Alea 6, 25–61 (2009)

A modified lookdown construction for the Xi-Fleming-Viot process with mutation and populations with recurrent bottlenecks

Matthias Birkner, Jochen Blath, Martin Möhle, Matthias Steinrücken and Johanna Tams

Matthias Birkner, Weierstraß-Institut für Angewandte Analysis und Stochastik, Mohrenstraße 39, 10117 Berlin, Germany
E-mail address: birkner@wias-berlin.de
URL: http://www.wias-berlin.de/~birkner

Jochen Blath, Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany
E-mail address: blath@math.tu-berlin.de
URL: http://www.math.tu-berlin.de/~blath

Martin Möhle, Mathematisches Institut, Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany
E-mail address: moehle@math.uni-duesseldorf.de (corresponding author)
URL: http://www.math.uni-duesseldorf.de/Personen/indiv/Moehle

Matthias Steinrücken, Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany
E-mail address: steinrue@math.tu-berlin.de
URL: http://www.math.tu-berlin.de/~steinrue

Johanna Tams
E-mail address: johannatams@gmx.de

Abstract. Let Λ be a finite measure on the unit interval. A Λ-Fleming-Viot process is a probability measure valued Markov process which is dual to a coalescent with multiple collisions (Λ-coalescent) in analogy to the duality known for the classical Fleming-Viot process and Kingman's coalescent, where Λ is the Dirac measure in 0.

We explicitly construct a dual process of the coalescent with simultaneous multiple collisions (Ξ-coalescent) with mutation, the Ξ-Fleming-Viot process with mutation, and provide a representation based on the empirical measure of an exchangeable particle system along the lines of Donnelly and Kurtz (1999). We establish pathwise convergence of the approximating systems to the limiting Ξ-Fleming-Viot process with mutation. An alternative construction of the semigroup based on the

Received by the editors August 4, 2008; accepted October 23, 2008.
2000 *Mathematics Subject Classification.* 60K35; 60G09; 92D10 60C05; 92D15 .
Key words and phrases. Coalescent, duality, Fleming-Viot process, measure-valued process, modified lookdown construction.

Hille-Yosida theorem is provided and various types of duality of the processes are discussed.

In the last part of the paper a population is considered which undergoes recurrent bottlenecks. In this scenario, non-trivial Ξ-Fleming-Viot processes naturally arise as limiting models.

1. Introduction and main results

1.1. *Motivation.* One of the fundamental aims of mathematical population genetics is the construction of population models in order to describe and to analyse certain phenomena which are of interest for biological applications. Usually these models are constructed such that they describe the evolution of the population under consideration forwards in time. A classical and widely used model of this kind is the Wright-Fisher diffusion, which can be used for large populations to approximate the evolution of the fraction of individuals carrying a particular allele. On the other hand it is often quite helpful to look from the present back into the past and to trace back the ancestry of a sample of n individuals, genes or particles. In many situations, the Kingman coalescent (Kingman, 1982a,b) turns out to be an appropriate tool to approximate the ancestry of a sample taken from a large population. It is well known that the Wright-Fisher diffusion is dual to the block counting process of the Kingman coalescent (Donnelly, 1986; Möhle, 2001). More general, the Fleming-Viot process (Fleming and Viot, 1979), a measure-valued extension of the Wright-Fisher diffusion, is dual to the Kingman coalescent.

Such and similar duality results are quite common in particular in the physics literature on interacting particle systems (Liggett, 1985) and in the more theoretical literature on mathematical population genetics (Alkemper and Hutzenthaler, 2007; Athreya and Swart, 2005; Donnelly and Kurtz, 1996, 1999; Ethier and Krone, 1995; Hiraba, 2000; Möhle, 1999, 2001). Donnelly and Kurtz (1996) established a so-called lookdown construction and used this construction to show that the Fleming-Viot process is dual to the Kingman coalescent. This construction and corresponding duality results have been extended (Donnelly and Kurtz, 1999; Bertoin and Le Gall, 2003, 2005, 2006) to the Λ-Fleming-Viot process, which is the measure-valued dual of a coalescent process allowing for multiple collisions of ancestral lineages. For more information on coalescent processes with multiple collisions, so-called Λ-coalescents, we refer to Pitman (1999) and Sagitov (1999).

There exists a broader class of coalescent processes (Möhle and Sagitov, 2001; Schweinsberg, 2000; Sagitov, 2003) in which many multiple collisions can occur with positive probability simultaneously at the same time. These processes can be characterized by a measure Ξ on an infinite simplex and are hence called Ξ-coalescents. It is natural to further extend the above constructions and results to this full class of coalescent processes and, in particular, to provide constructions of the dual processes, called Ξ-Fleming-Viot processes. Although such extensions have been briefly indicated in Donnelly and Kurtz (1999) and Bertoin and Le Gall (2003), these extensions have not been carried out in detail yet. Ξ-coalescents have also recently been applied to study population genetic problems, see Taylor and Véber (2009); Sargsyan and Wakeley (2008).

The motivation to present this paper is hence manifold. We explicitly construct the Ξ-Fleming-Viot process and provide a representation via empirical measures of an exchangeable particle system in the spirit of Donnelly and Kurtz (1996, 1999). We furthermore establish corresponding convergence results and pathwise duality to the Ξ-coalescent. We also provide an alternative, more classical functional-analytic construction of the Ξ-Fleming-Viot process based on the Hille-Yosida theorem and present representations for the generator of the Ξ-Fleming-Viot process. Our approaches include neutral mutations. The results give insights into the pathwise structure of the Ξ-Fleming-Viot process and its dual Ξ-coalescent. Examples and situations are presented in which certain Ξ-Fleming-Viot processes and their dual Ξ-coalescents occur naturally.

1.2. *Moran models with (occasionally) large families.* Consider a population of fixed size $N \in \mathbb{N} := \{1, 2, \ldots\}$ and assume that each individual is of a certain type, where the space E of possible types is assumed to be compact and Polish. Furthermore assume that for each vector $\mathbf{k} = (k_1, k_2, \ldots)$ of integers satisfying $k_1 \geq k_2 \geq \cdots \geq 0$ and $\sum_{i=1}^{\infty} k_i \leq N$ a non-negative real quantity $r_N(\mathbf{k}) \geq 0$ is given. The population is assumed to evolve in continuous time as follows. Given a vector $\mathbf{k} = (k_1, \ldots, k_m, 0, 0, \ldots)$, where $k_1 \geq \cdots \geq k_m \geq 1$ and $k_1 + \cdots + k_m \leq N$, with rate $r_N(\mathbf{k})$ we choose randomly m groups of sizes $k_1, \ldots, k_m$ from the present population. Inside each of these m groups we furthermore choose randomly a 'parent' which forces all individuals in its group to change their type to the type of that parent. We say that a $\mathbf{k}$-reproduction event occurs with rate $r_N(\mathbf{k})$. The classical Moran model corresponds to $r_N(2, 0, 0, \ldots) = N$.

Except for the fact that these models are formulated in continuous time, they essentially coincide with the class of neutral exchangeable population models with non-overlapping generations introduced by Cannings (1974, 1975). Starting with the seminal work of Kingman (1982a,b), the genealogy of samples taken from such populations is well understood, in particular for the situation when the total population size N tends to infinity.

1.3. *Genealogies and exchangeable coalescents.* For neutral population models of large, but fixed population size and finite-variance reproduction mechanism, Kingman (1982b) showed that the genealogy of a finite sample of size n can be approximately described by the so called n-coalescent $(\Pi_t^{\delta_{\mathbf{0}},(n)})_{t \geq 0}$. The n-coalescent is a time-homogeneous Markov process taking values in $\mathcal{P}_n$, the set of partitions of $\{1, \ldots, n\}$. If i and j are in the same block of the partition $\Pi_t^{\delta_{\mathbf{0}},(n)}$, then they have a common ancestor at time t ago. $\Pi_0^{\delta_{\mathbf{0}},(n)}$ is the partition of $\{1, \ldots, n\}$ into singleton blocks. The transitions are then given as follows: If there are b blocks at present, then each pair of blocks merges with rate 1, thus the overall rate of seeing a merging event is $\binom{b}{2}$. Note that only binary mergers are allowed and that at some random time, all individuals will have a (most recent) common ancestor.

Kingman (1982b) also showed that there exists a $\mathcal{P}_{\mathbb{N}}$-valued Markov process $(\Pi_t^{\delta_{\mathbf{0}}})_{t \geq 0}$, where $\mathcal{P}_{\mathbb{N}}$ denotes the set of partitions of $\mathbb{N}$. This process, the so-called Kingman coalescent, is characterised by the fact that for each n the restriction of $(\Pi_t^{\delta_{\mathbf{0}}})_{t \geq 0}$ to the first n natural numbers is the n-coalescent. The process can be constructed by an application of the standard Kolmogoroff extension theorem, since the restriction of every n-coalescent to $\{1, \ldots, m\}$, where $1 \leq m \leq n$, is an m-coalescent.

Whereas the Kingman coalescent allows only for binary mergers, the idea of a time-homogeneous $\mathcal{P}_{\mathbb{N}}$-valued Markov process that evolves by the coalescence of blocks was extended by Pitman (1999) and Sagitov (1999) to coalescents where multiple blocks are allowed to merge at the same time, so-called Λ-coalescents, which arise as the limiting genealogy of populations where the variance of the offspring distribution diverges as the population size tends to infinity. Möhle and Sagitov (2001) and Schweinsberg (2000) introduced the even larger class of coalescents with simultaneous multiple collisions, also called exchangeable coalescents or Ξ-coalescents, which describe the genealogies of populations allowing for large family sizes.

Schweinsberg (2000) showed that any exchangeable coalescent $(\Pi_t^{\Xi})_{t\geq 0}$ is characterised by a finite measure Ξ on the infinite simplex

$$\Delta \ := \ \{\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \dots) : \zeta_1 \geq \zeta_2 \geq \dots \geq 0, \textstyle\sum_{i=1}^{\infty} \zeta_i \leq 1\}.$$

Throughout the paper, for $\boldsymbol{\zeta} \in \Delta$, the notation $|\boldsymbol{\zeta}| := \sum_{i=1}^{\infty} \zeta_i$ and $(\boldsymbol{\zeta}, \boldsymbol{\zeta}) := \sum_{i=1}^{\infty} \zeta_i^2$ will be used for convenience. Note that Möhle and Sagitov (2001) provide an alternative (though somewhat less intuitive) characterisation of the Ξ-coalescent based on a sequence of finite symmetric measures $(F_r)_{r\in\mathbb{N}}$. Coalescent processes with multiple collisions (Λ-coalescents) occur if the measure Ξ is concentrated on the subset of all points $\boldsymbol{\zeta} \in \Delta$ satisfying $\zeta_i = 0$ for all $i \geq 2$. The Kingman-coalescent corresponds to the case $\Xi = \delta_{\mathbf{0}}$. It is convenient to decompose the measure Ξ into a 'Kingman part' and a 'simultaneous multiple collision part', that is, $\Xi = a\delta_{\mathbf{0}} + \Xi_0$ with $a := \Xi(\{\mathbf{0}\}) \in [0, \infty)$ and $\Xi_0(\{\mathbf{0}\}) = 0$. The transition rates of the Ξ-coalescent Π^{Ξ} are given as follows. Suppose there are currently b blocks. Exactly $\sum_{i=1}^{r} k_i$ blocks collide into r new blocks, each containing $k_1, \dots, k_r \geq 2$ original blocks, and s single blocks remain unchanged, such that the condition $\sum_{i=1}^{r} k_i + s = b$ holds. The order of $k_1, \dots, k_r$ does not matter. The rate at which the above collision happens is then given as (Schweinsberg, 2000, Theorem 2)

$$\begin{aligned}\lambda_{b;k_1,\dots,k_r;s} \quad = \quad & a\mathbb{1}_{\{r=1,k_1=2\}} \qquad\qquad (1.1)\\ & + \int_{\Delta} \sum_{l=0}^{s} \binom{s}{l} (1 - |\boldsymbol{\zeta}|)^{s-l} \sum_{i_1 \neq \dots \neq i_{r+l}} \zeta_{i_1}^{k_1} \cdots \zeta_{i_r}^{k_r} \zeta_{i_{r+1}} \cdots \zeta_{i_{r+l}} \frac{\Xi_0(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta}, \boldsymbol{\zeta})}.\end{aligned}$$

An intuitive explanation of (1.1) is given below in terms of Schweinsberg's (2000) Poisson process construction of the Ξ-coalescent. If $\Xi(\Delta) \neq 0$, then without loss of generality it can be assumed that Ξ is a probability measure, as remarked after Eq. (12) of Schweinsberg (2000). Otherwise simply divide each rate by the total mass $\Xi(\Delta)$ of Ξ.

1.4. *Poisson process construction of the Ξ-coalescent.* It is convenient to give an explicit construction of the Ξ-coalescent in terms of Poisson processes. Indeed, Schweinsberg (2000, Section 3) shows that the Ξ-coalescent can be constructed from a family of Poisson processes $\{\mathfrak{N}_{i,j}^K\}_{i,j\in\mathbb{N},i<j}$ and a Poisson point process $\mathfrak{M}^{\Xi_0}$ on $\mathbb{R}_+ \times \Delta \times [0,1]^{\mathbb{N}}$. The processes $\mathfrak{N}_{ij}^K$ have rate $a = \Xi(\{\mathbf{0}\})$ each and govern the binary mergers of the coalescent. The process $\mathfrak{M}^{\Xi_0}$ has intensity measure

$$dt \otimes \frac{\Xi_0(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta}, \boldsymbol{\zeta})} \otimes (\mathbb{1}_{[0,1]}(t)dt)^{\otimes\mathbb{N}}. \qquad (1.2)$$

These processes can be used to construct the Ξ-coalescent as follows: Assume that before the time t_m the process Π is in a state $\{B_1, B_2, \ldots\}$. If t_m is a point of increase of one of the processes $\mathfrak{N}_{i,j}^K$ (and there are at least $i \vee j$ blocks), then we merge the corresponding blocks B_i and B_j into a single block (and renumber). This mechanism corresponds to the Kingman-component of the coalescent.

The non-Kingman collisions are governed by the points

$$(t_m, \boldsymbol{\zeta}_m, \mathbf{u}_m) \;=\; (t_m, (\zeta_{m1}, \zeta_{m2}, \ldots), (u_{m1}, u_{m2}, \ldots)) \tag{1.3}$$

of the Poisson process $\mathfrak{M}^{\Xi_0}$. The random vector $\boldsymbol{\zeta}_m$ denotes the respective asymptotic family sizes in the multiple merger event at time t_m and the $\mathbf{u}_m$ are "uniform coins", determining the blocks participating in the respective merger groups; see (2.2) or Schweinsberg (2000, Section 3) for details.

1.5. *Ξ-Fleming-Viot processes.* An in many senses dual approach to population genetics is to view a population of finite size as a vector of types $(Y_1^N, \ldots, Y_N^N)$ with values in E^N or as an empirical measure of that vector $\frac{1}{N}\sum_{i=1}^N \delta_{Y_i^N}$ and look at the evolution under mutation and resampling forwards in time. When N tends to infinity one obtains the Fleming-Viot process (Fleming and Viot, 1979). This process has been extended to incorporate other important biological phenomena and has found wide applications, see Ethier and Kurtz (1993) for a survey.

Donnelly and Kurtz (1996) embedded an E^∞-valued particle system into the classical Fleming-Viot process, via a clever lookdown construction, and showed that it is dual to the Kingman-coalescent. This construction and the duality has been extended to the so-called Λ-Fleming-Viot processes, dual to the Λ-coalescents, and investigated by several authors, see, e.g., Donnelly and Kurtz (1999); Birkner et al. (2005); Bertoin and Le Gall (2003, 2005, 2006), or Birkner and Blath (2009) for an overview.

Let $f \in C_b(E^p)$, $\mu \in \mathcal{M}_1(E)$ and $G_f(\mu) := \langle f, \mu^{\otimes p}\rangle$. The generator of the Λ-Fleming-Viot process without mutation has the form (see Birkner et al., 2005, Equation (1.11))

$$L^\Lambda G_f(\mu) = \sum_{J\subset\{1,\ldots,p\},|J|\geq 2} \lambda_{p;|J|;p-|J|} \int \big(f(\mathbf{x}^J) - f(\mathbf{x})\big)\, \mu^{\otimes p}(d\mathbf{x}), \tag{1.4}$$

where

$$(\mathbf{x}^J)_i \;=\; \begin{cases} x_{\min(J)} & \text{if } i \in J, \\ x_i & \text{otherwise.} \end{cases} \tag{1.5}$$

Note that (1.4) includes the generator of the classical Fleming-Viot process (without mutation) if the summation is restricted to sets J satisfying $|J| = 2$.

Our aim in this paper is to present the modified lookdown construction for a measure-valued process that we will call the Ξ-Fleming-Viot process with mutation, or the (Ξ, B)-Fleming-Viot process. The symbol B stands here for an operator describing the mutation process. We will establish its duality to the Ξ-coalescent with mutation. The modified lookdown construction will also enable us to establish some path properties of the (Ξ, B)-Fleming-Viot process.

1.6. *A modified lookdown construction of the (Ξ, B)-Fleming-Viot process.* Consider a population described by a vector $Y^N(t) = (Y_1^N(t), \ldots, Y_N^N(t))$ with values in E^N, where $Y_i^N(t)$ is the type of individual i at time t. The evolution of this population (forwards in time) has two components, namely reproduction and mutation. During its lifetime, each particle undergoes mutation according to the bounded linear mutation operator

$$Bf(x) \ = \ r\int_E (f(y) - f(x))\, q(x, dy), \tag{1.6}$$

where f is a bounded function on E, $q(x, dy)$ is a Feller transition function on $E \times \mathcal{B}(E)$, and $r \geq 0$ is the global mutation rate.

The resampling of the population is governed by the Poisson point process $\mathfrak{M}^{\Xi_0}$, which was introduced as a driving process for the Ξ-coalescent. In particular, the resampling events allow for the simultaneous occurrence of one or more large families. The resampling procedure is described in detail in Section 2. An important fact is that this resampling is made such that it retains exchangeability of the population vector.

In Section 2, we introduce another particle system $X^N = (X_1^N, \dots, X_N^N)$ again with values in E^N. Each particle mutates according to the same generator (1.6) as before. For the resampling event, we will use the same driving Poisson point process $\mathfrak{M}^{\Xi_0}$, but we will use the modified lookdown construction introduced in Donnelly and Kurtz (1999), suitably adapted to our scenario. This (Ξ, B)-lookdown process will be introduced in Section 2.2. It is crucial that the resampling events retain exchangeability of the population vector and that the process $\{X^N(t)\}$ has the same empirical measure $\sum_{i=1}^N \delta_{X_i^N(t)}$ as the process $\{Y^N(t)\}$.

The construction of the resampling events allows us to pass to the limit as N tends to infinity and obtain an E^∞-valued particle system $X = (X_1, X_2, \dots)$. Since this particle system is also exchangeable, this procedure enables us to access the almost sure limit of the empirical measure as N tends to infinity by the de Finetti Theorem (which is not possible for the Y^N).

1.7. *Results.* Let $\mathcal{D}(B)$ denote the domain of the mutation generator B and let $f_1, f_2, \ldots \in \mathcal{D}(B)$ be uniformly bounded functions that separate points of $\mathcal{M}_1(E)$ in the sense that $\int f_k\, d\mu = \int f_k\, d\nu$ for all $k \in \mathbb{N}$ implies that $\mu = \nu$. Such sequences exist, see, e.g. Section 1 (Lemma 1.1 in particular) of Donnelly and Kurtz (1996). We use the metric d on $\mathcal{M}_1(E)$ defined via

$$d(\mu, \nu) \ := \ \sum_k \frac{1}{2^k}\Big| \int f_k\, d\mu - \int f_k\, d\nu \Big|, \qquad \mu, \nu \in \mathcal{M}_1(E) \tag{1.7}$$

and equip the topology of locally uniform convergence on $D_{\mathcal{M}_1(E)}([0,\infty))$ with the metric

$$d_p(\mu, \nu) \ := \ \int_0^\infty e^{-t} d(\mu(t), \nu(t))\, dt. \tag{1.8}$$

Theorem 1.1. *The $\mathcal{M}_1(E)$-valued process $(Z_t)_{t\geq 0}$, defined in terms of the* ordered *particle system $X = (X^1, X^2, \dots)$ by*

$$Z_t \ := \ \lim_{n\to\infty} Z_t^n \ = \ \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}, \quad t \geq 0,$$

is called the Ξ-Fleming-Viot process with mutation operator B *or simply the* (Ξ, B)-Fleming-Viot process. *Moreover, the empirical processes $(Z_t^n)_{t\geq 0}$ converge almost surely on the path space $D_{\mathcal{M}_1(E)}([0,\infty))$ to the càdlàg process $(Z_t)_{t\geq 0}$.*

Since the empirical measures of X^N and Y^N are identical, we arrive at the following corollary.

Corollary 1.2. *Define, for each n,*

$$\tilde{Z}^n_t := \frac{1}{n}\sum_{i=1}^n \delta_{Y_i(t)}, \quad t \geq 0,$$

the empirical process of the n-th unordered particle system, and assume that $\tilde{Z}^n_0 \to Z_0$ weakly as $n \to \infty$.

Then, $(\tilde{Z}^n_t)_{t\geq 0}$ converges weakly on the path space $D_{\mathcal{M}_1(E)}([0,\infty))$ to the (Ξ, B)-Fleming-Viot process $(Z_t)_{t\geq 0}$.

The Markov process $(Z_t)_{t\geq 0}$ is characterized by its generator as follows.

Proposition 1.3. *The (Ξ, B)-Fleming-Viot process $(Z_t)_{t\geq 0}$ is a strong Markov process. Its generator, denoted by L, acts on test functions of the form*

$$G_f(\mu) \ := \ \int_{E^n} f(x_1,\dots,x_n)\,\mu^{\otimes n}(dx_1,\dots,dx_n), \quad \mu \in \mathcal{M}_1(E), \tag{1.9}$$

where $f : E^n \to \mathbb{R}$ is bounded and measurable, via

$$LG_f(\mu) \ := \ L^{a\delta_{\mathbf{0}}}G_f(\mu) + L^{\Xi_0}G_f(\mu) + L^B G_f(\mu), \tag{1.10}$$

where

$$L^{a\delta_{\mathbf{0}}}G_f(\mu) := a \sum_{1\leq i<j\leq n} \int_{E^n} \Big(f(x_1,..,x_i,..,x_i,..,x_n) - f(x_1,..,x_i,..,x_j,..,x_n)\Big)\mu^{\otimes n}(d\mathbf{x}), \tag{1.11}$$

$$L^{\Xi_0}G_f(\mu) := \int_\Delta \int_{E^{\mathbb{N}}} \left[G_f\big((1-|\boldsymbol{\zeta}|)\mu + \textstyle\sum_{i=1}^\infty \zeta_i \delta_{x_i}\big) - G_f(\mu)\right] \mu^{\otimes \mathbb{N}}(d\mathbf{x})\frac{\Xi_0(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta},\boldsymbol{\zeta})}, \tag{1.12}$$

$$L^B G_f(\mu) := r\sum_{i=1}^n \int_{E^n} B_i(f(x_1,\dots,x_n))\mu^{\otimes n}(d\mathbf{x}), \tag{1.13}$$

and $B_i f$ is the mutation operator B, defined in (1.6), acting on the i-th coordinate of f.

Remark 1.4. 1) In 1999, Donnelly & Kurtz established a construction and pathwise duality for the Λ-Fleming-Viot process. In some sense, their paper works under the general assumption "allow simultaneous and/or multiple births and deaths, but we assume that all the births that happen simultaneously come from the same parent" (p. 166), even though they very briefly in Section 2.5 mention a possible extension to scenarios with simultaneous multiple births to multiple parents. In essence, the present paper converts these ideas into theorems.

2) Note that in a similar direction, Bertoin and Le Gall (2003) remark briefly on p. 277 how their construction of the Λ-Fleming-Viot process via flows of bridges can be extended to the simultaneous multiple merger context (but leave details to the interested reader). We are not following this approach, as it is hard to combine with a general type space and general mutation process.

3) The Ξ-Fleming-Viot process has recently been independently constructed by Taylor and Véber (personal communication, 2008) via Bertoin and Le Gall's flow of bridges (see Bertoin and Le Gall, 2003) and Kurtz and Rodriguez' Poisson representation of measure-valued branching processes (see Kurtz and Rodrigues, 2008).

In this context we refer to Taylor and Véber (2009) for a larger study of structured populations, in which Ξ-coalescents appear under certain limiting scenarios.

4) Note that the modified lookdown construction of the Λ-Fleming-Viot process contains all information available about the genealogy of the process and therefore also provides a pathwise embedding of the Λ-*coalescent measure tree* considered by Greven et al. (2009). A similar statement holds for the Ξ-coalescent.

The rest of the paper is organised as follows: In Section 2 we use the Poisson point process $\mathfrak{M}^{\Xi_0}$ to introduce the finite unordered (Ξ_0, B)-Moran model Y^N and the finite ordered (Ξ_0, B)-lookdown model X^N. It is shown that the ordered model is constructed in such a way that we can let N tend to infinity and obtain a well defined limit. We will also show that the reordering preserves the exchangeability property, which will be crucial for the proof in Section 3. In this section, we will introduce the empirical measures of the process Y^N and X^N, show that they are identical and converge to a limiting process having nice path properties, which is the statement of Theorem 1.1.

Section 4.2 will be concerned with the generator of the Ξ_0-Fleming-Viot process. We will give two alternative representations and show that it generates a strongly continuous Feller semigroup. Furthermore, we will show that the process constructed in Section 3 solves the martingale problem for this generator.

One representation of the generator will then be used in Section 5 to establish a functional duality between the Ξ-coalescent and the Ξ-Fleming-Viot process on the genealogical level. Due to the Poissonian construction, this duality can also be extended to a "pathwise" duality. We will also give a function-valued dual, which incorporates mutation.

In Section 6, we look at two examples: The first example is concerned with a population model with recurrent bottlenecks. Here, a particular Ξ-coalescent, which is a subordination of Kingman's coalescent, arises as a natural limit of the genealogical process. The second example discusses the Poisson-Dirichlet-coalescent and obtains explicit expressions for some quantities of interest.

2. Exchangeable E^∞-valued particle systems

2.1. *The canonical* (Ξ, B)*-Moran model.* We can use the Poisson process from Section 1.4 governing the Ξ-coalescent to describe a corresponding forward population model in a canonical way, simply reversing the construction of the coalescent by interpreting the merging events as birth events.

Consider the points

$$(t_m, \boldsymbol{\zeta}_m, \mathbf{u}_m) \;=\; (t_m, (\zeta_{m1}, \zeta_{m2}, \ldots), (u_{m1}, u_{m2}, \ldots)) \tag{2.1}$$

of $\mathfrak{M}^{\Xi_0}$ defined by (1.2). The t_m denote the times of reproduction events. Define

$$g(\boldsymbol{\zeta}, u) \;:=\; \begin{cases} \min\{j \,|\, \zeta_1 + \cdots + \zeta_j \geq u\} & \text{if } u \leq \sum_{i \in \mathbb{N}} \zeta_i, \\ \infty & \text{else.} \end{cases} \tag{2.2}$$

At time t_m, the N particles are grouped according to the values $g(\boldsymbol{\zeta}_m, u_{ml})$, $l = 1, \ldots, N$ as follows: For each $k \in \mathbb{N}$, all particles $l \in \{1, \ldots, N\}$ with $g(\boldsymbol{\zeta}_m, u_{ml}) = k$ form a family. Among each non-trivial family we uniformly pick a 'parent' and change the others' types accordingly. Note that although the jump times (t_m) may be dense in $\mathbb{R}_+$, the condition

$$\int_{\Delta} \sum_i \zeta_i^2 \frac{\Xi_0(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta}, \boldsymbol{\zeta})} \;=\; \Xi_0(\Delta) \;<\; \infty$$

guarantees that in a finite population, in each finite time interval only finitely many non-trivial reproduction events occur. As above, each particle follows an independent mutation process, according to (1.6), in between reproductive events.

We describe the population corresponding to the N-particle (Ξ, B)-Moran model at time $t \geq 0$ by a random vector

$$Y^N(t) \;:=\; (Y_1^N(t), \ldots, Y_N^N(t)) \tag{2.3}$$

taking values in E^N.

Remark 2.1. *Note that this model is completely symmetric, thus, for each t, the population vector $Y^N(t)$ is exchangeable if $Y^N(0)$ is exchangeable.*

2.2. *The ordered model and exchangeability.* We now define an ordered population model with the same family size distribution, extending the ideas of Donnelly and Kurtz (1999) in an obvious way. This time each particle will be attached a "level" from $\{1, 2, \ldots\}$ in such a way that we obtain a nested coupling of approximating (Ξ, B)-Moran models as N tends to infinity. It will be crucial to show that this ordered model retains initial exchangeability, so that the limit as $N \to \infty$ of the empirical measures of the particle systems, at each fixed time, exists by de Finetti's Theorem.

We will refer to this model as the (Ξ, B)-lookdown-model. If the population size is N, it will be described at time t by the E^N-valued random vector

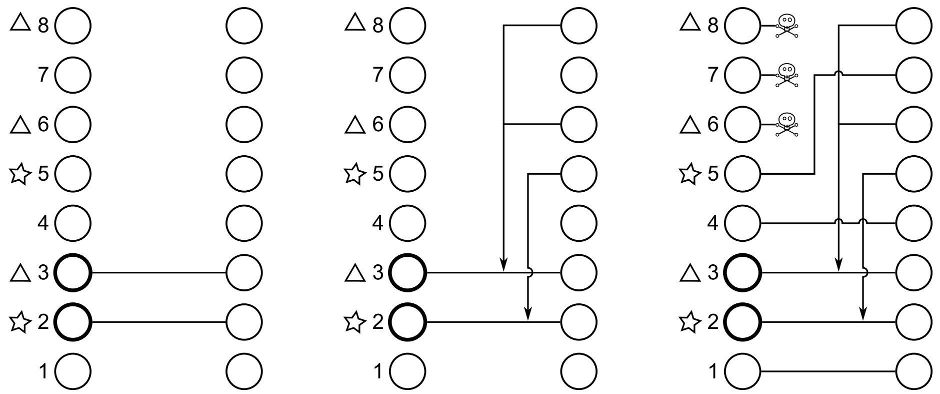
$$X^N(t) \;:=\; (X_1^N(t), \ldots, X_N^N(t)). \tag{2.4}$$

The dynamics works as in the (Ξ, B)-Moran model above, including the distribution of family sizes and the mutation processes for each particle.

In each reproduction step, for each family, a "parental" particle will be chosen, that then superimposes its type upon its family. This time, however, the parental particle will not be chosen uniformly among the members of each family (as in the (Ξ, B)-Moran model). Instead, the parental particle will always be the particle with the lowest level among the members of a family (hence each family member "looks down" to their relative with the lowest level). The attachment of types to levels is then rearranged as follows (see Figure 1 for an illustration):

a) All parental particles of all families (including the trivial ones) will retain their type and level.
b) All levels of members of families will assume the type of their respective parental particle.
c) All levels which are still vacant will assume the pre-reproduction types of non-parental particles retaining their initial order. Once all N levels are filled, the remaining types will be lost.

In this way, the dynamics of a particle, at level l, say, will only depend on the dynamics of the particles with *lower* levels. This consistency property allows to construct all approximating particle systems, as well as their limit as $N \to \infty$, *on the same probability space.*



(a) Parental particles retain type and level.

(b) Family members copy type of parental particle.

(c) Remaining particles retain their order and surplus particles get killed.

FIGURE 1. An illustration of the reproduction mechanism in the (Ξ, B)-lookdown model. The particles at levels 2 and 5 belong to the "star" family, whereas the particles at levels 3, 6 and 8 belong to the "triangle" family. The particles on the remaining levels belong to no family.

Exchangeability of the modified (Ξ, B)-lookdown model is crucial in order to pass to the De Finetti limit of the associated empirical particle systems. For each N, we will show that if $X^N(0)$ is exchangeable, then X^N is exchangeable at fixed times and at stopping times. The proof will rely on an explicit construction of uniform random permutations $\Theta(t)$ which maps X^N to Y^N.

Theorem 2.2. *If the initial distribution of the population vector* $(X_1^N(0), \ldots, X_N^N(0))$ *in the* (Ξ, B)*-lookdown-model is exchangeable, then* $(X_1^N(t), \ldots, X_N^N(t))$ *is exchangeable for each* $t \geq 0$.

For the rest of this section, we omit the superscript N for the population models in an attempt not to get *lost in notation*.

The proof of Theorem 2.2 follows that of Theorem 3.2 in Donnelly and Kurtz (1999). We will construct a coupling via a permutation-valued process $\Theta(t)$ such that

$$(Y_1(t), \ldots, Y_N(t)) = (X_{\Theta_1(t)}(t), \ldots, X_{\Theta_N(t)}(t)) \tag{2.5}$$

and $\Theta(t)$ is uniformly distributed on all permutations of $\{1, \ldots, N\}$ for each t and independent of the empirical process up to time t and the "demographic information" in the model (see (2.15) for a precise definition).

It suffices to construct the skeleton chain $(\theta_m)_{m \in \mathbb{N}_0}$ of Θ. As a guide through the following notation, we have found it useful to occasionally remember that $\Theta(t)$ (and its skeleton chain) is built to the following aim:

Θ *maps a position of an individual in the vector* Y
(the (Ξ, B)*-Moran model) to the level of the corresponding individual in*
the ordered vector X *(the* (Ξ, B)*-lookdown model).*

Notation and ingredients: For $N > 0$ let S_N denote the collection of all permutations of $\{1, \ldots, N\}$, let $P_N = \mathcal{P}(\{1, \ldots, N\})$, the set of all subsets of $\{1, \ldots, N\}$, and let $P_{N,k} \subset P_N$ be the subcollection of subsets with cardinality k. For a set M, $M(i)$ will denote the ith smallest element in M.

At time m (for the skeleton chain) let c_m denote the total number of children. Let a_m be the number of families and c_m^i the number of children born to family i, hence

$$\sum_{i=1}^{a_m} c_m^i = c_m. \tag{2.6}$$

Note that we allow $c_m^i = 0$ for some, but not all i. These are the trivial families where only the parental particle is below level N and all potential children are above. Furthermore, we need to keep track of these "one-member families" in order to match the rates of our model to those of the Ξ-coalescent later on.

Let θ_0 be uniformly distributed over S_N. For each $m \in \mathbb{N}$, pick (independently, and independent of θ_0)

- Φ_m a random set, uniformly chosen from P_{N,c_m+a_m},
- $\left(\phi_m^1, \dots, \phi_m^{a_m}\right)$ a random (ordered) partition of Φ_m, such that each ϕ_m^i has size $c_m^i + 1$,
- σ_m^i, $i = 1, \dots, a_m$ random permutations, each σ_m^i uniformly distributed over $S_{c_m^i+1}$, independently of Φ_m and the ϕ_m^i.

Denote

- $\mu_m^i := \min \phi_m^i$, $i \in \{1, \dots, a_m\}$, and
- write Δ_m for the set of the highest c_m integers from $\{1, \dots, N\} \setminus \bigcup_{i=1}^{a_m} \mu_m^i$.

Proceeding inductively we assume that θ_{m-1} has already been defined. We then construct θ_m as follows: Let

- $\nu_m^i := \theta_{m-1}^{-1}(\mu_m^i)$,
- $\psi_m := \theta_{m-1}^{-1}(\Delta_m)$, and
- a random ordered partition $\left(\psi_m^1, \dots, \psi_m^{a_m}\right)$ of ψ_m such that $|\psi_m^i| = c_m^i$, chosen independently of everything else.

In view of our intended application of θ_m to transfer from the Moran model to the lookdown model, we will later on interpret these quantities as follows: In the m-th event, μ_m^i will be the level of the parental particle of family i in the lookdown-model, and ν_m^i will be the corresponding index in the (unordered) Moran model. Δ_m will specify the levels in the lookdown-model at which individuals die. We do not just pick the highest c_m levels, because we wish to retain parental particles. ψ_m will be the corresponding indices in the Moran model. $\left(\phi_m^1, \dots, \phi_m^{a_m}\right)$ describes the family decomposition (including the respective parents) in this event in the lookdown model, and ψ_m^i are the indices of the children in the i-th family in the Moran model. Thus, θ_m will map ϕ_m^i to $\psi_m^i \cup \{\nu_m^i\}$ (in a particular order).

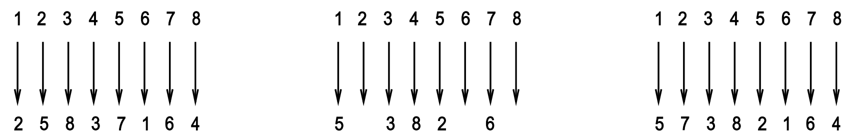
Finally, define θ_m as follows: Put $\Psi_m := \{\nu_m^1, \dots, \nu_m^{a_m}\} \cup \psi_m$. On Ψ_m,

$$\theta_m(\nu_m^i) := \phi_m^i(\sigma_m^i(1)),\ i = 1, \dots, a_m, \tag{2.7}$$

and

$$\theta_m(\psi_m^i(j)) := \phi_m^i(\sigma_m^i(j+1)) \quad \forall j \in \{1, \dots, c_m^i\} \tag{2.8}$$

for each $i \in \{1, \dots, a_m\}$ with $c_m^i \neq 0$. On $\{1, \dots, N\} \setminus \Psi_m$ let θ_m be the mapping onto $\{1, \dots, N\} \setminus \Phi_m$ with the same order as θ_{m-1} restricted to $\{1, \dots, N\} \setminus \Psi_m$, that is, whenever $\theta_{m-1}(i) < \theta_{m-1}(j)$ for some $i, j \in \{1, \dots, N\} \setminus \Psi_m$, then $\theta_m(i) < \theta_m(j)$ should also hold.



(a) Initial permutation θ_{m-1}

(b) The families are added

(c) The completed permutation in Example 2.3

FIGURE 2. The construction of the new permutation from the old permutation carried out in Example 2.3

Example 2.3. We consider a realisation of the m-th event of a population of size $N = 8$, as illustrated in Figure 1. There are $a_m = 2$ families (depicted by "triangle" and "star", respectively). The first family $\phi_m^1 = \{3, 6, 8\}$ has size $c_m^1 + 1 = 3$, the second, $\phi_m^2 = \{2, 5\}$, has size $c_m^2 + 1 = 2$. Hence, the set of levels involved in this birth event is $\Phi_m = \{2, 3, 5, 6, 8\}$, and $\mu_m^1 = 3$, $\mu_m^2 = 2$ are the levels of the parental particles. Since there is no parental particle among the highest three levels, the particles at levels $\Delta_m = \{6, 7, 8\}$ "die".

Now let us assume that θ_{m-1} is as given in Figure 2(a). Thus, $\nu_m^1 = 4$, $\nu_m^2 = 1$, $\psi_m = \{3, 5, 7\}$. The set of indices ψ_m of individuals in the Moran model who will get replaced by offspring in this event is partitioned according to the family sizes, for example let $\psi_m^1 = \{3, 7\}$ and $\psi_m^2 = \{5\}$.

We construct θ_m as follows: Let $\sigma_m^1 = \binom{1\,2\,3}{3\,1\,2}$ and $\sigma_m^2 = \binom{1\,2}{2\,1}$. For the restriction of θ_m to $\Psi_m = \{1, 3, 4, 5, 7\}$, we read from (2.7) that $\theta_m(4) = \phi_m^1(3) = 8$, $\theta_m(1) = \phi_m^2(2) = 5$ and from (2.8) that $\theta_m(3) = \theta_m(\psi_m^1(1)) = \phi_m^1(\sigma_m^1(1+1)) = \phi_m^1(1) = 3$, $\theta_m(7) = \theta_m(\psi_m^1(2)) = \phi_m^1(\sigma_m^1(2+1)) = \phi_m^1(2) = 6$ and $\theta_m(5) = \theta_m(\psi_m^2(1)) = \phi_m^2(\sigma_m^2(1+1)) = \phi_m^2(1) = 2$. This leads to the partial permutation which is given in Figure 2(b).

Restricted to the complementary set $\{2, 6, 8\}$, θ_m is a mapping onto $\{1, 4, 7\}$ with the same order as θ_{m-1} restricted to $\{2, 6, 8\}$. The resulting permutation θ_m is given in Figure 2(c). ■

For notational convenience, let

$$\chi_m := (\nu_m^1, \psi_m^1, \ldots, \nu_m^{a_m}, \psi_m^{a_m}), \tag{2.9}$$

which summarises the combinatorial information generated in the m-th step (namely, the family structure we would observe in the Moran model).

Lemma 2.4. *For each m, $\chi_1, \ldots, \chi_m, \theta_m$ are independent. Furthermore θ_m is uniformly distributed over S_N and*

$$\Upsilon_m := \bigcup_{i=1}^{a_m} \{\nu_m^i\} \cup \psi_m^i \tag{2.10}$$

is uniformly distributed over $P_{N, c_m + a_m}$, and each χ_m is, given Υ_m, uniformly distributed on all ordered partitions of Υ_m with family sizes consistent with the c_m^i.

Proof: We prove the statement by induction. Denoting $\mathcal{F}_m = \sigma(\theta_k, \chi_k : 0 \le k \le m)$, we have

$$\mathbb{E}[f(\theta_m, \chi_m) \mid \mathcal{F}_{m-1}] = \mathbb{E}[f(\theta_m, \chi_m) \mid \theta_{m-1}], \tag{2.11}$$

since θ_m and χ_m are only based on θ_{m-1} and additional independent random structure. This implies, for any choice of $h_k\colon \cup_{n=1}^N \big(\{1,\dots,N\} \times \mathcal{P}(\{1,\dots,N\})\big)^n \to \mathbb{R}$ and $f\colon S_n \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}\left[f(\theta_m)\prod_{k=1}^m h_k(\chi_k)\right] &= \mathbb{E}\left[\mathbb{E}[f(\theta_m)h_m(\chi_m) \mid \mathcal{F}_{m-1}]\prod_{k=1}^{m-1} h_k(\chi_k)\right]\\ &= \mathbb{E}\left[\mathbb{E}[f(\theta_m)h_m(\chi_m) \mid \theta_{m-1}]\prod_{k=1}^{m-1} h_k(\chi_k)\right]\\ &= \mathbb{E}[f(\theta_m)h_m(\chi_m)]\prod_{k=1}^{m-1} \mathbb{E}[h_k(\chi_k)],\end{aligned}$$

where we used (2.11) in the second and the induction hypothesis in the third equality. It remains to show that θ_m and χ_m are independent and have the correct distributions.

θ_{m-1} is uniformly distributed by the induction hypothesis and independent of the distributions of the parental-levels μ_m^i and the "death-levels" Δ_m by construction. It is immediate from the construction that Φ_m and Υ_m are uniformly distributed over P_{N,c_m+a_m} and the family structure χ_m is uniformly distributed among all admissible configurations.

Furthermore, conditioning on χ_m and Φ_m, θ_m is uniformly distributed over all permutations that map Υ_m onto Φ_m. This follows from the fact that Φ_m is uniform on P_{N,c_m+a_m} and that this set is uniformly divided into the families ϕ_m^i. Since uniform and independent permutations σ_m^i are used for the construction of θ_m and the non-participating levels remain uniformly distributed, θ_m is uniform under these conditions.

Finally, conditioning on χ_m does not alter the fact that Φ_m is uniformly distributed over P_{N,c_m+a_m}. This implies that given χ_m, θ_m is also uniformly distributed over S_N. Since

$$\mathcal{L}(\theta_m|\chi_m) = \mathrm{unif}(S_N) = \mathcal{L}(\theta_m), \tag{2.12}$$

θ_m and χ_m are independent of each other. □

Proof of Theorem 2.2: Suppose a realization X of the N-particle (Ξ, B)-lookdown-model is given and let $\{t_m\}$ denote the times at which the birth events occur. The families involved in the m-th birth event are denoted by ϕ_m^i. Note that by definition of the lookdown-dynamics, the "ingredients" $\Phi_m, c_m, a_m, c_m^i, \mu_m^i, \Delta_m$ introduced earlier can be obtained from this, and that their joint distributions is as discussed above.

Moreover, let the initial permutation θ_0 be independent of X and uniformly distributed on S_N. Let σ_m^i be independent of all other random variables and uniformly distributed on $S_{c_m^i+1}$, $1 \le i \le a_m$, $m \in \mathbb{N}$.

Define θ_m as above, and

$$\Theta(t) := \theta_m \quad \text{for } t_m \le t < t_{m+1}. \tag{2.13}$$

Observe that, by Lemma 2.4,

$$(Y_1(t), \dots, Y_N(t)) := (X_{\Theta_1(t)}(t), \dots, X_{\Theta_N(t)}(t)) \tag{2.14}$$

is a version of the (Ξ, B)-Moran-model. Note that "one-member families" are in this construction simply treated as non-participating individuals in the (Ξ, B)-Moran model.

$Y(t)$ depends only on $Y(0)$, $\{\chi_m\}_{t_m \leq t}$ and the evolution of the type processes between birth and death events, so $\Theta(t)$, and hence $\Theta(t)^{-1}$ is independent of

$$\mathcal{G}_t := \sigma\big((Y_1(s), \dots, Y_N(s)) : s \leq t\big) \vee \sigma(\chi_m : m \in \mathbb{N}) \tag{2.15}$$

due to Lemma 2.4. Therefore, we see from

$$(X_1(t), \dots, X_N(t)) = (Y_{\Theta_1^{-1}(t)}(t), \dots, Y_{\Theta_N^{-1}(t)}(t)) \tag{2.16}$$

that $(X_1(t), \dots, X_N(t))$ is exchangeable. □

Corollary 2.5. *Starting from the same exchangeable initial condition, the laws of the empirical processes of the (Ξ, B)-Moran-model and the (Ξ, B)-lookdown-model coincide.*

The exchangeability property does not only hold for fixed times, but also for stopping times.

Theorem 2.6. *Suppose that the initial population vectors $Y^N(0)$ in the (Ξ, B)-Moran-model and $X^N(0)$ in the (Ξ, B)-lookdown-model have the same exchangeable distribution, and let τ be a stopping time with respect to $(\mathcal{G}_t)_{t \geq 0}$ given by (2.15). Then, $(X_1^N(\tau), \dots, X_N^N(\tau))$ is exchangeable.*

Proof: We show that $\Theta(\tau)$ is independent of the σ-algebra $\mathcal{G}_\tau$ (the τ-past) and uniformly distributed over S_N.

First, assume that τ takes only countable many values t_k, $k \in \mathbb{N}$. Let $A \in \mathcal{G}_\tau$ and $h : S_N \to \mathbb{R}_+$, then

$$\begin{aligned} \mathbb{E}\Big(h\big(\Theta(\tau)\big)\mathbb{1}_A\Big) &= \mathbb{E}\Big(\sum_{k=1}^\infty h\big(\Theta(t_k)\big)\mathbb{1}_{A \cap \{\tau = t_k\}}\Big) \\ &= \sum_{k=1}^\infty \Big(\mathbb{E}h\big(\Theta(t_k)\big)\Big)\Big(\mathbb{E}\mathbb{1}_{A \cap \{\tau = t_k\}}\Big) \\ &= \int h(\Theta)\,\mathfrak{U}(d\Theta) \sum_{k=1}^\infty \mathbb{E}\mathbb{1}_{A \cap \{\tau = t_k\}} \\ &= \int h(\Theta)\,\mathfrak{U}(d\Theta)\,\mathbb{E}\mathbb{1}_A, \end{aligned} \tag{2.17}$$

where $\mathfrak{U}$ denotes the uniform distribution on S_N. To see that the second equality holds, observe that, for fixed t_k, $\Theta(t_k)$ is independent of $\mathcal{G}_{t_k}$ as defined in (2.15).

By approximating an arbitrary stopping time from above by a sequence of discrete stopping times, we see that (2.17) holds in the general case as well. Now, exchangeability of $(X_1^N(\tau), \dots, X_N^N(\tau))$ follows as in the proof of Theorem 2.2. □

Remark 2.7. One can also define a variant of the (Ξ, B)-lookdown model which is more in the spirit of the 'classical' lookdown construction from Donnelly and Kurtz (1996), where, instead of a)–c) on page 33, at a jump time each particle simply copies the type of that member of the family it belongs to with the lowest level (and no types get shifted upwards). This variant, which is (up to a renaming of levels by the points of a Poisson process on $\mathbb{R}$) also the one suggested by adapting Kurtz and Rodrigues (2008) to the 'simultaneous multiple merger'-scenario, has been considered by Taylor & Véber (2008, personal communication). The same results as above hold for this variant, with only minor modifications of the proofs. Note that the flavour of the lookdown process described above is easily adaptable to a set-up with time-varying total population size, which is not obvious for the other variant.

2.3. *The limiting population.* We now construct the limiting E^∞-valued particle system $X = (X_1, X_2, \ldots)$ by formulating a stochastic differential equation for each level l. These exist for each level and are well defined, since the equation for level l needs only information about lower levels.

The generator (1.6) of a pure jump process can be written in the form

$$Bf(x) = r \int_0^1 \big(f(m(x,u)) - f(x)\big)\, du,$$

where r is the global mutation rate and $m\colon E \times [0,1] \to E$ transforms a uniformly distributed random variable on $[0,1]$ into the jump distribution $q(x,dy)$ of the process. The random times and uniform "coins" for the mutation process at each level l are given by a Poisson point process $\mathfrak{N}_l^{\mathrm{Mut}}$ on $\mathbb{R}_+ \times [0,1]$ with intensity measure $rdt \otimes du$.

As in Section 2.1, denote by

$$(t_m, \boldsymbol{\zeta}_m, \mathbf{u}_m) = (t_m, (\zeta_{m1}, \zeta_{m2}, \ldots), (u_{m1}, u_{m2}, \ldots))$$

the points of the Poisson point process $\mathfrak{M}^{\Xi_0}$ and recall the definition (2.2) of the "colour" function g. Based on this, define

$$L_J^l(t) \;:=\; \sum_{m: t_m \le t} \prod_{j \in J} \mathbb{1}_{\{g(\boldsymbol{\zeta}_m, u_{mj}) < \infty\}} \prod_{j \in \{1,\ldots,l\} \setminus J} \mathbb{1}_{\{g(\boldsymbol{\zeta}_m, u_{mj}) = \infty\}}, \tag{2.18}$$

for $J \subset \{1,\ldots,l\}$ with $|J| \ge 2$. $L_J^l(t)$ counts how many times, among the levels in $\{1,\ldots,l\}$, exactly those in J were involved in a birth event up to time t. Moreover, let

$$L_{J,k}^l(t) \;:=\; \sum_{m: t_m \le t} \prod_{j \in J} \mathbb{1}_{\{g(\boldsymbol{\zeta}_m, u_{mj}) = k\}} \prod_{j \in \{1,\ldots,l\} \setminus J} \mathbb{1}_{\{g(\boldsymbol{\zeta}_m, u_{mj}) \ne k\}}. \tag{2.19}$$

$L_{J,k}^l(t)$ counts how many times, among the levels in $\{1,\ldots,l\}$, exactly those in J were involved in a birth event up to time t and additionally assumed "colour" k.

To specify the new levels of the individuals not participating in a certain birth event, we construct a function J_m as follows:

Denote by $\mu_m^k := \min\{l \in \mathbb{N} \,|\, g(\boldsymbol{\zeta}_m, u_{ml}) = k\}$ the level of the parental particle of family number k and by $M_m := \{\mu_m^k\}_{k \in \mathbb{N}}$ the set of all levels of parental particles involved in the m-th birth event. Furthermore $U_m := \{l \in \mathbb{N} \,|\, g(\boldsymbol{\zeta}_m, u_{ml}) = \infty\}$ denotes the set of the levels not participating in the birth event m. Define the mapping

$$J_m \;:\; U_m \to \mathbb{N} \setminus M_m \tag{2.20}$$

that maps the i-th smallest element of the set U_m to the i-th smallest element of the set $\mathbb{N} \setminus M_m$ for all i.

Assuming for the moment that E is an Abelian group, the (infinite) vector describing the types in the (Ξ, B)-lookdown-model is defined as the (unique) strong solution of the following system of stochastic differential equations. The lowest individual on level 1 just evolves according to mutation, i.e.,

$$X_1(t) \ := \ \int_{[0,t]\times[0,1]} (m(X_1(s-),u) - X_1(s-))\, d\mathfrak{N}_1^{\mathrm{Mut}}(s,u). \tag{2.21}$$

The individuals above level one can look down during birth events. Thus, for $l \geq 2$, define

$$\begin{aligned} X_l(t) :=& X_l(0) + \int_{[0,t]\times[0,1]} \big(m(X_l(s-),u) - X_l(s-)\big)\, d\mathfrak{N}_l^{\mathrm{Mut}}(s,u) \\ &+ \sum_{1\leq i<l} \int_0^t (X_i(s-) - X_l(s-))\, d\mathfrak{N}_{il}^K(s) \\ &+ \sum_{1\leq i<j<l} \int_0^t (X_{l-1}(s-) - X_l(s-))\, d\mathfrak{N}_{ij}^K(s) \\ &+ \sum_{k\in\mathbb{N}} \sum_{K\subset\{1,\ldots,l\},l\in K} \int_0^t (X_{\min(K)}(s-) - X_l(s-))\, dL_{K,k}^l(s) \\ &+ \sum_{K\subset\{1,\ldots,l\},l\notin K} \int_0^t (X_{J_m(l)}(s-) - X_l(s-))\, dL_K^l(s). \end{aligned} \tag{2.22}$$

The second and third lines describe the “Kingman events”, where only pairs of individuals are involved. The first part copies the type from level i when l looks down to this level, because it is involved in a birth event and the parental particle is at level i. The second part handles the event that the parental particle places a child on a level below l. In this case, l has to copy the type from the level $l-1$, since the new individual is inserted at some level below l and pushes all particles above that level one level up.

The fourth and fifth lines describe the change of types for a birth event with large families in a similar way. If the particle at level l is involved in the family k, it copies the type from the parental particle which resides at the lowest level of the family. If level l is not involved in any family, then $J_m(l)$ $(\leq l)$ gives the level from where the type is copied (which comes from shifting particles not involved in the lookdown event upwards).

Since the equation for X_l involves only $X_1, \ldots, X_l$ and finitely many Poisson processes, it is immediate that there exists a unique strong solution of (2.21)–(2.22).

In the case where E has no group structure, one may still construct suitable jump-hold processes X_i, using the driving Poisson processes in an obvious extension of (2.21)–(2.22).

These stochastic differential equations determine an infinitely large population vector

$$X(t) \ := \ (X_1(t), X_2(t), \ldots) \tag{2.23}$$

in a consistent way, and for each $N \in \mathbb{N}$, the dynamics of $(X_1, \ldots, X_N)$ is identical to that defined in Section 2.2. In particular, we see from Theorem 2.2 that, for each $t \geq 0$, $X(t)$ is exchangeable and the empirical distribution

$$Z(t) \ := \ \lim_{l\to\infty} Z^l(t) \ := \ \lim_{l\to\infty} \frac{1}{l} \sum_{i=1}^{l} \delta_{X_i(t)} \tag{2.24}$$

exists almost surely. Let F be the set of bounded measurable functions $\varphi : [0,\infty) \times [0,1]^{\mathbb{N}} \times [0,1]^{\infty} \to \mathbb{R}$ such that $\varphi(t, \boldsymbol{\zeta}, \mathbf{u})$ does not depend on $\mathbf{u}$, and put

$$\mathcal{H}_t \;:=\; \sigma\bigg(\big(Z(s) : s \le t\big), \Big(\int \varphi \, d\mathfrak{M}^{\Xi_0} : \varphi \in F\Big)\bigg). \tag{2.25}$$

Corollary 2.8. *Let τ be a stopping time with respect to $(\mathcal{H}_t)_{t\ge 0}$. Then*

$$X(\tau) \;=\; (X_1(\tau), X_2(\tau), \ldots) \tag{2.26}$$

is exchangeable.

Proof: We claim that for $t \ge 0$, $A \in \mathcal{H}_t$ with $\mathbb{P}\{A\} > 0$ and $n \in \mathbb{N}$,

$$(X_1(t), \ldots, X_n(t)) \text{ is exchangeable under } \mathbb{P}\{\cdot|A\}. \tag{2.27}$$

Observe that, taking $A = \{\tau = t_k\}$, (2.27) immediately implies the result for discrete stopping times τ, from which the general case can be deduced by approximation as in the proof of Theorem 2.6.

Obviously, (2.27) is equivalent to

$$\mathbb{P}\big\{A \cap \{(X_1(t), \ldots, X_n(t)) \in C\}\big\} \;=\; \mathbb{P}\big\{A \cap \{(X_{\sigma(1)}(t), \ldots, X_{\sigma(n)}(t)) \in C\}\big\} \quad \forall\, C \subset E^n, \sigma \in S_n. \tag{2.28}$$

As the collection of sets A from $\mathcal{H}_t$ satisfying (2.28) is a Dynkin system, it suffices to verify (2.28) for events of the form

$$A \;=\; \{Z(s_1) \in B_1, \ldots, Z(s_k) \in B_k\} \cap H', \tag{2.29}$$

where $H' \in \sigma\big(\int \varphi \, d\mathfrak{M}^{\Xi_0} : \varphi \in F\big)$, $k \in \mathbb{N}$, $s_1 < \cdots < s_k \le t$, $B_i \in \mathcal{B}(s_i)$ for $i \in \{1, \ldots, k\}$, and $\mathcal{B}(s_i)$ is a $\cap$-stable generator of $\mathcal{B}_{\mathcal{M}_1(E)}$ with the property that $\mathbb{P}\{Z(s_i) \in \partial B'\} = 0$ for all $B' \in \mathcal{B}(s_i)$.

For A as given in (2.29), $\varepsilon > 0$ and $n \in \mathbb{N}$, $\sigma \in S_n$, $C \subset E^n$ appearing in (2.28), by (2.24) there exists l ($l \gg n$) such that

$$A_l \;:=\; \{Z^l(s_1) \in B_1, \ldots, Z^l(s_k) \in B_k\} \cap H'$$

satisfies $\mathbb{P}\{(A \setminus A_l) \cup (A_l \setminus A)\} \le \varepsilon$. By the arguments given in the proof of Theorem 2.6, (2.28) holds with A replaced by A_l. Finally, take $\varepsilon \to 0$ to conclude. □

3. Pathwise convergence: Proof of Theorem 1.1

Recall the empirical processes Z^l, and their limit Z, from (2.24). Obviously, for each $l \in \mathbb{N}$, the process $(Z^l(t))_{t\ge 0}$ has càdlàg paths. To verify the corresponding property for Z, we introduce the following auxiliary (Lévy) process U, derived from the Poisson point process $\mathfrak{M}^{\Xi_0}$ which governs the large family birth events of the population X: If $\big\{(t_m, \boldsymbol{\zeta}_m, \mathbf{u}_m)\big\}$ are the points of the process $\mathfrak{M}^{\Xi_0}$, we define

$$U(t) \;:=\; \sum_{t_m \le t} v_m^2, \tag{3.1}$$

where $v_m := \sum_{i=1}^{\infty} \zeta_{mi}$. The jumps of $U := (U(t))_{t\ge 0}$ are the squared total fractions of the population which are replaced in large birth events. The generator of U is given by

$$Df(u) \;=\; \int_0^1 \big(f(u + v^2) - f(u)\big)\, \nu(dv), \tag{3.2}$$

where the measure ν on $[0,1]$, defined via

$$\nu(A) \;:=\; \int_\Delta \mathbb{1}_{\{\sum_{i=1}^\infty \zeta_i \in A\}} \frac{\Xi(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta},\boldsymbol{\zeta})}, \tag{3.3}$$

governs the jumps.

We need the following version of Lemma A.2 from Donnelly and Kurtz (1999).

Lemma 3.1. *a) Let $e_1, e_2, \ldots$ be exchangeable and suppose there exists a constant K such that $|e_i| \le K$ almost surely. Define*

$$M_k \;:=\; \frac{1}{k}\sum_{i=1}^k e_i \tag{3.4}$$

and let M_∞ be the almost sure limit of $(M_k)_{k\in\mathbb{N}}$, whose existence is guaranteed by the de Finetti Theorem. Let $\varepsilon > 0$. Then there exists $\eta_1 > 0$ depending only on K and ε, such that, for $l < n \in \mathbb{N}\cup\{\infty\}$,

$$\mathbb{P}\{|M_n - M_l| \ge \varepsilon\} \le 2e^{-\eta_1(K,\varepsilon)l}. \tag{3.5}$$

b) Let $(e_i(t))_{t\in[0,1]}$ be centered martingales such that $\sup_{i\in\mathbb{N},t\in[0,1]}|e_i(t)| \le K$ almost surely and $(e_1(1), e_2(1), \ldots)$ is exchangeable. Put

$$M_k(t) \;:=\; \frac{1}{k}\sum_{i=1}^k e_i(t).$$

Let $\varepsilon > 0$. Then there exists $\eta_2 > 0$ depending only on K and ε, such that, for $l \in \mathbb{N}$

$$\mathbb{P}\{\sup_{t\in[0,1]} |M_k(t)| \ge \varepsilon\} \le 2e^{-\eta_2(K,\varepsilon)l}. \tag{3.6}$$

Proof: The proof of part a) is a straightforward extension of that of Lemma A.2 from Donnelly and Kurtz (1999), which employs the fact that an infinite exchangeable sequence is conditionally i.i.d. together with standard arguments based on the moment generating function.

For part b) observe that by Doob's submartingale inequality,

$$\mathbb{P}\Big\{\sup_{0\le t<1} |M_k(t)| \ge \varepsilon\Big\} \;\le\; \inf_{\lambda>0} \frac{1}{e^{\varepsilon\lambda}}\mathbb{E}e^{\lambda|M_k(1)|} \;\le\; \inf_{\lambda>0}\frac{1}{e^{\varepsilon\lambda}}\mathbb{E}\exp\Big(\frac{\lambda}{k}\sum_{i=1}^k |e_i(1)|\Big). \tag{3.7}$$

Now proceed as in part a). □

The following lemma provides the technical core of the argument and replaces Lemma 3.4 and Lemma 3.5 in Donnelly and Kurtz (1999). The proof given below follows closely the arguments of Donnelly and Kurtz (1999).

Lemma 3.2. *In the setting of Theorem 1.1, for all $c, T, \epsilon > 0$ and $f \in \mathcal{D}(B)$ (the domain of the mutation generator) there exists a sequence δ_l such that $\sum_{l=1}^\infty \delta_l < \infty$ and*

$$\mathbb{P}\Big\{\sup_{0\le t\le T} \big|\langle f, Z(t)\rangle - \langle f, Z^l(t)\rangle\big| \ge 11\epsilon, U(T) \le c\Big\} \le \delta_l. \tag{3.8}$$

Proof: By Lemma 3.1 and the exchangeability properties of X, we have

$$\mathbb{P}\{|\langle f, Z(\alpha)\rangle - \langle f, Z^l(\alpha)\rangle| \geq \epsilon\} \ \leq \ 2e^{-\eta l}, \tag{3.9}$$

if α is a stopping time with respect to $\widetilde{\mathcal{H}} := (\widetilde{\mathcal{H}}_t)_{t\geq 0} := \big(\sigma(U(s) : s \geq 0) \vee \sigma(Z(s) : 0 \leq s \leq t)\big)_{t\geq 0}$ (observe that $\widetilde{\mathcal{H}}_t \subset \mathcal{H}_t$, where $\mathcal{H}_t$ is defined in (2.25)).

Now fix l and ϵ. Define the $\widetilde{\mathcal{H}}$-stopping times

$$\alpha_1 \ := \ \inf\left\{t : U(t) > \frac{1}{l^4}\right\} \wedge \frac{1}{l^4} \tag{3.10}$$

and

$$\alpha_{o+1} \ := \ \inf\left\{t : U(t) > U(\alpha_o) + \frac{1}{l^4}\right\} \wedge \left(\alpha_o + \frac{1}{l^4}\right), \quad o = 1, 2, \ldots, \tag{3.11}$$

which yield a decomposition of the interval $[0, T]$. Remark that on the event $\big\{U(T) \leq c\big\}$ there exist at most

$$o_l := 2(c+T)l^4 \tag{3.12}$$

such α_o, i.e., we have

$$\mathbb{P}\big\{\alpha_{o_l} < T, U(\alpha_{o_l}) < c, U(T) \leq c\big\} = 0. \tag{3.13}$$

We define a second kind of $\widetilde{\mathcal{H}}$-stopping times depending on α_o via

$$\tilde{\alpha}_o := \inf\{t > \alpha_o : |\langle f, Z(t)\rangle - \langle f, Z(\alpha_o)\rangle| \geq 6\epsilon\}. \tag{3.14}$$

We see from (3.9) that

$$H_o := |\langle f, Z(\alpha_o)\rangle - \langle f, Z^l(\alpha_o)\rangle| \vee |\langle f, Z(\tilde{\alpha}_o)\rangle - \langle f, Z^l(\tilde{\alpha}_o)\rangle| \tag{3.15}$$

satisfies

$$\mathbb{P}\Big\{\sup_{o\leq o_l} H_o \geq \varepsilon, U(T) \leq c\Big\} \leq \sum_{o=1}^{o_l} \mathbb{P}\left\{H_o \geq \varepsilon, U(T) \leq c\right\} \leq 8(c+T)l^4 e^{-\eta l}. \tag{3.16}$$

It remains to estimate the variation of Z^l and Z in between the stopping times α_o. For $u \in [\alpha_o, \alpha_{o+1})$ let $\beta_{jo}(u)$ denote the smallest index of a descendant of $X_j(\alpha_o)$, let the stopping time γ_{jo} be the time when the smallest descendant of $X_j(\alpha_o)$ is shifted above the level l. Put

$$\tilde{X}_j(u) = \begin{cases} X_{\beta_{jo}(u)}(u) & \text{if } u < \gamma_{jo}, \\ X_{\beta_{jo}(\gamma_{jo}-)}(\gamma_{jo}-) & \text{if } u \geq \gamma_{jo}. \end{cases}$$

Observe that

$$\langle f, Z^l(u)\rangle - \langle f, Z^l(\alpha_o)\rangle = \langle f, Z^l(u)\rangle - \frac{1}{l}\sum_{j=1}^{l} f(\tilde{X}_j(u)) + \frac{1}{l}\sum_{j=1}^{l}\Big(f(\tilde{X}_j(u)) - f(\tilde{X}_j(\alpha_o))\Big). \tag{3.17}$$

It will be useful to treat the two parts of the sum separately. Define

$$K_1 := \max_{o\leq o_l} \sup_{u\in[\alpha_o,\alpha_{o+1})} \left|\langle f, Z^l(u)\rangle - \frac{1}{l}\sum_{j=1}^{l} f(\tilde{X}_j(u))\right|$$

and

$$K_2 := \max_{o\leq o_l} \sup_{u\in[\alpha_o,\alpha_{o+1})} \left|\frac{1}{l}\sum_{j=1}^{l}\big(f(\tilde{X}_j(u)) - f(\tilde{X}_j(\alpha_o))\big)\right|.$$

Note that the law of K_2 depends only on the mutation mechanism, since $\tilde{X}_j(u)$ follows the line of the individual $\tilde{X}_j(\alpha_o) = X_j(\alpha_o)$ and thus only evolves independently according to a mutation process with generator B.

Begin with K_1 and note that, for $u \in [\alpha_o, \alpha_{o+1})$,

$$\langle f, Z^l(u)\rangle - \frac{1}{l}\sum_{j=1}^{l} f(\tilde{X}_j(u)) = \frac{1}{l}\Big(\sum_{j=1}^{l} f(X_j(u)) - \sum_{j=1}^{l} f(\tilde{X}_j(u))\Big) \le \frac{2\|f\|}{l} N^l[\alpha_o, \alpha_{o+1}), \tag{3.18}$$

where $N^l[\alpha_o, \alpha_{o+1})$ is the total number of births occurring in the time interval $[\alpha_o, \alpha_{o+1})$ with index less than or equal to l. To see this note that at time α_o the two sums in the second expression cancel. A birth event in the interval $[\alpha_o, \alpha_{o+1})$ means that one type is removed from the second sum and another one is added, thus the expression can be altered by up to $2||f||/l$.

There are two mechanisms which can increase $N^l[\alpha_o, \alpha_{o+1})$. It can either increase during a large birth event given by a "jump" of $\mathfrak{M}^{\Xi_0}$ or during a small birth event which is given by one of the "Kingman-related" Poisson processes $\mathfrak{N}_{ij}^K$.

We first consider large birth events. Let (v_i) be the jumps of U in the interval $[\alpha_o, \alpha_{o+1})$, and condition on this configuration for the rest of this paragraph. At the time of the m-th jump, a Binomial(l, v_m)-distributed number of levels $\le l$ participates in this event, hence k_m, the total number of children below level l in the m-th birth event, satisfies

$$k_m \le (b_m - 1)_+,$$

where b_m is Binomial(l, v_m)-distributed. Note that we can subtract 1 from the binomial random variable, since at least one of the levels participating in the birth event must be a mother. This subtraction will be crucial later on.

By elementary calculations with Binomial distributions, involving fourth moments, similar to Donnelly and Kurtz (1999, p. 186), we can estimate

$$\mathbb{P}\Big\{\sum_m k_m > \epsilon l\Big\} \le \mathbb{P}\Big\{\sum_m (b_m - 1)_+ > \epsilon l\Big\} \le \frac{C_1}{l^6} \tag{3.19}$$

for some $0 < C_1 < \infty$.

As we mentioned before, $N^l[\alpha_o, \alpha_{o+1})$ and thus K_1 can also be increased by the Kingman part of the birth process, but only if the parental particle and its offspring are placed below level l. The number of times this happens in the interval $[\alpha_o, \alpha_{o+1})$ is stochastically dominated by a Poisson distributed random variable R with parameter $\binom{l}{2} l^{-4}$ since the length of the interval is bounded by l^{-4}. So, the probability that $\frac{2\|f\|}{l} N^l[\alpha_o, \alpha_{o+1})$ exceeds 2ϵ due to this mechanism is bounded by the probability that R exceeds $\frac{l\epsilon}{\|f\|}$. By elementary estimates on the tails of Poisson random variables, we have

$$\mathbb{P}\Big\{R > \frac{l\epsilon}{\|f\|}\Big\} \le e^{-\eta_1 l}, \tag{3.20}$$

for some $\eta_1 > 0$ and l large enough.

Combining (3.19) and (3.20), we obtain

$$\begin{aligned}\mathbb{P}\Big\{K_1 > 2\epsilon, U(T) \le c\Big\} =& \mathbb{P}\Big\{\max_{o\le o_l}\sup_{u\in[\alpha_o,\alpha_{o+1})}\Big|\langle f, Z^l(u)\rangle - \frac{1}{l}\sum_{j=1}^{l} f(\tilde{X}_j(u))\Big| > 2\epsilon, U(T)\le c\Big\}\\ &\le o_l\Big(\frac{C_1}{l^6} + e^{-\eta_1 l}\Big),\end{aligned} \tag{3.21}$$

for l large enough. This controls the increments of $\langle f, Z^l\rangle$ in the intervals $[\alpha_o, \alpha_{o+1})$.

We now consider K_2. Observe that

$$\begin{aligned}\frac{1}{l}\sum_{j=1}^{l}(f(\tilde{X}_j(u)) - f(\tilde{X}_j(\alpha_o))) =& \frac{1}{l}\sum_{j=1}^{l}\Big(f(\tilde{X}_j(u)) - f(\tilde{X}_j(\alpha_o)) - \int_{\alpha_o}^{u} Bf(\tilde{X}_j(s))ds\Big)\\ &+ \frac{1}{l}\sum_{j=1}^{l}\int_{\alpha_o}^{u} Bf(\tilde{X}_j(s))ds,\end{aligned} \tag{3.22}$$

and that, for $u \ge \alpha_o$ and each o,

$$M_{lo}(u\wedge\alpha_{o+1}) := \frac{1}{l}\sum_{j=1}^{l}\Big(f(\tilde{X}_j(u\wedge\alpha_{o+1})) - f(\tilde{X}_j(\alpha_o)) - \int_{\alpha_o}^{u\wedge\alpha_{o+1}} Bf(\tilde{X}_j(s))ds\Big) \tag{3.23}$$

is a martingale. For l large enough so that $l^{-4}\|Bf\| \le \varepsilon$, we have

$$\begin{aligned}\mathbb{P}\Big\{K_2 \ge 2\varepsilon, U(T)\le c\Big\} \le& \sum_{o=0}^{o_l-1}\mathbb{P}\Big\{\sup_{\alpha_o\le u<\alpha_{o+1}}|M_{lo}(u)\\ &+ \frac{1}{l}\sum_{j=1}^{l}\int_{\alpha_o}^{u} Bf(\tilde{X}_j(s))ds| \ge 2\varepsilon, U(T)\le c\Big\}\\ \le& \sum_{o=0}^{o_l-1}\mathbb{P}\Big\{\sup_{\alpha_o\le u<\alpha_{o+1}}|M_{lo}(u)| + l^{-4}\|Bf\| \ge 2\varepsilon, U(T)\le c\Big\}\\ \le& \sum_{o=0}^{o_l-1}\mathbb{P}\Big\{\sup_{\alpha_o\le u<\alpha_{o+1}}|M_{lo}(u)| \ge \varepsilon, U(T)\le c\Big\}.\end{aligned} \tag{3.24}$$

We now need to bound each summand. Using the notation

$$M_{lo}(u) = \frac{1}{l}\sum_{j=1}^{l} e_j(u),$$

where

$$e_j(u) := f(\tilde{X}_j(\alpha_{o+1}\wedge u)) - f(\tilde{X}_j(\alpha_o)) - \int_{\alpha_o}^{\alpha_{o+1}\wedge u} Bf(\tilde{X}_j(s))ds, \quad u\in[0,1], \tag{3.25}$$

each $(e_j(u))_u$ is a martingale with $\mathbb{E}e_j(u) = 0$ and $|e_j(u)| \le 2\|f\| + \|Bf\|/l^4 =: K$ almost surely. Moreover, the $e_j(u)$ are exchangeable. We obtain from Lemma 3.1

$$\mathbb{P}\Big\{\sup_{\alpha_o\le u<\alpha_{o+1}}|M_{lo}(u)| \ge \varepsilon\Big\} \le 2e^{-\eta_2 l}, \tag{3.26}$$

for some $\eta_2 > 0$.

Combining this result with (3.24), we arrive at

$$\mathbb{P}\Big\{K_2 \ge 2\varepsilon, U(T)\le c\Big\} \le o_l C_2 e^{-\eta_2 l}. \tag{3.27}$$

Now observe that if $\max_{o\le o_l} H_o < \epsilon$, $K_1 < 2\epsilon$ and $K_2 < 2\epsilon$, then $\tilde{\alpha}_o \ge \alpha_{o+1}$. This can easily be seen by contradiction. Indeed, if we assume that $\tilde{\alpha}_o < \alpha_{o+1}$, this would imply

$$|\langle f, Z(\alpha_o)\rangle - \langle f, Z(\tilde{\alpha}_o)\rangle| \ge 6\epsilon, \tag{3.28}$$

according to (3.14). But on the other hand we know that

$$|\langle f, Z(\alpha_o)\rangle - \langle f, Z^l(\alpha_o)\rangle| < \epsilon \text{ and } |\langle f, Z(\tilde{\alpha}_o)\rangle - \langle f, Z^l(\tilde{\alpha}_o)\rangle| < \epsilon \quad \forall o \tag{3.29}$$

due to our bound on H_o. Since the distance between $\langle f, Z\rangle$ and $\langle f, Z^l\rangle$ was at most ϵ at the beginning of the interval and $\langle f, Z^l\rangle$ can only have moved by at most 4ϵ on the event $\{K_1 \le 2\epsilon\} \cap \{K_2 \le 2\epsilon\} \cap \{\max_{o\le o_l} H_o \le \epsilon\}$,

$$|\langle f, Z(\alpha_o)\rangle - \langle f, Z^l(\tilde{\alpha}_o)\rangle| < 5\epsilon \tag{3.30}$$

must hold if $\tilde{\alpha}_o \le \alpha_{o+1}$. But equation (3.28) states that $\langle f, Z(\tilde{\alpha}_o)\rangle$ is more than 6ϵ away from its starting point, so this contradicts that it can only be ϵ away from $\langle f, Z^l(\tilde{\alpha}_o)\rangle$ which is ensured by our condition on H_o. Thus $\tilde{\alpha}_o$ has to be greater than α_{o+1} which in turn implies that

$$\sup_{\alpha_o \le u < \alpha_{o+1}} \Big\{ \big| \langle f, Z(u)\rangle - \langle f, Z(\alpha_o)\rangle \big| \Big\} \le 6\epsilon \tag{3.31}$$

holds on the event $\{K_1 \le 2\epsilon\} \cap \{K_2 \le 2\epsilon\} \cap \{\max_{o\le o_l} H_o \le \epsilon\}$.

Putting observations (3.16) and (3.31), the bound (3.27) and the bound (3.21) together, we finally obtain

$$\mathbb{P}\Big\{ \sup_{0\le t\le T} \big|\langle f, Z(t)\rangle - \langle f, Z^l(t)\rangle\big| \ge 11\epsilon, U(T) \le c \Big\} \le \delta_l \tag{3.32}$$

with

$$\delta_l := 8(c+T)l^4 e^{-\eta l} + o_l C_1 l^{-6} + o_l e^{-\eta_1 l} + o_l C_2 e^{-\eta_2 l}, \tag{3.33}$$

which is the statement of the lemma since due to equation (3.12) $o_l \sim l^4$ holds and therefore the δ_l are summable. □

Proof of Theorem 1.1: Almost sure convergence of Z^l to Z with respect to the metric (1.8) follows directly from Lemma 3.2 and the Borel-Cantelli Lemma, completing the proof of Theorem 1.1. □

4. The Hille-Yosida approach

In this section we provide two alternative representations of the Ξ_0-Fleming-Viot generator, leading to the distributional duality to the Ξ-coalescent discussed in Section 5, and we show that they generate a Markov semigroup on $\mathcal{M}_1(E)$, hence leading to a classical construction of the Ξ_0-Fleming-Viot process as a Markov process.

4.1. *Two representations of the Ξ_0-Fleming-Viot generator.* Recall that if the type space E is a compact Polish space (which is assumed in this paper), then the set $\mathcal{M}_1(E)$ of all probability measures on E, equipped with the weak topology, is again a Polish space. We briefly recall the notation from Section 1. For $f : E^n \to \mathbb{R}$ bounded and measurable consider the test function

$$G_f(\mu) := \int_{E^n} f(x_1, \ldots, x_n)\, \mu^{\otimes n}(dx_1, \ldots, dx_n), \quad \mu \in \mathcal{M}_1(E). \tag{4.1}$$

The linear operator L^{Ξ_0} was defined via

$$L^{\Xi_0}G_f(\mu) = \int_\Delta \int_{E^{\mathbb{N}}} \left[G_f\big((1-|\boldsymbol{\zeta}|)\mu + \textstyle\sum_{i=1}^\infty \zeta_i\delta_{x_i}\big) - G_f(\mu)\right] \mu^{\otimes\mathbb{N}}(d\mathbf{x})\frac{\Xi_0(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta},\boldsymbol{\zeta})}. \tag{4.2}$$

This operator is the Ξ_0-Fleming-Viot generator from Proposition 1.3. The following representation will be useful to establish the duality with the Ξ_0-coalescent. Note that if Ξ is concentrated on $\{\boldsymbol{\zeta}\in\Delta : \zeta_i = 0 \text{ for all } i\geq 2\}$, i.e., if the corresponding coalescent is a Λ-coalescent, then this result has already been obtained by Bertoin and Le Gall (2003, Eqs. (16) and (17)).

For convenience, we will denote the transition rates by

$$\lambda(k_1,\ldots,k_p) \;=\; \lambda_{b;k_1,\ldots,k_r;s}, \tag{4.3}$$

where $k_1\geq\cdots\geq k_r\geq 2$, $p-r=s$ and $k_{r+1}=\ldots=k_p=1$. Furthermore, define for $p,n_1,\ldots,n_p\in\mathbb{N}$ such that $n_1+\cdots+n_p>p$ ($\Leftrightarrow$ not all $n_i=1$)

$$\lambda(n_1,\ldots,n_p) \;:=\; \lambda(k_1,\ldots,k_p), \tag{4.4}$$

where $k_1\geq\cdots\geq k_p$ is the re-arrangement of $n_1,\ldots,n_p$ in decreasing order.

Lemma 4.1. *The operator L^{Ξ_0} has the alternative representation*

$$L^{\Xi_0}G_f(\mu) \;=\; \sum_{\substack{\pi=\{A_1,\ldots,A_p\}\in\mathcal{P}_n\\ \text{not all singletons}}} \lambda(|A_1|,\ldots,|A_p|)\int_{E^n}\big(f\big(\mathbf{x}[\pi]\big)-f(\mathbf{x})\big)\,\mu^{\otimes n}(dx_1,\ldots,dx_n), \tag{4.5}$$

where $\mathbf{x}[\{A_1,\ldots,A_p\}]\in E^n$ has entries

$$(\mathbf{x}[\{A_1,\ldots,A_p\}])_i \;:=\; x_{\min A_j} \quad \text{if} \;\; i\in A_j,\; i=1,\ldots,n.$$

Remark 4.2. *Note that (4.5) basically boils down to (1.4), if $|A_i|=1$ for all but one A_i.*

Proof of Lemma 4.1: First note that for fixed $\boldsymbol{\zeta}$ and $\mathbf{x}$,

$$\begin{aligned}&G_f\big((1-|\boldsymbol{\zeta}|)\mu+\textstyle\sum_{i=1}^\infty\zeta_i\delta_{x_i}\big)\\ &\quad=\sum_{\phi:\{1,\ldots,n\}\to\mathbb{Z}_+}(1-|\boldsymbol{\zeta}|)^{a(\phi)}\prod_{j\leq n\,:\phi(j)>0}\zeta_{\phi(j)}\int_{E^{a(\phi)}}f\big(\eta(\phi,\mathbf{x},\mathbf{y})\big)\mu^{\otimes a(\phi)}(dy_1,\ldots,dy_{a(\phi)}),\end{aligned} \tag{4.6}$$

where $a(\phi):=\#\{1\leq j\leq n:\phi(j)=0\}$ and $\eta(\phi,\mathbf{x},\mathbf{y})\in E^n$ is given by

$$\eta(\phi,\mathbf{x},\mathbf{y})_j \;=\; \begin{cases} x_{\phi(j)} & \text{if } \phi(j)>0,\\ y_k & \text{if } \phi(j)=0, \text{ where } k=\#\{1\leq j'\leq j:\phi(j')=0\}.\end{cases}$$

Identity (4.6) can be understood as follows: Expanding the n-fold product of $(1-|\boldsymbol{\zeta}|)\mu+\sum_{i=1}^\infty\zeta_i\delta_{x_i}$, we put $\phi(j)=0$ if in the j-th factor, we use $(1-|\boldsymbol{\zeta}|)\mu$, and we put $\phi(j)=i$ if we use $\zeta_i\delta_{x_i}$ in the j-th factor.

Each $\phi:\{1,\ldots,n\}\to\mathbb{Z}_+$ is uniquely described by a partition $\pi=\{A_1,\ldots,A_p\}\in\mathcal{P}_n$ with labels $\ell_1,\ldots,\ell_p\in\mathbb{Z}_+$ by defining $j\sim_\phi j'$ if and only if $\phi(j)=\phi(j')>0$

and putting $\ell_i := \phi(A_i)$, $i = 1, \dots, p$. Note that for a given partition $\{A_1, \dots, A_p\}$, any vector $(\ell_1, \dots, \ell_p) \in \mathbb{Z}_+^p$ of labels with the properties

$$\ell_i = 0 \;\Rightarrow\; |A_i| = 1 \quad \text{and} \quad i \neq j, \ell_i, \ell_j \neq 0 \;\Rightarrow\; \ell_i \neq \ell_j$$

is admissible. Thus we have

$$\begin{aligned} &\int_{E^{\mathbb{N}}} G_f\big((1-|\boldsymbol{\zeta}|)\mu + \textstyle\sum_{i=1}^{\infty} \zeta_i \delta_{x_i}\big)\mu^{\otimes\mathbb{N}}(d\mathbf{x}) \\ &\quad = \sum_{\pi=\{A_1,\dots,A_p\}\in\mathcal{P}_n} \sum_{\substack{(\ell_1,\dots,\ell_p) \\ \text{admissible}}} (1-|\boldsymbol{\zeta}|)^{\#\{1\le i\le p:\ell_i=0\}} \prod_{\substack{i=1 \\ \ell_i>0}}^{p} \zeta_{\ell_i}^{|A_i|} \int_{E^n} f(\mathbf{x}[\pi])\,\mu^{\otimes n}(d\mathbf{x}). \end{aligned} \tag{4.7}$$

Note that, for a given partition with p blocks, the integration appearing in the last line runs effectively only over E^p. For further simplification assume that the blocks $A_1, \dots, A_p$ of $\pi = \{A_1, \dots, A_p\} \in \mathcal{P}_n$ are enumerated according to decreasing block size, and write $s(\pi)$ for the number of singleton blocks of the partition $\pi = \{A_1, \dots, A_p\}$. Then, for a given $\pi = \{A_1, \dots, A_p\} \in \mathcal{P}_n$, the last sum in (4.7) can be written as

$$\sum_{l=0}^{s(\pi)} \binom{s(\pi)}{l} (1-|\boldsymbol{\zeta}|)^{s(\pi)-l} \sum_{\substack{i_1,\dots,i_{p-s(\pi)+l}\in\mathbb{N} \\ \text{all distinct}}} \zeta_{i_1}^{|A_1|} \cdots \zeta_{i_{p-s(\pi)+l}}^{|A_{p-s(\pi)+l}|} \int_{E^n} f\big(\mathbf{x}[\pi]\big)\mu^{\otimes n}(d\mathbf{x}).$$

Furthermore, for any $\boldsymbol{\zeta} \in \Delta$ and $n \in \mathbb{N}$,

$$\begin{aligned} 1 &= \Big(\big(1-|\boldsymbol{\zeta}|\big) + \textstyle\sum_{i=1}^{\infty} \zeta_i\Big)^n \\ &= \sum_{\pi=\{A_1,\dots,A_p\}\in\mathcal{P}_n} \sum_{l=0}^{s(\pi)} \binom{s(\pi)}{l} (1-|\boldsymbol{\zeta}|)^{s(\pi)-l} \sum_{\substack{i_1,\dots,i_{p-s(\pi)+l}\in\mathbb{N} \\ \text{all disticnt}}} \zeta_{i_1}^{|A_1|} \cdots \zeta_{i_{p-s(\pi)+l}}^{|A_{p-s(\pi)+l}|}. \end{aligned}$$

This allows us to re-express the inner integral in (4.2) as

$$\begin{aligned} &\sum_{\pi=\{A_1,\dots,A_p\}\in\mathcal{P}_n} \sum_{l=0}^{s(\pi)} \binom{s(\pi)}{l} (1-|\boldsymbol{\zeta}|)^{s(\pi)-l} \quad \times \\ &\qquad \sum_{\substack{i_1,\dots,i_{p-s(\pi)+l}\in\mathbb{N} \\ \text{all distinct}}} \zeta_{i_1}^{|A_1|} \cdots \zeta_{i_{p-s(\pi)+l}}^{|A_{p-s(\pi)+l}|} \int_{E^n} \big[f\big(\mathbf{x}[\pi]\big) - f(\mathbf{x})\big]\,\mu^{\otimes n}(d\mathbf{x}) \\ &= \sum_{\substack{\pi=\{A_1,\dots,A_p\}\in\mathcal{P}_n \\ \text{not all singletons}}} \sum_{l=0}^{s(\pi)} \binom{s(\pi)}{l} (1-|\boldsymbol{\zeta}|)^{s(\pi)-l} \sum_{\substack{i_1,\dots,i_{p-s(\pi)+l}\in\mathbb{N} \\ \text{all distinct}}} \zeta_{i_1}^{|A_1|} \cdots \zeta_{i_{p-s(\pi)+l}}^{|A_{p-s(\pi)+l}|} \\ &\qquad \times \int_{E^n} \big[f\big(\mathbf{x}[\pi]\big) - f(\mathbf{x})\big]\,\mu^{\otimes n}(d\mathbf{x}), \end{aligned}$$

because $\mathbf{x}[\{\{1\}, \dots, \{n\}\}] = \mathbf{x}$. Integrating this equation over Δ with respect to the measure $(\boldsymbol{\zeta}, \boldsymbol{\zeta})^{-1}\Xi_0$ yields (4.5).

Note that (see also Sagitov, 2003, p. 844)

$$\begin{aligned}
&\sum_{\substack{\pi=\{A_1,\ldots,A_p\}\in\mathcal{P}_n\\ \text{not all singletons}}} \sum_{l=0}^{s(\pi)} \binom{s(\pi)}{l} (1-|\boldsymbol{\zeta}|)^{s(\pi)-l} \sum_{\substack{i_1,\ldots,i_{p-s(\pi)+l}\in\mathbb{N}\\ \text{all distinct}}} \zeta_{i_1}^{|A_1|}\cdots\zeta_{i_{p-s(\pi)+l}}^{|A_{p-s(\pi)+l}|}\\
&\leq \sum_{\substack{\pi=\{A_1,\ldots,A_p\}\in\mathcal{P}_n\\ \text{not all singletons}}} \Big(\sum_{i_1=1}^{\infty}\zeta_{i_1}^2\Big) \sum_{l=0}^{s(\pi)} \binom{s(\pi)}{l} (1-|\boldsymbol{\zeta}|)^{s(\pi)-l}\\
&\qquad\qquad \times \sum_{i_{p-s(\pi)+1},\ldots,i_{p-s(\pi)+l}\in\mathbb{N}} \zeta_{i_{p-s(\pi)+1}}\cdots\zeta_{i_{p-s(\pi)+l}}\\
&= \sum_{\substack{\pi=\{A_1,\ldots,A_p\}\in\mathcal{P}_n\\ \text{not all singletons}}} (\boldsymbol{\zeta},\boldsymbol{\zeta}) \sum_{l=0}^{s(\pi)} \binom{s(\pi)}{l} (1-|\boldsymbol{\zeta}|)^{s(\pi)-l}|\boldsymbol{\zeta}|^l \;=\; (|\mathcal{P}_n|-1)\,(\boldsymbol{\zeta},\boldsymbol{\zeta})
\end{aligned}$$

to verify that there is no singularity near $\boldsymbol{\zeta}=\mathbf{0}$. □

4.2. *Construction of the Markov semigroup and proof of Proposition 1.3.* The following proposition ensures that there exists a Markov process attached to the Ξ_0-Fleming-Viot generator.

Proposition 4.3. *The closure of*

$$\{(G_f, L^{\Xi_0}G_f) : n\in\mathbb{N}, f: E^n\to\mathbb{R} \text{ bounded and measurable}\}$$

generates a Markov semigroup on $\mathcal{M}_1(E)$.

Proof: We write G instead of G_f for convenience. By the Hille-Yosida Theorem (see, for example, Ethier and Kurtz, 1986, p. 165, Theorem 2.2) it is sufficient to verify that

- (i) the domain D is dense in $C(\mathcal{M}_1(E))$,
- (ii) the operator L^{Ξ_0} satisfies the positive maximum principle, i.e., $L^{\Xi_0}G(\mu)\leq 0$ for all $G\in D$, $\mu\in\mathcal{M}_1(E)$ with $\sup_{\nu\in\mathcal{M}_1(E)} G(\nu)=G(\mu)\geq 0$, and that
- (iii) the range of $\lambda - L^{\Xi_0}$ is dense in $C(\mathcal{M}_1(E))$ for some $\lambda>0$.

In order to verify (i) and (iii) we mimic the proof of Proposition 3.5 in Chapter 1 of Ethier and Kurtz (1986) and construct a suitable sequence $D_1, D_2, \ldots$ of finite-dimensional subspaces of $C(\mathcal{M}_1(E))$ such that $D := \bigcup_{k\in\mathbb{N}} D_k$ is dense in $C(\mathcal{M}_1(E))$ and $L^{\Xi_0}: D_k\to D_k$ for all $k\in\mathbb{N}$ as follows. For $n\in\mathbb{N}$ and $f: E^n\to\mathbb{R}$ bounded and measurable let D_f denote the set of all linear combinations of elements from the set

$$\{G : G(\mu)=\textstyle\int f(\mathbf{x}[\pi])\,\mu^{\otimes n}(d\mathbf{x}), \pi\in\mathcal{P}_n\}.$$

Since $|\mathcal{P}_n|<\infty$, it is easily seen that D_f is a finite-dimensional subspace of $C(\mathcal{M}_1(E))$. From (4.5) it follows that $L^{\Xi_0}: D_f\to D_f$. For each $n\in\mathbb{N}$ let $\{g_{nm}: m\in\mathbb{N}\}\subset C(E^n)$ be dense, and let $\{f_k : k\in\mathbb{N}\}$ be an enumeration of $\{g_{nm}: n,m\in\mathbb{N}\}$. Then, $D_k := D_{f_k}$, $k\in\mathbb{N}$, has the desired properties. Note that $D:=\bigcup_{k\in\mathbb{N}} D_k$ is dense in $C(\mathcal{M}_1(E))$ (Stone-Weierstrass), i.e. condition (i) holds.

We have $(\lambda-L^{\Xi_0})(D_k)=D_k$ for all λ not belonging to the set of eigenvalues of $L^{\Xi_0}|_{D_k}$, i.e., for all but at most finitely many $\lambda>0$. Thus, $(\lambda-L^{\Xi_0})(D) = (\lambda-L^{\Xi_0})(\bigcup_{k\in\mathbb{N}} D_k) = \bigcup_{k\in\mathbb{N}} D_k = D$ is dense in $C(\mathcal{M}_1(E))$ for all but at most countably many $\lambda>0$. In particular, condition (iii) is satisfied.

Condition (ii) follows from the fact that the expression inside the integrals in (1.12) satisfies

$$G((1-|\boldsymbol{\zeta}|)\mu+\sum_{i=1}^{\infty}\zeta_i\delta_{x_i})-G(\mu) \;\le\; \sup_{\nu\in\mathcal{M}_1(E)}G(\nu)-G(\mu) \;=\; G(\mu)-G(\mu) \;=\; 0$$

for all $\mathbf{x}=(x_1,x_2,\dots)\in E^{\mathbb{N}}$, $\boldsymbol{\zeta}\in\Delta$, $G\in D$ and $\mu\in\mathcal{M}_1(E)$ with $\sup_{\nu\in\mathcal{M}_1(E)}G(\nu)=G(\mu)$.

Thus, the Hille-Yosida Theorem ensures that the closure $\overline{L^{\Xi_0}}$ of L^{Ξ_0} on $C(\mathcal{M}_1(E))$ is single-valued and generates a strongly continuous, positive, contraction semigroup $\{T_t\}_{t\ge0}$ on $\mathcal{M}_1(E)$. Note that from (iii) it follows that D is a core for $\overline{L^{\Xi_0}}$ (Ethier and Kurtz, 1986, p. 166). The operator L^{Ξ_0} maps constant functions to the zero function, i.e., L^{Ξ_0} is conservative. Thus, $\{T_t\}_{t\ge0}$ is a Feller semigroup and corresponds to a Markov process with sample paths in $D_{\mathcal{M}_1(E)}([0,\infty))$. □

Remark 4.4. *i) If the finite measure Ξ on Δ allows for some mass $a:=\Xi(\{\mathbf{0}\})$ at zero, then L^{Ξ_0} has to be replaced by $L^{\Xi}:=L^{\Xi_0}+L^{a\delta_{\mathbf{0}}}$, where L^{Ξ_0} is defined as before and $L^{a\delta_{\mathbf{0}}}$ is the generator of the classical Fleming-Viot process (Fleming and Viot, 1979) given by (1.11). The existence of a Markov process $Z=(Z_t)_{t\ge0}$ with generator L^{Ξ} can be deduced as in the proof of Proposition 4.3 via the Hille-Yosida Theorem.*
ii) The construction of the Markov process attached to the 'full' generator L, including the Kingman component (1.11) and the mutation component (1.13), works via the standard Trotter approach.
iii) Note that $\int(L^{\Xi})G\,d\delta_{\delta_x}=0$, $x\in E$, where $\delta_\nu\in\mathcal{M}_1(\mathcal{M}_1(E))$ denotes the unit mass at $\nu\in\mathcal{M}_1(E)$. Thus, see Ethier and Kurtz (1986, p. 239, Proposition 9.2), the states δ_x, $x\in E$, are absorbing for the Ξ-Fleming-Viot process.

We now turn to the proof of Proposition 1.3. Indeed, we verify the following

Claim: The distribution of the measure valued Markov process with generator L, as defined in Remark 4.4 ii), coincides with the distribution of the (Ξ,B)-Fleming-Viot process, as defined in Theorem 1.1.

It suffices to verify the following lemma.

Lemma 4.5. *The (Ξ,B)-Fleming-Viot process defined in Theorem 1.1 solves the martingale problem for the generator L given in (1.10).*

To prepare this, let us concentrate on the case when there is no mutation and no Kingman-component ($L=L^{\Xi_0}$). Fix l and suppose we are at the m-th birth event. As in the previous section, let $\{\phi_m^1,\dots,\phi_m^{a_m}\}$ denote the assignments of the levels to one of the a_m families. So $\phi_m^i\subset\{1,\dots,l\}$ and $\phi_m^i\cap\phi_m^i\neq\emptyset$ for all i,j. Furthermore, we again denote by $\Phi_m:=\bigcup_{i=1}^{a_m}\phi_m^i$ all individuals participating in the birth event. Note, that this can be a strict subset of $\{0,\dots,l\}$, and $\{\phi_m^1,\dots,\phi_m^{a_m}\}$ holds all information about what is going on at the birth event. The function $g(\boldsymbol{\zeta},u)$ is defined as in (2.2). We introduce a Poisson process counting the number of times a specific birth event $\{\phi_m^1,\dots,\phi_m^{a_m}\}$ happens. With $(t_m,\boldsymbol{\zeta}_m,\mathbf{u}_m)$ denoting the points of the Poisson point process $\mathfrak{M}^{\Xi_0}$ we define

$$L_{\{\phi_m^1,\dots,\phi_m^{a_m}\}}(t):=\sum_{t_m\le t}\;\sum_{\substack{b_1,\dots,b_{a_m}\in\mathbb{N}\\ \text{all distinct}}}\;\prod_{i=1}^{a_m}\prod_{j\in\phi_m^i}\mathbb{1}_{\{g(\boldsymbol{\zeta}_m,u_{mj})=b_i\}}\prod_{j\in\{1,\dots,l\}\setminus\Phi_m}\mathbb{1}_{\{g(\boldsymbol{\zeta}_m,u_{mj})=\infty\}}. \tag{4.8}$$

To describe the effect of the birth event $\{\phi_m^1, \ldots, \phi_m^{a_m}\}$ on the population vector $x \in E^l$ we introduce the function $\mathfrak{T}$ defined by

$$\big(\mathfrak{T}_{\{\phi_m^1,\ldots,\phi_m^{a_m}\}}(\mathbf{x})\big)_i := \begin{cases} x_{\min(\phi_m^j)} & \text{if } k \in \phi_m^j, \\ x_{J_m(i)} & \text{else} \end{cases} \tag{4.9}$$

for all $k \in \{1, \ldots, l\}$, where J_m is the function defined in (2.20) that holds the information on where the non-participating particles should look down to.

With this notation we can use equation (2.22) and the dependence between the $L_{J,k}^l$ and L_J^l to show that

$$X^l(t) := X^l(0) + \sum_{\substack{\{\phi_m^1,\ldots,\phi_m^{a_m}\},\\ \dot\bigcup \phi_m^i \subset \{1,\ldots,l\}}} \int_0^t \Big(\mathfrak{T}_{\{\phi_m^1,\ldots,\phi_m^{a_m}\}}\big(X^l(s-)\big) - X^l(s-)\Big)\, dL_{\{\phi_m^1,\ldots,\phi_m^{a_m}\}}(s) \tag{4.10}$$

describes the evolution of the first l levels $X^l \in E^l$, if we assume no mutation and no Kingman part. Note that for simplicity we use the notation $X^l = (X_1, \ldots, X_l)$.

Since the $L_{\{\phi_m^1,\ldots,\phi_m^{a_m}\}}(t)$ are Poisson processes derived from the Poisson point process $\mathfrak{M}^{\Xi_0}$ it is straightforward to verify that their rates are given by

$$r\big(\{\phi_m^1, \ldots, \phi_m^{a_m}\}\big) := \sum_{\substack{i_1,\ldots,i_{a_m}\\ \text{all distinct}}} \int_\Delta \zeta_{i_1}^{k_m^1+1} \cdots \zeta_{i_r}^{k_m^r+1} \zeta_{i_{r+1}} \cdots \zeta_{i_{a_m}} (1-|\boldsymbol{\zeta}|)^{(l-|\Phi|)} \frac{\Xi_0(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta},\boldsymbol{\zeta})}, \tag{4.11}$$

where $k_m^i + 1 = |\phi_m^i|$ as before and the sets are ordered, such that $k_m^1 \geq \cdots \geq k_m^r \geq 1$ and $k_m^{r+1} = \cdots = k_m^{a_m} = 0$ hold. Assume that at least $k_m^1 \geq 1$ holds, because otherwise $\mathfrak{T}$ is the identity. Note that under this assumption the integral in (4.11) is finite (c.f. Schweinsberg, 2000 or Sagitov, 2003).

We now turn to the actual proof of the lemma.

Proof of Lemma 4.5: We will prove the result for the generator L^{Ξ_0}. The full result can then be obtained in analogy to the proof of Theorem 2.4 in Donnelly and Kurtz (1996).

Indeed, we have to show that for each function $G_f \in \mathcal{D}(L^{\Xi_0})$ of the form

$$G_f(\mu) = \langle f, \mu^{\otimes l} \rangle, \tag{4.12}$$

for $\mu \in \mathcal{M}_1(E)$ and $f\colon E^l \to \mathbb{R}$ bounded and measurable,

$$G_f(Z(t)) - G_f(Z(0)) - \int_0^t (L^{\Xi_0} G_f)(Z(s))\, ds \tag{4.13}$$

is a martingale with respect to the natural filtration of the Poisson point process $\mathfrak{M}^{\Xi_0}$ given by

$$\{\mathcal{J}_t\}_{t\geq 0} := \Big\{\sigma\big(\mathfrak{M}^{\Xi_0}\big|_{[0,t]\times\Delta\times[0,1]^{\mathbb{N}}}\big)\Big\}_{t\geq 0}. \tag{4.14}$$

Note that

$$\mathbb{E}\Big[f\big(X_1(s), \ldots, X_l(s)\big)\Big|\mathcal{J}_t\Big] = \mathbb{E}\Big[\big\langle f, Z(s)^{\otimes l}\big\rangle\Big|\mathcal{J}_t\Big] \tag{4.15}$$

holds for all $s, t \geq 0$, which will be crucial in the following steps.

We start by observing that, for $0 \le w \le t$, the representation (4.10) leads to

$$0 = \mathbb{E}\Big[f\big(X^l(t)\big) - f\big(X^l(w)\big) - \sum_{\substack{\{\phi_m^1,\dots,\phi_m^{a_m}\},\\ \dot\bigcup \phi_m^i \subset \{1,\dots,l\}}} \int_w^t \Big(f\big(\mathfrak{T}_{\{\phi_m^1,\dots,\phi_m^{a_m}\}}\big(X^l(s)\big)\big) - f\big(X^l(s)\big)\Big) r\big(\{\phi_m^1,\dots,\phi_m^{a_m}\}\big)\, ds \Big| \mathcal{J}_w \Big], \quad (4.16)$$

since this is a martingale.

Using the definition of the rates (4.11) and the fact that due to the exchangeability of X^l, the action of $\mathfrak{T}_{\{\phi_m^1,\dots,\phi_m^{a_m}\}}$ and the $[\pi]$ operation under the expectation is the same, we can now rewrite the last term (without the substraction of $f(X^l(s))$ from the integrand) as

$$\begin{aligned}
&\mathbb{E}\Big[\int_w^t \sum_{\substack{\{\phi_m^1,\dots,\phi_m^{a_m}\},\\ \dot\bigcup \phi_m^i \subset \{1,\dots,l\}}} r\big(\{\phi_m^1,\dots,\phi_m^{a_m}\}\big) f\Big(\mathfrak{T}_{\{\phi_m^1,\dots,\phi_m^{a_m}\}}\big(X^l(s)\big)\Big)\, ds \Big| \mathcal{J}_w \Big] \\
&= \mathbb{E}\Big[\int_w^t \sum_{\pi=\{A_1,\dots,A_p\}\in\mathcal{P}_n} \sum_{\substack{(r_1,\dots,r_p)\\ \text{admissible}}} \int_\Delta (1-|\boldsymbol\zeta|)^{\#\{r_i=0\}} \\
&\qquad \times \prod_{\substack{i=1\\ r_i>0}}^p \zeta_{r_i}^{|A_i|} \frac{\Xi_0(d\boldsymbol\zeta)}{(\boldsymbol\zeta,\boldsymbol\zeta)} f\Big(\big(X^l(s)\big)[\pi]\Big)\, ds \Big| \mathcal{J}_w \Big] \\
&= \mathbb{E}\Big[\int_w^t \int_\Delta \sum_{\pi=\{A_1,\dots,A_p\}\in\mathcal{P}_n} \sum_{\substack{(r_1,\dots,r_p)\\ \text{admissible}}} (1-|\boldsymbol\zeta|)^{\#\{r_i=0\}} \\
&\qquad \times \prod_{\substack{i=1\\ r_i>0}}^p \zeta_{r_i}^{|A_i|} \langle f\circ[\pi], Z(s)^{\otimes l}\rangle \frac{\Xi_0(d\boldsymbol\zeta)}{(\boldsymbol\zeta,\boldsymbol\zeta)}\, ds \Big| \mathcal{J}_w \Big] \\
&= \mathbb{E}\Big[\int_w^t \int_\Delta \int_{E^{\mathbb{N}}} G_f\big((1-|\boldsymbol\zeta|)Z(s) + \textstyle\sum_{i=1}^\infty \zeta_i \delta_{x_i}\big) Z(s)^{\otimes\mathbb{N}}(d\mathbf{x}) \frac{\Xi_0(d\boldsymbol\zeta)}{(\boldsymbol\zeta,\boldsymbol\zeta)}\, ds \Big| \mathcal{J}_w \Big],
\end{aligned} \quad (4.17)$$

since the sum about the configurations $\{\phi_m^1,\dots,\phi_m^{a_m}\}$ and the distinct indices $i_1,\dots,i_{a_m}$ can be rewritten as the sum about the partitions π and the admissible vectors $(r_1,\dots,r_p)$. The last equality holds due to equation (4.7).

Combining equation (4.16) with equation (4.17) we see that

$$\begin{aligned}
0 &= \mathbb{E}\Big[f\big(X^l(t)\big) - f\big(X^l(w)\big) \\
&\qquad - \int_w^t \int_\Delta \int_{E^{\mathbb{N}}} \big(G_f\big((1-|\boldsymbol\zeta|)Z(s) + \textstyle\sum_{i=1}^\infty \zeta_i\delta_{x_i}\big) \\
&\qquad\qquad - G_f(Z(s))\big) Z(s)^{\otimes\mathbb{N}}(d\mathbf{x}) \frac{\Xi_0(d\boldsymbol\zeta)}{(\boldsymbol\zeta,\boldsymbol\zeta)}\, ds \Big| \mathcal{J}_w \Big] \\
&= \mathbb{E}\Big[\langle f, Z(t)^{\otimes l}\rangle - \langle f, Z(w)^{\otimes l}\rangle - \int_w^t (L^{\Xi_0} G_f)(Z(s))\, ds \Big| \mathcal{J}_w \Big] \\
&= \mathbb{E}\Big[G_f\big(Z(t)\big) - G_f\big(Z(w)\big) - \int_w^t (L^{\Xi_0} G_f)(Z(s))\, ds \Big| \mathcal{J}_w \Big]
\end{aligned} \quad (4.18)$$

holds, where we use (4.15) in the second equality. Thus, (4.13) is a martingale. □

5. Dualities

5.1. *Distributional duality versus pathwise duality.* We first establish a *distributional duality* in the classical sense of Liggett (1985). Indeed, (4.5) and results about the classical Fleming-Viot process bring forth the following duality between a Ξ-coalescent $\Pi = (\Pi_t)_{t\geq 0}$ and a Ξ-Fleming-Viot process $Z = (Z_t)_{t\geq 0}$.

Lemma 5.1. *(Duality) For $n \in \mathbb{N}$, $f : E^n \to \mathbb{R}$ bounded and measurable, $\mu \in \mathcal{M}_1(E)$, $\pi \in \mathcal{P}_n$ and $t \geq 0$,*

$$\mathbb{E}^\mu\Big[\int_{E^n} f\big(\mathbf{x}[\pi]\big) Z_t^{\otimes n}(d\mathbf{x})\Big] = \mathbb{E}^\pi\Big[\int_{E^n} f\big(\mathbf{x}[\Pi_t^{(n)}]\big)\mu^{\otimes n}(d\mathbf{x})\Big], \tag{5.1}$$

where $\Pi_t^{(n)}$ is the restriction of Π_t to $\mathcal{P}_n$.

To obtain a *pathwise duality*, we use the driving Poisson processes of the modified lookdown construction to construct realisation-wise a Ξ-coalescent embedded in the Ξ-Fleming-Viot process.

More explicitly, recall the Poisson processes L_J^l and $L_{J,k}^l$ from equation (2.18) and (2.19) in Section 2.3 and the Poisson process $\mathfrak{N}_{ij}^K$ defined in Section 1.3. For each $t \geq 0$ and $l \in \mathbb{N}$, let $N_t^l(s), 0 \leq s \leq t$, be the level at time s of the ancestor of the individual at level l at time t. In terms of the L_J^l and $L_{J,k}^l$, the process $N_t^l(\cdot)$ solves, for $0 \leq s \leq t$,

$$\begin{aligned} N_t^l(s) \;=\; & l - \sum_{1\leq i<j<l} \int_{s-}^t \mathbb{1}_{\{N_t^l(u+)>j\}}\, d\mathfrak{N}_{ij}^K(u) \\ & - \sum_{1\leq i<j<l} \int_{s-}^t (j-i)\mathbb{1}_{\{N_t^l(u+)=j\}}\, d\mathfrak{N}_{ij}^K(u) \\ & - \sum_{K\subset\{1,\dots,l\}} \int_{s-}^t (N_t^l(u+) - J_m(N_t^l(u+)))\mathbb{1}_{\{N_t^l(u+)\notin K\}}\, dL_K^l(u) \\ & - \sum_{k\in\mathbb{N}} \sum_{K\subset\{1,\dots,l\}} \int_{s-}^t (N_t^l(u+) - \min(K))\mathbb{1}_{\{N_t^l(u+)\in K\}}\, dL_{K,k}^l(u), \end{aligned} \tag{5.2}$$

where $J_m(\cdot) = J_{m(u)}(\cdot)$ is defined by (2.20) and $m(u)$ is the index of the jump at time u. Fix $0 \leq T$ and, for $t \leq T$, define a partition Π_t^T of $\mathbb{N}$ such that k and l are in the same block of Π_t^T if and only if $N_T^l(T-t) = N_T^k(T-t)$. Thus, k and l are in the same block if and only if the two levels k and l at time T have the same ancestor at time $T-t$. Then (Donnelly and Kurtz, 1999, Section 5),

$$\text{the process } (\Pi_t^T)_{0\leq t\leq T} \text{ is a } \Xi\text{-coalescent run for time } T. \tag{5.3}$$

Note that by employing a natural generalisation of the lookdown construction using driving Poisson processes on $\mathbb{R}$ and e.g. using $T = 0$ above, one can use the same construction to find an Ξ-coalescent with time set $\mathbb{R}_+$. We would like to emphasise that the lookdown construction provides a realisation-wise coupling of the type distribution process $(Z_t)_{t\geq 0}$ and the coalescent describing the genealogy of a sample, thus extending (5.1), which is merely a statement about one-dimensional distributions.

5.2. *The function-valued dual of the* (Ξ, B)*-Fleming-Viot process.* The duality between the Ξ-Fleming-Viot process and the Ξ-coalescent established in Section 5.1 worked only on the genealogical level, the mutation was not taken into account. However, it is possible to define a function-valued dual to the (Ξ, B)-Fleming-Viot process such that not only the genealogical structure, but also the mutation is part of the duality. This kind of duality is well known for the classical Fleming-Viot process, see, e.g., Etheridge (2000, Chapter 1.12).

First note that due to Lemma 4.1 we can rewrite the generator of the (Ξ, B)-Fleming-Viot process given by equation (1.10) to obtain

$$\begin{aligned} LG_f(\mu) := a &\sum_{1\le i<j\le n} \int_{E^n} \Big(f(x_1,..,x_i,..,x_i,..,x_n) - f(x_1,..,x_i,..,x_j,..,x_n)\Big)\mu^{\otimes n}(d\mathbf{x}) \\ &+ \sum_{\substack{\pi=\{A_1,\dots,A_p\}\in\mathcal{P}_n \\ \text{not all singletons}}} \lambda(|A_1|,\dots,|A_p|) \int_{E^n} \big(f\big(\mathbf{x}[\pi]\big) - f(\mathbf{x})\big)\,\mu^{\otimes n}(d\mathbf{x}), \\ &+ r\sum_{i=1}^{n} \int_{E^n} B_i(f(x_1,\dots,x_n))\mu^{\otimes n}(d\mathbf{x}). \end{aligned} \tag{5.4}$$

We can now reinterpret the function $G_f(\mu)$ acting on measures as a function $G_\mu(f)$ acting on the functions $C_b(E^n)$. This reinterpretation transfers the operator L acting on $C\big(\mathcal{M}_1(E)\big)$ to an operator L^* acting on $C_b\big(C_b(E^n)\big)$. Let $\mathcal{C} := \bigcup_{n=1}^{\infty} C_b(E^n)$. A $\mathcal{C}$-valued Markov process $(\rho_t)_{t\ge 0}$ solving the martingale problem for L^* can then be constructed as follows:

- If $\rho_t(\mathbf{x}) \in C_b(E^n)$ and $n \ge 2$, then the process $(\rho_t)_{t\ge 0}$ jumps to $\rho_t\big(\mathbf{x}[\pi]\big)$ with rate $\lambda(|A_1|,\dots,|A_p|) + a\mathbb{1}_{\{\exists! |A_i|=2; \forall j\ne i: |A_j|=1\}}$, $\forall \pi = \{A_1,\dots,A_p\} \in \mathcal{P}_n$, where $|A_j| \ge 1$ for at least one j.
- If $\rho_t \in C_b(E)$, that is it is a function of a single variable, then no further jumps occur.
- Between jumps the process evolves deterministically according to the "heat flow" generated by the mutation operator (1.6), independently for each coordinate.

Note that this process is not literally a coalescent, but has coalescent-like features.

The duality relation between ρ_t and Z_t immediately follows from (5.4) and can be written in integrated form as

$$\mathbb{E}_{Z_0}\langle \rho_0, Z_t^{\otimes n}\rangle \;=\; \mathbb{E}_{\rho_0}\langle \rho_t, Z_0^{\otimes n}\rangle. \tag{5.5}$$

It can be used for example to show uniqueness of the martingale problem for L via the existence of $(\rho_t)_{t\ge 0}$ or to calculate the moments of the (Ξ, B)-Fleming-Viot process.

5.3. *The dual of the block counting process.* In this section, we specialise to the case where the type space E consists of two types only, say $E = \{0, 1\}$. Define the real-valued process $Y = (Y_t)_{t\ge 0}$ via $Y_t := Z_t(\{1\})$, $t \ge 0$. Define $g : \mathcal{M}_1(E) \to [0, 1]$ via $g(\mu) := \mu(\{1\})$. The generator A of Y is then given by $Af(x) = (L^{\Xi}(f \circ g))(\mu)$, $f \in C^2([0, 1])$, where μ depends on $x \in [0, 1]$ and can be chosen arbitrary, as long as $g(\mu) = x$. Thus,

$$Af(x) = a\frac{x(1-x)}{2}f''(x) + \int_\Delta \int_{\{0,1\}^{\mathbb{N}}} \big(f((1-|\boldsymbol{\zeta}|)x + \textstyle\sum_{i=1}^\infty \zeta_i y_i) - f(x)\big)(\mathcal{B}(1,x))^{\otimes \mathbb{N}}(d\mathbf{y})\frac{\Xi_0(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta},\boldsymbol{\zeta})}, \quad (5.6)$$

$x \in [0,1]$, $f \in C^2([0,1])$, where $\mathcal{B}(1,x)$ denotes the Bernoulli distribution with parameter x. For $x \in [0,1]$ let $V_1(x), V_2(x), \ldots$ be a sequence of independent and identically $\mathcal{B}(1,x)$-distributed random variables. Then,

$$Af(x) \;=\; a\frac{x(1-x)}{2}f''(x) + \int_\Delta \int_{[0,1]} \big(f((1-|\boldsymbol{\zeta}|)x + y) - f(x)\big)Q(\boldsymbol{\zeta},x,dy)\frac{\Xi_0(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta},\boldsymbol{\zeta})},$$

where $Q(\boldsymbol{\zeta},x,.)$ denotes the distribution of $\sum_{i=1}^\infty \zeta_i V_i(x)$. Hence the process can be considered as a Wright-Fisher diffusion with jumps. The situation where Ξ is concentrated on $[0,1] \times \{0\}^{\mathbb{N}}$, i.e., when the underlying Ξ-coalescent is a Λ-coalescent, has been studied in Bertoin and Le Gall (2005).

Note that $Af \equiv 0$ for $f(x) = x$, so Y is a martingale. Furthermore, the boundary points 0 and 1 are obviously absorbing.

In analogy to Lemma 5.1 it follows that Y is dual to the block counting process $D = (D_t)_{t\geq 0}$ of the Ξ-coalescent with respect to the duality function $H : [0,1]\times\mathbb{N} \to \mathbb{R}$, $H(x,n) := x^n$ (see, e.g., Liggett, 1985), i.e.,

$$\mathbb{E}^y[Y_t^n] \;=\; \mathbb{E}^n[y^{D_t}], \quad n \in \mathbb{N}, y \in [0,1], t \geq 0.$$

Thus, the moments of the ‘forward’ variable Y_t can be computed via the generating function of the ‘backward’ variable D_t and vice versa. Such and closely related moment duality relations are well known from the literature (Alkemper and Hutzenthaler, 2007; Athreya and Swart, 2005; Möhle, 1999). The duality can be used to relate the accessibility of the boundaries of Y and the existence of an entrance law for D with $D_{0+} = \infty$. Note that by the Markov property and the structure of the jump rates, we always have

$$\mathbb{P}^\infty(D_t = 1 \text{ eventually}) \;\in\; \{0,1\} \quad (5.7)$$

and either $\mathbb{P}^\infty(\bigcap_{t\geq 0}\{D_t = \infty\}) = 1$ (if the probability in (5.7) equals 0) or $\lim_{t\to\infty}\mathbb{P}^\infty(D_t = 1) = 1$ (if the probability in (5.7) equals 1).

Proposition 5.2. $\lim_{t\to\infty}\mathbb{P}^\infty(D_t = 1) = 1$ *if and only if* Y*, the dual of its block counting process, hits the boundary* $\{0,1\}$ *in finite time almost surely, starting from any* $y \in (0,1)$.

Proof: Fix $y \in (0,1)$, $T > 0$. Construct (Z_t) starting from $y\delta_1 + (1-y)\delta_0$ and no mutations, $Bf \equiv 0$, (and hence Y starting from y) by using the lookdown construction from Section 2.3: Let $X_1(0), X_2(0), \ldots$ be independent $\mathcal{B}(1,y)$-distributed random variables which are independent of the driving Poisson processes, and let $X_n(t)$, $t > 0$, $n \in \mathbb{N}$, be the solution of (2.22). Let

$$D'_t \;:=\; |\{N_T^n(T-t) : n \in \mathbb{N}\}|,$$

where $N_T^n(s)$ solves (5.2). By (5.3), the law of $(D'_t)_{0\leq t\leq T}$ is that of the block counting process of the (standard-)Ξ-coalescent run for time T. Then by construction (as there is no mutation),

$$X_n(T) = X_{N_T^n(0)}(0),$$

implying

$$\{D'_T = 1\} \subset \{Y_T \in \{0,1\}\} \quad \text{and} \quad \{D'_T = \infty\} \subset \{0 < Y_T < 1\} \text{ almost surely,}$$

which easily yields the claim. □

This is related to the so-called 'coming down from infinity'-property of the standard Ξ-coalescent (i.e., the property that starting from $D_0 = \infty$, $D_t < \infty$ almost surely for all $t > 0$). Recall Schweinsberg (2000, p. 39f), that a Ξ-coalescent may have infinitely many classes for a positive amount of time and then suddenly jumps to finitely many classes. This can occur if Ξ has positive mass on $\Delta_f := \{\mathbf{u} = (u_1, u_2, \dots) \in \Delta : u_1 + \cdots + u_n = 1 \text{ for some } n \in \mathbb{N}\}$. On the other hand by Lemma 31 of Schweinsberg (2000), if $\Xi(\Delta_f) = 0$, then the Ξ-coalescent either comes down from infinity immediately or always has infinitely many classes. Combining this with Proposition 5.2 we obtain

Remark 5.3. *Assume that $\Xi(\Delta_f) = 0$. Then the Ξ-coalescent comes down from infinity if and only if the dual of its block counting process hits the boundary $\{0,1\}$ in finite time almost surely.*

In general, there seems to be no 'simple' criterion to check whether a Ξ-coalescent comes down from infinity (see the discussion in Section 5.5 of Schweinsberg (2000)). On the other side, there seems to be also no 'handy' criterion for accessibility of the boundary of a process with jumps (and with values in $[0,1]$), but at least Proposition 5.2 allows to transfer any progress from one side to the other and vice versa.

We conclude this section with a simple toy example for which most quantities of interest, in particular the generator A, can be computed explicitly.

Example 5.4. Fix $l \in \mathbb{N}$. If the measure Ξ is concentrated on $\Delta_l := \{\boldsymbol{\zeta} \in \Delta : \zeta_1 + \cdots + \zeta_l = 1\}$, then (5.6) reduces to

$$Af(x) \;=\; \int_\Delta \sum_{y_1,\dots,y_l \in \{0,1\}} x^{y_1+\cdots+y_l}(1-x)^{l-(y_1+\cdots+y_l)}\big(f(\textstyle\sum_{i=1}^l \zeta_i y_i) - f(x)\big)\dfrac{\Xi(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta},\boldsymbol{\zeta})}.$$

For example, assume that the measure Ξ assigns its total mass $\Xi(\Delta) := 1/l$ to the single point $(1/l, \dots, 1/l, 0, 0, \dots) \in \Delta_l$. Then,

$$Af(x) \;=\; \sum_{k=0}^{l} \binom{l}{k} x^k (1-x)^{l-k} f(k/l) - f(x) \;=\; \int \big(f(y/l) - f(x)\big)\,\mathcal{B}(l,x)(dy),$$

where $\mathcal{B}(l,x)$ denotes the binomial distribution with parameters l and x. Note that the corresponding Ξ-coalescent never undergoes more than l multiple collisions at one time. The rates (4.3) are

$$\lambda(k_1,\dots,k_p) \;=\; \int_\Delta \sum_{\substack{i_1,\dots,i_p \in \mathbb{N} \\ \text{all distinct}}} \zeta_{i_1}^{k_1} \cdots \zeta_{i_p}^{k_p} \frac{\Xi(d\boldsymbol{\zeta})}{(\boldsymbol{\zeta},\boldsymbol{\zeta})} \;=\; \frac{(l)_p}{l^n},$$

where $(l)_p := l(l-1)\cdots(l-p+1)$ and $n := k_1 + \cdots + k_p$. The block counting process D has rates

$$g_{np} \;=\; \frac{n!}{p!} \sum_{\substack{m_1,\dots,m_p \in \mathbb{N} \\ m_1+\cdots+m_p = n}} \frac{\lambda(m_1,\dots,m_p)}{m_1! \cdots m_p!} \;=\; S(n,p)\frac{(l)_p}{l^n}, \quad 1 \le p < n,$$

where the $S(n,p)$ denote the Stirling numbers of the second kind. The total rates are $g_n = \sum_{p=1}^{n-1} g_{np} = 1 - (l)_n/l^n$, $n \in \mathbb{N}$. Note that the corresponding Ξ-coalescent stays infinite for a positive amount of time ('Case 2' on top of Schweinsberg, 2000, p. 39 with $\Xi_2 \equiv 0$). The dual of its block counting process hits the boundary in finite time. ∎

6. Examples

The first of the two examples in this section presents a model where the population size varies substantially due to recurrent bottlenecks. It is shown that the Ξ-coalescent appears naturally as the limiting genealogy of this model. In the second example we present the Poisson-Dirichlet-coalescent by choosing a particular measure for Ξ which has a density with respect to the Poisson-Dirichlet distribution. We provide explicit expressions for several quantities of interest.

6.1. *An example involving recurrent bottlenecks.* Consider a population, say with non-overlapping generations, in which the population size has undergone occasional abrupt changes in the past. Specifically, we assume that 'typically', each generation contains N individuals, but at several instances in the past, it has been substantially smaller for a certain amount of time, and then the population has quickly re-grown to its typical size N. This is related to the models considered by Jagers and Sagitov (2004), but we assume occasional much more radical changes in population size than Jagers and Sagitov (2004). Let us assume that the demographic history is described by three sequences of positive real numbers $(s_i)_{i\in\mathbb{N}}$, $(l_{i,N})_{i\in\mathbb{N}}$ and $(b_{i,N})_{i\in\mathbb{N}}$, where $0 < b_{i,N} \le 1$ holds for all i, and the population size t generations before the present is given by $G(t)$, where

$$G(t) \;=\; \begin{cases} b_{m,N}N & \text{if } N\big(\sum_{i=1}^{m-1}(s_i + l_{i,N}) + s_m\big) < t \le N\sum_{i=1}^{m}(s_i + l_{i,N}),\ m \in \mathbb{N}, \\ N & \text{otherwise.} \end{cases}$$

Thus, back in time the population stays at size N for some time s_iN. Then the size is reduced to $b_{i,N}N$ for the time $l_{i,N}N$. Thereafter it is again given by N, until the next bottleneck occurs after time $s_{i+1}N$. Note that for simplicity, we have assumed 'instantaneous' re-growth after each bottleneck. Furthermore, we assume that the reproduction behaviour is given by the standard Wright-Fisher dynamics, so each individual chooses its parent uniformly at random from the previous generation, independently of the other individuals. This is the case in every generation, also during the bottleneck and at the transitions between the bottleneck and the typical size.

We now want to keep track of the genealogy of a sample of n individuals from the present generation, and describe its dynamics in the limit $N \to \infty$. Denote by $\Pi^{(N,n)}(t)$ the ancestral partition of the sample t generations before the present.

Lemma 6.1. *Fix $(s_i)_{i\in\mathbb{N}}$ and assume that $b_{i,N} \to 0$ and that $l_{i,N} \to 0$ as $N \to \infty$. Furthermore assume that $b_{i,N}N \to \infty$ and that $l_{i,N}/b_{i,N} \to \gamma_i > 0$. Then*

$$\Pi^{(N,n)}(Nt) \;\to\; \Pi^{\delta_{\mathbf{0}},(n)}(R_t)$$

weakly as $N \to \infty$ on $D_{\mathcal{P}_n}([0,\infty))$, where $R_t := t + \sum_{i:s_1+\cdots+s_i \le t} \gamma_i$.

Note that we assume $l_{i,N} \to 0$ as $N \to \infty$, so the duration of the bottleneck is negligible on the timescale of the 'normal' genealogy. We also assume $b_{i,N} \to 0$

but $Nb_{i,N} \to \infty$, i.e., in the pre-limiting scenario, the population size during a bottleneck should be tiny compared to the normal size, but still large in absolute numbers. The ratio $l_{i,N}/b_{i,N}$ is sometimes called the *severity* of the (i-th) bottleneck in the population genetic literature.

Sketch of proof: Given sequences $(s_i), (b_{i,N})$ and $(l_{i,N})$, classical convergence results for samples of size n can be applied for the time-intervals between bottlenecks and “inside” the bottlenecks. Since $b_{i,N}N \to \infty$, the probability that any of the ancestral lines of the sample converge exactly at the transition to a bottleneck is $O((b_{i,N}N)^{-1}) = o(1)$, so that naïve “glueing” is feasible. □

Remark 6.2. *Note that bottleneck events with $\gamma_i = 0$ become invisible in the limit, whereas in a bottleneck with $\gamma_i = +\infty$ the genealogy necessarily comes down to only one lineage (and thus, all genetic variability is erased).*

Since we fixed the s_i and γ_i, the limiting process described in Lemma 6.1 is not a homogeneous Markov process and thus does not fit literally into the class of exchangeable coalescent processes considered in this paper. Assume that the waiting intervals s_i are exponentially distributed, say with parameter β, and that the γ_i are independently drawn from a certain law $\mathcal{L}_\gamma$. Thus, in the pre-limiting N-particle model forwards in time, in each generation there is a chance of $\sim \beta/N$ that a ‘bottleneck event’ with a randomly chosen severity begins. In this situation, the genealogy of an n-sample from the population at present is (approximately) described by

$$\Pi^{\delta_{\mathbf{0}},(n)}(S_t), \quad t \geq 0, \tag{6.1}$$

where $(S_t)_{t\geq 0}$ is a subordinator (in fact, a compound Poisson process with Lévy measure $\beta\mathcal{L}_\gamma$ and drift 1).

Proposition 6.3. *Let N_γ be the number of lineages at time $\gamma > 0$ in the standard Kingman coalescent starting with $N_0 = \infty$, and let D_j be the law of the re-ordering of a (j-dimensional) Dirichlet$(1, \dots, 1)$ random vector according to decreasing size, padded with infinitely many zeros. The process defined in (6.1) is the Ξ-coalescent restricted to $\{1, \dots, n\}$, where*

$$\Xi(d\boldsymbol{\zeta}) \;=\; \delta_{\mathbf{0}}(d\boldsymbol{\zeta}) + (\boldsymbol{\zeta}, \boldsymbol{\zeta}) \int_{(0,\infty)} \sum_{j=1}^{\infty} \mathbb{P}(N_\sigma = j) D_j(d\boldsymbol{\zeta})\, \beta\mathcal{L}_\gamma(d\sigma).$$

Proof: Recall that the number of families of the classical Fleming-Viot process without mutation after σ time units is N_σ. Given $N_\sigma = j$, the distribution of the family sizes is a uniform partition of $[0, 1]$, hence Dirichlet$(1, \dots, 1)$. Size-ordering thus leads to the above formula for Ξ. □

6.2. *The Poisson-Dirichlet case.* The Poisson-Dirichlet distribution PD_θ with parameter $\theta > 0$ is a distribution concentrated on the subset Δ^* of points $\boldsymbol{\zeta} \in \Delta$ satisfying $|\boldsymbol{\zeta}| = 1$. It can, for example, be obtained via size-ordering of the normalized jumps of a Gamma-subordinator at time θ. For more information on this distribution we refer to Kingman (1975) or Arratia et al. (1999). Sagitov (2003) considered the Poisson-Dirichlet coalescent $\Pi = (\Pi_t)_{t\geq 0}$ with parameter $\theta > 0$, where (by definition) the measure Ξ has density $\boldsymbol{\zeta} \mapsto (\boldsymbol{\zeta}, \boldsymbol{\zeta})$ with respect to PD_θ.

As the measure PD_θ is concentrated on Δ^*, the rates (4.3) reduce to

$$\lambda(k_1,\ldots,k_j) \;=\; \int_{\Delta^*} \sum_{\substack{i_1,\ldots,i_j\in\mathbb{N}\\ \text{all distinct}}} \zeta_{i_1}^{k_1}\cdots\zeta_{i_j}^{k_j}\,\mathrm{PD}_\theta(d\boldsymbol{\zeta}).$$

From the calculations of Kingman (1993) it follows that the Poisson-Dirichlet coalescent has rates

$$\lambda(k_1,\ldots,k_j) \;=\; \frac{\theta^j}{[\theta]_k}\prod_{i=1}^j (k_i-1)!,$$

$k_1,\ldots,k_j\in\mathbb{N}$ with $k:=k_1+\cdots+k_j>j$, where $[\theta]_k:=\theta(\theta+1)\ldots(\theta+k-1)$.

Möhle and Sagitov (2001) characterised exchangeable coalescents via a sequence $(F_j)_{j\in\mathbb{N}}$ of symmetric finite measures. For each $j\in\mathbb{N}$, the measure F_j lives on the simplex $\Delta_j:=\{(\zeta_1,\ldots,\zeta_j)\in[0,1]^j:\zeta_1+\cdots+\zeta_j\le 1\}$ and is uniquely determined via its moments

$$\lambda(k_1,\ldots,k_j) \;=\; \int_{\Delta_j}\zeta_1^{k_1-2}\cdots\zeta_j^{k_j-2}F_j(d\zeta_1,\ldots,d\zeta_j),\quad k_1,\ldots,k_j\ge 2.$$

For the Poisson-Dirichlet coalescent, an application of Liouville's integration formula shows that the measure F_j has density $f_j(\zeta_1,\ldots,\zeta_j):=\theta^j\zeta_1\cdots\zeta_j(1-\sum_{i=1}^j\zeta_i)^{\theta-1}$ with respect to the Lebesgue measure on Δ_j.

As Ξ is concentrated on Δ^*, it follows that

$$\int_\Delta \frac{|\boldsymbol{\zeta}|}{(\boldsymbol{\zeta},\boldsymbol{\zeta})}\,\Xi(d\boldsymbol{\zeta}) \;=\; \int_\Delta \frac{1}{(\boldsymbol{\zeta},\boldsymbol{\zeta})}\,\Xi(d\boldsymbol{\zeta}) \;=\; \int_{\Delta^*}\mathrm{PD}_\theta(d\boldsymbol{\zeta}) \;=\; 1 \;<\; \infty. \tag{6.2}$$

By Schweinsberg (2000, Proposition 29), the Poisson-Dirichlet coalescent is a jump-hold Markov process with bounded transition rates and step function paths. By Schweinsberg (2000, Proposition 30), for arbitrary but fixed $t>0$, Π_t does not have proper frequencies.

The block counting process $D:=(D_t)_{t\ge 0}$, where $D_t:=|\Pi_t|$ denotes the number of blocks of Π_t, is a decreasing process with rates

$$g_{nk} \;=\; \frac{n!}{k!}\sum_{\substack{n_1,\ldots,n_k\in\mathbb{N}\\ n_1+\cdots+n_k=n}}\frac{\lambda(n_1,\ldots,n_k)}{n_1!\cdots n_k!} \;=\; \frac{\theta^k}{[\theta]_n}\frac{n!}{k!}\sum_{\substack{n_1,\ldots,n_k\in\mathbb{N}\\ n_1+\cdots+n_k=n}}\frac{1}{n_1\cdots n_k} \;=\; \frac{\theta^k}{[\theta]_n}s(n,k),$$

$k,n\in\mathbb{N}$ with $k<n$, where the $s(n,k)$ are the absolute Stirling numbers of the first kind. The total rates are

$$g_n \;:=\; \sum_{k=1}^{n-1} g_{nk} \;=\; 1-\frac{\theta^n}{[\theta]_n},\quad n\in\mathbb{N}.$$

Note that $g_{nk}=\mathbb{P}\{K_n=k\}$, $k<n$, where K_n is a random variable taking values in $\{1,\ldots,n\}$ with distribution

$$\mathbb{P}\{K_n=k\} \;=\; \frac{\theta^k}{[\theta]_n}s(n,k),\qquad k\in\{1,\ldots,n\}.$$

We have

$$\gamma_n \;:=\; \sum_{k=1}^{n-1}(n-k)g_{nk} \;=\; \sum_{k=1}^{n-1}(n-k)\mathbb{P}\{K_n=k\} \;=\; n-\mathbb{E}K_n \;\le\; n.$$

In particular, $\sum_{n=2}^{\infty} \gamma_n^{-1} \geq \sum_{n=2}^{\infty} 1/n = \infty$. Together with (6.2) and $\Xi(\Delta_f) = 0$, where $\Delta_f := \{\boldsymbol{\zeta} \in \Delta \,|\, \zeta_1 + \cdots + \zeta_n = 1 \text{ for some } n\}$, it follows from Schweinsberg (2000, Proposition 33) that the Poisson-Dirichlet coalescent stays infinite.

If we assume no mutation, then the generator L^{Ξ} (defined in Remark 4.4) of the corresponding Fleming-Viot process reduces to

$$L^{\Xi} G_f(\mu) \;=\; \int_{\Delta^*} \int_{E^{\mathbb{N}}} \left[G_f\big(\textstyle\sum_{i=1}^{\infty} \zeta_i \delta_{x_i}\big) - G_f(\mu) \right] \mu^{\otimes \mathbb{N}}(d\mathbf{x}) \mathrm{PD}_{\theta}(d\boldsymbol{\zeta}).$$

Acknowledgement. We thank the referee for her/his careful reading of the manuscript and helpful suggestions.

References

R. Alkemper and M. Hutzenthaler. Graphical representations of some duality relations in stochastic population models. *Electron. Commun. Probab.* **12**, 206–229 (2007). (link).

R. Arratia, A.D. Barbour and S. Tavaré. The Poisson-Dirichlet distribution and the scale-invariant Poisson process. *Combin. Probab. Comput.* **8**, 407–416 (1999).

S. R. Athreya and J. M. Swart. Branching-coalescing particle systems. *Probab. Theory Relat. Fields* **131**, 376–414 (2005).

J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields* **126**, 261–288 (2003).

J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes ii: Stochastic differential equations. *Ann. Inst. Henri Poincaré Probabilités et Statistiques* **41**, 307–333 (2005).

J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes iii: Limit theorems. *Illinois Journal of Mathematics* **10**, 147–181 (2006).

M. Birkner and J. Blath. Measure-valued diffusions, general coalescents and genetic inference (2009). To appear in: Trends in Stochastic Analysis, Cambridge University Press.

M. Birkner, J. Blath, M. Capaldo, A. Etheridge, M. Möhle, J. Schweinsberg and A. Wakolbinger. Alpha-stable branching and Beta-coalescents. *Electron. J. Probab.* **10**, 303–325 (2005).

C. Cannings. The latent roots of certain Markov chains arising in genetics: a new approach, i. Haploid models. *Adv. Appl. Probab.* **6**, 260–290 (1974).

C. Cannings. The latent roots of certain Markov chains arising in genetics: a new approach, ii. Further haploid models. *Adv. Appl. Probab.* **7**, 264–282 (1975).

P. Donnelly. Dual processes in population genetics. In *Stochastic Spatial Processes*, volume 1212 of *Lecture Notes in Math*, pages 95–105. Springer, Berlin (1986).

P. Donnelly and T. Kurtz. A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.* **24**, 698–742 (1996).

P. Donnelly and T. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.* **27**, 166–205 (1999).

A. M. Etheridge. *An introduction to superprocesses.* American Mathematical Society (2000).

S. Ethier and S. M. Krone. Comparing Fleming-Viot and Dawson-Watanabe processes. *Stoch. Process. Appl.* **60**, 401–421 (1995).

S. Ethier and T. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley (1986).

S. Ethier and T. Kurtz. Fleming-Viot processes in population genetics. *SIAM J. Control Optim.* **31**, 345–386 (1993).

W. H. Fleming and M. Viot. Some measure-valued Markov processes in population genetics theory. *Indiana University Mathematics Journal* **28**, 817–843 (1979).

A. Greven, P. Pfaffelhuber and A. Winter. Convergence in distribution of random metric measures spaces: (Lambda-coalescent measure trees) (2009). To appear in *Probab. Theory Relat. Fields*. (link).

S. Hiraba. Jump-type Fleming-Viot processes. *Adv. Appl. Probab.* **32**, 140–158 (2000).

P. Jagers and S. Sagitov. Convergence to the coalescent in populations of substantially varying size. *J. Appl. Probab.* **41**, 368–378 (2004).

J. F. C. Kingman. Random discrete distributions. *J. Royal Statist. Soc. B* **37**, 1–22 (1975).

J. F. C. Kingman. The coalescent. *Stoch. Process. Appl.* **13**, 235–248 (1982a).

J. F. C. Kingman. On the genealogy of large populations. *J. Appl. Probab.* **19A**, 27–43 (1982b).

J. F. C. Kingman. *Poisson processes*. The Clarendon Press Oxford University Press (1993).

T. Kurtz and E. Rodrigues. Poisson representations of branching Markov and measure-valued branching processes (2008). Preprint.

T. M. Liggett. *Interacting Particle Systems*. Springer, Berlin (1985).

M. Möhle. The concept of duality and applications to Markov processes arising in neutral population genetics models. *Bernoulli* **5**, 761–777 (1999).

M. Möhle. Forward and backward diffusion approximations for haploid exchangeable population models. *Stoch. Process. Appl.* **95**, 133–149 (2001).

M. Möhle and S. Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29**, 1547–1562 (2001).

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

S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**, 1116–1125 (1999).

S. Sagitov. Convergence to the coalescent with simultaneous multiple mergers. *J. Appl. Probab.* **40**, 839–854 (2003).

O. Sargsyan and J. Wakeley. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies on many marine organisms. *Theor. Pop. Biol.* **74**, 104–114 (2008).

J. Schweinsberg. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5** (2000). (link).

J. E. Taylor and A. Véber. Coalescent processes in mass extinctions. *Electron. J. Probab.* **14**, 242–288 (2009). (link).