DOI: 10.1214/EJP.v10-290

Electronic Journal of Probability

Vol. 10 (2005), Paper no. 39, pages 1286-1358.

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

REPRESENTATION THEOREMS FOR INTERACTING MORAN MODELS, INTERACTING FISHER–WRIGHT DIFFUSIONS AND APPLICATIONS

ANDREAS GREVEN, VLADA LIMIC, AND ANITA WINTER

Abstract. We consider spatially interacting Moran models and their diffusion limit which are interacting Fisher-Wright diffusions. The Moran model is a spatial population model with individuals of different type located on sites given by elements of an Abelian group. The dynamics of the system consists of independent migration of individuals between the sites and a resampling mechanism at each site, i.e., pairs of individuals are replaced by new pairs where each newcomer takes the type of a randomly chosen individual from the parent pair. Interacting Fisher-Wright diffusions collect the relative frequency of a subset of types evaluated for the separate sites in the limit of infinitely many individuals per site. One is interested in the type configuration as well as the time-space evolution of genealogies, encoded in the so-called historical process. The first goal of the paper is the analytical characterization of the historical processes for both models as solutions of well-posed martingale problems and the development of a corresponding duality theory.

For that purpose, we link both the historical Fisher-Wright diffusions and the historical Moran models by the so-called look-down process. That is, for any fixed time, a collection of historical Moran models with increasing particle intensity and a particle representation for the limiting historical interacting Fisher-Wright diffusions are provided on one and the same probability space. This leads to a strong form of duality between spatially interacting Moran models, interacting Fisher-Wright diffusions on the one hand and coalescing random walks on the other hand, which extends the classical weak form of moment duality for interacting Fisher-Wright diffusions.

Our second goal is to show that this representation can be used to obtain new results on the long-time behavior, in particular (i) on the structure of the equilibria, and of the equilibrium historical processes, and (ii) on the behavior of our models on large but finite site space in comparison with our models on infinite site space. Here the so-called finite system scheme is established for spatially interacting Moran models which implies via the look-down representation also the already known results for interacting Fisher-Wright diffusions. Furthermore suitable versions of the finite system scheme on the level of historical processes are newly developed and verified.

In the long run the provided look-down representation is intended to answer questions about finer path properties of interacting Fisher-Wright diffusions.

Submitted to EJP on February 13, 2004. Final version accepted on March 17, 2005.

2000 *Mathematics Subject Classification.* Primary: 60K35; Secondary: 60G09, 92D25.

Key words and phrases. Spatially interacting Moran model, Interacting Fischer-Wright diffusions, historical process, historical martingale problem, look-down construction, exchangeability, equilibrium measure, large finite systems.

Contents

1. Introduction	1288
1.1. Motivation and background	1288
1.2. The models	1290
1.3. Historical processes	1292
1.4. The historical look-down process and the enriched historical IMM	1294
2. Main results I: Analytical characterizations	1297
2.1. The martingale problem characterization (Theorems 1 and 2)	1297
2.2. The particle representation (Theorem 3)	1301
3. Main results II: Applications	1303
3.1. Representation of genealogies via coalescents, strong duality	1303
3.2. Application I: Equilibria of IMM and IFWD (Theorems 4 and 5)	1306
3.3. Application II: Finite system scheme (Theorems 6 and 7)	1310
4. Wellposedness of the historical martingale problems	1317
4.1. The graphical construction of the enriched historical processes	1317
4.2. Proof of Theorems 1 and 2	1319
5. Particle representation of the historical processes (Proof of Theorem 3)	1327
6. Strong duality and the coalescent	1332
6.1. Construction of the coalescent (Proof of Proposition 3.4)	1332
6.2. Proofs of duality relations (including Proposition 3.5 and Corollary 3.7)	1335
6.3. Representations for the historical look-down (Proof of Theorem 4)	1336
7. The finite system scheme for historical IMM	1337
7.1. Coupled representations for the look-down genealogies	1337
7.2. Preliminaries: Asymptotic analysis of the coalescent	1338
7.3. Preliminaries: Asymptotic analysis of the historical coalescent	1347
7.4. Asymptotic analysis of the look-down process (Proof of Theorem 6)	1348
8. Transfer to IFWD (Proofs of Theorems 5 and 7)	1351
References	1355
Index	1357

1. Introduction

1.1. **Motivation and background.** In this paper we analyze two related classes of spatial multi-type population models, which we call the *spatially interacting Moran models* and the ensembles of *interacting Fisher-Wright diffusions*. In spatial population models one considers individuals having a genetic type and a location in geographic space where the number of particles and the genetic decomposition at any particular location change according to random dynamics.

The spatially interacting Moran models (IMM) on an Abelian group (with group elements representing possible locations in geographic space) can be described as follows: the group consists of finitely or countable many sites. Each site is populated by an a.s. finite number of individuals, which carry one of the possible genetic types. Individuals *migrate* between the sites independently of each other according to a continuous time random walk (RW). Furthermore, at a given site each pair of individuals is replaced according to a *resampling* mechanism by a new pair of individuals with types chosen independently at random from one of the parents.

For many purposes sufficient information about the population in the IMM is given by the functional of the proportions of a supported subset of the type space evaluated for the separate sites (see, for example, Shiga (1980) [43]). In the large local population limit of this functional an ensemble of *interacting Fisher–Wright diffusions* (IFWD) arises, that interact due to a drift of each single component to an average over the neighbors. The average is given by the jump rates of a random walk modeling migration in the finite or countable site-space.

The basic ergodic theory (on the long-term behavior) for IMM and IFWD is the same as that for a wide class of interacting spatial processes with components indexed by $\mathbb{Z}^d$, by $\mathbb{R}^d$, or by the hierarchical group. The class of processes with such long-term behavior includes on the one hand interacting particle models, for example the voter model (Holley and Liggett (1975) [34]), branching random walk (Kallenberg (1977) [35], Durrett (1979) [23]), branching Brownian motion (Fleischman (1978) [26], Gorostiza and Wakolbinger (1991) [29]), and on the other hand, interacting diffusions, for instance the Fisher–Wright stepping stone model (Shiga (1980) [43]), the Ornstein–Uhlenbeck process (Deuschel (1988) [17]), the case of a general diffusion function (Cox and Greven (1994) [9] and Cox, Fleischmann and Greven (1996) [6]) and finally super Brownian motion (Dawson (1977) [12], Etheridge (1993) [24]). It can be stated as a metatheorem that the long-term behavior for the interacting spatial systems depends on whether the underlying (symmetrized) migration kernel is recurrent or transient; hence it differs sharply in high and low dimensions. In high dimensions each process has a one-parameter family of invariant measures indexed by the "density" of the system which is preserved for every finite time t and in the limit as $t \to \infty$. In low dimensions the invariant measures are "degenerate", that is, steady states are concentrated on the traps of the stochastic evolution (e.g., for the two-type IMM these traps are the monotype configurations, for branching models the trap is the zero configuration). In this case one says that the systems *cluster*.

The above described dichotomy and the results for IFWD on the structure of the *equilibria* in the high-dimensional case, and the structure of the *clusters* in the low-dimensional case, were obtained via a moment calculation which is based on a duality to (delayed) coalescing random walks. This duality holds for IMM only under special initial conditions. Nevertheless this duality is more than a clever trick! In fact there is a much richer structure available, namely an embedding in a model where each individual and its historical evolution is defined. The objective to treat IMM and IFWD in a unified way, motivated by a few natural open problems, suggests that indeed we should change our point of view, and embed both processes into these so-called *historical processes* which allows one to define the genealogies corresponding to particles/masses evaluated at time t. This provides a powerful tool whenever these histories viewed backward from a reference time t form again a Markov process. In order to carry out this program we have to first lay some analytical foundations and provide characterizations of these random processes via *martingale problems*.

The idea of taking into account information about the evolution of the individuals in the system up to time t has been explored in the context of branching models (compare Dawson and Perkins (1991) [16], and Duquesne and Le Gall (2002) [22]) and for the voter model by Cox and Geiger (2000) [7]. Strong results on support and regularity properties were derived based on the decomposition into family clusters corresponding to different degrees of relationship. The previous analysis was based on the concept of infinite divisibility and graphical representations. However, exploiting the fact that, due to the *exchangeability* of the individuals, the time-reversed process becomes Markovian allows one to define such a family decomposition also for IMM and IFWD. It turns out that for both processes, IMM and IFWD, the genealogies evolve as a *spatially structured coalescent process*. This gives a much stronger form of duality than just the usual moment duality, resulting in a *historical form of duality* which provides a powerful tool.

Based on our construction we may think of clusters, which build up in the low-dimensional site spaces for $t \to \infty$, as groups of individuals having the same type due to the same ancestor back at time 0, and according to the same principle we can decompose the equilibrium state, arising in high-dimensional cases, into infinitely old family clusters. In order to analyze the historical process in the diffusion limit we use the particle approximations and random embeddings of the historical Moran processes in the historical Fisher-Wright process.

Hence the first goal of the paper is to give an analytical description for the historical processes, which will be via well-posed martingale problems. This allows us to construct a countable particle representation on the level of the historical process of both the approximating particle models and the diffusion limit in such a way that one can be naturally embedded in the other. We then apply this construction to answer questions on the long-time behavior of such systems. For the construction, we adapt the idea of *look-downs* given by Donnelly and Kurtz (1996) [19] to our setting with interacting components.

Our second goal is to study the longtime behavior of the IMM and IFWD processes by applying the strong duality on a historical level. Here we focus mainly on the high dimensional case, and in particular the following two applications: (i) Based on the genealogies we investigate the structure of the equilibrium states by providing a stochastic representation of the lines of descent in an infinitely old population. (ii) We describe the longtime behavior of the historical systems indexed by large but *finite* subsets of the infinite geographic space, and its connection to the equilibrium states of the corresponding system on $\mathbb{Z}^d$, $d \geq 3$. For performing this analysis Cox and Greven (1990) suggested a method, the *finite system scheme*, compare e.g. [8], and [10]. Our approach to IMM and IFWD based on the look-down construction allows us to verify this strong form of the finite system scheme for IMM and IFWD in a unified way and to develop the *historical finite system scheme*. At the same time this allows us to develop a rigorous renormalization analysis of such models including the historical structure. (Compare [30] for a survey on the problem of renormalization and universality).

Before outlining the rest of the paper, we describe some perspectives the new approach has in the the low-dimensional case. Much work on *clustering phenomena* has been centered on thinking of clusters as regions where most of the population – in the sense of averages over huge blocks – are of the same type. In this spirit the large scale correlation structure of IFWD on the hierarchical group is studied in Fleischmann and Greven (1994), (1996) [27] [28]. The results in these papers are proved by a moment calculation relying on the duality of the evolution of these moments to a system of delayed coalescing random walks. Again the look-down construction provides a powerful tool to explain the clustering phenomena for the IMM, and to give a unified approach to both IMM and IFWD yielding stronger convergence results in both cases. This analysis will be done in a forthcoming paper by the authors [31].

Moreover, it turns out that a number of interesting questions concerning the shape and the time structure of a cluster cannot be answered by moments or deduced fairly directly from the look-down construction. These include the following open problems:

- Let the underlying migration be recurrent. What is the behavior of the holding times between two sucessive upcrossings of a component over $[\varepsilon, 1-\varepsilon]$?
- Cox and Griffeath (1986) [11] observed for the voter model on $\mathbb{Z}^2$ that if a block average at time t on a box of side length t^γ is near zero, then a box of side length $t^{\gamma-\varepsilon}$ must have no 1's with overwhelming probability. Does a similar statement hold for interacting Fisher-Wright diffusions as well?

The look-down construction on the level of historical processes should open eventually new possibilities. For that purpose one has to extend the genealogy spanned by the population of a fixed time t to include the information on those individuals which had lived before time t but do not have descendants at time t. This leads to genealogical forests, which can be analysed based on the concept of $\mathbb{R}$-trees. An analytic characterization for this kind of genealogical forests is developed in the forthcoming paper [32] by Greven, Pfaffelhuber and Winter.

Outline We continue by specifying our models rigorously in Subsection 1.2. Then in Subsection 1.3 we construct the historical processes explicitly. Finally, in Subsection 1.4 the historical look-down process and the enriched historical IMM are introduced which will provide a suitable framework to discuss both the particle model and its diffusion limit.

In Sections 2 the historical processes are characterized analytically through well-posed martingale problems and as a consequence a particle representation based on the historical look-down process is given. In Section 3 we will apply the analytical representations to construct in Subsection 3.1 a strong duality which will be used to investigate the genealogical structure of equilibrium states in Subsection 3.2, and to study the behavior of large finite systems from the point of view of historical processes in Subsection 3.3. The proofs of the main results are given in Sections 4 through 8.

1.2. **The models.** We will study two models, a particle model called the spatially interacting Moran models (IMM), and its diffusion limit called the interacting Fisher-Wright diffusions (IFWD), which we define below in (i) and (ii), respectively.

The basic ingredient for our processes is a *random walk* (RW) on a countable Abelian group G: let $(R_n)_{n\in\mathbb{N}}$ be a non-degenerate and irreducible RW in discrete-time with transition probability $a(x,y) = a(0,y-x)$. We denote its n-step transition probabilities by $a^{(n)}(x,y)$, $x,y \in G$. The transition kernel of its continuous time version, $R := (R_t)_{t\geq 0}$, is then given by

$$a_t(x,y) := \sum_{n\geq 0} a^{(n)}(x,y)\frac{t^n e^{-t}}{n!}, \qquad x,y \in G. \tag{1.1}$$

(i) The *spatially interacting Moran model*, ξ, on G is a locally finite population model where individuals of the population are assumed to be of one of the possible genetic types chosen from the type-space, $K \subseteq \mathbb{R}$. It is defined by the following two mechanisms which act independently of each other as follows:

Migration Each individual moves in G independently of the other individuals according to the law of R.

Resampling Each pair of individuals situated at the same site dies at rate γ and is instantaneously replaced by a new pair of individuals, where each new individual adopts a type by choosing the parent independently at random from the "dying pair". All such resampling events are independent of each other. This defines a branching mechanism where the size of the population in a branching step is preserved.

Remark Note that the local population size remains constant during time intervals without a migration step. Therefore, the resampling part of the evolution during intervals without migration steps is the same as in the traditional fixed size population models. □

Following this description we can define the stochastic evolution of all the individuals and their descendants via collections of independent versions of R, of exponential waiting times, and of a sequence of $\{0,1\}$-valued random variables used for choosing parents, provided that the initial state is chosen suitably (we discuss this below). The result generates a *tree-indexed random walk* by associating with each individual the path of its descendants, the edges corresponding to the individuals and the edge length to the life-time.

Remark These rules of evolution have, due to the *exchangeability* of individuals, an important *projection property*. If we start with a set of types K and combine certain types to a new type resulting in a new set K' of fewer types, we obtain the same process by first running the dynamics with types K and then relabeling according to K', as by first relabeling according to K' and then running the dynamics. □

Very often we are only interested in the numbers of particles of a certain type at certain sites. For this purpose note that the individuals generate a random population at time $t > 0$, inducing an integer-valued measure whose restriction on finite sets are atomic, i.e.,

$$\eta_t \in \mathcal{N}(G \times K). \tag{1.2}$$

Of course, in order to construct a countable particle system with always finitely many particles per site, it is in general necessary to impose some growth restrictions on the initial configuration of the process at infinity: fix a strictly positive and finite measure α on G such that for some finite constant Γ we have for all $x \in G$,

$$\sum_{y \in G} a(x,y)\alpha(\{y\}) \le \Gamma\alpha(\{x\}). \tag{1.3}$$

Take for instance a strictly positive and finite measure β on G and set

$$\alpha(\{y\}) := \sum_{n \ge 0} \Gamma^{-n} \sum_{z \in G} a^{(n)}(y,z)\beta(\{z\}). \tag{1.4}$$

Use as possible initial configurations those configurations which give rise to a measure $\eta \in \mathcal{E}_G$ if each individual is associated with a δ-measure on its current position and type, and here $\mathcal{E}_G$ is given by

$$\mathcal{E}_G := \{\eta \in \mathcal{M}(G \times K) : \| \eta \|_{\ell^1_\alpha} := \sum_{x \in G} \eta(\{x\} \times K)\alpha(\{x\}) < \infty\}. \tag{1.5}$$

Note that in the context of particle models, $\mathcal{E}_G$ was first introduced in Liggett and Spitzer (1981) [40], and is therefore referred to as the *Liggett-Spitzer space*.

The stochastic process $\eta = (\eta_t)_{t \ge 0}$ can be described as a Markov process with state space $\mathcal{E}_G$. The well-posedness of the process can be established by coupling techniques (see Leopold (2001) [38]), or alternatively will be a consequence of our results in Subsection 2.1.

It is obvious that the projection of η on the location coordinate simply yields a countable system of independent random walks,

$$\chi := (\chi_t)_{t \ge 0}. \tag{1.6}$$

(ii) Next we derive the second model, namely *interacting Fisher-Wright diffusions* (IFWD) as the high density limit of a functional of the particle model previously discussed. The exact statement is given in Theorem 0 below.

We introduce the candidate for the scaling limit as follows: define the interacting Fisher-Wright diffusion (IFWD), $\zeta = (\zeta_t)_{t \ge 0}$, as the Markov process with the following pregenerator. Let $C_0^2([0,1]^G)$ be the set of twice continuously differentiable functions depending only on finitely many coordinates of $\varphi = (\varphi_x)_{x \in G} \in [0,1]^G$. Then for each $f \in C_0^2([0,1]^G)$ set

$$L_\zeta f = L_\chi f + \frac{\gamma}{2} \sum_{x \in G} (1 - \varphi_x)\varphi_x \frac{\partial^2 f}{\partial \varphi_x^2}, \tag{1.7}$$

where

$$L_\chi f := \sum_{x \in G} (\sum_{y \in G} (\bar{a}(x,y) - \delta(x,y))\varphi_y) \frac{\partial f}{\partial \varphi_x}, \tag{1.8}$$

and $\bar{a}(x,y) := a(y,x)$ denotes the reversed kernel. Note that $C_0^2([0,1]^G)$ is a core for L_ζ and L_χ, a fact which is established in Shiga (1980) [43]. Thus one has existence and uniqueness of the process ζ.

Remark The law of IFWD, ζ, is given by the solution of the $(L_\zeta, C_0^2([0,1]^G))$-martingale problem which is well-posed. That is, for each $f \in C_0^2([0,1]^G)$,

$$\left(f(\zeta_t) - \int_0^t L_\zeta f(\zeta_s) ds \right)_{t \ge 0} \tag{1.9}$$

is a continuous martingale for the canonical filtration. □

Remark IFWD can also be obtained as the law of the unique strong solution of a countable system of stochastic differential equations:

$$d\zeta_t(x) = \sum_{y \in G} (a(x,y) - \delta(x,y))\, \zeta_t(y)dt + \sqrt{g(\zeta_t(x))}dw_t^x, \qquad x \in G, \tag{1.10}$$

where $\zeta_0 \in [0,1]^G$, and $\{w^x := (w_t^x)_{t \geq 0};\, x \in G\}$ is a collection of independent standard Brownian motions on the real line, and $g(z) := \gamma z(1-z)$. □

We wish to relate the IMM and the IFWD via a limit theorem. Set $K := \{0,1\}$, fix $\theta \in [0,1]$, and let $(\eta^\rho)_{\rho \in \mathbb{R}^+}$ be a family of IMM on $G \times K$ satisfying the following set of conditions: (i) Assume that $\mathcal{L}[\eta(\cdot \times K)] \in \mathcal{E}_G$ is translation invariant and ergodic with total mass per site intensity $\rho > 0$, and (ii) that $\mathcal{L}[\eta(\cdot \times \{1\})] \in \mathcal{E}_G$ is translation invariant and ergodic with total mass per site intensity $\theta\rho$, $\theta \in [0,1]$.

Define then the functional $\widehat{\eta}^\rho$ of η^ρ by

$$\widehat{\eta}_t^\rho(x) := \frac{\eta_t^\rho(\{x\} \times \{1\})}{\eta_t^\rho(\{x\} \times K)} 1_{\{\eta_t^\rho(\{x\} \times K) \neq 0\}}, \tag{1.11}$$

i.e., $\widehat{\eta}_t^\rho$ represents the *relative frequency of type* 1 *individuals.*

Remark If the initial distribution satisfies the assumptions (i) and (ii) above, then at every time $t > 0$ the conditions are satisfied as well. This can be verified using truncation arguments together with moment calculations, which are tedious but standard and are omitted here. □

Then $\widehat{\eta}^\rho$ converges in law in Skorokhod topology. That is:

Theorem 0. *(Diffusion limit) Under the above assumption on the initial state η, we have*

$$\mathcal{L}^\eta[\widehat{\eta}^\rho] \underset{\rho \to \infty}{\overset{\mathcal{D}}{\Longrightarrow}} \mathcal{L}^{\underline{\theta}}[\zeta], \tag{1.12}$$

where the convergence means weak convergence on $\mathcal{M}_1(\mathcal{D}([0,\infty);[0,1]^G))$, and where $\mathcal{L}^{\underline{\theta}}$ on the right hand side indicates that the initial state of the IFWD ζ is $\zeta_0(x) = \theta$, for all $x \in G$.

Proof of Theorem 0 See Theorem 3 in Leopold (2001) [38]. □

Remark If the type space K is larger than $\{0,1\}$ the analog of (1.12) still holds where the limiting Markov process is the so-called interacting Fleming-Viot model. In order to keep notation as simple as possible we concentrate on the two type case, but in terms of the techniques, the case of general K can be handled by the very same methods. □

1.3. **Historical processes.** As motivated in Subsection 1.1, it is helpful to have detailed information on the genealogies which is contained in the description of IMM, $\xi := (\xi_t)_{t \geq 0}$, as defined in (i) in Subsection 1.2. This description was given by specifying migration and resampling of individual particles at random times but is lost after passing to the functional η. In (i) below we introduce therefore another functional of ξ, namely the *historical IMM*, $\eta^* = (\eta_t^*)_{t \geq 0}$, which does not contain all information about particular individuals but nevertheless allows us to give a complete description of the statistics of genealogies. A characterization of η^* via a martingale problem is given in Theorem 1 in Subsection 2.1. The latter characterization will then justify to pass in Paragraph (ii) below to the diffusion limit of historical IMM and obtain historical IFWD, $\zeta^* := (\zeta_t^*)_{t \geq 0}$. Both these historical processes are *time-inhomogeneous Markov processes.*

(i) The historical process of the IMM arises through the following construction which we present in three steps.

Step 1 (Lines of descent) For the first step fix a time horizon t. With every individual present in the population at time t we associate its *path of descent.* This path follows the random walk in reversed time from the time t location until the birth time of the individual. At that time the parent particle from whom the type has been inherited provides the continuation of the path back to its birth place. This is continued until we reach time 0.

Observe that this procedure gives at time t only information about those individuals which were able to pass on their type to at least one individual alive at time t. In order to obtain a state space for the path process which is independent of t we use the convention: for times $s < 0$ and $s > t$ the path is simply continued as the constant path.

Building blocks for the process describing the evolution of all paths of descent are the paths of descent processes for each individual. In this way we associate with every individual alive at time t a path in $\mathcal{D}(\mathbb{R}, G)$, that is its *line of descent*. At the same time the individual has a type in K. In order to keep our framework somewhat general, for example to allow for mutation, it is best to encode both the path of descent and the type as a path in $G \times K$ denoted by y, i.e.,

$$y \in \mathcal{D}(\mathbb{R}, G \times K). \tag{1.13}$$

The path of descent of a single tagged individual is a time-inhomogeneous Markov process on $\mathcal{D}(\mathbb{R}, G \times K)$, and is called the *path process* associated with the model.

Step 2 (Exchangeable measures on lines of descent) The next step is to associate for each time t with the path process of each individual a δ-measure on $\mathcal{D}(\mathbb{R}, G \times K)$. Then we can summarize the relevant information about the population at time t by forming the sum of these δ-measures. In this way, we obtain the random measure

$$\eta_t^* \in \mathcal{N}(\mathcal{D}(\mathbb{R}, G \times K)). \tag{1.14}$$

Step 3 (Historical IMM, η^*) Now we let the time horizon t vary. The associated stochastic process $\eta^* = (\eta_t^*)_{t \geq 0}$ is called the *historical IMM*. This process is a Markov process on a suitable subset of $\mathcal{N}(\mathcal{D}(\mathbb{R}, G \times K))$.

In the historical process the ordinary IMM are embedded by the relation

$$\eta_t(\{x\} \times \{k\}) = \eta_t^*(\{y \in \mathcal{D}(\mathbb{R}, G \times K) | y_t = (x, k)\}), \tag{1.15}$$

$(x, k) \in G \times K$. Hence, define the state space $\mathcal{E}_G^*$ in terms of the Liggett-Spitzer state space (recall $\mathcal{E}_G$ from (1.5)) by

$$\mu^* \in \mathcal{E}_G^* \subset \mathcal{M}(\mathcal{D}(\mathbb{R}, G \times K)) \tag{1.16}$$

if and only if, for each $t \in \mathbb{R}$,

$$\mu_t(\cdot) := \mu^*(\{y \in \mathcal{D}(\mathbb{R}, G \times K) | y_t \in \cdot \times K\}) \in \mathcal{E}_G. \tag{1.17}$$

This implies that if the process starts at time 0 with a population corresponding to a measure in $\mathcal{E}_G$ then for any $t > 0$, $\eta_t^* \in \mathcal{E}_G^*$, and in particular, η_t^* is a.s. locally finite (that is, σ-finite).

(ii) Motivated by Theorem 0 we define the *historical IFWD*, ζ^*, by the diffusion limit procedure: let $K := \{0, 1\}$ and let for each $A \subseteq G$,

$$E_{A,t} := \{y \in \mathcal{D}(\mathbb{R}, G \times K) : \, y_t \in A \times K\}, \tag{1.18}$$

and

$$E^{\{1\},t} := \{y \in \mathcal{D}(\mathbb{R}, G \times K) : \, y_t \in G \times \{1\}\} \tag{1.19}$$

denote the set of paths which are at site $x \in A$ at time t and the set of paths which have type 1 at time t, respectively. Consider a family of historical IMM, $(\eta^{*,\rho})_{\rho \in [0,\infty)}$, satisfying the following conditions: (i) The law of $\eta^{*,\rho}(\cdot)$ is translation invariant and the induced (via (1.15)) configuration η_t is shift ergodic for all t. (ii) The law of $\eta^{*,\rho}(\cdot \times K)$ has mass-(per site)-intensity $\rho > 0$, i.e., $\mathbf{E}[\eta_0^{*,\rho}(E_{\{x\},t})] = \rho$. (iii) Assume furthermore that $\mathbf{E}\left[\eta_0^{*,\rho}(E_{\{x\},t} \cap E^{\{1\},t})\right] = \theta\rho$.

Define then the functional $\hat{\eta}^{*,\rho}$ of $\eta^{*,\rho}$ in analogy to (1.11) as follows: for each measurable subset $A_{\{x\},t} \subseteq E_{\{x\},t}$,

$$\widehat{\eta}_t^{*,\rho}(A_{\{x\},t}) := \frac{\eta_t^{*,\rho}(A_{\{x\},t})}{\eta_t^{*,\rho}(E_{\{x\},t})} 1_{\{\eta_t^{*,\rho}(E_{\{x\},t}) > 0\}}. \tag{1.20}$$

Then the limiting process,

$$\zeta^* := (\zeta^*_t)_{t\geq 0}, \tag{1.21}$$

obtained by letting $\rho \to \infty$, is a $\mathcal{M}(\mathcal{D}(\mathbb{R}, G\times K))$-valued process with the following additional property: for each $x \in G$, $\zeta^*_t(E_{\{x\},t}) = 1$.

It turns out that the resulting limit process can again, as on the level of the non-historical processes η and ζ, be described by a martingale problem. We carry this out in Subsection 2.1.

Remark Once more we can use arbitrary type spaces, e.g. $K := [0,1]$, in order to obtain the so-called historical interacting *Fleming-Viot model* without much further mathematical input. □

1.4. **The historical look-down process and the enriched historical IMM.** The main tool for establishing analytical characterizations via well-posed martingale problems and for constructing duality relations for the historical processes will be the so-called *historical look-down process*. With this process one can realize on the underlying probability space a whole collection of historical IMM with a given type intensity θ and total mass intensities $\rho \in \mathbb{R}^+$, as well as the diffusion limit the historical IFWD. For such a construction to be possible it is essential to use the *exchangeability* between all lines of descent. The historical look-down process corresponds to a representation of the enriched IMM, ξ, as a historical process ξ^* (the enriched historical IMM), i.e., a collection of random paths, which we give subsequently.

(i) Fix a countable index set $\mathcal{I}$ representing the reservoir of individuals we shall have to consider for every *fixed* value of t. Note that we would need a larger set of indices if we wished to label all individuals present in the population at *any* time! The *historical look-down*, $X^* := (X^*_t)_{t\geq 0}$, is a process where

$$X^*_t = \{X^{*,\iota}_t;\, \iota \in \mathcal{I}\} \in (\mathcal{D}(\mathbb{R}, G\times K)\times \mathcal{U})^{\mathcal{I}}, \tag{1.22}$$

with *label set* $\mathcal{U} := \{u^\iota;\, \iota\in\mathcal{I}\}$ a countable subset of $[0,\infty)$ which will play a rôle in defining the dynamics, in particular it determines the *look-down* order and this in addition allows us to work with a “maximal” type space.

The initial state of X^* will be given by specifying initial positions in G, $\{x^\iota;\, \iota\in\mathcal{I}\}$, and labels, $\{u^\iota;\, \iota \in \mathcal{I}\}$, which have to be chosen such that for each $\rho \in (0,\infty)$, the spatial configuration arising from positions corresponding to labels less than or equal to ρ lies in the Liggett-Spitzer space (recall (1.5)). That is, for each $\rho \in (0,\infty)$,

$$\pi_G \circ \left(\sum_{\iota\in\mathcal{I}^\rho} \delta_{\{x^\iota\}}\right) \in \mathcal{E}_G, \tag{1.23}$$

where for each $\rho\in\mathbb{R}^+$,

$$\mathcal{I}^\rho := \{\iota \in \mathcal{I} :\, u^\iota \leq \rho\}. \tag{1.24}$$

Notice that condition (1.23) is preserved if the individuals perform independently RW with kernel $a(x,y)$.

Convention Moreover we choose as initial types $k^\iota := u^\iota$, that is, as mentioned above we use at this point

$$K = \mathcal{U} := \{u^\iota;\, \iota\in\mathcal{I}\}. \qquad \square \tag{1.25}$$

The initial state is then defined as follows:

$$X^*_0 := \{(\text{constant path through } (x^\iota, k^\iota), u^\iota);\, \iota \in \mathcal{I}\}\,. \tag{1.26}$$

The initial positions, types, and labels are random and defined on $(\Omega_0, \mathcal{F}_0, \mathbf{P}_0)$.

Next we give an explicit construction of the dynamics of the process X^*. The construction is split into three steps.

Step 1 (Migration) We assume there is a probability space $(\Omega_1, \mathcal{F}_1, \mathbf{P_1})$ supporting the following construction. There is a $\mathcal{P}(G)\times\mathcal{B}(\mathbb{R}^+)\times\mathcal{F}_1$–measurable mapping $R: G\times\mathbb{R}^+\times\Omega_1 \to G$ such that $R(x,0,\omega) = x$, for each

$x \in G$, $\mathbf{P}_1$-a.s., and $(R(z,t,\boldsymbol{\cdot}))_{t\geq 0}$ is a Markov process with transition function $a_t(x,y)$ (recall (1.1)) starting in z. We abbreviate $R(x,s)$ for the random variable $R(x,s,\boldsymbol{\cdot})$.

Let now on $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$

$$\{R^\iota;\, \iota \in \mathcal{I}\} \tag{1.27}$$

be a collection of independent realizations of R.

Step 2 (Resampling) Let

$$\{N^{\iota_1,\,\iota_2};\, \iota_1,\, \iota_2 \in \mathcal{I},\, u^{\iota_1} < u^{\iota_2}\}, \tag{1.28}$$

be independent realizations of rate γ Poisson processes, defined on a probability space $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$.

Each time $t \in N^{\iota_1,\,\iota_2}$, $\iota_1,\, \iota_2 \in \mathcal{I}$ such that $u^{\iota_1} < u^{\iota_2}$, is a candidate time for a resampling event at which the particle with index ι_1 copies its path of descent and type and pushes the path of descent and type with index ι_2 out of the configuration and this transition is realized if both individuals are sharing the same site, i.e., $R^{\iota_1}(x^{\iota_1}, t) = R^{\iota_2}(x^{\iota_2}, t)$.

Next we combine migration and resampling.

Step 3 (Migration and resampling combined) Consider the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ which is the product of the $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i)$, $i = 0, 0, 1$. That is, on $(\Omega, \mathcal{F}, \mathbf{P})$ the initial configuration X_0^*, the random walks R, and the Poisson processes N are all independent of each other.

We define each coordinate of $\{(X_t^{*,\,\iota})_{t\geq 0};\, \iota \in \mathcal{I}\}$ as a random element of $\mathcal{D}(\mathbb{R}, G \times K) \times \mathcal{U}$ in the following way. Between two look-down events the individual indexed by ι_2 moves according to the law of R on G, while at the jump times it looks down at the individual sharing the same site and indexed by ι_1 ($u^{\iota_1} < u^{\iota_2}$) in order to adopt both the path of descent and the type of ι_1 (which is actually a constant path through the type space). The individual **always keeps its label**. Formally this is done as follows.

Let for $\iota_1,\, \iota_2 \in \mathcal{I}$ such that $u^{\iota_1} < u^{\iota_2}$,

$$\begin{aligned} &B^{\iota_1,\,\iota_2}([0,t]) := \\ &\qquad \#\{s \in [0,t]:\ \ R^{\iota_1}(x^{\iota_1}, s) = R^{\iota_2}(x^{\iota_2}, s),\, N^{\iota_1,\,\iota_2}([0,s]) = N^{\iota_1,\,\iota_2}([0,s)) + 1\}. \end{aligned} \tag{1.29}$$

Then $B^{\iota_1,\,\iota_2}$ measures the number of look-down events between the individuals with index ι_1 and ι_2. Whenever the particle with index ι_2 looks down at the particle with index ι_1 the particle with index ι_2 adopts both the type and the spatial path of the particle with index ι_1, but it never changes its label. This defines effectively a resampling event.

For $\iota,\, \iota_1,\, \iota_2 \in \mathcal{I}$, $n \geq 1$, let us denote by

$$\gamma_\iota^1 := \min\{s > 0 : \sum_{\iota' \in \mathcal{I}:\, u^{\iota'} < u^\iota} B^{\iota',\,\iota}([0,s]) = 1\} \tag{1.30}$$

the time of the first look-down event of the particle with index ι to some particle (with a smaller label than u_ι), by

$$\tau^n_{\iota_1,\,\iota_2} := \min\{s > 0;\, B^{\iota_1,\,\iota_2}([0,s]) = n\} \tag{1.31}$$

the time of the n^{th} look-down event in which the particle indexed by ι_2 looks down at the particle indexed by ι_1, and by

$$\gamma^n_{\iota_1,\,\iota_2} := \min\{\tau^{n'}_{\iota',\,\iota_2} > \tau^n_{\iota_1,\,\iota_2};\, \iota' \in \mathcal{I}, u^{\iota'} < u^{\iota_2}, n' \geq 1\} \tag{1.32}$$

the first time after $\tau^n_{\iota_1,\,\iota_2}$ at which the particle with index ι_2 looks down at some particle.

At time $t \geq 0$, for any $\iota \in \mathcal{I}$, the random "path" $X_t^{*,\,\iota}$ is defined by requiring that $X_t^{*,\,\iota}$ is a "path" which is constant before time 0 and after time t, and is defined for intermediate times depending on t as follows:

- for $0 \leq t < \gamma_\iota^1$,

$$(X_t^{*,\iota})_s := (R^\iota(x^\iota, s), k^\iota, u^\iota) \qquad \forall\, s \in [0,t], \tag{1.33}$$

- and for $\tau^k_{\iota',\iota} \le t < \gamma^k_{\iota',\iota}$,

$$(X^{*,\iota}_t)_s := \begin{cases} (\pi^*_{G\times K}(X^{*,\iota'}_t)_s, u^\iota) & s \in [0, \tau^k_{\iota',\iota}) \\ (R^\iota(x^\iota, s), \pi^*_K(X^{*,\iota'}_{\tau^k_{\iota',\iota}}), u^\iota) & s \in [\tau^k_{\iota',\iota}, t] \end{cases}, \tag{1.34}$$

where $\pi^*_{G\times K}$ resp. π^*_K denotes the projection of elements in $\mathcal{D}(\mathbb{R}, G\times K)\times\mathcal{U}$ into $\mathcal{D}(\mathbb{R}, G\times K)$ resp. K.

This construction makes use of the facts that for each $\iota\in\mathcal{I}$, and fixed time $t\ge 0$, there exist $k\ge 1$, and ι' such that $u^{\iota'} < u^\iota$ and t fits in one of the two cases above, and that the particle with index ι looks down at most finitely often up to time t. The latter allows one to keep track all of its ancestors, and reconstruct the path of descent in a unique way.

The precise statement, which implies that the whole construction is well-defined, will be stated and proved in Section 4 (Proposition 4.1).

Remark The historical look-down denoted by X^* is the analogue of the *look-down process* by Donnelly and Kurtz (1996) first given in [19], extended in Donnelly, Evans, et. al. (1999) [18], and in Kurtz (1998) [37].

Our construction differs from the one in [19] in two aspects. First, in the setting of [19] the labels are smaller than a given value, and the total number of particles in each of the approximating particle systems is finite, which is not the case for IMM if G is infinite. Secondly, what is called migration in [19] describes mutation and is not the same kind of migration which we have here between colonies. That is, the resampling rate in [19] is the same for all particles, which is not the case for IMM since resampling takes place only if a pair of individuals is sharing the same site. The latter condition is dropped in [18] but previously the look-down approach has not been used in a historical setting. □

(ii) Having constructed the historical look-down, we construct next the *enriched historical IMM*, denoted by ξ^*. In fact we need a whole sequence of such historical processes, $\{\xi^{*,\rho}; \rho > 0\}$, to finally approximate the *historical IFWD* ζ^*. In other words we explicitly construct the IMM including all genealogical data for every parameter value ρ, and get from these data also the historical IFWD, as $\rho\to\infty$. Furthermore as a functional of ξ^* we get a specific version of η^* on that probability space.

Fix a countable index set $\widetilde{\mathcal{I}}$. The *enriched historical* IMM, i.e. the historical IMM which distinguishes between the individuals, $\xi^* := (\xi^*_t)_{t\ge 0}$, is a process with:

$$\xi^*_t = \{\xi^{*,\widetilde{\iota}}_t; \widetilde{\iota}\in\widetilde{\mathcal{I}}\} \in (\mathcal{D}(\mathbb{R}, G\times K))^{\widetilde{\mathcal{I}}}. \tag{1.35}$$

The law of ξ^* is constructed by prescribing initial state and dynamics as follows.

The *initial state* of ξ^* will be specifying the initial locations and the initial types. The initial positions, $\{\widetilde{x}^{\widetilde{\iota}}; \widetilde{\iota}\in\widetilde{\mathcal{I}}\}$, have to be chosen such that

$$\pi_G \circ \big(\sum_{\widetilde{\iota}\in\widetilde{\mathcal{I}}} \delta_{\{\widetilde{x}^{\widetilde{\iota}}\}}\big) \in \mathcal{E}_G. \tag{1.36}$$

Notice that if $\underline{x}\in (G\times K)^{\mathcal{I}}$ satisfies (1.36) then (1.23) holds as well. However, since (1.36) does not allow for locally infinite configurations, (1.23) is a stronger condition than (1.36).

The initial state is then given via the two above ingredients defined as follows:

$$\xi^*_0 := \Big\{\text{constant path through } (\widetilde{x}^{\widetilde{\iota}}, \widetilde{k}^{\widetilde{\iota}}); \widetilde{\iota}\in\widetilde{\mathcal{I}}\Big\}. \tag{1.37}$$

The positions and types are random and defined on $(\widetilde{\Omega}_0, \widetilde{\mathcal{F}}_0, \widetilde{\mathbf{P}}_0)$.

Next we specify the *dynamics* of the process consisting of the two mechanisms, migration and resampling. We assume therefore that there are two probability spaces $(\widetilde{\Omega}_1, \widetilde{\mathcal{F}}_1, \widetilde{\mathbf{P}}_1)$ and $(\widetilde{\Omega}_2, \widetilde{\mathcal{F}}_2, \widetilde{\mathbf{P}}_2)$ supporting a collection of independent realizations of RW with transition kernel $a_t(x,y)$, denoted by $\{\widetilde{R}^{\widetilde{\iota}}; \widetilde{\iota}\in\widetilde{\mathcal{I}}\}$, and a collection of rate $\gamma/2$ Poisson point processes, $\{\widetilde{N}^{\widetilde{\iota}_1,\widetilde{\iota}_2}; \widetilde{\iota}_1\ne\widetilde{\iota}_2\in\widetilde{\mathcal{I}}\}$, respectively.

Consider then the probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbf{P}})$ which is the product of $(\widetilde{\Omega}_i, \widetilde{\mathcal{F}}_i, \widetilde{\mathbf{P}}_i)$, $i = 0, 1, 2$. Again individuals perform random walks based on $\widetilde{R}^\iota$ and $\tilde{\iota}_1, \tilde{\iota}_2$ resample at the points of $\widetilde{N}^{\tilde{\iota}_1,\tilde{\iota}_2}$ and the initial state is given via (1.37). Precisely proceed as follows.

As done before, let $\widetilde{B}^{\widetilde{\iota}_1,\widetilde{\iota}_2}$ be the random measure which counts the jumps of $\widetilde{N}^{\tilde{\iota}_1,\tilde{\iota}_2}$ at which the positions $\widetilde{R}^{\widetilde{\iota}_1}$ and $\widetilde{R}^{\widetilde{\iota}_2}$ coincide. We denote by $\widetilde{\gamma}^1_{\widetilde{\iota}}$ the time of the first jump of $\sum_{\widetilde{\iota}'\neq\widetilde{\iota}} \widetilde{B}^{\widetilde{\iota}',\widetilde{\iota}}$, by $\widetilde{\tau}^n_{\widetilde{\iota}_1,\widetilde{\iota}_2}$ the n^{th} jump time of $\widetilde{B}^{\widetilde{\iota}_1,\widetilde{\iota}_2}$, and by $\widetilde{\gamma}^n_{\widetilde{\iota}_1,\widetilde{\iota}_2}$ the first jump time after $\widetilde{\tau}^n_{\widetilde{\iota}_1,\widetilde{\iota}_2}$ which involves the particle with index $\widetilde{\iota}_2$.

Define then for given t,

- for $0 \leq t < \widetilde{\gamma}^1_{\widetilde{\iota}}$,

$$(\xi^{*,\widetilde{\iota}}_t)_s := \left(\widetilde{R}^{\widetilde{\iota}}(\widetilde{x}^{\widetilde{\iota}}, s), \widetilde{k}^{\widetilde{\iota}}\right) \qquad \forall\, s \in [0,t], \tag{1.38}$$

- and for $\widetilde{\tau}^k_{\widetilde{\iota}',\widetilde{\iota}} \leq t < \widetilde{\gamma}^k_{\widetilde{\iota}',\widetilde{\iota}}$,

$$(\xi^{*,\widetilde{\iota}}_t)_s := \begin{cases} (\xi^{*,\widetilde{\iota}'}_t)_s & s \in [0, \widetilde{\tau}^k_{\widetilde{\iota}',\widetilde{\iota}}) \\ \left(\widetilde{R}^{\widetilde{\iota}}(\widetilde{x}^{\widetilde{\iota}}, s), \pi^*_K(\xi^{*,\widetilde{\iota}'}_{\widetilde{\tau}^k_{\widetilde{\iota}',\widetilde{\iota}}})\right) & s \in [\widetilde{\tau}^k_{\widetilde{\iota}',\widetilde{\iota}}, t] \end{cases}. \tag{1.39}$$

Then once more by Proposition 4.1, ξ^* is well-defined. We then consider

$$\eta^*_t := \sum\nolimits_{\widetilde{\iota}\in\widetilde{\mathcal{I}}} \delta_{\xi^{*,\widetilde{\iota}}_t}. \tag{1.40}$$

2. Main results I: Analytical characterizations

In this section we develop the concepts necessary to provide martingale problem characterizations for the various historical processes introduced in Subsections 1.3 and 1.4. In Subsection 2.1 we state in Theorems 1 and 2 the well-posedness of the martingale problems for X^*, ξ^*, η^*, respectively ζ^*. The given analytic representation allows one (cf. Theorem 3 in Subsection 2.2) to embed a family of enriched historical IMM, $\{\xi^{*,\rho}; \rho \in [0,\infty)\}$, into the historical look-down process, X^*, and consequently to represent the historical IFWD, ζ^*, and the historical IMM $\{\eta^{*,\rho};\ \rho \in [0,\infty)\}$ as its a.s. functional. Once we have established this structure we apply it to solving concrete problems which we then present in Section 3.

2.1. **The martingale problem characterization (Theorems 1 and 2).**

For many purposes it is necessary to have instead of the explicit construction an analytically more manageable description of the law of the historical processes X^*, ξ^*, η^*, and ζ^* at hand. Here the key observation is that we can describe the evolution of the historical process without referring back to the random evolution of all the individuals or to a graphical construction. Namely, we describe the processes as solutions to well-posed *martingale problems.* In this subsection we formulate in several steps the martingale problems for our various historical processes, and state the key results on the well-posedness.

Note that the path processes and hence the historical processes are time inhomogeneous. Therefore we enlarge the state space, by including time, to obtain a homogeneous Markov process which we can characterize by the generator of the corresponding semigroup. For a Markov process $X := (X_t)_{t\geq 0}$ the *time-space process of* X is given by $(t, X_t)_{t\geq 0}$.

The construction of the martingale problems is a bit complex and we proceed in several steps. We start giving the key ingredient for all of them, which is the historical migration process in Step 1, and then we write down the martingale problem for the various processes in Steps 2–5, first in Steps 2 and 3 specifying the generator and then in Steps 4 and 5 the martingale problems.

Step 1 (The historical migration process) The key element of the evolution in time and space is the *path process* describing the positions in space taken by the various individuals. The description of this evolution is the aim of Step 1.

The first object we need to define is the generator $\widetilde{A}$ of the (time-space) path process, which arises from the migration. We now specify the domain of $\widetilde{A}$. The action of the operator $\widetilde{A}$ is defined on functions of the special form,

$$\Phi(s, y), \qquad s \in \mathbb{R},\, y \in \mathcal{D}(\mathbb{R}, G \times K), \tag{2.1}$$

given as follows. Consider for $n \in \mathbb{N}$ the collection of functions:

$$g_j : \mathbb{R} \times G \times K \to \mathbb{R}, \qquad j = 1, \ldots, n, \tag{2.2}$$

where g_j are bounded and C^1 in the time variable, $j = 1, \ldots, n$. Fix a collection $0 < t_1 < t_2 < \cdots < t_n$, $n \in \mathbb{N}$, of time points. Now define the function

$$\Phi(t, y) = \prod\nolimits_{j=1}^{n} g_j(t, y_{t \wedge t_j}). \tag{2.3}$$

Then we look at the algebra $\widetilde{\mathcal{A}}$ of functions generated by the functions given in (2.3), where $n \in \mathbb{N}$, t_j, and g_j, $j = 1, ..., n$, vary over all possibilities described above. Observe that this algebra $\widetilde{\mathcal{A}}$ is measure determining on $\mathcal{D}(\mathbb{R}, G \times K)$.

To define the action of $\widetilde{A}$ (note that $\widetilde{A}$ and $\widetilde{\mathcal{A}}$ are different symbols) we use the generator A of the RW on $G \times K$ given by

$$Af(x, k) = \sum_{z \in G} (a(x, z) - \delta(x, z)) f(z, k), \tag{2.4}$$

for all f bounded on $G \times K$. The action of the operator $\widetilde{A}$ on a given function Φ in the algebra $\widetilde{\mathcal{A}}$, is given, for each $t \in (t_k, t_{k+1}]$, by

$$\begin{aligned} \widetilde{A}\Phi(t, y) = \ & \prod_{j=1}^{k} g_j(t, y_{t \wedge t_j}) [(\tfrac{\partial}{\partial t} + A) \prod_{j=k+1}^{n} g_j(t, y_{t \wedge t_j})] \\ & + [\tfrac{\partial}{\partial t} \prod_{j=1}^{k} g_j(t, y_{t \wedge t_j})] \, [\prod_{j=k+1}^{n} g_j(t, y_{t \wedge t_j})]. \end{aligned} \tag{2.5}$$

In order to formulate the martingale problems we need one more ingredient: denote by y^r the *path y stopped at time r*, i.e.,

$$y^r_\cdot = y_{\cdot \wedge r}, \qquad y \in \mathcal{D}(\mathbb{R}, G \times K). \tag{2.6}$$

This map of $\mathcal{D}(\mathbb{R}, G \times K)$ into itself induces a map on $\mathcal{M}(\mathcal{D}(\mathbb{R}, G \times K))$ into those measures on $\mathcal{D}(\mathbb{R}, G \times K)$ which are supported on paths staying constant after time r. That is, (2.6) induces the map

$$\eta^* \mapsto \eta^{*,r}, \tag{2.7}$$

and $\eta^{*,r}$ is concentrated on that subset of paths, which describes paths of descent of a possible population of individuals alive at time r and which therefore can evolve further at time r. We finally denote by

$$\pi^*_G y \text{ and } \pi^*_K y \tag{2.8}$$

the projections of a path y onto $\mathcal{D}(\mathbb{R}, G)$ and onto $\mathcal{D}(\mathbb{R}, K)$, respectively.

Step 2 (Generator for X^* and ξ^*) The basis of the characterization of X^* is the following observation concerning this process. By construction X^* is *Markovian* but time inhomogeneous, and the time-space process determines a Feller semigroup on $\mathcal{C}(\mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G \times K) \times \mathcal{U})^{\mathcal{I}})$. We can identify the generator $L_{(t,X^*)}$ of the time-space process of X^* on certain functions explicitly. The action of $L_{(t,X^*)}$ is determined via two ingredients related to migration and resampling, which we define next.

Start with the migration part. Recall next from (2.4) the migration generator A acting on bounded functions $f : G \times K \to \mathbb{R}^+$. In an obvious way we extend the generator to a generator $A^{\mathcal{I}}$ which acts now on bounded functions $f^{\mathcal{I}} : (G \times K)^{\mathcal{I}} \to \mathbb{R}^+$ which depend on finitely many coordinates (indices), $\iota \in \mathcal{I}$, only. Namely, for each $\iota \in \mathcal{I}$, we let A^ι denote the generator which acts on $f^{\mathcal{I}}$ as a function of the ι^{th} coordinate as defined in (2.4), and set then

$$A^{\mathcal{I}} := \sum_{\iota \in \mathcal{I}} A^\iota. \tag{2.9}$$

Having done this we extend the definition of the historical migration generator, $\widetilde{A}$, given in (2.5), which acts on functions of type (2.3), to an operator

$$\widetilde{A}^{\mathcal{I}} \quad \text{(historical migration generator)}. \tag{2.10}$$

The operator $\widetilde{A}^{\mathcal{I}}$ acts on functions Φ as in (2.5) where g_j are now defined on $\mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}, G \times K) \times \mathcal{U})^{\mathcal{I}}$ as a product of functions on $\mathbb{R}^+ \times \mathcal{D}(\mathbb{R}, G \times K) \times \mathcal{U}$ which have form (2.3) and which are constant in the $\mathcal{U}$-variable as functions on $\mathbb{R}^+ \times \mathcal{D}(\mathbb{R}, G \times K)$, and by replacing A by $A^{\mathcal{I}}$ in (2.5). This definition is extended to the algebra, denoted by $\mathcal{A}^{*,\text{look}}$, which is generated by functions of this type.

Now turn to the resampling part. We use the following notation. We write an element in $(\mathcal{D}(\mathbb{R}, G \times K) \times \mathcal{U})^{\mathcal{I}}$ as an underlined object with an entry for each $\iota \in \mathcal{I}$. Define for each $(t, \underline{(y,u)}) \in \mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G \times K) \times \mathcal{U})^{\mathcal{I}}$, and $\iota_1,\, \iota_2 \in \mathcal{I}$, $\widetilde{\theta}_{\iota_1,\,\iota_2}(t, \underline{(y,u)})$ as the element in $\mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G \times K) \times \mathcal{U})^{\mathcal{I}}$ obtained by replacing the ι_2-th path-coordinate of $(t, \underline{(y,u)})$ by its ι_1-th path-coordinate. That is, in symbols,

$$\begin{aligned} \widetilde{\theta}_{\iota_1,\,\iota_2} : \quad & (t, (..., (y^{\iota_1}, u^{\iota_1}), ..., (y^{\iota_2}, u^{\iota_2}), ...)) \\ & \mapsto (t, (..., (y^{\iota_1}, u^{\iota_1}), ..., (y^{\iota_1}, u^{\iota_2}), ...)) \,. \end{aligned} \tag{2.11}$$

We then define for a function F on the state space of X^*:

$$\widetilde{\Theta}_{\iota_1,\,\iota_2} F := F \circ \widetilde{\theta}_{\iota_1,\,\iota_2}. \tag{2.12}$$

With the above two ingredients the generator $L_{(t,X^*)}$ is defined on functions in $\mathcal{A}^{*,\text{look}}$ which are bounded functions $F : \mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G \times K) \times \mathcal{U})^{\mathcal{I}} \to \mathbb{R}^+$, the depend on finitely many coordinates only. We set then for $F \in \mathcal{A}^{*,\text{look}}$,

$$\begin{aligned} & L_{(t,X^*)} F(s, \underline{(y,u)}) \\ & \quad = \widetilde{A}^{\mathcal{I}} F(s, \underline{(y,u)}) + \gamma \sum_{\substack{\iota_1,\,\iota_2 \in \mathcal{I};\\ u^{\iota_1} < u^{\iota_2}}} 1_{\{(\pi^*_G y^{\iota_1})_s = (\pi^*_G y^{\iota_2})_s\}} (\widetilde{\Theta}_{\iota_1,\,\iota_2} F - F)(s, \underline{(y,u)}). \end{aligned} \tag{2.13}$$

Note that this specifies $L_{(t,X^*)}$ on a dense subset of $\mathcal{C}(\mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G \times K) \times \mathcal{U})^{\mathcal{I}})$.

Define the algebra, $\mathcal{A}^{*,\text{ind}}$, of functions consisting of functions in $\mathcal{A}^{*,look}$ which are constant in the $\mathcal{U}$-label variables. Note that if we symmetrize the resampling, we obtain on $\mathcal{A}^{*,\text{ind}}$ the operator which will generate the enriched historical IMM:

$$\begin{aligned} & L_{(t,\xi^*)} F(s, \underline{y}) \\ & \quad = \widetilde{A}^{\mathcal{I}} F(s, \underline{y}) + \frac{\gamma}{2} \sum_{\iota_1,\,\iota_2 \in \mathcal{I}} 1_{\{(\pi^*_G y^{\iota_1})_s = (\pi^*_G y^{\iota_2})_s\}} (\widetilde{\Theta}_{\iota_1,\,\iota_2} F - F)(s, \underline{y}). \end{aligned} \tag{2.14}$$

Step 3 (Generator for historical IMM) We can now describe the generator for the martingale problem defining η^*. We start by specifying the domain of the generator. Consider the algebra $\mathcal{A}^*$ which is generated by functions F on the state-space of the time-space process $(t, \eta^*)_{t \geq 0}$, i.e., on $\mathbb{R} \times \mathcal{N}(\mathcal{D}(\mathbb{R}, G \times K))$,

$$F(t, \eta^*) := f\left(\langle \eta^*, \Phi(t, \cdot) \rangle\right), \qquad f \in C^2_b(\mathbb{R}), \Phi \in \widetilde{\mathcal{A}}. \tag{2.15}$$

Introduce for every pair of parameters $y, y' \in \mathcal{D}(\mathbb{R}, G \times K)$ the operator $\Delta_{y,y'}$ acting on functions $F \in \mathcal{A}^*$ by

$$\Delta_{y,y'} F(t, \eta^*) := \begin{cases} F(t, \eta^* - \delta_y + \delta_{y'}) - F(t, \eta^*) & \text{if } \eta^*(\{y\}) \geq 1 \\ 0 & \text{otherwise} \end{cases} . \tag{2.16}$$

Denote for each $(t,z) \in \mathbb{R}^+ \times G$, and $y \in \mathcal{D}(\mathbb{R}, G \times K)$,

$$y_s^{(t,z)} := \begin{cases} y_s, & \text{if } s < t \\ y_s + (z,0), & \text{if } s \geq t \end{cases}. \tag{2.17}$$

Furthermore let $\{I^t;\, t \in \mathbb{R}^+\}$ be functions on $(\mathcal{D}(\mathbb{R}, G \times K))^2$ defined as follows:

$$I^t(y,y') = \begin{cases} 1 & \text{if } (\pi_G^* y)_t = (\pi_G^* y')_t \\ 0 & \text{otherwise} \end{cases}. \tag{2.18}$$

Define now the operator $L_{(t,\eta^*)}$ on $\mathcal{A}^*$ by (recall (2.7) for $\mu^{*,s}$)

$$\begin{aligned} L_{(t,\eta^*)} F(s,\mu^*) &:= \frac{\partial}{\partial t} F(s,\mu^*) \\ &+ \int_{\mathcal{D}(\mathbb{R}, G\times K)} \mu^{*,s}(dy) \sum_{z \in G} a(0,z) \Delta_{y, y^{(t,z)}} F(s,\mu^*) \\ &+ \frac{\gamma}{2} \int_{(\mathcal{D}(\mathbb{R}, G\times K))^2} \Delta_{y,y'} F(s,\mu^*)\, I^s(y,y')\, \mu^{*,s}(dy)(\mu^{*,s}(dy') - \delta_y(dy')). \end{aligned} \tag{2.19}$$

Step 4 (Martingale problems historical IMM, look-down) Recall the notion of a solution of a martingale problem. We say that a law on $\mathcal{D}([0,\infty), \mathcal{E}_G^*)$ satisfies the $(L_{(t,\eta^*)}, \mathcal{A}^*)$*-martingale problem with initial value* μ_0^*, if for each $F \in \mathcal{A}^*$ the canonical process $(\mu_s^*)_{s \geq 0}$ satisfies that:

$$\left(F(s,\mu_s^*) - F(0,\mu_0^*) - \int_0^s \mathrm{d}u\, L_{(t,\eta^*)} F(u,\mu_u^*) \right)_{s \geq 0} \tag{2.20}$$

is a martingale with respect to the canonical filtration. The martingale problems for X^*, ξ^* are defined similarly.

We then define

Definition 2.1. *(Historical look-down and IMM) The law of the historical look-down process X^*, the process ξ^* and the historical spatially interacting Moran models, η^*, are given by the unique solutions of the $(L_{(t,X^*)}, \mathcal{A}^{*,look})$, $(L_{(t,\xi^*)}, \mathcal{A}^{*,\mathrm{ind}})$ and $(L_{(t,\eta^*)}, \mathcal{A}^*)$-martingale problems, respectively.*

The above definition makes sense, if we can show that the martingale problems are well-posed.

Theorem 1. *(Martingale problem; historical IMM) The $(L_{(t,\eta^*)}, \mathcal{A}^*)$-martingale problem is well-posed and the process defined in (1.14) is a version of the solution. A similar statement holds for the $(L_{(t,\xi^*)}, \mathcal{A}^{*,\mathrm{ind}})$-martingale problem and the process $\xi^{*,\rho}$, respectively for $(L_{(t,X^*)}, \mathcal{A}^{*,look})$-martingale problem and the look-down process X^* constructed in Subsection 1.4.*

Step 5 (Martingale problem for historical IFWD) Next we give the analog description for the IFWD. Let the algebra $\tilde{\mathcal{A}}$ and the operator $\tilde{A}$ be defined as in (2.3) and (2.5), respectively.

Definition 2.2. *(Historical IFWD) For a given initial state at time s, $\zeta_s^* = \zeta_s^{*,s}$, such that for each $x \in G$, $\zeta_s^*(E_{\{x\},s}) = 1$, the historical process ζ^* is the solution of the following martingale problem on $C([s,\infty), \mathcal{E}_G^*)$. For every $\Phi \in \widetilde{\mathcal{A}}$ and every pair (t,s) with $t \geq s$:*

$$\left\{ \langle \zeta_t^*, \Phi(t,\cdot) \rangle - \langle \zeta_s^*, \Phi(s,\cdot) \rangle - \int_s^t \langle \zeta_r^{*,r}, (\widetilde{A}\Phi)(r,\cdot) \rangle dr \right\}_{t \geq s} \tag{2.21}$$

is a martingale with increasing process

$$\left(\int_s^t \int_{(\mathcal{D}(\mathbb{R}, G\times K))^2} I^r(y,y') \Phi(r,y) \Phi(r,y')\, \zeta^{*,r}(dy)(\zeta^{*,r}(dy') - \delta_y(dy'))\, dr \right)_{t \geq s}. \tag{2.22}$$

This definition makes sense since we shall prove:

Theorem 2. *(Martingale problem; historical IFWD) The martingale problem (2.21) to (2.22) is well-posed. The solution arises as the diffusion limit (see (1.21)) of* η^*.

2.2. **The particle representation (Theorem 3).** We are now ready to outline the philosophy in our approach to interacting Fisher-Wright diffusions, which has three aspects: First we use our knowledge about the Moran model to derive properties of the diffusion limit. In particular, we relate the long-term behavior of η and ζ. Secondly, in order to obtain representations for the family decomposition, we evaluate the whole information about genealogies contained in the historical processes. A third point is that the analysis of the historical process becomes more manageable if we construct the approximating IMM and their diffusion limit IFWD on the same probability space. The link is given by the historical look-down process, X^*, introduced in Subsection 1.4, i.e., a particle system with locally infinitely many particles whose genealogy contains the genealogies of the approximating IMM as random subgenealogies and is therefore rich enough to "generate" the genealogy of IFWD.

The idea behind such a construction, which goes back to Donnelly and Kurtz, is to reorder for a fixed time t the individuals in the population according to the time they will remain in the population. That means in the non-spatial case, the individuals are labeled in such a way that an individual with a smaller label will be replaced later than the one with a larger label, and therefore during any resampling event the type is always passed from the individual with the smaller label to the individual with the larger label. We will below develop the necessary concepts to make similar ideas work in a spatial context with countably many approximating particles.

More precisely we will give now in (i) a representation for a collection of $\{\xi^{*,\rho}, \rho > 0\}$ on the same probability space on which we introduced the historical look-down, X^*. The historical IFWD is then obtained as an a.s. limit of functionals of the historical look-down process X^*. This will be stated in Theorem 3 below. We conclude in (ii) with a discussion concerning the random configurations as initial states of our processes.

(i) The nice point observed first by Donnelly and Kurtz in the models without migration is that both processes, X^*_t and $\xi^{*,\rho}_t$ can be defined on the same probability space and we denote this space by

$$(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}}). \tag{2.23}$$

In order to construct this object we take the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ defined in the sequel of (1.26) and enrich it (independently from everything we already have) by an independent countable collection of coin flipping experiments.

On this space we can now, via the next theorem, also define a whole collection $\{(\xi^{*,\rho}_t)_{t\geq 0}; \rho > 0\}$ which then allows us to represent also the historical IFWD on that probability space.

Theorem 3. *(IMM and IFWD as functionals of the look-down process) Consider the probability space* $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}})$ *as basic space.*

(a) *Fix some* $\rho > 0$*. Then there exist random bijections* $\sigma^\rho_t := \{\sigma^\rho_t(\iota); \iota \in \mathcal{I}^\rho\}$ *from* $\mathcal{I}^\rho$ *onto* $\mathcal{I}^\rho$ *on the basic space such that the following holds: Define for each* $X^* \in (\mathcal{D}(\mathbb{R}, G \times K) \times \mathcal{U})^{\mathcal{I}}$ *and each* $t \geq 0$ *the random map:*

$$\Sigma^\rho_t X^* := \{\pi^*_{G\times K} X^{*,(\sigma^\rho_t)^{-1}(\widetilde{\iota})}; \widetilde{\iota} \in \mathcal{I}^\rho\}. \tag{2.24}$$

Furthermore introduce the random variable

$$\widetilde{\xi}^{*,\rho}_t := \Sigma^\rho_t X^*_t. \tag{2.25}$$

Then for fixed t, $\widetilde{\xi}^{*,\rho}_t$ *is a version of an enriched historical IMM with* $\widetilde{\mathcal{I}} = \mathcal{I}^\rho$ *at time* t .

The random bijections σ^ρ_t *are given explicitly in Proposition 5.1 in Section 5.*

(b) *Assume that we have an initial state for* X^* *such that*

$$\frac{1}{\rho} \sum_{\iota \in \mathcal{I}^\rho} \delta_{\pi^*_{G\times K}(X^{*,\iota}_0)} \text{ converges } \mathcal{L}(X^*_0)\text{-a.s., as } \rho \to \infty, \tag{2.26}$$

to the random element denoted by ζ_0^* with values in $\mathcal{M}(\mathcal{D}(\mathbb{R}, G \times K))$ concentrated on constant paths such that $\zeta_0^*(\{y \in \mathcal{D}(\mathbb{R}, G \times K) : y_0 \in \{x\} \times K\}) = 1$ for all $x \in G$.

Then for each $t \geq 0$,

$$\widetilde{\zeta}_t^* := \lim_{\rho\to\infty} \frac{1}{\rho} \sum_{\iota\in\mathcal{I}^\rho} \delta_{\pi^*_{G\times K}(X_t^{*,\iota})} \tag{2.27}$$

exists a.s. where the limit is taken in the weak topology on $\mathcal{M}(\mathcal{D}(\mathbb{R}, G))$ and $\mathcal{L}[\widetilde{\zeta}_t^*]$ is the law of the historical IFWD at time t started in ζ_0^*.

Remark The basic idea behind the construction of the random bijections σ_t^ρ is to successively and locally symmetrize the pecking order introduced by the look-down through coin-flipping experiments. □

(ii) The above theorem relates the historical look-down to the historical IMM and IFWD. In the applications that follow (on the strong form of duality, the long-term behavior, and the finite system scheme) we consider the latter starting in a *random configuration* which is typically associated with a random set of labels.

Another reason why one would like to introduce a random set of labels rather than a deterministic one arises when we consider as type space an uncountable set like $[0, 1]$ which is often done in Fleming-Viot models, and concentrate on the random countable set of types which exist in the system for strictly positive time.

However, in both cases one could try to transfer the situation to the case of a deterministic set of labels. There are in principle two strategies one can follow: either to code the random set of labels by a deterministic set of labels using a clever device, or to define the dynamics based on the set of possible labels in such a way that it is possible to answer the same questions as in a model with a fixed set of labels.

Since we are interested in a countable representation but would still like to have the Fleming-Viot model in mind, we follow a mixed strategy in enriching the notation by enumerating the individuals deterministically but choosing for each individual a random label from an uncountable set. However, the label set and the index set are supposed to be closely related by the fact that there is a concrete random bijection which maps one into the other.

We next describe the procedure of choosing jointly the initial positions and the labels corresponding to the initial types.

Step 1 (Initial state of X^*) Fix $\mathcal{I}$ countable, $\theta \in [0, 1]$, and let $K := \{0, 1\}$. Sample a Poisson field on $G \times K \times [0, \infty)$ with intensity $n \otimes \{(1 - \theta)\delta_0 + \theta\delta_1\} \otimes \lambda$, where n is counting measure on G, and λ is Lebesgue measure on $[0, \infty)$. Enumerate the countably many sampled points by $\mathcal{I}$, and let for the ιth point, $\iota \in \mathcal{I}$, x^ι, k^ι and u^ι be its G-valued, K-valued, and $\mathcal{U}$-valued coordinate. This defines by considering constant paths through an element in $G \times K$ uniquely a $\mathcal{I}$-indexed collection of $\mathcal{D}(\mathbb{R}, \Omega \times K) \times \mathcal{U}$-valued random variables which we can use as initial state for X^*. We denote this initial state as

$$\Psi(\theta) := \{(\text{constant path through } (x^\iota, k^\iota), u^\iota);\ \iota \in \mathcal{I}\}\,. \tag{2.28}$$

Step 2 (Initial state of $\xi^{*,\rho}, \eta^\rho$ and $\eta^{*,\rho}$) Fix $\rho \in \mathbb{R}^+$, $\widetilde{\mathcal{I}}$ countable, $\theta \in [0, 1]$, and let $K := [0, 1]$. Sample a Poisson field on $G \times K$ with intensity $\rho\, n \otimes \{(1 - \theta)\delta_0 + \theta\delta_1\}$. Enumerate the countably many sampled points by $\widetilde{\mathcal{I}}$. Once more for $\widetilde{\iota} \in \widetilde{\mathcal{I}}$, let $\widetilde{x}^{\widetilde{\iota}}$ and $\widetilde{k}^{\widetilde{\iota}}$ be $\widetilde{\iota}$'s G-valued and K-valued coordinate, respectively. Each of these points defines a constant path in $G \times K$. We then start the IMM distinguishing individuals in the corresponding element of $\mathcal{D}(\mathbb{R}, G \times K)^{\widetilde{\mathcal{I}}}$. The initial state of η^ρ is then given by:

$$\widetilde{\Psi}(\rho, \theta) := \sum\nolimits_{\widetilde{\iota}\in\widetilde{I}} \delta_{(\widetilde{x}^{\widetilde{\iota}}, \widetilde{k}^{\widetilde{\iota}})}. \tag{2.29}$$

Note that our sampling procedure ensures that the $\pi_{G\times K}$-projections of

$$\Psi(\rho, \theta) := \sum\nolimits_{\iota\in\mathcal{I}^\rho} \delta_{(x^\iota, k^\iota, u^\iota)} \tag{2.30}$$

equal in law $\widetilde{\Psi}(\theta, \rho)$.

Remarks Sometimes we choose Poisson random fields specifying initial positions only for notational convenience, while sometimes we are going to make use of the fact that for G countably infinite those laws are invariant under the migration dynamics. We will remark on this as we go on. □

3. Main results II: Applications

In the present section we establish, as the first application of our results, based on the main embedding theorem for the historical look-down process (see Theorem 3), a strong form of duality between the historical IMM and the historical IFDW on the one hand, and a spatially structured coalescent on the other. This duality relation is presented in Subsection 3.1, and then we give a representation of the genealogies via the spatially structured coalescent. All representation theorems are presented for (historical) IMM, IFDW and look-down process in a unified way. In Subsection 3.2 we will then apply the duality to state representations for the equilibrium historical processes. Finally, in Subsection 3.3 we present the so-called finite system scheme for the various historical processes.

3.1. **Representation of genealogies via coalescents, strong duality.** We next explain how the look-down construction can be used and extended to create a strong duality, i.e. to define the historical process and its dual process both on the same probability space which provides in particular a *historical form of duality.* This is a powerful tool for the analysis of the particle model and its diffusion limit, and is a key to the applications we give in the sequel.

The *duality relation* we have in mind will be a special case of a relation between the genealogies of the population alive at a reference time $t > 0$ which are an object obtained looking at the process in reversed time from t back to 0 and on the other hand a Markov process generating them, which is running forward in time.

Fix a reference time point $t > 0$. We are interested in the paths of descent of the population alive at time t. We wish to describe the evolution of these paths viewed *in reversed time* from the fixed reference time point t, in particular we record the times when they merge into one ancestral line, and their movement through space. Provided we can specify the dynamics for this random object , we will call the obtained process “the dual process”.

In order to answer these questions we proceed in steps. We start in Step 1 with considering the genealogy in the historical look-down X^*, and then in Step 2 the genealogy of the enriched historical IMM $\xi^{*,\rho}$. In Step 3 we construct the dual dynamics which is a spatially marked coalescent. Finally (Step 4) this can be used to derive stronger versions of duality than the moment duality between IMM or IFWD and the *coalescing random walks.*

Step 1 (Genealogy of X_t^*) Let $\mathcal{I}$ be the index set of all individuals. Fix a time t, and consider X_t^*. For each $s \in [0, t]$, define an *equivalence relation* $\approx_s^{X^*}$on $\mathcal{I}$: for ι_1, $\iota_2 \in \mathcal{I}$,

$$\iota_1 \approx_s^{X^*} \iota_2, \tag{3.1}$$

if at time t the individuals with indices ι_1 and ι_2 possess a common ancestor at time $t - s$ meaning that if we had assigned each individual a separate type at time $t - s$ (recall (1.25)), then types of these two individuals would agree at time t. This gives a decomposition into *family clusters* where the corresponding partition of $\mathcal{I}$ at time $0 \le s \le t$ is denoted by Γ_s^t. With each cluster (partition element) $\varpi \in \Gamma_s^t$ we can then associate its position $\ell_s^t(\varpi) \in G$, and denote the collection of positions by

$$\ell_s^t := \left\{ \ell_s^t(\varpi);\, \varpi \in \Gamma_s^t \right\}. \tag{3.2}$$

The latter allows us to define the *genealogical process*

$$(\Gamma_s^t, \ell_s^t)_{s \le t}, \tag{3.3}$$

which contains the information we need about the paths of descent of X^*.

Step 2 (Genealogy of $\xi_t^{*,\rho}$) Fix $\rho \in \mathbb{R}^+$. We define further equivalence relations $\approx_s^{\xi^{*,\rho}}$ in the set of labels $\widetilde{\mathcal{I}}$ for the enriched historical IMM $\xi^{*,\rho}$, by the same principle as in (3.1).

This results in the corresponding genealogical process,

$$(\widetilde{\Gamma}^{t,\rho}_s, \widetilde{\ell}^{t,\rho}_s)_{s\le t}. \tag{3.4}$$

Remark Suppose now we have constructed X^* and $\xi^{*,\rho}$ as described in Theorem 3. It will turn out that by applying the maps $((\sigma^\rho_{t-s})^{-1})_{0\le s\le t}$ from Theorem 3 we can get an equivalent way to describe this relation. Namely, for $\widetilde{\iota}_1, \widetilde{\iota}_2 \in \widetilde{\mathcal{I}}$, $0\le s\le t$,

$$\widetilde{\iota}_1 \approx^{\xi^{*,\rho}}_s \widetilde{\iota}_2, \qquad \text{iff} \qquad (\sigma^\rho_{t-s})^{-1}\widetilde{\iota}_1 \approx^{X^*}_s (\sigma^\rho_{t-s})^{-1}\widetilde{\iota}_2. \tag{3.5}$$

□

We will now relate the two genealogies for X^* and ξ^* just defined. Consider pairs (Γ,ℓ) of two objects, on one hand partitions Γ of $\mathcal{I}$, and on the other hand corresponding positions, $\ell(\pi)\in G$, where $\pi\in\Gamma$. Suppose now we have given a subset of indices, $\mathcal{I}_1\subset\mathcal{I}$. Then we define the *restriction of* (Γ,ℓ) *on* $\mathcal{I}_1$, denoted

$$\phi_{\mathcal{I}_1}(\Gamma,\ell) = (\phi_{\mathcal{I}_1}(\Gamma),\ell), \tag{3.6}$$

by removing all $\iota\notin\mathcal{I}_1$, and ignoring empty partition elements and their locations.

Then by Theorem 3, we can relate the objects defined in (3.3) and (3.4):

Corollary 3.1. *(Representation of family decomposition)*

$$\sum_{\varpi\in\phi_{\mathcal{I}^\rho}(\Gamma^t_s)} \delta_{(\varpi,\ell^t_s(\varpi))} \overset{d}{=} \sum_{\widetilde{\varpi}\in\widetilde{\Gamma}^{t,\rho}_s} \delta_{(\widetilde{\varpi},\widetilde{\ell}^{t,\rho}_s(\widetilde{\varpi}))}. \tag{3.7}$$

This allows us in many situations to work with the genealogy of the historical look-down process.

The decomposition of X^*_t into family clusters can be analyzed due to a "duality relation" involving a partition valued coalescent with spatial motion which we introduce next.

Step 3 (Coalescent) This process is introduced in two parts, we first introduce a locally finite process and then the process that starts from a locally countable state. Now we turn to the first construction. Fix a countable index set $\mathcal{I}$. Let $\Pi^{\mathcal{I}}$ define the partitions of $\mathcal{I}$. We shall equip this set with a metric (see (3.28)) which generates a topology together with the induced σ-algebra that gives us the needed structure to introduce probability measures on this metric space.

Definition 3.2. *(The $\mathcal{I}$-coalescent) The $\mathcal{E}_G$-*marked $\mathcal{I}$-coalescent*,*

$$(C_t, L_t)_{t\ge0}, \tag{3.8}$$

is a Markov process with values in the marked partitions $\mathcal{P}\in\Pi^{\mathcal{I}}$ of $\mathcal{I}$, where for a partition element $\pi\in\mathcal{P}$, the marks $L(\pi)$ are positions in G. They start at time 0 in the partition $C_0\in\Pi^{\mathcal{I}}$, where the partition elements $\pi\in\mathcal{P}$ are marked by $L_0(\pi)$ such that the corresponding configuration lies in the Liggett-Spitzer space, i.e.,

$$\sum\nolimits_{\pi\in C_0} \pi_G\circ\delta_{\{L_0(\pi)\}}\in\mathcal{E}_G. \tag{3.9}$$

*The dynamics of the $\mathcal{E}_G$-marked $\mathcal{I}$-*coalescent *with coalescence rate γ is given by the following two independent mechanisms:*

- Migration *The marks of the partition elements perform independent rate 1 RW on G with kernel $\bar a(x,y)$, where $\bar a(x,y):=a(y,x)$ denotes the reversed kernel.*
- Coalescence *Each pair of partition elements merges into one partition element at rate γ whenever their marks coincide.*

Remark If $\mathcal{I}$ is finite, then each possible spatial configuration lies in the Liggett-Spitzer space, and therefore the $\mathcal{I}$-coalescent is well-defined. □

A key property of the $\mathcal{E}_G$-marked $\mathcal{I}$-*coalescent* is that the definition can be extended to the setting where the starting configuration has countably many partition elements with the same mark. The restriction $\phi_{\mathcal{I}'}\mathcal{P}$ of a partition $\mathcal{P} \in \Pi^{\mathcal{I}}$ on a subset $\mathcal{I}' \subseteq \mathcal{I}$ is once more defined by removing all $\iota' \notin \mathcal{I}'$, and ignoring empty partition elements.

We now enrich our marked $\mathcal{I}$-coalescent by assigning to each individual a label $u \in \mathcal{U}$. This will allow us in particular to define locally finite coalescents embedded in coalescents which are locally countable at time $t = 0$. Fix a countable subset $\mathcal{U} := \{u^\iota;\, \iota \in \mathcal{I}\}$ of $[0, \infty)$. Once more, let for each $\rho \in [0, \infty)$, $\mathcal{I}^\rho := \{\iota \in \mathcal{I} : u^\iota \leq \rho\}$. We then define:

Definition 3.3. *(The $(\mathcal{I},\mathcal{U})$-coalescent) The $\mathcal{E}_G^{\uparrow}$-marked $(\mathcal{I},\mathcal{U})$-coalescent, $(C_t, L_t)_{t\geq 0}$, is a Markov process with values in marked partitions of $\mathcal{I}$ with marks in G such that for each $\rho \in \mathbb{R}^+$, the restriction $\phi_{\mathcal{I}^\rho}(C_t, L_t) = (\phi_{\mathcal{I}^\rho}(C_t), L_t)$ is a $\mathcal{E}_G$-marked $\mathcal{I}^\rho$-coalescent.*

The following proposition holds under an abstract assumption on the migration space G which is formulated in Condition 6.1 in Subsection 6.1, and which ensures that the group G may be approximated by finite groups.

Proposition 3.4. *(The coalescent is well-defined) Assume that G fulfills Condition 5.1, let $\mathcal{I}$ be finite or countable, and $\mathcal{U} := \{u^\iota : \iota \in \mathcal{I}\}$ be a finite or countable subset of $[0, \infty)$. Then the $\mathcal{E}_G^{\uparrow}$-marked $(\mathcal{I},\mathcal{U})$-coalescent starting in $C_0 = \{\{\iota\};\, \iota \in \mathcal{I}\}$ is well-defined, and for each $t > 0$, the coalescent yields a configuration in the Liggett-Spitzer space, i.e.,*

$$\sum\nolimits_{\pi\in C_t} \pi_G \circ \delta_{\{L_t(\pi)\}} \in \mathcal{E}_G. \tag{3.10}$$

In particular, for each $t > 0$, the coalescent is locally finite.

Remark Since the $\mathcal{E}_G^{\uparrow}$-marked $(\mathcal{I},\mathcal{U})$-coalescent is defined via a projective limit procedure, if (C_t, L_t) is a $\mathcal{E}_G^{\uparrow}$-marked $(\mathcal{I},\mathcal{U})$-coalescent, then for each $\rho \in \mathbb{R}^+$, the restriction $\phi_{\mathcal{I}^\rho}(C_t, L_t)$ on individuals with indices in $\mathcal{I}^\rho$ is the $\mathcal{E}_G$-marked $\mathcal{I}^\rho$-coalescent. That means, the coalescent and the "thinning" procedure commute. □

Step 4 (Strong duality) We obtain a strong duality relation between the ancestral tree of X_t^* and the above spatially structured coalescent as follows. Let (C, L) be the $\mathcal{E}_G^{\uparrow}$-marked $(\mathcal{I},\mathcal{U})$-coalescent. This duality relation requires a particular way of choosing the labels from $\mathcal{U}$. Then we get the following representation.

Proposition 3.5. *(Coalescent representation for the historical look-down) For a fixed time t, choose the initial state of a $(\mathcal{I},\mathcal{U})$-coalescent according to the device specified in the paragraph containing (2.28). Then*

$$\mathcal{L}\left[(\Gamma_s^t, \ell_s^t)_{s\in[0,t]}\right] = \mathcal{L}^{\{(\{\iota\},x^\iota);\,\iota\in\mathcal{I}\}}\left[(C_s, L_s)_{s\in[0,t]}\right]. \tag{3.11}$$

We next establish a relation giving a strong duality for the enriched IMM. Recall that Poisson systems are the equilibria for both: the system of independent random walks with the kernel $a(x, y)$ and the system of independent random walks with the reversed kernel $\bar{a}(x, y)$. Therefore, if the IMM starts in a Poisson system, then the genealogy of the IMM particle process is generated by the coalescent started in a Poisson system as stated next.

Corollary 3.6. *(Strong duality: historical IMM) For a fixed time t, and $\rho \in \mathbb{R}^+$ and initial positions given by locations from a Poisson system with intensity ρ,*

$$\mathcal{L}\left[(\widetilde{\Gamma}_s^{t,\rho}, \widetilde{\ell}_s^{t,\rho})_{s\in[0,t]}\right] = \mathcal{L}^{\{(\{\iota\},x^\iota);\,\iota\in\mathcal{I}\}}\left[\phi_{\mathcal{I}^\rho}(C_s, L_s)_{s\in[0,t]}\right]. \tag{3.12}$$

Remark If ξ^* would not start according to a Poisson configuration, provided that for each $\rho \in \mathbb{R}^+$,

$$\sum_{\iota' \in \mathcal{I}^\rho} \delta_{x^{\iota'}} \stackrel{d}{=} \sum_{\iota' \in \mathcal{I}^\rho} \delta_{\pi^*_G(\xi^{*,\iota'}_t)_t}, \tag{3.13}$$

(3.11) still holds if we replace the reversed kernel $\bar{a}(x,y)$ for the time-reversed migration by the law of a migrating system of particles conditioned on (3.13). However, we would loose time-homogeneity of the migrating particles. □

How does this statement relate to classical moment duality relations? By picking a generic k-tuple of different individuals located at sites x_1, x_2, ..., x_k we see that their genealogy is described by the corresponding finite coalescent. This implies the usual moment duality. Namely, define for ι_1, ι_2, ..., $\iota_k \in \mathcal{I}$, the quantity

$$\mu(x_1, ..., x_k) := \{(\{\iota_1\}, x_1), (\{\iota_2\}, x_2), ..., (\{\iota_k\}, x_k)\}. \tag{3.14}$$

Then we can conclude from (3.12) that:

Corollary 3.7. *(Weak duality: moment duality)*

(a) *Let the IMM start in a Poisson system with total intensity $\rho \in \mathbb{R}^+$, and type 1 intensity $\rho\theta$. Then*

$$\mathbf{E}\left[\prod_{j=1}^k \left(\eta_t(\{x_j\} \times \{1\}) - \sum_{l=j+1}^k 1_{\{x_l = x_j\}}\right)\right] = \rho^k \mathbf{E}^{\mu(x_1,\dots,x_k)}\left[\theta^{|C_t|}\right]. \tag{3.15}$$

(b) *Consequently, for the initial state $\zeta_0(\{x\}) \equiv \theta$, $x \in G$, we have:*

$$\mathbf{E}\left[\prod\nolimits_{j=1}^k \zeta_t(\{x_j\})\right] = \mathbf{E}^{\mu(x_1,\dots,x_k)}\left[\theta^{|C_t|}\right]. \tag{3.16}$$

Remark Notice that correction terms in the moment expression for the diffusion limit disappear since there are locally infinitely many particles which ensures that each particle is sampled at most once, almost surely. □

3.2. **Application I: Equilibria of IMM and IFWD (Theorems 4 and 5).** The aim of this subsection is to give a representation of the *family structure* of the equilibrium historical processes including the look-down, IMM and IFWD, in terms of the *spatially structured coalescent*. This is our first joint application of the representation given in Theorem 3 and of the strong historical duality relation. For that purpose we first recall in (i) the basic ergodic theory of IMM and IFWD with type space $K := \{0,1\}$, and we then state the results in (ii) for IMM and in (iii) for IFWD.

(i) It turns out that the long-time behavior of IMM depends on the strength of interaction between sites x and y described by the *symmetrized migration kernel*, i.e.,

$$\widehat{a}(x,y) := \frac{1}{2}(a(x,y) + a(y,x)). \tag{3.17}$$

We start the IMM with a Poisson configuration of particles with intensity $\rho \in \mathbb{R}^+$ and types 0 and 1 assigned independently with probability $1-\theta$ and θ, respectively (recall $\widetilde{\Psi}(\theta,\rho)$ from (2.29)). (Since it is known that pure random walk systems have as extremal invariant measures the Poisson systems, it would be possible to discuss general translation invariant measures with standard methods, using the knowledge from this special initial states, we omit this here).

If $\widehat{a}(x,y)$ is recurrent, then the duality in (3.12) implies that the system converges locally to one of the monotype configurations. That is,

$$\mathcal{L}^{\widetilde{\Psi}(\rho,\theta)}[\eta_t] \underset{t\to\infty}{\Longrightarrow} (1-\theta)\widetilde{\mathcal{H}}_{\rho,0} + \theta\widetilde{\mathcal{H}}_{\rho,1}, \tag{3.18}$$

where $\widetilde{\mathcal{H}}_{\rho,i}$, $i = 0,1$, are the laws of Poisson fields on $G \times \{0,1\}$ with intensity measures $\rho\, n_i$, and where n_i are the counting measures supported by the subsets $G \times \{i\}$, $i = 0,1$, respectively.

On the other hand, the transience of $\widehat{a}(x,y)$ implies that a coalescent restricted to any set of more than one individual has a number of partition elements converging to a random number which is bigger than one with positive probability. With the same reasoning as in [15] concerning the locations of this coalescent this gives equilibria with coexistence of both types, i.e.,

$$\mathcal{L}^{\widetilde{\Psi}(\rho,\theta)}\left[\eta_t\right] \underset{t\to\infty}{\Longrightarrow} \widetilde{\mathcal{H}}_{\rho,\theta}, \tag{3.19}$$

where $\widetilde{\mathcal{H}}_{\rho,\theta} \in \mathcal{P}(\mathcal{N}(G\times K))$ is a homogeneous, shift ergodic, and invariant distribution with

$$\mathbf{E}_{\widetilde{\mathcal{H}}_{\rho,\theta}}[\eta(\{x\}\times\{k\})] = \begin{cases} (1-\theta)\rho & \text{if } k=0 \\ \theta\rho & \text{if } k=1 \end{cases}. \tag{3.20}$$

The same dichotomy is known to be valid for the IFWD and follows again from the duality. Namely, if $\widehat{a}(x,y)$ is recurrent, then for all translation invariant initial states, Φ_θ, with intensity $\theta\in[0,1]$ the following holds:

$$\mathcal{L}^{\Phi_\theta}\left[\zeta_t\right] \underset{t\to\infty}{\Longrightarrow} (1-\theta)\delta_{\underline{0}} + \theta\delta_{\underline{1}}. \tag{3.21}$$

On the other hand, if $\widehat{a}(x,y)$ is transient, there exist exactly one spatially homogeneous and ergodic probability law with intensity θ, $\widehat{\mathcal{H}}_\theta$. Moreover, for all translation invariant initial states, Φ_θ, with intensity $\theta\in[0,1]$ the following holds:

$$\mathcal{L}^{\Phi_\theta}[\zeta_t] \underset{t\to\infty}{\Longrightarrow} \widehat{\mathcal{H}}_\theta, \tag{3.22}$$

where $\widehat{\mathcal{H}}_\theta$ is an equilibrium measure which is a translation invariant, shift ergodic with intensity θ.

(ii) A similar dichotomy as described in (i) can be observed for the long-term behavior of the historical look-down process, X^*, and hence of the historical processes, $\eta^{*,\rho}$ and ζ^* (recall Definition 2.1 and 2.2).

Let

$$\mathcal{H}^*_\theta(s,t), \qquad \text{and} \qquad \widetilde{\mathcal{H}}^*_{\rho,\theta}(s,t) \tag{3.23}$$

be the law of X^*_t and $\eta^{*,\rho}_t$ started at time s in a collection of constant paths, the constants generated by $\Psi(\theta)$, and by $\widetilde{\Psi}(\rho,\theta)$, respectively (recall (2.28) and (2.29)).

If one is interested in the structure of a process η or ζ in equilibrium, one can first construct the entrance law for that process (by letting the starting time $s\to-\infty$ in (3.23)) to obtain a stationary process with time running from $-\infty$ to $+\infty$ and then consider the corresponding historical process, which we call *equilibrium historical process.*

In order to define the convergence in law for a system described by $\mathcal{H}^*_\theta(s,0)$ as $s\to-\infty$ we need to find the topology which is suited best. Namely, we observe our system at time 0 locally, that is, we pick individuals at that time in a given spatial window and follow their ancestral lines back. This means the labels of the individuals we are interested in are actually random. To resolve this, we assign new indices which still allow us to use the product topology as follows.

Given that the system X^* started in $s\le 0$ and observed at time 0 as described above we define a new system $\widetilde{X}^*$ by reindexing. We assign indices in $G\times\mathbb{N}$ by observing for each path with index ι at time 0 its position in G and the rank of that index (according to the corresponding labels) at this random position. This results in the new system

$$\widetilde{X}^*_t := \{\widetilde{X}^{*,\iota}_t;\ \iota\in G\times\mathbb{N}\}, \tag{3.24}$$

which we can consider as an element of

$$\left(\mathcal{D}(\mathbb{R},G\times K)\times\mathcal{U}\right)^{G\times\mathbb{N}}. \tag{3.25}$$

On the space of paths we have the Skorokhod topology and on $\mathcal{U}$ the discrete topology. Then on the state space we use the product topology of $\left(\mathcal{D}(\mathbb{R},G\times K)\times\mathcal{U}\right)^{G\times\mathbb{N}}$ to define the weak topology of measures on that state space. This induces a natural topology of laws for the original objects.

Note that considering limits of $\mathcal{H}^*_\theta(s,t)$ for $s\to-\infty$ results in objects which are only different by a time shift so that w.l.o.g. we may put $t=0$. The following is easily verified using the duality of Proposition 3.5.

Lemma 3.8. *(Entrance laws of the historical processes) The laws (with w-lim as explained above)*

$$\mathcal{H}^*_\theta := w\text{-}\lim_{s\to-\infty} \mathcal{H}^*_\theta(s,0), \quad \text{and} \quad \widetilde{\mathcal{H}}^*_{\rho,\theta} := w\text{-}\lim_{s\to-\infty} \widetilde{\mathcal{H}}^*_{\rho,\theta}(s,0), \tag{3.26}$$

are well defined, and are supported on paths whose type coordinate is constant.

(iii) In order to prepare for our next theorem we need the long-time behavior of the coalescent. Choose from now on, unless stated otherwise,

$$\mathcal{I} := G \times \mathbb{N}, \tag{3.27}$$

and recall from Definition 3.3 the $\mathcal{E}^{\uparrow}_G$-marked $(\mathcal{I},\mathcal{U})$-coalescent, $(C_t, L_t)_{t\geq 0}$. Obviously, on the set of partitions $\Pi^{\mathcal{I}}$ of $\mathcal{I}$ a partial order is defined by declaring that $\mathcal{P} \leq \mathcal{P}'$ if $\mathcal{P}'$ is a refinement of $\mathcal{P}$, i.e., if the partition elements of $\mathcal{P}$ are obtained by aggregating one or more elements of $\mathcal{P}'$.

Define the distance between two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ of $\mathcal{I}$ by

$$\mathrm{d}(\mathcal{P}_1,\mathcal{P}_2) := \frac{1}{2}\sum_{(\iota_1,\,\iota_2)\in\mathcal{I}\times\mathcal{I}} \alpha_{\iota_1,\,\iota_2}\left\{1_{\{\iota_1\approx_{\mathcal{P}_1}\iota_2,\,\iota_1\not\approx_{\mathcal{P}_2}\iota_2\}} + 1_{\{\iota_1\not\approx_{\mathcal{P}_1}\iota_2,\,\iota_1\approx_{\mathcal{P}_2}\iota_2\}}\right\} \tag{3.28}$$

with (strictly positive) weights normalized such that $\sum_{\iota_1,\,\iota_2}\alpha_{\iota_1,\,\iota_2} = 1$ and $\approx_{\mathcal{P}}$ denoting the equivalence of individuals with respect to the partition $\mathcal{P}$. (Notice that since $\mathcal{I}$ is countable there is a bijection n from $\mathcal{I}$ to $\mathbb{N}$ which gives an order relation on $\mathcal{I}$. The above topology induces a convergence where for any such ordering the partitions restricted to the first k elements converge in law.)

We are going to start in the *maximal partition*, i.e.,

$$(C_0, L_0) := \{(\{\iota\}, x^\iota);\ \iota\in\mathcal{I}\}. \tag{3.29}$$

The dynamics of $(C_t, L_t)_{t\geq 0}$ results in a non-increasing sequence of partitions (with respect to $\leq$) and it is clear that

$$C_\infty := \lim_{t\to\infty} C_t, \qquad \text{exists } a.s. \tag{3.30}$$

Notice that C_∞ differs for transient migration kernel $\hat{a}(x,y)$ from the *minimal partition*, $\{\mathcal{I}\}$, with probability 1, and even still consists of countably many partition elements, while in the recurrent case $C_\infty = \{\mathcal{I}\}$, a.s.

It is now very natural to go beyond the sheer decomposition in family clusters and to ask where the individuals observed at time 0 came from, for a system that was started at time $-\infty$. For that we need to take into account the historical information in the coalescent. We therefore define the *historical coalescent*, $(C^*_t, L^*_t)_{t\geq 0}$, by setting (recall (3.29)):

$$C^*_t := \left(C_{(s\vee 0)\wedge t}\right)_{s\in\mathbb{R}}, \tag{3.31}$$

and $L^* := \{(L^{*,\{\iota\}}_t)_{t\geq 0};\ \iota\in\mathcal{I}\})$ where

$$L^{*,\{\iota\}}_t := \left(L_{(s\vee 0)\wedge t}\left(C^{\{\iota\}}_{(s\vee 0)\wedge t}\right)\right)_{s\in\mathbb{R}}, \tag{3.32}$$

and where the $C^{\{\iota\}}_s$, $s\geq 0$, $\iota\in\mathcal{I}$, denotes the partition element in C_s which contains ι. On the space where L^* is defined we use the product topology and in fact in $\mathcal{D}(\mathbb{R}, G\times K)$ the Skorokhod topology on the space of paths. Taking the product of the topology on pairs of a path of partitions (3.31) and its path through space defines the convergence for the object (C^*_t, L^*_t) as $t\to\infty$.

We can now let $t\to\infty$ to obtain with (3.30) an object we will need in the sequel and which we denote:

$$(C^*_\infty, L^*_\infty). \tag{3.33}$$

The notion of the historical coalescent allows one to even strengthen the duality relation further. Namely we can consider the paths of descent together with the resampling times, and obtain this way the full tree-indexed random walks describing the family relations of a given time t population. Their law is described by the historical coalescent. We will exploit the full strength of this object further in Section 7.

The historical process is then a functional of the historical coalescent. Therefore, in order to describe the equilibrium historical process, it suffices to consider a random functional of the historical coalescent. The extra randomness is used to capture the type structure associated with the population. Hence in the final step we enrich our probability space on which the coalescent is defined by a collection,

$$\{\Delta_\theta(\pi);\ \pi \subseteq \mathcal{I}\}, \tag{3.34}$$

of independent Bernoulli random variables with parameter θ. Now associate with the coalescent tags in $K = \{0,1\}$ which specify the particles' types. Namely

$$\Delta_\theta(C^*_\infty, L^*_\infty) := \left\{ \left(L^{*,\{\iota\}}_\infty, \Delta_\theta((C^{*,\{\iota\}}_\infty)_\infty) \right);\ \iota \in \mathcal{I} \right\}, \tag{3.35}$$

where for each $\iota \in \mathcal{I}$, $C^{*,\{\iota\}}_\infty$ takes values in the set of paths which map $\mathbb{R}$ into subsets of $\mathcal{I}$, and for each $s \in \mathbb{R}$, $(C^{*,\{\iota\}}_\infty)_s$ denotes the partition element of C^*_s which contains ι.

(iv) Now we can formulate our next result. Recall $\Psi(\theta)$ from (2.28), and denote by Ψ the configuration where we ignore the K-valued component, i.e.,

$$\Psi := \{(\text{constant path through } x^\iota, u^\iota);\ \iota \in \mathcal{I}\}. \tag{3.36}$$

We then have in equilibrium a representation of the *law of the paths of descent* in terms of paths of the coalescent as follows:

Theorem 4. *(Representation of equilibrium historical look-down via the coalescent) The look-down equilibrium historical process has the form:*

$$\mathcal{H}^*_\theta = \mathcal{L}^\Psi \left[\Delta_\theta(C^*_\infty, L^*_\infty) \right]. \tag{3.37}$$

This leads to the following dichotomy:

(a) *If $\widehat{a}(x,y)$ is recurrent then $C_\infty = \{\mathcal{I}\}$, a.s. Hence if we consider for θ the two extreme values i which are either 0 or 1, then the $\mathcal{H}^*_\theta$ are clan measures, i.e.,*

$$\mathcal{H}^*_i(\{\textit{for all pairs of paths } \exists\, t_0 :\ \textit{paths agree for } t \le t_0;\ \textit{and have type } i\}) = 1. \tag{3.38}$$

Thus the law of $\Delta_\theta(C^*_\infty, L^*_\infty)$ is a mixture of $\mathcal{H}^*_0$ and $\mathcal{H}^*_1$, i.e.,

$$\mathcal{H}^*_\theta = (1-\theta)\mathcal{H}^*_0 + \theta \mathcal{H}^*_1. \tag{3.39}$$

(b) *On the other hand, if $\widehat{a}(x,y)$ is transient then under each law $\mathcal{H}^*_\theta$ and all $\theta \in [0,1]$ the probability that for any two given paths starting at time 0 there is no $t_0 < 0$ such that the paths agree for all $t < t_0$ is positive. Thus the equilibrium state decomposes in countably many clusters with i.i.d. types.*

Now two questions arise. Firstly, what can one conclude from the representation of the historical look-down process for the equilibrium historical process of the IMM? The second question is related to the following observation. The theorem above gives a global representation of the genealogies of the equilibrium process. If we are only interested in the law of the population which we observe in a local spatial window, can one get more explicit results? To answer these questions we need some preparation.

We introduce the (random) map on the range of the historical coalescent by assigning to each path of $\Delta_\theta(C^*_\infty, L^*_\infty)$ a δ-measure, i.e.,

$$\widetilde{\Delta}_{\rho,\theta}(C^*_\infty, L^*_\infty) := \sum_{\iota \in \mathcal{I}^\rho} \delta_{(L^{*,\{\iota\}}_\infty, \Delta_\theta(C^{*,\{\iota\}}_\infty))}, \tag{3.40}$$

where once more $C^{*,\{\iota\}}_\infty$ denotes the path corresponding to the partition elements of C^*_∞ which contain ι. Recall $E_{A,t}$ from (1.18). For a measure η define $r_A(\eta)$ as a measure

$$r_A(\eta)(\cdot) = \eta(\cdot \cap E_{A,0}), \tag{3.41}$$

which is η restricted to paths which are located in $A \subseteq G$ at time 0. For a measure valued random variable Y taking values in $\mathcal{M}(\mathcal{D}(\mathbb{R}, G \times K))$ with law $\mathcal{L}$ write $\widehat{r}_A\mathcal{L}$ for the law of $r_A(Y)$.

Then we can represent the law of the paths of descent observed in a local window in the equilibrium distribution of IMM as follows.

Corollary 3.9. *(Equilibrium historical IMM and the coalescent) Theorem 4 stays valid if we replace the historical look-down equilibrium, $\mathcal{H}^*_\theta$, by the historical IMM equilibrium, $\widetilde{\mathcal{H}}^*_{\rho,\theta}$, and Δ_θ by $\widetilde{\Delta}_{\rho,\theta}$, respectively.*

Furthermore we get a simple local representation. For each fixed bounded $A \subseteq G$:

$$\widehat{r}_A \widetilde{\mathcal{H}}^*_{\rho,\theta} = \mathcal{L}^\mu \left[\widetilde{\Delta}_{\rho,\theta}(C^*_\infty, L^*_\infty)\right], \tag{3.42}$$

where the initial condition μ arises by putting

$$C_0 = \{\{1\}, \ldots, \{N\}\}, \tag{3.43}$$

N is Poisson distributed with intensity $\rho|A|$, and the initial positions of the N points are uniformly sampled from A.

(iii) Next we use Theorem 3 to obtain the analogue of Theorem 4 for IFWD, i.e. to represent the equilibrium configuration of the IFWD, ζ_∞, and of the historical IFWD, ζ^*_∞.

For this purpose, consider once more the $\mathcal{E}^\uparrow_G$-marked $(\mathcal{I},\mathcal{U})$-coalescent, $(C_t, L_t)_{t\geq 0}$, with (3.29), where $\mathcal{U}$, and $\{x^\iota;\, \iota \in \mathcal{I}\}$ are sampled according to Ψ (recall (3.36)), and let the $\mathcal{E}_G$-marked $\mathcal{I}^\rho$-coalescent be the functional, $\phi_{\mathcal{I}^\rho}(C_t, L_t)$, which is obtained by the *restriction* of (C_t, L_t) to the individuals with index $\iota \in \mathcal{I}^\rho$ (recall (1.24)). Note once more that in particular for this initial state the limit, $C_\infty := \lim_{t\to\infty} C_t$, exists a.s., and recall $\widetilde{\Delta}_{\rho,\theta}$ from (3.40).

Theorem 5. *(Representation of the equilibrium historical IFWD via the coalescent) With the above choices for (C_0, L_0), there exists almost surely a unique random variable $\widetilde{\zeta}^*_\infty \in \mathcal{M}(\mathcal{D}(\mathbb{R}, G \times K))$ such that for each Borel set $A \subseteq \mathcal{D}(\mathbb{R}, G \times K)$,*

$$\widetilde{\zeta}^*_{\theta,\infty}(A) := \lim_{\rho\to\infty} \rho^{-1} \widetilde{\Delta}_{\rho,\theta}(C^*_\infty, L^*_\infty)(A) \qquad a.s. \tag{3.44}$$

Moreover,

$$\mathcal{L}[\widetilde{\zeta}^*_{\theta,\infty}] = w\text{-}\lim_{s\to-\infty} \mathcal{L}^{\underline{\theta},s}[\zeta^*_0], \tag{3.45}$$

where $\mathcal{L}^{\underline{\theta},s}$ denotes the law of a historical Fisher-Wright diffusion started at time $s \leq 0$ in a measure $\zeta^* \in \mathcal{M}(\mathcal{D}(\mathbb{R}, G\times K))$ such that $\zeta^*\{y \in \mathcal{D}(\mathbb{R}, G\times K) : y_t \in \{x\}\times\{1\}\} = 1 - \zeta^*\{y \in \mathcal{D}(\mathbb{R}, G\times K) : y_t \in \{x\}\times\{0\}\} = \theta$ for all $x \in G$ and $t \geq s$.

Remark Note that the above implies that taking the diffusion limit with the IMM-equilibrium historical process gives the IFWD-equilibrium historical process. The limits $\rho \to \infty$ and $t \to \infty$ interchange. □

Corollary 3.10. *(Representation for the equilibrium of the IFWD) The law of the random measure on G, defined by*

$$\widetilde{\zeta}_{\theta,\infty} = \lim_{\rho\to\infty} \rho^{-1} \sum_{\iota\in\mathcal{I}^\rho} \delta_{(L_0^{\{\iota\}}, \Delta_\theta(C_\infty^{\{\iota\}}))}, \qquad a.s. \tag{3.46}$$

is equal to $\widehat{\mathcal{H}}_\theta$ (recall (3.22)).

3.3. **Application II: Finite system scheme (Theorems 6 and 7).** A question of fundamental interest is the relation between the long-time behavior of the IMM and the historical IMM defined on finite but large subsets of an infinite group G in comparison with behavior of the corresponding processes on the whole group G. Since on a finite subset of G, which has the group structure, the migration is always recurrent, at least for transient $\widehat{a}(x, y)$ the long-term behavior differs sharply in the two situations. The first question of interest is therefore *how does the system feel the finiteness of the group (geographical space)*?

The second point is that the determination of this behavior is the key step in the renormalization analysis of this class of models on the level of historical processes. We will not go into more detail on this topic here but refer an interested reader to [30].

Let us return to the first question above. We use as an approach the so-called *finite system scheme* which was used in Cox, Greven and Shiga (1995) [10] to answer the same question for the diffusion limit, IFWD. In that scheme we observe a sequence of finite systems of increasing size on a size dependent time scale. The starting point of these considerations is that if we relate the system size and the time of observation we expect the following: on short time scales (compared to the system size) a large system looks locally like an infinite one, and after a long time (compared to the system size) we see the typical behavior of a finite system. In between some intermediate behavior occurs, which can be described explicitly.

Our particular task here is twofold: (a) we develop the finite system scheme for the historical look-down process and apply our representation of the enriched IMM and the duality with the coalescent to carry out the analysis needed to establish the finite system scheme for the IMM in such a way that we obtain as a consequence the result for IFWD; (b) we extend this kind of analysis to the historical process and obtain the *historical finite system scheme* which will imply the results for the ordinary non-historical case. This construction has not been carried out previously in any model.

We proceed as follows. We construct next in (i) the "finite" systems, then we derive in (ii) the finite system scheme for the historical look-down process, historical IMM and historical IFWD. In part (iii) we formulate some consequences for the ordinary non-historical processes.

(i) In order to keep the arguments transparent and to minimize the notation we focus on the case $G = \mathbb{Z}^d$, even though we could treat fairly general groups including in particular the so-called hierarchical group (compare e.g. with [15]). We also assume from now on that $d \geq 3$, that is, we are in the transient regime.

We begin by defining the finite systems corresponding to IMM on $\mathbb{Z}^d$. Recall the migration kernel $a(x,y)$ from (1.1). Consider the subsets

$$G_N = [-N,N]^d \cap \mathbb{Z}^d, \tag{3.47}$$

and let

$$(\eta_t^N)_{t\geq 0}, \qquad \text{and} \qquad (\eta_t^{*,N})_{t\geq 0}, \tag{3.48}$$

be the IMM and the historical IMM on G_N with random walk kernel

$$a_N(x,y) = \sum_{z\sim y} a(x,z), \qquad x,y \in G_N, \tag{3.49}$$

where $\sim$ denotes equivalence modulo $2N+1$ in each coordinate.

Fix $\theta \in [0,1]$, and $\rho \in \mathbb{R}^+$. In order to avoid certain technicalities we assume for the remaining results that $\mathcal{L}[\eta_0^N]$, and $\mathcal{L}[\eta_0]$ are Poisson systems on $G_N \times K$, and $\mathbb{Z}^d \times K$, with intensity measure $\rho n^N \otimes ((1-\theta)\delta_0 + \theta\delta_1)$, and $\rho n \otimes ((1-\theta)\delta_0 + \theta\delta_1)$, where n^N and n are the counting measures on $G_N \times K$ and $\mathbb{Z}^d \times K$, respectively.

Furthermore, in order to be in a position to have the local central limit theorem for the symmetrized kernel, we restrict our attention to random walk kernels with

$$\sum_x \widehat{a}(0,x)|x|^{d+2} < \infty. \tag{3.50}$$

This way we obtain a sequence of finite systems $\{(\eta_t^N)_{t\geq 0},\ N \in \mathbb{N}\}$ or $\{(\eta_t^{*,N})_{t\geq 0},\ N \in \mathbb{N}\}$, and an infinite system $(\eta_t)_{t\geq 0}$ or $(\eta_t^*)_{t\geq 0}$, respectively, which are related as follows: for every fixed finite time T,

$$\mathcal{L}[(\eta_t^N)_{t\in[0,T]}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(\eta_t)_{t\in[0,T]}], \tag{3.51}$$

and

$$\mathcal{L}[(\eta_t^{*,N})_{t\in[0,T]}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}[(\eta_t^*)_{t\in[0,T]}], \tag{3.52}$$

where here the convergence is weak convergence in $\mathcal{D}([0,T],\mathcal{E}_G)$, and $\mathcal{D}([0,T],\mathcal{E}_G^*)$, respectively.

(ii) To formulate the main result for the historical processes, we need four closely related ingredients: namely, the proper *time scale*, the *resulting limits* of our spatially structured coalescent, the *local equilibria* of our system appearing in that scale and finally the *mixing measure* given the proportions at which the local equilibria appear. We conclude in the final step 5 with formulating the main result.

Step 1 (The critical time scale) The *critical time scale* distinguishing the behavior of the finite systems from the infinite system is given by

$$\beta_N(t) = (2N+1)^d\, t, \qquad d \geq 3, \tag{3.53}$$

rather than the order of the time needed for uniformly started particles to hit the boundary of the torus, N^2 or N (if the underlying random walks have drift), as one might naively guess. The intuitive reason is that in $d \geq 3$ two random walk particles on a large torus, started at randomly sampled points, meet (and therefore interact) only after time of the order of the volume. Observe that two independent particles spend time of order of the inverse volume of the torus in the same site, and therefore their interaction occurs on a scale $\beta_N(t)$.

More precisely, the mean time until they interact is of order κ times the volume (see Lemma 7.3 for a stronger statement), where

$$\kappa := (1/\gamma + g/2)^{-1}, \tag{3.54}$$

and g is the expected number of returns to 0 by the random walk on $\mathbb{Z}^d$. Note that $g < \infty$ if and only if $\widehat{a}(x,y)$ is transient. The constant is related to the resampling rate and the chance for two fixed individuals to be related. The following claim is standard and easy.

Lemma 3.11. *(Interpretation of κ) Let $\iota_1, \iota_2 \in \mathcal{I}$, and let (C,L) be the $\{\iota_1, \iota_2\}$-coalescent (with delay rate γ). Then*

$$\kappa = \gamma\, \mathbf{P}^{\{(\{\iota_1\},0),(\{\iota_2\},0)\}}[C_\infty = \{\{\iota_1\},\{\iota_2\}\}], \tag{3.55}$$

where C_∞ is defined in (3.30).

Step 2 (Kingman's coalescent) To describe the asymptotic behavior of the spatially structured coalescent processes we will need another coalescent process, namely the well-known *Kingman's coalescent*, $K := (K_t)_{t\geq 0}$. This process is a special case of the $\mathcal{I}$-coalescent where we identify all sites in G, that is, we use the site space with the trivial group $\{0\}$. The Kingman coalescent can therefore equivalently be described as a process with values in the partitions of $\mathcal{I}$ which is the entrance law (at time $t = 0$) of a Markov process with the following dynamics:

Kingman's coalescence Each pair of partition elements coalesces at rate 1.

We write

$$K_t(C) \tag{3.56}$$

for the Kingman's coalescent started in the partition C of $\mathcal{I}$.

This object arises in our model as follows. Consider the genealogies of two individuals, started at the origin, in the process on the (large) torus. Their ancestral paths try to coalesce as they would do on $\mathbb{Z}^d$. If they do not succeed they finally drift apart, and start wrapping around the torus. This probability is given by

$$\mathbf{P}^{\{(\{\iota_1\},0),(\{\iota_2\},0)\}}[C_\infty = \{\{\iota_1\},\{\iota_2\}\}] = (\gamma g/2 + 1)^{-1}. \tag{3.57}$$

Given that our two individuals have not coalesced without wrapping around, they have forgotten all the information about their initial spatial distance, they are both uniformly distributed on the torus. On times of the order of the volume they come to a common point again and then their coalescence dynamics restarts. The time intervals during which the particles are at distance $O(1)$ from each other are very short when compared to the time intervals until wrapping around. This means that they now try to coalesce (on the critical time scale) at rate γ according to the Kingman's coalescent.

Step 3 (The local equilibria on the scale $\beta_N(t)$) Above we have indicated the asymptotics for the genealogical relation of two individuals. Our goal is to understand the asymptotics of the full genealogical structure. For that purpose we come to the third ingredient, the collection $\mathcal{H}^*_{1,\theta}(C,\mathcal{P})$ of measures with $C,\mathcal{P}$ running over partitions.

Suppose we are given the family structures C which can be observed based on spatial paths before wrapping around the torus, and that we are also given $\mathcal{P}$ describing at the same instant the coarser picture based on the ancestral relations due to the wrapping around the torus. Conditioned on these observations the laws of the finite historical look-down processes observed on scales $\beta_N(t)$ look approximately like a law $\mathcal{H}^*_{1,\theta}(C,\mathcal{P})$ and unconditioned the state of the system at $\beta^N(t)$ has approximately a law which is the mixture (over $C,\mathcal{P}$) of such laws. It turns out that the $\mathcal{H}^*_{1,\theta}(C,\mathcal{P})$ arise as the entrance law of the historical look-down process at $-\infty$ if we condition the entrance configuration to have a particular family structure C and type structure induced by $\mathcal{P}$ (coarser than C). This suggests the following structure, which we introduce in definition 3.12 formally.

For that purpose we shall need here and in the sequel frequently the operator τ^*_t which is defined by the requirement that it is induced on measures on path space by the time-shift operator on paths, which maps a path y as follows:

$$(\tau^*_t y)_\cdot := y_{(\cdot - t)}. \tag{3.58}$$

Definition 3.12. *(Clan measures conditioned on family structures) Given a partition C of $G\times\mathbb{N}$, and $\theta\in[0,1]$, pick a partition $\mathcal{P}$ coarser than C with $|\mathcal{P}|<\infty$ and proceed as follows (Recall (3.3), (3.26), (3.34) for definitions).*

- *Condition $\mathcal{H}^*_1$ on having $\Gamma^0_{-\infty}=C$. Fix a regular version of this conditional law.*
- *Take each $\mathcal{D}(\mathbb{R}\times\{1\}\times\mathcal{U})^{G\times\mathbb{N}}$-valued realization according to the above (conditional) law, and reset in each family given by the partition $\mathcal{P}$ the type equal to $\Delta_\theta(\mathcal{P})$.*
- *Now consider the configuration induced on $(\mathcal{D}(\mathbb{R},G\times K)\times\mathcal{U})^{G\times\mathbb{N}}$ by this random map.*
- *Associate with $(C,\mathcal{P})$ the law of the above random collection.*

This whole procedure results in a measurable, probability measure-valued map:

$$(C,\mathcal{P})\mapsto\mathcal{H}^*_{1,\theta}(C,\mathcal{P}), \tag{3.59}$$

where the measure on the right hand side is a measure on the σ-algebra of the space of collections of labeled paths, i.e. $(\mathcal{D}(\mathbb{R},G\times K)\times\mathcal{U})^{G\times\mathbb{N}}$.

Remark We introduce the above object with the following picture in mind. Suppose we observe the system at time $\beta_N(t)$. Recalling the relationship we described between two individuals we have to realize that due to wrapping around the torus common ancestors occur on *two* different scales. Individuals are either unrelated or related due to resampling events which occur either in the time interval $[\beta_N(t)-L(N),\beta_N(t)]$, here $L(N)\uparrow\infty$ and $L(N)=o(N)$, or during the time interval $[0,\beta_N(t)-M(N)]$ (with $M(N)\uparrow\infty$ and $N^2\log N\ll M(N)\ll N^d$ as $N\to\infty$), or during the time interval $[\beta_N(t)-M(N),\beta_N(t)-L(N)]$ in which wrapping around the torus occurs. The first type of families will look exactly like families of the equilibrium historical process on the whole group, the second type of families will have a distribution which is spatially a superposition of independent versions of the first kind, and the individuals involved in resampling events of the third kind above will asymptotically become negligible. To give the whole picture we therefore still need a way to describe the second type of families. □

Step 4 (Mixing measure) Having defined the local limiting law for given family and type structure we come to the fourth ingredient. The following theorem states that the laws of the finite historical look-down processes observed on scales $\beta_N(t)$ look like a mixture over $(C,\mathcal{P})$ of the laws $\mathcal{H}^*_{1,\theta}(C,\mathcal{P})$. The *mixing measure*, $\bar{Q}^*_{\kappa t}$, is given by the joint law of the two coupled random variables C_∞ and Kingman's coalescent K at time κt and started at time 0 in state C_∞, i.e.,

$$\bar{Q}^*_t[\mathrm{d}(C,\mathcal{P})] := \mathbf{P}[C_\infty\in\mathrm{d}C;\ K_t(C_\infty)\in\mathrm{d}\mathcal{P}]. \tag{3.60}$$

Observe that $\bar{Q}_t^*[\mathrm{d}(\cdot,\cdot)]$ is supported on those pairs of partitions $(C, \mathcal{P})$, where $\mathcal{P}$ has *finitely many* partition elements for $t > 0$ and is always *coarser* than C.

Step 5 (Main result) Now we have all four ingredients needed to formulate the next theorem saying that on the given time scale the historical configurations consist of the clans of the infinite system but the types are kept equal in $\mathcal{P}$, which means they are not assigned simply as i.i.d. random variables to each element in C (as in the infinite system) but rather some of them are coupled as follows:

Theorem 6. *(Finite systems scheme for historical look-down) Let $\theta \in [0,1]$, $t \in [0,\infty]$, and $t(N) \uparrow \infty$ such that $t(N)/|G_N| \to t$, as $N \to \infty$. Then*

$$\mathcal{L}^{\Psi(\theta),0}\left[\tau^*_{t(N)} X^{*,N}_{t(N)}\right] \underset{N\to\infty}{\Longrightarrow} \int \bar{Q}^*_{\kappa t}[\mathrm{d}(C,\mathcal{P})]\, \mathcal{H}^*_{1,\theta}(C,\mathcal{P}). \tag{3.61}$$

Remark The result says that as macroscopic time increases in the critical time scale, the path of descent stops determining the genealogical relation alone, as time progresses the family structure on a large torus is made coarser compared to the infinite system via applying a Kingman coalescent which forces more and more "family clusters" in the infinite system to have the same type. □

Remark The scenario above can be even stated in more detail with stronger statements which also bring into view the build-up of the genealogies through the various degrees of kinship. See for details Proposition 7.6 in Subsection 7.3. □

As before we can also derive from this theorem on the look-down process X^*, results on η^* and also on $\xi^{*,\rho}$, which finally allows us to obtain results on ζ^*.

We first consider the historical IMM, i.e. η^*. Here we need a new object for the description replacing $\mathcal{H}^*_{1,\theta}(C,\mathcal{P})$ above, which will be denoted by $\widetilde{\mathcal{H}}^*_{\pi,\rho,1}(C)$. In order to understand the sequel recall the discussion at the beginning of Step 3 and the following fact. The weak topology of probability measures on paths on larger and larger tori leads to a convergence of end-pieces of arbitrary length but we never catch properties of the path after wrapping around. We distinguished three *degrees of kinship*, one of which is asymptotically negligible and we the other two prevail in the limit. The first degree of kinship occurs when two individuals have coalesced in near past, and so their ancestral paths have a common beginning piece, while the second degree of kinship occurs when two individuals managed to coalesce with the help of wrapping around the torus, so that even though their types agree one can not conclude their relation by observing (in the limit) their ancestral paths only. (The reader should be aware that there is a grey zone in which individuals manage to coalesce without wrapping around but still long time ago so that one can not conclude their relation by observing their ancestral paths only. However, after rescaling the coalescent events that occur during this time period will not contribute substantially.)

Now C runs through the set of partitions of $\mathcal{I}^\rho$ and should be read as the final partition restricted to individuals with label smaller than ρ where equivalence is with respect to first degree of kinship, while for a given C, π runs through the set of unions of partition elements of C and plays the role of a family patch in the rougher partitions than C where equivalence is with respect to the second degree of kinship. We derive $\widetilde{\mathcal{H}}^*_{\pi,\rho,1}(C)$ from $\mathcal{H}^*_{1,i}$, $i \in \{0,1\}$, by decomposing the random measures induced by the individuals in $\mathcal{I}^\rho$ into independent components. Precisely, consider the representation of $\widetilde{\mathcal{H}}^*_{\rho,i}$, $i = 0,1$, (recall (3.32)) by the coalescent. Let for $\rho \in \mathbb{R}^+$, C be a partition of $\mathcal{I}^\rho$, $\mathcal{P}$ be a coarser partition than C which satisfies $|\mathcal{P}| < \infty$, let $\pi \in \mathcal{P}$, and $i \in \{0,1\}$. Then carry out the following steps. (i) Consider under $\widetilde{\mathcal{H}}^*_{\rho,i}$ a regular version of the conditional distribution given $\Gamma^0_{-\infty} = C$, for each C this is a law on measures on $\mathcal{D}(\mathbb{R}, G \times K)$. (ii) When forming the empirical measure on path space, use only individuals with index $\iota \in \pi$. We denote the version of the conditional law taken in (i) resulting from this restriction by $\widetilde{\mathcal{H}}^*_{\pi,\rho,i}(C)$.

By the strong form of duality, we have the alternative description of $\widetilde{\mathcal{H}}^*_{\pi,\rho,i}(C)$:

$$\begin{aligned}&\widetilde{\mathcal{H}}^*_{\pi,\rho,i}(C)\\ &:= \mathcal{L}^{\Psi}\left[\sum_{\iota\in\pi\cap\mathcal{I}^\rho}\delta_{(L^{*,\{\iota\}}_\infty,\underline{i})}\Big|\forall\,\iota_1\approx^C_\infty\iota_2\ \ \exists\, t\geq 0:\ (L^{*,\{\iota_1\}}_\infty)_{s\geq t}=(L^{*,\{\iota_2\}}_\infty)_{s\geq t}\right],\end{aligned} \tag{3.62}$$

where $\underline{i}$ denotes the path through K which is constant $i\in K$. Observe that the map $(C,\mathcal{P})\mapsto \ast_{\pi\in\mathcal{P}}(\theta\tilde{\mathcal{H}}^*_{\pi,\rho,1}(C)+(1-\theta)\tilde{\mathcal{H}}^*_{\pi,\rho,0}(C))$ where $\ast$ denotes convolution is measurable.

In particular, we conclude then from Theorem 6 the following for η^*:

Corollary 3.13. *(Finite systems scheme for historical IMM) Let $\theta\in[0,1]$, $t\in[0,\infty]$, and $t(N)\uparrow\infty$ such that $t(N)/|G_N|\to t$, as $N\to\infty$. Then*

$$\mathcal{L}^{\Psi(\theta,\rho),0}\left[\tau^*_{t(N)}\eta^{*,N}_{t(N)}\right]\underset{N\to\infty}{\Longrightarrow}\int\bar{Q}^*_{\kappa t}[\mathrm{d}(C,\mathcal{P})]\left\{\underset{\pi\in\mathcal{P}}{\ast}\Big(\theta\widetilde{\mathcal{H}}^*_{\pi,\rho,1}(C)+(1-\theta)\widetilde{\mathcal{H}}^*_{\pi,\rho,0}(C)\Big)\right\}. \tag{3.63}$$

Analogously, as a second consequence of Theorem 6 we obtain a statement about large finite systems of historical IFWD. We first need to define the limiting distributions appearing, i.e., we need an analogue of (3.62). Since this object involves taking a limit, we cannot (and need not) define it for every pair $(C,\mathcal{P})$ but only for those pairs which appear as $K_t(C)$ and hence are due to an *exchangeable* selection among the elements of C.

Lemma 3.14. *(Conditioned historical IFWD equilibria) Let (C^*_∞, L^*_∞) be as introduced in (3.33). For a partition C of $\mathcal{I}$, take an exchangeable random selection (independent of everything else) of a partition $\mathcal{P}$ coarser than C with $|\mathcal{P}|<\infty$. Take $\pi\subseteq\mathcal{I}$ such that π is element of $\mathcal{P}$. Then the following limit exists a.s. w.r.t. the law of C^*_∞ and the random selection:*

$$\lim_{\rho\to\infty}\rho^{-1}\sum\nolimits_{\iota\in\pi\cap\mathcal{I}^\rho}\delta_{(L^{*,\{\iota\}}_\infty,\underline{i})}, \tag{3.64}$$

where once more $\underline{i}$ denotes the path through K which is constant $i\in K$. Then consider for $(C,\mathcal{P})$ and every $\pi\in\mathcal{P}$ a version of

$$\mathcal{L}\left[\lim_{\rho\to\infty}\rho^{-1}\sum\nolimits_{\iota\in\pi\cap\mathcal{I}^\rho}\delta_{(L^{*,\{\iota\}}_\infty,\underline{i})}\Big|\forall\,\iota_1\approx^C\iota_2\ \ \exists\, t\geq 0\,\forall\, s\geq t:\ (L^{*,\{\iota_1\}}_\infty)_s=(L^{*,\{\iota_2\}}_\infty)_s\right], \tag{3.65}$$

which is denoted by $\widehat{\mathcal{H}}^*_\pi$. Assigning to each path either type 0 or 1 results in $\widehat{\mathcal{H}}^*_{\pi,i}$ with $i=0,1$. This results in the transition kernel, say K:

$$K((C,\mathcal{P}),\cdot)=\underset{\pi\in\mathcal{P}}{\ast}\Big(\theta\widehat{\mathcal{H}}^*_{\pi,1}(C)+(1-\theta)\widehat{\mathcal{H}}^*_{\pi,0}(C)\Big)(\cdot). \tag{3.66}$$

Applying the fact that $\mathcal{P}$ can be selected through Kingman's coalescent $K_{\kappa t}$ yields the following.

Theorem 7. *(Finite systems scheme for historical IFWD) Let $\theta\in[0,1]$, $t\in[0,\infty]$, and $t(N)\uparrow\infty$ such that $t(N)/|G_N|\to t$, as $N\to\infty$. Then we have:*

$$\mathcal{L}^{\underline{\theta},0}\left[\tau^*_{t(N)}\zeta^{*,N}_{t(N)}\right]\underset{N\to\infty}{\Longrightarrow}\int\bar{Q}^*_{\kappa t}[\mathrm{d}(C,\mathcal{P})]\left(\underset{\pi\in\mathcal{P}}{\ast}\Big(\theta\widehat{\mathcal{H}}^*_{\pi,1}(C)+(1-\theta)\widehat{\mathcal{H}}^*_{\pi,0}(C)\Big)\right). \tag{3.67}$$

(iii) So far we have considered the historical processes. However, by projection of the historical process at time t on the time t configuration, the results on historical processes can be specialized to yield results on the configurations η_t and ζ_t. These already appear in the literature on particle systems [5] and [8] and interacting diffusions [10]. Let us revisit these results from a perspective of the historical look-down construction.

By projecting on the time t configuration the only relevant information should sit in the density of types and in the equilibria of the system η and ξ, respectively. To derive this from our theorem the key structural element is *exchangeability*. Namely start observing that given C_∞, the random partition $K_{\kappa t}(C_\infty)$ is exchangeable for each $t > 0$, that is, its distribution is invariant under renumbering finitely many of the indices. So the same holds for the collections of $\{0,1\}$-random variables $\{\Delta_\theta(K^\pi_{\kappa t}(C_\infty));\, \pi \in C_\infty\}$ and $\{\Delta_\theta(K^\pi_{\kappa t}(C_\infty));\, \pi \in \phi_{\mathcal{I}^\rho} C_\infty\}$ (recall (3.34)).

Hence by de Finetti's theorem, given C_∞, $\{\Delta_\theta(K^\pi_{\kappa t}(C_\infty));\, \pi \in C_\infty\}$ is a random *mixture* of i.i.d. $\{0,1\}$-random variables.

If $\widehat{a}(x,y)$ is transient, C_∞ is countably infinite, i.e., $C_\infty := \{\pi_1, \pi_2, ...\}$. Therefore, given C_∞, the frequency of 1's in the array, denoted z_t, arises as:

$$z_t := \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \Delta_\theta(K^{\pi_k}_{\kappa t}(C_\infty)), \qquad a.s., \tag{3.68}$$

and

$$z^\rho_t := \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \Delta_\theta(\phi_{\mathcal{I}^\rho}(K^{\pi_k}_{\kappa t}(C_\infty))), \qquad a.s., \tag{3.69}$$

respectively.

From the construction of Kingman's coalescent via the genealogy of a (single) Fleming-Viot process, which is the non-spatial and non-historical version of our strong duality, it is well-known (see Proposition 3.3 in [19]) that the laws of z_t and z^ρ_t are the same and equal to the law of the Fisher-Wright diffusion (FWD) with resampling rate κ, started in θ. A FWD is the unique strong solution of the stochastic initial value problem

$$dz_s = \sqrt{\kappa\, z_s(1-z_s)}dw_s; \qquad z_0 = \theta \in [0,1], \tag{3.70}$$

where $w = (w_s)_{s\geq 0}$ is a standard Brownian motion on the real line.

We denote the transition kernel of FWD by

$$Q_t[\theta, \mathrm{d}\,\boldsymbol{\cdot}] := \mathbf{P}^\theta[z_t \in \mathrm{d}\,\boldsymbol{\cdot}]. \tag{3.71}$$

In order to apply above observations to obtain the finite system scheme for η and ζ we need one more observation. The fact that on the critical time scale the number of families which are building up the equilibrium is finite but random, is reflected at a fixed time t in the phenomenon that *the empirical density*, $D^{\rho,N} \in [0,1]$, is fluctuating at that scale. Let for any $x \in G_N$, σ^N_x denote the image measure induced by the shift by x, that is,

$$(\sigma^N_x \eta)(y,k) = \eta(z,k), \qquad z = (x+y) \mod (2N+1). \tag{3.72}$$

Then define

$$D^{\rho,N}(\eta^N) := (\rho\,|G_N|)^{-1} \sum_{x\in G_N} \left(\sigma^N_x \eta^N\right)\left((0,1)\right), \qquad t \geq 0. \tag{3.73}$$

Recall $\widetilde{\Psi}(\theta,\rho)$ from (2.29). We are now in a position to formulate the following:

Corollary 3.15. *(Finite system scheme; IMM and IFWD) Assume that $\widehat{a}(x,y)$ is transient.*

(a) *Then given the density the system equilibrates locally into the same (in law) state as the infinite system, i.e.,*

$$\mathcal{L}^{\widetilde{\Psi}(\theta,\rho)}\left[\eta^N_{\beta_N(t)}\right] \underset{N\to\infty}{\Longrightarrow} \int_{[0,1]} Q_t[\theta, d\theta']\, \widetilde{\mathcal{H}}_{\rho,\theta'}, \tag{3.74}$$

and

$$\mathcal{L}^{\underline{\theta}}\left[\zeta^N_{\beta_N(t)}\right] \underset{N\to\infty}{\Longrightarrow} \int_{[0,1]} Q_t[\theta, d\theta']\, \widehat{\mathcal{H}}_{\theta'}. \tag{3.75}$$

(b) *The empirical density observed at the macroscopic time scale satisfies*

$$\mathcal{L}^{\widetilde{\Psi}(\theta,\rho)}\left[(D^{\rho,N}(\eta^N_{\beta_N(t)}))_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{\theta}\left[(z_t)_{t\geq 0}\right], \tag{3.76}$$

and

$$\mathcal{L}^{\underline{\theta}}\left[(D^{1,N}(\zeta^N_{\beta_N(t)}))_{t\geq 0}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{\theta}\left[(z_t)_{t\geq 0}\right]. \tag{3.77}$$

Remark This means the empirical density fluctuates in the scale $\beta_N(t)$ like the FWD and the local configuration is now coupled to that macroscopic variable and is approximately equal to the equilibrium corresponding to the current density. □

4. Wellposedness of the historical martingale problems

In this section we establish the well-posedness of the martingale problems for ξ^*, X^*, η^*, and ζ^* stated in Subsection 2.1. We prove the existence for the ξ^*, X^* and η^* in Subsection 4.1. The uniqueness statements together with the existence for the diffusion limit ζ^* are dealt with in in Subsection 4.2.

4.1. **The graphical construction of the enriched historical processes.** The key in proving the *existence* of a solution to the above martingale problems for $X^*, \xi^{*,\rho}$ and $\eta^{*,\rho}$ is to verify that the look-down construction we gave in Subsection 2.2 is well-defined and then check that the construction provides indeed a solution for the X^*-problem and induces as well solutions for the $\xi^{*,\rho}$ and $\eta^{*,\rho}$ problems. The existence of the solution of the ζ^* martingale problem is difficult since ζ^* arises as diffusion limit and therefore existence and uniqueness (approximations converge) are closely related. We therefore discuss this in Subsection 4.2 where we prove uniqueness to the martingale problems.

We first give in Subsection 4.1.1 the main tool, Subsection 4.1.2 contains the arguments showing that construction is well-defined and that the process constructed is a solution to the martingale problem.

4.1.1. *The graphical point of view.* In order to show that the look-down process is well-defined, we shall make use of a *graphical representation* in the spirit of Harris' graphical construction of the voter model (compare with [33]). Consider a random space-time diagram on $\mathcal{I}\times[0,\infty)$. Let $\mathrm{gen}(X^*)$ be a point process on $\mathcal{I}\times\mathcal{I}\times[0,\infty)$ satisfying the requirement that the random points should represent the information at which times and from whom the individuals adopt a type. Therefore we define

$$(\iota_1,\,\iota_2,t)\in\mathrm{gen}(X^*) \quad \text{iff} \quad B^{\iota_1,\,\iota_2}[0,t]=B^{\iota_1,\,\iota_2}[0,t)+1. \tag{4.1}$$

For $(\iota_1,\,\iota_2,t)\in\mathrm{gen}(X^*)$ we draw an *arrow* from (ι_1,t) to (ι_2,t) in our space-time diagram describing the "flow" of types.

From $\mathrm{gen}(X^*)$ we can then recover a family relationship by reversing the perspective. We call the individual ι_1 at time t_1 *related* to the individual ι_2 at time $t_2\geq t_1$ if we can reach (ι_1,t_1) by moving downwards along vertical lines starting at (ι_2,t_2), jumping to the tail of an arrow, whenever we encounter its head.

Notice that the look-down process is well-defined iff we can attach with each individual ι at time $t\geq 0$ a type in a unique way, which means actually that ι at t is related to exactly one individual, $A_{(\iota,t)}(0)$, at time 0. The latter will be shown in the next subsection.

4.1.2. *The look-down construction is well defined.* Recall from (1.30) to (1.32) the notation for the time points of look-down events.

In order to ensure that the construction of X^* makes sense we have to show that:

Proposition 4.1. *(Look-down construction is well-defined)*

(a) (i) *For each $\iota\in\mathcal{I}$, $\gamma^1_\iota>0$, $\mathbf{P}$-a.s.*
 (ii) *For each $\iota_1,\,\iota_2\in\mathcal{I}$, and $n\geq 1$, $\gamma^n_{\iota_1,\,\iota_2}>\tau^n_{\iota_1,\,\iota_2}$, $\mathbf{P}$-a.s.*
 (iii) *For each $\iota\in\mathcal{I}$, and $t>0$, there is a finite number $n=n(t,\,\iota,\omega)\in\mathbb{N}$ such that $\gamma^n_\iota\leq t<\gamma^{n+1}_\iota$.*

(b) *In particular, X^* is well-defined by (1.33) and (1.34).*

(c) *$\mathcal{L}(X^*)$ solves the $(L_{(t,X^*)},\mathcal{A}^*)$-martingale problem.*

(d) *Furthermore ξ^* and η^* are well-defined by the construction following (1.38).*

Proof **(a)** Since for $(\iota_1, \iota_2) \in \mathcal{I} \times \mathcal{I}$, B^{ι_1, ι_2} can jump only if the two individuals with index ι_1 and ι_2 are sharing the same site, it is sufficient to prove that up to each fixed time horizon a tagged particle meets a.s. only finitely many other individuals with a smaller label.

Fix $\iota \in \mathcal{I}$, and denote by $\chi^\iota := (\chi^\iota_t)_{t \geq 0}$ the random field which counts for each site $x \in G$ the current number of individuals with a label $\leq u^\iota$ regardless of their types, i.e.,

$$\chi^\iota_t(\boldsymbol{\cdot}) := \sum_{\iota' \in \mathcal{I};\, u^{\iota'} < u^\iota} 1\{R^{\iota'}(x^{\iota'}, t) \in \boldsymbol{\cdot}\}. \tag{4.2}$$

For a given $t > 0$, we follow the individual with index ι in reversed time. The migration of an individual with a specified label is independent of the evolution of all other particles. This means, the number of individuals with a smaller label which the individual with index ι meets along the way is stochastically smaller than the following object. Take a particle, say started in $y \in G$, which is wandering as a reversed random walk, $\bar{R}$, where $\bar{R}$ arises as R but with transition kernel $\bar{a}(x, y) := a(y, x)$ through a fluctuating configuration evolving as $(\chi^\iota_{t-s})_{s \geq 0}$ and observe the number of particles observed from the perspective of this particle. Then

$$\begin{aligned} \mathbf{P}[\gamma^1_\iota > 0] &= \lim_{t \downarrow 0} \mathbf{P}[\gamma^1_\iota > t] \\ &\geq \lim_{t \downarrow 0} \mathbf{E}\left[\exp -\left\{\gamma \int_0^t \mathrm{d}s\, \chi^\iota_{t-s}(\bar{R}_s)\right\}\right]. \end{aligned} \tag{4.3}$$

It suffices to show that the above time integral goes to zero in probability. Observe that for each $\iota \in \mathcal{I}$, $y \in G$,

$$\begin{aligned} \mathbf{E}^y[\chi^\iota_{t-s}(\bar{R}_s)] &= \sum_{z \in G} \mathbf{P}^y[\bar{R}_s = z] \mathbf{E}[\chi^\iota_{t-s}(\{z\})] \\ &= \sum_{\iota' \in \mathcal{I};\, u^{\iota'} < u^\iota} \sum_{z \in G} \bar{a}_s(y, z) a_{t-s}(x^{\iota'}, z) \\ &= \sum_{x \in G} \chi^\iota_0(\{x\}) a_t(x, y). \end{aligned} \tag{4.4}$$

The right hand side of (4.4) is finite since we know by assumption that the initial state is in the Liggett-Spitzer space, and hence

$$\sum_{y \in G} \mathbf{E}^y[\chi^\iota_{t-s}(\bar{R}_s)] \alpha(\{y\}) \leq e^{(\Gamma - 1)t} \sum_{x \in G} \chi^\iota_0(\{x\}) \alpha(\{x\}) < \infty. \tag{4.5}$$

Therefore $\int_0^t \mathrm{d}s\, \chi^\iota_{t-s}(\bar{R}_s) \downarrow 0$ in probability, as $t \downarrow 0$. Thus the right hand side of (4.3) is equal to 1.

The proofs of (ii) and (iii) can be based on similar reasoning which we skip here.

(b) Recall from (4.1) the random point process, $\mathrm{gen}(X^*)$. We start by showing that each individual ι at time t is related with exactly one individual at time 0, $\mathbf{P}$-a.s. Notice that we can find all individuals at some time $s \leq t$ related to ι at time t by moving downwards along the diagram.

For $\iota \in \mathcal{I}$, $t \geq 0$, let

$$\bar{\beta}^1_{(\iota,t)} := \min\{s > 0 : \exists\, \bar{\iota}^1_{(\iota,t)} \in \mathcal{I} \text{ s.t. } (\bar{\iota}^1_{(\iota,t)}, \iota, t - s) \in \mathrm{gen}(X^*)\}. \tag{4.6}$$

By part (a)(ii) $\bar{\iota}^1_{(\iota,t)}$ is uniquely determined, $\mathbf{P}$-a.s.

We therefore can define inductively $(\bar{\iota}^n_{(\iota,t)})_{n \in \mathbb{N}}$ and $(\bar{\beta}^n_{(\iota,t)})_{n \in \mathbb{N}}$ by

$$\bar{\beta}^{n+1}_{(\iota,t)} := \bar{\beta}^1_{(\bar{\iota}^n_{(\iota,t)}, t - \bar{\beta}^n_{(\iota,t)})}. \tag{4.7}$$

Once more, this definition makes sense on a set $C_{(\iota,t)}$ of full $\mathbf{P}$-measure. Obviously, for each $s \in [t - \bar{\beta}^{n+1}_{(\iota,t)}, t - \bar{\beta}^n_{(\iota,t)})$ the individual ι at time t is related to the individual $\bar{\iota}^n_{(\iota,t)}$ at time s. Moreover, $C_{(\iota,t)}$ can be chosen such that there exists a finite number,

$$m_{(\iota,t)}(0) \in \mathbb{N}, \tag{4.8}$$

referred to as the *degree of kinship* between

$$A_{(\iota,t)}(0) := \bar{\iota}_{(\iota,t)}^{m_{(\iota,t)}(0)} \tag{4.9}$$

at time 0 and ι at time t, such that there is no individual of a smaller label than $A_{(\iota,t)}(0)$ which is related with ι at some (non-negative) time. Hence, on $C_t := \cap_{\iota \in \mathcal{I}} C_{(\iota,t)}$, we may define

$$k^{\iota}(t) = k^{A_{(\iota,t)}(0)}, \tag{4.10}$$

and for $s \in [t - \bar{\beta}_{(\iota,t)}^{n+1}, t - \bar{\beta}_{(\iota,t)}^{n})$, $\bar{\beta}_{(\iota,t)}^{0} := 0$, and $0 \le n \le m_{(\iota,t)}(0)$,

$$(\pi_G^* X_t^{*,\iota})_s := R^{\bar{\iota}^n_{(\iota,t)}}(x^{\bar{\iota}^n_{(\iota,t)}}, s). \tag{4.11}$$

The latter fulfills obviously (1.34). Thus the given construction of X^* is well-defined.

(c) Next we deal with the problem of *existence* of a solution to the martingale problem. Observe that the existence of the historical look-down follows by verifying that indeed the process $(X_t^*)_{t \ge 0}$, which we constructed explicitly above satisfies the properties required by the martingale problem. This is a straightforward explicit calculation with the generator which we omit here. Recall that it is at this point that the initial state has to be chosen with some care. The tools for that are developed in the theory of interacting particle systems, namely in order to avoid explosion of rates we have to use the Liggett-Spitzer space $\mathcal{E}_G^*$ as we explained in detail in Subsection 1.3 (recall (1.16)).

(d) It remains to show that also ξ^* and η^* are well-defined by the construction we gave and that the solutions satisfy the respective martingale problems. This however is essentially the same as above, since η^* is a functional of ξ^* and in the construction of ξ^* the only difference to X^* is the symmetrization of the resampling mechanism which does not change the estimates which involve solely the random walks. □

4.2. **Proof of Theorems 1 and 2.** Here we have to establish the well-posedness of the historical martingale problem for the IMM, the look-down, the enriched IMM, and the IFWD. Based on these facts we show that the historical IFWD is the diffusion limit of the historical IMM. With the previous work this boils down to proving *uniqueness* of the solution which is the key problem. We proceed here in three parts. We begin in (i) with the process X^*. The uniqueness for ξ^* and η^* can be then derived very quickly in (ii) as a direct consequence of the look-down uniqueness. In (iii) we discuss existence and uniqueness of the diffusion limit.

(i) The proof for the uniqueness of the martingale problem for X^* proceeds in five steps. We will use knowledge about classical martingale problems which guarantees uniqueness for finite index sets and then utilize the special structure of our process to reduce the problem of uniqueness via coupling to the finite particle case. Recall the generator $L_{(t,X^*)}$ for the historical look-down process given in (2.13). We refer to the corresponding martingale problem as the $(\mathcal{I},\mathcal{U})$-martingale problem.

Step 1 (Finite index set and its consequences) The first observation is that for a situation where $|\mathcal{I}| < \infty$ we see that we can decompose $L_{(t,X^*)}$ into two components L_1 and L_2, where L_1 is the generator associated with the path process of the migration process and the operator L_2 corresponds to the resampling and involves pure jumps with bounded rates. Then by the results of Ethier and Kurtz (compare with Theorem 4.4.1, Corollary 4.4.4, and Problem 4.11.3 in [25]), $L_1 + L_2$ specifies a well-posed martingale problem. Now we use this $|\mathcal{I}| < \infty$ result to obtain uniqueness for the case $|\mathcal{I}| = \infty$ as well.

The reduction from $|\mathcal{I}| = +\infty$ to $|\mathcal{I}| < \infty$ comes in two steps. First we reduce everything to a locally finite system. Namely by restricting the index set in (2.13) from $\mathcal{I}$ to $\mathcal{I}^\rho$ one obtains a new generator with the corresponding martingale problem, which we call the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem, that has locally only finitely many individuals. Note that the evolution of any solution to the historical look-down martingale problem restricted to individuals with indices in $\mathcal{I}^\rho$ is a solution to the $\mathcal{I}^\rho$-martingale problem. Therefore, to show uniqueness of the solution for the $(\mathcal{I},\mathcal{U})$-martingale problem, it suffices to show uniqueness of the solution for the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem for each $\rho \in (0,\infty)$. Finally note that one solution to the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem is given by the evolution of all individuals with indices in $\mathcal{I}^\rho$ in the historical look-down process from Subsection 1.4.

The strategy of the proof for uniqueness of the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem is to reduce this infinite (locally finite) case to the finite case, for which we have uniqueness. This reduction is by approximation using that (as it turns out) any solution of the historical look-down martingale problem has the property that in any finite time interval one particle interacts with at most finitely many particles, and therefore, in any finite time interval, any finite collection of particles interacts with at most finitely many particles. Before giving a rigorous argument, we sketched the proof in the next three paragraphs.

Note that two solutions of the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem will be equal in law, if for any finite collection $\tilde\iota_1,\tilde\iota_2,\cdots,\tilde\iota_n$ of individuals, the laws of processes restricted to the evolution of this collection inside the two solutions are equal. Consider therefore a *finite number of tagged individuals*. Next choose a finite subset of the site space, say B. The idea is now that the dynamics of the tagged individuals in the two systems is for a finite time horizon very well approximated by the dynamics of the tagged individuals in the system consisting only of those individuals in the $\mathcal{I}^\rho$-system which start initially in B if we make B sufficiently large. To make this comparison precise we have to estimate the influence of all the individuals not initially in B on the evolution of the tagged ones which is the main content of the next steps.

The comparison estimate is based on a *coupling* of the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-evolution and the one of the individuals which are initially in B which we denote $X^{*,\rho,\mathrm{fin}}$ (which depends of course also on B). Coupling means here realization on some common probability space. More precisely, given any solution $X^{*,\rho}$ to the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem, one constructs a solution $(X^{*,\rho},X^{*,\rho,\mathrm{fin}})$ to the coupled martingale problem. This construction we carry out in Step 3 below. The migration process in $X^{*,\rho,\mathrm{fin}}$ is for the finitely many individuals considered to be the same as in $X^{*,\rho}$, but the resampling occurs only for individuals with indices initially located in the finite set B. Therefore the law of $X^{*,\rho,\mathrm{fin}}$ is uniquely determined via its martingale problem by the result on the finite case.

The key property however is that for the coupling $(X^{*,\rho},X^{*,\rho,\mathrm{fin}})$ the migration and resampling dynamics of given tagged particles $\tilde\iota_1$, ..., $\tilde\iota_n$ coincide in $X^{*,\rho}$ and $X^{*,\rho,\mathrm{fin}}$ except on events of arbitrarily small probability (due to Lemma 4.2 in Step 2) by choosing B sufficiently large.

Next we give a rigorous argument. Fix $n\ge 1$, indices $\tilde\iota_1$, ..., $\tilde\iota_n\in\mathcal{I}^\rho$, and $T>0$. Assume that the solution to the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem is not unique. The Steps 2 and 5 below will imply that for all fixed T and all fixed n, the law of the evolution of the individuals $\widetilde\iota_1$, ..., $\widetilde\iota_n$, on the interval $[0,T]$, is the same in both processes. Since this holds for every finite collection of tagged individuals , this yields a contradiction. Hence we have uniqueness.

Step 2 (Estimate for influence of infinite set onto finite set) For the coupled process we construct in Step 3 we will need to know for a solution of the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem the influence of the individuals not initially in B on those tagged individuals , which are assumed initially in B, in order to estimate the distance of the coupling between $X^{*,\rho}$ and $X^{*,\rho,\mathrm{fin}}$ from the diagonal. For convenience we choose B from the set of all balls with large radius. A ball with radius R is denoted by B_R.

Fix $\varepsilon>0$ and a time horizon T. Now consider solutions to the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem and denote by $\mathbf{P}$ any such law. In other words whenever we write an estimate involving $\mathbf{P}$, it holds for all such $\mathbf{P}$.

We shall now pick two parameters $0<r<R$. First we choose r large enough so that

$$\mathbf{P}[\{(\pi_G^* y^{\widetilde\iota_i})_s\in B_r,\,\forall\, s\in[0,T],\, i=1,...,n\}]\ge 1-\varepsilon. \tag{4.12}$$

Then choose $R>r$ such that

$$\sum_{\iota\in\mathcal{I}^\rho:\,(\pi_G^* y^\iota)_0\notin B_R}\mathbf{P}[\{(\pi_G^* y^\iota)_s\in B_r,\ \text{for some}\ \ s\in[0,T]\}]\le\varepsilon. \tag{4.13}$$

Observe that such an R always exists since we have assumed that the initial positions of X_0^* fulfill (1.23), a.s.

We have to estimate the influence of all individuals with label less than ρ and initial positions outside the large ball B_R during the fixed time interval $[0,T]$ on those individuals with indices $\iota\in\mathcal{I}^\rho$ and initial positions in B_r with $r<R$ which we have tagged. We need to show that this influence becomes small for fixed r as $R\to\infty$. For this purpose we estimate the *expected rate* at which a tagged particle interacts with individuals starting outside of B_R by time t on the event that the tagged individuals all stay in B_r during time $[0,T]$. We

call this quantity

$$d_{\{\widetilde{\iota}_1,\cdots,\widetilde{\iota}_n\},r,R}(t). \tag{4.14}$$

With this in mind we consider the random variable which describes at which rate interaction occurs between the *tagged individuals* with indices $\widetilde{\iota}_1, \ldots, \widetilde{\iota}_n$ on the one hand, and individuals which had interacted with a particle with index $\iota \in \mathcal{I}^\rho$ and initial positions outside B_R (also indirectly individuals which had interacted with individuals which had interacted with individuals initially outside B_R, and so on) on the other hand. Here we only need to count interactions which have the property that the individual from outside B_R has the smaller label, since only then a path starting outside of B_R may influence the path of a tagged individual. Define the event $A_{r,t}(t)$ that the path of the tagged particles remain in B_r till time $T \wedge t$.

$$D_{\{\widetilde{\iota}_1,\ldots,\widetilde{\iota}_n\},r,R}(t) := \gamma 1_{A_{r,t}(t)} \sum_{i=1}^{n} \sum_{\iota \in \mathcal{I}^\rho : (\pi_G^* y^\iota)_0 \notin B_R} \int_0^{T\wedge t} \mathrm{d}s \, 1_{\{y_s^{\widetilde{\iota}_i} = y_s^\iota\}}, \tag{4.15}$$

We claim that

$$d_{\{\widetilde{\iota}_1,\cdots,\widetilde{\iota}_n\},r,R}(t) = \mathbf{E}\left[D_{\{\widetilde{\iota}_1,\cdots,\widetilde{\iota}_n\},r,R}(t)\right]. \tag{4.16}$$

A quick look at the above formula may puzzle the reader, since the interactions of individuals $\widetilde{\iota}_1, \ldots, \widetilde{\iota}_n$ with individuals that started *inside* of B_R and furthermore interacted with individuals that started outside of B_R etc. do not seem to be accounted for. However, we are using here the "2 versus 0 rule" that says: after each resampling event involving a particle that started *inside* of B_R and a particle interacted with individuals started outside of B_R, with probability $1/2$ the number of individuals that started inside of B_R but interacted with individuals outside of B_R increases by 1, and with probability $1/2$ the same number decreases by 1, so on average there is exactly one interaction contribution term coming from each of the individuals started outside of B_R. This argument uses the fact that the positions and indices are chosen independently of the labels.

Note that for any solution of the $(\mathcal{I}^\rho, \mathcal{U}^\rho)$-martingale problem, the evolution of the path of an individual is given by a random walk with transition kernel $a(x,y)$ and jump rate 1 which is *uniquely determined* by these data and the starting point. Applying the martingale problem we shall see later that the migration of paths disregarding the type is indeed a system of independent random walks. This combined with standard random walk estimates immediately gives the following.

Lemma 4.2. *(Interaction with distant individuals is small) If the spatial configuration corresponding to X_0^* fulfills (1.23) then for every solution to the $(L_{(t,X^*)}, \mathcal{A}^{*,look})$-martingale problem we have:*

$$\mathbf{E}[D_{\{\widetilde{\iota}_1,\ldots,\widetilde{\iota}_n\},r,R}(T)] \le \gamma n T \varepsilon, \tag{4.17}$$

with R satisfying (4.13).

In particular, by (4.16),

$$d_{\{\widetilde{\iota}_1,\cdots,\widetilde{\iota}_n\},r,R}(t) \le \gamma n T \varepsilon. \tag{4.18}$$

This allows an important conclusion. Namely using $(1 - e^{-x}) \le x$, for $x > 0$, together with (4.12) and with C_R denoting the event that *a tagged particle changes type due to an (direct or indirect) interaction with a particle starting outside* B_R (recall (4.12)) we conclude that

$$\mathbf{P}[C_R] \le \varepsilon + \gamma n T \varepsilon. \tag{4.19}$$

Estimate (4.19) will be used later in step 5 for showing via coupling that finite index set processes approximate the infinite index set ones.

Step 3 (Construction of coupling) In this step we define the coupled dynamics explicitly. The coupled dynamics will be a process on

$$\mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G \times K) \times \mathcal{U})^{\mathcal{I}^\rho} \times (\mathcal{D}(\mathbb{R}^+, G \times K) \times \mathcal{U})^{\mathcal{I}_1^\rho}. \tag{4.20}$$

Here $\mathcal{I}_1^\rho$ is a finite subset of $\mathcal{I}^\rho$ which we choose (deterministically and hence not by simply taking individuals with position in B_R which would be random) according to the following observation. Note that since for all solutions $\mathbf{P}$ of the $\mathcal{I}^\rho$-martingale problem we have the prescribed initial law, we can find for R fixed as above an integer $\bar{n}(R) \geq n$ such that

$$\mathcal{I}_1^\rho := \{\widetilde{\iota}_1, \widetilde{\iota}_2, \ldots, \widetilde{\iota}_n, \widetilde{\iota}_{n+1}, \ldots, \widetilde{\iota}_{\bar{n}(R)}\} \tag{4.21}$$

has the property that

$$\mathbf{P}\left[\{\iota \in \mathcal{I}^\rho : (\pi_G^* y^\iota)_0 \in B_R\} \subseteq \mathcal{I}_1^\rho\right] \geq 1 - \varepsilon. \tag{4.22}$$

The idea for the coupled dynamics is to produce as marginals the processes we wish to compare and to make as many transitions as possible jointly in both processes. The process will be defined by specifying the generator consisting of a migration part and a resampling part. We discuss first the migration and then the resampling.

Recall "migration" generators A describing the motion of one individual and $\widetilde{A}$ describing the evolution of the path of one individual, both defined in (2.4)-(2.5). Furthermore recall from (2.9) and (2.10) how, based on these operators, we defined new operators $A^{\mathcal{I}}, \widetilde{A}^{\mathcal{I}}$ for collections of individuals respectively paths. In analogy to (2.9) define $A^{\mathrm{cou}, \mathcal{I}^\rho, \mathcal{I}_1^\rho}$ (where "cou" stands for coupling) as follows for bounded functions $F : (G \times K)^{\mathcal{I}^\rho} \times (G \times K)^{\mathcal{I}_1^\rho} \to \mathbb{R}^+$:

$$\begin{aligned} &A^{\mathrm{cou}, \mathcal{I}^\rho, \mathcal{I}_1^\rho} F(\underline{(x_1, k_1)}, \underline{(x_2, k_2)}) \\ &\quad = \sum_{\iota \in \mathcal{I}_1^\rho} \sum_{z \in G} (a(0,z) - \delta(0,z)) F(\underline{(x_1^{z,\iota}, k_1)}, \underline{(x_2^{z,\iota}, k_2)}) \\ &\quad + \sum_{\iota \in \mathcal{I}^\rho \setminus \mathcal{I}_1^\rho} \sum_{z \in G} (a(0,z) - \delta(0,z)) F(\underline{(x_1^{z,\iota}, k_1)}, \underline{(x_2, k_2)}), \end{aligned} \tag{4.23}$$

where for $\underline{x} \in G^{\mathcal{I}^\rho}$, $z \in G$, $\iota \in \mathcal{I}^\rho$, the object $\underline{x}^{z,\iota} \in G^{\mathcal{I}^\rho}$ is defined by

$$(\underline{x}^{z,\iota})_{\widetilde{\iota}} := \begin{cases} x^{\widetilde{\iota}} + z & \text{if } \iota = \widetilde{\iota} \\ x^{\widetilde{\iota}} & \text{if } \iota \neq \widetilde{\iota} \end{cases} \tag{4.24}$$

Now define based on $A^{\mathrm{cou}, \mathcal{I}^\rho, \mathcal{I}_1^\rho}$ the operator $\widetilde{A}^{\mathrm{cou}, \mathcal{I}^\rho, \mathcal{I}_1^\rho}$ in the way analogous to (2.10).

Next we discuss the resampling part. In analogy to (2.11)-(2.12), define for each

$$(t, \underline{(y_1, u_1)}, \underline{(y_2, u_2)}) \in \mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G \times K) \times \mathcal{U})^{\mathcal{I}^\rho} \times (\mathcal{D}(\mathbb{R}^+, G \times K) \times \mathcal{U})^{\mathcal{I}_1^\rho} \tag{4.25}$$

and $\iota_1 \in \mathcal{I}^\rho$, $\iota_2 \in \mathcal{I}_1^\rho$ the following operators

$$\begin{aligned} \widetilde{\theta}^{1,2}_{\iota_1, \iota_2} : \quad & (t, (..., (y_1^{\iota_1}, u_1^{\iota_1}), ..., (y_1^{\iota_2}, u_1^{\iota_2}), ...), (..., (y_2^{\iota_1}, u_2^{\iota_1}) ..., (y_2^{\iota_2}, u_2^{\iota_2}) ...)) \\ & \mapsto (t, (..., (y_1^{\iota_1}, u_1^{\iota_1}) ..., (y_1^{\iota_1}, u_1^{\iota_2}), ...), (..., (y_2^{\iota_1}, u_2^{\iota_1}), ..., (y_2^{\iota_1}, u_2^{\iota_2}), ...)), \end{aligned} \tag{4.26}$$

and

$$\begin{aligned} \widetilde{\theta}^{1}_{\iota_1, \iota_2} : \quad & \Big(t, (..., (y_1^{\iota_1}, u_1^{\iota_1}), ..., (y_1^{\iota_2}, u_1^{\iota_2}), ...), \underline{(y_2, u_2)}\Big) \\ & \mapsto \Big(t, (..., (y_1^{\iota_1}, u_1^{\iota_1}) ..., (y_1^{\iota_1}, u_1^{\iota_2}), ...), \underline{(y_2, u_2)}\Big), \end{aligned} \tag{4.27}$$

and analogously to (2.12) their corresponding operators $\widetilde{\Theta}^{1,2}_{\iota_1, \iota_2}$, $\widetilde{\Theta}^{1}_{\iota_1, \iota_2}$ on the space of functions given by (4.25).

With these two ingredients we define now the generator of the "bivariate" process generating the coupling. As in (2.13), denote by $\underline{(y_1, u_1)}$ the $\mathcal{I}^\rho$-indexed vectors of (y, u)'s, and let in addition $\underline{(y_2, u_2)}$ be a copy of $\underline{(y_1, u_1)}$ but restricted to indices in $\mathcal{I}_1^\rho$. Consider as test functions the members of the algebra of functions which is generated by products of functions in the two respective groups of variables as they appear in the formulation of the ordinary martingale problem. Define:

$$\begin{aligned}
&L^{\mathrm{cou}}_{(t,X^{*,\rho},X^{*,\rho,\mathrm{fin}})}F(s,\underline{(y_1,u_1)},\underline{(y_2,u_2)}) \\
&\qquad= \widetilde{A}^{\mathrm{cou},\mathcal{I}^\rho,\mathcal{I}_1^\rho}F(s,\underline{(y_1,u_1)},\underline{(y_2,u_2)}) \\
&\qquad+\gamma\sum_{\substack{\iota_1,\,\iota_2\in\mathcal{I}_1^\rho;\\ u^{\iota_1}<u^{\iota_2}}} 1_{\{(\pi_G^* y^{\iota_1})_s=(\pi_G^* y^{\iota_2})_s\}}(\widetilde{\Theta}^{1,2}_{\iota_1,\,\iota_2}F-F)(s,\underline{(y_1,u_1)},\underline{(y_2,u_2)}) \\
&\qquad+\gamma\sum_{\substack{\iota_1\in\mathcal{I}^\rho\setminus\mathcal{I}_1^\rho,\iota_2\in\mathcal{I}^\rho;\\ u^{\iota_1}<u^{\iota_2}}} 1_{\{(\pi_G^* y^{\iota_1})_s=(\pi_G^* y^{\iota_2})_s\}}(\widetilde{\Theta}^{1}_{\iota_1,\,\iota_2}F-F)(s,\underline{(y_1,u_1)},\underline{(y_2,u_2)}).
\end{aligned}\tag{4.28}$$

Call the martingale problem corresponding to the above generator $L^{\mathrm{cou}}_{(t,X^{*,\rho},X^{*,\rho,\mathrm{fin}})}$ the *coupled martingale problem.*

Note that by construction of the operator in (4.28) we know that for every solution of the coupled martingale problem the first marginal of the solution satisfies the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem and the second marginal satisfies the $(\mathcal{I}_1^\rho,\mathcal{U}_1^\rho)$-martingale problem, where $\mathcal{U}_1:=\{u^\iota;\,\iota\in\mathcal{I}_1^\rho\}$.

Step 4 (Coupled process exists for given first marginal) The next result is crucial for our uniqueness argument since it guarantees that for a given solution of the $\mathcal{I}^\rho$-martingale problem, we can construct a solution of the coupled martingale problem.

Lemma 4.3. *(Coupling exists and determines second coordinate uniquely) For any solution $X^{*,\rho}$ of the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem, there exists a process $X^{*,\rho,\mathrm{fin}}$ on the same probability space, so that $(t,X^{*,\rho},X^{*,\rho,\mathrm{fin}})$ is a solution of the coupled martingale problem and $X^{*,\rho,\mathrm{fin}}$ solves the $(\mathcal{I}_1^\rho,\mathcal{U}_1^\rho)$-martingale problem (with $\mathcal{I}_1^\rho$ as defined in (4.21)).*

Proof of Lemma 4.3 Proceed stepwise as follows. (i) We define on the same probability space *as the given solution* $X^{*,\rho}$ a new process $\widetilde{X}^{*,\rho}$ (with the same state space as $X^{*,\rho}$) by suppressing all resampling events between individuals where one of the individuals has an index *not* in $\mathcal{I}_1^\rho$. We will have to argue that this is possible, since countably many individuals contribute to the dynamics. (ii) We restrict the process $\widetilde{X}^{*,\rho}$ to indices in $\mathcal{I}_1^\rho$, to obtain a process denoted by $\widetilde{X}^{*,\rho,\mathrm{fin}}$ on the same probability space. (iii) We then form $(t,X^{*,\rho},\widetilde{X}^{*,\rho,\mathrm{fin}})$ and prove that the pair $(t,X^{*,\rho},\widetilde{X}^{*,\rho,\mathrm{fin}})$ is a version of the coupled process.

What has been achieved once we have carried out this construction? Since the second component has only finitely many indices involved, its law agrees then by our reasoning in Step 1 with the unique solution of the $\mathcal{I}_1^\rho$-martingale problem on the same probability space. Step 5 uses this to prove the Lemma.

Hence it remains to construct

$$(t,X^{*,\rho},\widetilde{X}^{*,\rho}) \text{ on } \mathbb{R}^+\times(\mathcal{D}(\mathbb{R},G\times K)\times\mathcal{U})^{\mathcal{I}^\rho}\times(\mathcal{D}(\mathbb{R},G\times K)\times\mathcal{U})^{\mathcal{I}^\rho}\tag{4.29}$$

on the probability space where $X^{*,\rho}$ is defined, such that the process in (4.29) satisfies the martingale problem with generator:

$$\begin{aligned}
&\widetilde{L}_{(t,X^{*,\rho},\widetilde{X}^{*,\rho})}F(s,\underline{(y_1,u_1)},\underline{(y_2,u_2)}) \\
&\qquad= \widetilde{A}^{\mathrm{cou},\mathcal{I}^\rho,\mathcal{I}^\rho}F(s,\underline{(y_1,u_1)},\underline{(y_2,u_2)}) \\
&\qquad+\gamma\sum_{\substack{\iota_1,\,\iota_2\in\mathcal{I}_1^\rho;\\ u^{\iota_1}<u^{\iota_2}}} 1_{\{(\pi_G^* y^{\iota_1})_s=(\pi_G^* y^{\iota_2})_s\}}(\widetilde{\Theta}^{1,2}_{\iota_1,\,\iota_2}F-F)(s,\underline{(y_1,u_1)},\underline{(y_2,u_2)}) \\
&\qquad+\gamma\sum_{\substack{\iota_1,\,\iota_2\in\mathcal{I}^\rho\times\mathcal{I}^\rho\setminus\mathcal{I}_1^\rho\times\mathcal{I}_1^\rho;\\ u^{\iota_1}<u^{\iota_2}}} 1_{\{(\pi_G^* y^{\iota_1})_s=(\pi_G^* y^{\iota_2})_s\}}(\widetilde{\Theta}^{1}_{\iota_1,\,\iota_2}F-F)(s,\underline{(y_1,u_1)},\underline{(y_2,u_2)}).
\end{aligned}\tag{4.30}$$

Fix a solution $X^{*,\rho}$ of the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem and realize it on some probability space. We now construct $\widetilde{X}^{*,\rho}$ and verify that its claimed properties are satisfied. This involves two arguments. We first deduce some useful properties of solution $X^{*,\rho}$ of the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem, in order to construct the process from (4.29) and (4.30).

We begin now to deduce properties of $X^{*,\rho}$ from the martingale problem. What we first need is to establish that the motion is a system of independent random walks and secondly that resampling events occur for each given pair at most finitely often at distinct times during a finite period $[0,T]$ in order to be able to define $\widetilde{X}^{*,\rho}$ as described in the beginning. Finally we have to verify that the constructed process satisfies the martingale problem (4.30). We establish for this purpose that the resampling events arise from appropriate Poisson processes. For that latter purpose we look at the migration part and the resampling part of the evolution separately and combine the argument with the first, respectively second, task just mentioned.

Start with the *migration*. As mentioned above, the migration dynamics of individuals in $X^{*,\rho}$, that is the process obtained by ignoring the types associated with the paths, is uniquely determined and is given by independent random walks, which we denote by $R = \{R^\iota(t), t \geq 0, \iota \in \mathcal{I}^\rho\}$. Let $R^* = \{R^{*,\iota}, \iota \in \mathcal{I}^\rho\}$ be the "historical process" corresponding to the process R. The law of R^* is uniquely determined by the martingale problem (this is classical, see [16]).

The following consequence of this fact on the migration part will be important for analyzing the resampling part. Fix indices ι_1, ι_2, corresponding to labels $u^{\iota_1} < u^{\iota_2}$. Then the total amount of time when individuals with indices ι_1, ι_2 share location during $[0,T]$ is well defined for $X^{*,\rho}$. That is, its law is uniquely determined. We denote this quantity by

$$D_{\iota_1,\iota_2}(t) := \int_0^t \mathrm{d}s\, 1_{\{(\pi_G^* X^{*,\rho,\iota_1})_s = (\pi_G^* X^{*,\rho,\iota_2})_s}. \tag{4.31}$$

Recall that these times are specifying time slots where resampling between ι_1, ι_2 is possible.

However, notice that not all look-down events are between paths of different types which is necessary if one would like to observe them. We will therefore introduce an enrichment of our process, namely we enrich the type by a second component such that throughout every individual has a different type initially and changes upon look-down only the first component of the type. This turns all possible look-down events in the original process into observable ones in the enriched process.

We turn now to the *resampling part* and verify three facts. (α) We shall show that since by assumption the initial state $X_0^{*,\rho}$ is in the Liggett-Spitzer space, there will be only finitely many look-down interactions between each pair of individuals during $[0,T]$, and that there are no simultaneous look-down interactions, almost surely. (β) We show that given $D_{\iota_1,\iota_2}(T)$, the number of look-down interactions between the individuals ι_1 and ι_2 during the time interval $[0,T]$ is Poisson (rate $\gamma D_{\iota_1,\iota_2}(T)$). (γ) Finally, the numbers of look-down interactions between different pairs of individuals are conditionally independent given the total location sharing times.

A basic tool in verifying these properties is to formulate them as martingale properties. Recall that the Poisson (rate λ) process $(N(t), t \geq 0)$ is characterized uniquely by the Laplace transforms specified at all times. Hence we have a unique characterization by requiring that the process $W := (W(t))_{t\geq 0}$ with

$$W(t) := \exp\left\{-\theta N(t) + \lambda t(1-e^{-\theta})\right\} \tag{4.32}$$

is a martingale for every $\theta > 0$.

Now we prove the three facts $(\alpha) - (\gamma)$ based on the definition of the $(\mathcal{I}^\rho,\mathcal{U}^\rho)$-martingale problem.

Define, for ι_1 and ι_2 with $u^{\iota_1} < u^{\iota_2}$, $x^* := (x_t^{*,\iota}; \iota \in \mathcal{I}^\rho)_{t\geq 0}$ in $(\mathcal{D}(\mathbb{R}^+, G\times K))^{\mathcal{I}^\rho}$ the function

$$f^{\iota_1,\iota_2}(t,x^*) := \#\{s \in [0,t] : \pi_K^* x_{s-}^{*,\iota_2} \neq \pi_K^* x_{s-}^{*,\iota_1}, \pi_K^* x_s^{*,\iota_2} = \pi_K^* x_s^{*,\iota_1}\}. \tag{4.33}$$

(α) Let $N^{\iota_1,\iota_2}(t) := f^{\iota_1,\iota_2}(t, X_t^{*,\rho})$ be the number of "look-down interactions" for the pair ι_1, ι_2 during $[0,t]$. From the martingale property (2.20), (2.13) applied to f^{ι_1,ι_2}, we get $EN^{\iota_1,\iota_2}(t) \leq \gamma t$ so there are only finitely many look-down interactions for $X^{*,\rho}$ involving ι_1 and ι_2 during $[0,t]$.

(β) The martingale property (2.20), (2.13) implies that

$$\left(\exp\{-\theta N^{\iota_1,\iota_2}(t) + \gamma D_{\iota_1,\iota_2}(t)(1-e^{-\theta})\}\right)_{t\geq 0} \tag{4.34}$$

is a martingale. Define as usual $D^{-1}_{\iota_1,\iota_2}(t) := \inf\{s > 0 : D_{\iota_1,\iota_2}(s) > t\}$, and let $\widetilde{N}^{\iota_1,\iota_2}(t) := N(D^{-1}_{\iota_1,\iota_2}(t))$. Then by (4.34) we have

$$\left(\exp\{-\theta\widetilde{N}^{\iota_1,\iota_2}(t) + \gamma t(1-e^{-\theta})\}\right)_{t\ge 0} \tag{4.35}$$

is a martingale for each $\theta > 0$, and therefore by (4.32) $(\widetilde{N}^{\iota_1,\iota_2}(t), t \ge 0)$ is a Poisson (rate γ) process.

(γ) Next we need to verify that different pairs interact independently given the location sharing times. Suppose that $\iota_3 < \iota_4$ are given in addition to ι_1 and ι_2, and suppose that either $\iota_1 \neq \iota_3$ or $\iota_2 \neq \iota_4$. Denote by N^{ι_3,ι_4} the analogue of N^{ι_1,ι_2} where ι_3 (resp. ι_4) replaces ι_1 (resp. ι_2). We introduce a further enrichment of our probability space. Define two auxiliary processes $(\hat{N}^{\iota_1,\iota_2}(t), t \ge 0)$ and $(\hat{N}^{\iota_3,\iota_4}(t), t \ge 0)$ so that $\hat{N}^{\iota_1,\iota_2}$ (resp. $\hat{N}^{\iota_3,\iota_4}$) given a solution $X^{*,\rho}$, is a Poisson (rate γ) process evolving only at times when ι_1 and ι_2 (resp. ι_3 and ι_4) individuals do not share the same location and assume without loss of generality that $\hat{N}^{\iota_1,\iota_2}$, $\hat{N}^{\iota_3,\iota_4}$ are otherwise independent of $X^{*,\rho}$. Hence $\widehat{N}^{\iota_1,\iota_2}$, (resp. $\widehat{N}^{\iota_3,\iota_4}$) have generator: for f bounded functions on $\mathbb{N}$,

$$\begin{aligned} L_{\hat{N}^{\iota_1,\iota_2}} f(x) &= \quad \gamma 1_{\{(\pi^*_G y^{\iota_1})_s \neq (\pi^*_G y^{\iota_2})_s\}}(f(x+1) - f(x)) \\ (\text{resp. } L_{\hat{N}^{\iota_3,\iota_4}} f(x) &= \quad \gamma 1_{\{(\pi^*_G y^{\iota_3})_s \neq (\pi^*_G y^{\iota_4})_s\}}(f(x+1) - f(x))). \end{aligned} \tag{4.36}$$

Then clearly for $(M^{\iota_1,\iota_2}(t), t \ge 0)$, where $M^{\iota_1,\iota_2}(t) := N^{\iota_1,\iota_2}(t) + \hat{N}^{\iota_1,\iota_2}(t)$, we have that

$$\exp\left\{-\theta M^{\iota_1,\iota_2}(t) + \lambda t(1-e^{-\theta})\right\} \tag{4.37}$$

is a martingale for each $\theta > 0$, so that M^{ι_1,ι_2} is a Poisson (rate γ) process. Similarly $(M^{\iota_3,\iota_4}(t), t \ge 0)$, where $M^{\iota_3,\iota_4}(t) := N^{\iota_3,\iota_4}(t) + \hat{N}^{\iota_3,\iota_4}(t)$, is a Poisson (rate γ) process.

Moreover calculating explicitly with the generator, the process

$$\exp\left\{-\theta_1 M^{\iota_1,\iota_2}(t) + \gamma t(1-e^{-\theta_1})\right\} \cdot \exp\left\{-\theta_2 M^{\iota_3,\iota_4}(t) + \gamma t(1-e^{-\theta_2})\right\} \tag{4.38}$$

is a martingale for all $\theta_1, \theta_2 > 0$ so M^{ι_1,ι_2} and M^{ι_3,ι_4} are independent processes. The above procedure can be performed for any k different pairs of labels $(\iota_1,\iota_2),\ldots,(\iota_{2k-1},\iota_{2k})$ to determine that their corresponding processes $M^{\iota_1,\iota_2},\ldots,M^{\iota_{2k-1},\iota_{2k}}$ are mutually independent.

Now we have verified properties $(\alpha)-(\gamma)$ for the process $X^{*,\rho}$ and we are ready to define $(\widetilde{X}^{*,\rho}_t)_{t\ge 0}$. Set first

$$\widetilde{X}^{*,\rho}_0 = X^{*,\rho}_0 \tag{4.39}$$

and then define $(\widetilde{X}^{*,\rho}_t)_{t\ge 0}$ separately for $\iota \in \mathcal{I}^\rho_1$, and for $\iota \in \mathcal{I}^\rho \setminus \mathcal{I}^\rho_1$, in the following way.

Case 1 For fixed $\iota \in \mathcal{I}^\rho$ denote by $\gamma^1_\iota, \gamma^2_\iota, \ldots$ the interaction times in $X^{*,\rho}$ of the particle ι with another particle having a lower label. These times are well defined by the discussion above which allows one to use the estimates we gave in constructing our process.

For fixed $\iota \notin \mathcal{I}^\rho_1$, let $\widetilde{X}^{*,\rho,\iota}_t = (R^{*,\iota}(x^\iota,t), k^\iota, u^\iota)$, $t \ge 0$, or equivalently, for $t \in [0,\gamma^1_\iota) \cap [0,T]$ let $\widetilde{X}^{*,\rho,\iota}_t := X^{*,\rho,\iota}_t$, for $t \in \{\gamma^1_\iota, \gamma^2_\iota, \ldots\} \cap [0,T]$, let $\widetilde{X}^{*,\rho,\iota}_t := \widetilde{X}^{*,\rho,\iota}_{t-}$, and for $t \in (\gamma^{j-1}_\iota, \gamma^j_\iota) \cap [0,T]$, $j \ge 2$, let

$$(\widetilde{X}^{*,\rho,\iota}_t)_s := \begin{cases} (\widetilde{X}^{*,\rho,\iota}_t)_s & s \in [0,\gamma^{j-1}_\iota) \\ (R^\iota(x^\iota,s), \pi^*_K(\widetilde{X}^{*,\rho,\iota}_{\gamma^{j-1}_\iota}), u^\iota) & s \in [\gamma^{j-1}_\iota, t] \end{cases}, \tag{4.40}$$

(compare with (1.34)).

Case 2 Here we have $\iota \in \mathcal{I}^\rho_1$. First reorder the indices in $\mathcal{I}^\rho_1$ in increasing order according to their labels. This is always possible since $\mathcal{I}^\rho_1$ is finite. The construction is now carried out piece by piece in increasing order of ι.

First take the index ι with the lowest label. Define $\widetilde{X}^{*,\rho,\iota}_t, t \in [0,T]$ as in the last paragraph. In words, the particle with index ι does not look-down at any other particle during the time interval $[0,T]$, and its migration dynamics is inherited from R^ι.

Inductively define $\widetilde{X}^{*,\rho,\iota}_t$, $t \in [0,T]$ for higher order indices as follows. For $t \in [0,\gamma^1_\iota)$ set $\widetilde{X}^{*,\rho,\iota}_t := X^{*,\rho,\iota}_t$. For $t \in \{\gamma^1_\iota, \gamma^2_\iota, \ldots\}$ let either $\widetilde{X}^{*,\rho,\iota}_t := X^{*,\rho,\iota}_{t-}$ if $X^{*,\rho,\iota}_t = X^{*,\rho,\iota_1}_{t-}$ for $\iota_1 \notin \mathcal{I}^\rho_1$ (i.e. in $X^{*,\rho}$ ι-indexed particle looks down at another particle with index in $\mathcal{I}^\rho \setminus \mathcal{I}_1^\rho$), or $\widetilde{X}_t^{*,\rho,\iota} := \widetilde{X}_{t-}^{*,\rho,\iota_1}$ if $X_t^{*,\rho,\iota} = X_{t-}^{*,\rho,\iota_1}$ for $\iota_1 \in \mathcal{I}_1^\rho$ (i.e. in $X^{*,\rho}$ the ι-indexed particle looks down at another particle with index in $\mathcal{I}_1^\rho$). For $t \in (\gamma_\iota^{j-1}, \gamma_\iota^j)$, $j \geq 2$, let

$$(\widetilde{X}_t^{*,\rho,\iota})_s := \begin{cases} (\widetilde{X}^{*,\rho,\iota}_{\gamma_\iota^{j-1}})_s & s \in [0, \gamma_\iota^{j-1}) \\ (R^\iota(x^\iota, s), \pi_K^*(\widetilde{X}^{*,\rho,\iota}_{\gamma_\iota^{j-1}}), u^\iota) & s \in [\gamma_\iota^{j-1}, t] \end{cases}. \tag{4.41}$$

This concludes the construction of $\widetilde{X}^{*,\rho}$.

By the discussion in (4.31) to (4.39) we obtain that the constructed process satisfies the martingale problem in (4.30) and hence as we also saw the coupled martingale problem if we restrict the second component to $\mathcal{I}_1^\rho$. This completes the proof of Lemma 4.3. □

Step 5 (Coupling estimate) Now we are ready to conclude the proof by showing that indeed the tagged individuals in both marginals of the coupled process agree with probability arbitrarily close to one.

Clearly the laws of $(X_t^{*,\rho,\text{fin},\iota}; \iota \in \{\widetilde{\iota}_1, \ldots, \widetilde{\iota}_n\})_{t\geq 0}$ and $(\widetilde{X}_t^{*,\rho,\iota}; \iota \in \{\widetilde{\iota}_1, \ldots, \widetilde{\iota}_n\})_{t\geq 0}$ are equal. Now by (4.19), (4.17) and (4.22) the laws of $(X_t^{*,\rho,\iota}; \iota \in \{\widetilde{\iota}_1, \ldots, \widetilde{\iota}_n\})_{t\in[0,T]}$ and $(X_t^{*,\rho,\text{fin},\iota}; \iota \in \{\widetilde{\iota}_1, \ldots, \widetilde{\iota}_n\})_{t\in[0,T]}$, differ by at most $(2+\gamma n T)\varepsilon$ in the variational distance. Therefore, the laws of $(X_t^{*,\rho,\iota,i}; \iota \in \{\widetilde{\iota}_1, \ldots, \widetilde{\iota}_n\})_{t\in[0,T]}$, $i = 1, 2$ coming from two different solutions of the $(\mathcal{I}^\rho, \mathcal{U}^\rho)$-martingale problem, differ by at most $2(2+\gamma n T)\varepsilon$ in the variational distance. Since $\varepsilon > 0$ is arbitrary the laws of the tagged individuals are in fact equal for both solutions of the $(\mathcal{I}^\rho, \mathcal{U}^\rho)$-martingale problem. Then the uniqueness follows as discussed above in Step 1. □

(ii) The proof of the well-posedness for $\xi^{*,\rho}$ is of course the same as the one given above since only the direction of the interaction (i.e. the look-down) becomes now symmetric, but here to begin with we only consider indices $\iota \in \mathcal{I}^\rho$.

The argument for the uniqueness of the solution of the martingale problem for η^* uses the above result for the X^*-, and the $\xi^{*,\rho}$-martingale problems in a standard way. Namely we need to associate with each η^*-solution a $\xi^{*,\rho}$-process which then are automatically different and hence we obtain a contradiction. We can now extend a solution η^* to obtain a process $\xi^{*,\rho}$. The state of η^* is an atomic measure on the state of paths. If for η^* we have initially that the state is purely atomic and weights of atoms are 1, this is straightforward by labeling the atoms with $\mathcal{U}$. If this is not the case and we have an atomic initial state with weights in $\mathbb{N}$ but not necessarily 1 we have to consider a refinement of the states by assigning randomly extra types, i.e., we have to enrich the probability space by independent randomizing experiments to sample the transitions of the types involved. Then use the projection property (recall the first remark in Subsection 1.2) of the model to see that we can construct a $\xi^{*,\rho}$-process for a larger set of labels, to which nevertheless we can apply the uniqueness results.

(iii) The existence of the historical IFWD is obtained by verifying that the prescription we gave for the diffusion limit gives a tight sequence of laws and that weak limit points satisfy the martingale problem. For that purpose one calculates the generator $L_{(t,\xi^*)}$ on functions F of finitely many individuals, which are symmetric and twice continuously differentiable in these variables. Using the Taylor expansion one verifies that all weak limit points must satisfy the $L_{(t,\zeta^*)}$-martingale problem. Since the latter turns out to have unique solutions it suffices to establish tightness. This is again a calculation involving generators. Since both these points are carried out in a non-spatial context in monographs (see for example in [25]) and in the spatial (but non-historical) situation in [38] we omit here the details.

A bit more subtle is the verification of the uniqueness. The basic idea here is to first consider only reduced information about the path, namely to fix a set of time points $s_1 < s_2 < \cdots < s_n$ and to record only the position of the path at these times. This information can be encoded as additional component of the type. However the evolution of this process between the picked types $s_1, \ldots, s_n$ can be described as the dynamics of some multitype IFWD. Here the type space is of course bigger than the original one and bigger than two. Nevertheless we can first use duality to get the uniqueness of the evolution between the time points $s_1, \ldots, s_n$ from which we get uniqueness of the historical process observed at the skeleton of time points. Passing then to finer and finer subdivisions will give in the limit the uniqueness for the real historical processes. This scheme was given in

Dawson and Greven (2003) in [14] for (even time inhomogeneous) Fleming-Viot processes. We refer the reader to that paper for details. □

5. Particle representation of the historical processes (Proof of Theorem 3)

In this chapter we prove the representation theorem (Theorem 3). The first part concerns the IMM where individuals are distinguished, and the second part shows a consequence for the historical IFWD, ζ^*.

In the theorem the process $\xi^{*,\rho}$ is given as a (random) functional of the look-down process X^* and some independent coin flipping. We will show that the process obtained by applying the bijections is a solution to the $(L_{(t,\xi^*)}, \mathcal{A}^{*,\mathrm{ind}})$-martingale problem given in Definition 2.1. In order to recognize the constructed process as the historical IMM, we then use that we have verified in Section 4 that the martingale problem is well-posed.

Fix $\rho \in \mathbb{R}^+$. The proof proceeds in steps. In the first one the random bijection process $\sigma^\rho := \{(\sigma_t^\rho(\iota))_{t\geq 0};\, \iota \in \mathcal{I}^\rho\}$ is constructed, and in the subsequent ones the claimed relation between $\Sigma^\rho X^*$ and $\xi^{*,\rho}$, and $(\Sigma^\rho X^*)_{\rho\in\mathbb{R}^+}$ and ζ^*, are established. A key element is the characterization of X^* as a solution to a martingale problem given in (2.13).

Step 1 (Construction of Σ^ρ and a version of $\xi^{*,\rho}$) Our starting point is the construction of the process X^* on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ given in Subsection 2.2. We shall enlarge this probability space by i.i.d. coin flipping experiments to $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$. We next construct on this probability space the process σ^ρ of random bijections from $\mathcal{I}^\rho$ to $\mathcal{I}^\rho$ such that for each $t \geq 0$, $\Sigma_t^\rho X_t^*$ is a version of the enriched historical IMM $\xi_t^{*,\rho}$. The two processes $\xi_t^{*,\rho}$ and X_t^* have the same migration mechanism. In order to remove the preference inherent in the look-down, the task we face in constructing σ^ρ is to symmetrize the components in a suitable way. For this purpose we need a "random permutation" among all individuals .

Observe there are at most countably many individuals which have interacted during time interval $[0,t]$ for some time $t > 0$. In order to handle this problem we construct a process generating the needed permutations by using the fact that there are *locally* only finitely many individuals . To carry this out we need a *second set of labels*. In other words with every individual ι we have to associate a new random label $v_t^\rho(\iota)$ depending on the history of the process X^* up to time t.

Start X^* in the same initial state as given in the construction, cf. Subsection 4.1. We provide a random "permutation" $\sigma^\rho : \mathcal{I}^\rho \to \mathcal{I}^\rho$ of our countable index set $\mathcal{I}^\rho$ such that for each time $t \geq 0$ and each $\iota_1 \neq \iota_2$ with $x_t^{\sigma_t^\rho(\iota_1)} = x_t^{\sigma_t^\rho(\iota_2)}$, we have:

$$\mathbf{P}[u^{\sigma_t^\rho(\iota_1)} < u^{\sigma_t^\rho(\iota_2)}] = \mathbf{P}[u^{\sigma_t^\rho(\iota_1)} > u^{\sigma_t^\rho(\iota_2)}] = \frac{1}{2}. \tag{5.1}$$

For that we initially sample for each individual, independently of the others and independent of the initial state, a second label uniformly from $[0,\rho]$. Set

$$v^{\rho,\,\iota} := \text{ second label of the individual with index } \iota. \tag{5.2}$$

Next we introduce the initial state of the random permutation process σ_t^ρ. Initially we set

$$\sigma_0^\rho(\iota, \underline{v}) = \sigma_0^\rho(\iota) := \widetilde{\iota} \tag{5.3}$$

iff $x_0^\iota = x_0^{\widetilde{\iota}}$, and ι has the same rank in $\{u^{\iota'};\, \iota' \in \mathcal{I}^\rho \text{ s.t. } x_0^{\iota'} = x_0^\iota\}$ as $\widetilde{\iota}$ in $\{v^{\widetilde{\iota}'};\, \widetilde{\iota}' \in \mathcal{I}^\rho \text{ s.t. } x_0^{\widetilde{\iota}'} = x_0^{\widetilde{\iota}}\}$.

In order to define the dynamics of σ_t^ρ we enrich our probability space $(\Omega, \mathcal{F}, \mathbf{P})$ by defining independently of everything else a family $L := \{L^{\iota_1,\,\iota_2,n};\, \iota_1,\, \iota_2 \in \mathcal{I}^\rho, n \in \mathbb{N}\}$ of independent $\{0,1\}$-valued random variables with

$$\mathbf{P}[L^{\iota_1,\,\iota_2,n} = 0] = \mathbf{P}[L^{\iota_1,\,\iota_2,n} = 1] = \frac{1}{2}. \tag{5.4}$$

We then define, based on L, random mappings $M = (M^{\iota_1,\,\iota_2,n};\, \iota_1,\, \iota_2 \in \mathbb{N})$ from $([0,\rho])^{\mathcal{I}^\rho}$ to $([0,\rho])^{\mathcal{I}^\rho}$ given by:

$$\begin{aligned} M^{\boldsymbol{\iota_1},\,\boldsymbol{\iota_2},n} &: (\ldots, v^{\boldsymbol{\iota_1}}, \ldots, v^{\boldsymbol{\iota_2}}, \ldots) \\ &\mapsto \begin{cases} (\ldots, v^{\iota_1}, \ldots, v^{\iota_2}, \ldots) & \text{if } L^{\iota_1,\,\iota_2,n} = 0 \\ (\ldots, v^{\iota_2}, \ldots, v^{\iota_1}, \ldots) & \text{if } L^{\iota_1,\,\iota_2,n} = 1 \end{cases}. \end{aligned} \tag{5.5}$$

Based on these ingredients we then define a process $(v_t^\rho)_{t\geq 0}$ with

$$v_t^\rho := \{v_t^\rho(\iota);\, \iota \in \mathcal{I}^\rho\}, \tag{5.6}$$

such that for each $\iota \in \mathcal{I}^\rho$, the ιth component of this configuration is $\mathcal{I}^\rho$-valued, starts in $v^\rho(\iota)$, and is *constant* except for possible jump times $\tau^n_{\iota',\,\iota}$ or $\tau^n_{\iota,\,\iota'}$, for some $n \geq 1$, $\iota' \in \mathcal{I}^\rho$ (recall the notation from (1.31) resp. the corresponding notation for ξ^* given in (1.37) through (1.39)). In particular, up to a finite time horizon $t \geq 0$, it jumps only finitely often. This means that the configuration $(v_t^\rho)_{t\geq 0}$ is *locally constant* up to positive jump times. Set for $\iota_1,\, \iota_2 \in \mathcal{I}^\rho$, $n \in \mathbb{N}$, (here ι_1, ι_2 are ordered according to their label)

$$T^n_{\iota_1,\,\iota_2} := \tau^n_{\iota_1\wedge\iota_2,\,\iota_1\vee\iota_2}. \tag{5.7}$$

Then the jumps of $v^{\rho,\,\iota}$ at time points $T^n_{\iota',\,\iota}$ are defined by the relation (and we verify in Proposition 5.1 that this is possible):

$$v^\rho_{T^n_{\iota',\,\iota}} := M^{\iota',\,\iota,n}(v^\rho_{T^n_{\iota',\,\iota}-}). \tag{5.8}$$

With these ingredients we can now define the random bijection σ_t^ρ of $\mathcal{I}^\rho$ where the randomness is through that dependence on $\underline{v}_t^\rho$. Namely we extend (5.3) for all $t \geq 0$ in such a way that (5.1) holds, i.e.,

$$\sigma_t^\rho(\iota) = \sigma_t^\rho(\iota, \underline{v}^\rho) := \sigma_0^\rho(\iota, \underline{v}_t^\rho). \tag{5.9}$$

In particular,

$$\sigma_t^\rho(\iota) := \widetilde{\iota}, \tag{5.10}$$

iff $x_t^\iota = x_t^{\widetilde{\iota}}$, and ι the rank of ι in $\{u^{\iota'};\, \iota' \in \mathcal{I}^\rho \text{ s.t. } x_t^{\iota'} = x_t^\iota\}$ and the rank of $\widetilde{\iota}$ in $\{v_t^{\widetilde{\iota}'};\, \widetilde{\iota}' \in \mathcal{I}^\rho \text{ s.t. } x_t^{\widetilde{\iota}'} = x_t^{\widetilde{\iota}}\}$ are equal.

This completes the construction of $\sum_t^\rho$ provided we can show:

Proposition 5.1. *(Second label process construction is well-defined) Fix $(\mathcal{I},\mathcal{U})$, $\mathcal{V}^\rho := \{v^\iota;\, \iota \in \mathcal{I}^\rho\}$, and let $\rho \in \mathbb{R}^+$.*

(a) *There is a uniquely determined process, $v^\rho := (v_t^\rho)_{t\geq 0}$, with values in the set of bijections from $\mathcal{I}^\rho$ onto $\mathcal{I}^\rho$ which starts in $\{v^{\rho,\iota};\, \iota \in \mathcal{I}^\rho\}$, and fulfills relation (5.8).*

(b) *For each time $t \geq 0$, σ_t^ρ defined by (5.10) is a bijection from $\mathcal{I}^\rho$ to $\mathcal{I}^\rho$.*

Proof **(a)** The reasoning is similar to that for Proposition 4.1 where we constructed the genealogies of the look-down via a graphical device. Recall from (4.1) the random point process $\mathrm{gen}(X^*)$ which encodes the genealogy of X^*.

Proceeding from $\mathrm{gen}(X^*)$, and using the $\{0,1\}$-valued random variables L defined in (5.4), we define a further point process $\mathcal{I}\times\mathcal{I}\times[0,\infty)$, denoted by $\mathrm{tr}(\sigma^\rho)$, which encodes the traces of second labels. Namely, with $T^n_{\iota_1,\,\iota_2}$ as defined in (5.7), for $t \geq 0$, and $\iota_1,\, \iota_2 \in \mathcal{I}^\rho$,

$$(\iota_1, \iota_2, t) \in \mathrm{tr}(\sigma^\rho) \quad \text{iff} \quad \begin{array}{c} \exists\, n \in \mathbb{N} \text{ such that } t = T^n_{\iota_1,\,\iota_2}, \\ \text{and } L^n_{\iota_1,\,\iota_2} = 1 \end{array}. \tag{5.11}$$

That is, in the present construction we first ignore the direction of an arrow $(\iota_1,\, \iota_2, t)$, and establish *potential bridges* between two individuals no matter what the order of the labels is. Secondly we open or close independently of each other every potential bridge depending on whether the random variable associated with the potential bridge, say $L^{\iota_1,\,\iota_2,n}$, satisfies $L^{\iota_1,\,\iota_2,n} = 0$ or 1.

In order to determine now the value $\sigma_t^{\rho,\,\iota}$ we proceed analogously to the construction of the look-down process. Start from the point $(\iota, t) \in \mathcal{I}\times\mathbb{R}^+$, and look back in time for the first (ι', s) related to (ι, t), that is, and follow the diagram downwards along vertical lines until in encountering a bridge which is open. To make this precise define

$$\tilde{\beta}^1_{(\iota,t)} := \min\{s > 0 :\, \exists\, \tilde{\iota}^1_{(\iota,t)} \in \mathcal{I}^\rho \text{ s.t. } (\tilde{\iota}^1_{(\iota,t)},\, \iota, t-s) \in \mathrm{tr}(\sigma^\rho)\}. \tag{5.12}$$

Since by Assumption (1.23) for all times $t \geq 0$ there are locally only a finite number of individuals with label $\leq \rho$, the same reasoning as in the proof of Proposition 4.1(a) yields that $\tilde{\beta}^1_{(\iota,t)}$ is associated with an, **P**-a.s., uniquely determined label $\tilde{\iota}^1_{(\iota,t)} \in \mathcal{I}^\rho$ satisfying the condition on the right hand side of (5.12). Next continue from $(\tilde{\iota}^1_{(\iota,t)}, t-\tilde{\beta}^1_{(\iota,t)})$, going downwards, crossing bridges (whenever they are not turned down), and so on, until we stop at $(\tilde{\iota}, 0)$, **P**-a.s., for some uniquely determined $\tilde{\iota} \in \mathcal{I}$. We then set

$$v^{\rho,\iota}_t := v^{\rho,\tilde{\iota}}. \tag{5.13}$$

This way the process v^ρ is obviously well-defined.

(b) Notice that the above construction is such that if we start with different labels at time t then we end up with different labels at time 0. Hence for each $t \geq 0$, $v^\rho_t : \mathcal{V}^\rho \to \mathcal{V}^\rho$ is injective. Surjectivity follows via a similar argument by using upward traversal of the graph. Therefore by (5.10) in particular, for each $t \geq 0$, $\sigma^\rho_t : \mathcal{I}^\rho \to \mathcal{I}^\rho$ is a bijection. □

Step 2 (Relations between $\widetilde{\xi}^{*,\rho} = \Sigma^\rho X^*$ and $\xi^{*,\rho}$) . We begin this step by defining a candidate for a version of $\xi^{*,\rho}_t$. Let $(\sigma^\rho_t)^{-1} : \mathcal{I}^\rho \to \mathcal{I}^\rho$ denote the inverse of σ^ρ_t. Based on the above construction, for each $t > 0$, σ^ρ_t induces a map $\Sigma^\rho_t : (\mathcal{D}(\mathbb{R}^+, G\times K)\times\mathcal{U})^{\mathcal{I}^\rho} \to (\mathcal{D}(\mathbb{R}^+, G\times K))^{\mathcal{I}^\rho}$ defined by (recall that $\pi^*_{G\times K}$ means projection on the $\mathcal{D}(\mathbb{R}, G\times K)$-component)

$$\Sigma^\rho_t : \{X^{*,\iota};\ \iota \in \mathcal{I}^\rho\} \mapsto \{\pi^*_{G\times K} X^{*,(\sigma^\rho_t)^{-1}(\widetilde{\iota})};\ \widetilde{\iota} \in \mathcal{I}^\rho\}. \tag{5.14}$$

We then define a process, $\xi^{*,\rho}$, such that for each $t \geq 0$, $\xi^{*,\rho}_t$ is a collection $\{\xi^{*,\rho,\widetilde{\iota}}_t;\ \widetilde{\iota} \in \mathcal{I}\}$ of $\mathcal{D}(G\times K)$-valued components. Namely, for given $t \geq 0$, we set

$$\widetilde{\xi}^{*,\rho}_t := \Sigma^\rho_t X^{*,\rho}_t. \tag{5.15}$$

Now we have to prove that $\widetilde{\xi}^{*,\rho}$ is a version of our process $\xi^{*,\rho}$, this means based on the construction on the right hand side in (5.15) we have to show that $\widetilde{\xi}^{*,\rho}$ satisfies the martingale problem given in Subsection 2.1. The following Lemma is similar to that of Lemma 2.1 in [19] and completes the proof of the first part of the representation theorem.

Recall $\widetilde{\Theta}$ from (2.11), and let $\Theta : \mathcal{C}(\mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G\times K))^{\mathcal{I}^\rho}) \to \mathcal{C}(\mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G\times K))^{\mathcal{I}^\rho})$ be $\widetilde{\Theta}$ restricted to functions which are constant in the $\mathcal{U}$-valued coordinates.

Lemma 5.2. *(Characterization of the permuted process)*

(a) *The process $\widetilde{\xi}^{*,\rho}$ as defined in (5.15) is Markovian and its time-space process determines a Feller semigroup on $\mathcal{C}(\mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G\times K))^{\mathcal{I}^\rho})$.*

(b) *We denote the generator of the corresponding time-space process by $L_{(t,\widetilde{\xi}^{*,\rho})}$. This operator has a domain which includes functions $F : \mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G\times K))^{\mathcal{I}^\rho} \to \mathbb{R}^+$, which are bounded, and depend only on finitely many coordinates. The specified set of functions are dense in $\mathcal{C}(\mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G\times K))^{\mathcal{I}^\rho})$. The action of the operator on F is given by:*

$$\begin{aligned} &L_{(t,\widetilde{\xi}^{*,\rho})}F(s,\underline{y}) \\ &\quad= \widetilde{A}^{\mathcal{I}^\rho} F(s,\underline{y}) + \frac{\gamma}{2}\sum_{\widetilde{\iota}_1\neq\widetilde{\iota}_2\in\mathcal{I}^\rho} 1_{\{(\pi^*_G y^{\widetilde{\iota}_1})_s=(\pi^*_G y^{\widetilde{\iota}_2})_s\}}(\Theta_{\widetilde{\iota}_1,\widetilde{\iota}_2}F - F)(s,\underline{y}). \end{aligned} \tag{5.16}$$

Proof of Part (a) of Theorem 3 Lemma 5.2 gives a characterization which identifies $\Sigma^\rho X^*$ as unique solution to a martingale problem. This already implies that we have a Markov process with the prescribed generator. That is, we have proved the first assertion of Theorem 3. □

Proof of Lemma 5.2 Fix $(\mathcal{I},\mathcal{U})$, and $\rho \in \mathbb{R}^+$. Observe that our construction of the random permutations is only based on the graphical representation for the Markov process X^* up to a fixed time and on the initial choice of second labels, $\mathcal{V}^\rho := \{v^\iota;\ \iota \in \mathcal{I}^\rho\}$ (recall (5.2)). Denote the σ-algebra generated by these random elements up to time t by $\mathcal{F}_t$. In addition the character of the construction is such that at time $t+s$ the construction given $\mathcal{F}_t$ relies in fact only on X^*_t, σ^ρ_t (or by (5.9) equivalently on $\underline{v}^\rho_t$ defined in (5.13)), and random variables which are independent of everything what happened up to time t. According to (5.15) we consider the new process (recall $X^{*,\rho}_t = (X^{*,\iota}_t, \iota \in \mathcal{I}^\rho)$)

$$(t, X_t^{*,\rho,(\sigma^\rho(\cdot,\underline{v}_t^\rho))^{-1}}, \underline{v}_t^\rho)_{t\geq 0} \tag{5.17}$$

which is then automatically Markovian with respect to the canonical σ-algebra.

As a consequence, the distribution of this trivariate process $(t, X_t^{*,(\sigma^\rho(\cdot,\underline{v}_t^\rho))^{-1}}, \underline{v}_t^\rho)_{t\geq 0}$ can be characterized by a corresponding martingale problem of the form: $\left(f(Y_s) - \int_0^s (Lf)(Y_u)du\right)_{s\geq 0}$ is a martingale for f varying in a distribution determining set of functions. We shall determine the generator of the martingale problem denoted $L_{(t,X^{*,(\sigma^\rho(\cdot,\underline{v}))^{-1}},\underline{v})}$, on certain functions next.

We specify first the set of functions on which we shall write the generator. Consider $F : \mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G\times K)\times \mathcal{U}\times[0,\rho])^{\mathcal{I}^\rho}$, bounded and depending on finitely many individuals only.

Next we give the ingredients needed to write the generator. Define $\widehat{\Theta}^0_{\iota_1,\iota_2}F$, and $\widehat{\Theta}^1_{\iota_1,\iota_2}F$ similar as in (2.11) and (2.12): for $(t, \underline{(y,u,\widetilde{\iota})}) \in \mathbb{R}^+ \times (\mathcal{D}(\mathbb{R}^+, G\times K)\times\mathcal{U}\times[0,\rho])^{\mathcal{I}^\rho}$, let

$$\begin{aligned}\hat{\theta}^0_{\iota_1,\iota_2}: \quad & (t, (\ldots, (y^{\iota_1}, u^{\iota_1}, v^{\iota_1}), \ldots, (y^{\iota_2}, u^{\iota_2}, v^{\iota_2}), \ldots)) \\ & \mapsto (t, (\ldots, (y^{\iota_1}, u^{\iota_1}, v^{\iota_1}), \ldots, (y^{\iota_1}, u^{\iota_2}, v^{\iota_2}), \ldots)),\end{aligned} \tag{5.18}$$

and

$$\begin{aligned}\hat{\theta}^1_{\iota_1,\iota_2}: \quad & (t, (\ldots, (y^{\iota_1}, u^{\iota_1}, v^{\iota_1}), \ldots, (y^{\iota_2}, u^{\iota_2}, v^{\iota_2}), \ldots)) \\ & \mapsto (t, (\ldots, (y^{\iota_1}, u^{\iota_1}, v^{\iota_2}), \ldots, (y^{\iota_1}, u^{\iota_2}, v^{\iota_1}), \ldots)),\end{aligned} \tag{5.19}$$

and based on these maps we obtain the induced operators on functions F on these path spaces, which we denote by $\widehat{\Theta}^0_{\iota_1,\iota_2}$, resp. $\widehat{\Theta}^1_{\iota_1,\iota_2}$. Furthermore since the migration step is decoupled from the random permutation step we define (misusing notation a bit) $\widetilde{A}^{\mathcal{I}}$ by the same recipe as before simply ignoring the $[0,\rho]^{\mathcal{I}^\rho}$-valued coordinate in the argument of F.

The generator $L_{(t,X^{*,(\sigma^\rho(\cdot,\underline{v}))^{-1}},\underline{v})}$ acts on F as follows:

$$\begin{aligned}& L_{(t,X^{*,(\sigma^\rho(\cdot,\underline{v}))^{-1}},\underline{v})}F(s, \underline{(y,u,v)}) \\ & = \widetilde{A}^{\mathcal{I}^\rho}F(s, \underline{(y,u,v)}) \\ & + \frac{\gamma}{2} \sum_{\substack{\iota_1,\iota_2\in\mathcal{I}^\rho; \\ u^{\sigma^\rho(\cdot,\underline{v})^{-1}(\iota_1)} < u^{\sigma^\rho(\cdot,\underline{v})^{-1}(\iota_2)}}} 1_{\{(\pi^*_G y^{\iota_1})_s = (\pi^*_G y^{\iota_2})_s\}} (\widehat{\Theta}^0_{\iota_1,\iota_2}F - F)(s, \underline{(y,u,v)}) \\ & + \frac{\gamma}{2} \sum_{\substack{\iota_1,\iota_2\in\mathcal{I}^\rho; \\ u^{\sigma^\rho(\cdot,\underline{v})^{-1}(\iota_1)} < u^{\sigma^\rho(\cdot,\underline{v})^{-1}(\iota_2)}}} 1_{\{(\pi^*_G y^{\iota_1})_s = (\pi^*_G y^{\iota_2})_s\}} (\widehat{\Theta}^1_{\iota_1,\iota_2}F - F)(s, \underline{(y,u,v)})\end{aligned} \tag{5.20}$$

Recall from (5.9) that there is a one to one correspondence between $\underline{u}$ and $\underline{v}^\rho$. In particular, the sums on the right hand side of (5.20) run over indices ι_1, ι_2 such that $v^{\rho,\iota_1} < v^{\rho,\iota_2}$.

The next step now is to see what we can conclude from the characterization of the process defined in (5.17) by a martingale problem for the process $(t, \widetilde{\xi}^{*,\rho}_t)_{t\geq 0}$ we are interested in. We start with the following observation. Since F depends on finitely many individuals only, we may renumber the indices. To make this precise, we define a probability measure on $([0,\rho]^{\mathcal{I}^\rho}, \mathcal{B}[0,\rho])^{\mathcal{I}^\rho}$ with density

$$\Upsilon^\rho[d\underline{v}] := \bigotimes\nolimits_{\iota\in\mathcal{I}^\rho} \Phi^{\rho,\iota}[d\widehat{v}^\iota], \tag{5.21}$$

where $\Phi^{\rho,\iota}[\mathrm{d}\cdot]$ are uniform distributions on $[0,\rho]$. Integration with respect to Υ gives the identity,

$$\begin{aligned}\frac{\gamma}{2}\int \Upsilon^{\rho}[\mathrm{d}\underline{\widehat{v}}] \sum_{\substack{\iota_1,\iota_2\in\mathcal{I}^{\rho};\\ \widehat{v}^{\iota_1}<\widehat{v}^{\iota_2}}} 1_{\{(\pi^*_G y^{\widetilde{\iota}_1})_s=(\pi^*_G y^{\widetilde{\iota}_2})_s\}}(\widehat{\Theta}^1_{\iota_1,\iota_2}F-F)(s,\underline{(y,u,\widehat{v})})\\ =\frac{\gamma}{2}\int \Upsilon^{\rho}[\mathrm{d}\underline{\widehat{v}}] \sum_{\substack{\iota_1,\iota_2\in\mathcal{I}^{\rho};\\ \widehat{v}^{\iota_1}>\widehat{v}^{\iota_2}}} 1_{\{(\pi^*_G y^{\widetilde{\iota}_1})_s=(\pi^*_G y^{\widetilde{\iota}_2})_s\}}(\widehat{\Theta}^0_{\iota_1,\iota_2}F-F)(s,\underline{(y,u,\widehat{v})}).\end{aligned} \tag{5.22}$$

We can use this relation in (5.20) if we integrate in (5.22) with respect to Υ^{ρ}, and then eliminate the transition in (5.19). We apply this fact next by replacing F by functions G depending on t and $\underline{y}$ only.

Since $(t, X^{*,(\sigma^{\rho}(\cdot,\underline{v}^{\rho}))^{-1}}, (\underline{v}^{\rho})^{-1})$ is Markovian, and the functional $(t,\widetilde{\xi}^{*,\rho})$ is obtained from the graph defining X^* by random permutations depending only on the paths evolved up to time t, and on a collection of random variables independent of everything else, $(t,\widetilde{\xi}^{*,\rho})$ also solves a martingale problem for a specific set of functions. Namely, consider all functions $G:\mathbb{R}^+\times(\mathcal{D}(\mathbb{R}^+,G\times K))^{\mathcal{I}^{\rho}}$ which arise from the functions F, which do not depend on $\underline{u}$, by integration as follows:

$$G(t,\underline{y}) := \int \Upsilon^{\rho}[\mathrm{d}\underline{\widehat{v}}]\, F(t,\underline{(y,u,\widehat{v})}). \tag{5.23}$$

Then by (5.20) the process given by:

$$\left(G(t,\widetilde{\xi}^{*,\rho}_t)-\int_0^t \mathrm{d}s\, \widetilde{A}^{\mathcal{I}^{\rho}}G(s,\widetilde{\xi}^{*,\rho}_s) - \frac{\gamma}{2}\int_0^t \mathrm{d}s \sum_{\widetilde{\iota}_1\neq\widetilde{\iota}_2\in\mathcal{I}^{\rho}} 1_{\{(\pi^*_G y^{\widetilde{\iota}_1})_s=(\pi^*_G y^{\widetilde{\iota}_2})_s\}}(\Theta_{\widetilde{\iota}_1,\widetilde{\iota}_2}G-G)(s,\widetilde{\xi}^{*,\rho}_s)\right)_{t\geq 0}, \tag{5.24}$$

is a martingale with respect to the filtration $\mathcal{F}^{N,R,M}_t := \sigma((N_s,R_s,M_s)_{s\leq t})$. Here N_s,R_s,M_s are the collections of Poisson processes and random walks used to construct X^* up to time s respectively the process of the $\{0,1\}$-valued variables used till time s (recall (1.28), (1.27), (5.5), and (2.12)). Since $(s,\widetilde{\xi}^*_s)_{s\leq t}$ is $\mathcal{F}^{N,R,M}_t$-measurable by construction, the processes in (5.24) are also martingales w.r.t. the filtration generated by the process $\xi^{*,\rho}$ itself.

Now in order to conclude that $\widetilde{\xi}^{*,\rho}$ solves the martingale problem given in (2.14), we have to observe that by (5.23) we have produced a class of functions, which is sufficiently large to be measure determining for random measures on the path space under consideration. To see this observe that we start from a class reach enough to be distribution determining for paths *and* permutations.

Hence the law of the process $(t,\widetilde{\xi}^{*,\rho})$ solves the martingale problem corresponding to the generator specified in (2.14). The coefficients of these operators imply a unique solution of the martingale problem (recall Theorem 1). Furthermore the coefficients depend only on the current state of $(t,\widetilde{\xi}^{*,\rho})$ and hence the solution has the Markov property.

We need to verify the Feller property. Since we use the product topology and one path resamples only with finitely many others in finite time, and since the sharing times of a finite collection of paths is a continuous functional of the corresponding path processes, this follows from the Feller property of the path process. The Feller property of the path process holds here, since the migration is given by random walks. This finally proves the claim of Lemma 5.2. □

Step 3 (The empirical measures coincide) By (5.15), for any $t\geq 0$, there is a random bijection from the set of paths $\{\widetilde{\xi}^{*,\rho,\widetilde{\iota}}_t;\,\widetilde{\iota}\in\mathcal{I}^{\rho}\}$ to that of $\{X^{*,\iota}_t;\,\iota\in\mathcal{I}^{\rho}\}$ based on additional random coin flipping independent of everything. Hence the empirical measures of both processes coincide. To be more precise, consider finite random subsets $\mathcal{I}_i^\rho \subseteq \mathcal{I}^\rho$, $i = 1, 2$, and let

$$Z_t^{X^*}(\mathcal{I}_1^\rho) := \sum_{\iota \in \mathcal{I}_1^\rho} \delta_{\pi^*_{G\times K} X_t^{*,\iota}}, \qquad \text{and} \qquad Z_t^{\widetilde{\xi}^{*,\rho}}(\mathcal{I}_2^\rho) := \sum_{\widetilde{\iota} \in \mathcal{I}_2^\rho} \delta_{\widetilde{\xi}_t^{*,\rho,\widetilde{\iota}}}. \tag{5.25}$$

Then for any $t \geq 0$,

$$Z_t^{X^*}(\mathcal{I}_1^\rho) = Z_t^{\xi^{*,\rho}}(\sigma_t^\rho(\mathcal{I}_1^\rho)). \tag{5.26}$$

Step 4 (Particle representation of ζ^*) We are now in a position to state the main consequence.

Assume that $K := \{0, 1\}$.

Proposition 5.3. *(Particle representation of historical IFWD) Suppose we have constructed X^* and $\{\xi^{*,\rho}; \rho > 0\}$ on the same probability space as described above. Given $X^* = (X_t^{*,\iota}, \iota \in \mathcal{I})_{t\geq 0}$, let $Z^{*,\rho} := (\sum_{x\in G} Z_t^{*,\rho,x})_{t\geq 0}$ where for each $x \in G$, $Z_t^{*,\rho,x}$ is the finite random measure supported on those paths in $\mathcal{D}(\mathbb{R}, G \times K)$ which are located at x at time t given by*

$$Z_t^{*,\rho,x} := \rho^{-1} \sum_{\iota \in \mathcal{I}^\rho : \pi_G^*(X_t^{*,\iota})_t = x} \delta_{\pi^*_{G\times K} X_t^{*,\iota}} = \rho^{-1} \sum_{\widetilde{\iota} \in \sigma_t^\rho(\{\iota \in \mathcal{I}^\rho : \pi_G^*(X_t^{*,\iota})_t = x\})} \delta_{\widetilde{\xi}_t^{*,\rho,\widetilde{\iota}}}. \tag{5.27}$$

Then there exists a $\mathcal{C}(\mathbb{R}, \mathcal{M}(\mathcal{D}(\mathbb{R}, G \times K)))$-valued process, $Z^* := (\sum_{x\in G} Z_t^{*,x})_{t\geq 0}$, such that for each $t > 0$ (here $\mathcal{M}$ is equipped with the weak topology),

$$Z_t^{*,\rho} \underset{\rho\to\infty}{\longrightarrow} Z_t^*, \quad a.s., \tag{5.28}$$

In particular, $Z^* := (Z_t^*)_{t\geq 0}$ is a version of the historical IFWD.

Proof of Part (b) of Theorem 3. The above proposition is the second part of the assertion of Theorem 3. Therefore, the proof of Theorem 3 is now completed. □

Proof of Proposition 5.3. Recall $E_{\{x\},t}$ from (1.18). Notice that for each $x \in G$ and $t \geq 0$, the family $\{X_t^{*,\iota}; \iota \text{ such that } X^{*,\iota} \in E_{\{x\},t}\}$ is exchangeable. Almost surely convergence follows then by the formulation of de Finetti's theorem for random measures (compare e.g. Theorem 11.2.1 in [13]). Then Proposition 5.3 is a consequence of Theorem 0 and (5.26) and the fact that ζ^* arises as the diffusion limit of $\xi^{*,\rho}$ as $\rho \to \infty$ as described earlier. □

6. Strong duality and the coalescent

In this section we prove the results concerning our first set of applications. We begin in Subsection 6.1 by showing that the coalescent is well-defined and continue in Subsection 6.2 with verifying the strong duality. In Subsection 6.3 we prove the representation theorem for the genealogy of the equilibrium historical process via the historical coalescent.

6.1. Construction of the coalescent (Proof of Proposition 3.4).

In this section we prove Proposition 3.4. The idea is to show that each $\mathcal{E}_G^\uparrow$-marked $(\mathcal{I}, \mathcal{U})$-coalescent gives a configuration in the Liggett-Spitzer space for each positive time t provided that the geographical space G is finite. We then approximate the coalescent on the infinite Abelian group G with the coalescents on arbitrarily large finite sets, and show that the configurations for a positive time stay uniformly stochastically bounded in terms of their sizes (numbers of partitions).

In order to make the approximation work, we introduce an additional condition on the group.

Condition 6.1 (Condition on the group). *The group G is assumed to have the following property. For each $N \in \mathbb{N}$ there exists a sub-group $U := U_N \subseteq G$ and a set $\Lambda := \Lambda_N$ of representatives of the quotient group $G/_U$ such that $(\Lambda_N)_{N\in\mathbb{N}}$ is increasing to the full group.*

Remark Notice that Condition 6.1 is obviously fulfilled for the groups one is mainly interested in, i.e., $G := \mathbb{Z}^d$ or the *hierarchical group* $G = \Omega_N$ given as $\bigoplus_{\mathbb{N}} \mathbb{Z}_N$, with $\mathbb{Z}_N = \{0, 1, \dots, N-1\}$ where addition is modulo N. For the latter group we have the n-balls $\{\xi \in \oplus_{\mathbb{N}} \mathbb{Z}_N | \xi_k = 0 \text{ for } k > n\}$ as an exhausting increasing system of subgroups. To see that Condition 6.1 holds for $G = \mathbb{Z}^d$, choose $U_N := (2N+1)\mathbb{Z}^d$, and $\Lambda_N := [-N, N]^d$. □

Proof of Proposition 3.4 Let $(C_t, L_t)_{t \geq 0}$ be a $\mathcal{E}_G^{\uparrow}$-marked $(\mathcal{I}, \mathcal{U})$-coalescent.

Recall the finite reference measure α from the definition of the Liggett-Spitzer space (1.5) and denote by $|\alpha|$ the total mass of α. For $t > 0$ and $g \in G$, let

$$\#g_t := \#\{\pi \in C_t : L_t^\pi = g\} \tag{6.1}$$

be the number of partition elements located at site g at time t. Consider the coalescent (C, L) at time $t > 0$ and the Liggett-Spitzer norm of its spatial configuration $\sum_{\pi \in C_t} \delta_{L_t^\pi}$,

$$\| (C_t, L_t) \|_{\ell_\alpha^1} = \sum_{g \in G} \#g_t \alpha(g). \tag{6.2}$$

Proposition 3.4 will be proved if we show that this object is almost surely finite, since this implies in particular local finiteness. In order to obtain bounds on $||(C_t, L_t)||_{\ell_\alpha^1}$ we use the approximation of G by finite sets and coalescents on these finite sets. We shall proceed in three steps. We begin by reducing the problem to a uniform estimate (6.7) of spatially finite coalescents, and then we obtain the estimate in steps 2 and 3.

Step 1 (Approximation) Choose for each $N \in \mathbb{N}$ a subgroup U_N and a finite set Λ_N which represents the quotient group $G_{/U_N}$ from Condition 6.1, such that $\Lambda_N \uparrow G$.

We define then a family of Λ_N-marked coalescents, $(C_t^N, L_t^N)_{t \geq 0}$, as follows:

- We start the N^{th} approximating coalescent by restricting the initial state of the coalescent to the configuration corresponding to Λ_N.
- Recall that for an Abelian subgroup $U \subseteq G$ an equivalence relation in G is given by

$$g_1 \equiv_U g_2 \qquad \text{iff} \qquad g_1 - g_2 \in U. \tag{6.3}$$

  Based on the $\{\equiv_{U_N}; N \in \mathbb{N}\}$ we couple for each N the migration steps in each of the approximating coalescents taking the migration on G and then identifying sites which are equivalent modulo $\equiv_{U_N}$.
- We couple the coalescence events by using for each pair ι_1, ι_2 the exponential clocks of the coalescent on G and in addition exponential clocks for the additional sharing times occurring on sites that are equivalent with respect to $\equiv_U$.

Since $\Lambda_N \uparrow G$ we have with respect to the metric $\mathbb{d}(\cdot, \cdot)$ defined in (3.28) that

$$(C^N, L^N) \underset{N \to \infty}{\longrightarrow} (C, L), \qquad \text{a.s.}. \tag{6.4}$$

Fix $t > 0$ and let $\mathbf{P}$ and $\mathbf{P}^{U_N}$ denote the laws of the coalescent on G and the one induced by U_N (and living on the geographic space Λ_N), respectively and for each time $t > 0$ and $N \in \mathbb{N}$ define

$$\mu_t^N := \sum_{\pi \in C_t^{U_N}} \delta_{L_t^{U_N}(\pi)} \alpha(\{L^{U_N}(\pi)\}) \tag{6.5}$$

to be the partition location counting measure relative to α. We now write (compare with (6.2))

$$|\mu_t^N| = \sum_{g \in \Lambda_N} \#g_t^N \, \alpha(g) \tag{6.6}$$

for the norm (possibly infinite) of μ_t^N.

The goal is to show that $|\mu_t|$ is a proper random variable, i.e., $|\mu_t| < \infty$, a.s. by showing that $|\mu_t^N|$ is stochastically bounded *uniformly* in the parameter N. For this purpose we construct in the next two steps a deterministic function $b : (0, 1] \times (0, \infty) \mapsto (0, \infty)$ such that

$$\liminf_{N \to \infty} \mathbf{P}^{U_N} [|\mu_t^N| < b(\delta, t)] \geq (1 - \delta)^2. \tag{6.7}$$

Step 2 (Upper bound on the mean time to reach finite intensity) Fix N, identify $g_t^N = g_t$, and suppose there are countably many individuals present at each site of Λ_N at the initial time 0. For each $m \geq 1$ define τ_m to be the first time at which there are at most $m|\Lambda_N|$ partition elements present in Λ_N. We think of τ_m as the moment when the "density of partition elements in Λ_N" first drops down to m.

The next calculation shows that uniformly in $N \in \mathbb{N}$, a sequence $c(m)$ can be chosen so that:

$$\mathbf{E}^{U_N}[\tau_{m-1} - \tau_m] \leq c(m), \quad m \geq 4, \qquad \text{and} \quad \sum_{m\geq 4} c(m) < \infty. \tag{6.8}$$

This implies of course in a straightforward way that in positive time only finitely many elements remain a.s.

Denote by $A_m(s)$ the event that $\#g_s \in \{m, m-1\}$ for each $g \in U_N$. Equivalently,

$$A_m(s) = \{\tau_m \leq s \leq \tau_{m-1}\} \cap \left\{ \left| \#\bar{g}_s - \frac{\sum_g \#g_s}{|\Lambda_N|} \right| < 1, \text{ for all } \bar{g} \in \Lambda_N \right\}. \tag{6.9}$$

Observe that on $A_m(s)$ the infinitesimal rate of a coalescence is greater than $|\Lambda_N|(m-1)(m-2)/2$ since there are $|\Lambda_N|$ sites and each contributes a rate greater than $(m-1)(m-2)/2$. Now we argue that in fact on all of $\{\tau_m \leq s \leq \tau_{m-1}\}$ the infinitesimal coalescence rate is higher than $(m-1)(m-2)/2$ for all $m \geq 4$.

For this it suffices to show that

$$|\Lambda_N| \frac{\frac{\sum_g x_g}{|\Lambda_N|}\left(\frac{\sum_g x_g}{|\Lambda_N|} - 1\right)}{2} \leq \sum_g \frac{x_g(x_g-1)}{2}, \tag{6.10}$$

for all $x_g \geq 0$, $g \in \Lambda_N$, or equivalently

$$\sum_g x_g\Big(\sum_g x_g - |\Lambda_N|\Big) \leq |\Lambda_N| \sum_g x_g(x_g - 1). \tag{6.11}$$

By Jensen's inequality

$$|\Lambda_N| \sum_g x_g^2 \geq \Big(\sum_g x_g\Big)^2, \tag{6.12}$$

and now (6.12) implies (6.11) and (6.10) for all $s \leq \tau_3$.

Now we bound the expectation of $\tau_{m-1} - \tau_m$ by splitting it according to the successive jump times. Let $\tau_m =: \tau_m^0 < \tau_m^1 < \tau_m^2 < \ldots < \tau_m^{|\Lambda_N|} := \tau_{m-1}$ be the times of successive coalescents of pairs of individuals during the period where the density drops down from m to $m-1$. The above calculation implies that for each $i = 1, \ldots, |\Lambda_N|$, and $m \geq 4$ the random variable $\tau_m^i - \tau_m^{i-1}$ is stochastically dominated by an exponential (rate $|\Lambda_N|(m-1)(m-2)/2)$ random variable. Therefore

$$\mathbf{E}^{U_N}[\tau_m^i - \tau_m^{i-1}] \leq 2/(|\Lambda_N|(m-1)(m-2)), \tag{6.13}$$

and

$$\mathbf{E}^{\Lambda_N}[\tau_m - \tau_{m-1}] \leq \frac{2|\Lambda_N|}{|\Lambda_N|(m-1)(m-2)} = \frac{2}{(m-1)(m-2)} =: c(m). \tag{6.14}$$

Step 3 (Density drops uniformly in $N \in \mathbb{N}$) For $\delta > 0$ and $t > 0$, choose $m_0 := m_0(t, \delta)$ such that

$$\liminf_{|U_N| \to \infty} \mathbf{P}^{\Lambda_N}[\tau_{m_0} \leq t] \geq 1 - \delta. \tag{6.15}$$

For example, use (6.14) and Chebyshev inequality and take m large enough such that $\sum_{k\geq m} c(k) < t\delta$. Let k be large enough such that $1/k < \delta$ and set

$$b(\delta, t) := km|\alpha|. \tag{6.16}$$

Now we use the group structure. Observe next that

$$\mathbf{E}^{U_N}[\#0_t|\tau_n \le t < \tau_{n-1}] = \mathbf{E}^{U_N}[\#g_t|\tau_n \le t < \tau_{n-1}], \tag{6.17}$$

for all $g \in \Lambda_N$, by the translation invariance of the coalescent transitions. Then since $\mathbf{E}^{\Lambda_N}[\sum_{g\in\Lambda_N} \#g_t|\tau_n \le t < \tau_{n-1}] \le n|\Lambda_N|$ one gets $\mathbf{E}^{U_N}[\#0_t|\tau_n \le t < \tau_{n-1}] \le n$, and therefore

$$\mathbf{E}^{U_N}\left[|\mu_t^N|\,|\tau_n \le t < \tau_{n-1}\right] = \mathbf{E}^{U_N}\Big[\sum_{g\in\Lambda_N} \#g_t\alpha(g)\,\big|\tau_n \le t < \tau_{n-1}\Big] \le \sum_{g\in\Lambda_N} n\,\alpha(g) \le n|\alpha|. \tag{6.18}$$

This implies via Chebyshev inequality that:

$$\mathbf{P}^{U_N}\left[|\mu_t^N| < kn|\alpha|\,\big|\tau_n \le t < \tau_{n-1}\right] \ge 1 - 1/k. \tag{6.19}$$

Now we get the following chain of estimates:

$$\begin{aligned}\mathbf{P}^{U_N}\left[|\mu_t^N| < b(\delta,t)\right] &\ge \mathbf{P}^{U_N}\left[|\mu_t^N| < b(\delta,t), \tau_m \le t\right]\\ &\ge \sum_{n=1}^m \mathbf{P}^{U_N}\left[|\mu_t^N| < km|\alpha|\,\big|\tau_n \le t < \tau_{n-1}\right]\mathbf{P}^{U_N}\left[\tau_n \le t < \tau_{n-1}\right]\\ &\ge \sum_{n=1}^m \mathbf{P}^{U_N}\left[|\mu_t^N| < kn|\alpha|\,\big|\tau_n \le t < \tau_{n-1}\right]\mathbf{P}^{U_N}\left[\tau_n \le t < \tau_{n-1}\right]\\ &\ge \sum_{n=1}^m \left(1-\frac{1}{k}\right)\mathbf{P}^{U_N}\left[\tau_n \le t < \tau_{n-1}\right] \ge \left(1-\frac{1}{k}\right)\mathbf{P}^{U_N}\left[\tau_m \le t\right] \ge (1-\delta)^2.\end{aligned} \tag{6.20}$$

Since $\delta > 0$ was arbitrary and $b = b(\delta, t) < \infty$ does not depend on N, we have (6.7) and the proposition follows. □

6.2. Proofs of duality relations (including Proposition 3.5 and Corollary 3.7).

Proof of Proposition 3.5 First consider the system with $\gamma = 0$ (i.e. pure migration). Note that under the law of X_0^*, the position at time t of the individual indexed by $\iota \in \mathcal{I}$ is precisely $R^\iota(x^\iota, t)$ and its reversed path is generated by $\overline{a}(x, y)$. Moreover, the joint law for the paths of any collection of individuals in reversed time from t back to 0, is the same as the joint law for the paths of this collection of individuals in forward time from 0 to t, if started from $R^u(x^u, t), u \in \mathcal{I}$ and using $\overline{R}^u(x^u, t)$. Now we use the special form of the construction of the individuals' labels to get information about the initial state of the reversed system.

Since Poisson systems are equilibria both for $a(x, y)$ and $\bar{a}(x, y)$, the positions are Poisson at time t if they are Poisson at time $t = 0$.

Therefore if we now incorporate the resampling, we see immediately from the construction of the resampling times via Poisson point processes that:

$$(\Gamma_s^t, \ell_s^t)_{s\in[0,t]} \stackrel{d}{=} (C_s, L_s)_{s\in[0,t]}. \quad \square \tag{6.21}$$

Proof of Corollary 3.6 This is a consequence of Identity (3.7) if we note that by construction of the initial state X_0^*, the restriction to individuals with labels less than ρ gives positions forming a Poisson system with intensity ρ. □

Proof of Corollary 3.7 **(a)** Fix time $t > 0$ and sites $x_1, x_2, ..., x_k$, and let

$$\chi_t^\rho := \{\eta_t^\rho(\{x\} \times K);\, x \in G\} \tag{6.22}$$

be the total number of individuals located at the sites of G at time t. W.l.o.g. we assume that we are on the event $\{\prod_{j=1}^k \left(\chi_t^\rho(\{x_j\}) - \sum_{l=j+1}^k 1_{\{x_l = x_j\}}\right) \ge 1\}$ (otherwise the functional under consideration is zero anyway). We then order the individuals at each site in some arbitrary but fixed way. Now use the fact that the probability that a k-tuple of different individuals sampled at time t from $x_1, ..., x_k$ has type 1 equals the probability that all

Γ_t^t independent families, which this k-tuple forms by tracing back the genealogy, are of type 1. Hence since the spatial configurations are independent of the initial types,

$$
\begin{aligned}
&\mathbf{E}\Big[\prod_{j=1}^k\Big(\eta_t^\rho(\{x_j\}\times\{1\})-\sum_{l=j+1}^k 1_{\{x_l=x_j\}}\Big)\Big]\\
&=\mathbf{E}\big[\sum_{\iota_1\in\mathcal{I}^\rho}\sum_{\iota_2\neq\iota_1\in\mathcal{I}^\rho}\cdots\sum_{\iota_k\notin\{\iota_1,\dots,\iota_{k-1}\}\in\mathcal{I}^\rho}\prod_{j=1}^k 1\{x_t^{\iota_j}=x_j;\, k_t^{\iota_j}=1\}\big]\\
&=\mathbf{E}\Big[\prod_{j=1}^k\Big(\chi_t^\rho(\{x_j\})-\sum_{l=j+1}^k 1_{\{x_l=x_j\}}\Big)\,\mathbf{E}^{\mu(x_1,\dots,x_k)}\Big[\theta^{|\Gamma_t^t|}\Big]\Big]\\
&=\rho^k\,\mathbf{E}^{\mu(x_1,\dots,x_k)}\Big[\theta^{|C_t|}\Big].
\end{aligned}
\tag{6.23}
$$

Notice that the second equality above follows by Proposition refL4.1a, since $\theta^{|\Gamma_t^t|}$ is the probability that family clusters of size $|\Gamma_t^t|$ have the same type 1. For the third equality we use that Poisson systems are preserved under migration and that $E[X(X-1)\cdots(X-k+1)]=\rho^k$ for X Poisson distributed with parameter ρ.

(b) Part(b) is a consequence of part(a), Theorem 0 and uniform integrability, combined with the fact that, for a fixed t and θ,

$$
\frac{\prod_{j=1}^k\Big(\chi_t^\rho(\{x_j\})-\sum_{l=j+1}^k 1_{\{x_l=x_j\}}\Big)}{\prod_{j=1}^k\chi_t^\rho(\{x_j\})}\underset{\rho\to\infty}{\longrightarrow}1,\qquad a.s.\quad\square
\tag{6.24}
$$

6.3. **Representations for the historical look-down (Proof of Theorem 4).** In this section we prove Theorem 4 and Corollary 3.9.

Proof of Theorem 4. We have to prove here (3.37) only, since the remaining claims in (a) and (b) are immediate consequences of the definitions and the properties of the coalescent C, and in particular C_∞, in the case of recurrent and transient kernel $\widehat{a}(x,y)$.

The basic idea now is to obtain the result from the result for systems started in Poisson systems at time $-s$ via the strong form of duality and the restriction property of coalescents. For that purpose we need to approximate our system X^* by such special ones.

Recall the metric $\mathbb{d}(\cdot,\cdot)$ defined in (3.28), via non-negative numbers α_{ι_1,ι_2}, $\iota_1,\,\iota_2\in\mathcal{I}$ such that $\sum_{\iota_1,\,\iota_2}\alpha_{\iota_1,\,\iota_2}=1$.

Fix a time $-s$. Each individual in G has index $\iota\in\mathcal{I}$, label u^ι, some initial type k^ι in K, some initial position x^ι in G, and also some rank in $\mathbb{N}$ such that no two individuals located initially at the same site are of equal rank. To the individual located at g of rank $i\in\mathbb{N}$ assign the random variable $\mathcal{E}_{g,i}$, where $\{\mathcal{E}_{g,i}: g\in G, i\in\mathbb{N}\}$ is a family of independent exponential (rate 1) random variables.

Pick $\rho\in[0,\infty)$ large, use the exponential random variables $\mathcal{E}_{g,i}$ to *tag* only Poisson(ρ) many individuals from each site. More precisely, a particle at site g of rank j is tagged iff $\sum_{i\le j}\mathcal{E}_{g,i}\le\rho$. Call the (random) subset of indices picked in this way $\mathcal{V}(\rho)$. With probability higher than $1-\varepsilon(\rho)$,

$$
\sum_{\iota_1,\iota_2\in\mathcal{I}\cap\mathcal{V}(\rho)}\alpha_{\iota_1,\,\iota_2}\ge 1-\varepsilon(\rho),
\tag{6.25}
$$

where $\varepsilon(\rho)\to 0$ as $\rho\to\infty$.

Since the initial (time $-s$) configuration of individuals with indices in $\mathcal{V}(\rho)$ is a Poisson(ρ) field, the paths of these individuals followed in reversed time are independent random walk paths with transitions according to the reversed kernel $\bar{a}$. Therefore it is clear that the distribution of the system of tagged individuals is equal to the distribution of

$$
\Delta_\theta(\phi_{\mathcal{V}(\rho)}(C_s^*,L_s^*)).
\tag{6.26}
$$

As $s \to \infty$ we get that

$$\Delta_\theta(\phi_{\mathcal{V}(\rho)}(C_s^*, L_s^*)) \Rightarrow \Delta_\theta(\phi_{\mathcal{V}(\rho)}(C_\infty^*, L_\infty^*)). \tag{6.27}$$

Since with $\rho \to \infty$, $\varepsilon(\rho) \to 0$, (6.25) implies that for every partition $\mathcal{P}$ of $\mathcal{I}$ we have for the restriction $\mathbb{d}(\phi_{\nu(\rho)}\mathcal{P}, \mathcal{P}) \to 0$ in probability, as $\rho \to \infty$. Therefore as $\rho \to \infty$, we get

$$\Delta_\theta((C_\infty^*, L_\infty^*)), \tag{6.28}$$

which proves the statement (3.37) of the theorem. □

Proof of Corollary 3.9. Let $\{\Delta_\theta(\pi);\, \pi \subseteq \mathcal{I}\}$ be a family of independent and identically distributed $\{0,1\}$-valued random variables with $\mathbf{P}[\Delta_\theta(\pi) = 1] = \theta$. For each $\rho \in [0, \infty)$, define $f^\rho : (\mathcal{D}([-\infty, 0], G \times K))^{\mathcal{I}^\rho} \to \mathcal{N}(\mathcal{D}([-\infty, 0], G \times K))$ by

$$f^\rho(\underline{y}) := \sum\nolimits_{\iota \in \mathcal{I}^\rho} \delta_{y^\iota}. \tag{6.29}$$

By continuity of f^ρ with respect to the metric on $(\mathcal{D}((-\infty, 0], G \times K))^{\mathcal{I}}$ introduced above Lemma 3.8 (which is the product topology of Skorokhod topologies on the path space obtained by a suitable choice of the index set $\mathcal{I}$ and weak topology on $\mathcal{N}(\mathcal{D}([-\infty, 0], G \times K))$, (6.27) implies also that, for $s \to \infty$,

$$(f^\rho)^{-1}\Delta_\theta(\phi_{\mathcal{V}(\rho)}(C_s^*, L_s^*)) \Rightarrow (f^\rho)^{-1}\Delta_\theta(\phi_{\mathcal{V}(\rho)}(C_\infty^*, L_\infty^*)). \tag{6.30}$$

Since $\widetilde{\Delta}^*_{\rho,\theta} = (f^\rho)^{-1}\Delta_\theta(\phi_{\mathcal{V}(\rho)}(C_\infty^*, L_\infty^*))$, the latter yields $\widetilde{\mathcal{H}}^*_{\rho,\theta} = \mathcal{L}^\Psi[\widetilde{\Delta}]$.

Fix $A \subseteq G$. Recall r_A from (3.41) and (1.18). Then

$$\widehat{r}_A \widetilde{\mathcal{H}}^*_{\rho,\theta} = \mathcal{L}^\Psi\left[(f^\rho)^{-1}\Delta_\theta(\phi_{\mathcal{V}(\rho)}(C_\infty^*, L_\infty^*))(E^{A,0})\right]. \tag{6.31}$$

Now under Ψ, $N := \#\{\iota \in \mathcal{I}^\rho : y^\iota \in E^{A,0}\}$ is Poisson distributed with parameter $\rho|A|$, and given N the set of positions $\{(y^\iota)_0;\, \iota \in \mathcal{I}^\rho : y^\iota \in E^{A,0}\}$ consists of independent and on A uniformly distributed points. This gives (3.42). □

7. The finite system scheme for historical IMM

In this chapter we prove the theorems about large finite systems. We begin by constructing the whole sequence of processes and the limiting infinite system on one probability space, i.e., we give an explicit coupling. Then we analyze the corresponding look-down processes using the strong duality with the coalescent. We define the coupling in Subsection 7.1, we analyze the coalescent in Subsection 7.2 and Subsection 7.3, and, based on these results, we prove Theorem 6 and two corollaries in Subsection 7.4.

7.1. Coupled representations for the look-down genealogies.

Recall the look-down process as introduced in Subsection 2.2. The goal of this subsection is to construct the look-down process, X^*, on the countable group $\mathbb{Z}^d$, and a sequence of look-down processes, $X^{*,N}$, on the finite groups $G_N = [-N, N] \cap \mathbb{Z}^d$, $N \in \mathbb{N}$, with addition modulo $2N+1$, on one and the same probability space. More specifically, this will be on the probability space for the infinite system.

For a given $N \in \mathbb{N}$, the coupling is achieved

- by using as initial state the initial state of X^* but restricted to all individual paths initially positioned on G_N. In particular, use as index set $\mathcal{I}^{G_N} \subset \mathcal{I}$ obtained by that restriction,
- by sampling for each index $\iota \in \mathcal{I}$ independently of the others a realization of the RW, and restrict the achieved collection to the sub-collection associated with a index $\iota \in \mathcal{I}^{G_N}$ belonging to the restricted configuration on G_N,
- by identifying two positions of an individual for the process $X^{*,N}$ if they are equal modulo $(2N+1)$,
- and by choosing the sampling symmetrization variables L as in (5.4) for the infinite system, and using for the finite system the ones obtained by restricting to pairs of individuals both located initially on G_N.

This gives us all the ingredients we need for the representation of the genealogy processes gen(X^*) and gen$(X^{*,N})$, (compare with (4.1)) on one and the same probability space. As done before in (3.3), for a fixed time t, we define for each $s \in [0,t]$ the equivalence relation, $\approx_s^{X^*}$, on $\mathcal{I}$ by relating labels which are going back to a common ancestor at time $t-s$. This defines again the genealogy process, $(\Gamma_s^t, \ell_s^t)_{s \le t}$, which reflects the equivalence classes, called family clusters, and the clusters' positions. The genealogy process,

$$(\Gamma_s^{t,N}, \ell_s^{t,N})_{s \le t}, \tag{7.1}$$

for the system on the torus is then achieved by defining $\approx_s^{X^{*,N}}$ based on gen$(X^{*,N})$.

We apply now the strong duality. Recall from Definition 3.3 the $\mathcal{E}_G^{\uparrow}$-marked $(\mathcal{I},\mathcal{U})$-coalescent, $(C_t, L_t)_{t \ge 0}$. If $(C_t^N, L_t^N)_{t \ge 0}$ is an $\mathcal{E}_{G_N}^{\uparrow}$-marked $(\mathcal{I}^{G_N},\mathcal{U}^{G_N})$-coalescent then it is clear from the construction that the strong duality stated in Proposition 3.5 applies as follows:

Lemma 7.1. *(The $\mathcal{I}^{G_N}$-coalescent is generated by the $\mathcal{I}^{G_N}$-genealogy) Assume that the look-down process starts from initial state such that the labels from $\mathcal{U}$ are chosen as explained before Proposition 3.5. Then*

$$\mathcal{L}^{\mu^N}\left[\left(\Gamma_{t(N)}^{t(N),N}, \ell_{t(N)}^{t(N),N}\right)\right] = \mathcal{L}^{\mu^N}\left[\left(C_{t(N)}^N, L_{t(N)}^N\right)\right], \tag{7.2}$$

where μ arises by the construction given in (3.36) and

$$\mu^N := \mu\big|_{G_N}. \tag{7.3}$$

We therefore see that in order to prove our theorems on large finite systems all we have to do is to analyze the behavior of the coalescent on the right hand side of (7.2) and combine it with what we know about the coalescent $(C_t, L_t)_{t \ge 0}$ for the infinite system.

Remark Note furthermore that again we can construct all processes

$$\{(C^N, L^N);\, N \in \mathbb{N}\} \quad \text{and} \quad (C, L) \tag{7.4}$$

on one and the same probability space, following the device given above for the IMM. The straightforward modification is left to the reader. We shall in the sequel in this chapter refer to $\mathbf{P}$ as the probability law of these coupled processes. □

7.2. **Preliminaries: Asymptotic analysis of the coalescent.** Since by Proposition 3.5 the scaling behavior for the systems on the tori is translated to the scaling behavior for the genealogy, it is enough to study the asymptotic behavior of the coalescent. In this subsection we therefore focus on a description of the coalescent. Recently the coalescent results of this section were extended and generalized in the setting of spatial coalescents with multiple collisions, cf. [41].

Recall Kingman's coalescent, $(K_t)_{t \ge 0}$, from (3.56).

Recall the topology introduced previous to (3.33). Now we can study the behavior of $(C_{t(N)}^N, L_{t(N)}^N)$ in various time scales $t(N)$ as $N \to \infty$ and get the following.

Proposition 7.2. *(Coalescent asymptotics) Fix a sequence $t(N) \uparrow \infty$, as $N \to \infty$. If K is a version of the Kingman's coalescent K independent of C_∞, then*

$$\mathcal{L}^{(C_0,L_0)|_{G_N}}[C_{t(N)}^N] \underset{N\to\infty}{\Longrightarrow} \begin{cases} \mathcal{L}^{(C_0,L_0)}[C_\infty], & \text{if } t(N) = o((2N+1)^d), \\ \mathcal{L}^{(C_0,L_0)}[K_{\kappa t}(C_\infty)], & \text{if } t(N) \sim t(2N+1)^d, \\ \mathcal{L}[K_\infty], & \text{if } t(N) \gg (2N+1)^d. \end{cases} \tag{7.5}$$

Proof The proof splits into several steps: The basic idea is to assume in the beginning that $|C_0|$ is finite and then to upgrade the results to the general case. We treat the finite case in Steps 2 and 3 after some preparation in Step 1, and conclude with the general case in Steps 4 and 5 where we verify an estimate on the decay of $\mathbf{E}[C_t^N]$ needed during the reasoning. Throughout the argument we can use extensions of some ideas and calculations which were developed by Bramson and Griffeath [4] and Cox [5] to study the longtime behavior of the Voter model.

Step 1 (Preparation) The argument relies on coalescing random walk estimates given in Cox (1989) [5] for the special case of simple symmetric random walk and for instantaneously coalescing individuals rather than coalescence with a rate γ delay as in our situation. These two obstacles have to be removed. The first one is technical, while the second one changes things quantitatively, i.e., the constant g has to be replaced by $g + 2/\gamma$.

Let us turn to the first point. By checking the proofs in [5], one sees that what is really needed is the expansion of the transition kernel of the random walk by Corollary 22.3 in Bhattacharya and Rao (1976) [2]. These are however valid, even in continuous time, if we assume (3.50). Therefore we can use the assertions in [5] throughout our argument.

Now we come to the new part of the argument, namely analyzing the delayed coalescent. Assume that for every $N \geq 1$,

$$|C_0^N| = k. \tag{7.6}$$

Suppose furthermore that N is large enough so that the given initial positions, $z_i(0)$, $i = 1, ..., k$, are contained in G_N. Say that a random walk particle on the torus G_N *wraps around* G_N *by time* t, if it visits the complement of the box $[-N, N]^d$ before time t. The steps that this particle makes before it wraps around G_N are identical to the steps of a random walk particle on $\mathbb{Z}^d$ before the first visit to the complement of $[-N, N]^d$. Moreover, we will see that if we let

$$\widehat{\tau}^N := \inf\{s > 0 : \sup_{\pi \in C_s^N} |L_s^N(\pi)| \geq N\}, \tag{7.7}$$

then

$$\lim_{\varepsilon \to 0} \inf_N \mathbf{P}[\widehat{\tau}^N > \varepsilon N^2] = 1. \tag{7.8}$$

In words, by time εN^2 no partition elements have positions that have wrapped around the torus G_N, with high probability. Again the paths of (C^N, L^N) and (C, L) can be made identical until time $\hat{\tau}^N$. Since there are only k individuals present at time 0 (due to coalescence there may be even fewer partition elements present at time εN^2), it suffices to show (7.8) for a single particle, i.e.,

$$\lim_{\varepsilon \to 0} \inf_N \mathbf{P}[\sup_{0 \leq s \leq \varepsilon N^2} |X_s| < N] = 1\,, \tag{7.9}$$

where X_s is a random walk on $\mathbb{Z}^d$ started at 0. The last statement is true due to the functional CLT.

Let a_N be a sequence such that $a_N = o(N)$, as $N \to \infty$. In order to make use of estimates in [5], we need for every fixed $\varepsilon > 0$:

$$\lim_{N \to \infty} \mathbf{P}[|L^N_{\varepsilon N^2}(\pi) - L^N_{\varepsilon N^2}(\pi')| \geq a_N, \;\; \forall \pi, \pi' \in C^N_{\varepsilon N^2}, \pi \neq \pi'] = 1. \tag{7.10}$$

In words, at time εN^2, all the partition individuals are at mutual distance greater than a_N, with high probability. For (7.10) it is again enough to show that

$$\lim_{N \to \infty} \mathbf{P}[|X^1_{\varepsilon N^2} - X^2_{\varepsilon N^2}| < a_N] = 0, \tag{7.11}$$

where X^1, X^2 are two independent random walks on $\mathbb{Z}^d$ started at 0, which again holds due to functional CLT, and the assumption that $a_N = o(N)$.

Step 2 (Finite case with distant individuals) The next ingredient in the proof of Proposition 7.2 is the following lemma analogous to Theorem 5 in [5].

Lemma 7.3. *(Asymptotics for distant individuals) Let* (C^N, L^N) *be the* $\{1, ..., n\}$*-coalescent, and fix a sequence* $(a_N) \underset{N\to\infty}{\longrightarrow} \infty$ *such that* $a_N = o(N)$*, as* $N \to \infty$*, and* $T \in (0, \infty)$*. Then uniformly in* $0 \leq t \leq T$*, and initial* *positions* $L_0^N := \{L_0^N(\{i\}),\, i = 1, ..., n\}$ *such that* $|L_0^N(\{i\}) - L_0^N(\{j\})| \geq a_N$*, for* $i \neq j$*, and uniformly in cylinder events* B *on the space of partitions of* $\{1, 2, \dots, n\}$*,*

$$\mathbf{P}[C^N_{t(2N+1)^d} \in B] \underset{N\to\infty}{\longrightarrow} \mathbf{P}[K_{2t/(g+2/\gamma)} \in B]. \tag{7.12}$$

Proof The argument is based on estimates given in [5], but it is adapted to the delayed coalescence. First consider the above system with $n = 2$. Then it suffices to show

$$\begin{aligned}\mathbf{P}[C^N_{t(2N+1)^d} = \{1,2\}] &= \mathbf{P}[|C^N_{t(2N+1)^d}| = 2] \\ &\underset{N\to\infty}{\longrightarrow} 1 - e^{-2t/(g+2/\gamma)} \\ &= \mathbf{P}[K_{2t/(g+2/\gamma)} = \{1,2\}]\,.\end{aligned} \tag{7.13}$$

Let

$$\tau^N = \inf\{s \geq 0 : |C^N_s| = 1\} \tag{7.14}$$

be the *coalescing time for the two individuals*. We turn now to the look-down processes and claim that

$$\mathbf{P}[\tau^N > (2N+1)^d t] \underset{N\to\infty}{\longrightarrow} e^{-2t/(g+2/\gamma)}. \tag{7.15}$$

In our setting the coalescing random walks coalesce only after an exponential time if located at the same site. In order to estimate the total length of time "in contact" up to time $t(2N+1)^d$, we make use of the following coupling.

Let X^1, X^2 (we suppress here the N dependence in the notation) be two non coalescing independent random walks on the torus G_N with initial states $X_0^i = L^N(\{i\})$, $i = 1, 2$. Let Z be an exponential (rate γ) random variable independent of X^1, X^2. One can construct a pair $(\widetilde{Y}^1, \widetilde{Y}^2)$ of coalescing random walks as a function of X^1, X^2 and Z by letting $S_s := \int_0^s 1_{\{X_u^1 = X_u^2\}}\, du$, and putting $(\widetilde{Y}_s^1, \widetilde{Y}_s^2) = (X_s^1, X_s^2)$ for $S_s < Z$ and $(\widetilde{Y}_s^1, \widetilde{Y}_s^2) = (X_s^1, X_s^1)$ for $S_s \geq Z$. Clearly we have $\tau^N \overset{d}{=} \inf\{s : S_s = Z\}$. Consider the difference process

$$Y := (X_s^1 - X_s^2)_{s\geq 0}. \tag{7.16}$$

Note that $|Y_0| > a_N$. The time interval $[0, \tau^N]$ decomposes into excursions of Y away from 0 where the first "excursion" is incomplete. We call excursions "small" if Y stays bounded by a_N during their lifetime, and we call it "large" otherwise. Let

$$\sigma_1^N := \inf\{s \geq 0 : Y_s = 0\}, \tag{7.17}$$

and $\beta_1^n := \sigma_1^N$, and for $k \geq 2$,

$$\alpha_k^N := \inf\{s \geq \beta_{k-1}^N : |Y_s| > a_N\}, \tag{7.18}$$

and

$$\beta_k^N := \inf\{s \geq \alpha_k^N : Y_s = 0\}. \tag{7.19}$$

In this way the stopping times $\beta_1^N, \beta_2^N, \beta_3^N, \dots$ form a sequence of end-times of successive "large" excursions for the process Y. Define for $k \geq 2$,

$$\sigma_k^N := \beta_k^N - \alpha_k^N. \tag{7.20}$$

During the intervals $[\beta_k^N, \alpha_{k+1}^N]$ there may (and typically will) be many small excursions of the process Y. We will soon estimate their number and combined lengths.

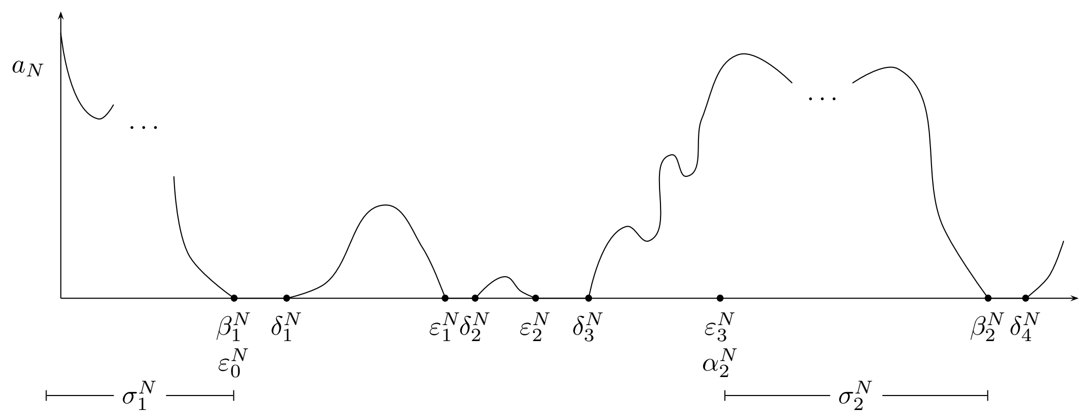

Figure 1

Let $\varepsilon_0^N := \sigma_1^N$, $\delta_1^N := \inf\{s > \beta_1^N : |Y_s| > 0\}$ and define recursively the following stopping times: for $k \geq 1$, let $\varepsilon_k^N := \inf\{s > \delta_k^N : Y_s = 0 \text{ or } |Y_s| > a_N\}$ and if $Y_{\varepsilon_k^N} = 0$ then let $\delta_{k+1}^N := \inf\{s > \varepsilon_k^N : |Y_s| > 0\}$, while if $|Y_{\varepsilon_k^N}| > a_N$ in which case ε_k^N must equal α_j^N for some (random) j almost surely, let $\delta_{k+1}^N := \inf\{s > \beta_j^N : |Y_s| > 0\}$. The realization of Y in Figure 1 has $\varepsilon_3^N = \alpha_2^N$.

Note that almost surely σ_k^N and all the stopping times above are finite for all $k \geq 1$.

Moreover the random number $J_N := \inf\{j \geq 1 : |Y_{\varepsilon_j^N}| > a_N\}$ has geometric distribution with success probability

$$p^N := \mathbf{P}[|Y_{\varepsilon_1^N}| > a_N]. \tag{7.21}$$

Also note that given $J_N = m$ the sequence $(Z_j := \delta_j^N - \varepsilon_{j-1}^N;\ m \geq j \geq 1)$, i.e., the time lengths of successive intervals during which Y equals 0 prior to the first large excursion, are independent and identically distributed according to a rate 2 exponential random variable. Let g denote the expected number of returns of Y to zero, and

$$p^* := (g-1)/g \tag{7.22}$$

be the probability that the underlying random walk on $\mathbb{Z}^d$ ever returns to its starting point. Then obviously

$$p^N \underset{N\to\infty}{\longrightarrow} p^*. \tag{7.23}$$

Let $K_N := \max\{k : \beta_k^N \leq \tau^N\}$ be the number of large excursions before τ^N (recall from (7.14)). Since the migration and resampling are independent, for an i.i.d. sequence $(Z_i)_{i\geq 1}$ with $\mathbf{P}[Z_1 \geq t] = \exp[-2t]$, it easy to see that K_N is a geometric random variable, independent of σ_j^N, $j \geq 1$, and $(Z_i)_{i\geq 1}$, with success probability

$$\begin{aligned} p(K_N) &= 1 - \mathbf{E}[\exp[-\gamma \sum_{j=1}^{J_N} Z_j]] \\ &= 1 - \sum_{k\geq 1} (1-p^N)(p^N)^{k-1} \left(\frac{2}{2+\gamma}\right)^k \\ &= \gamma\,(2(1-p^N)+\gamma)^{-1} \underset{N\to\infty}{\longrightarrow} g(g+2/\gamma)^{-1}. \end{aligned} \tag{7.24}$$

Let $L_N := \max\{k : \varepsilon_k^N \leq \tau^N\}$ be the total number of excursions of Y before time τ^N. Since the migration and resampling are independent, L_N is a geometric random variable with success probability $\gamma/(2+\gamma)$ (which is the probability to leave distance zero before the coalescence may happen). Finally note that with Z an independent rate γ exponential random variable which bounds from above the time Y spends in zero without coalescing,

$$\sum_{j=1}^{K_N}\sigma_j^N \le \tau^N \le \sum_{j=1}^{K_N}\sigma_j^N + \sum_{j=1}^{L_N}(\varepsilon_j^N - \delta_j^N) + Z. \tag{7.25}$$

The asymptotic result (1.9) in [5] says that uniformly in $0 \le u \le T$ and k we have

$$\mathbf{P}[\sigma_k^N > (2N+1)^d u] \underset{N\to\infty}{\longrightarrow} e^{-2u/g} \tag{7.26}$$

which implies that uniformly in $0 \le t \le T$,

$$\mathbf{P}[\sum_{k=1}^{K_N}\sigma_k^N > t(2N+1)^d] \underset{N\to\infty}{\longrightarrow} e^{-2t/(g+2/\gamma)}. \tag{7.27}$$

At the same time, $\mathbf{E}[\varepsilon_j^N - \delta_j^N] = o(a_N^2)$ so that $\mathbf{E}[\sum_{j=1}^{L_N}(\varepsilon_j^N - \delta_j^N)] = o(a_N^2)$.

Therefore

$$\mathbf{P}[\sum_{j=1}^{L_N}(\varepsilon_j^N - \delta_j^N) + Z > \varepsilon(2N+1)^d] \underset{N\to\infty}{\longrightarrow} 0, \qquad \text{for all } \varepsilon > 0. \tag{7.28}$$

The above estimates together with (7.25) gives (7.15), or equivalently (7.13). Hence we are done with the case $|C_0^N| = 2$.

Now the same induction argument as in the proof of [5], Theorem 5 gives for $n = |C_0^N| \ge 2$ and $1 \le k \le n$ that

$$\mathbf{P}\left[|C^N_{t(2N+1)^d}| = k\right] \to \mathbf{P}\left[|K_t| = k\right], \tag{7.29}$$

uniformly in $0 \le t \le T$, and the initial positions as in the assumptions of the lemma. A simple calculation shows that under the assumptions of the lemma, the conditional probability

$$\mathbf{P}\left[C^N_{t(2N+1)^d} = \pi_0^{i,j} \big| |C^N_{t(2N+1)^d}| = n-1\right] \underset{N\to\infty}{\longrightarrow} \binom{n}{2}^{-1}, \qquad i < j, \tag{7.30}$$

where $\pi_0^{i,j} := \{\{i,j\},\{1\},\ldots,\{i-1\},\{i+1\},\ldots,\{j-1\},\{j+1\},\ldots,\{n\}\}$.

Furthermore

$$\mathbf{P}\left[C^N_{t(2N+1)^d} \in B \big| |C^N_{t(2N+1)^d}| = l\right] \underset{N\to\infty}{\longrightarrow} \mathbf{P}\left[K_t \in B \big| |K_t| = l\right], \tag{7.31}$$

for all B sets of partitions with l equivalence classes, implying the lemma. □

Step 3 (Conclusion finite case) We now prove Proposition 7.2 for finite initial states. Recall $\widehat{\tau}^N$ from (7.7), and define events

$$\begin{aligned} E_1^N &= \{\widehat{\tau}^N > \varepsilon N^2\} \\ E_2^N &= \{|L^N_{\varepsilon N^2}(\pi) - L^N_{\varepsilon N^2}(\pi')| \ge a_N, \forall \pi,\pi' \in C^N_{\varepsilon N^2}, \pi \ne \pi'\} \\ E_3^N &= \{C_\infty = C_{\varepsilon N^2}\}. \end{aligned} \tag{7.32}$$

Recall (7.8) and (7.10) to see that we can fix a small $\delta > 0$, and choose some small enough $\varepsilon > 0$ and large enough $N_\varepsilon < \infty$ so that (using transience in $d \ge 3$) for this ε and for all $N \ge N_\varepsilon$,

$$\inf_N \mathbf{P}[E_1^N] \ge 1 - \delta/4 \tag{7.33}$$

and

$$\min\{\inf_N \mathbf{P}[E_2^N], \inf_N \mathbf{P}[E_3^N]\} \ge 1 - \delta/4. \tag{7.34}$$

We have then by the construction that there exists $\varepsilon > 0$ and a large enough $\widetilde{N}_\varepsilon < \infty$ such that for all $N \geq \widetilde{N}_\varepsilon$

$$\begin{aligned}
1 - 3\frac{\delta}{4} &\leq \mathbf{P}[E_1^N \cap E_2^N \cap E_3^N] \\
&\leq \mathbf{P}\left[\left\{(C^N_{\varepsilon N^2}, L^N_{\varepsilon N^2}) = (C_{\varepsilon N^2}, L_{\varepsilon N^2})\right\} \cap E_2^N \cap E_3^N\right] \\
&= \mathbf{P}\left[\left\{(C^N_{\varepsilon N^2}, L^N_{\varepsilon N^2}) = (C_\infty, L_{\varepsilon N^2})\right\} \cap E_2^N \cap E_3^N\right] \\
&\leq \mathbf{P}\left[\left\{C^N_{t(N)-\varepsilon^2 N}(C^N_{\varepsilon N^2}, L^N_{\varepsilon N^2}) = C^N_{t(N)-\varepsilon^2 N}(C_\infty, L^N_{\varepsilon N^2})\right\} \cap E_2^N \cap E_3^N\right].
\end{aligned} \tag{7.35}$$

Therefore, if B is a cylinder event on the space of partitions of $\{1, \dots, k\}$, then

$$\mathbf{P}[C^N_{t(N)} \in B] = \mathbf{P}[\{C^N_{t(N)-\varepsilon^2 N}(C_\infty, L^N_{\varepsilon N^2}) \in B\} \cap E_2^N \cap E_3^N] + r^N(B), \tag{7.36}$$

where the remainder $r^N(B)$ satisfies:

$$\sup_{N \geq N_\varepsilon} \sup_B |r^N(B)| \leq 3\delta/4. \tag{7.37}$$

Now suppose that $t(N)/(2N+1)^d \underset{N\to\infty}{\longrightarrow} s$, where $s > 0$ is fixed. Since $\frac{(t(N)-\varepsilon N^2)}{(2N+1)^d} \underset{N\to\infty}{\longrightarrow} s$ (recall $d \geq 3$), Lemma 7.3 yields that on $E_2^N \cap E_3^N$,

$$\Big(\mathbf{P}[C^N_{t(N)-\varepsilon N^2}(C_\infty) \in B | C_\infty, L^N_{\varepsilon N^2}] - \mathbf{P}[K_{2s/(g+2/\gamma)}(C_\infty) \in B | C_\infty, L^N_{\varepsilon N^2}]\Big) \underset{N\to\infty}{\longrightarrow} 0. \tag{7.38}$$

Hence

$$\begin{aligned}
&\mathbf{E}\left[\mathbf{P}[C^N_{t(N)-\varepsilon N^2}(C_\infty) \in B | C_\infty, L^N_{\varepsilon N^2}] 1_{E_2^N \cap E_3^N}\right] \\
&\qquad - \mathbf{P}[K_{2s/(g+2/\gamma)}(C_\infty) \in B, E_2^N \cap E_3^N] \underset{N\to\infty}{\longrightarrow} 0,
\end{aligned} \tag{7.39}$$

which finally implies that

$$\lim_{N\to\infty} \Big|\mathbf{P}[\{C^N_{t(N)-\varepsilon N^2}(C_\infty) \in B\} \cap E_2^N \cap E_3^N] - \mathbf{P}[K_{2s/(g+2/\gamma)}(C_\infty) \in B]\Big| \leq 3\delta/4. \tag{7.40}$$

The last equation together with (7.36) gives

$$\lim_{N\to\infty} \Big|\mathbf{P}[C^N_{t(N)} \in B] - \mathbf{P}[K_{2s/(g+2/\gamma)}(C_\infty) \in B]\Big| \leq |\tilde{r}^N(B)|, \tag{7.41}$$

where

$$\sup_{N \geq N_\varepsilon} \sup_B |\tilde{r}^N(B)| \leq 6\delta/4. \tag{7.42}$$

Letting $\delta \to 0$ gives the intermediate behavior in (7.5).

The behaviors for $t(N) = o((2N+1)^d)$ and $t(N) \gg (2N+1)^d$ now follow from the properties of Kingman's coalescent and the sandwich principle. □

Step 4 (General case) From earlier arguments on the coalescent and its local finiteness we know already that at positive time and hence in particular by time $\varepsilon(2N+1)^d$, with $\varepsilon > 0$ arbitrary, we have only finitely many partition elements left. We can analyze the behavior of this configuration using the previous step, provided we can prove the following two facts.

(i) The number of the partition elements at time $\varepsilon(2N+1)^d$ is stochastically bounded in N, and
(ii) the positions of these partition elements at time $\varepsilon(2N+1)^d$ are at mutual spatial distance of order N.

Then the assertion of part(a) of Proposition 7.2 follows from Lemma 7.3 by first establishing for each $\varepsilon > 0$ the desired convergence of the rescaled process for $t \geq \varepsilon$, and then taking $\varepsilon \to 0$ relying on existence and uniqueness of Kingman's coalescent. Part (b) is established separately in Step 5 below. Now we conclude this step by verifying (i) and (ii).

Fact (i) To establish the first fact, we need some control over the mean number of partition elements left in the coalescent C_t^N given a configuration of partition elements, $\sum_{\pi\in C_0^N}\delta_{L_0(\pi)}$. The following lemma is similar to Theorem 1 in [4] and the proposition in Section 5 of [5]. We are going to prove it in Step 5.

Lemma 7.4. *(Number of partition elements in the coalescent) There is a finite constant c_d such that uniformly in sequences $(|C_0^N|)_{N\in\mathbb{N}}$ such that $|C_0^N|\ge(2N+1)^d$, and uniformly in $N\in\mathbb{N}$,*

$$\mathbf{E}\left[|C_t^N|\right]\le c_d\max\left\{1,\frac{|C_0^N|}{t}\right\}. \tag{7.43}$$

By Proposition 3.4, for each given $\varepsilon>0$, $C^N_{\varepsilon(2N+1)^d}$ is locally finite. Now Lemma 7.4 implies in particular, that for each given $\varepsilon>0$, $|C^N_{\varepsilon(2N+1)^d}|$ is moreover uniformly stochastically bounded in N since we can for every $\delta>0$ control $|C^N_\delta|/(2N+1)^d$ uniformly in N, using the spatial homogeneity, and local finiteness (recall the proof of Proposition 3.4).

Fact (ii) The second fact follows immediately observing that by times of order δN^2, where δ is large, (i.e., $t(N)=o((2N+1)^d)$) two independent random walks are with probability tending to 1 in distance of order N, as $N\to\infty$.

Step 5 (Proof of Lemma 7.4.) W.l.o.g. we may assume that $|C_0^N|<\infty$. Otherwise the estimate (7.43) would hold trivially.

The reasoning relies on that for the coalescing random walk estimates given in [5]. Once more it differs due to the facts that our partition elements coalesce with a rate γ delay, and that we may have more than 1 particle per site.

Observe that

$$|C_t^N|=|C_0^N|-Z_t^N, \tag{7.44}$$

where Z_t^N denotes the number of jumps of $(C_s^N)_{s\in[0,t]}$. Since for a fixed $\pi_0\in C_0^N$,

$$Z_t^N\ge\sum_{\pi\in C_0^N\setminus\{\pi_0\}}1_{\{\pi\approx_{C_t^N}\pi_0\}}, \tag{7.45}$$

its expectation is greater than or equal to

$$\mathbf{E}[Z_t^N]\ge(|C_0^N|-1)\min_{\pi\neq\pi_0\in C_0^N}\mathbf{P}[\pi\approx_{C_t^N}\pi_0], \tag{7.46}$$

In particular, if we start the coalescent such that all positions of partition elements are contained in a subset $A\subseteq G_N$, then the right hand side of (7.46) can be bounded below as follows.

Let X^N be a rate 2 continuous time random walk on G_N, σ^N the hitting time of the origin, furthermore Y_1, Y_2 independent rate γ and rate 2 exponential random variables and

$$h_t(A):=\min_{x\in A}\mathbf{P}^x[\sigma^N\le t]. \tag{7.47}$$

Then

$$\mathbf{P}[\pi\approx_{C_t^N}\pi_0]\ge\mathbf{P}[Y_1\le t/2;Y_1\le Y_2]\,h_{t/2}\left(\chi(C_0^N,L_0^N)\right), \tag{7.48}$$

where

$$\chi(C_0^N,L_0^N):=\{x-y:\exists\,\pi_1\neq\pi_2\in C_0^N:L_0(\pi_1)=x,L_0(\pi_2)=y\}. \tag{7.49}$$

Hence since $\mathbf{P}[Y_1\le t/2;Y_1\le Y_2]=\int_0^{t/2}\mathrm{d}s\,\gamma\,e^{-\gamma s}\,e^{-2s}$,

$$\begin{aligned}&\mathbf{E}[|C_t^N|]\\&\le|C_0^N|-(|C_0^N|-1)\frac{\gamma}{2+\gamma}\left(1-\exp\left[-\frac{2+\gamma}{2}t\right]\right)h_{t/2}\left(\chi(C_0^N,L_0^N)\right).\end{aligned} \tag{7.50}$$

In the next step we will use the latter estimate to control the decay rate of $\frac{\mathbf{E}[|C_t^N|]}{|C_0^N|}\,t$. The following lemma has again an analogue in [5].

Lemma 7.5. *(Upper bound for the decay rate of partition elements) If*

$$f(t) := \frac{\mathbf{E}[|C_t^N|]}{|C_0^N|}\,t, \tag{7.51}$$

then there exists a finite constant $M_d \geq 1$ such that uniformly in $N \in \mathbb{N}$ and uniformly in $t \in [0, |C_0^N|]$,

$$f(t) \leq M_d, \tag{7.52}$$

or

$$f(2t) \leq \max\{M_d, f(t)\}. \tag{7.53}$$

Proof of Lemma 7.4 (Continuation) Observe that Conditions (7.52) and (7.53) together imply immediately that for each $t \in [0, |C_0^N|]$, $f(t) \leq M_d$. This proves then Lemma 7.4 with $c_d := M_d$. Indeed, define $n(t) := \inf\{m \in \mathbb{N} \cup \{0\} : f(t/2^m) \leq M_d\}$, $t \geq 0$, and notice that $n(t) < \infty$ for all $t \geq 0$ since $f(t) \leq t$ for all t. Hence, for all $t \in [0, |C_0^N|]$ we have

$$\begin{aligned} f(t) \leq \max\{M_d, \max\{M_d, f(t/2)\}\} &= \max\{M_d, f(t/2)\} \\ &\leq ... \\ &\leq \max\{M_d, f(t/2^{n(t)})\} = M_d. \quad \square \end{aligned} \tag{7.54}$$

Proof of Lemma 7.5 Once more if $C_0^N = \infty$ then $f(t) \equiv 0$ and the conditions hold trivially. Assume therefore that $C_0^N < \infty$.

Fix $t_0 \in [0, |C_0^N|]$. We shall show the assertion for a constant $M_d \geq 16^d$ to be constructed below. For that we may assume w.l.o.g. that

$$\mathbf{E}[|C_{t_0}^N|] \geq 8^d, \tag{7.55}$$

and

$$\mathbf{E}[|C_{2t_0}^N|] \geq 2 \cdot 4^{-d}\mathbf{E}[|C_{t_0}^N|]. \tag{7.56}$$

Indeed, if one of the two conditions would fail, then the assertion of Lemma 7.5 holds for t_0 with any $M_d \geq 8^d$.

Next we are going to cover the torus G_N by n_{t_0} disjoint cubes, $\{B_{t_0}^i;\, i = 1, ..., n_{t_0}\}$, where side length not larger than some $a_{t_0} \geq 8$ with

$$a_{t_0} := \lfloor 8\big(\frac{|C_0^N|}{\mathbf{E}[|C_{t_0}^N|]}\big)^{\frac{1}{d}}\rfloor \wedge (2N+1). \tag{7.57}$$

Let now $t_0 \leq r \leq r+s \leq 2t_0$. If we ignore the coalescence of partition elements starting in different boxes, $B_{t_0}^i$, then by the Markov property:

$$\mathbf{E}\left[|C_{r+s}^N|\right] \leq \mathbf{E}\left[\sum_{i=1}^{n_{t_0}} \mathbf{E}^{\phi_{B_t^i}(C_r^N, L_r^N), r}\left[|C_{r+s}^N|\right]\right], \tag{7.58}$$

where $\phi_{B^i_{t_0}}(C^N_r, L^N_r)$ is the restriction of C^N_r to those partition elements π with $L^N_r(\pi) \in B^i_{t_0}$. Then by (7.50),

$$\begin{aligned}(7.59)\quad &\mathbf{E}[|C^N_{r+s}|] \\ &\le \mathbf{E}[|C^N_r|] - (\mathbf{E}[|C^N_r|] - n_{t_0})\,\frac{\gamma}{2+\gamma}\left(1-\exp[-\frac{2+\gamma}{2}s]\right) h_{s/2}\left(B^1_{t_0}\right) \\ &\le \mathbf{E}[|C^N_r|]\left(1-\frac{\gamma}{2(2+\gamma)}(1-\exp[-\frac{2+\gamma}{2}s])h_{s/2}\left(B^1_{t_0}\right)\right) \\ &\le \mathbf{E}[|C^N_r|]\,\exp\left[-\frac{\gamma}{2(2+\gamma)}(1-\exp[-\frac{2+\gamma}{2}s])\,h_{s/2}\left(B^1_{t_0}\right)\right],\end{aligned}$$

where we have used that by (7.55) and (7.56), $\mathbf{E}[|C^N_r|] \ge \mathbf{E}[|C^N_{2t_0}|] \ge 2\mathbf{E}|C^N_{t_0}|/4^d$, and therefore $(\mathbf{E}[|C^N_r|]-n_{t_0}) \ge \frac{1}{2}\mathbf{E}[|C^N_r|]$, and that $1-e^{-x} \le x$ for all $x \ge 0$. Furthermore observe that $h_{s/2}(B^1_{t_0})$ depends on $B^1_{t_0}$ only via the side length a_{t_0}, and $B^1_{t_0}$ represents therefore a box of side length a_{t_0}.

Iterating the latter one obtains

$$\begin{aligned}(7.60)\quad &\mathbf{E}[|C^N_{2t_0}|] \\ &\le \mathbf{E}[|C^N_{t_0}|]\,\exp\left[-\left\lfloor\frac{t_0}{s}\right\rfloor\frac{\gamma}{2(2+\gamma)}(1-\exp[-\frac{2+\gamma}{2}s])\,h_{s/2}\left(B^1_{t_0}\right)\right].\end{aligned}$$

To employ (7.60) effectively, we still may choose $s := s_{t_0}$. By the rescaling analysis for random walks with finite second moments there exists a constant $\alpha_d \in (0,\infty)$ such that for a box $B := [-b,b]^d \cap \mathbb{Z}^d$ of side length $b \ge 8$,

$$(7.61)\qquad h_{b^2}(B) \ge \alpha_d b^{2-d}.$$

To be in a position to apply the latter on (7.60), we choose

$$(7.62)\qquad s_{t_0} := 2\,a^2_{t_0},$$

and consider w.l.o.g. the case $s_{t_0} \le \frac{t_0}{2}$. Otherwise we are done since by (7.57),

$$(7.63)\qquad f(t_0) \le 8^d\,a^{-d}_{t_0}\,t_0 \le 8^d\cdot 2^{\frac{d}{2}}\cdot s^{-\frac{d}{2}}_{t_0}\cdot t_0 \le 8^d\cdot 2^{\frac{d}{2}}\cdot 2^{\frac{d}{2}}\cdot t_0^{-\frac{d}{2}}\cdot t \le 16^d.$$

Then by (7.57) and (7.62),

$$(7.64)\qquad \lfloor t_0/s_{t_0}\rfloor \ge t_0/2s_{t_0} \ge \frac{\mathbf{E}[|C^N_{t_0}|]^{2/d}}{256\,|C^N_0|^{2/d}}\,t_0.$$

Therefore by (7.61),

$$(7.65)\qquad h_{s_{t_0}/2}(B^1_{t_0}) \ge \alpha_d\,a^{2-d}_{t_0} \ge \alpha_d\,8^{2-d}\left(\frac{\mathbf{E}[|C^N_{t_0}|]}{|C^N_0|}\right)^{1-\frac{2}{d}},$$

and since $s_{t_0}/2 = a^2_{t_0} \ge 64$ (for N sufficiently large),

$$(7.66)\qquad \frac{\gamma}{2+\gamma}(1-\exp[-\frac{2+\gamma}{2}s_{t_0}]) \ge \frac{\gamma}{2+\gamma}(1-\exp[-(2+\gamma)\,64]) =: \beta.$$

This finally yields

$$(7.67)\qquad f(2t_0) \le f(t_0)\exp\left[\log 2 - \alpha_d\,\beta\,8^{-(d+1)}\,f(t_0)\right],$$

and so (7.52) or (7.53) holds with $M_d := \max\{16^d, \frac{8^{(d+1)}\,\log 2}{\alpha_d\,\beta}\}$. Together with (7.63) this gives the claim. □

7.3. **Preliminaries: Asymptotic analysis of the historical coalescent.** In this subsection we lift the results of the previous subsection to the level of the historical coalescent which actually gives stronger results than we have claimed in our theorems. Recall the historical coalescent (C_t^*, L_t^*) and its topology from (3.31) and (3.32) and define the *historical Kingman's coalescent*, K_t^*, analogously.

Denote by C_∞^* the limit of C_t^* as $t \to \infty$ in the following topology on paths. On paths with values in partitions we use the Skorokhod topology induced by the metric on the partitions defined in (3.28), but we consider here only the paths on $(0, \infty)$ *excluding* the point 0.

We now have to scale paths of the partition-valued process C_t^N in time between times 0 and times of order $(2N+1)^d$ or in the language of the historical coalescent we have to rescale C_t^*. Therefore we define for a path $y \in \mathcal{D}(\mathbb{R}, E)$, where E is a Polish space, the scaling operation

$$s_\beta : \ y(\cdot) \mapsto y(\beta \cdot). \tag{7.68}$$

Proposition 7.6. *(Historical coalescent asymptotics) Fix a sequence $t(N) \uparrow \infty$ as $N \to \infty$.*

(a) *Let K^* be a version of the historical Kingman's coalescent started in $\widetilde{C}_\infty^*$ which is the constant path equal to C_∞. Let $\beta_N(t) = t(2N+1)^d$. Then*

$$\mathcal{L}^{(C_0, L_0)|_{G_N}} \left[s_{\beta_N(1)} C^{*,N}_{\beta_N(t)} \right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{(C_0,L_0)} \left[K^*_{\kappa t} (\widetilde{C}^*_\infty) \right]. \tag{7.69}$$

If $t(N) \to \infty$ but $t(N) = o((2N+1)^d)$, then

$$\mathcal{L}^{(C_0, L_0)|_{G_N}} \left[C^{*,N}_{t(N)} \right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{(C_0,L_0)} \left[C^*_\infty \right]. \tag{7.70}$$

(b) *For the path process we obtain no matter on which scale $t(N)$ tends to ∞ the following:*

$$\mathcal{L}^\Psi [L^{*,N}_{t(N)}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^\Psi [L^*_\infty]. \tag{7.71}$$

Proof We start with the first assertion of part (a). Observe that the partition valued paths of the coalescents are processes whose paths have a monotonicity property. Namely, as time proceeds we can only jump to coarser partitions. This means that the sequence of laws of the paths is tight if we can establish that in finite time intervals $[s(2N+1)^d, t(2N+1)^d]$ with $t > s > 0$ only finitely many jumps can occur. For this it suffices to know that

$$\limsup_{N\to\infty} \mathbf{P}[|C^N_{\varepsilon(2N+1)^d}| \geq n] \underset{n\to\infty}{\longrightarrow} 0. \tag{7.72}$$

However we have proved that the partition valued part of the coalescent converges to $K_s(C_\infty)$ and we know that $K_\varepsilon(C_\infty)$ has only finitely many partition elements. Therefore the probability to have more than K jumps after time $\varepsilon(2N+1)^d$ goes to zero as $K \to \infty$ uniformly in N. This establishes tightness of the process $(C^{*,N}_{t(2N+1)^d})_{t>0}$ (note 0 is excluded).

We are now left with proving that the finite dimensional distributions of the processes $(C^{*,N}_{t(2N+1)^d})_{t\geq 0}$ converge to those of $K_t(C^*_\infty)$. In Proposition 7.2 we proved the assertion for the one-dimensional marginal distributions. But the proof of this statement showed more. First of all the locations of the partition elements at times $t(2N+1)^d$ are uniformly distributed on the torus for any such configuration we get that if we follow the partition for an additional time $s(2N+1)^d$ then the law conditioned on the configuration at $t(2N+1)^d$ to be C (some finite partition) converges to $K_s(C)$. This gives (7.69).

For the second assertion of part (a) we need that the law of C^N_T converges to C_T as $N \to \infty$ by the coupling we have constructed. This gives immediately that $(C_t^{*,N})_{t\leq T}$ converges in law to $(C_t^*)_{t\leq T}$ for $N \to \infty$. Furthermore we know that $C_T \to C_\infty$ a.s. as $T \to \infty$. Hence we are done if we can show that the probability that a coalescence occurs between time T and time $t(N)$ is bounded by ε_T with $\varepsilon_T \to 0$ as $T \to \infty$. This we have done in the previous Subsection 7.3.

The weak convergence in (7.71) is trivial due to our coupling construction in which the path for the process on $[-N, N] \cap \mathbb{Z}^d$ is obtained from the path of the infinite system by taking the position modulo $(2N+1)$. □

7.4. **Asymptotic analysis of the look-down process (Proof of Theorem 6).** Our goal is to prove here Theorem 6. This means that we need to describe asymptotically the time-space picture of the genealogy of the time t population. Namely, for the time t population consider for every $s \le t$ the ancestors of the present population and their decomposition in related individuals and their paths followed in geographic space. We assert that the resulting time-space picture can be rescaled as a function of the final time horizon t such that it converges to the historical coalescent. We make this precise in the proposition below.

Consider the look-down process started in the configuration given by (2.28). We next focus on the genealogy, $(\Gamma^{t(N),N}_{t(N)}, \ell^{t(N),N}_{t(N)})$, of the look-down process (recall (7.1)). The object Γ^t_t is the partition at time t induced by merging all related individuals of the time t population in one partition element. The object Γ^t_s is the partition arising by merging all individuals which have a common ancestor at the time s.

Recall from (3.31) and (3.32) the historical coalescent, $(C^*_t, L^*_t)_{t\ge 0}$, and its topology from prior of (3.33). Define similarly,

$$\Gamma^{*,t}_t := \left(\Gamma^t_{(s\vee 0)\wedge t}\right)_{s\in\mathbb{R}}, \tag{7.73}$$

and

$$\ell^{*,t}_t := \left(\ell^t_{(s\vee 0)\wedge t}\left(\Gamma^t_{(s\vee 0)\wedge t}(\{\iota\});\ \iota\in\mathcal{I}\right)\right)_{s\in\mathbb{R}}, \tag{7.74}$$

where $\Gamma^t_s(\{\iota\})$ denotes the family at time $t-s$ consisting of ι and its relatives. Moreover, define $\Gamma^{*,t,N}_t$ and $\ell^{*,t,N}_t$ analogously on the torus.

With the duality formulated in Proposition 3.5 we can conclude immediately from Proposition 7.6 that one has (recall τ^*_t is the time shift of paths by t and s_β the rescaling given in (7.68)):

Proposition 7.7. *(Asymptotics of the genealogy in $d\ge 3$.)*
Assume that the labels are chosen as in (2.28) and that $d \ge 3$.

(a) With the same notation as used in Proposition 7.6 we have

$$\mathcal{L}\left[(s_{\beta_N(1)}\tau^*_{\beta_N(t)}\Gamma^{*,t(N),N}_{\beta_N(t)})\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[K^*_{\kappa t}(\widetilde{C}^*_\infty)\right]. \tag{7.75}$$

For $t(N)\to\infty$ but $t(N) = o((2N+1)^d)$,

$$\mathcal{L}\left[\tau^*_{t(N)}\Gamma^{*,t(N),N}_{t(N)}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[C^*_\infty\right]. \tag{7.76}$$

(b) For the second component we obtain for each $t(N)\uparrow\infty$ the following:

$$\mathcal{L}\left[\ell^{*,t(N),N}_{t(N)}\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}\left[L^*_\infty\right]. \tag{7.77}$$

Based on the above Proposition we prove Theorem 6, i.e. the statement on the look-down process. We then conclude Corollary 3.13 by a projection on the time horizon at which the first interaction takes place after the individuals start feeling the boundary.

Proof of Theorem 6 Once more we restrict our attention to the intermediate scale, $\beta_N(t) := t\,(2N+1)^d$. The other cases will then follow from the properties of the historical Kingman's coalescent and the sandwich principle.

Observe that $X^{*,N}_{\beta_N(t)}$ can be reconstructed from the genealogy $(\Gamma^{*,\beta(N),N}_{\beta(N)}, \ell^{*,\beta(N),N}_{\beta(N)})$ and the assignment of the types at time 0. In fact we don't need the full "historical genealogy", we only need the value of this "path" at time $\beta_N(t)$ and then we have to add the types of the family clusters as follows. The law of $\tau^*_{\beta_N(t)}X^{*,N}_{\beta_N(t)}$ is given by assigning each family cluster independently of the others the type 1 or 0 with probability θ or $1-\theta$, respectively. Hence (recall Δ_θ from (3.35))

$$\mathcal{L}^{\Psi(\theta)}[\tau^*_{\beta_N(t)}X^{*,N}_{\beta_N(t)}] = \mathcal{L}^{\Psi(\theta)}[\Delta_\theta\left((\Gamma^{*,\beta_N(t),N}_{\beta_N(t)}, \ell^{*,\beta_N(t),N}_{\beta_N(t)})_{\beta_N(t)}\right)]. \tag{7.78}$$

Then by Proposition 7.7 the right hand side of (7.78) converges as $N \to \infty$ to

$$\mathcal{L}^{\Psi(\theta)}[\Delta_\theta(K_{\kappa t}(C_\infty), L^*_\infty)] = \int \bar{Q}^*_{\kappa t}[\mathrm{d}(C,\mathcal{P})]\, \mathcal{L}^{\Psi(\theta)}[\Delta_\theta(K_{\kappa t}(C_\infty), L^*_\infty) | C_\infty = C, K_{\kappa t}(C_\infty) = \mathcal{P}] \tag{7.79}$$

and by Definition 3.12 this equals

$$\int \bar{Q}^*_{\kappa t}[\mathrm{d}(C,\mathcal{P})]\, \mathcal{H}^*_{1,\theta}(C,\mathcal{P}). \tag{7.80}$$

This is the claim of the Theorem in the intermediate time scale. □

Proof of Corollary 3.13 By Theorem 3 the genealogy of the enriched historical IMM, $\xi^{*,\rho,N}$, which we denoted by $(\widetilde{\Gamma}^{*,t,\rho,N}_s, \widetilde{\ell}^{*,t,\rho,N}_s)_{s\le t}$ (compare also with (3.4)), can be reconstructed from the genealogy of the look-down process together with an additional construction of random permutations of the labels. This allows analog to the situation for the proof of Theorem 6 above to represent the law of the configuration in terms of Δ_θ and the genealogy. This genealogy is then in law equal to the one which is defined on the probability space of the look-down process enriched with the coin flipping variables for generating the symmetrization. These variables are independent of everything else. Hence again

$$\begin{aligned}\mathcal{L}^{\Psi(\rho,\theta)}[\tau^*_{\beta_N(t)}\xi^{*,\rho,N}_{\beta_N(t)}] &= \mathcal{L}^{\Psi(\theta)}[\Delta_\theta((\widetilde{\Gamma}^{*,\beta_N(t),\rho,N}_{\beta_N(t)}, \widetilde{\ell}^{*,\beta_N(t),\rho,N}_{\beta_N(t)})_{\beta_N(t)})] \\ &= \mathcal{L}^{\Psi(\theta)}[\Delta_\theta(\phi_{\mathcal{I}^\rho}(\Gamma^{*,\beta_N(t),N}_{\beta_N(t)}, \ell^{*,\beta_N(t),N}_{\beta_N(t)})_{\beta_N(t)})] \\ &\underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{\Psi(\theta)}[\Delta_\theta\phi_{\mathcal{I}^\rho}(K_{\kappa t}(C_\infty), L^*_\infty)].\end{aligned} \tag{7.81}$$

Next we turn to $\eta^{*,\rho,N}$. Recall $\widetilde{\Delta}_{\rho,\theta}$ from (3.40). Then by (1.40), the law of $\eta^{*,\rho,N}_{\beta_N(t)}$ satisfies

$$\mathcal{L}^{\Psi(\rho,\theta)}[\tau^*_{\beta_N(t)}\eta^{*,N}_{\beta_N(t)}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{\Psi(\theta)}[\widetilde{\Delta}_{\rho,\theta}(K_{\kappa t}(C_\infty), L^*_\infty)]. \tag{7.82}$$

Applying the family decomposition on the random variable on the right hand side, we get

$$\begin{aligned}\widetilde{\Delta}_{\rho,\theta}(K_{\kappa t}(C_\infty), L^*_\infty) &= \sum_{\iota\in\mathcal{I}^\rho} \delta_{(L^{*,\{\iota\}}_\infty, \Delta_\theta(K^{\{\iota\}}_{\kappa t}(C_\infty)))} \\ &= \sum_{\pi\in K_{\kappa t}(C_\infty)} \sum_{\iota\in\pi\cap\mathcal{I}^\rho} \delta_{(L^{*,\{\iota\}}_\infty, \Delta_\theta(K^{\{\iota\}}_{\kappa t}(C_\infty)))} \\ &= \sum_{\pi\in\phi_{\mathcal{I}^\rho}K_{\kappa t}(C_\infty)} \widetilde{\Delta}_{\rho,\theta}\phi_\pi(K_{\kappa t}(C_\infty), L^*_\infty).\end{aligned} \tag{7.83}$$

Hence, by independence in the assignment of types to the different families, and the fact that $\phi_\pi K_{\kappa t}(C_\infty)$ is maximal, i.e., $\phi_\pi K_{\kappa t}(C_\infty) = \{\pi\}$, if $\pi \in K_{\kappa t}(C_\infty)$,

$$\begin{aligned}&\mathcal{L}^{\Psi(\rho,\theta)}[\tau^*_{\beta_N(t)}\eta^{*,N}_{\beta_N(t)}] \\ &\underset{N\to\infty}{\Longrightarrow} \int \bar{Q}^*_{\kappa t}[d(C,\mathcal{P})] \left\{ \mathop{\ast}_{\pi\in\phi_{\mathcal{I}^\rho}\mathcal{P}} \Big(\theta\widetilde{\mathcal{H}}^*_{\pi,1}(C) + (1-\theta)\widetilde{\mathcal{H}}^*_{\pi,0}(C)\Big)\right\}. \quad \square\end{aligned} \tag{7.84}$$

Proof of Corollary 3.15. At this stage we will give the proof for IMM only. The transfer from IMM to IFWD is given in Section 8. Before we prove the two assertions of the corollary, we verify a key relation (7.86) below.

First observe that given a realization of $\{\Delta_\theta(\pi); \pi \subseteq \mathcal{I}\}$ (recall (3.34)), $X^{*,N}_{\beta_N(t)}$ can be constructed from the genealogy, $(\Gamma^{\beta_N(t),N}_{\beta_N(t)}, \ell^{\beta_N(t),N}_{\beta_N(t)})$, by assigning the type $\Delta_\theta(\varpi)$ to the final family cluster $\varpi \in \Gamma^{\beta_N(t),N}_{\beta_N(t)}$. Hence, by Proposition 7.7,

$$\mathcal{L}^{\Psi(\theta)}[(X^{*,N}_{\beta_N(t)})_{\beta_N(t)}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{\Psi}[\Delta_\theta(K_{\kappa t}(C_\infty), (L^*_\infty)_0)]. \tag{7.85}$$

Recall $\widetilde{\Delta}_{\rho,\theta}$ from (3.40). Then by Proposition 3.5 the same holds for IMM, i.e.,

$$\mathcal{L}^{\Psi(\rho,\theta)}[\eta^N_{\beta_N(t)}] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{\Psi}[\widetilde{\Delta}_{\rho,\theta}(K_{\kappa t}(C_\infty), L_0)]. \tag{7.86}$$

Now we are ready to prove the two parts of the Corollary concerning IMM.

(a) Recall that $\phi_{\mathcal{I}^\rho} C_\infty$ consists of countably many partition elements, a.s. Hence by Proposition 3.3(c) in [19] for each $t > 0$, $K_{\kappa t}(C_\infty)$ is exchangeable, that is, its distribution is invariant under renumbering finitely many of the labels. So the same holds for the collection, $(\Delta_\theta(K^{\{\iota\}}_{\kappa t}(C_\infty)); \iota \in \mathcal{I})$, of $\{0,1\}$-random variables (recall (3.34)). Recall from (3.68) to (3.70) that no matter in which way we number the patches of $C_\infty := \{\pi_1, \pi_2, ...\}$, the de Finetti measure on $\{0,1\}$ determined by,

$$z_t := \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Delta_\theta(\phi_{\mathcal{I}^\rho}(K^{\pi_k}_{\kappa t}(C_\infty))), \qquad a.s., \tag{7.87}$$

exists, and its law equals the law of a Fisher-Wright diffusion with resampling rate κ, and started in θ. Notice in particular that this law does not depend on ρ.

Given the de Finetti measure, z_t, the collection,

$$(\Delta_\theta(K^{\pi}_{\kappa t}(C_\infty)); \pi \in \phi_{\mathcal{I}^\rho}(C_\infty)), \tag{7.88}$$

consists of independent $\{0,1\}$-elements with probability for a one equal to z_t, where as before for each $\pi \in \phi_{\mathcal{I}^\rho}(C_\infty)$, $K^{\pi}_{\kappa t}(C_\infty)$ denotes the patch in $\phi_{\mathcal{I}^\rho}(K_{\kappa t}(C_\infty)$ containing π. We know that

$$\mathcal{L}^{\Psi}[\widetilde{\Delta}_{\rho,\theta}(K_{\kappa t}(C_\infty), L_0)\big|z_t] = \mathcal{L}^{\Psi}[\widetilde{\Delta}_{\rho,z_t}(C_\infty, L_0)] = \widetilde{\mathcal{H}}_{\rho,z_t}. \tag{7.89}$$

We therefore obtain from (7.86) that for every $t > 0$,

$$\mathcal{L}^{\Psi(\rho,\theta)}[\eta^N_{\beta_N(t)}] \underset{N\to\infty}{\Longrightarrow} \int_{[0,1]} Q_t[\theta, \mathrm{d}\theta'] \widetilde{\mathcal{H}}_{\rho,\theta'}. \tag{7.90}$$

(b) We show first the convergence of the finite-dimensional marginals of the density (of type one) process and then the tightness based on martingale arguments.

In order to show the *f.d.d. convergence* we begin with the one-dimensional marginals. By (3.20), the equilibrium measures $(\widetilde{\mathcal{H}}_{\theta,\rho})_{\theta\in[0,1]}$ have again the initial density of mass ρ and relative frequency of type 1 given by θ, and by the representation via the strong duality which is stated in Proposition 3.5 they are spatially ergodic and in fact mixing. In Cox, Greven, Shiga [10] it was proved that in this context the densities (of ones) converge to the density of the weak limit. Hence by (7.89), for each $t \geq 0$,

$$\mathcal{L}^{\widetilde{\Psi}(\theta,\rho)}\big[D^{\rho,N}(\eta^N_{\beta_N(t)})\big] \underset{N\to\infty}{\Longrightarrow} \mathcal{L}^{\theta}[z_t]. \tag{7.91}$$

This establishes the convergence claimed in (3.74) for one-dimensional marginals of the process. More generally, calculating the space-time moments and showing their convergence to the ones of the standard Fisher-Wright diffusion, we get convergence of the finite dimensional marginals (abbreviated as *f.d.d.*-convergence), i.e.,

$$\mathcal{L}^{\widetilde{\Psi}(\theta,\rho)}\Big[\Big(D^{\rho,N}(\eta^N_{\beta_N(t)})\Big)_{t\geq 0}\Big] \underset{N\to\infty}{\overset{f.d.d.}{\Longrightarrow}} \mathcal{L}^{\theta}[(z_t)_{t\geq 0}]. \tag{7.92}$$

(Such a moment calculation has been carried out e.g. in [11], and will therefore be omitted here.)

In order to show tightness we rely on Proposition 1.2 in Aldous (1989) [1] which is really tailored for our situation since the densities are bounded martingales. In this case weak convergence in the path space follows from *f.d.d.*-convergence and the fact that the limit is continuous in t. The latter is clear since $(z_t)_{t\geq 0}$ is the Fisher-Wright diffusion. $\square$

8. Transfer to IFWD (Proofs of Theorems 5 and 7)

In this section we prove the results concerning the application of our techniques to the historical IFWD. The proofs will be based on what we have already shown about the IMM, knowing that IFWD not only occurs as the diffusion limit of IMM but both can be constructed via a common particle representation.

The key point in transferring our results from IMM to IFWD is to show that we can interchange the limit $\rho \to \infty$ (giving rise to the diffusion limit) with the limit $t \to \infty$ for the infinite system and with the $N \to \infty$ limits for the finite systems. Here the key structure is given by our representation theorems and structural properties of the coalescent.

The basic reason why the interchangeability of limits should hold is as follows. By the strong duality we can pass to the dual process which is a spatially marked coalescent on $\mathbb{Z}^d$ (respectively on the tori of $\mathbb{Z}^d$). But again, as pointed out before in Subsection 7.1, in $\mathbb{Z}^d$ we can define the coalescents on the torus as a functional of the coalescent on $\mathbb{Z}^d$. Furthermore we define this coalescent starting with countable many individuals at each site, and then we can embed all the $\mathcal{I}^\rho$-coalescents in this object. In other words, we have a coupling of all the dual processes we need for all t, N and ρ on the same probability space. However, to verify the needed uniformity we will need to go further and use another approximation by an object we call the *pruned coalescent* that has finitely many partition elements for positive times.

We now proceed in four steps, we construct the pruned coalescent, represent frequency functionals in terms of the coalescent, and finally derive the uniformity for the two situations.

Step 1 (Construction of a pruned coalescent) Here we show that we can construct representations of the quantities in question in terms of coalescent processes for which we then can utilize certain coupling constructions to get the needed uniformity in the convergence relations to justify the exchange in the order of limits.

Let us first observe that the limit in (2.27) for the basic particle representation exists a.s. due to the exchangeability of the individuals starting at one site in the coalescents and the independence of the variables assigning the types to the patches.

The basic convergence result we need to transfer is (3.26). Recall that for the historical Moran model the convergence of the laws in the topology we consider means that we consider the subpopulation of individuals which are at time t in some finite spatial window of observation. This means we need to apply the strong duality relation to coalescence processes starting from a state with all marks in a fixed finite subset of G. We therefore investigate coalescents of this type.

Fix therefore a set $A \subseteq G$ with $|A| < \infty$ and start a coalescent with countably many individuals in each point in A and denote it by $(C_t^A, L_t^A)_{t \geq 0}$. We embed in this object the coalescents corresponding to labels which do not exceed ρ. In the case where we study finite systems proceed as follows: For all $G_N \supseteq A$ we can then consider the corresponding coalescent denoted by $(C_t^{A,N}, L_t^{A,N})_{t \geq 0}$ on G_N as well as the one on G. Furthermore we couple the coalescents on the size N tori and on $\mathbb{Z}^d$ on one probability space as explained in Subsection 7.1.

The next point is to reduce this coalescent introduced above on G, which has countably many partition elements, to a coalescent with *finitely* many partition elements. The problem is that even though a non-spatially structured coalescent collapses in positive time to finitely many partition elements, and our spatially structured coalescent is known to collapse in positive time to locally finitely many partition elements, the number of different locations of the partition elements belonging to the spatially structured coalescent started with the individuals living initially in a fixed finite window may stay infinite for all times. However, the idea is that still most, in the sense of asymptotic frequency, individuals coalesce to finitely many partition elements such that the functionals giving the diffusion limit are not changed much if we restrict to those. In order to make this idea precise we proceed as follows.

Fix a $\delta > 0$, and then choose a suitable $t_0 > 0$ and define coalescents starting with countably many individuals initially on each site of A, while removing from it all individuals which migrate before time t_0. The resulting object is called *pruned* coalescent denoted $(\widehat{C}_t^{A,\delta}, \widehat{L}_t^{A,\delta})_{t \geq 0}$ (and $(\widehat{C}_t^{A,N,\delta}, \widehat{L}_t^{A,N,\delta})_{t \geq 0}$ on the torus, respectively). Since a non spatially structured coalescent collapses to a finite partition in any positive time, and since A is finite, the pruned coalescent is finite for $t > t_0$.

By the above construction we can achieve that the probability of every single particle to be removed is less than δ at each site by choosing t_0 small enough. This means that the density of the ones removed among all is bounded by δ.

This means in particular that we get a uniform approximation of averages concerning the type of individuals as in (2.27) by making the parameter δ small. Therefore we can now continue working with these pruned coalescents and consider them only for times $t \geq t_0$.

We write $(\widehat{C}_t^{A,\rho,\delta}, \widehat{L}_t^{A,\rho,\delta})$ for the pruned $\mathcal{I}$-coalescent which is reduced to labels less or equal to ρ and which is a $\mathcal{I}^\rho$-coalescent, and we write $(\widehat{C}_t^{A,\rho,N,\delta}, \widehat{L}_t^{A,\rho,N,\delta})$ for the corresponding objects on the torus.

We define now the finite stopping times

$$T^{\rho,\delta} := \inf\{t > t_0;\, |\widehat{C}_t^{A,\rho,\delta}| = |\widehat{C}_t^{A,\delta}|\}, \tag{8.1}$$

and similarly,

$$T^{\rho,N,\delta} := \inf\{t > t_0;\, |\widehat{C}_t^{A,\rho,N,\delta}| = |\widehat{C}_t^{A,N,\delta}|\}. \tag{8.2}$$

From this time on all transitions occur jointly in the pruned $\mathcal{I}$-coalescent and in the pruned $\mathcal{I}^\rho$-coalescent. In other words we have some sort of coupling, even though of course the partition elements themselves do not agree.

Now define the events that we see no new partition elements if we increase ρ. That is, for $t \geq t_0 = t_0(\delta)$, and $\rho \in [0,\infty)$,

$$E(t,\rho) := \{\omega : T^{\widetilde{\rho},\delta}(\omega) \leq t, \text{ for all } \widetilde{\rho} \geq \rho\}. \tag{8.3}$$

Similarly, let for $t \geq t_0 = t_0(\delta)$, $\rho \in [0,\infty)$, and $N \in \mathbb{N}$,

$$E(N,t,\rho) := \{\omega : T^{\widetilde{\rho},\widetilde{N},\delta}(\omega) \leq t, \text{ for all } \widetilde{\rho} \geq \rho, \widetilde{N} \geq N\}. \tag{8.4}$$

Recall that until time t_0 all individuals we consider behave like $|A|$-independent Kingman coalescents which each consists at time t_0 of only finitely many partition elements. We know therefore that for every δ we have fixed that by construction as $\rho \uparrow \infty$,

$$E(t,\rho) \uparrow E(t,\infty), \quad \text{and} \quad \mathbf{P}[E(t,\infty)] = 1, \,\forall\, t \geq t_0, \tag{8.5}$$

and

$$E(N,t,\rho) \uparrow E(N,t,\infty), \quad \text{and} \quad \mathbf{P}[E(N,t,\infty)] = 1 \,\forall\, t \geq t_0. \tag{8.6}$$

Note that by definition, on the events $E(t,\rho)$ and $E(N,t,\rho)$ the pruned $\mathcal{I}^{\widetilde{\rho}}$-coalescents and the pruned $\mathcal{I}^{\widetilde{\rho},G_N}$-coalescents have for $\widetilde{\rho} \geq \rho$ completely coupled location vectors and the same holds for the processes giving the respective numbers of partition elements.

Step 2 (Representation by the coalescent) After this preparation, we return to the historical look-down process. We can consider the functionals $\widetilde{\zeta}_t^*$, and $\widetilde{\zeta}_t^{*,N}$ respectively, of the historical look-down process for $t \in \mathbb{R} \cup \{\infty\}$,

$$\widetilde{\zeta}_t^* := \lim_{\rho\to\infty} \rho^{-1} \sum_{\iota \in \mathcal{I}^\rho} \delta_{\pi^*_{G\times K} X_t^{*,\iota}}, \tag{8.7}$$

and for $N \in \mathbb{N}$,

$$\widetilde{\zeta}_t^{*,N} := \lim_{\rho\to\infty} \rho^{-1} \sum_{\iota \in \mathcal{I}^\rho} \delta_{\pi^*_{G\times K} X_t^{*,N,\iota}}. \tag{8.8}$$

These objects exist due to the exchangeability of the individuals including their type in the historical look-down at each fixed site.

On the other hand by the *strong duality* in (3.12) we can express these functionals equivalently in terms of the coalescent. Having done this, we can make use of the events we defined earlier. This means, in particular, that we can now replace the functionals given in (8.7) and (8.8) by the analogous objects based on the paths of the coalescent. That is, for all $N \in \mathbb{N}$, and $t \geq 0$,

$$\widehat{\zeta}_t^* := \lim_{\rho\to\infty} \rho^{-1} \sum_{\iota\in\mathcal{I}^\rho} \delta_{(L_t^{*,\iota},\Delta_\theta(C_t^{*,\iota}))}, \quad a.s., \tag{8.9}$$

and

$$\widehat{\zeta}_t^{*,N} := \lim_{\rho\to\infty} \rho^{-1} \sum_{\iota\in\mathcal{I}^\rho} \delta_{(L_t^{*,N,\iota},\Delta_\theta(C_t^{*,N,\iota}))}, \quad a.s. \tag{8.10}$$

The new objects have for each fixed t the same distribution as the original ones in (8.7) and (8.8).

Finally note that we have in (8.9) and (8.10) measures on paths in $G \times K$. We shall below approximate them uniformly in other parameters via a coalescent which consists for positive times only of finitely many partition elements and on the events $E(t,\rho)$ and $E(N,t,\rho)$ we see no new partition elements if we increase ρ. The K-part of the path depends in its description explicitly only on the partitions and only indirectly on the locations of the partition elements.

Focus first on the assertion concerning $\widehat{\zeta}_t^*$. We observe that we can approximate $\widehat{\zeta}_t^*$ by a quantity

$$\widehat{\zeta}_t^{*,\delta} := \lim_{\rho\to\infty} \rho^{-1} \sum_{\iota\in\mathcal{I}^\rho;\, \exists\, \pi\in\widehat{C}_t^{A,\delta}\,:\iota\in\pi} \delta_{(\widehat{L}_{t\vee t_0}^{*,\delta,\iota},\Delta_\theta(\widehat{C}_{t\vee t_0}^{*,\delta,\iota}))}, \quad a.s., \tag{8.11}$$

by passing from the coalescent to the pruned coalescent and letting then $\delta \to 0$. This approximation is uniform in t, the way it is constructed by pruning the initial state. This means that independent of everything else and independently for each individual, such an individual is removed with probability less than or equal to δ. Therefore if we consider empirical measures over the corresponding configuration it is changed in variational norm by at most δ and hence this difference tends to 0 as $\delta \to 0$ uniformly in all other parameters.

Analogously we proceed with $\widehat{\zeta}^{*,N}$.

Step 3 (Uniform convergence for models on $\mathbb{Z}^d$) On the events $E(t,\rho)$ respectively $E(N,t,\rho)$ the above almost sure convergence in ρ is, as we shall see below, by construction uniform in t or N, respectively. We then obtain from the above mentioned uniform a.s. convergence a statement about a certain uniformity in the weak convergence as $\rho \to \infty$ in t and N, respectively.

What type of convergence do we need? Return to the IMM on the infinite group and the convergence statement for it first. The weak convergence means that we consider the ancestral path of individuals which are at time t in some finite spatial window A of observation and we observe their paths on a finite grid of time points viewed backwards. Let A from the construction in the previous step be such a set of sites and fix time points $t_1, t_2, \cdots, t_n$ such that $0 \leq t_1 < t_2 < \cdots < t_n \leq t$. Then consider the random measure induced by the historical look-down process as on the right hand side of (4.20) restricted to the sub-σ-algebra corresponding to A and the finite set of observation points in time $\{t, t-t_1, \cdots, t-t_n\}$. This new random measure has to be analyzed now.

If we translate this into the dual process, we have to start with the individuals in A and observe them at times $t_1, t_2, \cdots, t_n$ and t. The last time index is needed to obtain the type components. This means that for the dual process we look at the measure generated by the skeleton for the location path given by $t_1, \cdots, t_n$ and the partition at time t.

We know that the coalescent converges as $t \to \infty$ and this holds even for the historical coalescent and due to the pruning we have for $t \geq t_0$ a *finite* coalescent. Hence we know $\widehat{C}_t^{A,\delta} = \widehat{C}_\infty^{A,\delta}$ for large enough t with high probability (that is, the probability is at least $1-\varepsilon$, for some prescribed $\varepsilon > 0$).

Next fix a set of time points for the coalescent corresponding in the original process to the selected set of time-points viewed backwards from time t, i.e. time points of the form $\{t, t-t_1, t-t_2, ..., t-t_n\}$ with $t_1, \ldots, t_n$ prescribed. Observe the coalescent path only on that skeleton of time points and on the final point t. This defines a sub-σ-algebra on $\mathcal{M}(\mathcal{D}(\mathbb{R}, G \times K))$. Then restrict the functional in (8.11) to this sub-σ-algebra, which results in some new random measure on $\mathcal{D}(\mathbb{R}, G \times K)$ equipped with this sub-σ-algebra. We denote this random measure by $Y_{\rho,t}$.

We claim that based on our construction we can write our restricted random measures for this fixed $\varepsilon > 0$ and for fixed $\delta > 0$ (the dependence in δ is not apparent from the notation) in the form:

$$Y_{\rho,t} := Y^{1,\varepsilon}_{\rho,\infty} + Y^{2,\varepsilon}_{\rho,t}, \tag{8.12}$$

by considering $Y_{\rho,t}$ on the event $\widehat{C}^{A,\delta}_t = \widehat{C}^{A,\delta}_\infty$ and its component. Then (0 is here the 0-measure) by the definition of $Y^{2,\varepsilon}_{\rho,t}$

$$\mathbf{P}[Y^{2,\varepsilon}_{\rho,t} \neq 0] \leq \varepsilon, \quad \forall\, t \geq t_0(\varepsilon), \rho \in [0,\infty), \tag{8.13}$$

and by (8.5)

$$Y^{1,\varepsilon}_{\rho,\infty} \underset{\rho\to\infty}{\longrightarrow} Y^{1,\varepsilon}_{\infty,\infty}, \quad Y^{1,\varepsilon}_{\infty,\infty} \underset{\varepsilon\to 0}{\longrightarrow} Y_{\infty,\infty}. \tag{8.14}$$

Namely there are no further coalescence events on $E(\rho,t)$ for $\tilde{t} \geq t,\ \rho \geq \tilde{\rho}$, but only the finitely many end points of the locations still fluctuate, they are however independent of ρ.

We see that we then can conclude that for every fixed value of $\delta > 0$,

$$\lim_{\rho\to\infty} \lim_{t\to\infty} \mathcal{L}[Y^{1,\varepsilon}_{\rho,t}] = \lim_{t\to\infty} \lim_{\rho\to\infty} \mathcal{L}[Y^{1,\varepsilon}_{\rho,t}]. \tag{8.15}$$

This applies to the pruned coalescent and the functional $\widehat{\zeta}^*_t$ for every restriction to spatial sets A and finite sets of time-points of the described type and for every $\varepsilon > 0$, and $\delta > 0$ prescribed. This implies in turn the equality in law of $\widehat{\zeta}^{*,\rho}_t$ and $\zeta^{*,\rho}_t$ and letting ε and δ tend to zero (recall that the approximation $\delta \to 0, \varepsilon \to 0$ is uniformly in t), that:

$$\mathcal{L}[\zeta^{*,\rho}_\infty] = \lim_{\rho\to\infty} \{ \lim_{t\to\infty} \mathcal{L}[\rho^{-1}\eta^{*,\rho,t}_0]\}, \tag{8.16}$$

where $\eta^{*,\rho,t}$ is the historical IMM started at time $-t$.

This gives immediately the assertion of Theorem 3 by what we have shown already for $(\eta^*_t)_{t\geq 0}$. Namely we know that the limit in (3.45) can be obtained by first carrying out the limit $t \to \infty$ in the IMM, which we can represent using Theorem 4 and then form the diffusion limit $\rho \to \infty$ in (3.44) which exists due to de Finetti's theorem.

Step 4 *(Uniform convergence on torus)*

Next we turn to the assertion concerning $\zeta^{*,N}_t$. We need here

$$\lim_{N\to\infty} \mathcal{L}[\zeta^{*,N}_{t(2N+1)^d}] = \lim_{\rho\to\infty} \lim_{N\to\infty} \mathcal{L}\left[\rho^{-1}\eta^{*,\rho,N}_{t(2N+1)^d}\right]. \tag{8.17}$$

Knowing this, we have to show that if we take the r.h.s. in (3.63) and perform for a realization the transformation given in (1.20) and then let $\rho \to \infty$, so that we obtain the r.h.s. in (3.67).

For this purpose it suffices to verify that carrying out the transformation (1.20) for a realization of $\mathcal{H}^*_{\pi,\rho,1}$ that for $\rho \to \infty$ we obtain as weak limit $\widehat{\mathcal{H}}^*_{\pi,1}$. This however was the way these objects had been defined. Hence the result (8.17) would imply Theorem 5.

In the proof of (8.17) we have, compared to the argument in (8.16), to use $E(N,t,\rho)$ instead of $E(t,\rho)$ and to replace the fact that $\widehat{C}^{A,\delta}_t = \widehat{C}^{A,\delta}_\infty$ for t large with high probability used previously, now by the complementary fact that the number of partition elements in $\widehat{C}^{A,N}_{t(2N+1)^d}$ converges in law to the quantity $K_{\kappa t}(\widehat{C}^{A,\delta}_\infty)$ as $N \to \infty$. Furthermore the random mechanism $K_{\kappa t}$ has a distribution which is independent of the locations of our paths as we observe them in any time interval between $[\widetilde{t} - T, \widetilde{t}]$ for every finite T and $\widetilde{t} = \widetilde{t}(N) = (2N+1)^d t$.

Therefore we can conclude that with probability at least $1 - \varepsilon$ the coalescence events represent a random variable which can be obtained by restricting $K_{\kappa t}(C^A_\infty)$ to a suitable large event and hence the law of our coalescent has, for N large enough, variational distance at most ε from $\mathcal{L}[K_{\kappa t}(C^A_\infty)]$. Then we can proceed as in the previous case and conclude the argument.

Acknowledgement We are grateful to Anja Sturm for some very useful comments, and to an anonymous referee for numerous valuable comments and suggestions that helped in improving the presentation of the paper.

References

[1] **Aldous, D. (1989)** Stopping times and tightness II, Ann. Probab., **17(2)**, 586-595.

[2] **Bhattacharya, R.N. and Rao, R.R. (1976)** Normal approximation and asymptotic expansions. Wiley New-York.

[3] **Bramson, M., Cox, J.T., and Griffeath, D. (1986)** Consolidation rates for two interacting systems in the plane, Probab. Th. Rel. Fields, **73**, 613-625.

[4] **Bramson, M., and Griffeath, D. (1980)** Asymptotics for interacting particle systems on $\mathbb{Z}^d$, Z. Wahrscheinlichkeit verw. Gebiete, **53**, 183-196.

[5] **Cox, J.T. (1989)** Coalescing Random Walks and Voter model consensus times on the torus in $\mathbb{Z}^d$, Ann. Probab., **17**, 1333-1366.

[6] **Cox, J.T., Fleischmann, K. and Greven, A. (1996)** Comparison of interacting diffusions and an application to their ergodic theory, Probab.Theory Rel. Fields **105**(4), 513-528,

[7] **Cox, J. T., and Geiger, J. (2000)** The genealogy of a cluster in the multitype voter model, Ann. Probab., **28**, 1588-1619.

[8] **Cox, J.T., and Greven, A. (1990)** On the long-term behavior of some finite particle systems, Probab. Theory Rel. Fields, **85**, 195-237.

[9] **Cox, J.T. and Greven, A. (1994)** Ergodic theorems for systems of locally interacting diffusions, Ann. of Prob. **22**(2), 833-853.

[10] **Cox, J.T., Greven, A. and Shiga, T. (1995)** Finite and infinite systems of interacting diffusions, Probab. Rel. Fields, **103**, 165-197.

[11] **Cox, J.T., and Griffeath, D. (1986)** Diffusive clustering in the two dimensional voter model. Ann. Prob., **14**(2), 347-370.

[12] **Dawson, D.A. (1977)** The critical measure diffusion process, Z. Wahrscheinlichkeitstheorie verw. Gebiete, **40**, 125-145.

[13] **Dawson, D.A. (1993)** Measure-valued Markov processes, École d'Été de Probabilités de Saint Flour XXI – 1991, Lect. Notes in Math. **1541**, 1-260, Springer-Verlag.

[14] **Dawson, D.A., and Greven A. (2003)** State dependent multitype spatial branching processes and their long-time behavior, EJP **8 (4)**, 1-93.

[15] **Dawson, D.A., Greven, A. and Vaillancourt, J. (1995)**, *Equilibria and Quasi-equilibria for infinite systems of Fleming-Viot processes.* Transactions of the American Math. Society, Vol. 347, Number 7, 2277-2360.

[16] **Dawson, D.A. and Perkins, E.A. (1991)** Historical Processes, Memoirs of the AMS **93**, 454.

[17] **Deuschel, J. D. (1988)** Central limit theorem for an infinite lattice system of interacting diffusion processes, Ann. Probab. **16**, 700-716.

[18] **Donnelly, P., Evans, S.N., Fleischmann, K., Kurtz, T.G., and Zhou, X. (1999)** Continuum-Sites Stepping–Stone Models, Coalescing Exchangeable Partitions, and Random Trees, Ann Probab., 28, 1063-1110.

[19] **Donnelly, P., and Kurtz, T.G. (1996)** A countable representation of the Fleming-Viot processes, Ann. Probab., **24(2)**, 698-742.

[20] **Donnelly, P. and Kurtz, T.G. (1998)** Particle representation for measure-valued population models, Ann. Probab., **27**, 166-205.

[21] **Donnelly, P. and Kurtz, T.G. (1999)** Genealogical processes for Fleming-Viot models with selection and recombination, Ann. Appl. Probab., **9**, 1091-1148.

[22] **Duquesne, T., and Le Gall, J.-F. (2002)** Random trees, Lévy processes and spatial branching processes, Astérique, **281**.

[23] **Durrett, R. (1979)** A infinite particle system with additive interactions. Adv. Appl. Prob. **11**, 355-383.

[24] **Etheridge, A. (1993)** Asymptotic behavior of measure-valued critical branching processes, Proc. AMS, **118(4)**, 1251-1261.

[25] **Ethier, S.N., and Kurtz, G.T. (1985)** Markov processes, characterization and convergence, Wiley.

[26] **Fleischman, J. (1978)** Limiting distributions for branching random fields, Trans. AMS, **239**, 353-389.

[27] **Fleischmann, K., and Greven, A. (1994)** Diffusive clustering in an infinite system of hierarchically interacting diffusions, Probab. Theory Rel. Fields, **98**, 517-566.

[28] **Fleischmann, K., and Greven, A. (1996)** Time–space analysis of the cluster formation in interacting diffusions, EJP, **1**, Paper 6, 1-46, http://www.emis.de/journals/EJP-ECP/ejp1contents.html,

[29] **Gorostiza, L.G. and Wakolbinger, A. (1991)** Persistence criteria for a class of critical branching particle systems in continuous time, Ann. Probab., **19(1)**, 266-288.

[30] **Greven, A. (2000)**: Interacting stochastic systems: Long-Time behavior and its renormalization analysis. Jber. d. Dt. Math. Verein, **102** (2000), 149-170.

[31] **Greven, A., Limic, V. and Winter A.** Cluster formation in spatial Moran models in criticial dimension via particle representations, (in preparation 2005).

[32] **Greven, A., Pfaffelhuber, P., and Winter, A.** Time-space structure of genealogies of spatially interacting Moran models and applications to interacting Fisher-Wright diffusions, (in preparation).

[33] **Harris, T. (1978)** Additive set-valued Markov processes and graphical methods, Ann. Probab., **6**, 355-378.

[34] **Holley, R. and Liggett, T.M. (1975)** Ergodic theorems for weakly interacting systems and the voter model, Ann. Probab., **3**, 643-663.

[35] **Kallenberg, O. (1977)** Stability of critical cluster fields, Math. Nachr., **77**, 7-43.
[36] **Kingman, J.F.C. (1982)** The coalescent, Stochastic Process. Appl., **13**, 235-248.
[37] **Kurtz, T. G. (1998)** Particle representation for measure-valued population processes with spatially varying birth rates, Proceedings of International Conference on Stochastic Models (Ottawa, June, 1998), CMS Conference Proceedings, AMS, Providence, **26**, 299-317.
[38] **Leopold, K. (2001)** Das Moranmodell mit Selektion, PhD-Thesis, Universität Erlangen-Nürnberg.
[39] **Liggett, T.(1985)** Infinite particle systems, Springer-Verlag.
[40] **Liggett, T. and Spitzer, F. (1981)** Ergodic theorems for coupled random walk and other systems with locally interacting components, Z. Wahrscheinlichkeitstheorie verw. Gebiete, , 443-468.
[41] **Limic, V. and Sturm, A. (2005)** The spatial Λ-coalescent. Preprint.
[42] **Rebolledo, R. (1980)** Sur l'existence de solution a certains problemes de semimartingales, C R Acad Sci Paris, **290**.
[43] **Shiga, T. (1980)** An interacting system in population genetics, J. Math. Kyoto Univ., **20**, 213-242.
[44] **Winter, A. (2002)** Multiple Scale Analysis of Spatial Branching Processes under the Palm Distribution, EJP, **7(13)**, 1-72.

Index

(C_t, L_t), spatially structured coalescent, 1304
(C_t^N, L_t^N), spatially structured coalescent restricted on torus, 1333
(Γ_s^t, ℓ_s^t), genealogy of the look-down process, 1303
$(\Pi^{\mathcal{I}}, \mathbb{d})$, metrized space of partitions of the index set, 1308
(χ_t), system of independent RWs, 1291
$(\widetilde{\Gamma}_s^{t,\rho}, \widetilde{\ell}_s^{t,\rho})$, genealogy of the enriched IMM, 1304
(z_t), Fisher-Wright diffusion, 1316
(K_t), Kingman's coalescent, 1312
A, RW generator, 1298
$A^{\mathcal{I}}$, generator of $\mathcal{I}$-indexed system of RWs, 1298
C_∞, final family partition, 1308
$D^{\rho,N}$, empirical density, 1316
G, geographic space, 1290
G_N, torus of side length $(2N+1)$, 1311
K, type space, 1290
Q_t, transition kernel of a Fisher-Wright diffusion, 1316
$\mathcal{H}_\theta^*$, entrance law of the historical look-down, 1308
$\Psi(\theta)$, initial state of the historical look-down, 1302
α, weight on G, 1291
$\bar{Q}_t^*$, joint law of the final partition of a spatially structured coalescent and a Kingman's coalescent started in that final partition, 1313
$\bar{a}(x, y)$, reversed discrete time RW kernel, 1291
$\beta_N(t)$, the critical time scale, 1312
$\ast$, convolution, 1315
$\circ$, composition, 1294
η, IMM, 1290
η^*, historical IMM, 1293
γ, resampling rate, 1290
ι, index of an individual, 1295
κ, probability for non-coalescence of two partition elements in the spatially structured coalescent, 1312
π_G, projection on G, 1294
ρ, spatial density, 1292
τ_t^*, shift operator on paths, 1313
$\widehat{a}(x, y)$, symmetrized discrete time RW kernel, 1306
$\widehat{r}_A$, restriction operator on probability laws on paths, 1309
$\widehat{\mathcal{H}}_\theta$, equilibrium measure for IFWD, 1307
$\widetilde{\mathcal{H}}_{\rho,\theta}^*$, entrance law of the historical IMM, 1308
$\widetilde{\Psi}(\theta, \rho)$, Poisson configuration with spatial intensity ρ and type 1 probability θ, 1306
$\widetilde{\mathcal{H}}_{\rho,0}, \widetilde{\mathcal{H}}_{\rho,1}$, Poisson fields on $G \times \{0, 1\}$ supported on monotype paths, 1306
$\widetilde{\mathcal{H}}_{\rho,\theta}$, equilibrium measure for IMM, 1307
$\widetilde{\mathcal{H}}_{\rho,\theta}^*(s, t)$, law of historical IMM at time t started at time s in collection of constant paths generated by $\Psi(\rho, \theta)$, 1307
ξ, enriched IMM which distinguishes between individuals, 1290
ξ^*, enriched historical IMM which distinguishes between individuals, 1296
ζ, IFWD, 1291
ζ^*, historical IFWD, 1294
$\{\Delta_\theta(\pi)\}$, reservoir of types for the possible family patches, 1309
$a(x, y)$, discrete time RW kernel, 1290
$a_N(x, y)$, discrete time RW kernel on the torus, 1311
$a_t(x, y)$, continuous time RW kernel, 1290
g, expected number of returns, 1312
r_A, restriction operator on paths, 1309
s_β, scaling operator on paths, 1347
$\mathcal{D}([0, \infty)$,E), cadlag maps from $[0, \infty)$ to E, 1292
$\mathcal{E}_G$, (Liggett-Spitzer) state space, 1291
$\mathcal{E}_G^*$, (Liggett-Spitzer) state space for the historical processes, 1293
$\mathcal{H}_\theta^*(s, t)$, law of historical look-down at time t started at time s in collection of constant paths generated by $\Psi(\theta)$, 1307
$\mathcal{I}$, set of indices representing individuals, 1294
$\mathcal{I}^\rho$, set of indices restricted to individuals with label smaller than ρ, 1294
$\mathcal{M}(\cdot)$, $\mathbb{R}^+$-valued locally finite measures, 1291
$\mathcal{N}(\cdot)$, $\mathbb{N}$-valued locally finite measures, 1290
$\mathcal{P}$, partition, 1304
$\mathcal{U}$, set of labels, 1294

Mathematisches Institut, Universität Erlangen–Nürnberg, Bismarckstrasse $1\frac{1}{2}$, 91054 Erlangen, GERMANY
E-mail address: greven@mi.uni-erlangen.de

Department of Mathematics, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada
E-mail address: limic@math.ubc.ca

Mathematisches Institut, Universität Erlangen–Nürnberg, Bismarckstrasse $1\frac{1}{2}$, 91054 Erlangen, GERMANY
E-mail address: winter@mi.uni-erlangen.de