of the recursion relation in (3.17). Let $\underline{d}^0, \underline{d}^1$ be the solutions when $\mathcal{L}_\rho$ is replaced by δ_0, δ_1 (recall that ρ has mean 1 under $\mathcal{L}_\rho$). As initial values take $d_0^0 \le d_0 \le d_0^1$. We use induction on k to show that $d_k^0 < d_k < d_k^1$ for all $k \in \mathbb{N}$.

Define (see Fig. 7.4)

$$f_k^0(x) = \frac{c_k x}{c_k+x}, \qquad f_k(x) = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{c_k(\mu_k\rho+x)}{c_k+(\mu_k\rho+x)}\right], \qquad f_k^1(x) = \frac{c_k(\mu_k+x)}{c_k+(\mu_k+x)}. \tag{7.1}$$

Because $a \mapsto c_k(\mu_k a+x)/[c_k+(\mu_k a+x)]$ is strictly increasing and strictly concave on $[0,\infty)$ for all $x \in [0,\infty)$, it follows that $f_k^0(x) < f_k(x) < f_k^1(x)$ for all $x \in [0,\infty)$. Hence, if $d_k^0 \le d_k \le d_k^1$, then $d_{k+1}^0 = f_k^0(d_k^0) < f_k(d_k^0) \le f_k(d_k) = d_{k+1}$ and $d_{k+1} = f_k(d_k) < f_k^1(d_k) \le f_k^1(d_k^*) = d_{k+1}^1$. □

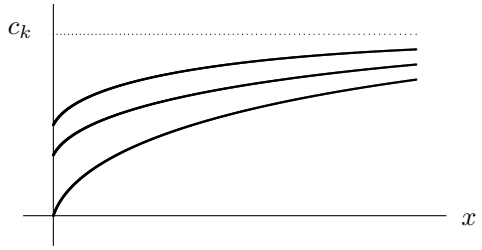

FIGURE 7.4. Qualitative pictures of $x \mapsto f_k^0(x)$ (bottom), $x \mapsto f_k(x)$ (middle) and $x \mapsto f_k^1(x)$ (top). All three functions are strictly increasing and strictly concave on $[0,\infty)$, and tend to c_k as $x \to \infty$.

The same argument proves the claim made in Section 3.3 that $M < M^*$ for the fixed points of (3.35) (random environment) and its analogue with $\mathcal{L}_\rho$ replaced by δ_1 (average environment).

7.2. *Scaling of the volatility: polynomial coefficients.*

Proof of Theorem 3.15: We look at each of the four parameter regimes separately. Recall (3.33)–(3.34).

(a) Let $K_k = \mu_k/c_{k-1}$, $R_k = c_k/c_{k-1}$ and $\mho_k = d_k/c_{k-1}$. Rewrite (3.17) as

$$\mho_{k+1} = g_k(\mho_k) \quad \text{with} \quad g_k(x) = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{(K_k\rho + x)}{R_k + (K_k\rho + x)}\right]. \tag{7.2}$$

Since g_k is non-decreasing on $[0,\infty)$, we have the sandwich

$$g_k(0) \le \mho_{k+1} \le g_k(\infty) = 1. \tag{7.3}$$

We are in the regime where $\lim_{k\to\infty} K_k = K = \infty$ and $\lim_{k\to\infty} R_k = R = 1$. Hence $\lim_{k\to\infty} g_k(0) = 1$, and so (7.3) yields $\lim_{k\to\infty} d_k/c_k = \lim_{k\to\infty} \mho_k/R_k = 1/R = 1$.

(b) Again use (7.2). We are in the regime where $\lim_{k\to\infty} K_k = K \in (0,\infty)$ and $\lim_{k\to\infty} R_k = R = 1$. Hence, we see that g_k converges point-wise to g given by

$$g(x) = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{(K\rho + x)}{R + (K\rho + x)}\right]. \tag{7.4}$$

Both g and g_k are strictly increasing and strictly concave on $[0,\infty)$, with $g([0,\infty]) \subseteq [0,1]$ and $g_k([0,\infty]) \subseteq [0,1]$, with unique attracting fixed points $M \in (0,1)$ and $M_k \in (0,1)$, and with M the solution of (3.35). To show that $\lim_{k\to\infty} \mho_k = M$, we need two facts.

Lemma 7.1. *Let* $s_k = \sup_{x\in[0,1]} |g_k(x) - g(x)|$. *Then* $\lim_{k\to\infty} s_k = 0$. □

Proof: Estimate

$$\begin{aligned} &\frac{(K_k\rho + x)}{R_k + (K_k\rho + x)} - \frac{(K\rho + x)}{R + (K\rho + x)} = \frac{R}{R + (K\rho + x)} - \frac{R_k}{R_k + (K_k\rho + x)} \\ &= \frac{(RK_k - R_kK)\rho + (R - R_k)x}{[R + (K\rho + x)][R_k + (K_k\rho + x)]} \le \frac{(RK_k - R_kK)}{KR_k} + \frac{R - R_k}{RR_k}\,x. \end{aligned} \tag{7.5}$$

This gives

$$s_k \leq \frac{R}{K}\left|\frac{K_k}{R_k} - \frac{K}{R}\right| + \left|\frac{1}{R_k} - \frac{1}{R}\right|. \tag{7.6}$$

Let $k \to \infty$ to get the claim. □

Lemma 7.2. *Function g is a strict contraction around M, i.e., there exists a $\beta \in (0,1)$ such that $\sup_{x\in[0,\infty)}(g(x)-M)/(x-M) = \beta$.* □

Proof: Consider the linear function $L(x) = g(0) + [1 - \frac{g(0)}{M}]x$, $x \in [0,\infty)$, which satisfies $L(0) = g(0)$ and $L(M) = M = g(M)$ (see Fig. 7.5). Note that $g \geq L$ on $[0, M]$ while $g \leq L$ on $[M, \infty)$. Hence, we have

$$0 \leq \frac{g(x)-M}{x-M} \leq \frac{L(x)-M}{x-M} = 1 - \frac{g(0)}{M}. \tag{7.7}$$

Since $g(0) > 0$, we get the claim with $\beta = 1 - \frac{g(0)}{M}$. □

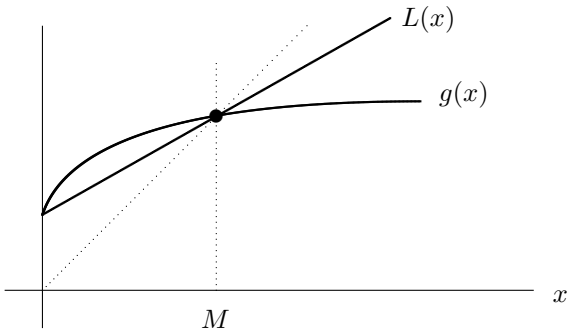

FIGURE 7.5. Comparison of g and L.

We can now complete the proof as follows. Let $\Delta_k = |\mho_k - M|$. Then

$$\begin{aligned}\Delta_{k+1} &= |\mho_{k+1} - M| \leq |\mho_{k+1} - g(\mho_k)| + |g(\mho_k) - M| \\ &= |g_k(\mho_k) - g(\mho_k)| + |g(\mho_k) - M| \leq s_k + \beta\Delta_k.\end{aligned} \tag{7.8}$$

Iteration yields

$$\Delta_{k+1} \leq \sum_{l=0}^{k} \beta^l s_{k-l} + \beta^{k+1}\Delta_0. \tag{7.9}$$

It follows from Lemma 7.1–7.2 that $\lim_{k\to\infty}\Delta_k = 0$. Hence $\lim_{k\to\infty} d_k/c_k = \lim_{k\to\infty}\mho_k/R_k = M/R = M$.

(c–d) Like in Case (a), the scaling turns out to be the *same* as for the average environment. The proof is based on a *comparison* between the recursions for the random environment and the average environment (last two items in (7.1)). The key idea is the following lemma, which can be viewed as a *stability property*.

Lemma 7.3. *Let $d_0 = d_0^1$. Then, the solution of the recursion $d_{k+1} = f_k(d_k)$, $k \in \mathbb{N}_0$, is the same as the solution of the recursion $d_{k+1}^1 = f_k^1(d_k^1)$, $k \in \mathbb{N}_0$, when in the latter recursion the coefficient μ_k is replaced by $\mu_k r_k$ with*

$$r_k = \frac{N_k}{D_k}, \qquad N_k = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{c_k\rho}{c_k(1+K_k\rho)+d_k}\right], \qquad D_k = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{c_k}{c_k(1+K_k\rho)+d_k}\right]. \tag{7.10}$$

□

Proof: Check that

$$\frac{c_k(\mu_k r_k + d_k)}{c_k + (\mu_k r_k + d_k)} = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{c_k(\mu_k\rho + d_k)}{c_k + (\mu_k\rho + d_k)}\right] = d_{k+1}, \tag{7.11}$$

and use induction on k. □

Since $\rho \mapsto c_k/[c_k(1+K_k\rho)+d_k]$ is non-increasing, we have $N_k \le D_k\mathbb{E}_{\mathcal{L}_\rho}[\rho] = D_k$, and so $r_k \le 1$. The following result shows that r_k tends to 1 as $k \to \infty$ in Cases (c) and (d).

Lemma 7.4. *If $\lim_{k\to\infty} K_k = K = 0$, then $\lim_{k\to\infty} r_k = 1$.* □

Proof: For any $C \in (0,\infty)$, we may estimate

$$N_k \ge \frac{c_k}{c_k(1+K_kC)+d_k}\,\mathbb{E}_{\mathcal{L}_\rho}\left[\rho\,1_{\{\rho\le C\}}\right], \qquad D_k \le \frac{c_k}{c_k+d_k}. \tag{7.12}$$

Since $\lim_{k\to\infty} K_k = 0$, we have $\lim_{k\to\infty}(c_k+d_k)/[c_k(1+K_kC)+d_k] = 1$, and hence

$$\liminf_{k\to\infty} \frac{N_k}{D_k} \ge \mathbb{E}_{\mathcal{L}_\rho}\left[\rho\,1_{\{\rho\le C\}}\right]. \tag{7.13}$$

Now let $C \to \infty$ and use that $\lim_{C\to\infty}\mathbb{E}_{\mathcal{L}_\rho}[\rho\,1_{\{\rho\le C\}}] = \mathbb{E}_{\mathcal{L}_\rho}[\rho] = 1$ by monotone convergence. □

Lemma 7.3 implies that the scaling of d_k is the same as the scaling of d_k^1 after μ_k is replaced by $\mu_k r_k$. But the latter scaling was derived in Greven et al. (2014), and a glance at the results for Cases (c) and (d) obtained there shows that the scaling is unaffected by the extra factor r_k because of Lemma 7.4. □

A technical remark is in order, for which we refer the reader to Greven et al. (2014, Section 11.3). We have assumed that $k \mapsto \mu_k$ is regularly varying at infinity (recall (3.33)). Because $\lim_{k\to\infty} r_k = 1$, also $k \mapsto r_k\mu_k$ is regularly varying at infinity. Therefore, $(r_k\mu_k)_{k\in\mathbb{N}_0}$ can be approximated from above and from below by sequences that have the same scaling behaviour but are *smoothly varying*, i.e., for all $n \in \mathbb{N}$ their n-th order discrete differences are regularly varying as well. This approximation is harmless because the maps $\underline{c} \mapsto \underline{d}$ and $\underline{\mu} \mapsto \underline{d}$ are component-wise non-decreasing (a fact that is immediate from (3.17)), and so the approximating sequences provide a *sandwich* for the scaling. Now, if the tail exponent of $r_k\mu_k$ is non-integer, i.e., $b \notin \mathbb{N}$ in (3.33), then for all $n \in \mathbb{N}$ the n-th order discrete differences are *asymptotically monotone*. This observation is important because it implies that certain sequences arising in Greven et al. (2014, Section 11.3) have *summable variation*, a property that is crucial for the proof of the scaling. If the tail exponent is integer, i.e., $b \in \mathbb{N}$ in (3.33), then the asymptotic monotonicity still holds for all $n \le b$, which turns out to be enough for the argument.

The extra regularity conditions on L_c, L_μ in (3.33), which are stated in Greven et al. (2014, Eqs. (1.79)–(1.81)), need no modification: $(r_k\mu_k)_{k\in\mathbb{N}_0}$ has the same slowly varying function L_μ as $(\mu_k)_{k\in\mathbb{N}_0}$.

7.3. *Scaling of the volatility: exponential coefficients.*

Proof of Theorem 3.17: We look at each of the five parameter regimes (= universality classes) separately. Recall (3.37–3.38).
(A). Use (7.2). We are in the regime where $\lim_{k\to\infty} K_k = K = \infty$ and $\lim_{k\to\infty} R_k = c$. The same argument as in the proof of Case (a) yields

$$\lim_{k\to\infty} d_k/c_k = \lim_{k\to\infty} \mathfrak{V}_k/R_k = 1/c. \tag{7.14}$$

(B). Let $\bar{K}_k = \bar{\mu}_k/\bar{c}_{k-1}$ and $\bar{R}_k = \bar{c}_k/\bar{c}_{k-1}$. Then $K_k = c\bar{K}_k$ and $R_k = c\bar{R}_k$ by (3.37), and so (7.2) becomes

$$\mathfrak{V}_{k+1} = \bar{g}_k(\mathfrak{V}_k) \quad \text{with} \quad \bar{g}_k(x) = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{(c\bar{K}_k\rho + x)}{c\bar{R}_k + (c\bar{K}_k\rho + x)}\right]. \tag{7.15}$$

We are in the regime where $\lim_{k\to\infty} \bar{K}_k = \bar{K} \in (0,\infty)$ and $\lim_{k\to\infty} \bar{R}_k = \bar{R} = 1$. The same argument as in Case (b) therefore yields $\lim_{k\to\infty} d_k/c_k = \lim_{k\to\infty} \mathfrak{V}_k/R_k = \bar{M}/c\bar{R} = \bar{M}/c$ with $\bar{M}$ the unique attracting fixed point of

$$\bar{g}(x) = \mathbb{E}_{\mathcal{L}_\rho}\left[\frac{(c\bar{K}\rho + x)}{c + (c\bar{K}\rho + x)}\right], \tag{7.16}$$

which is the analogue of (7.4).
(C1). This case is the same as Case (B), but with $\bar{K} = 0$. The analogue of (7.16) reads $\bar{g}(x) = x/(c+x)$. Since $\bar{g}$ has $\bar{M} = 1 - c \in (0,1)$ as unique attracting fixed point, we can copy the proof of Case (b) to get $\lim_{k\to\infty} d_k/c_k = \lim_{k\to\infty} \mathfrak{V}_k/R_k = (1-c)/c\bar{R} = (1-c)/c$. *Note*: In the proof of Case (b) we used that $g(0) > 0$, which fails here. However, even when $d_0 = 0$, the iterates d_k, $k \in \mathbb{N}$, are bounded away from 0 because the attracting fixed points of f_k, $k \in \mathbb{N}$, are bounded away from 0. Hence we may restrict the entire argument to $[\epsilon, 1]$ for some $\epsilon > 0$ instead of $[0,1]$, and use that $g(\epsilon) > 0$ (recall Fig. 7.5).
(C2). This case is like Case (c). Since $\bar{K} = 0$, we can copy the proof of Case (c) and show that the same scaling holds as in the average environment.
(C3). This case is like Case (d). Since $\bar{K} = 0$, we can copy the proof of Case (d) and show that the same scaling holds as in the average environment. □

8. Identification of the universality classes of cluster formation

In this section, we prove Theorem 3.19. In Section 8.1, we deal with cases (a), (A) and (b), (B), (C1), in Section 8.2 with cases (c), (C2) and (d), (C3). The strategy of proof is the same as for the homogeneous environment, except at a few points where the random environment comes into play seriously. We focus on the necessary modifications. Like Section 6, this section is *not completely autonomous*, and for an understanding of the fine details the reader must check the relevant passages in Greven et al. (2014).

Before we begin, we recall why we may choose the starting configuration to be identically equal to θ, the mean of the starting configuration. The initial state and the environment of our Cannings process are such (recall Theorem 3.6) that the scaling limit in (3.22) yields on average θ on level $j+1$.

8.1. *Random cluster size.*

Proof of cases (b), (B), (C1), (C3)[first subcase]: In Step 1, we give the proof for an i.i.d. random environment. In Step 2, we extend the proof to a stationary and ergodic random environment.
Step 1. We consider the set $\mathcal{M}_f([0,1]) \times \mathcal{P}(E)$, describing the environment and the state of a block. If the random environment is i.i.d., then the sequence

$$\left(\chi^{(\eta,j+1-\alpha)}(\omega), M^{(j)}_{-(j+1-\alpha)}\right)_{\alpha\in\mathbb{N}_0} \tag{8.1}$$

is a time-inhomogeneous Markov chain. Let $(K^{*,(j)}_\alpha)_{\alpha\in\mathbb{N}_0}$ be its sequence of transition kernels. (We suppress the index η from $M^{(j)}_{\eta,-(j+1-\alpha)}$ because its law is the same for all $\eta \in \Omega_N$.) It suffices to prove three properties:

(1) The sequence of transition kernels $(K^{*,(j)}_\alpha)_{\alpha\in\mathbb{N}_0}$ converges as $j \to \infty$ to the sequence $(K^{*,\infty}_\alpha)_{\alpha\in\mathbb{N}_0}$ of transition kernels given by

$$K^{*,\infty}_\alpha((\chi,\theta),\cdot) = \mathcal{L}\left[\chi^\alpha \otimes \nu_\theta^{1,\widetilde{M},2\widetilde{K}\chi^\alpha(\omega)}\right](\cdot), \tag{8.2}$$

where $\widetilde{M}, \widetilde{K}$ are defined in (3.47) and $(\chi_\alpha(\omega))_{\alpha\in\mathbb{N}_0}$ in (3.48).

(2) The map

$$((0,\infty)\times(0,\infty)\times\mathcal{P}([0,1])) \ni (c,d,\Lambda) \mapsto \nu_\theta^{c,d,\Lambda} \in \mathcal{P}(\mathcal{P}(E)) \tag{8.3}$$

is continuous.

(3) The map

$$\mathcal{P}(E) \ni \theta \mapsto \nu_\theta^{c,d,\Lambda} \in \mathcal{P}(\mathcal{P}(E)) \tag{8.4}$$

is continuous.

Items (1) and (3) imply the convergence of the process in (8.1), while item (2) is needed in the proof item (1).

Proof of (1)–(3)*:* Here a key is the *duality relation* for the McKean-Vlasov limit process. This duality arises as a special case of our duality relation by choosing a suitable geographic space. This coalescent is obtained by taking as space $\{0,*\}$, where the rates for all transitions in $*$ are zero (cemetery) state and jumps occur from 0 to $*$ at rate c. Kingman coalescence occurs at rate d and the Λ-coalescence is given via Λ (all as long as we are in 0). For a detailed discussion, see Greven et al. (2014, Section 4).

With the help of duality we can identify the equilibrium measure $\nu_\theta^{c,d,K\chi}$ by using a measure-determining sequence of test functions. The parameters c, d, χ enter via the rate of jump to the cemetery state (parameter c), the rate of pairwise coalescence (parameter d), and the rate of coalescence (parameter χ). In the latter, the ratio $\chi/\chi((0,1])$ determines the probability for partition elements to coalesce in groups (Λ-coalescence). In this equilibrium representation, the coalescent has run for *infinite* time.

(1) With $\mathcal{L}$ acting on $\chi_{j+1-\alpha}(\omega)$, we have

$$K^{*,(j)}_\alpha((\chi,\theta),\cdot) = \mathcal{L}\left[(\chi^{j+1-\alpha}) \otimes \nu_\theta^{1,d_{j+1-\alpha}/c_{j+1-\alpha},2K_{j+1-\alpha}\chi^{j+1-\alpha}(\omega)}\right](\cdot). \tag{8.5}$$

From Theorems (3.15) and (3.17), we know that $d_{j+1-\alpha}/c_{j+1-\alpha}$ and $K_{j+1-\alpha}$ converge to $\widetilde{M}$ and $\widetilde{K}$ as $j \to \infty$. The point is to show for every $\alpha \in \mathbb{N}_0$ the equilibrium

measure in the right-hand side converges as $j \to \infty$. By the stationarity of the random environment, the law of $\chi_{j+1-\alpha}(\omega)$ is independent of j. Hence (2) and (3) yield the claim.

(2) The continuity in (8.3) can be deduced from the dual representation in the McKean-Vlasov limit dynamic, in particular, from the fact that the coalescent has run for infinite time, and depends continuously on the migration rate c and the Kingman coalescence rate d, respectively, the rates for the Λ-coalescence. The coalescent has a monotone decreasing number of partition elements off the cemetery where all rates are zero and reaches the cemetery state after a finite time. This means we have a Markov chain hitting a trap in finite time and therefore depends continuously on the finitely many involved jump-rates.

(3) The continuity in (8.4) is deduced from the dual representation. We have to show that the dual expectation depends continuously on θ, which goes as follows. First note that the monomials $\{\langle \cdot, f\rangle^\ell \colon\ f \in C_b(E, \mathbb{R}),\ \ell \in \mathbb{N}\}$ are measure-determining on $(E, \mathcal{B})$. The dual expectation is a finite sum over terms arising from partition elements that are coalescing before jumping to the cemetery state. If ℓ partition elements remain, then the θ-dependence is via $\langle \theta, f\rangle^\ell$, which is a continuous function of θ. □

Step 2. To deal with a stationary and ergodic random environment, we *condition* on the sequence $(\chi_\alpha)_{\alpha\in\mathbb{N}_0}$. This leads to a sequence of Markov chains in random environment, indexed by j, for which the result in (1) holds, as explained above. After that we argue that (1)–(3) again imply the claim, because of the stationarity and the fact that we need only consider finite α.

Next, we consider the finite-dimensional laws of the Markov chain in random environment conditional on $(\chi_\alpha)_{\alpha\in\mathbb{N}_0}$ and we verify the appropriate versions of (1)–(3). To this end, we extend the duality to a *space-time duality* and obtain an expression for the *mixed space-time moments* in terms of triples of parameters

$$(c_k, d_k, \chi_k)_{k=j+1,j,\dots,j+1-L}, \tag{8.6}$$

with L being the order of the marginal distribution we consider.

In the *space-time dual*, we work with frozen partition elements which are activated (then once and forever) at a present time. Namely, we add partition elements marked by a label in $[0, \infty]$, which indicates from which time on the mechanisms of the coalescent are activated. Before this time, the partition element neither moves nor coalesces. This allows us to characterize the finite-dimensional marginals of the forward process. Suppose that we want to study the finite-dimensional distributions associated with times $0 \le t_1 < t_2 < t_3 \dots < t_n < t$. Then, we take individuals marked with $0, t - t_n, t - t_{n-1}, \dots, t - t_1$, consider the test functions in the duality relations for the time horizon $t_1, t_2, \dots, t_n, t$, and form the product. The duality relation holds again. Compare with Greven et al. (2016, Corollary 1.20).

In this setting, (1)–(3) turn into claims about the expectation of the duality expression under the law of the space-time coalescent, after which the argument proceeds as above. □

Proof of cases (a), (A): The limiting transition kernel of the rescaled interaction chain for a given environment degenerates to a transition kernel concentrated on the traps. We have

$$K = \infty, \qquad \lim_{k\to\infty} d_k/c_k = 1. \tag{8.7}$$

We must therefore show that

$$\lim_{K\to\infty} \mathcal{L}\left[\nu_\theta^{1,1,2K\chi}\right] = \mathcal{L}\left[\int_E \theta(\mathrm{d}u)\delta_{\delta_u}\right]. \tag{8.8}$$

Taking the dual representation, we see that as $K \to \infty$ the rate of the Λ-coalescence tends to infinity, implying that the coalescent converges before it jumps, and coalesces into a single partition element. The duality relation says that the original McKean-Vlasov process is in a mono-type equilibrium, where the type is chosen at random according to θ. The claim now follows because for $K(\theta,\cdot) = \int_E \theta(\mathrm{d}u)\delta_{\delta_u}(\cdot)$ the state δ_u is a *trap*, so that the limiting Markov chain is constant for every $\alpha \neq 0$, the constant being chosen according to θ for every realization of the random environment. □

8.2. *Random cluster order.*

Proof of cases (c), (C2) and (d), (C3)[second subcase]: In cases (c) and (d), averaging takes place via a law of large numbers and the situation is similar to the homogeneous environment, for which the results in Theorem 3.19 are of the same type, and it is only the formula for d that changes.

The claim is that the interaction chain, which is a space-time rescaled Markov chain and a measure-valued square-integrable martingale, converges to a limit that is a measure-valued diffusion and a square-integrable martingale. In Dawson et al. (1995, Section 6(b)), it was pointed out how, for the case of the Fleming-Viot process, this convergence reduces to the study of the process of conditioned variances along the path, which in turn reduces to showing the following asymptotic relations for these objects. Pick $\alpha_1, \alpha_2 \in I$ with $\alpha_2 < \alpha_1$, and suppose that $\lim_{j\to\infty} k_\alpha(j)/j = \beta(\alpha)$ with $0 \le \beta(\alpha_1) < \beta(\alpha_2) \le 1$. If the scaled Markov chain is such that

$$\lim_{j\to\infty} \mathrm{Var}\left(\left\langle f, M^{(j)}_{-k_{\alpha_1}(j)}\right\rangle \;\middle|\; M^{(j)}_{-k_{\alpha_2}(j)} = \theta\right) = \frac{\beta(\alpha_2)-\beta(\alpha_1)}{\beta(\alpha_2)}\,\mathrm{Var}_\theta(f), \qquad f \in C_b(E,\mathbb{R}), \tag{8.9}$$

with $\beta(\alpha) = 1-\alpha$, then by applying the transformation $\beta(\alpha) = \mathrm{e}^{-s}$ the right-hand side turns into the expression $(1-e^{-(s_1-s_2)})\,\mathrm{Var}_\theta(f)$. Since this scales like $(s_1-s_2)\,\mathrm{Var}_\theta(f)$ for $s_1 \downarrow s_2$, we see that the standard Fleming-Viot process $Y(s)_{s\ge 0}$ appears as the scaling limit. Since $s = \log(1/(1-\alpha))$, we get the time-scaled Fleming-Viot process $Y(\log(1/(1-\alpha))_{\alpha\in[0,1)}$ (see Dawson et al., 1995, Section 6).

With suitable time transformations, we can also handle the other forms of scaling $j \to k_\alpha(j)$ in Definition 3.18. Namely, we have to identify the function $F(\alpha_1,\alpha_2)$ appearing in front of $\mathrm{Var}_\theta(f)$ and find the transformation $\alpha = L(s)$ such that

$$\Delta F(s_2) = \lim_{s_1\downarrow s_2} \frac{F(L(s_1),L(s_2))}{s_1-s_2} \equiv 1, \tag{8.10}$$

so that again the standard Fleming-Viot process $(Y(s))_{s\ge 0}$ appears as the scaling limit. Since $s = L^{-1}(\alpha)$, we get the time-scaled Fleming-Viot process $(Y(L^{-1}(\alpha))_{\alpha\in I}$.

It was pointed out in Greven et al. (2014, Section 9.3) how (8.9) is established for the homogeneous hierarchical Cannings process by using the scaling analysis of the coefficients $\underline{d} = (d_k)_{k\in\mathbb{N}_0}$. In our case, we need to work with a random

sequence $(\mu_k\rho_k(\omega))_{k\in\mathbb{N}_0}$ instead of $(\mu_k)_{k\in\mathbb{N}_0}$, where $\rho_k(\omega)$ arises from the term $\Lambda = \Lambda^{(\eta,k)}((0,1])(\omega)$ in the following variance formula

$$\int_E \mu_\theta^{c,d,\Lambda}(\mathrm{d}x)\big(\langle f,x\rangle^2 - \langle f,\theta\rangle^2\big) = \frac{2c}{2c+\lambda\rho_k(\omega)+2d}\ \mathrm{Var}_\theta(\langle f,x\rangle), \quad f\in C_b(E,\mathbb{R}). \tag{8.11}$$

We thus have to see whether the product (with $\rho_k(\omega) = \rho^{\mathrm{MC}_k(0)}(\omega)$)

$$\prod_{k=j_1}^{j_2} \frac{2c_k}{2c_k+\lambda_k\rho_k(\omega)+2d_k}, \tag{8.12}$$

appearing in the expression for the variance in (8.9), does indeed exhibit averaging based on the tail triviality of the random sequence $(\rho_k(\omega))_{k\in\mathbb{N}_0}$ (see Greven et al., 2014, Eq. (8.14)).

To that end, we abbreviate

$$m_k(\omega) = \frac{\mu_k\rho_k(\omega)+d_k}{c_k}, \tag{8.13}$$

consider the relation

$$\mathrm{Var}\left(\left\langle M_{j_1}^{(j_2)}, f\right\rangle \mid M_{j_2+1}^{(j_2)} = \theta\right) = \left[\sum_{k=j_1}^{j_2}\frac{d_{k+1}}{c_k}\prod_{l=k+1}^{j_2}\frac{1}{1+m_l(\omega)}\right]\mathrm{Var}(\langle\theta,f\rangle) \tag{8.14}$$

and analyse its behaviour as $j\to\infty$ for appropriate choices of $j_1 = j_1(j)$ and $j_2 = j_2(j)$. We must show that, for $\mathbb{P}$ almost all ω, (8.14) behave asymptotically like the right-hand side of (8.9), and we must identify the associated F, ΔF and L.

In order to decide how the product scales as $j_2 - j_1\to\infty$, we take logarithms to turn this into the question whether the sum

$$\sum_{k=j_1}^{j_2}\frac{\mu_k\rho_k(\omega)+d_k}{c_k} = \sum_{k=j_1}^{j_2} m_k(\omega) \tag{8.15}$$

has a certain scaling behaviour, and we link this to the scaling behaviour of μ_k/c_k and d_k/c_k for $k\to\infty$ (which we know from Theorems 3.15 and 3.17) to derive the relevant asymptotics. We have to show that this asymptotics does *not* depend on ω and is equal to that with $\rho_k(\omega)$ replaced by its mean 1. To achieve the latter, we use the stationarity of $(\rho_k(\omega))_{k\in\mathbb{N}_0}$, plus the fact that it has bounded and decaying covariances (recall (2.41)–(2.42)). The key is the following lemma.

Lemma 8.1. *Define* $S(j_1,j_2)(\omega) = \sum_{k=j_1}^{j_2} m_k(\omega)$. *Then,*

$$\lim_{j_2-j_1\to\infty}\frac{S(j_1,j_2)(\omega)}{\mathbb{E}[S(j_1,j_2)(\omega)]} = 1 \quad \textit{in } \mathbb{P}\textit{-probability.} \tag{8.16}$$

Proof: Define

$$\chi_k(j_1,j_2) = \frac{\mu_k/c_k}{\sum_{k=j_1}^{j_2}(\mu_k+d_k)/c_k}, \qquad j_1\le k\le j_2. \tag{8.17}$$

Then

$$\frac{S(j_1,j_2)(\omega)}{\mathbb{E}[S(j_1,j_2)(\omega)]} - 1 = \sum_{k=j_1}^{j_2}\chi_k(j_1,j_2)[\rho_k(\omega)-1]. \tag{8.18}$$

With the help of Chebyshev's inequality we see that it suffices to show that

$$\lim_{j\to\infty} \sum_{k=j_1}^{j_2} \sum_{l=j_1}^{j_2} \chi_k(j_1,j_2)\chi_l(j_1,j_2)\,\mathbb{C}\mathrm{ov}[\rho_k(\omega),\rho_l(\omega)] = 0. \tag{8.19}$$

We have $\mathbb{C}\mathrm{ov}[\rho_k(\omega),\rho_l(\omega)] = C_{|k-l|}$ with $\lim_{m\to\infty} C_m = 0$. Since, by our assumptions on $(c_k)_{k\in\mathbb{N}_0}$ and $(\mu_k)_{k\in\mathbb{N}_0}$, we have

$$\lim_{j_2-j_1\to\infty} \max_{j_1\le k\le j_2} \chi_k(j_1,j_2) = 0, \qquad \sum_{j_1\le k\le j_2} \chi_k(j_1,j_2) \le 1, \tag{8.20}$$

the claim follows. □

Remark 8.2. The role of Lemma 8.1 is to show that the same clustering behaviour occurs in the random environment as in the homogeneous environment. We are only able to prove convergence in $\mathbb{P}$-probability and not $\mathbb{P}$-a.s. In the prefactor in the right-hand side of (8.14) weighted averages over j_1, j_2-dependent sliding windows of the random environment appear, which would need to be shown to converge $\mathbb{P}$-a.s. It is unclear how to do this, even for an i.i.d. random environment.

Lemma 8.1 implies that the term between square brackets in (8.14) scales like

$$\Delta(j_1,j_2) = \sum_{k=j_1}^{j_2} \frac{d_{k+1}}{c_k} \exp\left[-\sum_{l=k+1}^{j_2} \frac{\mu_l+d_l}{c_l}\right] \qquad \text{in } \mathbb{P}\text{-probability as } j_2-j_1\to\infty, \tag{8.21}$$

where we use that $\lim_{l\to\infty}(\mu_l+d_l)/c_l = 0$ in all cases of interest. In the remainder of the proof, we pick $j_1 = k_{\alpha_1}(j)$ and $j_2 = k_{\alpha_2}(j)$ with $\alpha_2 < \alpha_1$, with $k_\alpha(j)$ as in Definition 3.18, and compute the limit of (8.21) as $j\to\infty$. We omit writing $\lfloor\cdot\rfloor$ at places where labels are obviously integer. We determine k_α and identify F, L (recall the discussion leading up to (8.10)) for the different cases, in the order (c), (C2), (d), (C3). Recall that $K_k = \frac{\mu_k}{c_k}$ and $\bar K_k = \frac{\bar\mu_k}{\bar c_k}$.

Case (c). Pick $k_\alpha(j) = j+1-\alpha h(j)$ with $h(j) = 1/\sqrt{K_j}$, and insert $d_k \sim \sqrt{c_k\mu_k} = c_k\sqrt{K_k}$ and $d_{k+1}\sim d_k$, to obtain that (8.21) scales like

$$\Delta(j) = \sum_{k=j+1-\alpha_1/\sqrt{K_j}}^{j+1-\alpha_2/\sqrt{K_j}} \sqrt{K_k}\ \exp\left[-\sum_{l=k+1}^{j+1-\alpha_2/\sqrt{K_j}} \left(K_l+\sqrt{K_l}\right)\right]. \tag{8.22}$$

Putting $x = (j+1-k)\sqrt{K_j}$, and using that $\lim_{k\to\infty} K_k = 0$, $\lim_{k\to\infty} k^2K_k = \infty$ and $K_k\sim K_l\sim K_j$ uniformly in k,l in both sums, we get

$$\lim_{j\to\infty}\Delta(j) = \int_{\alpha_2}^{\alpha_1} \mathrm{d}x\ \exp[-(x-\alpha_2)] = 1-\exp[-(\alpha_1-\alpha_2)]. \tag{8.23}$$

Pick $\alpha = L(s) = s$. Then $\Delta F\equiv 1$. Since $s = L^{-1}(\alpha) = \alpha$, this proves the claim.

Case (C2)[subcase $\lim_{k\to\infty} k\bar K_k = \infty$]. Pick $k_\alpha(j) = j+1-\alpha h(j)$ with $h(j) = 1/\bar K_j$, and insert $d_k\sim \mu_k/(\mu-1) = \bar K_k c_k/(\mu-1)$ and $d_{k+1}\sim \mu d_k$, to obtain that (8.21) scales like

$$\Delta(j) = \frac{\mu}{\mu-1} \sum_{k=j+1-\alpha_1/\bar K_j}^{j+1-\alpha_2/\bar K_j} \bar K_k\ \exp\left[-\frac{\mu}{\mu-1}\sum_{l=k+1}^{j+1-\alpha_2/\bar K_j} \bar K_l\right]. \tag{8.24}$$

Putting $x = (j+1-k)\bar{K}_j$, and using that $\lim_{k\to\infty} \bar{K}_k = 0$, $\lim_{k\to\infty} k\bar{K}_k = \infty$ and $\bar{K}_k \sim \bar{K}_l \sim \bar{K}_j$ uniformly in k, l in both sums, we get

$$\lim_{j\to\infty} \Delta(j) = \frac{\mu}{\mu-1} \int_{\alpha_2}^{\alpha_1} \mathrm{d}x \ \exp\left[-\frac{\mu}{\mu-1}(x-\alpha_2)\right] = 1 - \exp\left[-\frac{\mu}{\mu-1}(\alpha_1-\alpha_2)\right]. \tag{8.25}$$

Pick $\alpha = L(s) = \frac{\mu-1}{\mu} s$. Then $\Delta F \equiv 1$. Since $s = L^{-1}(\alpha) = \frac{\mu}{\mu-1}\alpha$, this proves the claim.

Case (d). Pick $k_\alpha(j) = (1-\alpha)(j+1)$, and insert $d_k \sim M/\sigma_k$, $\sigma_k c_k \sim k/(1-a)$ and $d_{k+1} \sim d_k$, to obtain that (8.21) scales like

$$\Delta(j) = M(1-a) \sum_{k=(1-\alpha_1)(j+1)}^{(1-\alpha_2)(j+1)} \frac{1}{k} \ \exp\left[-\sum_{l=k+1}^{(1-\alpha_2)(j+1)} \left(K_l + \frac{M(1-a)}{l}\right)\right]. \tag{8.26}$$

Putting $x = (j+1-k)/(j+1)$, and using that $\lim_{k\to\infty} k^2 K_k = 0$, we get

$$\begin{aligned} \lim_{j\to\infty} \Delta(j) &= M(1-a) \int_{\alpha_2}^{\alpha_1} \frac{\mathrm{d}x}{1-x} \ \exp\left[-M(1-a)\int_{\alpha_2}^{x} \frac{\mathrm{d}y}{1-y}\right] \\ &= M(1-a)\,(1-\alpha_2)^{-M(1-a)} \int_{\alpha_2}^{\alpha_1} \mathrm{d}x\,(1-x)^{-1+M(1-a)} \\ &= 1 - \left(\frac{1-\alpha_1}{1-\alpha_2}\right)^{M(1-a)}. \end{aligned} \tag{8.27}$$

Pick $\alpha = L(s) = 1 - e^{-s/R}$ with $R = M(1-a)$. Then $\Delta F \equiv 1$. Since $s = L^{-1}(\alpha) = \log(1/(1-\alpha)^R)$ we get the claim.

Case (C2)[subcase $\lim_{k\to\infty} k\bar{K}_k = \bar{N}$]. This is the same as case (d) with $M(1-a)$ replaced by $\bar{N}\frac{\mu}{\mu-1}$.

Case (C3)[second subcase]. This is the same as case (d) with M replaced by 1. □

Acknowledgements

The authors are grateful to Evgeny Verbitskiy for help with the renormalization analysis, and to the anonymous referees for critical remarks.

References

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, 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.

J. Blath, A. Etheridge and M. Meredith. Coexistence in locally regulated competing populations and survival of branching annihilating random walk. *Ann. Appl. Probab.* **17** (5-6), 1474–1507 (2007). MR2358631.

A. Bovier and F. den Hollander. *Metastability*, volume 351 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham (2015). ISBN 978-3-319-24775-5; 978-3-319-24777-9. MR3445787.

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 and D. Griffeath. Diffusive clustering in the two-dimensional voter model. *Ann. Probab.* **14** (2), 347–370 (1986). MR832014.

J. T. Cox and A. Klenke. Recurrence and ergodicity of interacting particle systems. *Probab. Theory Related Fields* **116** (2), 239–255 (2000). MR1743771.

D. Dawson and A. Greven. Multiple time scale analysis of interacting diffusions. *Probab. Theory Related Fields* **95** (4), 467–508 (1993). 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. Multiple space-time scale analysis for interacting branching models. *Electron. J. Probab.* **1**, no. 14, approx. 84 pp. (1996). MR1423467.

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.

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.

S. N. Evans and K. Fleischmann. Cluster formation in a stepping-stone model with continuous, hierarchically structured sites. *Ann. Probab.* **24** (4), 1926–1952 (1996). MR1415234.

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.

A. Greven, F. den Hollander, S. Kliem and A. Klimovsky. Renormalisation of hierarchically interacting Cannings processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **11** (1), 43–140 (2014). MR3194010.

A. Greven, R. Sun and A. Winter. Continuum space limit of the genealogies of interacting Fleming-Viot processes on $\mathbb{Z}$. *Electron. J. Probab.* **21**, Paper No. 58, 64 (2016). MR3563886.

V. Limic and A. Sturm. The spatial Λ-coalescent. *Electron. J. Probab.* **11**, no. 15, 363–393 (2006). MR2223040.

J. Pitman. Coalescents with multiple collisions. *Ann. Probab.* **27** (4), 1870–1902 (1999). MR1742892.

S. Sawyer and J. Felsenstein. Isolation by distance in a hierarchically clustered population. *J. Appl. Probab.* **20** (1), 1–10 (1983). MR688075.