isid/ms/2003/22
August 8, 2003
http://www.isid.ac.in/~statmath/eprints

Branching-coalescing particle systems

Siva R. Athreya
Jan M. Swart

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Branching-coalescing particle systems

Siva R. Athreya Jan M. Swart*

August 8, 2003

Abstract

We study the ergodic behavior of systems of particles performing independent random walks, binary splitting, coalescence and deaths. Such particle systems are dual to systems of linearly interacting Wright-Fisher diffusions, used to model a population with resampling, selection and mutations. We use this duality to prove that the upper invariant measure of the particle system is the only homogeneous nontrivial invariant law and the limit started from any homogeneous nontrivial initial law. An interesting tool in our proofs is an ergodic theorem for countable groups that need not be amenable.

Contents

*Research supported in part by the German Science Foundation.

MSC 2000. Primary: 60K35, 92D25; Secondary: 60J80, 60J60.

Keywords. First Schlögl model, reaction-diffusion process, autocatalytic reaction, branching, coalescence, resampling, selection, mutation, contact process, spatial ergodic theorem, nonamenable group.

1 Introduction and Main Results

1.1 Introduction

This paper studies systems of particles subject to a stochastic dynamics with the following description. 1° Each particle moves independently of the others according to a continuous time Markov process on a lattice Λ, which jumps from site i to site j with rate $a(i,j)$. 2° Each particle splits with rate $b \geq 0$ into two new particles, created on the position of the old one. 3° Each pair of particles, present on the same site, coalesces with rate $2c$ (with $c \geq 0$) to one particle. 4° Each particle dies with rate $d \geq 0$. Throughout this paper, we make the following assumptions.

(i) Λ is a finite or countably infinite set.

(ii) The transition rates $a(i,j)$ are irreducible, i.e., if $\Delta \subset \Lambda$ is neither Λ nor $\emptyset$, then there exist $i \in \Delta$ and $j \in \Lambda \backslash \Delta$ such that $a(i,j) > 0$ or $a(j,i) > 0$.

(iii) $\sup_i \sum_j a(i,j) < \infty$.

(iv) $\sum_j a^\dagger(i,j) = \sum_j a(i,j)$, where $a^\dagger(i,j) := a(j,i)$.

(v) b, c, and d are nonnegative constants.

Here and elsewhere sums and suprema over i,j always run over Λ, unless stated otherwise. Assumption (iv) says that the counting measure is an invariant σ-finite measure for the Markov process with jump rates a. With respect to this invariant measure, the time-reversed process jumps from i to j with rate $a^\dagger(i,j)$.

Let $X_t(i)$ denote the number of particles present at site $i \in \Lambda$ and time $t \geq 0$. Then $X = (X_t)_{t \geq 0}$, with $X_t = (X_t(i))_{i \in \Lambda}$, is a Markov process with formal generator

$$\begin{aligned} Gf(x) := &\sum_{ij} a(i,j)x(i)\{f(x+\delta_j-\delta_i)-f(x)\} + b\sum_i x(i)\{f(x+\delta_i)-f(x)\} \\ &+c\sum_i x(i)(x(i)-1)\{f(x-\delta_i)-f(x)\} + d\sum_i x(i)\{f(x-\delta_i)-f(x)\}, \end{aligned} \tag{1.1}$$

where $\delta_i(j) := 1$ if $i = j$ and $\delta_i(j) := 0$ otherwise. The process X can be defined for finite initial states and also for some infinite initial states in an appropriate Liggett-Spitzer space (see Section 1.3). We call $(X_t)_{t\geq 0}$ a branching coalescing particle system with underlying motion (Λ, a), branching rate b, coalescence rate c and death rate d, or shortly the (a,b,c,d)-braco-process.

Some typical examples of underlying motions we have in mind are nearest neighbour random walk on $\Lambda = \mathbb{Z}^d$ and on $\Lambda = \mathbb{T}^d$, the homogeneous tree of degree $d+1$. We will not restrict ourselves to symmetric underlying motions (i.e., $a = a^\dagger$) but also allow $a(i,j) = 1_{\{j=i+1\}}$ on $\mathbb{Z}$, for example. The reason why we do not restrict ourselves to graphs, is that we also want to include the case $\Lambda = \Omega_d$, the hierarchical group with freedom d, i.e.,

$$\Omega_d := \{i = (i_1, i_2, \ldots) : i_\alpha \in \{0, \ldots, d-1\} \ \forall \alpha \geq 1, \ i_\alpha \neq 0 \text{ finitely often}\}, \tag{1.2}$$

equipped with componentwise addition modulo n. On Ω_d, one typically chooses transition rates $a(i,j)$ that depend only on the hierarchical distance $|i-j| := \min\{\alpha \geq 0 : i_\beta = j_\beta \ \forall \beta > \alpha\}$. The hierarchical group has found widespread applications in population biology and is therefore a natural choice for the underlying space.

1.2 Motivation

Our motivation for studying branching-coalescing particle systems comes from three directions.

Reaction diffusion models, Schlögl's first model. Branching-coalescing particle systems are known in the physics literature as a reaction diffusion models. More precisely, our model is a special case of Schlögl's first model [Sch72], where in the latter there is an additional rate with which particles are spontaneously created. For $d = 0$, our model is known as the autocatalytic reaction. Reaction diffusion models have been studied extensively by physicists and more recently also by probabilists [DDL90, Mou92, Neu90]. However, all work that we are aware of is restricted to the case $\Lambda = \mathbb{Z}^d$.

Population dynamics, the contact process. Branching-coalescing particle systems may be thought of as a more or less realistic model for the spread and growth of a population of organisms. Here, the underlying motion models the migration of organisms, births and deaths have their obvious interpretations, while coalescence of particles should be thought of as additional deaths, caused by local overpopulation. In this respect, our model is similar to the contact process. The latter is often referred to as a model for the spread of an infection, but in fact it is a reasonable model for the population dynamics of many organisms, from trees in a forest to killer bees. There are two striking differences between the contact process and branching-coalescing particle systems. First, whereas the total population at one site is subject to a rigid bound in the contact process (namely one), it may reach arbitrarily high values in a branching-coalescing system. However, when the local population is high, the coalescence (which grows quadratically in the number of organisms) dominates the branching (which grows linearly), and in this way the population is reduced. A second difference is that in the contact process, if one site infects its neighbor, the original site is still infected. As opposed to this, even when the death rate is zero, it is possible that a branching coalescing particle system goes to local extinction due to migration only. Thus, we can say that the gain from infection is guaranteed in the contact process, whereas the reward for migration is uncertain in a branching-coalescing particle system.

Resampling with selection and negative mutations. Our third motivation also comes from population dynamics, but from a different perspective. Assume that at each site $i \in \Lambda$ there lives a large, fixed number of organisms, and that each of these organisms carries a gene that comes in two types: a healthy and a defective one. Let us model the evolution of the population as follows. 1° with rate $a(i,j)$, we let an organism at site i migrate to site j. 2° to model the effect of natural selection, we let each organism with rate b choose another organism, living on the same site. If the first organism carries a healthy gene and the second organism a defective gene, then the latter is replaced by an organism with a healthy gene. 3° to model the effect of random mating, we resample each pair of organisms living at the same site with rate $2c$, i.e., we choose one of the two at random and replace it by an organism with the type of the other one. 4° with rate d, we let a healthy gene mutate into a defective gene. In the limit that the number of organisms at each site is large, the frequencies $\mathcal{X}_t(i)$ of healthy organisms at site i and time t are described by the unique pathwise solution to the infinite dimensional stochastic differential equation (SDE):

$$\begin{aligned} \mathrm{d}\mathcal{X}_t(i) = & \sum_j a(j,i)(\mathcal{X}_t(j) - \mathcal{X}_t(i))\,\mathrm{d}t + b\mathcal{X}_t(i)(1-\mathcal{X}_t(i))\,\mathrm{d}t - \mathrm{d}\mathcal{X}_t(i)\,\mathrm{d}t \\ & + \sqrt{2c\mathcal{X}_t(i)(1-\mathcal{X}_t(i))}\,\mathrm{d}B_t(i) \qquad (t \geq 0,\ i \in \Lambda). \end{aligned} \tag{1.3}$$

We call the $[0,1]^\Lambda$-valued process $\mathcal{X} = (\mathcal{X}_t)_{t\geq 0}$ the resampling-selection process with underlying motion (Λ, a), selection rate b, resampling rate c and mutation rate d, or shortly the (a,b,c,d)-resem-process (the letters in 'resem' standing for resampling, selection and mutation).

It is known that branching-coalescing particle systems are dual to resampling-selection processes. To be precise, for any $\phi \in [0,1]^\Lambda$ and $x \in \mathbb{N}^\Lambda$, write

$$\phi^x := \prod_i \phi(i)^{x(i)}, \tag{1.4}$$

where $0^0 := 1$. Let $\mathcal{X}$ be the (a,b,c,d)-resem-process and let X be the $(a^\dagger,b,c,d)$-braco-process. Then (see Theorem 1 (a) below)

$$E^\phi[(1-\mathcal{X}_t)^x] = E^x[(1-\phi)^{X_t^\dagger}]. \tag{1.5}$$

Formula (1.5) has the following interpretation: $E^\phi[(1-\mathcal{X}_t)^x]$ is the probability that x organisms, sampled from the population at time t, all have defective genes. If we want to calculate this probability, we must follow back in time those organisms that could possibly be healthy ancestors of these x organisms. In this way we end up with a system of branching coalescing $a^\dagger$-random walks, which die when a mutation occurs, coalesce when two potential ancestors descend from the same ancestor, and branch when a selection event takes place. If we end up with at least one healthy potential ancestor at time zero, then we know that not all the x particles have defective genes. As far as we know, the idea of modeling selection in this way by introducing branching into the usual coalescent was developed first (in a non-spatial setting) in [KN97]. The idea was later applied in [DK99, DG99, BES02]. A SPDE version of (1.3) (with $d = 0$) has been derived as the rescaled limit of long-range biased voter models in [MT95, Theorem 2].

Note that for $c = 0$, the process $\mathcal{X}$ is deterministic. In this case, the semigroup $(U_t)_{t\geq 0}$ defined by $U_t\phi := \mathcal{X}_t$ $(t \geq 0)$, where $\mathcal{X}$ is the deterministic solution of (1.3) with initial state $\mathcal{X}_0^\dagger = \phi \in [0,1]^\Lambda$, is called the generating semigroup of the branching particle system X. (For this terminology, see for example [FS03b].) Thus, the duality relation (1.5) says that, loosely speaking, branching coalescing particle systems have a random generating semigroup. The SDE (1.3) will be our main tool for studying branching coalescing particle systems.

1.3 Preliminaries

In this section we introduce the notation and definitions that we will use throughout the paper.

(Inner product and norm notation) For $\phi, \psi \in [-\infty, \infty]^\Lambda$, we write

$$\langle \phi, \psi \rangle := \sum_i \phi(i)\psi(i) \qquad \text{and} \qquad |\phi| := \sum_i |\phi(i)|, \tag{1.6}$$

whenever the infinite sums are defined.

(Poisson measures) If ϕ is a $[0, \infty)^\Lambda$-valued random variable, then by definition a Poisson measure with random intensity ϕ is an $\mathbb{N}^\Lambda$-valued random variable $\mathrm{Pois}(\phi)$ whose law is uniquely determined by

$$E[(1-\psi)^{\mathrm{Pois}(\phi)}] = E[e^{-\langle \phi, \psi \rangle}] \qquad (\psi \in [0,1]^\Lambda). \tag{1.7}$$

In particular, when ϕ is nonrandom, then the components $(\mathrm{Pois}(\phi)(i))_{i\in\Lambda}$ are independent Poisson distributed random variables with intensity $\phi(i)$.

(Thinned point measures) If x and ϕ are random variables taking values in $\mathbb{N}^\Lambda$ and $[0,1]^\Lambda$, respectively, then by definition a ϕ-thinning of x is an $\mathbb{N}^\Lambda$-valued random variable $\mathrm{Thin}_\phi(x)$ whose law is uniquely determined by

$$E[(1-\psi)^{\mathrm{Thin}_\phi(x)}] = E[(1-\phi\psi)^x] \qquad (\psi \in [0,1]^\Lambda). \tag{1.8}$$

In particular, when x and ϕ are nonrandom, and $x = \sum_{n=1}^m \delta_{i_n}$, then a ϕ-thinning of x can be constructed as $\mathrm{Thin}_\phi(x) := \sum_{n=1}^m \chi_n \delta_{i_n}$ where the χ_n are independent random variables with $P[\chi_n = 1] = 1 - P[\chi_n = 0] = \phi(i_n)$.

If ϕ and x are both random, then it will always be understood that they are independent. Thus, $\mathcal{L}(\mathrm{Thin}_\phi(x))$ depends on the laws $\mathcal{L}(\phi)$ and $\mathcal{L}(x)$ alone, and it is only the map $(\mathcal{L}(\phi), \mathcal{L}(x)) \mapsto \mathcal{L}(\mathrm{Thin}_\phi(x))$ that is of interest to us. We have chosen the present notation in terms of random variables instead of their laws to keep things simple if ϕ and x are nonrandom.

We leave it to the reader to check the elementary relations

$$\mathrm{Thin}_\psi(\mathrm{Thin}_\phi(x)) \stackrel{\mathcal{D}}{=} \mathrm{Thin}_{\psi\phi}(x) \quad \text{and} \quad \mathrm{Thin}_\psi(\mathrm{Pois}(\phi)) \stackrel{\mathcal{D}}{=} \mathrm{Pois}(\psi\phi), \tag{1.9}$$

where $\stackrel{\mathcal{D}}{=}$ denote equality in distribution.

(Weak convergence) We let $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ denote the one-point compactification of $\mathbb{N}$, and equip $\overline{\mathbb{N}}^\Lambda$ with the product topology. We say that probability measures ν_n on $\overline{\mathbb{N}}^\Lambda$ converge weakly to a limit ν, denoted as $\nu_n \Rightarrow \nu$, when $\int \nu_n(\mathrm{d}x) f(x) \to \int \nu(\mathrm{d}x) f(x)$ for every $f \in \mathcal{C}(\overline{\mathbb{N}}^\Lambda)$, the space of continuous real functions on $\overline{\mathbb{N}}^\Lambda$. One has $\nu_n \Rightarrow \nu$ if and only if $\nu_n(\{x : x(i) = y(i)\ \forall i \in \Delta\}) \to \nu(\{x : x(i) = y(i)\ \forall i \in \Delta\})$ for all finite $\Delta \subset \Lambda$ and $y \in \mathbb{N}^\Delta$.

We equip the space $[0,1]^\Lambda$ with the product topology, and we say that probability measures μ_n on $[0,1]^\Lambda$ converge weakly to a limit μ, denoted as $\mu_n \Rightarrow \mu$, when $\int \mu_n(\mathrm{d}\phi) f(\phi) \to \int \mu(\mathrm{d}\phi) f(\phi)$ for every $f \in \mathcal{C}([0,1]^\Lambda)$.

(Monotone convergence) If ν_1, ν_2 are probability measures on $\overline{\mathbb{N}}^\Lambda$, then we say that ν_1 and ν_2 are stochastically ordered, denoted as $\nu_1 \leq \nu_2$, if $\overline{\mathbb{N}}^\Lambda$-valued random variables Y_1, Y_2 with laws $\mathcal{L}(Y_i) = \nu_i$ $(i = 1, 2)$ can be coupled such that $Y_1 \leq Y_2$. We say that a sequence of probability measures ν_n on $\mathbb{N}^\Lambda$ decreases (increases) stochastically to a limit ν, denoted as $\nu_n \downarrow \nu$ $(\nu_n \uparrow \nu)$, if random variables Y_n, Y with laws $\mathcal{L}(Y_n) = \nu_n$ and $\mathcal{L}(Y) = \nu$ can be coupled such that $Y_n \downarrow Y$ $(Y_n \uparrow Y)$. It is not hard to see that $\nu_n \downarrow \nu$ $(\nu_n \uparrow \nu)$ implies $\nu_n \Rightarrow \nu$. Stochastic ordering and monotone convergence of probability measures on $[0, 1]^\Lambda$ are defined in the same way.

(Finite systems) We denote the set of finite particle configurations by $\mathcal{N}(\Lambda) := \{x \in \mathbb{N}^\Lambda : |x| < \infty\}$ and let

$$\mathcal{S}(\mathcal{N}(\Lambda)) := \{f : \mathbb{N}^\Lambda \to \mathbb{R} : |f(x)| \leq K|x|^k + M \text{ for some } K, M, k \geq 0\} \tag{1.10}$$

denote the space of real functions on $\mathcal{N}(\Lambda)$ satisfying a polynomial growth condition. For finite initial conditions, the (a, b, c, d)-braco-process X is well-defined as a Markov process in $\mathcal{N}(\Lambda)$ (in particular, X does not explode), $f(X_t)$ is absolutely integrable for each $f \in \mathcal{S}(\mathcal{N}(\Lambda))$ and $t \geq 0$, and the semigroup

$$S_t f(x) := E^x[f(X_t)] \qquad (t \geq 0,\ x \in \mathcal{N}(\Lambda),\ f \in \mathcal{S}(\mathcal{N}(\Lambda))) \tag{1.11}$$

maps $\mathcal{S}(\mathcal{N}(\Lambda))$ into itself (see Proposition 8 below).

(Liggett-Spitzer space) Set $a_{\rm s}(i, j) := a(i, j) + a^\dagger(i, j)$. It follows from our assumptions on a that there exist (strictly) positive constants $(\gamma_i)_{i\in\Lambda}$ such that

$$\sum_i \gamma_i < \infty \quad \text{and} \quad \sum_j a_{\rm s}(i, j)\gamma_j \leq K\gamma_i \quad (i \in \Lambda) \tag{1.12}$$

for some $K < \infty$. We fix such $(\gamma_i)_{i\in\Lambda}$ throughout the paper and define the Liggett-Spitzer space (after [LS81])

$$\mathcal{E}_\gamma(\Lambda) := \{x \in \mathbb{N}^\Lambda : \|x\|_\gamma < \infty\}, \tag{1.13}$$

where for $x \in \mathbb{Z}^\Lambda$ we put

$$\|x\|_\gamma := \sum_i \gamma_i |x(i)|. \tag{1.14}$$

We let $\mathcal{C}_{\rm Lip}(\mathcal{E}_\gamma(\Lambda))$ denote the class of Lipschitz functions on $\mathcal{E}_\gamma(\Lambda)$, i.e., $f : \mathcal{E}_\gamma(\Lambda) \to \mathbb{R}$ such that $|f(x) - f(y)| \leq L\|x - y\|_\gamma$ for some $L < \infty$.

(Infinite systems) It is known ([Che87], see also Proposition 11 below) that for each $f \in \mathcal{C}_{\rm Lip}(\mathcal{E}_\gamma(\Lambda))$ and $t \geq 0$, the function $S_t f$ defined in (1.11) can be extended to a unique Lipschitz function on $\mathcal{E}_\gamma(\Lambda)$, also denoted by $S_t f$. Moreover, there exists a time-homogeneous Markov process X in $\mathcal{E}_\gamma(\Lambda)$ (also called (a, b, c, d)-braco-process) with transition laws given by

$$E^x[f(X_t)] = S_t f(x) \qquad (f \in \mathcal{C}_{\rm Lip}(\mathcal{E}_\gamma(\Lambda)),\ x \in \mathcal{E}_\gamma(\Lambda),\ t \geq 0). \tag{1.15}$$

We will show (in Proposition 11 below) that X has a version with componentwise cadlag sample paths, a fact that may seem obvious but to our knowledge has not been proved before.

(Survival and extinction) We say that the (a, b, c, d)-braco-process survives if

$$P^x[X_t \neq 0 \ \forall t \geq 0] > 0 \quad \text{for some} \quad x \in \mathcal{N}(\Lambda). \tag{1.16}$$

If X does not survive we say that X dies out. Note that the process with death rate $d = 0$ survives, since the number of particles can no longer decrease once only one particle is left. If Λ is finite then the (a, b, c, d)-braco-process survives if and only if $d = 0$, but for infinite Λ survival often holds also for some $d > 0$. We plan to discuss sufficient conditions for survival in a forthcoming paper.

(Nontrivial measures) We say that a probability measure ν on $\overline{\mathbb{N}}^\Lambda$ is nontrivial if $\nu(\{0\}) = 0$, where $0 \in \overline{\mathbb{N}}^\Lambda$ denotes the zero configuration. Likewise, we say that a probability measure μ on $[0, 1]^\Lambda$ is nontrivial if $\mu(\{0\}) = 0$.

(Homogeneous lattices) For some of our results, we will have to assume that the underlying motion (Λ, a) has some sort of translation invariant structure. For our purposes, the easiest way to do this is to assume that Λ can be equipped with a group structure in such a way that $a(ki, kj) = a(i, j)$ for all $i, j, k \in \Lambda$. In this case we will say that (Λ, a) is homogeneous. This is the case, for example, for nearest neighbor random walk on $\mathbb{Z}^d$ and $\mathbb{T}^d$ and for the random walks on the hierarchical group Ω_d mentioned at (1.2). (There are, in fact, several ways to equip the tree $\mathbb{T}^d$ with a group structure that is consistent with its graph structure.)

Define shift operators $T_i : \mathbb{N}^\Lambda \to \mathbb{N}^\Lambda$ by

$$T_i x(j) := x(ij) \qquad (i, j \in \Lambda, \ x \in \mathbb{N}^\Lambda). \tag{1.17}$$

We say that a probability measure ν on $\mathbb{N}^\Lambda$ is homogeneous if $\nu \circ T_i^{-1} = \nu$ for all $i \in \Lambda$.

1.4 Main results

Theorem 1 (Dualities and Poissonization) *Let X and $\mathcal{X}$ be the (a, b, c, d)-braco-process and the (a, b, c, d)-resem-process, respectively, and let $\mathcal{X}^\dagger$ denote the $(a^\dagger, b, c, d)$-resem-process. Then the following holds:*

(a) (Duality)

$$P^x[\mathrm{Thin}_\phi(X_t) = 0] = P^\phi[\mathrm{Thin}_{\mathcal{X}_t^\dagger}(x) = 0] \qquad (t \geq 0, \ \phi \in [0, 1]^\Lambda, \ x \in \mathcal{E}_\gamma(\Lambda)). \tag{1.18}$$

(b) (Self-duality) *Assume $c > 0$, then*

$$P^\phi[\mathrm{Pois}(\tfrac{b}{c}\mathcal{X}_t\psi) = 0] = P^\psi[\mathrm{Pois}(\tfrac{b}{c}\phi\mathcal{X}_t^\dagger) = 0] \qquad (t \geq 0, \ \phi, \psi \in [0, 1]^\Lambda). \tag{1.19}$$

(c) (Poissonization) *Assume $c > 0$, then*

$$P^{\mathcal{L}(\mathrm{Pois}(\frac{b}{c}\phi))}[X_t \in \cdot\,] = P^\phi[\mathrm{Pois}(\tfrac{b}{c}\mathcal{X}_t) \in \cdot\,] \qquad (t \geq 0, \ \phi \in [0, 1]^\Lambda), \tag{1.20}$$

i.e., if X is started in the initial law $\mathcal{L}(\mathrm{Pois}(\frac{b}{c}\phi))$ and $\mathcal{X}$ is started in ϕ, then X_t and $\mathrm{Pois}(\frac{b}{c}\mathcal{X}_t)$ are equal in law.

Note that $P[\mathrm{Thin}_\phi(x) = 0] = (1-\phi)^x$ ($\phi \in [0,1]^\Lambda$, $x \in \mathbb{N}^\Lambda$). Therefore, Theorem 1 (a) is just a reformulation of the duality relation (1.5). Theorem 1 (b) says that resampling-selection processes are in addition dual with respect to each other. In particular, if the underlying motion is symmetric, i.e., $a = a^\dagger$, then this is a self-duality. Since $P[\mathrm{Pois}(\phi) = 0] = e^{-|\phi|}$, formula (1.19) can be rewritten as

$$E^\phi\big[e^{-\frac{b}{c}\langle \mathcal{X}_t, \psi\rangle}\big] = E^\psi\big[e^{-\frac{b}{c}\langle \phi, \mathcal{X}^\dagger_t\rangle}\big] \qquad (t \geq 0,\ \phi, \psi \in [0,1]^\Lambda). \tag{1.21}$$

To convince the reader that the notation in (1.18) and (1.19), which may feel a little uneasy in the beginning, is convenient, we give here the proof of the Poissonization formula (1.20).

Proof of Theorem 1 (c) By (1.9) and the duality relations (1.18) and (1.19),

$$\begin{aligned} &P^{\mathcal{L}(\mathrm{Pois}(\frac{b}{c}\phi))}[\mathrm{Thin}_\psi(X_t) = 0] = P^\psi[\mathrm{Thin}_{\mathcal{X}^\dagger_t}(\mathrm{Pois}(\tfrac{b}{c}\phi)) = 0] \\ &\quad = P^\psi[\mathrm{Pois}(\tfrac{b}{c}\mathcal{X}^\dagger_t\phi) = 0] = P^\phi[\mathrm{Pois}(\tfrac{b}{c}\psi\mathcal{X}_t) = 0] = P^\phi[\mathrm{Thin}_\psi(\mathrm{Pois}(\tfrac{b}{c}\mathcal{X}_t)) = 0]. \end{aligned} \tag{1.22}$$

Since this is true for all $\psi \in [0,1]^\Lambda$, the random variables X_t and $\mathrm{Pois}(\frac{b}{c}\mathcal{X}_t)$ are equal in distribution. ∎

Our next result shows that it is possible to start the (a,b,c,d)-braco-process with infinitely many particles at each site. This result (except for parts (b) and (f)) has been proved for branching-coalescing particle systems on $\mathbb{Z}^d$, but with more general branching and coalescing mechanisms, in [DDL90]. Their methods are not restricted to the case $\Lambda = \mathbb{Z}^d$, but we will give an independent proof using duality, which has the additional appeal of yielding the explicit bound in part (b).

Theorem 2 (The maximal branching-coalescing process) *Assume that $c > 0$. Then there exists an $\mathcal{E}_\gamma(\Lambda)$-valued process $X^{(\infty)} = (X^{(\infty)}_t)_{t>0}$ with the following properties:*

(a) *For each $\varepsilon > 0$, $(X^{(\infty)}_t)_{t\geq\varepsilon}$ is the (a,b,c,d)-braco-process starting in $X^{(\infty)}_\varepsilon$.*

(b) *Set $r := b - d + c$. Then*

$$E[X^{(\infty)}_t(i)] \leq \begin{cases} \frac{r}{c(1-e^{-rt})} & \text{if } r \neq 0, \\ \frac{1}{ct} & \text{if } r = 0 \end{cases} \qquad (i \in \Lambda). \tag{1.23}$$

(c) *If $X^{(n)}$ are (a,b,c,d)-braco-processes starting in initial states $x^{(n)} \in \mathcal{E}_\gamma(\Lambda)$ such that*

$$x^{(n)}(i) \uparrow \infty \quad \text{as } n \uparrow \infty \qquad (i \in \Lambda), \tag{1.24}$$

then

$$\mathcal{L}(X^{(n)}_t) \uparrow \mathcal{L}(X^{(\infty)}_t) \quad \text{as } n \uparrow \infty \qquad (t > 0). \tag{1.25}$$

(d) *There exists an invariant measure $\overline{\nu}$ of the (a,b,c,d)-braco-process such that*

$$\mathcal{L}(X^{(\infty)}_t) \downarrow \overline{\nu} \quad \text{as } t \uparrow \infty. \tag{1.26}$$

(e) *If ν is another invariant measure for the (a,b,c,d)-braco-process, then $\nu \leq \overline{\nu}$.*

(f) *The measure $\overline{\nu}$ is uniquely characterised by*

$$\int \overline{\nu}(\mathrm{d}x)(1-\phi)^x = P^\phi[\exists t \geq 0 \text{ such that } \mathcal{X}_t^\dagger = 0] \qquad (\phi \in [0,1]^\Lambda), \tag{1.27}$$

where $\mathcal{X}^\dagger$ denotes the $(a^\dagger, b, c, d)$-resem-process.

We call $X^{(\infty)}$ the maximal (a,b,c,d)-braco process and we call $\overline{\nu}$ the upper invariant measure. To see why Theorem 2 (f) holds, note that by Theorem 1 (a) and Theorem 2 (c),

$$P[\mathrm{Thin}_\phi(X_t^{(\infty)}) = 0] = \lim_{n\uparrow\infty} P^\phi[\mathrm{Thin}_{\mathcal{X}^\dagger}(x^{(n)}) = 0] = P^\phi[\mathcal{X}_t^\dagger = 0] \qquad (\phi \in [0,1]^\Lambda,\ t > 0). \tag{1.28}$$

Now 0 is an absorbing state for the (a,b,c,d)-resem-process, and therefore $P^\phi[\mathcal{X}_t^\dagger = 0] = P^\phi[\exists s \leq t \text{ such that } \mathcal{X}_s^\dagger = 0]$. Therefore, taking the limit $t \uparrow \infty$ in (1.28) we arrive at (1.27).

The (a,b,c,d)-resem process has an upper invariant measure too. Of our next theorem, parts (a)–(c) are simple, but part (d) lies somewhat deeper.

Theorem 3 (The maximal resampling-selection process) *Let $\mathcal{X}^1$ denote the (a,b,c,d)-resem-process started in $\mathcal{X}_0^1(i) = 1$ $(i \in \Lambda)$. Then the following holds.*

(a) *There exists an invariant measure $\overline{\mu}$ of the (a,b,c,d)-resem process such that*

$$\mathcal{L}(\mathcal{X}_t^1) \downarrow \overline{\mu} \quad \text{as } t \uparrow \infty. \tag{1.29}$$

(b) *If μ is another invariant measure, then $\mu \leq \overline{\mu}$.*

(c) *Let $X^\dagger$ denote the $(a^\dagger, b, c, d)$-braco-process. Then*

$$\int \overline{\mu}(\mathrm{d}\phi)(1-\phi)^x = P^x[\exists t \geq 0 \text{ such that } X_t^\dagger = 0] \qquad (x \in \mathcal{N}(\Lambda)), \tag{1.30}$$

and the measure $\overline{\mu}$ is nontrivial if and only if the $(a^\dagger, b, c, d)$-braco-process survives.

(d) *Assume that $c > 0$ and that Λ is infinite. If $\mathcal{Y}$ is a random variable such that $\overline{\mu} = \mathcal{L}(\mathcal{Y})$, and $\overline{\nu}$ is the upper invariant measure of the (a,b,c,d)-braco-process, then $\overline{\nu} = \mathcal{L}(\mathrm{Pois}(\frac{b}{c}\mathcal{Y}))$. If $\overline{\mu}$ is nontrivial then so is $\overline{\nu}$.*

Note that $\int \overline{\mu}(\mathrm{d}\phi)(1-\phi)^x$ is the probability that x individuals, sampled from a population with resampling and selection in the equilibrium measure $\overline{\mu}$, all have defective genes.

The following is our main result.

Theorem 4 (Convergence to the upper invariant measure) *Assume that (Λ, a) is infinite and homogeneous and that $c > 0$. Let X be the (a,b,c,d)-braco process started in a homogeneous nontrivial initial law $\mathcal{L}(X_0)$. Then $\mathcal{L}(X_t) \Rightarrow \overline{\nu}$ as $t \to \infty$, where $\overline{\nu}$ is the upper invariant measure.*

1.5 Methods

A key ingredient in the proofs of Theorem 3 (d) and Theorem 4 is the following property of resampling-selection processes, which is of some interest on its own.

Lemma 5 (Extinction versus unbounded growth) *Assume that $c > 0$. Let $\mathcal{X}$ be the (a,b,c,d)-resem-process starting in an initial state $\phi \in [0,1]^\Lambda$ with $|\phi| < \infty$. Then $e^{-\frac{b}{c}|\mathcal{X}_t|}$ is a submartingale, and a martingale if $d = 0$. If moreover Λ is infinite, then*

$$\mathcal{X}_t = 0 \quad \text{for some } t \geq 0 \quad \text{or} \quad \lim_{t\to\infty} |\mathcal{X}_t| = \infty \quad \text{a.s.} \tag{1.31}$$

Note that by Theorem 1 (b),

$$E^\phi\big[e^{-\frac{b}{c}\langle \mathcal{X}_t, 1\rangle}\big] = E^1\big[e^{-\frac{b}{c}\langle \phi, \mathcal{X}^\dagger_t\rangle}\big] \geq e^{-\frac{b}{c}\langle \phi, 1\rangle} \qquad (\phi \in [0,1]^\Lambda), \tag{1.32}$$

with equality if $d = 0$, since 1 is a stationary state for the $(a^\dagger, b, c, 0)$-resem-process. This shows that $e^{-\frac{b}{c}|\mathcal{X}_t|}$ is a submartingale, and a martingale if $d = 0$. By submartingale convergence, $|\mathcal{X}_t|$ converges a.s. to a limit in $[0,\infty]$. All the hard work of Lemma 5 consists of proving that this limit is a.s. either 0 or ∞, and that $\mathcal{X}$ gets extinct in finite time if the limit is zero.

Once Lemma 5 is established the proof of Theorem 3 (d) is simple.

Proof of Theorem 3 (d) Let $\mathcal{Y}$ be a random variable such that $\overline{\mu} = \mathcal{L}(\mathcal{Y})$ and let Y be a random variable such that $\overline{\nu} = \mathcal{L}(Y)$. By (1.9), Theorem 1 (b), and Theorem 2 (f)

$$\begin{aligned} &P[\mathrm{Thin}_\phi(\mathrm{Pois}(\tfrac{b}{c}\mathcal{Y}) = 0] = \lim_{t\to\infty} P^1[\mathrm{Pois}(\tfrac{b}{c}\phi\mathcal{X}_t) = 0] = \lim_{t\to\infty} P^\phi[\mathrm{Pois}(\tfrac{b}{c}\mathcal{X}^\dagger_t) = 0] \\ &\stackrel{!}{=} P^\phi[\exists t \geq 0 \text{ such that } \mathcal{X}^\dagger_t = 0] = P[\mathrm{Thin}_\phi(Y) = 0], \end{aligned} \tag{1.33}$$

where we have used Lemma 5 in the equality marked with '!'. Since (1.33) holds for all $\phi \in [0,1]^\Lambda$, the random variables $\mathrm{Pois}(\frac{b}{c}\mathcal{Y})$ and Y are equal in distribution. By Lemma 5, $|\mathcal{Y}| \in \{0,\infty\}$ a.s. and therefore if $\overline{\mu}$ is nontrivial then $\mathcal{L}(\mathrm{Pois}(\frac{b}{c}\mathcal{Y}))$ is nontrivial. ∎

In view of Theorem 3 (d), it is natural to ask if for infinite lattices, every invariant law of the (a,b,c,d)-braco-process is the Poissonization of an invariant law of the (a,b,c,d)-resem-process. We do not know the answer to this question.

In order to give a very short proof of Theorem 4, we need one more lemma.

Lemma 6 (Systems with particles everywhere) *Assume that (Λ, a) is infinite and homogeneous. Let X be the (a,b,c,d)-braco process started in a homogeneous nontrivial initial law $\mathcal{L}(X_0)$. Then*

$$\lim_{n\to\infty} P[\mathrm{Thin}_{\phi_n}(X_1) = 0] = 0, \tag{1.34}$$

for all $\phi_n \in [0,1]^\Lambda$ satisfying $|\phi_n| \to \infty$.

Proof of Theorem 4 Let $\mathcal{X}^\dagger$ denote the $(a^\dagger, b, c, d)$-resem-process. By Theorem 1 (a), Lemmas 5 and 6, and Theorem 2 (f),

$$\begin{aligned} \lim_{t\to\infty} P[\mathrm{Thin}_\phi(X_t) = 0] &= \lim_{t\to\infty} P[\mathrm{Thin}_{\mathcal{X}^\dagger_{t-1}}(X_1) = 0] \\ &= P[\exists t \geq 0 \text{ such that } \mathcal{X}^\dagger_t = 0] = \int \overline{\nu}(\mathrm{d}x)\,(1-\phi)^x. \end{aligned} \tag{1.35}$$

Since this holds for all $\phi \in [0,1]^\Lambda$, it follows that $\mathcal{L}(X_t) \Rightarrow \overline{\nu}$. ∎

Of course, all the hard work is in proving Lemmas 5 and 6, as well as the more basic Theorems 1 and 2. The heart of the proof of Theorem 2 is the bound in part (b). We derive this bound using a 'duality' relation with a nonnegative error term, between the (a,b,c,d)-braco-process and a super random walk (Proposition 24). We call this relation a subduality. Theorem 2 (b) yields a lower bound on the finite time extinction probabilities of the (a,b,c,d)-resem-process started with small initial mass (Lemma 25, in particular formula (5.1)), which plays a key role in the proof of Lemma 5.

1.6 A spatial ergodic theorem

An essential ingredient in the proof of Lemma 6 is a spatial ergodic theorem for countable groups that need not be amenable. Roughly speaking, a lattice is amenable if the surface of large blocks is small compared to their volume. More precisely, a countable group Λ is called amenable if for every finite nonempty $B \subset \Lambda$ and $\varepsilon > 0$, there exists a finite $A \subset \Lambda$ such that $|BA \vartriangle A| \leq \varepsilon|A|$. Here $BA := \{ba : b \in B,\ a \in A\}$ and $\vartriangle$ denotes the symmetric difference. (See [Pat88, Theorem 4.13].) For example, $\mathbb{Z}^d$ is amenable but the regular tree $\mathbb{T}^d$ is not.

It is a common misconception that ergodic theorems hold on amenable groups only. For example, one could get this impression from Krengel's book on ergodic theorems [Kre85], which devotes a section (Section 6.4) to ergodic theorems on amenable groups but doesn't consider nonamenable groups. Indeed, block averages on nonamenable lattices do in general not converge to their ergodic mean, but it turns out that averages with respect to sufficiently smooth probability distributions always converge. Since this fact seems not to have been noticed before, we formulate here a spatial L^2-ergodic theorem for general countable groups.

Let $(E, \mathcal{B})$ be a measurable space and let Λ be a countable group with unit element 0. Define shift-operators as in (1.17) and let $\mathcal{T} := \{A \in \mathcal{B}^\Lambda : T_i^{-1}(A) = A\ \forall i \in \Lambda\}$ denote the σ-field of shift-invariant events. Denote by $L^2(\mu)$ the L^2-space of with respect to μ square integrable real functions, equipped with the L^2-norm $\|f\|_{\mu,2} := (\int |f|^2 \mathrm{d}\mu)^{\frac{1}{2}}$. Assume that $P : \Lambda \times \Lambda \to [0,1]$ satisfies $\sum_i P(0,i) = 1$ and $P(i,j) = P(ki,kj)$ for all $k \in \Lambda$, and that $\{i : P(0,i) > 0\}$ generates Λ. For any bounded real function ϕ on Λ, write $P\phi(i) := \sum_j P(i,j)\phi(j)$.

Theorem 7 (L^2-ergodic theorem) *Assume that μ is a homogeneous probability measure on E^Λ, $f \in L^2(\mu)$, and that π_n are probability distributions on Λ such that*

$$\lim_{n\to\infty} |\pi_n - P\pi_n| = 0. \tag{1.36}$$

Then

$$\sum_i \pi_n(i)\, f \circ T_i \longrightarrow E[f|\mathcal{T}] \quad \text{in } L^2(\mu) \quad \text{as } n \to \infty, \tag{1.37}$$

where $E[f|\mathcal{T}]$ denotes the conditional expectation (with respect to μ) of f given $\mathcal{T}$.

Probability distributions satisfying (1.36) always exist. For example, one can take:

$$\pi_n = \frac{1}{n}\sum_{k=1}^{n} P^k \pi, \tag{1.38}$$

where π is any probability distribution on Λ. (If the π_n are chosen as in (1.38), then the left-hand side of (1.37) is the expected average of f as seen during the first n steps of a random walk that steps from i to j with probability $P(j,i)$.) However, if the lattice is not amenable, it may not always be possible to take π_n of the form $\pi_n = \frac{1}{|\Delta_n|}1_{\Delta_n}$, where the Δ_n are finite subsets of Λ. Thus, block averages may not always be the right object to look at for the ergodic mean.

1.7 Discussion

Generalizing our model, let X be a process in a Liggett-Spitzer subspace of $\mathbb{N}^\Lambda$, with local jump rates

$$\begin{array}{ll} x \mapsto x + \delta_j - \delta_i & \text{with rate } a(i,j) \\ x \mapsto x + \delta_i & \text{with rate } \sum_{n=0}^{k} b_n x^{(n)}, \\ x \mapsto x - \delta_i & \text{with rate } \sum_{n=1}^{k+1} c_n x^{(n)}, \end{array} \tag{1.39}$$

where $x^{(0)} := 1$ and $x^{(n)} := x(x-1)\cdots(x-n+1)$ $(n \geq 1)$. In particular, the (a,b,c,d)-braco-process corresponds to the case $k = 1$, $b_0 = 0$, $b_1 = b$, $c_1 = d$, and $c_2 = c$. Processes with jump rates as in (1.39) are known as reaction-diffusion systems. It has been known for a long time that if the coefficients satisfy

$$a = a^\dagger \quad \text{and} \quad b_n = \lambda c_n \quad \text{for some} \quad \lambda \geq 0, \tag{1.40}$$

then $\mathcal{L}(\mathrm{Pois}(\lambda))$ is a reversible equilibrium for the corresponding reaction-diffusion system. Note that the (a,b,c,d)-braco-process satisfies (1.40) if and only if $a = a^\dagger$ and $d = 0$.

The ergodic behavior of reaction-diffusion systems on $\Lambda = \mathbb{Z}^d$ satisfying the reversibility condition (1.40) was studied by Ding, Durrett and Liggett in [DDL90]. For our model with $a = a^\dagger$ and $d = 0$ on $\mathbb{Z}^d$, they show that all homogeneous invariant measures are convex combinations of δ_0 and $\mathcal{L}(\mathrm{Pois}(\frac{b}{c}))$. Their proof uses the fact that for a large block in $\mathbb{Z}^d$, surface terms are small compared to volume terms, i.e., $\mathbb{Z}^d$ is amenable. Such arguments typically fail on nonamenable lattices such as trees, and therefore it is at least not immediately obvious if their methods can be generalized to such lattices. Our Theorem 4 shows that all homogeneous invariant measures of the (a,b,c,d)-braco-process are convex combinations of δ_0 and $\overline{\nu}$, also in the non-reversible case $d > 0$ and for nonamenable lattices. Thus, neither reversibility nor amenability are essential here.

On the other hand, we believe that amenability is essential for more subtle ergodic properties of reaction-diffusion processes. In analogy with the contact process, let us say that a reaction-diffusion process with $b_0 = 0$ exhibits complete convergence, if

$$P^x[X_t \in \cdot\,] \Rightarrow \rho(x)\overline{\nu} + (1-\rho(x))\delta_0 \quad \text{as} \quad t \to \infty \qquad (x \in \mathcal{N}(\Lambda)), \tag{1.41}$$

where $\rho(x) := P^x[X_t \neq 0 \ \forall t \geq 0]$ denotes the survival probability. It has been shown by Mountford [Mou92] that complete convergence holds for reversible reaction-diffusion systems on $\Lambda = \mathbb{Z}^d$ satisfying (1.40), $b_0 = 0$, and a first moment condition on a. We conjecture that complete convergence holds more generally if $a = a^\dagger$ and Λ is amenable, but not in general on nonamenable lattices. As a motivation for this conjecture, we note that complete convergence holds for the contact process on $\mathbb{Z}^d$ but not in general on $\mathbb{T}^d$; see Liggett [Lig99].

The self-duality of resampling-selection processes (Theorem 1 (b)) is reminiscent of the self-duality of the contact process. It is an interesting question whether our methods can be adapted to the contact process, to show that the upper invariant measure of the contact process on a countable group is the limit started from any homogeneous nontrivial initial law.

Other interesting processes that some of our techniques might be applied to are multitype branching-coalescing particle systems. For example, it seems natural to color the particles in a branching-coalescing particle system in two (or more) colors, with the rule that in coalescence of differently colored particles, the newly created particle chooses the color of one of its parents with equal probabilities (neutral selection) or with a prejudice towards one color (positive selection). More difficult questions refer to what happens when the two colors have different parameters b, c, d or even different underlying motions a.

Another interesting question is whether the techniques in this paper can be generalized to reaction-diffusion processes with higher-order branching and coalescence as in (1.39). It seems plausible that these systems have some sort of resampling-selection dual too, now with 'resampling' and 'selection' events involving three and more particles.

1.8 Outline

In Section 2 we construct (a, b, c, d)-braco-processes and (a, b, c, d)-resem-processes and prove some of their elementary properties, such as comparison, approximation with finite systems, moment estimates and martingale problems. Section 3 contains the proof of Theorem 1 and of the subduality between branching-coalescing particle systems and super random walks. In Section 4 we prove Theorems 2 and 3. In Section 5, finally, we prove Lemma 5, Theorem 7, and Lemma 6, thereby completing the proof of Theorem 4.

Acknowledgements We thank Klaus Fleischmann who played a stimulating role during the early stages of this project, Claudia Neuhauser for answering questions about branching-coalescing processes, and Olle Häggström for answering questions on nonamenable groups. Part of this work was carried out during the visits of Siva Athreya to the Weierstrass Institute for Applied Analysis and Stochastics, Berlin and to the Friedrich-Alexander University Erlangen-Nuremberg, and of Jan Swart to the Indian Statistical Institute, Delhi. We thank all these places for their kind hospitality.

2 Construction and Comparison

2.1 Finite branching-coalescing particle systems

For finite initial conditions, the (a,b,c,d)-braco-process X can be constructed explicitly using exponentially distributed random variables. The only thing one needs to check is that X does not explode. This is part of the next proposition. Recall the definitions of $\mathcal{N}(\Lambda)$ and $\mathcal{S}(\mathcal{N}(\Lambda))$ from (1.10) and of G from (1.1).

Proposition 8 (Finite braco-processes) *Let X be the (a,b,c,d)-braco-process started in a finite state x. Then X does not explode. Moreover, with $z^{\langle k\rangle} := z(z+1)\cdots(z+k-1)$, one has*

$$E^x\big[|X|_t^{\langle k\rangle}\big] \leq |x|^{\langle k\rangle} e^{kbt} \qquad (k=1,2,\ldots,\ t\geq 0). \tag{2.1}$$

For each $f\in\mathcal{S}(\mathcal{N}(\Lambda))$, one has $\int_0^t E[|Gf(X_s)|\mathrm{d}s<\infty$ for all $t\geq 0$ and the process $(M_t)_{t\geq 0}$ given by

$$M_t := f(X_t) - \int_0^t Gf(X_s)\mathrm{d}s \qquad (t\geq 0) \tag{2.2}$$

is a martingale with respect to the filtration generated by X.

Proof Introduce stopping times $\tau_N := \inf\{t\geq 0 : |X_t|\geq N\}$. Put $f_t^k(x) := |x|^{\langle k\rangle}e^{-kbt}$. It is easy to see that

$$\{G+\tfrac{\partial}{\partial t}\}f_t^k(x) \leq kb|x|^{\langle k\rangle}e^{-kbt} - kb|x|^{\langle k\rangle}e^{-kbt} = 0. \tag{2.3}$$

The stopped process $(X_{t\wedge\tau_N})_{t\geq 0}$ is a jump process in $\{x\in\mathbb{N}^\Lambda : |x|\leq N\}$ with bounded jump rates, and therefore standard theory tells us that the process $(M_t)_{t\geq 0}$ given by

$$M_t := f^k_{t\wedge\tau_N}(X_{t\wedge\tau_N}) - \int_0^{t\wedge\tau_N}\big(\{G+\tfrac{\partial}{\partial s}\}f_s^k\big)(X_s)\,\mathrm{d}s \qquad (t\geq 0) \tag{2.4}$$

is a martingale. By (2.3), it follows that $E^x\big[|X_{t\wedge\tau_N}|^{\langle k\rangle}e^{-kb(t\wedge\tau_N)}\big]\leq |x|^{\langle k\rangle}$ and therefore

$$E^x\big[|X_{t\wedge\tau_N}|^{\langle k\rangle}\big] \leq |x|^{\langle k\rangle}e^{kbt} \qquad (k=1,2,\ldots,\ t\geq 0). \tag{2.5}$$

In particular, setting $k=1$, we see that

$$NP^x[\tau_N\leq t] \leq E^x\big[|X_{t\wedge\tau_N}|\big] \leq |x|e^{bt} \qquad (t\geq 0), \tag{2.6}$$

which shows that $\lim_{N\to\infty}P^x[\tau_N\leq t]=0$ for all $t\geq 0$, i.e., the process does not explode. Taking the limit $N\uparrow\infty$ in (2.5), using Fatou, we arrive at (2.1).

If $f\in\mathcal{S}(\mathcal{N}(\Lambda))$ then f is bounded on sets of the form $\{x\in\mathbb{N}^\Lambda : |x|\leq N\}$, and therefore Gf is well-defined. By standard theory, the processes $(M_t^N)_{t\geq 0}$ given by

$$M_t^N := f(X_{t\wedge\tau_N}) - \int_0^{t\wedge\tau_N} Gf(X_s)\mathrm{d}s \qquad (t\geq 0) \tag{2.7}$$

are martingales. It is easy to see that $f\in\mathcal{S}(\mathcal{N}(\Lambda))$ implies $Gf\in\mathcal{S}(\mathcal{N}(\Lambda))$, and therefore $\int_0^t E[|Gf(X_s)|\mathrm{d}s<\infty$ for all $t\geq 0$ by (2.1). Using (2.5), one can now check that for fixed $t\geq 0$, the random variables $\{M_t^N\}_{N\geq 1}$ are uniformly integrable. Therefore, taking the pointwise limit in (2.7) we see that the process in (2.2) is a martingale. ∎

2.2 Monotonicity and subadditivity

In this section we prove two simple comparison results for finite branching-coalescing particle systems.

Lemma 9 (Comparison of branching-coalescing particle systems) *Let X and $\tilde{X}$ be the (a,b,c,d)-braco-process and the $(a,\tilde{b},\tilde{c},\tilde{d})$-braco-process started in finite initial states x and $\tilde{x}$, respectively. Assume that*

$$x \leq \tilde{x}, \quad b \leq \tilde{b}, \quad c \geq \tilde{c}, \quad d \geq \tilde{d}. \tag{2.8}$$

Then X and $\tilde{X}$ can be coupled in such a way that

$$X_t \leq \tilde{X}_t \qquad (t \geq 0). \tag{2.9}$$

Proof We will construct a bivariate process (B, W), say of black and white particles, such that $X = B$ are the black particles and $\tilde{X} = B + W$ are the black and white particles together. To this aim, we let the particles evolve in such a way that black and white particles branch with rates b and $\tilde{b}$, respectively, and additionally black particles give birth to white particles with rate $\tilde{b} - b$. Moreover, all pairs of particles coalesce with rate $2\tilde{c}$, where the new particle is black if at least one of its parents is black, and additionally each pair of black particles is with rate $2c - 2\tilde{c}$ replaced by a pair consisting of one black and one white particle. Finally, all particles die with rate $\tilde{d}$, and additionally, black particles change into white particles with rate $d - \tilde{d}$. It is easy to see that with these rules, X and $\tilde{X}$ are the (a,b,c,d)-braco-process and the $(a,\tilde{b},\tilde{c},\tilde{d})$-braco-process, respectively. ∎

Lemma 10 (Subadditivity) *Let X, Y, Z be (a,b,c,d)-braco-processes started in finite initial states x, y, and $x + y$, respectively. Then X, Y, Z may be coupled in such a way that X and Y are independent and*

$$Z_t \leq X_t + Y_t \qquad (t \geq 0). \tag{2.10}$$

Proof Consider a trivariate process (B, W, R), say of black, white and red particles, with initial condition $(x, y, 0)$, such that each color evolves as an autonomous (a,b,c,d)-braco-processes, and additionally, pairs of black and white particles are with rate $2c$ replaced by a black and a red particle, and pairs of white and red particles are with rate $2c$ replaced by one white particle. It is not hard to see that $X := B$ (black), $Y := W + R$ (white + red) and $Z := B + W$ (black + white) are (a,b,c,d)-braco-systems, and that X and Y are independent. ∎

2.3 Infinite branching-coalescing particle systems

In this section we carry out the construction of branching-coalescing particle systems for infinite initial conditions. We will also derive two results on the approximation of infinite systems with finite systems, that are needed later on. Except for the statement about sample paths, the next proposition has been proved in [Che87], but we give a proof here for the sake of completeness.

Proposition 11 (Construction of branching-coalescing particle systems) *For each $f \in \mathcal{C}_{\rm Lip}(\mathcal{E}_\gamma(\Lambda))$ and $t \geq 0$, the function $S_t f$ defined in (1.11) can be extended to a unique Lipschitz function on $\mathcal{E}_\gamma(\Lambda)$, also denoted by $S_t f$. There exists a unique (in distribution) time-homogeneous Markov process in $\mathcal{E}_\gamma(\Lambda)$, with componentwise cadlag sample paths, such that*

$$E^x[f(X_t)] = S_t f(x) \qquad (f \in \mathcal{C}_{\rm Lip}(\mathcal{E}_\gamma(\Lambda)),\ x \in \mathcal{E}_\gamma(\Lambda),\ t \geq 0). \tag{2.11}$$

We start with the following lemma.

Lemma 12 (Action of the semigroup on Lipschitz functions) *If $f : \mathcal{N}(\Lambda) \to \mathbb{R}$ is Lipschitz continuous in the norm $\|\cdot\|_\gamma$ from (1.14), with Lipschitz constant L, and K is the constant from (1.12), then*

$$|S_t f(x) - S_t f(y)| \leq L e^{(K+b-d)t} \|x - y\|_\gamma \qquad (x, y \in \mathcal{N}(\Lambda),\ t \geq 0). \tag{2.12}$$

Proof It follows from Propostion 8 that $\frac{\partial}{\partial t} E[f(X_t)] = E[Gf(X_t)]$ for all $f \in \mathcal{S}(\mathcal{N}(\Lambda))$, $t \geq 0$. Applying this to the function $f(x) := \|x\|_\gamma$ we see that

$$\begin{aligned} \tfrac{\partial}{\partial t} E^x[\|X_t\|_\gamma] = & \sum_{ij} a(i,j)(\gamma_j - \gamma_i) E[X_t(i)] + (b-d) E^x[\|X_t\|_\gamma] \\ & - c \sum_i \gamma_i E[X_t(i)(X_t(i) - 1)] \ \leq \ (K + b - d) E[\|X\|_\gamma], \end{aligned} \tag{2.13}$$

and therefore

$$E^x[\|X_t\|_\gamma] \leq e^{(K+b-d)t} \|x\|_\gamma \qquad (x \in \mathcal{N}(\Lambda)). \tag{2.14}$$

Let X^x denote the (a,b,c,d)-braco-process started in x. By Lemma 9, we can couple X^x, X^y, $X^{x\wedge y}$, and $X^{x\vee y}$ such that $X^{x\wedge y}_t \leq X^x_t, X^y_t \leq X^{x\vee y}_t$ for all $t \geq 0$. It follows that

$$E[\|X^x_t - X^y_t\|_\gamma] \leq E[\|X^{x\vee y}_t - X^{x\wedge y}_t\|_\gamma]. \tag{2.15}$$

By Lemma 10, we can couple $X^{x\wedge y}$ and $X^{x\vee y}$ to the process $X^{|x-y|}$ such that $X^{x\vee y}_t \leq X^{x\wedge y}_t + X^{|x-y|}_t$ for all $t \geq 0$. Therefore, by (2.15) and (2.14),

$$E[\|X^x_t - X^y_t\|_\gamma] \leq E[\|X^{|x-y|}_t\|_\gamma] \leq \|x - y\|_\gamma e^{(K+b-d)t}, \tag{2.16}$$

which implies that

$$|S_t f(x) - S_t f(y)| \leq E[|f(X^x_t) - f(X^y_t)|] \leq L E[\|X^x_t - X^y_t\|_\gamma] \leq L \|x - y\|_\gamma e^{(K+b-d)t}, \tag{2.17}$$

as required. ∎

Lipschitz functions on $\mathcal{N}(\Lambda)$ have a unique Lipschitz extension to $\mathcal{E}_\gamma(\Lambda)$. Thus, by Lemma 12, $S_t f$ can be uniquely extended to a function in $\mathcal{C}_{\rm Lip}(\mathcal{E}_\gamma(\Lambda))$ for each $f \in \mathcal{C}_{\rm Lip}(\mathcal{E}_\gamma(\Lambda))$.

Lemma 13 (Construction of the process for fixed times) *Let $X^{(n)}$ be (a,b,c,d)-braco-processes started in initial states $x^{(n)} \in \mathcal{N}(\Lambda)$ such that $x^{(n)} \uparrow x$ for some $x \in \mathcal{E}_\gamma(\Lambda)$. Then the $X^{(n)}$ may be coupled such that $X^{(n)} \uparrow X$ for some $\overline{\mathbb{N}}^\Lambda$-valued process $X = (X_t)_{t\geq 0}$. The process X satisfies $X_t \in \mathcal{E}_\gamma(\Lambda)$ a.s. $\forall t \geq 0$ and X is a Markov process with semigroup $(S_t)_{t\geq 0}$.*

Proof It follows from Lemma 9 that the $X^{(n)}$ can be coupled such that $X_t^{(n)} \leq X_t^{(n+1)}$ $(t \geq 0)$, and therefore $X_t^{(n)} \uparrow X_t$ $(t \geq 0)$ for some $\overline{\mathbb{N}}^\Lambda$-valued random variable X_t. By (2.16),

$$E\big[\|X_t - X_t^{(n)}\|_\gamma\big] = \lim_{m\uparrow\infty} E\big[\|X_t^{(m)} - X_t^{(n)}\|_\gamma\big] \leq \|x - x^{(n)}\|_\gamma e^{(K+b-d)t}. \tag{2.18}$$

This shows in particular that $E[\|X_t\|_\gamma] < \infty$ and therefore $X_t \in \mathcal{E}_\gamma(\Lambda)$ a.s. $\forall t \geq 0$. If $f \in \mathcal{C}_{\rm Lip}(\mathcal{E}_\gamma(\Lambda))$ has Lipschitz constant L, then by (2.18),

$$\begin{aligned} |E[f(X_t)] - E[f(X_t^{(n)})]| &\leq E[|f(X_t) - f(X_t^{(n)})|] \\ &\leq LE[\|X_t - X_t^{(n)}\|_\gamma] \leq L\|x - x^{(n)}\|_\gamma e^{(K+b-d)t}, \end{aligned} \tag{2.19}$$

and therefore

$$E[f(X_t)] = \lim_{n\uparrow\infty} E[f(X_t^{(n)})] = \lim_{n\uparrow\infty} S_t f(x^{(n)}) = S_t f(x). \tag{2.20}$$

This proves that for each $x \in \mathcal{E}_\gamma(\Lambda)$ and $t \geq 0$ there exists a probability measure $P_t(x,\cdot)$ on $\mathcal{E}_\gamma(\Lambda)$ such that $\int P_t(x,\mathrm{d}y)f(y) = S_t f(x)$ for all $f \in \mathcal{C}_{\rm Lip}(\mathcal{E}_\gamma(\Lambda))$. We need to show that X is the Markov process with transition probabilities $P_t(x,\mathrm{d}y)$. Let $\mathcal{C}_{\rm Lip,b}(\mathcal{E}_\gamma(\Lambda))$ denote the class of bounded Lipschitz functions on $\mathcal{E}_\gamma(\Lambda)$. Then $\mathcal{C}_{\rm Lip,b}(\mathcal{E}_\gamma(\Lambda))$ is closed under multiplication and S_t maps $\mathcal{C}_{\rm Lip,b}(\mathcal{E}_\gamma(\Lambda))$ into itself. Therefore, for all $0 \leq t_0 < \cdots < t_k$ and $f_1,\ldots,f_k \in \mathcal{C}_{\rm Lip,b}(\mathcal{E}_\gamma(\Lambda))$, one has

$$E\big[f_1(X_{t_1}^{(n)})\cdots f_k(X_{t_k}^{(n)})\big] = S_{t_1}f_1S_{t_2-t_1}f_2\cdots S_{t_k-t_{k-1}}f_k(x^{(n)}). \tag{2.21}$$

It follows from (2.18) that

$$\big|E\big[f_1(X_{t_1})\cdots f_k(X_{t_k})\big] - E\big[f_1(X_{t_1}^{(n)})\cdots f_k(X_{t_k}^{(n)})\big]\big| \leq \|x - x^{(n)}\|_\gamma \sum_{i=1}^k L_i e^{(K+b-d)t_k} \prod_{j\neq i} \|f_j\|_\infty, \tag{2.22}$$

where L_i is the Lipschitz constant of f_i. Taking the limit $n \uparrow \infty$ in (2.21), using (2.22), we see that

$$E\big[f_1(X_{t_1})\cdots f_k(X_{t_k})\big] = S_{t_1}f_1S_{t_2-t_1}f_2\cdots S_{t_k-t_{k-1}}f_k(x), \tag{2.23}$$

i.e., X is the Markov process with semigroup $(S_t)_{t\geq 0}$. ∎

Proof of Proposition 11 We need to show that the process X from Lemma 13 satisfies $X_t \in \mathcal{E}_\gamma(\Lambda)$ $\forall t \geq 0$ a.s. (and not just for fixed times) and that $(X_t(i))_{t\geq 0}$ has cadlag sample paths a.s. for each $i \in \Lambda$. It suffices to prove these facts on the time interval $[0,1]$. We will do this by constructing an $\mathcal{E}_\gamma(\Lambda)$-valued process Z such that Z makes only upward jumps, and the number of upward jumps of Z dominates the number of upward jumps of X.

Couple the process $X^{(n)}$ from Lemma 13 to a process $Y^{(n)}$ such that the joint process $(X^{(n)}, Y^{(n)})$ is the Markov process in $\mathcal{N}(\Lambda) \times \mathcal{N}(\Lambda)$ with generator

$$\begin{aligned}
&G_{X,Y} f(x,y) := \\
&\sum_{ij} a(i,j)x(i)\{f(x+\delta_j-\delta_i, y+\delta_i) - f(x,y)\} + \sum_{ij} a(i,j)y(i)\{f(x,y+\delta_j) - f(x,y)\} \\
&+b\sum_i x(i)\{f(x+\delta_i,y) - f(x,y)\} + b\sum_i y(i)\{f(x,y+\delta_i) - f(x,y)\} \\
&+c\sum_i x(i)(x(i)-1)\{f(x-\delta_i, y+\delta_i) - f(x,y)\} + d\sum_i x(i)\{f(x-\delta_i, y+\delta_i) - f(x,y)\}.
\end{aligned} \tag{2.24}$$

and initial state $(X_0^{(n)}, Y_0^{(n)}) = (x^{(n)}, 0)$. Indeed, it is not hard to see that the first component of the process with generator $G_{X,Y}$ is the (a,b,c,d)-braco-process, and that $Z^{(n)} := X^{(n)} + Y^{(n)}$ is the Markov process in $\mathcal{N}(\Lambda)$ with generator

$$G_Z f(z) := \sum_{ij} a(i,j)z(i)\{f(z+\delta_j) - f(z)\} + b\sum_i z(i)\{f(z+\delta_i) - f(z)\} \tag{2.25}$$

and initial state $Z_0^{(n)} = x^{(n)}$. In analogy with (2.14) it is easy to check that

$$E^z[\|Z_t^{(n)}\|_\gamma] \leq \|x^{(n)}\|_\gamma e^{(K+b)t} \qquad (z \in \mathcal{N}(\Lambda),\ t \geq 0). \tag{2.26}$$

$Z^{(n)}$ makes only upward jumps and $Z^{(n)}(i)$ makes at least as many upward jumps as $X^{(n)}(i)$. Since $X^{(n)}(i)$ cannot become negative, it follows that

$$|\{t \in [0,1] : X_{t-}^{(n)}(i) \neq X_t^{(n)}(i)\}| \leq x^{(n)}(i) + 2Z_1^{(n)}(i). \tag{2.27}$$

Summing with respect to the γ_i, taking expectations, using (2.26), we see that

$$\sum_i \gamma_i E\big[|\{t \in [0,1] : X_{t-}^{(n)}(i) \neq X_t^{(n)}(i)\}|\big] \leq \|x^{(n)}\|_\gamma (1 + 2e^{K+b}). \tag{2.28}$$

Let Z be the increasing limit of the processes $Z^{(n)}$. It follows from (2.26) that $Z_1 \in \mathcal{E}_\gamma(\Lambda)$ a.s. Now

$$X_t \leq Z_t \leq Z_1 \qquad \forall t \in [0,1] \quad \text{a.s.}, \tag{2.29}$$

and therefore $X_t \in \mathcal{E}_\gamma(\Lambda)$ $\forall t \in [0,1]$ a.s. Since a.s. all jumps occur at different times,

$$|\{t \in [0,1] : X_{t-}^{(n)}(i) \neq X_t^{(n)}(i)\}| \uparrow |\{t \in [0,1] : X_{t-}(i) \neq X_t(i)\}| \quad \text{as } n \uparrow \infty. \tag{2.30}$$

Thus, taking the limit $n \uparrow \infty$ in (2.28) we see that

$$\sum_i \gamma_i E\big[|\{t \in [0,1] : X_{t-}(i) \neq X_t(i)\}|\big] \leq \|x\|_\gamma (1 + 2e^{K+b}). \tag{2.31}$$

This proves that X has a.s. componentwise cadlag sample paths. ∎

The proof of Proposition 11 yields a useful corollary.

Corollary 14 (Locally finite number of jumps) *The* (a,b,c,d)*-braco-process* X *satisfies*

$$\sum_i \gamma_i E^x\big[|\{t\in[0,1]: X_{t-}(i)\neq X_t(i)\}|\big] \leq \|x\|_\gamma(1+2e^{K+b}). \tag{2.32}$$

We can now prove two approximation lemmas.

Lemma 15 (Convergence of finite dimensional distributions) *Let* X^{x_n}, X^x *be the* (a,b,c,d)*-braco-process started in initial states* $x_n, x\in\mathcal{E}_\gamma(\Lambda)$*, respectively, such that*

$$\lim_{n\to\infty}\|x_n-x\|_\gamma=0. \tag{2.33}$$

Then, for all $0\leq t_1<\cdots<t_k$*, one has*

$$(X^{(n)}_{t_1},\ldots,X^{(n)}_{t_k})\Rightarrow(X_{t_1},\ldots,X_{t_k})\qquad \text{as } n\to\infty. \tag{2.34}$$

Proof Use (2.23) for x_n and then let $n\to\infty$. ∎

Lemma 16 (Monotonicities for infinite systems) *Lemmas 9 and 10 also hold for infinite initial states. If* X^x, X^{x_n} *are* (a,b,c,d)*-braco-process started in initial states* $x, x_n\in\mathcal{E}_\gamma(\Lambda)$*, such that* $x_n\uparrow x$*, then* X^x, X^{x_n} *may be coupled such that*

$$X^{x_n}_t(i)\uparrow X^x_t(i)\quad \text{as } n\uparrow\infty\quad \forall i\in\Lambda,\ t\geq 0\quad \text{a.s.} \tag{2.35}$$

Proof The proof of Proposition 11 shows that (2.35) holds if the x_n are finite. To generalize Lemma 9 to infinite initial states $x,\tilde{x}$, it therefore suffices to note that if $x\leq\tilde{x}$, then there exist finite $x_n\leq\tilde{x}_n$ such that $x_n\uparrow x$ and $\tilde{x}_n\uparrow\tilde{x}$, and then take the limit $n\uparrow\infty$ in (2.9) using (2.35). Lemma 10 can be generalized to infinite x,y by approximation with finite x_n,y_n in the same way. Finally, to see that (2.35) remains valid if the x_n are infinite, note that by Lemma 9 (which has now been proved in the infinite case), the processes X^{x_n} can be coupled such that $X^{x_n}_t(i)\leq X^{x_{n+1}}_t(i)$ for all $i\in\Lambda$ and $t\geq 0$. Denote the increasing limit of the X^{x_n} by X^x. Lemma 15 shows that X^x has the same finite dimensional distributions as the (a,b,c,d)-braco-process started in x and it follows from Corollary 14 that X^x has componentwise cadlag sample paths. ∎

2.4 Construction and comparison of resampling-selection processes

We equip the space $[0,1]^\Lambda$ with the product topology and let $\mathcal{C}([0,1]^\Lambda)$ denote the space of continuous real functions on $[0,1]^\Lambda$, equipped with the supremum norm. By $\mathcal{C}^2_{\rm fin}([0,1]^\Lambda)$ we denote the space of $\mathcal{C}^2$ functions on $[0,1]^\Lambda$ depending on finitely many coordinates. By definition, $\mathcal{C}^2_{\rm sum}([0,1]^\Lambda)$ is the space of continuous functions f on $[0,1]^\Lambda$ such that the partial derivatives $\frac{\partial}{\partial\phi(i)}f(\phi)$ and $\frac{\partial^2}{\partial\phi(i)\partial\phi(j)}f(\phi)$ exist for each $x\in(0,1)^\Lambda$ and such that the functions

$$\phi\mapsto\big(\tfrac{\partial}{\partial\phi(i)}f(\phi)\big)_{i\in\Lambda}\quad\text{and}\quad\phi\mapsto\big(\tfrac{\partial^2}{\partial\phi(i)\partial\phi(j)}f(\phi)\big)_{i,j\in\Lambda} \tag{2.36}$$

can be extended to continuous functions from $[0,1]^\Lambda$ into the spaces $\ell^1(\Lambda)$ and $\ell^1(\Lambda^2)$ of absolutely summable sequences on Λ and Λ^2, respectively, equipped with the ℓ^1-norm. Define an operator $\mathcal{G}:\mathcal{C}^2_{\rm sum}([0,1]^\Lambda)\to\mathcal{C}([0,1]^\Lambda)$ by

$$\begin{aligned}\mathcal{G}f(\phi):=&\sum_{ij}a(j,i)(\phi(j)-\phi(i))\tfrac{\partial}{\partial\phi(i)}f(\phi)+b\sum_i\phi(i)(1-\phi(i))\tfrac{\partial}{\partial\phi(i)}f(\phi)\\&+c\sum_i\phi(i)(1-\phi(i))\tfrac{\partial^2}{\partial\phi(i)^2}f(\phi)-d\sum_i\phi(i)\tfrac{\partial}{\partial\phi(i)}f(\phi)\qquad(\phi\in[0,1]^\Lambda).\end{aligned}\tag{2.37}$$

One can check that for $f\in\mathcal{C}^2_{\rm sum}([0,1]^\Lambda)$, the infinite sums converge in the supremumnorm and the result does not depend on the summation order [Swa99, Lemma 3.4.4]. We say that a $[0,1]^\Lambda$-valued process $\mathcal{X}$ solves the martingale problem for $\mathcal{G}$, if $\mathcal{X}$ has cadlag sample paths and for each $f\in\mathcal{C}^2_{\rm sum}([0,1]^\Lambda)$, the process $(M_t)_{t\geq0}$ defined by

$$M_t:=f(\mathcal{X}_t)-\int_0^t\mathcal{G}f(\mathcal{X}_s)\mathrm{d}s\qquad(t\geq0)\tag{2.38}$$

is a martingale with respect to the filtration generated by $\mathcal{X}$. It suffices to check this condition for all $f\in\mathcal{C}_{\rm fin}([0,1]^\Lambda)$ (see [Swa99, Lemma 3.4.5]).

Let $\mathcal{C}_{[0,1]^\Lambda}[0,\infty)$ denote the space of continuous functions from $[0,\infty)$ into $[0,1]^\Lambda$, equipped with the topology of uniform convergence on compacta. If $\mathcal{X}^{(n)},\mathcal{X}$ are $\mathcal{C}_{[0,1]^\Lambda}[0,\infty)$-valued random variables, then we say that $\mathcal{X}^{(n)}$ converges in distribution to $\mathcal{X}$, denoted as $\mathcal{X}^{(n)}\Rightarrow\mathcal{X}$, when $\mathcal{L}(\mathcal{X}^{(n)})$ converges weakly to $\mathcal{L}(\mathcal{X})$. Convergence in distribution implies convergence of the finite-dimensional distributions (see [EK86, Theorem 3.7.8]). The fact that a $\mathcal{C}_{[0,1]^\Lambda}[0,\infty)$-valued random variable $\mathcal{X}$ solves the martingale problem for $\mathcal{G}$ is a property of the law of $\mathcal{X}$ only. Standard results from [EK86] yield the following (for the details, see for example Lemma 4.1 in [Swa00]):

Lemma 17 (Existence and compactness of solutions to the martingale problem) *For each $\phi\in[0,1]^\Lambda$, there exists a solution $\mathcal{X}$ to the martingale problem for $\mathcal{G}$ with initial state $\mathcal{X}_0=\phi$, and each solution to the martingale problem for $\mathcal{G}$ has continuous sample paths. Moreover, the space $\{\mathcal{L}(\mathcal{X}):\mathcal{X}$ solves the martingale problem for $\mathcal{G}\}$ is compact in the topology of weak convergence.*

If $\mathcal{X}$ solves the SDE (1.3), then $\mathcal{X}$ solves the martingale problem for $\mathcal{G}$. Conversely, each solution to the martingale problem for $\mathcal{G}$ is equal in distribution to some (weak) solution of the SDE (1.3). Thus, existence of (weak) solutions to (1.3) follows from Lemma 17. Distribution uniqueness of solutions to (1.3) follows from pathwise uniqueness, which is in turn implied by the following comparison result.

Lemma 18 (Monotone coupling of linearly interacting diffusions) *Let $I\subset\mathbb{R}$ be a closed interval, let $\sigma:I\to\mathbb{R}$ be Hölder-$\frac{1}{2}$-continuous, and let $b_1,b_2:I\to\mathbb{R}$ Lipschitz continuous functions such that $b_1\leq b_2$. Let $\mathcal{X}^\alpha$ ($\alpha=1,2$) be solutions, relative to the same system of Brownian motions, of the SDE*

$$\mathrm{d}\mathcal{X}^\alpha_t(i)=\sum_j a(j,i)(\mathcal{X}^\alpha_t(j)-\mathcal{X}^\alpha_t(i))\mathrm{d}t+b_\alpha(\mathcal{X}^\alpha_t(i))\mathrm{d}t+\sigma(\mathcal{X}^\alpha_t(i))\mathrm{d}B_t(i).\tag{2.39}$$

$(i \in \Lambda,\ t \geq 0)$. *Then*

$$\mathcal{X}_0^1 \leq \mathcal{X}_0^2 \quad \textit{implies} \quad \mathcal{X}_t^1 \leq \mathcal{X}_t^2 \quad \forall t \geq 0 \quad \text{a.s.} \tag{2.40}$$

Proof We adapt a technique due to Yamada and Watanabe [YW71] to the infinite dimensional case, very much in the spirit of Theorem 3.2 in [SS80]. Set $\Delta_t(i) := \mathcal{X}_t^1(i) - \mathcal{X}_t^2(i)$ and write $x^+ := x \vee 0$. We will show that $\Delta_0^+(i) = 0$ $(i \in \Lambda)$ implies $\Delta_t^+(i) = 0$ for all $t \geq 0$ and $i \in \Lambda$.

Since $\int_{0+} \frac{\mathrm{d}x}{x} = \infty$, it is not hard to see that we can choose continuous functions $\rho_n : \mathbb{R} \to \mathbb{R}$ such that

$$0 \leq \rho_n(x) \leq \frac{1}{nx} 1_{(0,\frac{1}{n})}(x) \quad \text{and} \quad \int_0^{\frac{1}{n}} \mathrm{d}x\, \rho_n(x) = 1. \tag{2.41}$$

Define twice continuously differentiable functions ϕ_n by

$$\phi_n(x) := \int_0^x \mathrm{d}y \int_0^y \mathrm{d}z\, \rho_n(z) \qquad (x \in \mathbb{R}). \tag{2.42}$$

Then,

$$\begin{aligned} &\text{(i)} \quad 0 \leq \phi_n(x) \uparrow x^+ && \text{as } n \uparrow \infty, \\ &\text{(ii)} \quad 0 \leq \phi_n'(x) \uparrow 1_{\{x>0\}} && \text{as } n \uparrow \infty, \\ &\text{(iii)} \quad 0 \leq |x|\phi_n''(x) \leq \tfrac{1}{n}. \end{aligned} \tag{2.43}$$

Now

$$\mathrm{d}\Delta_t(i) = \sum_j a(j,i)(\Delta_t(j) - \Delta_t(i))\mathrm{d}t + \big(b_1(\mathcal{X}_t^1(i)) - b_2(\mathcal{X}_t^2(i))\big)\mathrm{d}t + \big(\sigma(\mathcal{X}_t^1(i)) - \sigma(\mathcal{X}_t^2(i))\big)\mathrm{d}B_t(i). \tag{2.44}$$

Itô's formula gives

$$\begin{aligned} \mathrm{d}\phi_n(\Delta_t(i)) = &\sum_j a(j,i)(\Delta_t(j) - \Delta_t(i))\phi_n'(\Delta_t(i))\mathrm{d}t + \big(b_1(\mathcal{X}_t^1(i)) - b_2(\mathcal{X}_t^2(i))\big)\phi_n'(\Delta_t(i))\mathrm{d}t \\ &+ \tfrac{1}{2}\big(\sigma(\mathcal{X}_t^1(i)) - \sigma(\mathcal{X}_t^2(i))\big)^2 \phi_n''(\Delta_t(i))\mathrm{d}t \quad + \quad \text{martingale terms.} \end{aligned} \tag{2.45}$$

Since σ is Hölder-$\frac{1}{2}$-continuous, by (2.43) (iii)

$$\big(\sigma(\mathcal{X}_t^1(i)) - \sigma(\mathcal{X}_t^2(i))\big)^2 \phi_n''(\Delta_t(i)) \leq K|\Delta_t(i)|\phi_n''(\Delta_t(i)) \leq \frac{1}{n}. \tag{2.46}$$

Taking expectations in (2.45) and letting $n \uparrow \infty$, using (2.43), we get

$$\begin{aligned} E[\Delta_t^+(i)] = &\int_0^t E\big[\sum_j a(j,i)(\Delta_s(j) - \Delta_s(i))1_{\{\Delta_s(i)>0\}}\mathrm{d}s\big] \\ &+ \int_0^t E\big[\big(b_1(\mathcal{X}_s^1(i)) - b_2(\mathcal{X}_s^2(i))\big)1_{\{\Delta_s(i)>0\}}\mathrm{d}s\big] \\ \leq &\int_0^t E\big[\sum_j a(j,i)\Delta_s^+(j)\mathrm{d}s\big] + L\int_0^t E\big[\Delta_s^+(i)\mathrm{d}s\big], \end{aligned} \tag{2.47}$$

Here we have used that $E[\Delta_0^+(i)] = 0$, that $(\Delta_s(j) - \Delta_s(i))1_{\{\Delta_s(i)>0\}}$ can only be positive when $\Delta_s(j) > \Delta_s(i) > 0$, and that $\big(b_1(\mathcal{X}_s^1(i)) - b_2(\mathcal{X}_s^2(i))\big)1_{\{\Delta_s(i)>0\}}$ can only be positive when $\mathcal{X}_s^1(i) > \mathcal{X}_s^2(i)$, in which case

$$b_1(\mathcal{X}_s^1(i)) - b_2(\mathcal{X}_s^2(i)) \leq b_2(\mathcal{X}_s^1(i)) - b_2(\mathcal{X}_s^2(i)) \leq L|\mathcal{X}_s^1(i) - \mathcal{X}_s^2(i)| = L\Delta_s^+(i), \tag{2.48}$$

where L is the Lipschitz-constant of b_2. Recall the norm $\|\cdot\|_\gamma$ from (1.14) and let K be the constant from (1.12). It is easy to see that (2.47) implies

$$E[\|\Delta_t^+\|_\gamma] \leq (K+L)\int_0^t E[\|\Delta_s^+\|_\gamma]\mathrm{d}s. \tag{2.49}$$

The result now follows from Gronwall's lemma. ∎

Corollary 19 (Comparison of resampling-selection processes) *Assume that $\mathcal{X}, \tilde{\mathcal{X}}$ are solutions to the SDE (1.3), relative to the same collection of Brownian motions, with parameters (a,b,c,d) and $(a,\tilde{b},c,\tilde{d})$ and starting in initial states $\phi, \tilde{\phi}$, respectively. Assume that*

$$\phi \leq \tilde{\phi}, \quad d - b \geq \tilde{d} - \tilde{b}, \quad d \geq \tilde{d}. \tag{2.50}$$

Then

$$\mathcal{X}_t \leq \tilde{\mathcal{X}}_t \quad \forall t \geq 0 \quad \text{a.s.} \tag{2.51}$$

Proof Immediate from Lemma 18 and the fact that by (2.50), $bx(1-x) - dx \leq \tilde{b}x(1-x) - \tilde{d}x$ for all $x \in [0,1]$. ∎

Our next lemma shows that resampling-selection processes with finite initial mass have finite mass at all later times. The estimate (2.52) is not very good if $b - d < 0$, but it suffices for our purposes.

Lemma 20 (Summable resampling-selection processes) *Let $\mathcal{X}$ be the (a,b,c,d)-resem-process started in $x \in [0,1]^\Lambda$ with $|x| < \infty$. Set $r := (b-d) \vee 0$. Then, for all $t \geq 0$,*

$$E^x\big[|\mathcal{X}_t|\big] \leq |x|e^{rt}, \tag{2.52}$$

and $|\mathcal{X}_t| < \infty \ \forall t \geq 0$ a.s.

Proof Without loss of generality we may assume that $b \geq d$; otherwise, using Corollary 19, we can bound $\mathcal{X}$ from above by a braco-process with a higher b. Set $r := b - d$ and put $\mathcal{Y}_t(i) := \mathcal{X}_t(i)e^{-rt}$. By Itô's formula,

$$\mathrm{d}\mathcal{Y}_t(i) = \sum_j a(j,i)(\mathcal{Y}_t(j) - \mathcal{Y}_t(i))\,\mathrm{d}t - be^{-rt}\mathcal{X}_t(i)^2\mathrm{d}t + e^{-rt}\sqrt{c\mathcal{X}_t(i)(1-\mathcal{X}_t(i))}\,\mathrm{d}B_t(i). \tag{2.53}$$

Set $\tau_N := \inf\{t \geq 0 : |\mathcal{X}_t| \geq N\}$. Integrate (2.53) up to $t \wedge \tau_N$ and sum over i. The motion terms yield

$$\begin{aligned}&\int_0^{t\wedge\tau_N}\sum_{ij} a(j,i)(\mathcal{Y}_s(j) - \mathcal{Y}_s(i))\,\mathrm{d}s\\ &\quad = \int_0^{t\wedge\tau_N}\sum_j \big(\sum_i a(j,i)\big)\mathcal{Y}_s(j)\,\mathrm{d}s - \int_0^{t\wedge\tau_N}\sum_i \big(\sum_j a^\dagger(i,j)\big)\mathcal{Y}_s(i)\,\mathrm{d}s = 0,\end{aligned} \tag{2.54}$$

where the infinite sums converge in a bounded pointwise way since $|Y_s| \leq N$ for $s \leq \tau_N$. It follows that

$$|\mathcal{Y}_{t\wedge\tau_N}| = |x| - b\sum_i \int_0^{t\wedge\tau_N} \mathcal{X}_s(i)^2 e^{-rs}\mathrm{d}s + \sum_i \int_0^{t\wedge\tau_N} \sqrt{c\mathcal{X}_s(i)(1-\mathcal{X}_s(i))}\, e^{-rs}\mathrm{d}B_s(i), \quad (2.55)$$

provided we can show that the infinite sum of stochastic integrals converges. Indeed, for any finite $\Delta \subset \Lambda$, by the Itô isometry,

$$\begin{aligned} &\sum_{i\in\Delta} E\Big[\Big|\int_0^{t\wedge\tau_N} \sqrt{c\mathcal{X}_s(i)(1-\mathcal{X}_s(i))}\, e^{-rs}\mathrm{d}B_s(i)\Big|^2\Big] \\ &\quad = c\sum_{i\in\Delta} E\Big[\int_0^{t\wedge\tau_N} \mathcal{X}_s(i)(1-\mathcal{X}_s(i))e^{-2rs}\mathrm{d}s\Big] \leq cE\Big[\int_0^{t\wedge\tau_N} |\mathcal{X}_s|\mathrm{d}s\Big] \leq ctN, \end{aligned} \quad (2.56)$$

which shows that the stochastic integrals in (2.55) are absolutely summable in L^2-norm. It follows from (2.55) that

$$E^x[|\mathcal{X}_{t\wedge\tau_N}|]e^{-rt} \leq E^x[|\mathcal{X}_{t\wedge\tau_N}|e^{-r(t\wedge\tau_N)}] = E^x[|\mathcal{Y}_{t\wedge\tau_N}|] \leq |x|. \quad (2.57)$$

Now $NP^x[\tau_N \leq t] \leq |x|e^{rt}$ for all $t \geq 0$, which shows that $\tau_N \uparrow \infty$ as $N \uparrow \infty$ a.s. Letting $N \uparrow \infty$ in (2.57) we arrive at (2.52). ∎

We conclude this section with two results on the continuity of $\mathcal{X}$ in its initial state.

Lemma 21 **(Convergence in law)** *Assume that $\mathcal{X}^{(n)}, \mathcal{X}$ are (a,b,c,d)-resem-processes, started in $x^{(n)}, x \in [0,1]^\Lambda$, respectively. Then $x^{(n)} \to x$ implies $\mathcal{X}^{(n)} \Rightarrow \mathcal{X}$.*

Proof By Lemma 17, the laws $\mathcal{L}(\mathcal{X}^{(n)})$ are tight and each cluster point of the $\mathcal{L}(\mathcal{X}^{(n)})$ solves the martingale problem for $\mathcal{G}$ with initial state x. Therefore, by uniqueness of solutions to the martingale problem, $\mathcal{X}^{(n)} \Rightarrow \mathcal{X}$. ∎

Lemma 22 **(Monotone convergence)** *Let $\mathcal{X}^{(n)}, \mathcal{X}$ be (a,b,c,d)-resem-processes started in $x^{(n)}, x \in [0,1]^\Lambda$, respectively, such that*

$$x^{(n)} \uparrow x \qquad \text{as} \quad n \uparrow \infty. \quad (2.58)$$

Then $\mathcal{X}^{(n)}, \mathcal{X}$ may be defined on the same probablity space such that

$$\mathcal{X}^{(n)}_t(i) \uparrow \mathcal{X}_t(i) \qquad \forall i \in \Lambda,\ t \geq 0 \quad \text{as} \quad n \uparrow \infty \quad \text{a.s.} \quad (2.59)$$

Proof Let $\mathcal{X}^{(n)}, \mathcal{X}$ be solutions of the SDE (1.3) relative to the same system of Brownian motions. By Corollary 19, $\mathcal{X}^{(n)} \leq \mathcal{X}^{(n+1)}$ and $\mathcal{X}^{(n)} \leq \mathcal{X}$ for all n. Write $\Delta^{(n)}_t := \mathcal{X}_t - \mathcal{X}^{(n)}_t$ and set $\tau^{(n)}_\varepsilon := \inf\{t \geq 0 : \Delta^{(n)}_t \geq \varepsilon\}$. A calculation as in (2.47) shows that

$$\mathrm{d}\|\Delta^{(n)}_t\|_\gamma \leq (K+L)\|\Delta^{(n)}_t\|\mathrm{d}t \quad + \quad \text{martingale terms.} \quad (2.60)$$

It follows that

$$E\big[\|\Delta^{(n)}_{t\wedge\tau^{(n)}_\varepsilon}\|_\gamma\big] \leq \|x - x^{(n)}\|_\gamma e^{(K+L)t}. \quad (2.61)$$

Now $\varepsilon P[\tau^{(n)}_\varepsilon \leq t] \leq \|x - x^{(n)}\|_\gamma e^{(K+L)t}$ from which we conclude that $\tau^{(n)}_\varepsilon \uparrow \infty$ as $n \uparrow \infty$ for every $\varepsilon > 0$. ∎

3 Duality

3.1 Duality with error term

Before we prove the two duality formulas in Theorem 1 (a) and (b), as well as the subduality between branching-coalescing particle systems and super random walks, we formulate a general theorem giving sufficient conditions for two martingale problems to be dual to each other possible error term. Although the techniques for proving Theorem 23 below are well-known (see, for example, [EK86, Section 4.4]), we don't know a good reference for it.

If E be a metrizable space, we denote by $M(E), B(E)$ the spaces of real Borel measurable and bounded real Borel measurable functions on E, respectively. If A is a linear operator from a domain $\mathcal{D}(A) \subset B(E)$ into $M(E)$ and X is an E-valued process with cadlag sample paths, then we say that X solves the martingale problem for A if for each $f \in \mathcal{D}(A)$,

$$\int_0^t E\big[|Af(X_s)|\big]\mathrm{d}s < \infty \qquad (t \geq 0), \tag{3.1}$$

and the process $(M_t)_{t\geq 0}$ defined by

$$M_t := f(X_t) - \int_0^t Af(X_s)\mathrm{d}s \qquad (t \geq 0) \tag{3.2}$$

is a martingale with respect to the filtration generated by X.

Theorem 23 (Duality with error term) *Assume that E_1, E_2 are metrizable spaces and that for $i = 1, 2$, A_i is a linear operator from a domain $\mathcal{D}(A_i) \subset B(E_i)$ into $M(E_i)$. Assume that $\Psi \in B(E_1 \times E_2)$ satisfies $\Psi(\cdot, x_2) \in \mathcal{D}(A_1)$ and $\Psi(x_1, \cdot) \in \mathcal{D}(A_2)$ for each $x_1 \in E_1$ and $x_2 \in E_2$, and that*

$$\Phi_1(x_1, x_2) := A_1\Psi(\cdot, x_2)(x_1) \quad \textit{and} \quad \Phi_2(x_1, x_2) := A_2\Psi(x_1, \cdot)(x_2) \qquad (x_1 \in E_1,\ x_2 \in E_2) \tag{3.3}$$

are jointly measurable in x_1 and x_2. Assume that X^1 and X^2 are independent solutions to the martingale problems for A_1 and A_2, respectively, and that

$$\int_0^T \mathrm{d}s \int_0^T \mathrm{d}t\, E\big[|\Phi_i(X^1_s, X^2_t)|\big] < \infty \qquad (T \geq 0,\ i = 1, 2). \tag{3.4}$$

Then

$$E[\Psi(X^1_T, X^2_0)] - E[\Psi(X^1_0, X^2_T)] = \int_0^T \mathrm{d}t\, E[R(X^1_t, X^2_{T-t})] \qquad (T \geq 0), \tag{3.5}$$

where $R(x_1, x_2) := \Phi_1(x_1, x_2) - \Phi_2(x_1, x_2) \quad (x_1 \in E_1,\ x_2 \in E_2)$.

Proof Put

$$F(s, t) := E[\Psi(X^1_s, X^2_t)] \qquad (s, t \geq 0). \tag{3.6}$$

Then, for each $T > 0$,

$$\begin{aligned}\int_0^T \mathrm{d}t \,\big\{F(t,0) - F(0,t)\big\} &= \int_0^T \mathrm{d}t \,\big\{F(T-t,t) - F(0,t) - F(T-t,t) + F(t,0)\big\} \\ &= \int_0^T \mathrm{d}t \,\big\{F(T-t,t) - F(0,t)\big\} - \int_0^T \mathrm{d}t \,\big\{F(t,T-t) - F(t,0)\big\},\end{aligned} \tag{3.7}$$

where we have subsituted $t \mapsto T-t$ in the term $-F(T-t,t)$. Since X^1 solves the martingale problem for A_1,

$$E\big[\Psi(X^1_{T-t}, x_2)\big] - E\big[\Psi(X^1_0, x_2)\big] = \int_0^{T-t} \mathrm{d}s \; E\big[\Phi_1(X^1_s, x_2)\big] \qquad (x_2 \in E_2), \tag{3.8}$$

and therefore, integrating the x_2-variable with respect to the law of X^2_t, using the independence of X^1 and X^2 and (3.4), we find that

$$\begin{aligned}\int_0^T \mathrm{d}t \,\big\{F(T-t,t) - F(0,t)\big\} &= \int_0^T \mathrm{d}t \,\big\{E\big[\Psi(X^1_{T-t}, X^2_t)\big] - E\big[\Psi(X^1_0, X^2_t)\big]\big\} \\ = \int_0^T \mathrm{d}t \int_0^{T-t} \mathrm{d}s \; E\big[\Phi_1(X^1_s, X^2_t)\big] &= \int_0^T \mathrm{d}t \int_0^t \mathrm{d}s \; E\big[\Phi_1(X^1_{t-s}, X^2_s)\big].\end{aligned} \tag{3.9}$$

Treating the second term in the right-hand side of (3.7) in the same way, we find that

$$\int_0^T \mathrm{d}t \,\big\{F(t,0) - F(0,t)\big\} = \int_0^T \mathrm{d}t \int_0^t \mathrm{d}s \; E\big[\Phi_1(X^1_{t-s}, X^2_s)\big] - \int_0^T \mathrm{d}t \int_0^t \mathrm{d}s \; E\big[\Phi_2(X^1_{t-s}, X^2_s)\big]. \tag{3.10}$$

Differentiating with respect to T we arrive at (3.5). ∎

3.2 Duality and self-duality

Proof of Theorem 1 (a) We first prove the statement for finite x. We apply Theorem 23. Our duality function is

$$\Psi(x, \phi) := (1-\phi)^x \qquad (x \in \mathcal{N}(\Lambda),\; \phi \in [0,1]^\Lambda). \tag{3.11}$$

We need to check that the right-hand side in (3.5) is zero, i.e., that

$$G\Psi(\cdot, \phi)(x) = \mathcal{G}^\dagger \Psi(x, \cdot)(\phi) \qquad (\phi \in [0,1]^\Lambda,\; x \in \mathcal{N}(\Lambda)), \tag{3.12}$$

where G be the generator of the (a,b,c,d)-braco-process, defined in (1.1), and $\mathcal{G}^\dagger$ is the generator of the $(a^\dagger,b,c,d)$-resem-process, defined in (2.37). Note that since x is finite, $\Psi(x,\cdot) \in$

$\mathcal{C}^2_{\rm fin}([0,1]^\Lambda)$. We check that

$$
\begin{aligned}
&G\Psi(\cdot,\phi)(x)\\
&=\sum_{ij} a(i,j)x(i)\{(1-\phi(j))-(1-\phi(i))\}(1-\phi)^{x-\delta_i}+b\sum_i x(i)\{(1-\phi(i))-1\}(1-\phi)^x\\
&\quad+c\sum_i x(i)(x(i)-1)\{1-(1-\phi(i))\}(1-\phi)^{x-\delta_i}+d\sum_i x(i)\{1-(1-\phi(i)\}(1-\phi)^{x-\delta_i}\\
&=-\sum_{ij} a^\dagger(j,i)(\phi(j)-\phi(i))x(i)(1-\phi)^{x-\delta_i}-b\sum_i\phi(i)(1-\phi(i))x(i)(1-\phi)^{x-\delta_i}\\
&\quad+c\sum_i\phi(i)(1-\phi(i))x(i)(x(i)-1)(1-\phi)^{x-2\delta_i}+d\sum_i\phi(i)x(i)(1-\phi)^{x-\delta_i}\\
&=\mathcal{G}^\dagger\Psi(x,\cdot)(\phi)\qquad(\phi\in[0,1]^\Lambda,\ x\in\mathcal{N}(\Lambda)).
\end{aligned}
\tag{3.13}
$$

Set

$$
\Phi(x,\phi):=G\Psi(\cdot,\phi)(x)=\mathcal{G}^\dagger\Psi(x,\cdot)(\phi)\qquad(\phi\in[0,1]^\Lambda,\ x\in\mathcal{N}(\Lambda)). \tag{3.14}
$$

It is not hard to see that there exists a constant K such that

$$
|\Phi(x,\phi)|\le K\Big(1+|x|^2\Big)\qquad(\phi\in[0,1]^\Lambda,\ x\in\mathcal{N}(\Lambda)). \tag{3.15}
$$

Therefore, condition (3.4) is satisfied by (2.1).

To generalize the statement from finite x to general $x\in\mathcal{E}_\gamma(\Lambda)$, we apply Lemma 16. Choose finite $x^{(n)}$ such that $x^{(n)}\uparrow x$ and couple the (a,b,c,d)-braco-processes $X^{(n)},X$ with initial conditions $x^{(n)},x$, respectively, such that $X^{(n)}\uparrow X$. Then, for each $t\ge 0$ and $\phi\in[0,1]^\Lambda$,

$$
E^\phi[(1-\mathcal{X}_t)^{x^{(n)}}]\downarrow E^\phi[(1-\mathcal{X}_t)^x]\quad\text{as } n\uparrow\infty, \tag{3.16}
$$

and

$$
E[(1-\phi)^{X^{(n)}_t}]\downarrow E[(1-\phi)^{X_t}]\quad\text{as } n\uparrow\infty, \tag{3.17}
$$

where we used the continuity of the function $x\mapsto(1-\phi)^x$ with respect to increasing sequences.∎

Proof of Theorem 1 (b) We first prove the statement under the additional assumption that ϕ and ψ are summable. Recall that by Lemma 20, if $\mathcal{X}_0$ is summable then $\mathcal{X}_t$ is summable for all $t\ge 0$ a.s. Let $S:=\{\phi\in[0,1]^\Lambda:|\phi|<\infty\}$ denote the space of summable states. We apply Theorem 23. Our duality function is

$$
\Psi(\phi,\psi):=e^{-\frac{b}{c}\langle\phi,\psi\rangle}\qquad(\phi,\psi\in S). \tag{3.18}
$$

Let $\mathcal{G},\mathcal{G}^\dagger$ denote the generators of the (a,b,c,d)-resem-process and the $(a^\dagger,b,c,d)$-resem-process, as in (2.37), respectively. We need to show that the right-hand side in (3.5) is zero, i.e., that $\mathcal{G}\Psi(\cdot,\psi)(\phi)=\mathcal{G}^\dagger\Psi(\phi,\cdot)(\psi)$. It is not hard to see that $\Psi(\cdot,\psi),\Psi(\phi,\cdot)\in\mathcal{C}_{\rm sum}([0,1]^\Lambda)$ for each

$\psi, \phi \in S$. We calculate

$$
\begin{aligned}
\mathcal{G}\Psi(\cdot,\psi)(\phi) &= \Big\{ \sum_{ij} a(j,i)(\phi(j)-\phi(i))(-\tfrac{b}{c})\psi(i) + b\sum_i \phi(i)(1-\phi(i))(-\tfrac{b}{c})\psi(i) \\
&\qquad + c\sum_i \phi(i)(1-\phi(i))(-\tfrac{b}{c})^2\psi(i)^2 - d\sum_i \phi(i)(-\tfrac{b}{c})\psi(i)\Big\} e^{-\frac{b}{c}\langle\phi,\,\psi\rangle} \\
&= -\tfrac{b}{c}\Big\{ \sum_{ij} a(j,i)\phi(j)\psi(i) - \Big(\sum_j a(j,i)\Big)\sum_i \phi(i)\psi(i) \\
&\qquad + b\sum_i \phi(i)(1-\phi(i))\psi(i)(1-\psi(i)) - d\sum_i \phi(i)\psi(i)\Big\} e^{-\frac{b}{c}\langle\phi,\,\psi\rangle} \\
&= \mathcal{G}^\dagger\Psi(\phi,\cdot)(\psi).
\end{aligned}
\tag{3.19}
$$

It is not hard to see that there exists a constant K such that

$$
|\mathcal{G}\Psi(\cdot,\psi)(\phi)| \leq K|\phi|\,|\psi| \qquad (\phi,\psi\in S). \tag{3.20}
$$

Therefore, condition (3.4) is implied by Lemma 20, and Theorem 23 is applicable. To generalize the result to general $\phi,\psi\in[0,1]^\Lambda$, we apply Lemma 22. ∎

3.3 Subduality

Fix constants $\beta\in\mathbb{R}$, $\gamma\geq 0$. Let $\mathcal{M}(\Lambda) := \{\phi\in[0,\infty)^\Lambda : |\phi|<\infty\}$ be the space of finite measures on Λ, equipped with the topology of weak convergence, and let $\mathcal{Y}$ be the Markov process in $\mathcal{M}(\Lambda)$ given by the unique pathwise solutions to the SDE

$$
\mathrm{d}\mathcal{Y}_t(i) = \sum_j a(j,i)(\mathcal{Y}_t(j)-\mathcal{Y}_t(i))\,\mathrm{d}t + \beta\mathcal{Y}_t(i)\,\mathrm{d}t + \sqrt{2\gamma\mathcal{Y}_t(i)}\,\mathrm{d}B_t(i) \tag{3.21}
$$

($t\geq 0,\ i\in\Lambda$). Then $\mathcal{Y}$ is the well-known super random walk with underlying motion a, growth parameter β and activity γ. One has

$$
E^\phi\big[e^{-\langle\mathcal{Y}_t,\,\psi\rangle}\big] = e^{-\langle\phi,\mathcal{U}_t\psi\rangle} \tag{3.22}
$$

for any $\phi\in\mathcal{M}(\Lambda)$ and bounded nonnegative $\psi:\Lambda\to\mathbb{R}$, where $u_t = \mathcal{U}_t\psi$ solves the differential equation

$$
\tfrac{\partial}{\partial t}u_t(i) = \sum_j a(j,i)(u_t(j)-u_t(i)) + \beta u_t(i) - \gamma u_t(i)^2 \qquad (i\in\Lambda,\ t\geq 0) \tag{3.23}
$$

with initial condition $u_0=\psi$. The semigroup $(\mathcal{U}_t)_{t\geq 0}$ acting on bounded nonnegative functions ψ on Λ is called the log-Laplace semigroup of $\mathcal{Y}$.

We will show that (a,b,c,d)-braco-process and the super random walk with underlying motion $a^\dagger$, growth parameter $b-d+c$ and activity c are related by a duality formula with a nonnegative error term. In analogy with words such as subharmonic and submartingale, we call this a subduality relation.

Proposition 24 (Subduality with a branching process) *Let X be the (a,b,c,d)-braco-process and let $\mathcal{Y}$ be the super random walk with underlying motion $a^\dagger$, growth parameter $b-d+c$ and activity c. Then*

$$E^x\big[e^{-\langle\phi,\,X_t\rangle}\big]\geq E^\phi\big[e^{-\langle\mathcal{Y}_t,\,x\rangle}\big]\qquad(x\in\mathcal{E}_\gamma(\Lambda),\ \phi\in\mathcal{M}(\Lambda)).\tag{3.24}$$

Proof We first prove the statement for finite x. We apply Theorem 23 to X and $\mathcal{Y}$ considered as processes in $\mathcal{N}(\Lambda)$ and $\mathcal{M}(\Lambda)$, respectively. The process $\mathcal{Y}$ solves the martingale problem for the operator

$$\begin{aligned}\mathcal{H}f(\phi):=&\sum_{ij}a^\dagger(j,i)(\phi(j)-\phi(i))\tfrac{\partial}{\partial\phi(i)}f(\phi)+(b-d+c)\sum_i\phi(i)\tfrac{\partial}{\partial\phi(i)}f(\phi)\\&+c\sum_i\phi(i)\tfrac{\partial^2}{\partial\phi(i)^2}f(\phi)\qquad(\phi\in[0,1]^\Lambda),\end{aligned}\tag{3.25}$$

defined for functions ϕ in the space $\mathcal{C}^2_{\rm fin,b}[0,\infty)^\Lambda$ of bounded $\mathcal{C}^2$ functions on $[0,\infty)^\Lambda$ depending on finitely many coordinates. Our duality function is $\Psi(x,\phi):=e^{-\langle\phi,x\rangle}$. We observe that $\Psi(x,\cdot)\in\mathcal{C}^2_{\rm fin,b}[0,\infty)^\Lambda$ for all $x\in\mathcal{N}(\Lambda)$ and calculate

$$\begin{aligned}G\Psi(\cdot,\phi)(x)=\Big\{&\sum_{ij}a(i,j)x(i)\big(e^{\phi(i)-\phi(j)}-1\big)+b\sum_i x(i)\big(e^{-\phi(i)}-1\big)\\&+c\sum_i x(i)(x(i)-1)\big(e^{\phi(i)}-1\big)+d\sum_i x(i)\big(e^{\phi(i)}-1\big)\Big\}e^{-\langle\phi,\,x\rangle},\end{aligned}\tag{3.26}$$

and

$$\begin{aligned}\mathcal{H}\Psi(x,\cdot)(\phi)=\Big\{&\sum_{ij}a^\dagger(j,i)x(i)(\phi(i)-\phi(j))-(b-d+c)x(i)\phi(i)\\&+c\sum_i x(i)^2\phi(i)\Big\}e^{-\langle\phi,\,x\rangle}\end{aligned}\tag{3.27}$$

$(x\in\mathcal{N}(\Lambda),\ \phi\in\mathcal{M}(\Lambda))$. It is not hard to see that there exists a constant K such that

$$|G\Psi(\cdot,\phi)(x)|\leq K|x|^2\quad\text{and}\quad|\mathcal{H}\Psi(x,\cdot)(\phi)|\leq K|x|^2\,|\phi|\qquad(x\in\mathcal{N}(\Lambda),\ \phi\in\mathcal{M}(\Lambda)).\tag{3.28}$$

and therefore condition (3.4) is implied by (2.1) and the elementary estimate $E[|\mathcal{Y}_t|]\leq e^{(b-d+c)t}|\phi|$. One has

$$\begin{aligned}G\Psi(\cdot,\phi)(x)-\mathcal{H}\Psi(x,\cdot)(\phi)=&\Big\{\sum_{ij}a(i,j)x(i)\big(e^{\phi(i)-\phi(j)}-1-(\phi(i)-\phi(j))\big)\\&+b\sum_i x(i)\big(e^{-\phi(i)}-1+\phi(i)\big)+c\sum_i x(i)(x(i)-1)\big(e^{\phi(i)}-1-\phi(i)\big)\\&+d\sum_i x(i)\big(e^{\phi(i)}-1-\phi(i)\big)\Big\}e^{-\langle\phi,\,x\rangle}\geq 0,\end{aligned}\tag{3.29}$$

and therefore, for finite x, (3.24) is implied by Theorem 23. The general case follows by approximation, using Lemma 16. ∎

4 The Maximal Processes

4.1 The maximal branching-coalescing process

Using Proposition 24 we can now prove Theorem 2.

Proof of Theorem 2 Choose $x^{(n)} \in \mathcal{E}_\gamma(\Lambda)$ such that $x^{(n)}(i) \uparrow \infty$ for all $i \in \Lambda$. By Lemma 16, the (a,b,c,d)-braco processes $X^{(n)}$ started in $x^{(n)}$, respectively, can be coupled such that $X^{(n)}_t \leq X^{(n+1)}_t$ for each $t \geq 0$. Define $X^{(\infty)} = (X^{(\infty)}_t)_{t\geq 0}$ as the $\overline{\mathbb{N}}^\Lambda$-valued process that is the pointwise increasing limit of the $X^{(n)}$. By Proposition 24 and (3.22),

$$E\big[1 - e^{-\langle \varepsilon\delta_i,\, X^{(n)}_t\rangle}\big] \leq 1 - e^{-\langle \varepsilon\delta_i, \mathcal{U}_t x^{(n)}\rangle} \qquad (t, \varepsilon \geq 0,\ i \in \Lambda). \tag{4.1}$$

where $(\mathcal{U}_t)_{t\geq 0}$ is the log-Laplace semigroup of the super random walk with underlying motion $a^\dagger$, growth parameter $r := b - d + c$ and activity c. It follows that

$$E[X^{(n)}_t(i)] = \lim_{\varepsilon\downarrow 0} \varepsilon^{-1} E\big[1 - e^{-\langle \varepsilon\delta_i, X^{(n)}_t\rangle}\big] \leq \lim_{\varepsilon\downarrow 0} \varepsilon^{-1}\big(1 - e^{-\langle \varepsilon\delta_i, \mathcal{U}_t x^{(n)}\rangle}\big) = \mathcal{U}_t x^{(n)}(i) \tag{4.2}$$

$(t \geq 0,\ i \in \Lambda)$. Using the explicit solution of (3.23) for constant initial conditions, it is easy to see that $\mathcal{U}_t x^{(n)} \uparrow \mathcal{U}_t\infty$, where

$$\mathcal{U}_t\infty := \begin{cases} \frac{r}{c(1-e^{-rt})} & \text{if } r \neq 0, \\ \frac{1}{ct} & \text{if } r = 0. \end{cases} \tag{4.3}$$

(See, for example, [FS03a, formula (32)].) Letting $n \uparrow \infty$ in (4.2) we arrive at Theorem 2 (b). Moreover, we see that

$$E\big[\|X^{(\infty)}_t(i)\|_\gamma\big] \leq \mathcal{U}_t\infty \sum_i \gamma_i < \infty \qquad (t > 0), \tag{4.4}$$

and therefore $X^{(\infty)}_t \in \mathcal{E}_\gamma(\Lambda)$ a.s. for each $t > 0$. Part (a) of the theorem now follows from Lemma 16. Using Theorem 1 (a) and the continuity of the function $x \mapsto (1-\phi)^x$ with respect to increasing sequences, reasoning as in (1.28), we see that

$$P[\mathrm{Thin}_\phi(X^{(\infty)}_t) = 0] = P^\phi[\mathcal{X}^\dagger_t = 0] \qquad (\phi \in [0,1]^\Lambda,\ t \geq 0), \tag{4.5}$$

where $\mathcal{X}^\dagger$ denotes the $(a^\dagger, b, c, d)$-resem-process. Since formula (4.5) determines the distribution of $X^{(\infty)}_t$ uniquely, the law of $X^{(\infty)}_t$ does not depend on the choice of the $x^{(n)} \uparrow \infty$ $(t \geq 0)$. This completes the proof of part (c) of the theorem.

To prove part (d), fix $0 \leq s \leq t$. Choose $y_n \in \mathcal{E}_\gamma(\Lambda)$, $y_n(i) \uparrow \infty$ $\forall i \in \Lambda$ and let $\tilde{X}^{(n)}$ be the (a,b,c,d)-braco-process started in $\tilde{X}^{(n)}_0 := X^{(\infty)}_{t-s} \vee y_n$. Then $\tilde{X}^{(n)}_0 \geq X^{(\infty)}_{t-s}$ and therefore, by Lemma 9, $\tilde{X}^{(n)}_s$ and $X^{(\infty)}_t$ may be coupled such that $\tilde{X}^{(n)}_s \geq X^{(\infty)}_t$. By part (c) of the theorem, $\tilde{X}^{(n)}_s$ and $X^{(\infty)}_s$ may be coupled such that $\tilde{X}^{(n)}_s \uparrow X^{(\infty)}_s$ and therefore $X^{(\infty)}_s$ and $X^{(\infty)}_t$ may be coupled such that $X^{(\infty)}_s \geq X^{(\infty)}_t$.

It follows that $\mathcal{L}(X_t^{(\infty)}) \downarrow \overline{\nu}$ for some probability measure $\overline{\nu}$ on $\mathcal{E}_\gamma(\Lambda)$. Set $\rho := \mathcal{L}(X_1^{(\infty)})$ and let $(S_t)_{t\geq 0}$ denote the semigroup of the (a,b,c,d)-braco-process. Recall the definition of $\mathcal{C}_{\rm Lip,b}(\mathcal{E}_\gamma(\Lambda))$ above (2.21). One has

$$\int \overline{\nu}(\mathrm{d}x) f(x) = \lim_{t\to\infty} \int \rho(\mathrm{d}x) S_t f(x) \tag{4.6}$$

for every $f \in \mathcal{C}_{\rm Lip,b}(\mathcal{E}_\gamma(\Lambda))$. Therefore, since S_t maps $\mathcal{C}_{\rm Lip,\,b}(\mathcal{E}_\gamma(\Lambda))$ into itself,

$$\int \overline{\nu}(\mathrm{d}x) S_s f(x) = \lim_{t\to\infty} \int \rho(\mathrm{d}x) S_t S_s f(x) = \int \overline{\nu}(\mathrm{d}x) f(x) \qquad (s \geq 0), \tag{4.7}$$

for every $f \in \mathcal{C}_{\rm Lip,\,b}(\mathcal{E}_\gamma(\Lambda))$, which shows that $\overline{\nu}$ is an invariant measure. If ν is another invariant measure, then $\mathcal{L}(X_t^{(\infty)}) \geq \nu$ for all $t \geq 0$. Letting $t \to \infty$, we see that $\overline{\nu} \geq \nu$, proving part (e) of the theorem. Part (f) has already been proved in the introduction. ∎

4.2 The maximal resampling-selection process

The proof of Theorem 3 (a)–(c) is similar to the proof of Theorem 2, but easier. Recall that Theorem 3 (d) is proved in Section 1.5.

Proof of Theorem 3 (a)–(c) Part (a) can be proved in the same way as Theorem 2 (d), using Lemma 22. The proof of part (b) goes analogue to the proof of Theorem 2 (e). To see why (1.30) holds, note that for any $\phi \in [0,1]^\Lambda$, by Theorem 1 (a),

$$\int \overline{\mu}(\mathrm{d}\phi)(1-\phi)^x = \lim_{t\to\infty} P^1[\mathrm{Thin}_{\mathcal{X}_t}(x) = 0] = \lim_{t\to\infty} P^x[\mathrm{Thin}_1(X_t^\dagger) = 0]. \tag{4.8}$$

To complete the proof of part (c) we must show that $\overline{\mu}$ is nontrivial if and only if the $(a^\dagger,b,c,d)$-process survives. Using subadditivity (Lemma 10) it is easy to see that the $(a^\dagger,b,c,d)$-process survives if and only if $P^{\delta_i}[X_t^\dagger \neq 0\ \forall t \geq 0] > 0$ for some $i \in \Lambda$. Formula (1.30) implies that $\int \overline{\mu}(\mathrm{d}\phi)\phi(i) = P^{\delta_i}[X_t^\dagger \neq 0\ \forall t \geq 0]$, which shows that $\overline{\mu} = \delta_0$ if and only if the $(a^\dagger,b,c,d)$-process survives. If $\overline{\mu} \neq \delta_0$ then the measure $\overline{\mu}$ conditioned on $\{\phi : \phi \neq 0\}$ is an invariant measure of the (a,b,c,d)-resem-process that is stochastically larger than $\overline{\mu}$. By part (b), this conditioned measure is $\overline{\mu}$ itself, thus $\overline{\mu}(\{0\}) = 0$, i.e., $\overline{\mu}$ is nontrivial. ∎

5 Convergence to the Upper Invariant Measure

5.1 Extinction versus unbounded growth

In this section we prove Lemma 5. It has already been proved in Section 1.5 that $e^{-\frac{b}{c}|\mathcal{X}_t|}$ is a submartingale. Therefore, if $b > 0$, then $|\mathcal{X}_t|$ converges a.s. to a limit in $[0,\infty]$. If $b = 0$ then it is easy to see that $|\mathcal{X}_t|$ is a nonnegative supermartingale and therefore also in this case $|\mathcal{X}_t|$ converges a.s. Thus, all we have to do is to show that $\lim_{t\to\infty} |\mathcal{X}_t|$ takes values in $\{0,\infty\}$ a.s. (Proposition 26 below), and that $\mathcal{X}$ gets extinct in finite time if the limit is zero (Lemma 25). Throughout this section, $c > 0$ and $\mathcal{X}$ is the (a,b,c,d)-resem-process starting in an initial state $\phi \in [0,1]^\Lambda$ with $|\phi| < \infty$.

Lemma 25 (Finite time extinction) *One has* $\mathcal{X}_t = 0$ *for some* $t \geq 0$ *a.s. on the event* $\lim_{t\to\infty} |\mathcal{X}_t| = 0$.

Proof Choose $x^{(n)} \in \mathcal{E}_\gamma(\Lambda)$ such that $x^{(n)}(i) \uparrow \infty$ for all $i \in \Lambda$. Let $X^{(n)\dagger}$ denote the $(a^\dagger, b, c, d)$-braco-process started in $x^{(n)}$ and let $X^{(\infty)\dagger}$ denote the maximal $(a^\dagger, b, c, d)$-braco-process. By Theorem 1 (a) and Theorem 2 (b),

$$\begin{aligned} P^\phi[\mathcal{X}_t \neq 0] &= \lim_{n\uparrow\infty} P^\phi[\mathrm{Thin}_{\mathcal{X}_t}(x^{(n)}) \neq 0] = \lim_{n\uparrow\infty} P[\mathrm{Thin}_\phi(X^{(n)\dagger}_t) \neq 0] \\ &= P[\mathrm{Thin}_\phi(X^{(\infty)\dagger}_t) \neq 0] \leq E\big[|\mathrm{Thin}_\phi(X^{(\infty)\dagger}_t)|\big] = \langle \phi, E[X^{(\infty)\dagger}_t]\rangle \leq |\phi| \mathcal{U}_t\infty, \end{aligned} \tag{5.1}$$

where $\mathcal{U}_t\infty$ is the function on the right-hand side in (1.23). Choose $\varepsilon > 0$ and $t_0 > 0$ such that $\varepsilon\mathcal{U}_{t_0}\infty \leq \frac{1}{2}$. Let $(\mathcal{F}_t)_{t\geq 0}$ denote the filtration generated by $\mathcal{X}$. By (5.1),

$$\tfrac{1}{2} 1_{\{|\mathcal{X}_t| \leq \varepsilon\}} \leq P[\mathcal{X}_{t+t_0} = 0|\mathcal{F}_t] \leq P[\exists s \geq 0 \text{ s.t. } \mathcal{X}_s = 0|\mathcal{F}_t]. \tag{5.2}$$

Now

$$1_{\{\lim_{s\to\infty} \mathcal{X}_s = 0\}} \leq \liminf_{t\to\infty} 1_{\{|\mathcal{X}_t| \leq \varepsilon\}}, \tag{5.3}$$

while

$$P[\exists s \geq 0 \text{ s.t. } \mathcal{X}_s = 0|\mathcal{F}_t] \to 1_{\{\exists s \geq 0 \text{ s.t. } \mathcal{X}_s = 0\}} \quad \text{as } t \to \infty \quad \text{a.s.}, \tag{5.4}$$

by a general property of strong Markov processes (see, for example, [FS03a, Lemma A.1]). Letting $t \to \infty$ in (5.2), using (5.3) and (5.4), we find that $\frac{1}{2} 1_{\{\lim_{s\to\infty} \mathcal{X}_s = 0\}} \leq 1_{\{\exists s \geq 0 \text{ s.t. } \mathcal{X}_s = 0\}}$ a.s. ∎

To finish this section, we need to prove:

Proposition 26 (Convergence to zero or infinity) *Assume that* Λ *is infinite. Then* $\lim_{t\to\infty} |\mathcal{X}_t| \in \{0, \infty\}$ *a.s.*

Since the proof of Proposition 26 is rather long we will skip some of the boring details. Our first step is:

Lemma 27 (Integrable fluctuations) *One has*

$$\int_0^\infty \sum_i \mathcal{X}_t(i)(1 - \mathcal{X}_t(i))\,\mathrm{d}t < \infty \tag{5.5}$$

a.s. on the event $\lim_{t\to\infty} |\mathcal{X}_t| \in [0, \infty)$.

Proof For any $\psi \in [0, \infty)^\Lambda$ with $|\psi| < \infty$ one has $e^{-\langle \cdot, \psi\rangle} \in \mathcal{C}^2_{\mathrm{sum}}([0, 1]^\Lambda)$ and (compare (3.19))

$$\begin{aligned} \mathcal{G} e^{-\langle \cdot,\, \psi\rangle}(\phi) = \Big\{ &- \sum_i \phi(i) \sum_j a^\dagger(j, i)(\psi(j) - \psi(i)) \\ &+ \sum_i \phi(i)(1 - \phi(i))\big(c\psi(i)^2 - b\psi(i)\big) + d \sum_i \phi(i)\psi(i) \Big\} e^{-\langle \phi,\, \psi\rangle}. \end{aligned} \tag{5.6}$$

Since $\mathcal{X}$ solves the martingale problem for $\mathcal{G}$,

$$E\Big[\int_0^t \mathcal{G} e^{-\langle \cdot,\, \psi\rangle}(\mathcal{X}_s)\mathrm{d}s\Big] = E\big[e^{-\langle \mathcal{X}_t,\, \psi\rangle}\big] - e^{-\langle \phi,\, \psi\rangle} \qquad (t \geq 0). \tag{5.7}$$

Choose $\lambda > 0$ such that $c\lambda^2 - b\lambda =: \mu > 0$ and $\psi_n \in [0,\infty)^\Lambda$ with $|\psi_n| < \infty$ such that $\psi_n \uparrow \lambda$. Then the bounded pointwise limit of the function $i \mapsto \sum_j a^\dagger(j,i)(\psi_n(j) - \psi_n(i))$ is zero and therefore, taking the limit in (5.7), using Lemma 20, we find that

$$E\Big[\int_0^t \sum_i \Big\{\mu\mathcal{X}_s(i)(1-\mathcal{X}_s(i)) + \lambda d\mathcal{X}_s(i)\Big\}e^{-\lambda|\mathcal{X}_s|}\mathrm{d}s\Big] = E\big[e^{-\lambda|\mathcal{X}_t|}\big] - e^{-\langle\phi,\psi\rangle}. \tag{5.8}$$

Letting $t \uparrow \infty$, using the fact that the right-hand side of (5.8) is bounded by one, we see that

$$\int_0^\infty \sum_i \Big\{\mu\mathcal{X}_t(i)(1-\mathcal{X}_t(i)) + \lambda d\mathcal{X}_t(i)\Big\}e^{-\lambda|\mathcal{X}_t|}\mathrm{d}t < \infty \quad \text{a.s.}, \tag{5.9}$$

which implies (5.5). ∎

Lemma 28 (Process not started with only zeros and ones) *For every $0 < \varepsilon < \frac{1}{4}$ there exists a $\delta, r > 0$ such that*

$$P^\phi\big[\mathcal{X}_t(i) \in (\varepsilon, 1-\varepsilon)\ \forall t \in [0,r]\big] \geq \delta \qquad (i \in \Lambda,\ \phi \in [0,1]^\Lambda,\ \phi(i) \in (2\varepsilon, 1-2\varepsilon)). \tag{5.10}$$

Proof Since $\sup_i \sum_j a(i,j) < \infty$ and all the components of the (a,b,c,d)-resem-process take values in $[0,1]$, the maximal drift that the i-th component $\mathcal{X}_t(i)$ can experience (both in the positive and negative direction) can be uniformly bounded. Now the proof of (5.10) is just a boring calculation, which we skip. ∎

Lemma 29 (Convergence to zero or one) *Almost surely on the event $\lim_{t\to\infty}|\mathcal{X}_t| \in [0,\infty)$, there exists a set $\Delta \subset \Lambda$ such that $\lim_{t\to\infty}\mathcal{X}_t(i) = 1$ for all $i \in \Delta$ and $\lim_{t\to\infty}\mathcal{X}_t(i) = 0$ for all $i \in \Lambda\backslash\Delta$.*

Proof Imagine that the statement does not hold. Then, by the continuity of sample paths, with positive probability $\lim_{t\to\infty}|\mathcal{X}_t| \in [0,\infty)$ while there exists $0 < \varepsilon < \frac{1}{4}$ such that for every $T > 0$ there exists $t \geq T$ and $i \in \Lambda$ with $\mathcal{X}_t(i) \in (2\varepsilon, 1-2\varepsilon)$. Using Lemma 28 and the strong Markov property, it is then not hard to check that with positive probability $\lim_{t\to\infty}|\mathcal{X}_t| \in [0,\infty)$ while there exist infinitely many disjoint time intervals $[t_k, t_k + r]$ and points $i_k \in \Lambda$ such that $\mathcal{X}_t(i_k) \in (\varepsilon, 1-\varepsilon)$ for all $t \in [t_k, t_k + r]$. This contradicts Lemma 27. ∎

Lemma 30 (Convergence to one on a finite nonempty set) *Almost surely on the event $\lim_{t\to\infty}|\mathcal{X}_t| \in (0,\infty)$, the set Δ from Lemma 29 is finite and nonempty.*

Proof It is clear that Δ is finite a.s. on the event $\lim_{t\to\infty}|\mathcal{X}_t| < \infty$. Now imagine that Δ is empty. Then, a.s. on the event $\lim_{t\to\infty}|\mathcal{X}_t| > 0$, there exists a random time T such that $\mathcal{X}_t(i) \leq \frac{1}{2}$ for all $t \geq T$ and $i \in \Lambda$. Since $z(1-z) \geq \frac{1}{2}z$ on $[0,\frac{1}{2}]$, it follows that a.s. on the event $\lim_{t\to\infty}|\mathcal{X}_t| > 0$,

$$\int_T^\infty \sum_i \mathcal{X}_t(i)(1-\mathcal{X}_t(i))\mathrm{d}t \geq \frac{1}{2}\int_T^\infty |\mathcal{X}_t| = \infty. \tag{5.11}$$

We arrive at a contradiction with Lemma 27. ∎

Proof of Proposition 26 Let Δ be the random set from Lemma 29. We will show that $\Delta = \Lambda$ a.s. on the event $\lim_{t\to\infty} |\mathcal{X}_t| \in (0,\infty)$. In particular, by Lemma 30, if Λ is infinite this implies that the event $\lim_{t\to\infty} |\mathcal{X}_t| \in (0,\infty)$ has zero probability. Assume that with positive probability $\lim_{t\to\infty} |\mathcal{X}_t| \in (0,\infty)$ and $\Delta \neq \Lambda$. By Lemma 30, Δ is nonempty, and therefore by irreducibility there exist $i \in \Lambda\backslash\Delta$ and $j \in \Delta$ such that $a(i,j) > 0$ or $a(j,i) > 0$. If $a(i,j) > 0$ then by the fact that the counting measure is an invariant measure for the Markov process with jump rates a and by the finiteness of Δ, there must also be an $i' \in \Lambda\backslash\Delta$ and $j' \in \Delta$ such that $a(j',i') > 0$. Thus, there exist $i,j \in \Lambda$ such that $a(j,i) > 0$ and with positive probability $\lim_{t\to\infty} \mathcal{X}_t(i) = 0$, and $\lim_{t\to\infty} \mathcal{X}_t(j) = 1$. It is not hard to see that this violates the evolution in (1.3). (We skip the details.) ∎

5.2 The spatial ergodic theorem

In this section we prove the spatial ergodic theorem Theorem 7 and a simple consequence of it. Our proof of Theorem 7 follows the usual pattern for an L^2-ergodic theorem, but with one subtle point. If one proceeds too naively, then it seems that instead of (1.36) one needs the stronger condition

$$\lim_{n\to\infty} |\pi_n - \pi_n \circ T_i| = 0 \quad \forall i \in \Lambda, \tag{5.12}$$

which can probably not be satisfied on general groups. To see that in fact (1.36) is all one needs, an extra argument is needed to show that

$$\sum_i P(i,0)\, f \circ T_i = f \quad \text{implies} \quad f = f \circ T_i \quad \forall i \in \Lambda \qquad \text{a.s.} \tag{5.13}$$

Note that functions P satisfying the assumptions of Theorem 7 always exist: if $\Delta \subset \Lambda$ is a generating set and Q is a probability distribution with support Δ, then one can take $P(i,j) := Q(i^{-1}j)$.

Proof of Theorem 7 Note that $\sum_i P^\dagger(0,i) = 1$ and that $\{i : P^\dagger(0,i) > 0\}$ generates Λ. Set $P^\dagger(i,j) := P(j,i)$ and define $S : L^2(\mu) \to L^2(\mu)$ by

$$Sf := \sum_i P^\dagger(0,i) f \circ T_i. \tag{5.14}$$

Note that, since μ is homogeneous,

$$\|Sf\|_{\mu,2} \leq \sum_i P^\dagger(0,i) \|f \circ T_i\|_{\mu,2} = \|f\|_{\mu,2}, \tag{5.15}$$

i.e., S is a contraction. Set

$$\begin{aligned} H &:= \{f \in L^2(\mu) : f \circ T_i = f\ \forall i \in \Lambda\}, \\ H' &:= \{f \in L^2(\mu) : Sf = f\}. \end{aligned} \tag{5.16}$$

As announced in (5.13), we will prove that $H = H'$. The inclusion $H \subset H'$ is obvious, so assume that $f \in H'$. Let $Y = (Y(i))_{i\in\Lambda}$ be an E^Λ-valued random variable with law $\mathcal{L}(Y) = \mu$. Fix a point $k \in \Lambda$ and let $\xi = (\xi_n)_{n\geq 0}$ be a random walk on Λ, independent of Y, with transition probabilities

$$P[\xi_{n+1} = j|\xi_n = i] = P^\dagger(i,j) \qquad (n \geq 0,\ i,j \in \Lambda), \tag{5.17}$$

and initial state $\xi_0 = k$. Define an E-valued process $Z = (Z_n)_{n\geq 0}$ by

$$Z_n := f \circ T_{\xi_n}(Y) \qquad (n \geq 0). \tag{5.18}$$

Note that, since μ is homogeneous,

$$\begin{aligned} E\big[|Z_n|^2\big] &= \sum_i (P^\dagger)^n(k,i) \int \mu(\mathrm{d}y)\, |f \circ T_i(y)|^2 \qquad (n \geq 0) \\ &= \sum_i (P^\dagger)^n(k,i) \|f\|_{\mu,2} = \|f\|_{\mu,2}. \end{aligned} \tag{5.19}$$

Let $\mathcal{F}_n := \sigma(Y, \xi_m : m \leq n)$ $(n \geq 0)$ be the σ-field generated by Y and the process ξ up to time $n \geq 0$. Then, since $f \in H'$,

$$\begin{aligned} E[Z_{n+1}|\mathcal{F}_n] &= \sum_j P^\dagger(\xi_n, j) f \circ T_j(Y) = \sum_i P^\dagger(\xi_n, \xi_n i) f \circ T_{\xi_n i} \\ &= \sum_i P^\dagger(0,i) f \circ T_i \circ T_{\xi_n} = f \circ T_{\xi_n} = Z_n, \end{aligned} \tag{5.20}$$

where we have used that

$$T_i T_j x(k) = T_j x(ik) = x(jik) = T_{ji} x(k). \tag{5.21}$$

Formulas (5.19) and (5.20) show that Z is a square integrable martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$. It follows that

$$E[|Z_{n+1} - Z_n|^2] = E[|Z_{n+1}|^2] - E[|Z_n|^2] \qquad (n \geq 0). \tag{5.22}$$

By (5.19), $E[|Z_n|^2]$ does not depend on n (in fact, Z is stationary) and therefore $Z_n = Z_0$ a.s. for all $n \geq 0$. Since $\{i \in \Lambda : P^\dagger(0,i) > 0\}$ generates Λ and the starting point k is arbitrary, this is possible only if $f = f \circ T_i$ a.s. for all $i \in \Lambda$. This proves that $H = H'$.

The rest of the proof is standard (compare the proof of Von Neumann's mean ergodic theorem in [Kre85, Theorem 1.1.4]). Note that

$$T_i T_j x(k) = T_j x(ik) = x(jik) = T_{ji} x(k). \tag{5.23}$$

Since $H = H'$ and by [Kre85, Lemma 1.1.3], the orthogonal complement $H^\perp$ of H is the closure of the space $\mathrm{span}\{h - Sh : h \in L^2(\mu)\}$. If π_n are probability distributions on Λ satisfying (1.36) and $f = h - Sh$, then

$$\begin{aligned}\|\sum_i \pi_n(i) f\circ T_i\|_{\mu,2} &= \|\sum_i \pi_n(i) h\circ T_i - \sum_i \pi_n(i)\big(\sum_j P^\dagger(0,j) h\circ T_j\big)\circ T_i\|_{\mu,2}\\ &= \|\sum_i \pi_n(i) h\circ T_i - \sum_{i,j} P(ij,i)\pi_n(i) h\circ T_{ij}\|_{\mu,2}\\ &= \|\sum_i \pi_n(i) h\circ T_i - \sum_k \big(\sum_i P(k,i)\pi_n(i)\big) h\circ T_k\|_{\mu,2}\\ &= \|\sum_i \big(\pi_n(i) - P\pi_n(i)\big) h\circ T_i\|_{\mu,2} \le \sum_i |\pi_n(i) - P\pi_n(i)|\,\|f\circ T_i\|_{\mu,2}\\ &= |\pi_n - P\pi_n|\,\|h\|_{\mu,2} \to 0.\end{aligned} \tag{5.24}$$

By approximation, $\|\sum_i \pi_n(i) f\circ T_i\|_{\mu,2} \to 0$ for all $f \in H^\perp$. On the other hand, $\sum_i \pi_n(i) f\circ T_i = f$ for all $f \in H$. It follows that $\sum_i \pi_n(i) f\circ T_i$ converges in $L^2(\mu)$ to the orthogonal projection of f on H, which equals the conditional expectation of f given $\mathcal{T}$. ∎

The following simple consequence of Theorem 7 is actually all that we need it for. Here $\stackrel{\mathrm{P}}{\to}$ denotes convergence in probability.

Lemma 31 (Positive density of occupied sites) *Assume that Y is an $\mathbb{N}^\Lambda$-valued random variable with a homogeneous nontrivial law, and assume that $\phi_n \in [0,1]^\Lambda$ satisfy $|\phi_n| \to \infty$ and $|\phi_n - P\phi_n|/|\phi_n| \to 0$. Then $\langle Y \wedge 1, \phi_n\rangle \stackrel{\mathrm{P}}{\to} \infty$.*

Proof Set $\mu := \mathcal{L}(Y)$ and define $f \in L^2(\mu)$ by $f(x) := x(0) \wedge 1$ ($x \in \mathbb{N}^\Lambda$). Let $F := E[f|\mathcal{T}]$ be the conditional expectation (with respect to μ) of f given $\mathcal{T}$. Note that if $\mu(\{F \in D\}) > 0$ for some measurable $D \subset [0,\infty)$, then the conditioned measure $\mu(A|F \in D) := \mu(A \cap \{F \in D\})/\mu(\{F \in D\})$ is homogeneous and F is also (a version of) the conditional expectation of f given $\mathcal{T}$ with respect to the conditioned measure.

Since μ is nontrivial, $F > 0$ a.s. $[\mu]$. For assume that $\mu(\{F \le 0\}) > 0$. Then the conditioned measure $\tilde\mu := \mu(\,\cdot\,|F \le 0)$ is homogeneous and satisfies $\int (x(0) \wedge 1)\tilde\mu(\mathrm{d}x) = \int F(x)\tilde\mu(\mathrm{d}x) \le 0$, and therefore $\tilde\mu = \delta_0$, which contradicts the assumption that μ is nontrivial.

Define probability distributions on Λ by $\pi_n := \phi_n/|\phi_n|$. Theorem 7 shows that

$$\mu\big(\sum_i \pi_n(i) f\circ T_i \ge \varepsilon/2 \big| F \ge \varepsilon\big) \to 1 \quad \text{as } n \to \infty \quad \forall \varepsilon > 0. \tag{5.25}$$

Since $|\phi_n| \to \infty$ it follows that

$$\mu\big(\sum_i \phi_n(i) f\circ T_i \ge M \big| F \ge \varepsilon\big) \to 1 \quad \text{as } n \to \infty \quad \forall \varepsilon > 0,\ M < \infty. \tag{5.26}$$

Therefore

$$\liminf_{n\to\infty} \mu\big(\sum_i \phi_n(i) f\circ T_i \ge M\big) \ge \mu(F \ge \varepsilon) \quad \forall \varepsilon > 0,\ M < \infty. \tag{5.27}$$

Since $F > 0$ a.s. $[\mu]$ and $\langle Y \wedge 1, \phi_n\rangle = \sum_i \phi_n(i) f\circ T_i(Y)$, we arrive at the claim in Lemma 31. ∎

5.3 Convergence to the upper invariant measure

In this section we complete the proof of Theorem 4, started in Section 1.5, by proving Lemma 6. We start with some preparatory lemmas. Throughout this section, $P(i,j)$ is the transition kernel defined by

$$P(i,j) = \tfrac{1}{|a|}a(i,j), \quad \text{where } |a| := \sum_j a(i,j) \quad \forall i \in \Lambda. \tag{5.28}$$

Our first lemma is just a simple observation.

Lemma 32 (Zero thinning) *Let Y_n be $\mathbb{N}^\Lambda$-valued random variables and $\phi_n \in [0,1]^\Lambda$. If $\langle Y_n, \phi_n\rangle \stackrel{\mathrm{P}}{\to} \infty$, then $P[\mathrm{Thin}_{\phi_n}(Y_n) = 0] \to 0$.*

Proof Just note that $(1-\phi_n)^{Y_n} = e^{\langle Y_n, \log(1-\phi_n)\rangle} \leq e^{-\langle Y_n, \phi_n\rangle}$ since $\log(z) \leq z - 1$ $(z \geq 0)$. Thus, $\langle Y_n, \phi_n\rangle \stackrel{\mathrm{P}}{\to} \infty$ implies $P[\mathrm{Thin}_{\phi_n}(Y_n) = 0] = E[(1-\phi_n)^{Y_n}] \to 0$. ∎

Combining Lemmas 31 and 32 we are almost at the statement in Lemma 6. We must only show that by letting the process X run during a time interval of length 1, we can get rid of the assumption from Lemma 31 that $|\phi_n - P\phi_n|/|\phi_n| \to 0$.

For any finite $\Delta \subset \Lambda$, let $\xi^\Delta = (\xi^\Delta_t)_{t\geq 0}$ be the continuous time Markov process in $\Delta \cup \{\dagger\}$ with the following description: For $i, j \in \Delta$, ξ^Δ jumps from i to j with rate $a(i,j)$. Moreover, ξ^Δ jumps from any site i to the cemetary point $\dagger$ with rate

$$d + \sum_{j \in \Lambda\setminus\Delta} a(i,j). \tag{5.29}$$

In other words, ξ^Δ is the underlying motion of the (a,b,c,d)-braco-process, which is additionally killed with rate d in any point $i \in \Delta$ and killed immediately outside Δ.

Lemma 33 (Lower estimate with killed random walks) *Assume that $\Delta \subset \Lambda$ is finite, $0 \in \Delta$, and that*

$$\varepsilon_\Delta := \min_{i\in\Delta} P^i[\xi^\Delta_1 = 0] > 0. \tag{5.30}$$

Let $A \subset \Lambda$ be a set such that the sets $a\Delta := \{ai : i \in \Delta\}$ with $a \in A$ are disjoint. Let X be the (a,b,c,d)-braco-process and define

$$Y(a) := \begin{cases} 1_{\{\langle X_0, 1_{a\Delta}\rangle \neq 0\}} & \text{for } a \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{5.31}$$

Then $\mathcal{L}(X_1) \geq \mathcal{L}(\mathrm{Thin}_{\varepsilon_\Delta}(Y))$.

Proof We can estimate $(X_t)_{t\geq 0}$ from below by a process of independent particles, one in each set $a\Delta$ with $a \in A$ that contains at least one particle of X_0, where the particle in $a\Delta$ jumps from $i \in a\Delta$ to $j \in a\Delta$ with rate $a(i,j)$, is killed with rate d in every site, and is killed immediately outside $a\Delta$. The particle in $a\Delta$ has a probability of at least ε_Δ to be in a at time 1, and therefore X_1 can be estimated from below by a random variable Z of the form

$$Z(a) := \begin{cases} \chi_a 1_{\{\langle X_0, 1_{a\Delta}\rangle \neq 0\}} & \text{for } a \in A, \\ 0 & \text{otherwise,} \end{cases} \tag{5.32}$$

where the $\{\chi_a\}_{a\in A}$ are independent $\{0,1\}$-valued random variables, independent of X_0, with $P[\chi_a = 1] = \varepsilon_\Delta$. This says that Z is an ε_Δ-thinning of Y. ∎

Lemma 34 (Smooth probability distributions) *There exist probability distributions π_n on Λ such that $\lim_{n\to\infty} |\pi_n - P\pi_n| = 0$, such that $\Delta_n := \{i \in \Lambda : \pi_n(i) > 0\}$ is finite, $0 \in \Delta_n$ and*

$$\varepsilon_n := \min_{i\in\Delta_n} P^i[\xi_1^{\Delta_n} = 0] > 0 \tag{5.33}$$

for all n.

Proof Set $\tilde\pi_n := \frac{1}{n}\sum_{k=0}^{n-1} P^k\delta_0$. It is easy to see that $|\tilde\pi_n - P\tilde\pi_n| \leq 2/n$. Set $\Lambda' := \{i : \exists k \geq 0 \text{ s.t. } P^k(i,0) > 0\}$. Note that the $\tilde\pi_n$ are concentrated on Λ'. We will show that it is possible to choose finite $\Delta_k \uparrow \Lambda'$ such that

$$\min_{i\in\Delta_k} P^i[\xi_1^{\Delta_k} = 0] > 0 \qquad \forall k \geq 0. \tag{5.34}$$

Then, for fixed n, $1_{\Lambda\setminus\Delta_k}\tilde\pi_n \downarrow 0$ and therefore

$$|(1-P)(1_{\Lambda\setminus\Delta_k}\tilde\pi_n)| \downarrow 0 \quad \text{and} \quad |1_{\Lambda\setminus\Delta_k}\tilde\pi_n| \downarrow 0 \quad \text{as } k\uparrow\infty. \tag{5.35}$$

We can choose $k(n)$ such that

$$|(1-P)(1_{\Lambda\setminus\Delta_{k(n)}}\tilde\pi_n)| \leq \tfrac{1}{n} \quad \text{and} \quad |1_{\Lambda\setminus\Delta_{k(n)}}\tilde\pi_n| \leq \tfrac{1}{n}. \tag{5.36}$$

Put $\pi_n := 1_{\Delta_{k(n)}}\tilde\pi_n / |1_{\Delta_{k(n)}}\tilde\pi_n|$. Then the π_n are finitely supported, the ε_n in (5.33) are positive, and

$$|\pi_n - P\pi_n| = \frac{|(1-P)1_{\Delta_{k(n)}}\tilde\pi_n|}{|1_{\Delta_{k(n)}}\tilde\pi_n|} \leq \frac{|(1-P)\tilde\pi_n| + |(1-P)1_{\Lambda\setminus\Delta_{k(n)}}\tilde\pi_n|}{1 - |1_{\Lambda\setminus\Delta_{k(n)}}\tilde\pi_n|} \leq \frac{2n^{-1}+n^{-1}}{1-n^{-1}} \to 0. \tag{5.37}$$

We are left with the task of showing that there exist $\Delta_k \uparrow \Lambda'$ such that (5.34) holds. By (5.28), $P^k(i,0) > 0$ if and only if there exist $i_0,\ldots,i_k$ such that $i_0 = i$, $i_k = 0$, and $a(i_{n-1},i_n) > 0$ $\forall n = 1,\ldots,k$. Therefore, if we put

$$\Delta_k := \bigcup_{l=0}^{k} \big\{i : \exists i_0,\ldots,i_l \text{ s.t. } i_0 = i,\ i_l = 0,\ a(i_{n-1},i_n) > \tfrac{1}{k}\ \forall n = 1,\ldots,l\big\}, \tag{5.38}$$

then it is not hard to see that $\Delta_k \uparrow \Lambda'$ and (5.34) holds. ∎

Lemma 35 (Sparse thinning functions) *Assume that* $\phi_m \in [0,1]^\Lambda$ *satisfy* $|\phi_m| \to \infty$. *Let* $\Delta_n \subset \Lambda$ *be finite, and let* $\varepsilon_n > 0$ *be positive constants. Then there exist finite sets* $A_n \subset \Lambda$ *such that the sets* $a\Delta_n$ *with* $a \in A_n$ *are disjoint and there exist* $m(n) \to \infty$ *such that*

$$\varepsilon_n \sum_{a \in A_n} \phi_{m(n)}(a) \to \infty \quad \text{as } n \to \infty. \tag{5.39}$$

Proof Set $\Gamma_n := \{ij^{-1} : i,j \in \Delta_n\}$. Note that $i\Delta_n$ and $j\Delta_n$ are disjoint if and only if $i \not\in j\Gamma_n$. Choose $m(n) \to \infty$ fast enough such that

$$|\phi_{m(n)}| \geq \frac{n|\Gamma_n|}{\varepsilon_n}. \tag{5.40}$$

Choose inductively $a^n_1, a^n_2, \ldots \in \Lambda$ such that

$$\phi_{m(n)} \text{ assumes its maximum over } \Lambda \backslash \bigcup_{l=1}^{k} a^n_l \Gamma_n \text{ in } a^n_{k+1}. \tag{5.41}$$

Set $B^n_1 := a^n_1 \Gamma_n$ and $B^n_{k+1} := a^n_{k+1}\Gamma_n \backslash \bigcup_{l=1}^{k} a^n_l \Gamma_n$. Then $|\phi_{m(n)}| = \sum_{k=1}^{\infty} \sum_{a \in B^n_k} \phi_{m(n)}(a)$ and $\sum_{a \in B^n_k} \phi_{m(n)}(a) \leq |\Gamma_n| \phi_{m(n)}(a^n_k)$. Therefore, we can choose $k(n)$ large enough such that

$$\sum_{k=1}^{k(n)} \phi_n(a^n_k) \geq \tfrac{1}{2} \frac{|\phi_{m(n)}|}{|\Gamma_n|}. \tag{5.42}$$

It follows from (5.40) and (5.42) that $A_n := \{a^n_1, \ldots, a^n_{k(n)}\}$ satisfies

$$\varepsilon_n \sum_{a \in A_n} \phi_{m(n)}(a) = \varepsilon_n \sum_{l=1}^{k(n)} \phi_n(a^n_k) \geq \frac{\varepsilon_n |\phi_{m(n)}|}{2|\Gamma_n|} \geq n/2 \to \infty. \tag{5.43}$$

∎

Proof of Lemma 6 Assume that $\phi_m \in [0,1]^\Lambda$ satisfy $|\phi_m| \to \infty$. Choose Δ_n, ε_n as in Lemma 34 and $A_n, m(n)$ as in Lemma 35. Define

$$\pi^a_n(i) := \pi_n(ai) \qquad (i \in \Lambda, a \in A_n), \tag{5.44}$$

and put

$$\psi_n := \varepsilon_n \sum_{a \in A_n} \phi_{m(n)}(a) \pi^a_n. \tag{5.45}$$

Then $\psi_n \in [0,1]^\Lambda$ since the π^a_n are supported on the sets $a\Delta_n$, which are disjoint. Moreover, $|\psi_n| = \varepsilon_n \sum_{a \in A_n} \phi_{m(n)}(a) \to \infty$ by Lemma 35, and

$$|P\psi_n - \psi_n|/|\psi_n| \leq \varepsilon_n \sum_{a \in A_n} \phi_{m(n)}(a) |P\pi^a_n - \pi^a_n|/|\psi_n| = |P\pi_n - \pi_n| \to 0, \tag{5.46}$$

by Lemma 34. Therefore, by Lemma 31, $\langle X_0 \wedge 1, \psi_n \rangle \overset{\mathrm{P}}{\to} \infty$, i.e.,

$$\varepsilon_n \sum_{a \in A_n} \phi_{m(n)}(a) \langle X_0 \wedge 1, \pi^a_n \rangle \overset{\mathrm{P}}{\to} \infty. \tag{5.47}$$

Define

$$Y_n(a) := \begin{cases} 1_{\{\langle X_0, 1_{a\Delta_n}\rangle \neq 0\}} & \text{for } a \in A_n, \\ 0 & \text{otherwise.} \end{cases} \tag{5.48}$$

Since $\langle X_0 \wedge 1, \pi_n^a \rangle \leq Y(a)$ $(a \in A_n)$, (5.47) implies

$$\varepsilon_n \langle Y_n, \phi_{m(n)} \rangle \overset{\mathrm{P}}{\to} \infty. \tag{5.49}$$

By Lemma 32, it follows that

$$P[\mathrm{Thin}_{\varepsilon_n \phi_n}(Y_n) = 0] \to 0. \tag{5.50}$$

Therefore, by Lemma 33,

$$P[\mathrm{Thin}_{\phi_{m(n)}}(X_1) = 0] \leq P[\mathrm{Thin}_{\phi_n}(\mathrm{Thin}_{\varepsilon_n}(Y_n)) = 0] \to 0. \tag{5.51}$$

We have proved that if $\phi_m \in [0,1]^\Lambda$ satisfy $|\phi_m| \to \infty$, then there exists a subsequence $\phi_{m(n)}$ such that $P[\mathrm{Thin}_{\phi_{m(n)}}(X_1) = 0] \to 0$. Now if $\phi_n \in [0,1]^\Lambda$ satisfy $|\phi_n| \to \infty$, then every subsequence ϕ'_n of the ϕ_n contains a further subsequence ϕ''_n such that $P[\mathrm{Thin}_{\phi''_n}(X_1) = 0] \to 0$. This implies that $P[\mathrm{Thin}_{\phi_n}(X_1) = 0] \to 0$. ∎

References

[BES02] N.H. Barton, A.M. Etheridge, and A.K. Sturm. Coalescence in a random background. WIAS Berlin, Preprint No. 756, 2002. To appear in *Ann. Appl. Probab.*

[Che87] M.F. Chen. Existence theorems for interacting particle systems with non-compact state space. *Sci. China Ser. A* 30, 148–156, 1987.

[DG99] D.A. Dawson and A. Greven. Hierarchically interacting Fleming-Viot processes with selection and mutation: Multiple space time scale analysis and quasi-equilibria. *Electron. J. Probab.* 4, Paper No.4, 81 p., 1999.

[DDL90] W. Ding, R. Durrett, and T.M. Liggett. Ergodicity of reversible reaction diffusion processes. *Probab. Theory Relat. Fields* 85(1), 13–26, 1990.

[DK99] P. Donelly and T.G. Kurtz. Genealogical processes for Fleming-Viot models with selection and recombination. Ann. Appl. Probab. 9, 1091-1148, 1999.

[EK86] S.N. Ethier and T.G. Kurtz. *Markov Processes; Characterization and Convergence.* John Wiley & Sons, New York, 1986.

[FS03a] K. Fleischmann and J.M. Swart. Extinction versus exponential growth in a supercritical super-Wright-Fisher diffusion. *Stochastic Processes Appl.* 106(1), 141-165, 2003.

[FS03b] K. Fleischmann and J.M. Swart. Trimmed trees and embedded particle systems. WIAS Berlin, Preprint No. 793, 2002. To appear in *Ann. Probab.*

[Kre85] U. Krengel. *Ergodic Theorems.* Walter de Gruyter, Berlin, 1985.

[KN97] S.M. Krone and C. Neuhauser. Ancestral processes with selection. *Theor. Popul. Biol.* 51(3), 210–237, 1997.

[Lig99] T.M. Liggett. *Stochastic Interacting Systems: Contact, Voter and Exclusion Process.* Springer-Verlag, Berlin, 1999.

[LS81] T.M. Liggett and F. Spitzer. Ergodic theorems for coupled random walks and other systems with locally interacting components. *Z. Wahrsch. verw. Gebiete* 56, 443–468, 1981.

[Mou92] T.S. Mountford. The ergodicity of a class of reversible reaction-diffusion processes. *Probab. Theory Relat. Fields* 92(2), 259–274, 1992.

[MT95] C. Müller and R. Tribe. Stochastic p.d.e.'s arising from the long range contact and long range voter processes. *Probab. Theory Relat. Fields* 102(4), 519–545, 1995.

[Neu90] C. Neuhauser. An ergodic theorem for Schlögl models with small migration. *Probab. Theory Relat. Fields* 85(1), 27–32, 1990.

[Pat88] A.L.T. Paterson. *Amenability.* American Mathematical Soc., Providence, 1988.

[Sch72] F. Schlögl. Chemical reaction models and non-equilibrium phase transitions. *Z. Phys.* 253, 147–161, 1972.

[SS80] T. Shiga and A. Shimizu. Infinite dimensional stochastic differential equations and their applications. *J. Math. Kyoto Univ.* 20, 395–416, 1980.

[Swa99] J.M. Swart. *Large Space-Time Scale Behavior of Linearly Interacting Diffusions.* PhD thesis, Katholieke Universiteit Nijmegen, 1999 http://helikon.ubn.kun.nl/mono/s/swart_j/largspscb.pdf.

[Swa00] J.M. Swart. Clustering of linearly interacting diffusions and universality of their long-time limit distribution. *Probab. Theory Relat. Fields* 118, 574–594, 2000.

[YW71] T. Yamada and S. Watanabe, On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* 11, 155–167, 1971.

Siva R. Athreya

Stat. Math. Unit
Indian Statistical Institute
Delhi Centre 7
S.J.S. Sansanwal Marg
New Delhi -110016, India
e-mail: athreya@isid.ac.in

Jan M. Swart

Mathematical Institute
University Erlangen-Nuremberg
Bismarckstraße $1\frac{1}{2}$
D–91054 Erlangen, Germany
e-mail: swart@math.tu-berlin.de