Alea 2, 1–66 (2006)

The Finite System Scheme for State-dependent interacting multitype Branching Systems

Peter Pfaffelhuber

LMU Munich, Department Biologie II, Großhaderner Straße 2, 82131 Planegg-Martinsried, Germany
E-mail address: p.p@lmu.de
URL: http://www.zi.biologie.uni-muenchen.de/~pfaffelhuber

Abstract. Consider a branching process which models the evolution of a population of particles that have different types. Particles move in the site space $\mathbb{Z}^d$, according to some random walk and branch binary critically. The branching rate of a particle at $\xi \in \mathbb{Z}^d$ is a function h of the total mass present at site ξ. This multitype model has been introduced in Dawson and Greven (2003).

As in many interacting systems the long-time behaviour differs sharply for low and high dimensions. In low dimensions and also for any finite system local extinction occurs while in high dimensions there are ergodic equilibrium measures.

Our main task is to develop new techniques to prove the finite system scheme for this model with interaction between types. The finite system scheme compares large finite and infinite systems in a time scale proportional to the size of the finite system. Asymptotically large finite systems are on their way dying out locally in the equilibrium of the infinite system where the parameter of the equilibrium is given by the density of the finite system.

In order to prove the finite system scheme a coupling of state-dependent interacting branching systems is introduced which should be of interest in a wider context.

0. Introduction

Multitype branching processes are today extensively studied. In these processes each type of particles has its own branching mechanism which can depend e.g. on the frequencies of other types. The simplest non-trivial examples with two types are catalytic branching and mutually catalytic branching. See Dawson and Fleischmann (1997) and Mytnik (1998) for papers on these models.

Received by the editors 10.6.2005, accepted 19.12.2005.

2000 *Mathematics Subject Classification.* 60K35 (Primary), 60J80, 60G57 (Secondary).

Key words and phrases. Spatial branching process, finite and infinite interacting processes, multitype branching process with type interaction, coupling of multidimensional processes.

Supported by a DFG Schwerpunkt.

We study here another multitype process introduced in Dawson and Greven (2003). Each particle no matter which type branches binary critically with a branching rate which is a function h of the total mass at the site. One extreme case is the case of independent branching (h being constant). A consequence of h being a function of the total mass is that the total mass per site process is an autonomous branching process.

Our focus is on studying this branching system in a discrete spatial setting, i.e. $\mathbb{Z}^d$. Uniqueness and the long-time-behaviour which reflects the long-time behaviour of the total mass per site process were derived in Dawson and Greven (2003).

The long-time behaviour of infinite systems, i.e. systems on $\mathbb{Z}^d$ is similar to many other interacting systems: one finds local extinction in low and nontrivial equilibria in high dimensions. But systems with only finitely many components will be extinct in the long run independent of the dimension. The deeper comparison of these two regimes is the finite system scheme which is subject of this paper for the above model. To be exact it states that there is a time scale β_N for which asymptotically infinite and finite systems begin to behave differently. Recall that in the equilibrium case for infinite systems the spatial density of the components is a constant for all times. On the other side this is not true for finite systems. Here the density fluctuates and it fluctuates exactly in the time scale β_N. The main result is that infinite and finite systems locally behave similarly up to times β_N. For times of order β_N finite systems are locally in the equilibrium measures of the infinite system but the intensity is given by the random density of the finite process.

The first author to compare finite and infinite systems was Dobrushin Dobrushin (1971). Whereas he concentrated on approximations of infinite systems by finite systems and of their stationary distributions the finite system scheme presented here is dealing with the question how long finite and infinite systems can be compared depending on the size of the finite system. The finite system as described here was introduced in Cox and Greven (1990), first for particle systems as the voter model and critical branching random walk. Later the results have been established for interacting diffusions Cox et al. (1995), Fleming-Viot processes Dawson et al. (1995) and mutually catalytic branching Cox et al. (2004).

In Greven et al. (2005) the finite system scheme is applied to the historical version of spatial interacting Moran models and interacting Fisher-Wright diffusions. One key idea of the proofs of the finite system scheme in the above mentioned papers is the analysis of a spatial coalescent by means of a comparison of the random walk paths for the finite and infinite system. With the formulation on a historical level this idea is made rigorous by considering not only states at a fixed time but whole paths of descent of particles.

Closely related to the finite system scheme for models on the lattice $\mathbb{Z}^d$ is the mean field finite system scheme for models on the hierarchical group. This is part of the multiple space time scale analysis carried out e.g. in Dawson and Greven (1996), Cox et al. (2004) and the forthcoming paper Dawson and Greven (2006).

Another topic is the question of universality. For the comparison of finite and infinite systems both in the lattice case and in the case of the hierarchical group a nonlinear map can be defined. It can have fixed points or fixed shapes. These fixed points and their stability is also of great interest because the corresponding models will be the prototypes for spatial evolutionary models. See Baillon et al. (1995), Baillon et al. (1997) for the one-type case on the hierarchical group and Cox et al. (1998) for the one-type case on the lattice.

There is one methodological point that makes our model interesting and challenging. As we want to compare finite and infinite versions of our systems a natural thing to do is coupling (see Chapter 3). For all one-type models it is possible to construct a coupling with the property that the expected distance of the two coupled spatial systems will decrease over time. This fact is crucial for the proofs of various statements (see e.g. Shiga (1992), Cox and Greven (1994b) and Cox et al. (1995), Dawson and Greven (2003)). But it is due to the multitype situation that any coupling for our model will not have this property.

To understand this consider the branching mechanism. In one-type-situations once there is a perfect matching in the coupled process this matching will not be destroyed by the branching mechanism. But in multitype situations a perfect match in one type can be destroyed by the branching mechanism by a mismatch in one of the other types and this is a general problem for multitype models.

We overcome this difficulty by giving minimal assumptions we have to make in order to have successful couplings. By a successful coupling we mean one where the distance measured becomes small as time evolves.

As a consequence of these difficulties arising in a multitype context our proof of the finite system scheme will be different from the proofs in earlier papers. Details can be found at the beginning of Chapter 4. Our main new technique in proving the finite system scheme is that for the coupled processes it is enough that the increase of the expected distance of the two processes becomes small. To derive this result we need the fact that the expected distance of the total masses decreases. So it is extremely helpful that the process of total masses has already been studied in Cox and Greven (1994b) and Cox et al. (1995).

0.0.1. *Outline.* We begin in Chapter 1 with the definition of our models, infinite and finite. In Chapter 2 we introduce the finite system scheme, state our main result and discuss it. The coupling we mentioned above and which is quite involved is subject to Chapter 3. Therefore we have to define the matching of finite (non-random and random) measures and define a coupled dynamics for the infinite and finite systems. Our main theorem, Theorem A, is proved in Chapter 4 by carrying out the above program. The proof consists of several steps. On overview over the proof can be found at the beginning of Chapter 4.

This paper covers the main part of the author's PhD thesis. If the reader would like some related material or some more details (s)he should consult Pfaffelhuber (2003).

1. The Models

The model we consider throughout this paper was introduced in Dawson and Greven (2003) and put forth in Dawson and Greven (2006). It is a model of measure-valued interacting stochastic processes and also a spatial model with site space $\Lambda := \mathbb{Z}^d$.

A state of the process (in the diffusion limit) will be written informally

$$x_\xi(t)(\{u\}) = \text{ mass at time } t \text{ in } \xi \in \Lambda \text{ of type } u \in I. \tag{1.1}$$

In the particle model which is easier to visualise we observe different particles which have types in $I := [0,1]$ at all sites. These types are best imagined by thinking of particles on the lattice $\mathbb{Z}^d$ having different colours.

These particles not only migrate according to the rates of a random walk kernel but also branch binary critically. The particles at a site $\xi \in \mathbb{Z}^d$ interact by a common state-dependent branching rate. That means that the branching rate for particles at ξ is $h(x_\xi(I))$ for some function h. Therefore particles (and the same holds for types) are not independent if h is non-trivial, i.e. not a constant.

As we will see our model gives more structure to a model already studied in detail, interacting (one-type) diffusions (see Shiga (1992), Cox and Greven (1994a), Cox et al. (1995), Cox et al. (1996) and Cox et al. (1998)). This model arises when we forget about the different types and consider total masses at each site we come to interacting diffusions.

We will also consider a finite version of our model with finite site space. These finite systems are easily obtained from the infinite systems by changing the random walk kernel that defines the migration between colonies.

1.1. *Infinite systems.* The interacting diffusions we want to study are diffusion limits of particle processes. Furthermore we will need later (in Chapter 3) again the approximation scheme of interacting diffusions by particle processes. These are two good reasons to give a motivation by informally describing a particle model. Then we come to its diffusion limit.

Motivation: a particle system. We consider particles on $\Lambda = \mathbb{Z}^d$. Let here

$$x_\xi(t)(\{u\}) = \#\text{ particles in } \xi \in \Lambda \text{ of type } u \in I \text{ at time } t \tag{1.2}$$

such that $x_\xi(t) \in \mathcal{M}_f(I)$.

The dynamics of the process is given as follows:

(i) Each particle moves independently of the others according to a random walk with transposed irreducible transition kernel $a(\cdot,\cdot)$, which gives the rates for a Markov chain with state space Λ. The rate $a(\eta,\xi)$ is the rate at which a particle located in ξ moves to the site η. The random walk kernel satisfies for $\xi,\eta \in \Lambda$

$$0 \le a(\eta,\xi) < \infty, \qquad a(\eta,\xi) = a(0,\xi-\eta), \qquad \sum_{\eta\in\Lambda} a(\eta,\xi) =: c < \infty. \tag{1.3}$$

(ii) If a particle is at time t at site ξ, it branches with rate $h(x_\xi(t)(I))$ for some

$$h : [0,\infty) \to [0,\infty) \text{ locally Lipschitz, } h(x) = o(x) \text{ as } x \to \infty. \tag{1.4}$$

This means that all particles at a site wait an exponentially distributed amount of time and then by a fair coin toss it is decided which particle splits and whether the particle splits in two particles of the same type or disappears. The rate of this exponential distribution is given by the function h that depends only on the total mass at site ξ. We assume

$$\exists \mathfrak{a} \in (0,\infty] : \qquad h(\theta) > 0 \quad \iff \quad \theta \in (0,\mathfrak{a}). \tag{1.5}$$

(iii) The transitions in (i) and (ii) are independent.

Observe that only in the case $\mathfrak{a} = \infty$ in (1.5) the assumption $h(x) = o(x)$ of (1.4) really plays a role. The assumption of h being locally Lipschitz is not necessary for the particle process. It is needed for the uniqueness result in the diffusion limit. However we don't want to distinguish these two cases in our assumptions.

Considering spatial branching multitype processes that are Markov. The most restricting assumption we made in our model is that the branching rate only depends on the total mass. Therefore no particle and no type of particles has a selective advantage.

As we have just seen the branching rates for the particle process only depend on the total masses per site. As we will need total masses very often we define for any $X = (x_\xi)_{\xi\in\Lambda} \in (\mathcal{M}_f(I))^\Lambda$

$$\overline{X} = (\overline{X}(t))_{t\geq 0}, \qquad \overline{X}(t) = (\overline{x}_\xi(t))_{\xi\in\Lambda}, \qquad \overline{x}_\xi(t) := x_\xi(t)(I) \in \mathbb{R}_+. \tag{1.6}$$

The diffusion limit for the infinite model. Now we want to consider the diffusion limit of the above particle model. Therefore imagine particles having masses and the process no longer counts particles but their masses. The diffusion limit means all particles have mass ϵ, we increase the number of initial particles by a factor ϵ^{-1}, accelerate the branching by a factor ϵ^{-1} and let ϵ go to 0. We will derive a Markov process with continuous paths which we will denote $X := (X(t))_{t\geq 0}$ with $X(t) = (x_\xi(t))_{\xi\in\Lambda}$. The set of initial states which will also be the state space of the process is denoted $\mathcal{E} \subseteq (\mathcal{M}_f(I))^\Lambda$ and introduced after having defined the dynamics of the process.

To characterise that process we have to introduce the corresponding pregenerator. For that consider smooth functions $F : \mathcal{E} \to \mathbb{R}$ of the form

$$F(X) = \prod_{i=1}^{n} g_i(\langle x_{\xi_i}, f_i\rangle), \qquad \xi_i \in \Lambda, f_i \in \mathcal{C}(I), g_i \in \mathcal{C}_b^2(\mathbb{R}), (i = 1, \dots, n) \tag{1.7}$$

for some $n \in \mathbb{N}$. Here $\mathcal{C}(I)$ is the set of all continuous functions on I, $\mathcal{C}_b^2(\mathbb{R})$ is the set of all bounded continuous functions with second derivatives that are also bounded and continuous. Furthermore $\mathcal{A}$ is the algebra generated by such functions which is separating in $\mathcal{P}(\mathcal{E})$.

To introduce differential operators on $\mathcal{A}$ we define derivatives for functions $F \in \mathcal{A}$ by

$$\frac{\partial F(X)}{\partial x_\xi}(u) = \lim_{\epsilon\to 0} \frac{F_{X,\xi}(x_\xi + \epsilon\delta_u) - F(X)}{\epsilon} \tag{1.8}$$

with

$$F_{X,\xi}(x) := F(X_\xi(x)), \qquad (X_\xi(x))_{\xi'} := \begin{cases} x, & \xi' = \xi \\ x_{\xi'}, & \text{else.} \end{cases} \tag{1.9}$$

Simply speaking $F_{X,\xi}$ is derived from F by holding all variables of $X = (x_{\xi'})_{\xi'\in\Lambda}$ except x_ξ constant. The second derivative is defined by

$$\frac{\partial^2 F(X)}{\partial x_\xi \partial x_\eta}(u, v) := \frac{\partial}{\partial x_\xi}\Big(\frac{\partial F(X)}{\partial x_\eta}(u)\Big)(v). \tag{1.10}$$

For $F \in \mathcal{A}$ these derivatives are weakly continuous (i.e. continuous with respect to the topology of weak convergence) by their form given in (1.7).

1.1.1. *Remark.* Instead of the definitions (1.8) and (1.10) it is also possible to use the definition

$$\frac{\partial_\star F(X)}{\partial_\star x_\xi}(u) = \lim_{\epsilon\to 0} \frac{F_{X,\xi}(\pi_\epsilon x_\xi + \epsilon\delta_{u_\epsilon}) - F_{X,\xi}(\pi_\epsilon x_\xi)}{\epsilon} \tag{1.11}$$

for functions of the form (1.7). This definition will be helpful for defining a coupled process in Section 3. Here

$$u_\epsilon := \sup\{n\epsilon : n\epsilon \le u, n \in \mathbb{N}\}. \tag{1.12}$$

and $\pi_\epsilon x$ is given by its distribution function

$$\pounds_{\pi_\epsilon x} := \inf\{F : I \to \epsilon\mathbb{N}_0 : F(u) = F(u_\epsilon), F(u_\epsilon) \ge \pounds_x(u_\epsilon),\ u \in I\}. \tag{1.13}$$

So $\pounds_{\pi_\epsilon x}$ is the smallest function which is constant on each $[k\epsilon, (k+1)\epsilon)$ and lies above $\pounds_x$ in the points u_ϵ. The second derivative is defined accordingly as in (1.10). The fact that this definition coincides with the definitions in (1.8) and (1.10) for a function $F(x) = g(\langle x, f\rangle)$ can be established by using a Taylor series approximation of g around $\langle x, f\rangle$.

We define the following differential operator acting on functions $F \in \mathcal{A}$

$$\begin{aligned}(GF)(X) := &\sum_{\xi,\eta\in\Lambda} a(\xi,\eta) \int_I \frac{\partial F(X)}{\partial x_\xi}(u)(x_\eta(du) - x_\xi(du)) \\ &+ \sum_{\xi\in\Lambda} h(\overline{x}_\xi) \int_I \int_I \frac{1}{2}\frac{\partial^2 F(X)}{\partial x_\xi^2}(u,v) x_\xi(du)\delta_u(dv).\end{aligned} \tag{1.14}$$

By $S(t)$ we denote the semigroup corresponding to this pregenerator which is uniquely given by Theorem 1. But before we come to this we give the initial states of our process.

Initial states. We recall the set of initial states of our process. We restrict the initial state to a subset $\mathcal{E} \subseteq (\mathcal{M}_f(I))^\Lambda$ which is defined in (1.17). For this, fix a map $\gamma : \Lambda \to (0,\infty)$ with $\sum_{\xi\in\Lambda}\gamma(\xi) < \infty$. Then define for $X = (x_\xi)_{\xi\in\Lambda} \in (\mathcal{M}_f(I))^\Lambda$ the γ-norms

$$||X||^p := ||\overline{X}||^p := \sum_{\xi\in\Lambda} \gamma(\xi)\overline{x}_\xi^p \qquad (p = 1, 2, \dots) \tag{1.15}$$

For γ, we require

$$\exists M < \infty : \quad \sum_{\xi\in\Lambda} \gamma(\xi) a(\xi,\eta) \le M\gamma(\eta). \tag{1.16}$$

Such a γ can always be found; (see e.g. Liggett and Spitzer (1981)). Then

$$\begin{aligned}\mathcal{E} := \mathcal{E}_\gamma &:= \{X \in (\mathcal{M}_f(I))^\Lambda : ||X||^1 < \infty\}, \\ \mathcal{L}_p(\mathcal{E}) &:= \{\mu \in \mathcal{P}(\mathcal{E}) : \langle \mu, ||X||^p\rangle < \infty\} \qquad (p \in \mathbb{N}).\end{aligned} \tag{1.17}$$

For our convergence results we equip $\mathcal{M}_f(I)$ with the topology of weak convergence (see Ethier and Kurtz (1986) and Billingsley (1999) for details). Furthermore we take the product topology on the product space $(\mathcal{M}_f(I))^\Lambda$. For particle models we denote by $\mathcal{E}^p$ the restriction of $\mathcal{E}$ to counting measures.

One frequently used possibility to define a Markov process is to characterise it via a martingale problem. This is an idea of Stroock and Varadhan.

Definition 1.1. *The solution of the (G,μ)-(local)martingale problem for a pregenerator G on an algebra $\mathcal{A}$ and $\mu \in \mathcal{P}(E)$ is a process X which is the canonical process for some probability distribution P on $\mathcal{C}((0,\infty),\mathcal{E})$ with $\mathcal{L}(X_0)=\mu$. The distribution of X_0 and of $(X_t - X_0)_{t\geq 0}$ are independent and for all $F \in \mathcal{A}$*

$$\Big((F(X_t) - F(X_0) - \int_0^t (GF)(X_s)ds\Big)_{t\geq 0} \tag{1.18}$$

is a P-(local) martingale with respect to the canonical filtration.

Next we recall results according to Theorem 1 and Lemma 2.1 of Dawson and Greven (2003).

Theorem 1.2. *Let $\mu \in \mathcal{L}_1(\mathcal{E})$ and G from (1.14). Then the (G,μ) local martingale problem is well posed. The solution is a Markov process X on $\mathcal{E}$. If either $\mu \in \mathcal{L}_2(\mathcal{E})$ or h is bounded then also the (G,μ) martingale problem is well-posed.*

Remarks.

1. We point out that for $\mu \in \mathcal{P}(\mathcal{E})$, writing expectations if they exist, we distinguish two cases
   - For $f:\mathcal{E}\to\mathbb{R}$ we define $\langle\!\langle \mu, f\rangle\!\rangle$ to be the expectation of f under μ.
   - For $f:\mathcal{E}^{\mathbb{R}_+}\to\mathbb{R}$ we define $\mathbf{E}^\mu[f]$ to be the expectation of f under the distribution of the process, i.e. the path measure started in μ.

   Any other expectation or integration will be denoted by $\langle\cdot,\cdot\rangle$.
2. Note that if $\nu \in \mathcal{P}((\mathcal{M}_f(I))^\Lambda)$ is space-shift invariant, i.e. $\sigma_\xi \nu = \nu$ with $(\sigma_\xi X)_\eta = x_{\xi+\eta}$ with finite intensity, i.e. $\langle\!\langle \nu, \overline{x}_0\rangle\!\rangle < \infty$ then $\langle\!\langle \nu, \|X\|^1\rangle\!\rangle < \infty$, so ν is supported in $\mathcal{E}$.
3. The process of total masses was already studied in Cox et al. (1995). The set $\mathcal{E}$ only depends on total masses of the process and so it can be deduced as mentioned in that paper that
$$\mathbf{P}[X(t)\in\mathcal{E},\ t\geq 0]=1. \tag{1.19}$$
   So $\mathcal{E}$ is not only the set of initial states but also the state space of the process.

1.2. *Finite systems.* First of all it is important to note that the notion of *finite* systems always means that only *finitely many sites* are involved. We saw in Section 1.1 for the infinite systems that the interaction between sites is realized by a random walk kernel. To derive a finite system all we do is restrict this migration to a finite site space. We want to do this without losing spatial homogeneity.

We take

$$\Lambda^N := (-N,N]^d \cap \mathbb{Z}^d =: \mathbb{Z}^d_N. \tag{1.20}$$

Addition mod $2N$ makes Λ^N a group. The model with space Λ^N will be called $X^N = (X^N(t))_{t\geq 0}$ with $X^N(t) = (x^N_\xi(t))_{\xi\in\Lambda^N}$. For any $\xi \in \mathbb{Z}^d_N$ the branching mechanism is exactly the same as for the infinite system. For the migration we restrict the random walk to $\mathbb{Z}^d_N$ by taking the random walk kernel

$$a^N(\xi,\eta) := \sum_{k\in\Lambda} a(\xi,\eta+2Nk) \qquad (\xi,\eta\in\mathbb{Z}^d_N), \tag{1.21}$$

so the paths of the random walk on $\mathbb{Z}^d_N$ emerge from those of the random walk on $\mathbb{Z}^d$ by projection on the torus, i.e. by considering each site mod $2N$.

Let us define the finite system of size N via their pregenerators. As the only difference to the infinite system is the migration mechanism it should be clear that the pregenerator is given by

$$\begin{aligned}(G^N F)(X^N) := &\sum_{\xi,\eta\in\Lambda^N} \int_I \frac{\partial F(X^N)}{\partial x_\xi^N}(u) a^N(\xi,\eta)(x_\eta^N(du) - x_\xi^N(du)) \\ &+ \sum_{\xi\in\Lambda^N} h(\overline{x}_\xi^N) \int_I \int_I \frac{1}{2} \frac{\partial^2 F(X^N)}{\partial x_\xi^N \partial x_\xi^N}(u,v) x_\xi^N(du)\delta_u(dv)\end{aligned} \tag{1.22}$$

acting on functions that are contained in the algebra $\mathcal{A}^N$ that is generated by functions $F : \mathcal{E}^N \to \mathbb{R}$ of the form

$$F(X^N) = \prod_{i=1}^n g_i(\langle x_{\xi_i}^N, f_i\rangle), \qquad \xi_i \in \Lambda^N, f_i \in \mathcal{C}(I), g_i \in \mathcal{C}_b^2(\mathbb{R}) \ (i = 1,\dots,n) \tag{1.23}$$

for some $n \in \mathbb{N}$.

For the finite systems the choice of the initial states is easier than for infinite systems. Any $X^N(0) \in (\mathcal{M}_f(I))^{\Lambda^N}$ will do. That is because finite systems always start with finite total mass and so no explosion in finite time can occur. To be more exact the explosions that could occur when starting the infinite system in some $X(0) \in (\mathcal{M}_f(I))^\Lambda \setminus \mathcal{E}$ at $\xi \in \Lambda$ result from infinite mass that moves to ξ in finite time. So we have the set of initial states for the finite system as well as the state space of the process

$$\mathcal{E}^N := (\mathcal{M}_f(I))^{\Lambda^N} \tag{1.24}$$

and define analogously to the infinite case for $p \in \mathbb{N}$

$$\begin{aligned}||X^N||^p &:= \sum_{\xi\in\Lambda^N} \overline{x}_\xi^p, \\ \mathcal{L}_p(\mathcal{E}^N) &:= \{\mu \in \mathcal{P}(\mathcal{E}^N) : \langle\!\langle \mu, ||X^N||^p \rangle\!\rangle\} < \infty.\end{aligned} \tag{1.25}$$

As the proof of Theorem 1.2 of Dawson and Greven (2003) does not rely on the fact that $\mathbb{Z}^d$ is infinite we immediately obtain:

Theorem 1.3. *Let $\mu \in \mathcal{L}_1(\mathcal{E}^N)$ and G^N from (1.22). Then the (G^N,μ) local martingale problem is well posed. The solution is a Markov process X^N on $\mathcal{E}^N$. If either $\mu \in \mathcal{L}_2(\mathcal{E}^N)$ or h is bounded then also the (G^N,μ) martingale problem is well posed.*

We make the convention throughout the whole paper that any time a superscript N appears it is implicit that there is some finite system of size N underlying.

1.3. *Review of some basic properties.* Here we want to make precise the already mentioned autonomy of the process of total masses and the structure of the relative weights process. Furthermore in order to be self-contained in our formulation of Theorem A we recall some properties, especially statements from Dawson and Greven (2003). The long-time behaviour differs sharply in low and high dimensions as is already well known for other interacting systems (e.g. the voter model, branching random walk, interacting diffusions or interacting Fleming-Viot systems).

The total mass process. For the process of total masses we forget about the multitype situation we are dealing with. For $x \in \mathcal{M}_f(I)$ we already introduced the notion of total masses in (1.6).

As seen above the space $\mathcal{E}$ only restricts initial states according to their total masses at the sites. Therefore for the total mass process we define $||.||$ as in (1.15) and set

$$\overline{\mathcal{E}} = \{X \in \mathbb{R}_+^\Lambda : ||X|| < \infty\} \tag{1.26}$$

as the state space of the process.

Moreover (as shown in Dawson and Greven (2003)) the process of total masses is a system of interacting diffusions which satisfies the system of stochastic differential equations

$$d\overline{x}_\xi(t) = \sum_{\eta \in \Lambda} a(\xi, \eta)(\overline{x}_\eta(t) - \overline{x}_\xi(t))dt + \sqrt{h(\overline{x}_\xi(t))\overline{x}_\xi(t)}dw_\xi(t) \qquad (\xi \in \Lambda) \tag{1.27}$$

where $(w_\xi)_{\xi \in \Lambda}$ is a family of independent Brownian motions. This system of stochastic differential equations has a unique strong solution which was proved in Shiga and Shimizu (1980). This solution is a Markov process whose semigroup will be denoted by $\overline{S}(t)$. It is studied in Shiga (1992), Cox and Greven (1994a), Cox et al. (1995), Cox et al. (1996) and Cox et al. (1998).

Each component x_ξ of X can be written as $x_\xi = \overline{x}_\xi \hat{x}_\xi$ where x_ξ is a probability measure giving the relative weights of types. As was shown in Dawson and Greven (2003) the process $\hat{X} = (\hat{x}_\xi)_{\xi \in \Lambda}$ conditioned on the process of total masses is a time-inhogeneous Fleming-Viot process. The result that interacting Fleming-Viot processes are concentrated on atomic measures for any $t > 0$ also carries over here. That means that the process X is also concentrated on atomic measures. (see Dawson and Greven (2003), Lemma 0.1).

Long-time behaviour. To give the results on the long-time behaviour for the process X we introduce

$$\mathcal{T}_q := \mathcal{T}_q(\mathcal{E}) := \{\mu \in \mathcal{P}(\mathcal{E}) \text{ space-shift invariant, } \langle\!\langle \mu, |\overline{x}_0|^q \rangle\!\rangle < \infty\} \qquad (q > 0), \tag{1.28}$$

$$\mathcal{M}_\theta := \{\mu \in \mathcal{P}(\mathcal{E}) \text{ space-shift ergodic, } \langle\!\langle \mu, x_0 \rangle\!\rangle = \theta\} \qquad (\theta \in \mathcal{M}_f(I)), \tag{1.29}$$

$$\mathcal{I} := \{\mu \in \mathcal{P}(\mathcal{E}) : \mu S(t) = \mu,\ t > 0\} \tag{1.30}$$

We say that a space-shift invariant probability measures μ with $\langle\!\langle \mu, x_0 \rangle\!\rangle = \theta$ has intensity θ and for a convex set $\mathfrak{D}$ the set $\mathfrak{D}_e$ is the set of extreme points of $\mathfrak{D}$. For spatially constant states with intensity θ we write $\delta_{\underline{\theta}} := \bigotimes_{\xi \in \Lambda} \delta_\theta$.

Note that in (1.29) it is $\langle\!\langle \mu, x_0 \rangle\!\rangle \in \mathcal{M}_f(I)$ and so x_0 is a random finite measure with first moment measure θ. Furthermore we set

$$\hat{a}(\xi, \eta) := \frac{a(\xi, \eta) + a(\eta, \xi)}{2} \qquad (\xi, \eta \in \Lambda) \tag{1.31}$$

which we call the symmetrised random walk kernel.

The behaviour of the system differs drastically in the case of a transient versus recurrent $\hat{a}$. We give here only the result for the transient case, the full treatment of the recurrent case can be found in Theorem 2 of Dawson and Greven (2003).

Theorem 1.4. *Assume* $\hat{a}(\cdot,\cdot)$ *is transient and* h *satisfies* (1.4) *and* (1.5) *with* $\mathfrak{a} \in (0,\infty]$ *from* (1.5) *and*
$\theta \in \mathcal{M}_f(I)$ *with* $\overline{\theta} \in [0,\mathfrak{a}]\}$.

- *If* $\mathcal{L}(X(0)) = \delta_{\underline{\theta}}$ *then the weak limit* $\nu_\theta = \lim_{t\to\infty} \mathcal{L}(X(t))$ *exists.*
- *If* $\mathcal{L}(X(0)) \in \mathcal{M}_\theta$ *then* $\mathcal{L}(X(t)) \stackrel{t\to\infty}{\Longrightarrow} \nu_\theta$.
  In this case also $\mathcal{L}(\overline{X}(t)) \stackrel{t\to\infty}{\Longrightarrow} \overline{\nu}_\theta$
  where $\overline{\nu}_\theta$ *depends only on the total mass* $\overline{\theta}$.
- *Each* ν_θ *is space-shift invariant and space-shift ergodic (even spatially mixing) and has intensity* θ.
- *Each* ν_θ *is invariant under the semigroup of the process* X *and*
  $(\mathcal{I} \cap \mathcal{T}_1)_e = \{\nu_\theta : \theta \in \mathcal{M}_f(I), \overline{\theta} \in [0,\mathfrak{a}]\}$.

2. Results: The finite system scheme

The finite systems scheme is a comparison scheme between spatially infinite and spatially finite models. It was first in Cox and Greven (1990) carried out for particle models like the Voter model and branching random walk. Then it has been shown that similar statements hold for interacting diffusions in Cox et al. (1995), interacting Fleming-Viot systems in Cox and Greven (1994b) and mutually catalytic branching in Cox et al. (2004). We carry out here the analogous result for the model introduced in the last section. This is the first time where it is proved for a real multitype branching system.

As the branching mechanism is critical and finite systems start with finite initial mass it is clear that in the long-time limit they die out, i.e. converge to $\delta_{\underline{0}}$. On the other hand we saw in Theorem 1.4 that for transient migration mechanism, i.e. for $d \geq 3$, infinite systems converge to some non-trivial equilibrium. So a natural question is when significant differences in the behaviour of finite and infinite systems will take place depending on the size of the finite system. In addition to that, what is the qualitative difference between the behaviour of infinite and finite systems after the systems began to behave differently?

It will turn out that for times $t \ll |\Lambda^N|$ the finite and the infinite system behave similarly. For times $t \approx |\Lambda^N|$ the behaviour of the finite and infinite system can still be compared and for times $t \gg |\Lambda^N|$ the finite system is already too far on its way to extinction to be compared with the infinite system. Crucial in this context is the (random) density of the finite system. Recall that the density in the infinite system is not random and is given by θ if the initial state is $\mu \in \mathcal{M}_\theta$.

Next we will need some ingredients in order to formulate our main theorem.

- For some $C < \infty$ we define
  $$\mathcal{M}^N := \{\mu^N \in \mathcal{T}_2(\mathcal{E}^N) : \langle\!\langle \mu^N, (\overline{x}_0^N)^2 \rangle\!\rangle \leq C\} \qquad (N \in \mathbb{N}), \tag{2.1}$$
  i.e. for a family $(\mu^N)_{N\in\mathbb{N}}$ in $\mathcal{M}^N$ second moments exist uniformly. Here for $q > 0$
  $$\mathcal{T}_q(\mathcal{E}^N) := \{\mu^N \in \mathcal{P}(\mathcal{E}^N) \text{ space-shift invariant}, \langle\!\langle \mu^N, (\overline{x}_0^N)^q \rangle\!\rangle < \infty\}. \tag{2.2}$$
  Note that we use two different notions of space-shift invariance. First the space-shift invariance in Λ, i.e. invariance according to σ_ξ ($\xi \in \Lambda$) (see Remark 1.1.1) and second as in the above definition space-shift invariance

in Λ^N, i.e. invariance according to σ_ξ^N ($\xi \in \Lambda^N$) with

$$(\sigma_\xi^N X)_\eta = x_\zeta \text{ with } \zeta = \xi + \eta \mod N. \tag{2.3}$$

- Define the scaling constant $\beta_N := (2N)^d$ and the *time scale* $\beta_N(t) := \beta_N t$. This scale determines the time for which finite and infinite systems begin to behave differently.
- We define the *densities* for systems of size N

$$\Theta^N : \begin{cases} \mathcal{E}^N & \to \mathcal{M}_f(I) \\ X^N & \mapsto \dfrac{1}{|\Lambda^N|} \displaystyle\sum_{\xi \in \Lambda^N} x_\xi^N \end{cases} \tag{2.4}$$

and *empirical measures*

$$U^N : \begin{cases} \mathcal{E}^N & \to \mathcal{P}(\mathcal{E}^N) \\ X^N & \mapsto \dfrac{1}{|\Lambda^N|} \displaystyle\sum_{\xi \in \Lambda^N} \delta_{\sigma_\xi^N X^N}. \end{cases} \tag{2.5}$$

- Given X^N the process

$$Z^N(t) = \Theta^N(X^N(t\beta_N)) \tag{2.6}$$

is the *rescaled process of densities* with state space $\mathcal{M}_f(I)$.
- In order to state our theorem we introduce another measure-valued process $Z = (Z(t))_{t \geq 0}$. Its existence and uniqueness is given by the next lemma. This is the Markov process with pregenerator defined for the algebra generated by functions $F : \mathcal{M}_f(I) \to \mathbb{R}$ of the form

$$F(\theta) = g(\langle \theta, f \rangle), \quad f \in \mathcal{C}(I), g \in \mathcal{C}_b^2(\mathbb{R}), \tag{2.7}$$

partial derivatives as defined in (1.8) and (1.10) and

$$(G_Z F)(\theta) := \int_I \int_I \frac{1}{2} \frac{\partial^2 F(\theta)}{\partial x^2}(u, v) h^\sharp(\overline{\theta}) \theta(du) \delta_u(dv). \tag{2.8}$$

with

$$h^\sharp(\overline{\theta}) := \frac{1}{\overline{\theta}} \langle\!\langle \nu_{\overline{\theta}}, h(\overline{x}_0) \overline{x}_0 \rangle\!\rangle. \tag{2.9}$$

We refer to Z as the *limit process.*

The above process Z is a crucial ingredient for our theorem. To state the main theorem we need that the limit process is well-defined.

Lemma 2.1. *Let $\theta \in \mathcal{M}_f(I)$ and G_Z taken from* (2.8). *Then the* (G_Z, θ) *local martingale problem is well posed.*

Proof. First observe that the pregenerator (2.8) has a form similar to the pregenerator in (1.14), except there is no migration part because there is no spatial structure for Z. So to apply Theorem 1.2 we would have to prove the Lipschitz continuity of $h^\sharp$ from (2.9).

Unfortunately we only know this result if $\mathfrak{a} < \infty$ (see Cox et al. (1995), Lemma 2.12) but not for the general case. But in the proof of Theorem 1.2 the Lipschitz continuity is only needed for two things:

First, for existence and uniqueness of the total mass process. But in our case as we are dealing with a one-dimensional diffusion existence and uniqueness of the total mass process follows from a theorem by Engelbert and Schmidt (see Karatzas and Shreve (1991), Theorems 5.5.4 and 5.5.7).

Second the Lipschitz continuity of $h^\sharp$ in 0 is needed to guarantee the non-integrability of zeros of the migration and resampling rates for the dual process of relative weights. But the non-integrability for the resampling rate in our case follows directly from Lemma 1.6 of Chapter 9 of Ethier and Kurtz (1986) which was used to prove the non-integrability in the spatial case. (See Lemma 1.7 of Dawson and Greven (2003)). Moreover we don't have a migration in the dual process for the relative weights so we do not need to prove the non-integrability for the migration rates.

Therefore we don't need the Lipschitz continuity of $h^\sharp$ in 0 at all and Theorem 1.2 carries over to our assumptions. □

We now come to our main result:

Theorem A. *Assume* $\hat{a}(\cdot,\cdot)$ *is transient and* h *satisfies* (1.4) *and* (1.5) *with* $\mathfrak{a} \in (0,\infty]$ *from* (1.5) *and* $\theta \in \mathcal{M}_f(I)$ *with* $\overline{\theta} \in [0,\mathfrak{a}]$.

Let $\mathcal{L}(X^N(0)) = \mu^N \in \mathcal{M}^N$ *and* $t_N \to \infty, \frac{t_N}{\beta_N} \to t \in [0,\infty)$. *Assume one of the following:*

- $\frac{t_N}{N^2} \to \infty, \Theta^N(X^N(0)) \Rightarrow \theta \in \mathcal{M}_f(I)$,
- $t_N = \mathcal{O}(N^2), \mu^N \Rightarrow \mu \in \mathcal{M}_\theta$.

Then if $Z(0) = \theta$

$$\mathcal{L}((Z^N(t))_{t\geq 0}) \overset{N\to\infty}{\Longrightarrow} \mathcal{L}((Z(t))_{t\geq 0}), \quad \mathcal{L}(\Theta^N(X^N(t_N))) \overset{N\to\infty}{\Longrightarrow} \mathcal{L}(Z(t)) \tag{2.10}$$

$$\mathcal{L}(X^N(t_N)) \overset{N\to\infty}{\Longrightarrow} \int_{\mathcal{M}_f(I)} \nu_\theta \mathbf{P}[Z(t) \in d\theta], \tag{2.11}$$

$$\mathcal{L}((U^N(X^N(t\beta_N)))_{t>0}) \underset{FDD}{\overset{N\to\infty}{\Longrightarrow}} \mathcal{L}((\nu_{Z(t)})_{t>0}). \tag{2.12}$$

The above theorem states the following: if we take an infinite and a series of finite systems of size N, observe the finite systems in time scales $t_N = o(\beta_N)$ such that the assumptions of the theorem are fulfilled then the finite and infinite system behave similarly. That is because by the assumptions we have $\mathcal{L}(Z^N(0)) \Rightarrow \theta$ and then by (2.11)

$$\mathcal{L}(X^N(t_N)) \overset{N\to\infty}{\Longrightarrow} \int_{\mathcal{M}_f(I)} \nu_\theta \mathbf{P}[Z(0) \in d\theta] = \nu_\theta \tag{2.13}$$

where ν_θ is the equilibrium measure for the infinite system. Here we rely on the fact that the map $\theta \to \nu_\theta$ is continuous, a fact given in Lemma 4.14.

If we observe the finite system in the time scale t_N of order β_N then the finite systems will diffuse through the ergodic states of the infinite system. For large N at time $t\beta_N$ the system will locally be in an equilibrium measure where the parameter of this distribution is given by the density of the process.

For the case t_N is larger than $t\beta_N$ assume $\mathfrak{a} = \infty$. As Z is a critical multitype branching system it can only converge to δ_0. So $\mathcal{L}(Z^N(\frac{t_N}{\beta_N})) \Rightarrow \mathcal{L}(Z(\infty)) = \delta_0$ by

(2.10) and according to (2.11)

$$\mathcal{L}(X^N(t_N)) \overset{\mathrm{N}\to\infty}{\Longrightarrow} \int_{\mathcal{M}_f(I)} \nu_\theta \mathbf{P}[Z(\infty) \in d\theta] = \nu_0 = \delta_{\underline{0}}. \tag{2.14}$$

Remarks.

1. To understand the difference between (2.11) and (2.12) take a bounded Lipschitz function $F : \mathcal{E} \to \mathbb{R}$ that only depends on finitely many coordinates. As we will see in our proof (2.11) is equivalent to

$$\mathbf{E}\Big[F(X^N(t\beta_N)) - \int_{\mathcal{M}_f(I)} \langle\!\langle \nu_\theta, F \rangle\!\rangle \mathbf{P}[Z(t) \in d\theta]\Big] \to 0, \tag{2.15}$$

   whereas (see the proof of Proposition (4.18)) because of $Z^N(t) \Rightarrow Z(t)$ (2.12) becomes

$$\mathbf{E}\Big[\Big|\frac{1}{|\Lambda^N|} \sum_{\xi\in\Lambda^N} F(\sigma_\xi^N X^N(t\beta_N)) - \langle\!\langle \nu_{Z^N(t)}, F \rangle\!\rangle\Big|\Big] \to 0. \tag{2.16}$$

   Note that X^N and Z^N are defined on the same probability space but X^N and Z are not. So (2.11) says that any local function of the finite process is approximately the same as the function applied to an infinite system in a mixed equilibrium measure where the mixture is given by the densities of the finite processes. So this equation can be seen as a local picture due to the fixed window it deals with. But (2.12) means that the mean of all shifts in Λ^N of the local function are approximately the same as applying the local function to an infinite system with intensity which is exactly the density of the finite system. That means that this equation gives by all σ^N-shifts a global version of the finite system scheme as the mean of all σ^N-shifts can be seen as making a statistics on the finite state space. So (and can be seen by omitting the modulus signs in (2.16)) equation (2.12) is a stronger result than (2.11).
2. The assumption $\mu^N \Rightarrow \mu \in \mathcal{M}_\theta$ is stronger than $\Theta^N(X^N(0)) \Rightarrow \theta$ because it makes a statement about global convergence and not only on convergence of densities. (See the corresponding result Lemma 2.5 in Cox et al. (1995)). Why do we need this stronger assumption in the case $t_N = \mathcal{O}(N^2)$? The reason is that an irreducible random walk on $\mathbb{Z}^d$ with $d \geq 3$ needs a time t_N, $\frac{t_N}{N^2} \to \infty$ to be distributed uniformly on the torus, i.e. for such a time scale we have that (see Lemma 2.1 in Cox et al. (1995))

$$\lim_{N\to\infty} \sup_{t\geq t_N} \sup_{\xi,\eta\in\Lambda^N} (2N)^d |a_t^N(\xi,\eta) - (2N)^{-d}| = 0. \tag{2.17}$$

3. Let us also say something about the convergence in (2.11). On the left hand side we have finite systems, i.e. distributions on $\mathcal{E}^N$ while on the right side a probability measure on $\mathcal{E}$ appears. Formally we introduce extension operators

$$\Phi_N : \mathcal{E}^N \to \mathcal{E} \text{ with } (\Phi_N(X^N))_\xi = x_\eta^N \text{ if } \eta = \xi \mod N. \tag{2.18}$$

   Using these we set for $\mu \in \mathcal{P}(\mathcal{E}),\ \mu^N \in \mathcal{P}(\mathcal{E}^N)$

$$\mu^N \overset{N\to\infty}{\Longrightarrow} \mu \;:\Longleftrightarrow\; \Phi_N \mu^N \overset{N\to\infty}{\Longrightarrow} \mu. \tag{2.19}$$

4. An important question in that setting is: why is the time scale β_N the right time scale? There are some reasons why β_N should be the time scale for which the finite and infinite systems stay comparable. As the interaction of sites only occurs via the migration mechanism and the finite and infinite processes only differ by their migration kernel the time scale β_N in our model should only be dependent on the migration mechanisms and not on the branching. In Cox et al. (1995) a statement about the comparison of random walks on the torus and on the whole lattice is given. It states that in a time scale t_N that is of order β_N with $\frac{t_N}{\beta_N} \to t > 0$ two random walks on the torus will have spent t units of time longer at the same site than the corresponding random walks on $\mathbb{Z}^d$. Exactly for such a time scale
$$\lim_{N\to\infty} \int_0^{t_N} \hat{a}_{2s}^N(\xi,\eta)ds = \hat{A}(\xi,\eta) + t \quad \text{where} \quad \hat{A}(\xi,\eta) = \int_0^\infty \hat{a}_{2s}(\xi,\eta)ds \tag{2.20}$$
is the Green's function of $\hat{a}(\cdot,\cdot)$.
   Therefore in that time scale two particles on the torus will have had a stronger interaction than two particles on $\mathbb{Z}^d$ leading to a different behaviour in that time scale.
5. In Greven et al. (2005) the first proof of a *historical* finite system scheme will appear. The treated models in that paper are interacting Moran models and interacting Fisher-Wright diffusions. Both are related to the so-called look-down process introduced in Donnelly and Kurtz (1996) which is fundamental for the treatment of interacting Moran- and Fisher-Wright models in Greven et al. (2005). Other techniques used there are a strong (i.e. historical) duality between these (i.e. look-down-, Moran- and Fisher-Wright-) processes and coalescing processes. With help of this duality the scenario described in 3. above can be made rigorous. The key point here is the extension of known results about the comparison of random walks on the infinite lattice and on corresponding finite lattices (i.e. the statements from Remarks 2. and 4. about random walks) to these coalescent processes to show the relation of the genealogies of the infinite model and the finite model.
6. In Cox et al. (1995) it was proved an analogous statement to Theorem A for the process of total masses. This will be needed in the proof of Theorem A. To be precise the above theorem is the same as the one proved in Cox et al. (1995) if you replace all measures by the corresponding measures in $\mathcal{P}(\overline{\mathcal{E}})$ and $\mathcal{P}(\overline{\mathcal{E}}^N)$, and use $\overline{X}^N$ the processes of total masses instead of X^N. Furthermore the limit process $\overline{Z}$ is in this case the solution of the stochastic differential equation
$$d\overline{Z}(t) = \sqrt{h^\sharp(\overline{Z}(t))\overline{Z}(t)}dw(t). \tag{2.21}$$
Here the measures $\nu_{\overline{\theta}}$ are the equilibrium measures of the process of total masses and arise by the long-time behaviour of the system of total masses, i.e. the theorem analogous to Theorem 1.4. (See Cox and Greven (1994a), Shiga (1992) and Cox et al. (1996).)
   Actually in Cox et al. (1995) the assumption
$$\lim_{\theta\to\infty} \frac{h(\theta)}{\theta} < \frac{1}{\hat{A}(0,0)} \tag{2.22}$$

where $\hat{a}(\cdot,\cdot)$ is the Green's function of the symmetrised random walk kernel is made together with the uniform existence of qth moments for some $q > 2$. But as remarked on p. 195 of Cox et al. (1995) their calculations also give a proof of the theorem in our setting, i.e. assumptions (1.4), (1.5) and uniform existence of second moments. We could also prove Theorem A in their setting. But as the regime of $h = \mathcal{O}(x)$ opens new questions we do not formulate our theorem for this case. E.g. it is not clear if $h^\sharp$ fulfils (2.22) if h does.

7. Note that the branching rate of the limit process only depends on the total mass and not on finer properties. We say that this property (of dependency only on the total mass) is reproduced under the finite system scheme. As a consequence

$$\cdot^\sharp : h \mapsto h^\sharp \text{ given by } h^\sharp(\overline{\theta}) = \frac{1}{\overline{\theta}} \langle\!\langle \nu_{\overline{\theta}}, h(\overline{x}_0)\overline{x}_0 \rangle\!\rangle \tag{2.23}$$

gives the dependence of the diffusion coefficient of the limit process denoted by $h^\sharp$ on the function h. As the diffusion constant only depends on the total mass as we just discussed we can ask for more, i.e. fixed points and fixed shapes of the map $\hat{F}$. By a fixed point we mean some h with $\hat{F}h = h$ and by a fixed shape a function h with $\hat{F}h = ch$ for some constant $c > 0$. As h^* only depends on total masses results for fixed points and fixed shapes carry over from the process of total masses. See Pfaffelhuber (2003) for a detailed discussion.

8. In the proof of Theorem A we will need coupled processes, both for two infinite systems and one infinite and a finite system. This is because by these couplings we can obtain estimates of how distant two processes are after some time. The construction of these is not standard because we are dealing with measure-valued processes. The main estimates we obtain are given in Propositions 3.10 for two infinite systems and Proposition 3.14 for one infinite and one finite system.

3. Techniques: Coupling

Coupling has been proven to be an extremely powerful tool for the proof of various statements concerning the long-time behaviour of stochastic processes. This is in particular true for systems with one-dimensional components. To develop this tool for multidimensional components leads to new challenges and this is focus of current research. Here we introduce a new coupling method for measure-valued processes inspired by Dawson and Greven (2003). This coupling will be a main ingredient in our proofs and all of our comparison estimates in the proof of Theorem A rely on it. We believe that this coupling can be developed further and be adapted to other processes as well. Here we think not only of multitype branching processes with more complex branching rates but also of resampling systems.

Coupling for two processes X^1, X^2 that both have a metric state space (S, d) means the construction of a process $Y = (Y^1, Y^2)$ on S^2 such that $\mathcal{L}(Y^i) = \mathcal{L}(X^i)$ $(i = 1, 2)$. Non-trivial couplings introduce a dependency structure between the components. Then both components make certain correlated (coupled) transitions given e.g. by transition functions or pregenerators.

In order to use the coupled process it is necessary to define a distance between its two components. Then if the coupling is defined properly one hopes to show that the distance between its components will become smaller as time evolves. Differing from standard probability language we call such a coupling *successful* though both processes need not be in the same state for any time t.

The coupling we introduce is a process Y on $((\mathcal{M}_f(I))^\Lambda)^2 = ((\mathcal{M}_f(I))^2)^\Lambda$. At each site ξ the state of the coupled process is $y_\xi = (y_\xi^1, y_\xi^2) \in (\mathcal{M}_f(I))^2$. The space $\mathcal{M}_f(I)$ is a metric space according to the Wasserstein[1] distance which will be defined in (3.2).

Consider atomic measures and the induced particle system. Matching of measures means the matching of particles. That means that each particle in the first component is assigned to a certain particle in the second component. Second, as often as possible, every time a transition (branching or migration) occurs for the particle in the first component the same transition occurs for the particle in the second component that is matched to it. There is a great variety of choices for this matching. As we are dealing with finite measures in both components it is possible that the total masses do not coincide. Therefore we have to introduce a new type of particle $\star \notin I$ and $I^\star := I \cup \{\star\}$. It is important to notice that any matching of particles that are given by finite measures y^1, y^2 can be described by a measure $\chi \in \mathcal{M}_f((I^\star)^2)$ with $\pi_1\chi|_I = y^1, \pi_2\chi|_I = y^2$ where π_1 and π_2 are projection operators. Given such a measure, i.e. a measure on the square $(I^\star)^2$ with $\chi((u,v)) > 0$ for some $u, v \in I$, a particle of type u in the first component is matched to a particle of type v in the second component.

We point out that we do not stick to the case $\Lambda = \mathbb{Z}^d$ in this chapter. Also any other countable Abelian group Λ will do. We define coupled processes for two infinite and for the system of one infinite and one finite process. Both coupled processes will rely on a concept we call matching of measures what will be described in Section 3.1. In Section 3.2 we do that for two infinite systems and in Section 3.3 for one infinite and one finite system.

Remark. The starting point of the whole construction of our coupling was the coupling introduced in Dawson and Greven (2003). But some problems are solved differently here. Let us mention the most important at this place and come to minor ones where they occur during our construction.

In Dawson and Greven (2003) the state space for the coupled process at each site was $\mathcal{M}_f((I^\star)^2)$ whereas we have $(\mathcal{M}_f(I))^2$. This difference comes from the fact that Dawson and Greven start by matching particles what leads to a state in $\mathcal{M}_f((I^\star)^2)$ and then they define their dynamics for these matched particles. In our construction the matching of particles is only implicit in the transition rates of the process. Here the matching of particles is only an auxiliary construction. The main advantage of our approach is that we do not have to deal with extra local time terms in the generator of the diffusion limit. For more on this see page Remark 3.2.

Both constructions start with particle processes and derive the diffusion limit by an approximation procedure. We already encountered in Chapter 1 two equivalent formulations of these diffusion limits (see equations (1.11) and (1.8)). When constructing our coupling we will use an equivalent version of (1.11) whereas Dawson and Greven used the version corresponding to (1.8).

In both couplings a metric on I is needed. Here we take the euclidean metric $\rho(u,v) = |u - v|$ whereas Dawson and Greven take the discrete metric $\rho(u,v) = 1_{u \neq v}$. The latter metric does not induce the canonical (euclidean) topology on $I^\star$ and therefore they have to approximate at the end of their construction finite measures by atomic measures with finitely many atoms. On the other side the

[1] Some authors also write Vasherstein or Vasershtein

discrete metric respects the property of the system that types are exchangeable whereas the euclidean metric imposes geometrical properties on the type space which are not founded on properties of the object to be modelled.

Although our proof of Theorem A could also be achieved with the help of the coupling introduced in Dawson and Greven (2003) we believe that the coupling described here will find a wider application. On the one hand our coupling has advantages in the construction because the generator is simpler than for the coupling of Dawson and Greven. On the other hand it can be considered a better coupling meaning that the distance of two coupled processes in our coupling can be bounded from above by the distance of the coupling of Dawson and Greven.

3.1. *Matching measures.* The reasons why we want to match finite measures are twofold. First for the dynamics of the coupled process we need a matching of finite measures to define the transitions and the rates of transitions. But a coupling not only consists of a coupled dynamics but also of coupled initial distributions. Secondly we want to make this coupling of initial distributions as good as possible. To derive this we have to match random finite measures because initial distributions can be random. As the second point is accomplished by matching non-random measures we begin with these and later come to initial distributions, i.e. random finite measures.

Matching two finite measures. Since we want to compare two finite measures θ_1 and θ_2 we have to deal with the problem of deciding whether two measures on I are close to each other or are far away. There are several ways of defining distance functions for finite measures. (See Rachev (1991) or Gibbs and Su (2001) for the case of probability measures.) We take here the approach of the Wasserstein metric.

The distance of two finite measures in this metric relies on a metric ρ on the space these measures are defined on. Therefore the matching of particles is good in the sense of the Wasserstein metric if we match particles (u,v) with $\sum_{u,v}\rho(u,v)$ as small as possible. Since we want the Wasserstein metric to induce the weak topology in the space of finite measures we have to choose a metric that defines the canonical topology on $I^\star$. We take

$$\rho(u,v)=\begin{cases}|u-v|, & u,v\in I,\\ |2-v|, & u=\star, v\in I,\\ |u-2|, & u\in I, v=\star\\ 0, & u=v=\star.\end{cases} \tag{3.1}$$

In the sequel we make the agreement that if not otherwise stated we calculate with the state $\star$ as if it had the value 2.

Then we define for $\theta_1,\theta_2\in\mathcal{M}_f(I)$ the Wasserstein distance

$$\begin{aligned}\rho_W^I(\theta_1,\theta_2) &:= \rho_W(\theta_1,\theta_2)\\ &:= \inf\Big(\int_{(I^\star)^2}\rho(u,v)\chi(d(u,v)),\ \chi\in\mathcal{M}_f((I^\star)^2), \pi_1\chi|_I=\theta_1, \pi_2\chi|_I=\theta_2\Big).\end{aligned} \tag{3.2}$$

Since the set

$$\{\chi\in\mathcal{M}_f((I^\star)^2):\pi_1\chi|_I=\theta_1,\pi_2\chi|_I=\theta_2,\chi((\star,\star))=0\} \tag{3.3}$$

is compact and the weights in $(\star,\star)$ play no role for the infimum of (3.2) the infimum is attained. Therefore there is a measure $\chi_{\theta_1,\theta_2} \in \mathcal{M}_f((I^\star)^2)$ such that

$$\rho_W(\theta_1,\theta_2) = \int_{(I^\star)^2} \rho(u,v)\chi_{\theta_1,\theta_2}(d(u,v)). \tag{3.4}$$

We will call every $\chi \in \mathcal{M}_f((I^\star)^2)$ for which (3.4) holds, a minimiser of the Wasserstein metric.

There are more ways to describe the Wasserstein metric. We state one of them that does not rely on a minimiser of the right side of (3.2) but deals with distribution functions. Let $\pounds_\theta$ be the distribution function of $\theta \in \mathcal{M}_f(I)$. The next Lemma is an old result obtained in Dall'Aglio (1956), Theorema 1. See also Rachev (1991), equations (5.1.32) and (5.1.33).

Lemma 3.1. *Take $\theta_1,\theta_2 \in \mathcal{M}_f(I)$. Then*

$$\rho_W(\theta_1,\theta_2) = \int_{I^\star} |\pounds_{\theta_1}(u) - \pounds_{\theta_2}(u)|du. \tag{3.5}$$

Here we have set for $f: I^\star \to \mathbb{R}$

$$\int_{I^\star} f(u)du := \int_I f(u)du + f(\star). \tag{3.6}$$

We mentioned that χ_{θ_1,θ_2} is generally not unique. However we will deal with a certain minimiser that has some nice properties. Therefore we make the following definition.

Definition 3.2. *Let $\psi : \mathcal{M}_f(I) \times \mathcal{M}_f(I) \to \mathcal{M}_f((I^\star)^2)$ be defined by*

$$\psi(\theta_1,\theta_2)([0,u]\times[0,v]) := \tilde\theta_1([0,u]) \wedge \tilde\theta_2([0,v]) \qquad (u,v \in I^\star) \tag{3.7}$$

with

$$\begin{aligned} \tilde\theta_1 &:= \theta_1 + (\theta_1(I) - \theta_2(I))^- \delta_\star \in \mathcal{M}_f(I^\star), \\ \tilde\theta_2 &:= \theta_2 + (\theta_1(I) - \theta_2(I))^+ \delta_\star \in \mathcal{M}_f(I^\star). \end{aligned} \tag{3.8}$$

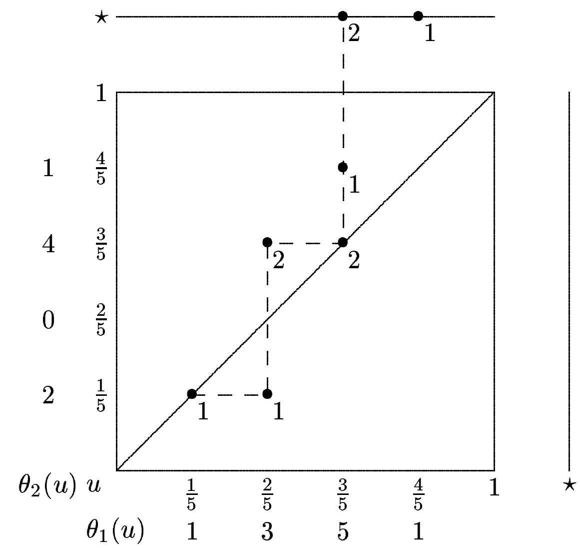
We define $\psi(\theta_1,\theta_2)$ on a generator of the σ-algebra of $(I^\star)^2$ and that is enough to define $\psi(\theta_1,\theta_2)$ uniquely on $\mathcal{B}((I^\star)^2)$.

Before proving that $\psi(\cdot,\cdot)$ has certain properties, e.g. that $\psi(\theta_1,\theta_2)$ is indeed a minimiser for the Wasserstein metric, we give an example. It deals with counting measures, i.e. particle systems.

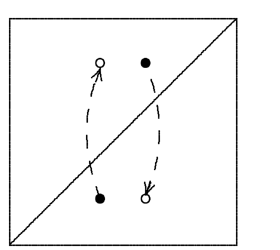
3.1.1. *Example.* We take θ_1,θ_2, defined by

$$\begin{aligned} &\theta_1(\tfrac{1}{5}) = 1, \quad \theta_1(\tfrac{2}{5}) = 3, \quad \theta_1(\tfrac{3}{5}) = 5, \quad \theta_1(\tfrac{4}{5}) = 1, \\ &\theta_2(\tfrac{1}{5}) = 2, \quad \theta_2(\tfrac{2}{5}) = 0, \quad \theta_2(\tfrac{3}{5}) = 4, \quad \theta_2(\tfrac{4}{5}) = 1, \end{aligned} \tag{3.9}$$

Then we can represent the measure $\psi(\theta_1,\theta_2)$ graphically:



In the case of atomic measures there is a certain algorithm how to put mass on $(I^{\star})^2$ such that one obtains $\psi(\theta_1, \theta_2)$. It is simplest to first consider the point $(\frac{1}{5}, \frac{1}{5})$. As the above formula indicates we have $\psi(\theta_1, \theta_2)([0, \frac{1}{5}], [0, \frac{1}{5}]) = \theta_1([0, \frac{1}{5}]) \wedge \theta_2([0, \frac{1}{5}]) = 1$. As $(\frac{1}{5}, \frac{1}{5})$ is the lowest leftmost point in the square that will carry mass for $\psi(\theta_1, \theta_2)$ because of the projections we have $\psi(\theta_1, \theta_2)(\frac{1}{5}, \frac{1}{5}) = 1$. Then we still have to put mass 1 on $I^{\star} \setminus \{\frac{1}{5}\} \times \{\frac{1}{5}\}$ because of the projection on the second component. Therefore we take the next possibility, i.e. $(\frac{2}{5}, \frac{1}{5})$ and put mass 1 to this point. This procedure corresponds to the calculation $\psi(\theta_1, \theta_2)([0, \frac{2}{5}] \times [0, \frac{1}{5}]) = \theta_1([0, \frac{2}{5}]) \wedge \theta_2([0, \frac{1}{5}]) = 2$ and we already have mass 1 on $(\frac{1}{5}, \frac{1}{5})$. Then we again must put the remaining mass 2 on $\{\frac{2}{5}\} \times I^{\star} \setminus \{\frac{1}{5}\}$. The next possibility to do this is $u = \frac{3}{5}$. We put mass 2 there and in the next step mass 2 on $(\frac{3}{5}, \frac{3}{5})$. Then we are left to put mass 3 on $\{\frac{3}{5}\} \times I^{\star} \setminus \{\frac{3}{5}\}$. We chose the next possibility $\frac{4}{5}$, but now we can only put mass 1 to $(\frac{3}{5}, \frac{4}{5})$, because $\theta_2(\frac{4}{5}) = 1$. So we must distribute mass 2 on $\{\frac{3}{5}\} \times I^{\star} \setminus \{\frac{3}{5}, \frac{4}{5}\}$, but we cannot put it on $\frac{3}{5} \times I$ because the second marginal is already correct. Therefore we put the rest on $I \times \{\star\}$ and everything is done. So the marginals $\pi_i \psi(\theta_1, \theta_2)|_I$ agree with θ_i $(i = 1, 2)$.



It is also intuitive (and will be proved in Lemma 3.4) that the given measure $\psi(\theta_1, \theta_2)$ is a minimiser of the Wasserstein metric. To see this notice that for given marginals θ_1, θ_2 and any measure on $(I^{\star})^2$ one can obtain any other measure on

$(I^\star)^2$ with the same marginals by flipping weights as shown in the above diagram. These flips clearly don't change the marginals. For our above example neither single flips nor consecutive flips make the Wasserstein metric smaller, so we gave a minimiser.

With the help of the map ψ we want to define our transition rates. Hence $\psi(\cdot,\cdot)$ must be at least measurable. Next we prove the stronger result that it is even continuous.

Lemma 3.3. *The map ψ is continuous with respect to the topology of weak convergence. Especially it is measurable.*

Proof. Take $(\theta_1, \theta_2) \in (\mathcal{M}_f(I))^2$ and a sequence $(\theta_1^n, \theta_2^n) \Rightarrow (\theta_1, \theta_2)$. We want to show that then also $\psi(\theta_1^n, \theta_2^n) \Rightarrow \psi(\theta_1, \theta_2)$. As the family $\psi(\theta_1^n, \theta_2^n)$ is relatively compact we take any limit point $\tilde{\theta}$. As θ_1 and θ_2 are finite measures they have at most countably many atoms. So there is a discrete set $J \subseteq I$ such that $(u,v) \notin J^2 \Rightarrow \tilde{\theta}(\{u\} \times I) = \tilde{\theta}(I \times \{v\}) = 0$. Therefore the sets

$$[0,u] \times [0,v] \quad (u, v \notin J) \tag{3.10}$$

are $\tilde{\theta}$-continuity sets (i.e. $\tilde{\theta}(\partial[0,u] \times [0,v]) = 0$). By the Portmanteau theorem (Ethier and Kurtz (1986), Theorem 3.3.1) we find that for $u, v \in I \setminus J$ that

$$\begin{aligned}\tilde{\theta}([0,u] \times [0,v]) &= \lim_{n\to\infty} \psi(\theta_1^n, \theta_2^n)([0,u] \times [0,v]) = \lim_{n\to\infty} \theta_1^n([0,u]) \wedge \theta_2^n([0,v]) \\ &= \theta_1[(0,u]) \wedge \theta_2([0,v]) = \psi(\theta_1,\theta_2)([0,u] \times [0,v]).\end{aligned} \tag{3.11}$$

A similar calculation holds for $u = \star$ or $v = \star$. So we have proved that $\tilde{\theta}$ and $\psi(\theta_1, \theta_2)$ coincide on a generating subset of $\mathcal{B}((I^\star)^2)$ Thus, $\tilde{\theta} = \psi(\theta_1, \theta_2)$ and we finished the proof. $\square$

The following lemma shows that our map ψ defines a minimiser and gives some nice descriptions of the Wasserstein metric.

Lemma 3.4. *Let $\theta_1, \theta_2 \in \mathcal{M}_f(I)$.*

1. For any $(u,v) \in supp(\psi(\theta_1, \theta_2))$

$$\psi(\theta_1,\theta_2)([0,u) \times (v,\star]) = \psi(\theta_1,\theta_2)((u,\star] \times [0,v)) = 0. \tag{3.12}$$

2. Let

$$\Delta_{\theta_1,\theta_2} : \begin{cases} I^\star & \to \mathbb{R} \\ u & \mapsto \theta_1([0,u]) - \theta_2([0,u]). \end{cases} \tag{3.13}$$

Then

$$\rho_W(\theta_1,\theta_2) = \int_{I^\star} |\Delta_{\theta_1,\theta_2}(u)| du = \int_{(I^\star)^2} |u' - v'| \psi(\theta_1,\theta_2)(d(u',v')). \tag{3.14}$$

Especially $\psi(\theta_1, \theta_2)$ defines a minimiser of the Wasserstein metric.

Proof. For 1. take $(u,v) \in \text{supp}(\psi(\theta_1,\theta_2))$ and assume $\psi(\theta_1,\theta_2)([0,u) \times (v,\star]) > 0$. Then there exists $(u',v') \in \text{supp}(\psi(\theta_1,\theta_2))$ with $u' < u, v' > v$. Set $\epsilon := \frac{u-u'}{2} \wedge \frac{v'-v}{2}$.

Since $(u'+\epsilon,\star]\times[0,v+\epsilon]$ is a neighbourhood of (u,v) we have

$$\begin{aligned}
0&<\psi(\theta_1,\theta_2)((u'+\epsilon,\star]\times[0,v+\epsilon])\\
&=\psi(\theta_1,\theta_2)([0,\star]\times[0,v+\epsilon])-\psi(\theta_1,\theta_2)([0,u'+\epsilon]\times[0,v+\epsilon])\\
&=\tilde{\theta}_1([0,\star])\wedge\tilde{\theta}_2([0,v+\epsilon])-\tilde{\theta}_1([0,u'+\epsilon])\wedge\tilde{\theta}_2([0,v+\epsilon])\\
&=\tilde{\theta}_2([0,v+\epsilon])-\tilde{\theta}_1([0,u'+\epsilon])\wedge\tilde{\theta}_2([0,v+\epsilon]),
\end{aligned}\tag{3.15}$$

since $\tilde{\theta_2}([0,v+\epsilon])\leq\tilde{\theta}_2([0,\star])=\tilde{\theta}_1([0,\star])$. Therefore, necessarily $\tilde{\theta}_1([0,u'+\epsilon])<\tilde{\theta}_2([0,v+\epsilon])$.

On the other hand $[0,u'+\epsilon]\times(v+\epsilon,\star]$ is a neighbourhood of (u',v') and so

$$\begin{aligned}
0&<\psi(\theta_1,\theta_2)([0,u'+\epsilon]\times(v+\epsilon,\star])\\
&=\psi(\theta_1,\theta_2)([0,u'+\epsilon]\times[0,\star])-\psi(\theta_1,\theta_2)([0,u'+\epsilon]\times[0,v+\epsilon])\\
&=\tilde{\theta}_1([0,u'+\epsilon])\wedge\tilde{\theta}_2([0,\star])-\tilde{\theta}_1([0,u'+\epsilon])\wedge\tilde{\theta}_2([0,v+\epsilon])\\
&=\tilde{\theta}_1([0,u'+\epsilon])-\tilde{\theta}_1([0,u'+\epsilon])\wedge\tilde{\theta}_2([0,v+\epsilon]),
\end{aligned}\tag{3.16}$$

and we arrive at $\tilde{\theta}_2([0,v+\epsilon])<\tilde{\theta}_1([0,u'+\epsilon])$, a contradiction. So $\psi(\theta_1,\theta_2)([0,u)\times(v,\star])=0$. The same reasoning holds to prove $\psi(\theta_1,\theta_2)((u,\star]\times[0,v))=0$.

The first equality of 2. follows from Lemma 3.1. Since by 1. $\psi(\theta_1,\theta_2)([0,u]\times(u,\star])=0$ or $\psi(\theta_1,\theta_2)((u,\star]\times[0,u])=0$ we have for the second equality

$$\begin{aligned}
\rho_W(\theta_1,\theta_2)&=\int_{I^\star}|\theta_1([0,u])-\theta_2([0,u])|du\\
&=\int_{I^\star}|\psi(\theta_1,\theta_2)([0,u]\times[0,\star])-\psi(\theta_1,\theta_2)([0,\star]\times[0,u])|du\\
&=\int_{I^\star}|\psi(\theta_1,\theta_2)([0,u]\times[0,u])+\psi(\theta_1,\theta_2)([0,u]\times(u,\star])\\
&\qquad-\psi(\theta_1,\theta_2)([0,u]\times[0,u])-\psi(\theta_1,\theta_2)((u,\star]\times[0,u])|du\\
&=\int_{I^\star}\psi(\theta_1,\theta_2)([0,u]\times(u,\star])+\psi(\theta_1,\theta_2)((u,\star]\times[0,u])du\\
&=\int_{I^\star}\int_{(I^\star)^2}(1(v'\geq u\geq u')+1(u'\geq u\geq v'))\psi(\theta_1,\theta_2)(d(u',v'))du\\
&=\int_{(I^\star)^2}\big((v'-u')^++(v'-u')^-\big)\psi(\theta_1,\theta_2)(d(u',v'))\\
&=\int_{(I^\star)^2}|v'-u'|\psi(\theta_1,\theta_2)(d(u',v')).
\end{aligned}\tag{3.17}$$

□

The last property, that $\psi(\theta_1,\theta_2)$ is a minimiser of the Wasserstein metric, is nice to have but not necessary for our construction. Nevertheless the matching defined by $\psi(\theta_1,\theta_2)$ makes the distance of two components implicit. We introduce a coupling because we want to measure the distance of two components and we will do this measurement also according to the Wasserstein metric. So the Wasserstein distance appears here in two different contexts, first the matching of particles and second measuring the distances of components. We point out that these two contexts need not rely on the same procedure in general.

Matching two random finite measures. Nevertheless we are not yet done with matching measures. Until now we only matched finite measures. Later on we want to couple two of our processes that already evolved for a while. Therefore we also have to couple their initial distributions (which are distributions of random finite measures). So we also have to deal with matching random finite measures. We already proved in Lemma 3.3 that the map $(\theta_1, \theta_2) \mapsto \psi(\theta_1, \theta_2)$ is continuous. We will need that in the next lemma when we will define a function denoted by $\tilde{\Psi}$ which is a minimiser of the (random) Wasserstein metric for two random finite measures which is defined in (3.18). We do not need this matching of random finite measures for the construction of our coupled process because we could also start the coupled process in the product measure but we shall need it later on for statements about convergence in the proof of Theorem A.

To carry this out we define in correspondence to the case of finite measures the Wasserstein metric on the space of random finite measures denoted by $\rho_W^{\mathcal{M}_f(I)}$ (which can be compared to the Wasserstein metric in (3.2)). We set for $\mu^1, \mu^2 \in \mathcal{P}(\mathcal{M}_f(I))$

$$\rho_W^{\mathcal{M}_f(I)}(\mu^1, \mu^2) := \inf\Big\{ \int_{(\mathcal{M}_f(I))^2} \rho_W(\theta_1, \theta_2)\mu(d(\theta_1, \theta_2)), \mu \in \mathcal{P}((\mathcal{M}_f(I))^2), \pi_i\mu = \mu^i \ (i = 1, 2)\Big\} \tag{3.18}$$

and again call any $\mu \in \mathcal{P}((\mathcal{M}_f(I))^2)$ with marginals μ^1 and μ^2 where this infimum is attained a minimiser of the Wasserstein metric in $\mathcal{P}((\mathcal{M}_f(I))^2)$ for μ^1 and μ^2. Again we construct a certain minimiser $\Psi(\mu^1, \mu^2)$ of the Wasserstein metric, but this time we do not construct it explicitly but according to the next lemma via a measurable selection.

We write for a metric space E (which will be either I or $(I^\star)^2$)

$$\mathcal{L}_1(E) := \Big\{\mu \in \mathcal{P}(\mathcal{M}_f(E)) : \int_E \langle\theta, 1\rangle \mu(d\theta) < \infty\Big\}. \tag{3.19}$$

Lemma 3.5. *There is a measurable function*

$$\begin{aligned} \tilde{\Psi} : \mathcal{L}_1(I) \times \mathcal{L}_1(I) &\to \mathcal{L}_1((I^\star)^2) \\ (\mu^1, \mu^2) &\mapsto \tilde{\Psi}(\mu^1, \mu^2) \end{aligned} \tag{3.20}$$

such that $\tilde{\Psi}(\mu^1, \mu^2)$ is a minimiser of the Wasserstein metric in $\mathcal{L}_1((I^\star)^2)$ for μ^1 and μ^2. Furthermore the Wasserstein metric is finite for any $\mu^1, \mu^2 \in \mathcal{L}_1(I)$.

The proof uses the measurable selection theorem (see Ethier and Kurtz (1986), Appendix, Theorem 10.1). The only difficulty is that one has to take care of possibly signed measures. For all details consult Pfaffelhuber (2003).

We state one extension of this lemma which will be needed in the proof of Theorem A. The map $\tilde{\Psi}$ we have just constructed was only auxiliary for the next corollary so we now come to the construction of the maps Ψ and Ψ^N we shall need in our proofs.

Define

$$\begin{aligned} \mathcal{T}_1(\mathcal{E} \times \mathcal{E}) &:= \{\mu \in \mathcal{P}(\mathcal{E} \times \mathcal{E}) \text{ shift invariant}, \pi_i\mu \in \mathcal{T}_1(\mathcal{E})\}, \\ \mathcal{T}_1(\mathcal{E} \times \mathcal{E}^N) &:= \{\mu \in \mathcal{P}(\mathcal{E} \times \mathcal{E}^N) : \hat{\mu} \circ \Phi_N^{-1} \in \mathcal{T}_1(\mathcal{E} \times \mathcal{E})\} \end{aligned} \tag{3.21}$$

with $\hat{\Phi}_N((Y,Y^N)) := (Y, \Phi_N Y^N)$ and a Wasserstein distance for $\mu^1, \mu^2 \in \mathcal{T}_1(\mathcal{E})$ similar to (3.18) defined by

$$\rho_W^{\mathcal{E}}(\mu^1,\mu^2) := \inf\Big\{ \int_{\mathcal{E}\times\mathcal{E}} \rho_W(y_0^1,y_0^2)\mu(d(y_0^1,y_0^2)), \mu \in \mathcal{T}_1(\mathcal{E}\times\mathcal{E}), \pi_i\mu = \mu^i \ (i=1,2)\Big\} \tag{3.22}$$

with the same terminology of the minimiser of the Wasserstein metric as above. Note that the minimiser of this metric must also be space-shift invariant.

Lemma 3.6. *There are measurable functions*

$$\begin{aligned} \Psi : \mathcal{T}_1(\mathcal{E}) \times \mathcal{T}_1(\mathcal{E}) &\to \mathcal{T}_1(\mathcal{E}\times\mathcal{E}) & \Psi^N : \mathcal{T}_1(\mathcal{E}) \times \mathcal{T}_1(\mathcal{E}^N) &\to \mathcal{T}_1(\mathcal{E}\times\mathcal{E}^N) \\ (\mu^1,\mu^2) &\mapsto \Psi(\mu^1,\mu^2) & (\mu,\mu^N) &\mapsto \Psi^N(\mu,\mu^N) \end{aligned} \tag{3.23}$$

such that $\Psi(\mu^1,\mu^2)$ is a minimiser of the Wasserstein metric in $\mathcal{T}_1(\mathcal{E}\times\mathcal{E})$ for μ^1 and μ^2 and Ψ^N is a minimiser of the Wasserstein metric in $\mathcal{T}_1(\mathcal{E}\times\mathcal{E})$ for μ and $\Phi_N\mu^N$. Furthermore these Wasserstein metrics are finite.

Proof. The proof is similar to the proof of Lemma 3.5 and can be found in Pfaffelhuber (2003). □

By defining this matching of the initial laws of two processes and then defining the coupled dynamics for the coupled process we derive for every time t an upper bound of the Wasserstein distance of the laws of the two processes. This is the way in which we derive results on weak convergence.

3.2. *Coupling of two infinite systems.* When constructing a coupled process one has to do two things. First match the initial states of the coupling (what we have done by now) and second describe the dynamics of the coupled process such that it has the right marginals. If the first process starts in y^1 and the second in y^2 then the coupled process starts in $\psi(y^1,y^2)$ or if the processes start with distribution μ^1 and μ^2 then the coupled process starts in $\Psi(\mu^1,\mu^2)$ from Corollary 3.6. Before we construct our process in the diffusion limit we will construct a particle process.

The coupled particle process. Consider first the branching dynamics. Here only one site ξ is involved. Since the components y_ξ^1 and y_ξ^2 must have the same law as an uncoupled system we have to take care of the frequencies of branching transitions in the components. Therefore we introduce two branching rates, namely

$$h(\overline{y}_\xi^1) \wedge h(\overline{y}_\xi^2) \qquad \text{and} \qquad |h(\overline{y}_\xi^1) - h(\overline{y}_\xi^2)|. \tag{3.24}$$

With the first rate a branching transition occurs in both components, i.e. here our matching is important. This rate is the maximal possible rate at which transitions in both components can occur simultaneously. So one can think of the particles in the first component and assume that a particle u decides to branch. Then this particle looks via the measure $\psi(y_\xi^1,y_\xi^2)$ with which particle of the second component it is matched. Then a birth or death occurs in both components with probability $\frac{1}{2}$. To define our dynamics the transitions according to the second rate are even simpler. Take for example $h(\overline{y}_\xi^1) > h(\overline{y}_\xi^2)$. Then with rate $h(\overline{y}_\xi^1) - h(\overline{y}_\xi^2)$ for any particle in the first component a branching transition occurs but not in the second component. Using these transitions we can define the pregenerator of the coupled process.

The migration mechanism is similar to the branching with the first rate. Here two sites ξ,η are involved. Again if a particle u from the first component moves from ξ to η it looks via $\psi(y^1_\xi, y^2_\xi)$ with which particle of the second component in ξ it is matched and then both migrate to η. So it is easy to write down the generator but when we want to calculate differences of the Wasserstein metric at η we again are led to the problem of finding $\psi(y^1_\eta+\delta_u, y^2_\eta+\delta_v)$ even if $\psi(y^1_\eta, y^2_\eta)((u,v))=0$.

The main task in comparing states after and before a transition is to calculate the effect on the Wasserstein metric of y^1_ξ and y^2_ξ. This is because mainly we want to derive a successful coupling which translates to the question if the expected Wasserstein metric becomes small. This will be our task in the next subsection. In this subsection we only construct our process.

In the following three steps we will define the transition mechanisms for the coupled process, i.e. a process $Y=(y_\xi)_{\xi\in\Lambda}=((y^1_\xi,y^2_\xi))_{\xi\in\Lambda}$. Every time we only deal with one site we write $y=(y^1,y^2)$. We will give for every possible transition the corresponding generator term. Combining all this we have constructed the coupled particle process and then we will come to the diffusion limit.

Step 1: The branching dynamics. To define the branching mechanism we have to take care of the rates. On the one hand we want to have as many transitions as possible that give a branching event in both components and on the other hand the rates in the components might be different so in one component more transitions occur. To overcome this difficulty we introduce the two branching rates given in (3.24). We will define, according to the two cases two generator terms, $G^p_{\mathrm{br},1}$ and $G^p_{\mathrm{br},2}$ and then

$$G^p_{\mathrm{br}} := G^p_{\mathrm{br},1} + G^p_{\mathrm{br},2}. \tag{3.25}$$

Superscript p indicates that we are dealing with the generator of a particle process. With the first rate branching in both components occur. In this case no rematching of particles is necessary after the transition. To define these generator terms we will take functions

$$\tilde F : (\mathcal{M}_f(I))^2 \to \mathbb{R}, \qquad F : ((\mathcal{M}_f(I))^2)^\Lambda \to \mathbb{R} \tag{3.26}$$

where the functions F only depend on finitely many sites. We will simply write $\tilde F$ and F for generic functions of this type.

If we fix $Y=(y_\eta)_{\eta\in\Lambda}$ and set

$$(Y_{y,\xi})_\eta := \begin{cases} y_\eta, & \eta\neq\xi \\ y, & \eta=\xi, \end{cases} \tag{3.27}$$

then every function $F : ((\mathcal{M}_f(I)^2))^\Lambda \to \mathbb{R}$ defines functions $F_{Y,\xi} : (\mathcal{M}_f(I))^2 \to \mathbb{R}$ by $F_{Y,\xi}(y) := F(Y_{y,\xi})$, i.e. $F_{Y,\xi}$ is the function derived from F by letting all components except y_ξ unchanged.

We define a modified point mass

$$\delta^\star_u := 1_{u\neq\star}\delta_u. \tag{3.28}$$

Then we have for one site

$$\begin{aligned} G^p_{\mathrm{br},1}\tilde F(y) = h(\overline{y}^1)\wedge h(\overline{y}^2)\int_{(I^\star)^2} &\big(\tfrac12 \tilde F(y^1+\delta^\star_u, y^2+\delta^\star_v) \\ &+ \tfrac12 \tilde F(y^1-\delta^\star_u, y^2-\delta^\star_v) - \tilde F(y^1,y^2)\big)\psi(y^1,y^2)(d(u,v)) \end{aligned} \tag{3.29}$$

and the generator term for all sites

$$G^p_{\mathrm{br},1}F(Y) = \sum_{\xi\in\Lambda} \tilde{G}^p_{\mathrm{br},1} F_{Y,\xi}(y_\xi). \tag{3.30}$$

One possible difficulty may be the appearance of signed measures. As we have defined $\psi(y^1, y^2)$ for unsigned measures $y^1, y^2 \in \mathcal{M}_f(I)$ we have to ensure that both marginals are also unsigned after every transition. This could at most occur at a death event. But if e.g. $y^1(u) = 0$ then necessarily $\psi(y^1, y^2)(\{u\} \times I^\star) = 0$, so no transition can occur leading to signed measures. The same reasoning holds for y^2.

With the second rate a branching transition occurs only in the component with the bigger rate. We have

$$\begin{aligned} &G^p_{\mathrm{br},2}\tilde{F}(y) \\ &\quad = (h(\overline{y}^1) - h(\overline{y}^2))^+ \int_I \big(\tfrac{1}{2}\tilde{F}(y^1+\delta_u, y^2) + \tfrac{1}{2}\tilde{F}(y^1-\delta_u, y^2) - \tilde{F}(y^1,y^2)\big) y^1(du) \\ &\quad + (h(\overline{y}^1) - h(\overline{y}^2))^- \int_I \big(\tfrac{1}{2}\tilde{F}(y^1, y^2+\delta_v) + \tfrac{1}{2}\tilde{F}(y^1, y^2-\delta_v) - \tilde{F}(y^1,y^2)\big) y^2(dv), \\ &G^p_{\mathrm{br},2}F(Y) = \sum_{\xi\in\Lambda} G^p_{\mathrm{br},2} F_{Y,\xi}(y_\xi). \end{aligned} \tag{3.31}$$

Again we must take care of possibly signed measures. But again a transition leading to $y^1 - \delta_u$ can only occur if $y^1(u) \geq 1$. So we don't obtain signed measures. The same holds for y^2.

Step 2: The migration dynamics. Here always two sites ξ, η are involved. Recall that in the particle process the particles jump according to the reversed kernel, i.e. the jumping rate from ξ to η is given by $a(\eta, \xi)$ (see Section 1.1). To define the generator term for the migration define for $Y = (y_{\xi'})_{\xi'\in\Lambda} \in ((\mathcal{M}_f(I))^2)^\Lambda$ and $\xi, \eta \in \Lambda, u', v' \in I^\star$ with $y_\eta((u', v')) > 0$ the collection $Y_{\xi,\eta,(u',v')}$ via

$$(Y_{\xi,\eta,(u',v')})_{\xi'} := \begin{cases} (y^1_\eta + \delta^\star_{u'}, y^2_\eta + \delta^\star_{v'}), & \xi' = \eta, \\ (y^1_\xi - \delta^\star_{u'}, y^2_\xi - \delta^\star_{v'}), & \xi' = \xi, \\ y_{\xi'}, & \text{else} \end{cases}. \tag{3.32}$$

Then we have the generator term for the migration from ξ to η

$$G^p_{\mathrm{mig},\xi,\eta}F(Y) = \int_{(I^\star)^2} \big(F(Y_{\xi,\eta,(u',v')}) - F(Y)\big)\psi(y^1_\xi, y^2_\xi)(d(u',v')) \tag{3.33}$$

and the complete generator term for the migration

$$G^p_{\mathrm{mig}}F(Y) = \sum_{\xi,\eta\in\Lambda} a(\eta,\xi) G^p_{\mathrm{mig},\xi,\eta}F(Y). \tag{3.34}$$

Again as explained above, signed measures cannot arise.

Step 3: The coupled particle process. Here we put together what we have till now. To define the particle process we consider the generator

$$G^p_{\mathrm{coup}} := G^p_{\mathrm{br}} + G^p_{\mathrm{mig}}. \tag{3.35}$$

Of course we have to prove that this generator in fact defines the coupled process.

Proposition 3.7. *Let $\mu \in \mathcal{P}(\mathcal{E}^p \times \mathcal{E}^p)$ with $\pi_i\mu \in \mathcal{L}_1(\mathcal{E}^p), i = 1, 2$. Then the (μ, G^p_{coup}) local martingale problem has a unique solution. This solution Y^p is a coupling for two of our processes X^p, i.e. $\mathcal{L}(\pi_i Y^p) = \mathcal{L}(X^p)$ given that $\mathcal{L}(X(0)) = \pi_i\mu$ $(i = 1, 2)$.*

Proof. A Markov process with the generator given in (3.35) can be constructed explicitly by independent Poisson processes giving the jumps. Due to the assumption $h(x) = o(x)$ second moments exist. That also means that there cannot be infinitely many jumps in finite time. Especially this Markov process exists. A general result then guarantees uniqueness for the solution of the martingale problem (see Ethier and Kurtz (1986), Theorem 4.4.1). □

The diffusion limit. We want to derive a diffusion limit of our coupled particle process for small masses, accelerated dynamics and an increased number of particles. In this section we will obtain that the martingale problem for the diffusion limit has a solution. We will neither need nor obtain uniqueness.

In order to obtain existence of a solution of the martingale problem all we have to do is to approximate the process in the diffusion limit by particle systems for decreasing masses of the particle systems. This is done by the convergence of pregenerators for the diffusion limit by approximating systems. Moreover we have to establish tightness for the approximating particle processes.

In order to formulate our statements recall the definition of u_ϵ from (1.12). Moreover $\pi_\epsilon x$ with $x \in \mathcal{M}_f(I)$ from (1.13) can be lifted to $y \in (\mathcal{M}_f(I))^2$ and $Y \in ((\mathcal{M}_f(I))^2)^\Lambda$ by $\pi_\epsilon y := (\pi_\epsilon y^1, \pi_\epsilon y^2)$ and $\pi_\epsilon Y := (\pi_\epsilon y_\xi)_{\xi\in\Lambda}$ respectively.

Now we define discrete derivatives for all functions $F \in \mathcal{C}_b(((\mathcal{M}_f(I))^2)^\Lambda)$

$$\begin{aligned}\frac{\partial_\epsilon F(Y)}{\partial_\epsilon y_\xi}(u,v) &:= \frac{F_{\pi_\epsilon Y,\xi}(\pi_\epsilon(y^1_\xi + \epsilon\delta^\star_{u_\epsilon}, y^2_\xi + \epsilon\delta^\star_{v_\epsilon})) - F_{\pi_\epsilon Y,\xi}(\pi_\epsilon(y^1_\xi, y^2_\xi))}{\epsilon},\\ \frac{\partial^2_\epsilon F(Y)}{\partial_\epsilon y^2_\xi}(u,v) &= \frac{1}{\epsilon^2}\Big(F_{\pi_\epsilon Y,\xi}(\pi_\epsilon(y^1_\xi + \epsilon\delta^\star_{u_\epsilon}, y^2_\xi + \epsilon\delta^\star_{v_\epsilon}))\\ &\qquad + F_{\pi_\epsilon Y,\xi}(\pi_\epsilon(y^1_\xi - \epsilon\delta^\star_{u_\epsilon}, y^2_\xi - \epsilon\delta^\star_{v_\epsilon})) - 2F_{\pi_\epsilon Y,\xi}(\pi_\epsilon(y^1_\xi, y^2_\xi))\Big).\end{aligned} \tag{3.36}$$

For all functions in the algebra $\mathcal{A}_{\text{coup}}$ generated by functions of the form

$$F(Y) = \prod_{i=1}^n g_i(\langle y^1_{\xi_i}, f^1_i\rangle, \langle y^2_{\xi_i}, f^2_i\rangle),\ f^j_i \in \mathcal{C}(I)\ (j=1,2), g_i \in \mathcal{C}^2_b(\mathbb{R}^2), (i = 1,\dots,n) \tag{3.37}$$

for some $n \in \mathbb{N}$ we define the operators

$$\begin{aligned}\frac{\partial F(Y)}{\partial y_\xi}(u,v) &:= \lim_{\epsilon\to 0}\frac{\partial_\epsilon F(Y)}{\partial_\epsilon y_\xi}(u,v),\\ \frac{\partial^2 F(Y)}{\partial y^2_\xi}(u,v) &:= \lim_{\epsilon\to 0}\frac{\partial^2_\epsilon F(Y)}{\partial_\epsilon y^2_\xi}(u,v).\end{aligned} \tag{3.38}$$

We obtain by continuity of the derivatives of the g_is in (3.37)

$$\frac{\partial^2 F(Y)}{\partial y^2_\xi}(u,v) = \frac{\partial}{\partial y_\xi}\Big(\frac{\partial F(Y)}{\partial y_\xi}(u,v)\Big)(u,v). \tag{3.39}$$

In order to obtain a coupling of two infinite systems it is enough to define the generator of the coupled process on the separating algebra $\mathcal{A}_{\text{coup}}$. We do not claim to have uniqueness for the solution of the martingale problem for the coupled process in the diffusion limit. That also means that we cannot rely on the coupled process being a Markov process.

The generator terms for the branching are given by

$$\begin{aligned} G^d_{\text{br},1}F(Y) &= \sum_{\xi\in\Lambda} h(\overline{y}^1_\xi) \wedge h(\overline{y}^2_\xi) \int_{(I^\star)^2} \frac{1}{2}\frac{\partial^2 F(Y)}{\partial y^2_\xi}(u,v)\psi(y^1_\xi, y^2_\xi)(d(u,v)), \\ G^d_{\text{br},2}F(Y) &= \sum_{\xi\in\Lambda} (h(\overline{y}^1_\xi) - h(\overline{y}^2_\xi))^+ \int_I \frac{1}{2}\frac{\partial^2 F(Y)}{\partial y^2_\xi}(u,\star) y^1_\xi(du) \\ &\qquad + (h(\overline{y}^1_\xi) - h(\overline{y}^2_\xi))^- \int_I \frac{1}{2}\frac{\partial^2 F(Y)}{\partial y^2_\xi}(\star,v) y^2_\xi(dv). \end{aligned} \tag{3.40}$$

For the migration, we define

$$\begin{aligned} &G^d_{\text{mig}}F(Y) \\ &\quad = \sum_{\xi,\eta\in\Lambda} a(\xi,\eta) \int_{(I^\star)^2} \frac{\partial F(Y)}{\partial y_\xi}(u,v)\big(\psi(y^1_\eta, y^2_\eta)(d(u,v)) - \psi(y^1_\xi, y^2_\xi)(d(u,v))\big) \end{aligned} \tag{3.41}$$

and altogether

$$G^d_{\text{coup}}F(Y) = \sum_{\xi\in\Lambda}(G^d_{\text{br},1,\xi}F(Y) + G^d_{\text{br},2,\xi}F(Y)) + G^d_{\text{mig}}F(Y). \tag{3.42}$$

Remark. Again we must be cautious because of the appearance of signed measures. For the particle processes we obtained that once a type has zero mass no branching transition will occur for that type. But for the diffusion limit this reasoning no longer holds. (See Karatzas and Shreve (1991), p. 293 for an example in the case of stochastic differential equations.)

But the next proposition will help us. Namely it states that any solution Y of the martingale problem corresponding to G^d_{coup} is a coupling of two of our processes. Recall that paths of Y have state space $((\mathcal{M}_f(I))^2)^\Lambda$ where the measures could be signed at first sight.

But as Y is a coupling for two of our processes each of its components Y^i has paths in $D(\mathbb{R}_+, (\mathcal{M}_f(I))^\Lambda)$ and so almost surely no signed measures appears in any of the two components and that is enough for the process Y to stay in the set of positive measures, i.e. the coupled process has paths in $D(\mathbb{R}_+, ((\mathcal{M}_f(I))^2)^\Lambda)$.

We point out that this reasoning cannot be applied to the coupling of Dawson and Greven (2003) because the state space of this coupling is $(\mathcal{M}_f((I^\star))^2)^\Lambda$. Then the positivity of the marginals of each measure $y_\xi \in \mathcal{M}_f((I^\star)^2)$ is not enough to ensure the positivity of the measure y_ξ.

Proposition 3.8. *Let $\mu = \mathcal{P}(\mathcal{E}\times\mathcal{E})$ with $\pi_i\mu \in \mathcal{L}_2(\mathcal{E}), i = 1,2$. Then the (μ, G^d_{coup}) martingale problem has a solution. Any solution Y of this martingale problem is a coupling for two of our processes X, i.e. $\mathcal{L}(\pi_i Y) = \mathcal{L}(X)$ given that $\mathcal{L}(X_0) = \pi_i\mu$ $(i = 1,2)$.*

Remark. If we take functions F that only depend on the first (respectively second) component and plug them into the generator (3.42) we see that the generator is the generator of a coupling for two of our processes. Namely for these functions the generator gives exactly the expression in (1.14). The main task of this whole construction is now to show that the (μ, G^d_{coup}) local martingale problem has a solution. We do this by standard methods, i.e. by approximating the diffusion limit by particle processes with decreasing masses of the particles.

We could weaken the assumptions to $\pi_i\mu \in \mathcal{L}_1(\mathcal{E}).i = 1,2$. In this case our statement would be that the (μ, G^d_{coup}) local martingale problem is well posed. Our proof carries over to this case by truncation by stopping times which are w.l.o.g. exit times of compact sets and observe that we are dealing with continuous processes.

As we will deal with weak convergence throughout we state a lemma about two facts on weak convergence (Lemma 2.10 of Cox et al. (1995)) which we will use several times.

Lemma 3.9. *Let (E,ρ) be complete and separable with one-point compactification $E \cup \{\infty\}$. Let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{P}(E)$ with $\mu_n \Rightarrow \mu$. Then*

- *If there is $F \in \mathcal{C}(E)$, $F \geq 0$ with $\langle \mu_n, F\rangle \to \langle \mu, F\rangle$ then $\langle \mu_n, F'\rangle \to \langle \mu, F'\rangle$ for any $F' \in \mathcal{C}(E)$ with $|F'| \leq F$.*
- *If there is $F \in \mathcal{C}(E)$ such that $\langle \mu_n, F\rangle$ is bounded then $\langle \mu_n, F'\rangle \to \langle \mu, F'\rangle$ for any $F' \in \mathcal{C}(E)$ with $\frac{F'(x)}{F(x)} \to 0$ as $x \to \infty$.*

For our proof we will use calculations similar to the proof of Ethier and Kurtz (1986), Lemma 4.5.1. and Remark 4.5.2. First, we have to define pregenerators which will be called $G^{p,\epsilon}_{\text{coup}}$ which uniquely define processes Y_ϵ such that the family $(Y_\epsilon)_{\epsilon>0}$ is tight. By Ethier and Kurtz (1986), equation 4.3.4 we have to establish that the following calculation is satisfied for any weak limit $Y^{\epsilon_n} \Rightarrow Y$. For F of the form (3.37), $t_1, \ldots < t_n \leq s \leq t$ and $H_i \in \mathcal{C}_b(\mathcal{E})$

$$\begin{aligned}
&\mathbf{E}\Big[\Big(F(Y_t) - F(Y_s) - \int_s^t G^d_{\text{coup}}F(Y_r)dr\Big)\prod_{i=1}^n H_i(Y_{t_i})\Big] \\
&= \lim_{n\to\infty} \mathbf{E}\Big[\Big(F(Y^{\epsilon_n}_t) - F(Y^{\epsilon_n}_s) - \int_s^t G^d_{\text{coup}}F(Y^{\epsilon_n}_r)dr\Big)\prod_{i=1}^n H_i(Y^{\epsilon_n}_{t_i})\Big] \qquad (3.43)\\
&= \lim_{n\to\infty} \int_s^t \mathbf{E}\Big[\Big(G^{p,\epsilon}_{\text{coup}}F(Y^{\epsilon_n}_r) - G^d_{\text{coup}}F(Y^{\epsilon_n}_r)dr\Big)\prod_{i=1}^n H_i(Y^{\epsilon_n}_{t_i})\Big]
\end{aligned}$$

as

$$\mathbf{E}\Big[\Big(F(Y_\epsilon(t)) - F(Y_\epsilon(s)) - \int_s^t G^{p,\epsilon}_{\text{coup}}F(Y_\epsilon(u))du\Big)\prod_{i=1}^n H_i(Y_\epsilon(t_i))\Big] = 0 \qquad (3.44)$$

as Y_ϵ solves the martingale problem for $G^{p,\epsilon}_{\text{coup}}$. The first equality in (3.43) is a consequence of the fact that $Y_{\epsilon_n} \Rightarrow Y$ (use Lemma 3.9). So we must establish

$$\lim_{\epsilon\to 0} \mathbf{E}\Big[\Big|G^{p,\epsilon}_{\text{coup}}F(Y_\epsilon(r)) - G^d_{\text{coup}}F(Y_\epsilon(r))\Big|\Big] = 0. \qquad (3.45)$$

We carry out the proof in three steps. In the first step we define the approximating particle processes. In the second step we establish tightness of $(Y_\epsilon)_{\epsilon>0}$ and in the third step we prove (3.45).

Step 1: Approximating particle systems. We will define pregenerators $G^{p,\epsilon}_{\text{coup}}$ for particle systems with particles of arbitrary mass. But before we look at the pregenerators we have to say for which functions they are defined. First we take

$$\mathcal{N}_{f,\epsilon}(I) := \{y \in \mathcal{M}_f(I) : y = \pi_\epsilon y\}, \tag{3.46}$$

i.e. counting measures which count masses of size ϵ and support $\epsilon\mathbb{N}_0 \cap I$ so that the state space for one site for the coupled process with particles of size ϵ will be $((\mathcal{N}_{f,\epsilon}(I))^2)^\Lambda$.

Then we define the pregenerators $G^{p,\epsilon}_{\text{coup}}$ for functions $F \in \mathcal{C}_b(((\mathcal{N}_{f,\epsilon}(I))^2)^\Lambda)$. We must make the restriction to $\mathcal{N}_{f,\epsilon}(I)$ instead of $\mathcal{M}_f(I)$ to be sure that the states in the approximating particle processes are unsigned measures. If we would not do this a transition $x_\xi \to x_\xi - \epsilon\delta_u$ for $x_\xi(u) > 0$ which can occur in our dynamics can lead to a negative mass at u.

Informally we now have for $\epsilon > 0$ the following Markov process: any particle carries mass ϵ and the state Y of the process no longer counts particles but masses. Therefore at site ξ the measure $\frac{1}{\epsilon}y_\xi$ is the measure that counts particles. Furthermore all transitions are accelerated which means that all transition rates are multiplied by $\frac{1}{\epsilon}$. So we have for $\epsilon > 0$ and $F \in \mathcal{C}_b(((\mathcal{N}_{f,\epsilon}(I))^2)^\Lambda)$

$$\begin{aligned} G^{p,\epsilon}_{\text{br},1}F(Y) &:= \sum_{\xi\in\Lambda} h(\overline{y}^1_\xi) \wedge h(\overline{y}^2_\xi) \int_{(I^\star)^2} \tfrac{1}{2}\frac{\partial^2_\epsilon F(Y)}{\partial_\epsilon y^2_\xi}(u,v)\psi(y^1_\xi, y^2_\xi)(d(u,v)), \\ G^{p,\epsilon}_{\text{br},2}F(Y) &:= \sum_{\xi\in\Lambda} (h(\overline{y}^1_\xi) - h(\overline{y}^2_\xi))^+ \int_I \tfrac{1}{2}\frac{\partial^2_\epsilon F(Y)}{\partial_\epsilon y^2_\xi}(u,\star)y^1_\xi(du) \\ &\qquad + (h(\overline{y}^1_\xi) - h(\overline{y}^2_\xi))^- \int_I \tfrac{1}{2}\frac{\partial^2_\epsilon F(Y)}{\partial_\epsilon y^2_\xi}(\star,v)y^2_\xi(dv), \end{aligned} \tag{3.47}$$

and

$$\begin{aligned} G^{p,\epsilon}_{\text{mig}}F(Y) = \sum_{\xi,\eta\in\Lambda} a(\xi,\eta) \int_{(I^\star)^2} \frac{\partial_\epsilon F(Y_{y_\xi - \epsilon(\delta_u,\delta_v),\xi})}{\partial_\epsilon y_\xi}(u,v)\big(\psi(y^1_\eta, y^2_\eta)(d(u,v)) \\ - \psi(y^1_\xi, y^2_\xi)(d(u,v))\big). \end{aligned} \tag{3.48}$$

Putting all terms together

$$G^{p,\epsilon}_{\text{coup}}F(Y) = (G^{p,\epsilon}_{\text{br},1}F(Y) + G^{p,\epsilon}_{\text{br},2}F(Y)) + G^{p,\epsilon}_{\text{mig}}F(Y). \tag{3.49}$$

With these pregenerators we define uniquely processes Y_ϵ. We will write $Y_\epsilon = (Y^1_\epsilon, Y^2_\epsilon) = (y^\epsilon_\xi)_{\xi\in\Lambda} = (y^{\epsilon,1}_\xi, y^{\epsilon,2}_\xi)_{\xi\in\Lambda}$. These processes are started in $\pi_\epsilon\mu$.

Step 2: Tightness. By now we defined a family of processes $(Y_\epsilon)_{\epsilon>0}$. To show that this family is tight it suffices according to Ethier and Kurtz (1986), Remark 4.5.2. to show that for $\delta, T > 0$ there exists $K_{\delta,T} \subseteq \mathcal{E} \times \mathcal{E}$ such that

$$\inf_{\epsilon>0} \mathbf{P}[Y_\epsilon(t) \in K_{\delta,T} \ \ (0 \le t \le T)] \ge 1 - \delta. \tag{3.50}$$

To prove this observe that $(Y^i_\epsilon)_{\epsilon>0}$ $(i = 1,2)$ are tight families and therefore there exist $K^i_{\delta,T} \subseteq \mathcal{E}$ $(i = 1,2)$ with

$$\inf_{\epsilon>0} \mathbf{P}[Y^i_\epsilon(t) \in K^i_{\delta,T} \ \ (0 \le t \le T)] \ge 1 - \tfrac{\delta}{2}. \tag{3.51}$$

Now take $\epsilon > 0$ and calculate

$$\begin{aligned}\mathbf{P}[Y_\epsilon(t) \in K^1_{\delta,T} \times K^2_{\delta,T} \ \ (0 \le t \le T)] \\ &= \mathbf{P}[Y^1_\epsilon(t) \in K^1_{\delta,T} \ \ (0 \le t \le T)] + \mathbf{P}[Y^2_\epsilon(t) \in K^2_{\delta,T} \ \ (0 \le t \le T)] \\ &\qquad - \mathbf{P}[Y_\epsilon(t) \in K^1_{\delta,T} \times \mathcal{E} \cup \mathcal{E} \times K^2_{\delta,T} \ \ (0 \le t \le T)] \\ &\ge 1 - \tfrac{\delta}{2} + 1 - \tfrac{\delta}{2} - 1 = 1 - \delta.\end{aligned} \tag{3.52}$$

Hence we are done.

3.2.1. *Step 3: Proof of* (3.45). As F only depends on finitely many coordinates it suffices to prove that (3.45) holds for $F(Y) = g(\langle f^1, y^1_0\rangle . \langle f^2, y^2_0\rangle)$. This is proved for the different generator terms separately. We focus here on $G_{\text{br},1}$. To show (3.45) for this term fix $\delta > 0$ and take a compact set K large enough such that

$$\mathbf{E}\Big[\Big|G^{p,\epsilon}_{\text{br},1} F(Y_\epsilon(r)) - G^d_{\text{br},1} F(Y_\epsilon(r))\Big|, y^\epsilon_0(r) \notin K\Big] \le \tfrac{\delta}{2}. \tag{3.53}$$

This K exists by the uniform integrability of $(h(\overline{y}^{\epsilon,1}_0(r))\overline{y}^{\epsilon,1}_0(r) + h(\overline{y}^{\epsilon,2}_0(r))\overline{y}^{\epsilon,2}_0(r))_{\epsilon>0}$ which is true since second moments are uniformly bounded.

Define $M := \sup_{\epsilon>0,t\ge 0} \mathbf{E}[(\overline{y}^{\epsilon,1}_0(t) + \overline{y}^{\epsilon,2}_0(t))^2] < \infty$.

Next observe that $\frac{\partial^2 F}{\partial y_0^2}$ is uniformly continuous on K. That means that there is a $\epsilon' > 0$ such that

$$\sup_{y',y''\in K, \rho_W(y',y'')\le\epsilon'} \Big|\frac{\partial^2 F(y')}{\partial y_0^2}(u,v) - \frac{\partial^2 F(y'')}{\partial y_0^2}(u,v)\Big| \le \tfrac{\delta}{2M}. \tag{3.54}$$

A Taylor series expansion of the function F reveals that

$$\frac{\partial^2_\epsilon F(y)}{\partial_\epsilon y_0^2}(u,v) = \frac{\partial^2 F(y_\epsilon)}{\partial y_0^2}(u,v) \tag{3.55}$$

for some y_ϵ with $\rho_W(y,y_\epsilon) \xrightarrow{\epsilon\to 0} C\epsilon$ where C only depends on F. Finally

$$\begin{aligned}\lim_{\epsilon\to 0}\mathbf{E}\Big[\Big|G^{p,\epsilon}_{\text{br},1} F(Y_\epsilon(r)) - G^d_{\text{br},1} F(Y_\epsilon(r))\Big|, y^\epsilon_0(r) \in K\Big] \\ = \lim_{\epsilon\to 0}\mathbf{E}\Big[\Big|h(\overline{y}^{\epsilon,1}_0(r)) \wedge h(\overline{y}^{\epsilon,2}_0(r)) \int_{(I^*)^2} \frac{\partial^2_\epsilon F(Y_\epsilon(r))}{\partial_\epsilon y_0^2}(u,v) \\ - \frac{\partial^2 F(Y_\epsilon(r))}{\partial y_0^2}(u,v)\psi(y^{\epsilon,1}_0(r), y^{\epsilon,2}_0(r))(d(u,v))\Big|, y^\epsilon_0(r) \in K\Big] \\ \le \sup_{y',y''\in K,\rho_W(y',y'')\le\epsilon'} \Big|\frac{\partial^2 F(y')}{\partial y_0^2} - \frac{\partial^2 F(y'')}{\partial y_0^2}\Big| \mathbf{E}[(\overline{y}^{\epsilon,1}_0(r) + \overline{y}^{\epsilon,2}_0(r))^2] \le \tfrac{\delta}{2}.\end{aligned} \tag{3.56}$$

Combining this with (3.53) gives (3.45) for $G_{\text{br},1}$. For the other terms the proof works analogously.

Functionals of the coupled process. By now we know that a solution of the martingale problem for our coupled process in the diffusion limit exists. We take any weak limit point $Y^{\epsilon_n} \Rightarrow Y$ which is automatically a solution of the martingale problem for G^d_{coup} in the sequel and will approximate it by particle systems. Therefore in this subsection we write superscript d for the diffusion limit and superscripts p, ϵ

for the particle model with particles having mass ϵ. Such particle processes will also be called ϵ-particle processes.

So in this section we state which functionals of the coupled process are of special interest for us. As we want our coupling to be successful we want to measure the distance between the two coupled processes and do that according to the Wasserstein metric. There is one other functional we examine, the change in total masses. This was already studied in Cox et al. (1995).

Recall the definition of Δ from (3.13). As we are mainly interested in the Wasserstein metric of the two components we are concerned with the functions

$$\begin{aligned}
&(1)\ \ k(y) := \rho_W(y^1, y^2) = \int_{(I^\star)^2} \rho(u,v)\psi(y^1,y^2)(d(u,v)) = \int_{I^\star} |\Delta_{y^1,y^2}(u)|du,\\
&(2)\ \ \bar{k}(y) := \int_{I^\star\setminus I} |\Delta_{y^1,y^2}(u)|du = |\overline{y}^1 - \overline{y}^2|.
\end{aligned}$$

The reason for the last equality is that for any point $u \in I^\star \setminus I$ the function $|\Delta_{y^1,y^2}(u)|$ is exactly the mass on $I \times \{\star\} \cup \{\star\} \times I$ and that is the difference in the total mass of the two components.

For notational convenience we will from now on write

$$\Delta_\xi(u) := \Delta_{y^1_\xi, y^2_\xi}(u), \qquad \Delta_\xi(u)(t) := \Delta_{y^1_\xi(t), y^2_\xi(t)}(u). \tag{3.57}$$

We want to examine the functions

$$K(t) := \mathbf{E}[k(y_\xi(t))] = \mathbf{E}[(k\circ\pi_\xi)(Y(t))] = \mathbf{E}\Big[\int_{I^\star} |\Delta_\xi(u)(t)|du\Big] \tag{3.58}$$

and $\bar{K}(t)$ respectively for our coupled process $Y = (y_\xi)_{\xi\in\Lambda}$ where π_ξ is the projection onto the ξ-component of Λ. These functions can be studied in both the ϵ-particle processes, then denoted by $K^{p,\epsilon}$ and $\overline{K}^{p,\epsilon}$ and the diffusion limit, denoted by K^d and $\bar{K}^d$. These functions will not depend on ξ because of space-shift invariance. Therefore we take $\xi = 0$ in our considerations. To examine these functions we will derive a differential equation obtained from a generator calculation. We have

$$\begin{aligned}
&(1)\ \ \frac{d}{dt}K^{p,\epsilon}(t) = \mathbf{E}[G^{p,\epsilon}_{\text{coup}}(k\circ\pi_0)(Y(t))], \qquad \frac{d}{dt}K^d(t) = \mathbf{E}[G^d_{\text{coup}}(k\circ\pi_0)(Y(t))],\\
&(2)\ \ \frac{d}{dt}\bar{K}^{p,\epsilon}(t) = \mathbf{E}[G^{p,\epsilon}_{\text{coup}}(\bar{k}\circ\pi_0)(Y(t))], \qquad \frac{d}{dt}\bar{K}^d(t) = \mathbf{E}[G^d_{\text{coup}}(\bar{k}\circ\pi_0)(Y(t))].
\end{aligned}$$

It is not clear that the functions $k\circ\pi_0$ and $\bar{k}\circ\pi_0$ are in the domain of the generator for the diffusion limit. But they are in the domain of the generator for the coupled approximating particle process for any particle size. Therefore an approximation argument yields we can approximate the pregenerator acting on these functions for the diffusion limit via the action of the generators for the particle process. Therefore we apply every generator term of the ϵ-particle process to the functions $k\circ\pi_0$ and $\bar{k}\circ\pi_0$ and let $\epsilon \to 0$.

We derive the following

Proposition 3.10. *1. Let $\mu \in \mathcal{T}_2(\mathcal{E}\times\mathcal{E})$. Then for the mean effect of the dynamics on the Wasserstein metric one obtains*[2]

$$\frac{d}{dt}K^d(t) = -2\sum_{\xi\in\Lambda} a(0,\xi)\int_{I^*}\mathbf{E}[1(sgn(\Delta_0(u)(t)) \neq sgn(\Delta_\xi(u)(t))) \\ 1(\Delta_0(u)(t)\neq 0)|\Delta_\xi(u)(t)|]du \\ + \int_{I^*}\mathbf{E}[|h(\overline{y}_0^1(t)) - h(\overline{y}_0^2(t))|y_0^1(t)([0,u])\,1(\Delta_0(u)(t)=0)]du, \tag{3.59}$$

2. For $\bar{K}^d$ one has

$$\frac{d}{dt}\bar{K}^d(t) = -2\sum_{\xi\in\Lambda} a(0,\xi)\mathbf{E}[1(sgn(\bar{y}_0^1(t)-\bar{y}_0^2(t)) \neq sgn(\bar{y}_\xi^1(t)-\bar{y}_\xi^2(t))) \\ 1(\bar{y}_0^1(t)\neq \bar{y}_0^2(t))|\bar{y}_\xi^1(t)-\bar{y}_\xi^2(t)|]. \tag{3.60}$$

3.2.2. *Remark.* One might wish to prove this Proposition by a Tanaka like formula. But we point out that we do not have uniqueness of the martingale for the coupled process and therefore do not know if the process Y is a Markov process. Therefore an approach via the Tanaka formula is not feasible and we must prove the Proposition by direct approximation by particle processes.

Proof. We will prove that all statements hold for ϵ-particle processes. Then the proposition is immediate by the approximation

$$K^{p,\epsilon}(\cdot) \xrightarrow{\epsilon\downarrow 0} K^d(\cdot), \quad \bar{K}^{p,\epsilon}(\cdot) \xrightarrow{\epsilon\downarrow 0} \bar{K}^d(\cdot) \tag{3.61}$$

which is true since we only consider weak limit points of $(Y_\epsilon)_{\epsilon>0}$. The fact that k is an unbounded function does not play a role since we assumed that second moments exist (see Lemma 3.9).

The proof of 1. is the crucial thing of this proposition. 2. will follow quickly having understood the dynamics of 1. We distinguish the effects of the branching and migration dynamics. The proof will take about eight pages and is therefore given in several subsections and lemmata.

Let now $\epsilon > 0$ be fixed and write

$$\frac{d}{dt}K^{p,\epsilon}(t) =: \dot{K}^{p,\epsilon}(t) = \dot{K}^{p,\epsilon}_{\text{br}}(t) + \dot{K}^{p,\epsilon}_{\text{mig}}(t), \qquad \dot{K}^{p,\epsilon}_{\text{br}}(t) = \dot{K}^{p,\epsilon}_{\text{br},1}(t) + \dot{K}^{p,\epsilon}_{\text{br},2}(t) \tag{3.62}$$

with

$$\dot{K}^{p,\epsilon}_{\text{br},1}(t) := \mathbf{E}[G^{p,\epsilon}_{\text{br},1}(k\circ\pi_0)(Y(t))], \qquad \dot{K}^{p,\epsilon}_{\text{br},2}(t) := \mathbf{E}[G^{p,\epsilon}_{\text{br},2}(k\circ\pi_0)(Y(t))], \\ \dot{K}^{p,\epsilon}_{\text{mig}}(t) := \mathbf{E}[G^{p,\epsilon}_{\text{mig}}(k\circ\pi_0)(Y(t))] \tag{3.63}$$

and analogously for $\overline{K}^{p,\epsilon}(t)$. We treat the branching and migration terms separately.

[2]We take $\text{sgn}(x) = \begin{cases} 1, & x>0 \\ 0, & x=0 \\ -1, & x<0. \end{cases}$

Effects of the branching mechanism.

1. As we want to compare the two components we have to take care of changes in the Wasserstein metric and therefore have to see how the minimiser of the Wasserstein metric $\psi(y_0^1, y_0^2)$ is affected by these transitions. Here the first rate corresponding to $G^{p,\epsilon}_{\text{br},1}$ causes no problem because $\psi(y_0^1 \pm \epsilon\delta_u, y_0^2 \pm \epsilon\delta_v) = \psi(y_0^1, y_0^2) \pm \epsilon\delta_{(u,v)}$ if $\psi(y_0^1, y_0^2)((u,v)) > 0$. So the absolute change of the Wasserstein distance is $|v-u|$ in any case with either a plus or a minus sign and so does not play any role in mean. Therefore we have

$$\dot{K}^{p,\epsilon}_{\text{br},1}(t) = 0. \tag{3.64}$$

The second rate corresponding to $G^{p,\epsilon}_{\text{br},2}$ is more complex because the effect on $\psi(y^1, y^2)$ is non-trivial if a birth or a death in only one component occurs. To analyse this effect we have to examine $\psi(y^1, y^2)$ and changes therein. We sum our results in the following lemma.

Lemma 3.11.

$$\dot{K}^{p,\epsilon}_{br,2}(t) = \mathbf{E}\Big[|h(\overline{y}_0^1(t)) - h(\overline{y}_0^2(t))| \int_{I^*} 1(y_0^1(t)([0,u]) = y_0^2(t)([0,u]))y_0^1(t)([0,u])du\Big] \tag{3.65}$$

$$= \int_{I^*} \mathbf{E}[|h(\overline{y}_0^1(t)) - h(\overline{y}_0^2(t))|y_0^1(t)([0,u]), \Delta_0(u)(t) = 0]du. \tag{3.66}$$

Proof. We will omit the subscript 0 and the dependence of t in the proof. The second equality is clear since we have

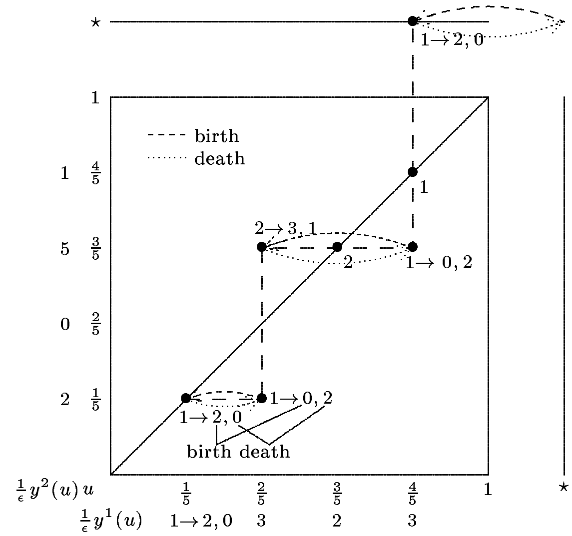
$$y^1([0,u]) = y^2([0,u]) \qquad \Longleftrightarrow \qquad \Delta(u) = 0 \tag{3.67}$$

and the integral and the expectation commute.

To see the first equality we have to consider two cases. The reason why we distinguish the two cases is, that in the first case the effects of a birth and a death on the Wasserstein metric annihilate each other in expectation but not in the second case.

Case A: $\forall u, v \in I: \; y^1([0,u]) \neq y^2([0,v])$

We already mentioned an example of this case in step 1 (see Example 3.1.1). First recall that in this case spoken in the graphical representation the support of $\psi(y^1, y^2)$ is part of a union of horizontal and vertical lines that are connected and in each intersection point $\psi(y^1, y^2)$ has positive weight. Look at the next figure to see what happens in case of a birth and a death respectively in the particle model. In the square we have the measure $\frac{1}{\epsilon}\psi(y^1, y^2)$.



In this figure a branching transition takes place in the first component at $u = \frac{1}{5}$. This is the case if $h(\overline{y}^1) > h(\overline{y}^2)$ and a transition with rate $(h(\overline{y}^1) - h(\overline{y}^2))^+$ occurs. We have to analyse the effect of a transition on the minimiser $\frac{1}{\epsilon}\psi(y^1, y^2)$ because we want to know how the distance of y^1, y^2 changes. In case of a birth look at the dashed arrows and in the case of a death at the dotted ones. Here we have the picture in mind that at $(\star, \star)$ an unlimited reservoir of particles is present. It is clear that horizontal jumps, i.e. rematches of particles do not influence y^2 and if some particle jumps to (u, v) and another one jumps away from (u, v') the net change of y^1 at u is 0. Therefore only the weight of y^1 at $\frac{1}{5}$ changes in the figure in case of a birth and a death respectively. Another important point to notice is that we gave the transitions that not only produce the right changes in the components but also immediately give a new minimiser of the Wasserstein metric. In contrast to the second case no matches that didn't occur before the transition are produced.

The case of $\overline{y}^1 > \overline{y}^2$ but $h(\overline{y}^1) < h(\overline{y}^2)$, i.e. transitions in the second component occur is completely the same except that we have vertical arrows and no particle of the unlimited reservoir is involved. Other possible cases are $\overline{y}^1 < \overline{y}^2$, $h(\overline{y}^1) < h(\overline{y}^2)$ and $\overline{y}^1 < \overline{y}^2$, $h(\overline{y}^1) > h(\overline{y}^2)$. For these cases only the roles of the first and second components are interchanged.

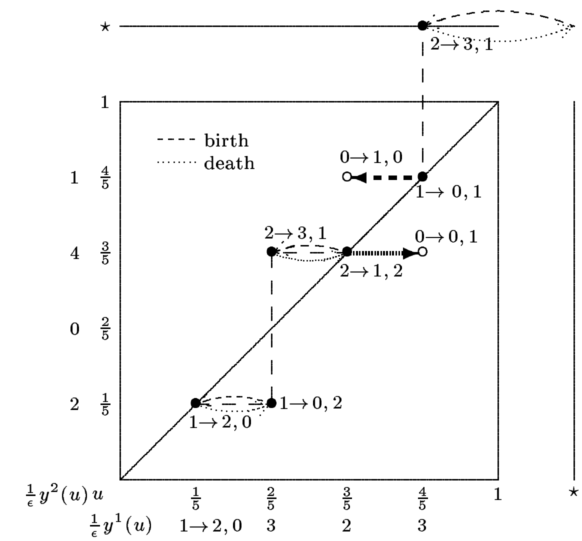
As is seen from the above figure and from the above explanations we obtain in this case

$$\dot{K}^{p,\epsilon}_{\text{br},2,A}(t) := \mathbf{E}[G^{p,\epsilon}_{\text{br},2}(k \circ \pi_0)(Y(t)), \forall u, v \in I : \; y^1_0(t)([0, u]) \neq y^2_0(t)([0, v])] = 0, \tag{3.68}$$

because all arrows for an occurring birth and death annihilate each other in expectation.

Case B: $\exists u, v \in I : \; y^1([0, u]) = y^2([0, v])$

Here some unpleasant things happen. The next figure shows an example again for an ϵ-particle model. Look at the transitions that occur at $\frac{3}{5}$ and $\frac{4}{5}$.



Again we took $\overline{y}^1 > \overline{y}^2$, $h(\overline{y}^1) > h(\overline{y}^2)$ and a transition with rate $(h(\overline{y}^1) - h(\overline{y}^2))^+$ happens at $u = \frac{1}{5}$. We have $y^1([0, \frac{3}{5}]) = y^2([0, \frac{3}{5}])$ before the transition. As we have seen already in the first case only horizontal jumps occur. Nevertheless in this case we must take care of the matched particles (u, v) where $y^1([0, u]) = y^2([0, v])$. Here as can be seen from the picture not all arrows annihilate each other. (There are even particles that are matched to a new type in the other component, marked with a '∘' in the figure.) In spite of these effects almost all arrows annihilate each other except for the thicker ones showing to the '∘'s. No matter if a birth or a death occurs the jumps according to these arrows lead to an increase of the Wasserstein metric.

To make these observation rigorous we remark first that whenever there exist types u, v with $y^1([0, u]) = y^2([0, v])$ and a transition in y^1 occurs for some $u' > u$ or in y^2 for some $v' > v$ we are in the picture of case A and no unpleasant things happen. So we only have to consider branching for $u' \le u$ or $v' \le v$. Without loss of generality we may assume that $y^1(u) > 0$ and $y^2(v) > 0$ by making u and v eventually smaller. Let

$$u_i^+ := \inf\{u' > u : u' \in \operatorname{supp}(y^i)\} \qquad (i = 1, 2). \tag{3.69}$$

It is then clear from the last figure that $\psi(y^1, y^2)((u, v_2^+)) = \psi(y^1, y^2)((u_1^+, v)) = 0$. Now we distinguish five cases

α: $u = v, u_1^+ = v_2^+$,
β: $u = v, u_1^+ \neq v_2^+$,
γ: $u \neq v, u_1^+ = v_2^+$,
δ: $u > v, u_1^+ > v_2^+$ or $u < v, u_1^+ < v_2^+$,
ϵ: $u > v, u_1^+ < v_2^+$ or $u < v, u_1^+ > v_2^+$

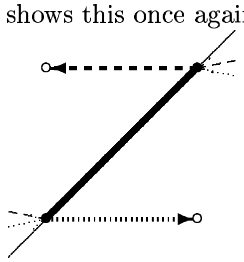
and we will see that in every case only such u' (that do not necessarily lie in the support of y^1 or y^2) for which $y^1([0, u']) = y^2([0, u'])$ play a role. In other words in either of these five cases (3.65) holds. We will treat all cases graphically by cutting out the interesting part of the whole figure. In each figure the thick part of the diagonal corresponds to the increases that occur in both cases birth and death.

α: $u = v, u_1^+ = v_2^+$

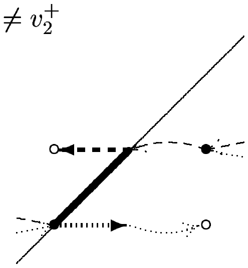
This is the case of our last figure. In this case the expected increase of the Wasserstein metric is $\epsilon(u_1^+ - u) = \epsilon(v_2^+ - v)$, which is the same as

$$\epsilon \int_{I^*} 1(y^1([0,u]) = y^2([0,u]))du. \tag{3.70}$$

The figure shows this once again:



β: $u = v, u_1^+ \neq v_2^+$



We take $u_1^+ > v_2^+$ but $u_1^+ < v_2^+$ is similar. The particle (u_1^+, v_2^+) in case of a birth makes two transitions in our picture. First it jumps on the diagonal to (v_2^+, v_2^+) (which produces a decrease of the Wasserstein metric) and then jumps to (u, v_2^+) what produces an increase. In case of a death the (u, v) particle it first jumps to (v_2^+, v) and then to (u_1^+, v). Altogether the jump of (u_1^+, v_2^+) to (v_2^+, v_2^+) in case of a birth and of (v_2^+, v) to (u_1^+, v) annihilate each other and the other arrows sum up to expression (3.70)

For the next three cases the reasoning is the same as in β, so we leave out the details. All cases give the same expression (3.70).

The rate for these transitions in any of the five cases is

$$\frac{1}{2}\frac{(h(\overline{y}^1) - h(\overline{y}^2))^+}{\epsilon} \tag{3.71}$$

for birth and death as we deal with ϵ-particle processes. Transitions in the second component can be treated analogously but then have rate

$$\frac{1}{2}\frac{(h(\overline{y}^1) - h(\overline{y}^2))^-}{\epsilon}. \tag{3.72}$$

Altogether that leads to

$$\begin{aligned}\dot{K}^{p,\epsilon}_{\text{br},2,B}(t) &:= \mathbf{E}[G^{p,\epsilon}_{\text{br},2}(k \circ \pi_0)(Y(t)), \exists u, v \in I:\ y_0^1(t)([0,u]) = y_0^2(t)([0,v])] \\ &= \mathbf{E}\Big[|h(\overline{y}^1) - h(\overline{y}^2)| \int_{I^*} 1(\Delta(u) = 0) y^1([0,u]) du\Big].\end{aligned} \tag{3.73}$$

As

$$\dot{K}^{p,\epsilon}_{\text{br},2} = \dot{K}^{p,\epsilon}_{\text{br},2,A} + \dot{K}^{p,\epsilon}_{\text{br},2,B} \tag{3.74}$$

the lemma follows. □

2. For $\bar{k}$ the situation is the same as if we would only have one type. So

$$\bar{k}(y^1, y^2) = k(\overline{y}^1\delta_1, \overline{y}^2\delta_1). \tag{3.75}$$

Observe that $\dot{\bar{K}}^{p,\epsilon}_{\text{br},1}(t) = \dot{\bar{K}}^{p,\epsilon}_{\text{br},2,A}(t) = 0$ as in the multitype case and

$$\begin{aligned}\dot{\bar{K}}^{p,\epsilon}_{\text{br},2,B}(t) &= \mathbf{E}\Big[|h(\bar{y}^1) - h(\bar{y}^2)| \int_{I^\star} 1(\Delta(u) = 0)(\overline{y}^1\delta_1)([0,u])du\Big] \\ &= \mathbf{E}[|h(\bar{y}^1) - h(\bar{y}^2)|1(\bar{y}^1 = \bar{y}^2)\bar{y}^1] = 0,\end{aligned} \tag{3.76}$$

because either $|h(\bar{y}^1_0(t)) - h(\bar{y}^2_0(t))| > 0$ or $1(\bar{y}^1_0(t) = \bar{y}^2_0(t)) > 0$, but not both at the same time. So we arrive at

$$\dot{\bar{K}}^{p,\epsilon}_{\text{br}}(t) = \dot{\bar{K}}^{p,\epsilon}_{\text{br},1}(t) + \dot{\bar{K}}^{p,\epsilon}_{\text{br},2,A}(t) + \dot{\bar{K}}^{p,\epsilon}_{\text{br},2,B}(t) = 0. \tag{3.77}$$

Effects of the migration dynamics.

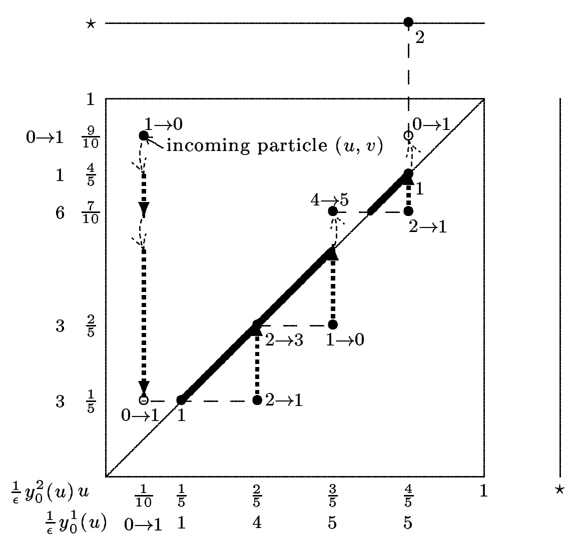
1. Again we consider the mean effect of the migration on the Wasserstein metric. Clearly

$$\mathbf{E}[G^{p,\epsilon}_{\text{mig},\xi,\eta}(k \circ \pi_0)(Y(t))] = 0 \tag{3.78}$$

if $\xi \neq 0$ and $\eta \neq 0$. For $\xi = 0$ we have

$$\begin{aligned}\mathbf{E}[G^{p,\epsilon}_{\text{mig},0,\eta}k \circ \pi_0(Y(t))] &= -\mathbf{E}\Big[\int_{(I^\star)^2} |u - v|\psi(y^1_0(t), y^2_0(t))(d(u,v))\Big] \\ &= -\mathbf{E}\Big[\int_{I^\star} |\Delta_0(u)(t)|du\Big]\end{aligned} \tag{3.79}$$

due to particles jumping away from 0. For $\eta = 0$ the effects are more complex. Again we omit the dependence on t and look at $\psi(y^1_0, y^2_0)$ and changes therein due to rematching. However the next figure illustrates what we have in mind.



As a particle moves from ξ to 0 this incoming particle causes rematches in $\psi(y_0^1, y_0^2)$ in a way that the new $\psi(y_0^1, y_0^2)$ is again a minimiser of the Wasserstein metric. We took the case $u < v$ in the figure. Several jumps occur. First the particle (u,v) jumps to $(u, \min\{v' \in A_{(u,v)} : y_0^2(v') > 0\})$ where $A_{(u,v)} = [u,\star] \times [0,v]$.

With this first jump the second marginal has changed so further jumps must occur to compensate this and with this first jump the matched particle (u,v) has come closer to the diagonal leading to a decrease of the Wasserstein metric. Next successively particles jump upwards to compensate the defect in the second marginal. The crucial point here is that if a particle (u',v') with $u' > v'$ jumps in the direction to the diagonal the Wasserstein metric decreases and if the particle jumps in the direction off the diagonal the Wasserstein metric increases. In this way part of the decrease that was caused by the first jump of (u,v) is annihilated and another part is doubled. The net effect of these transitions is indicated by the thicker lines in the diagram. The above reasoning yields

$$\begin{aligned} &\mathbf{E}[G^{p,\epsilon}_{\mathrm{mig},\xi,0} k \circ \pi_0(Y)] = \\ &\quad \mathbf{E}\Big[\int_{I^\star} |\Delta_\xi(u)| du - 2 \int_{(I^\star)^2} \int_{u' \wedge v'}^{u' \vee v'} \Big(1(v' \geq u') 1(y_0^2([0,u]) > y_0^1([0,u])) \\ &\qquad + 1(v' \leq u') 1(y_0^1([0,u]) > y_0^2([0,u]))\Big) du \;\; \psi(y_\xi^1, y_\xi^2)(d(u',v'))\Big], \end{aligned} \tag{3.80}$$

independently of ϵ. This can be simplified significantly what we do in the next lemma.

Lemma 3.12. *For the effect of migration,*[3]

$$\begin{aligned} \dot{K}^{p,\epsilon}_{mig}(t) = \sum_{\xi \in \Lambda} a(0,\xi) \int_{I^\star} \mathbf{E}[&sgn(\Delta_0(t)(u)) \Delta_\xi(t)(u) + 1(\Delta_0(t)(u) = 0) |\Delta_\xi(t)(u)| \\ &- |\Delta_0(t)(u)|] du. \end{aligned} \tag{3.81}$$

and in case of space-shift invariance

$$\begin{aligned} \dot{K}^{p,\epsilon}_{mig}(t) = -2 \sum_{\xi \in \Lambda} a(0,\xi) \int_{I^\star} \mathbf{E}[&1(sgn(\Delta_0(t)(u)) \neq sgn(\Delta_\xi(t)(u))) \\ &1(\Delta_0(t)(u) \neq 0) |\Delta_\xi(t)(u)|] du. \end{aligned} \tag{3.82}$$

Proof. To see (3.81) all we have to do is to simplify (3.80). Moreover we use Lemma 3.4. First note that we have (omitting dependencies on t)

$$\begin{aligned} \dot{K}^{p,\epsilon}_{\mathrm{mig}} &= \sum_{\xi,\eta \in \Lambda} a(\eta,\xi) \mathbf{E}[G^{p,\epsilon}_{\mathrm{mig},\xi,\eta} k \circ \pi_0(Y)] \\ &= \sum_{\xi \in \Lambda} a(0,\xi) \mathbf{E}[G^{p,\epsilon}_{\mathrm{mig},\xi,0} k \circ \pi_0(Y)] - \mathbf{E}\Big[\int_{I^\star} |\Delta_0(u)| du\Big]. \end{aligned} \tag{3.83}$$

[3]We would obtain the same result if we take $\mathrm{sgn}(x) := \begin{cases} 1, & x \geq 0 \\ -1, & x < 0. \end{cases}$

So we are left with showing

$$\mathbf{E}[G^{p,\epsilon}_{\mathrm{mig},\xi,0}k\circ\pi_0(Y)] = \mathbf{E}\Big[\int_{I^*}(\mathrm{sgn}(\Delta_0(u)))\Delta_\xi(u) + 1(\Delta_0(u)=0)|\Delta_\xi(u)|du\Big]. \tag{3.84}$$

Write

$$\begin{aligned}
\mathbf{E}[G^{p,\epsilon}_{\mathrm{mig},\xi,0}k\circ\pi_0(Y)] &= \mathbf{E}\Big[\int_{I^*}\Big((y^1_\xi([0,u]) - y^2_\xi([0,u]))^+ + (y^1_\xi([0,u]) - y^2_\xi([0,u]))^- \\
&- 2\cdot 1(y^2_0([0,u]) > y^1_0([0,u]))\underbrace{\int_{(I^*)^2} 1(u'\le u < v')\psi(y^1_\xi, y^2_\xi)(d(u',v'))}_{=(y^1_\xi([0,u])-y^2_\xi([0,u]))^+} \\
&- 2\cdot 1(y^1_0([0,u]) > y^2_0([0,u]))\underbrace{\int_{(I^*)^2} 1(v'\le u < u')\psi(y^1_\xi, y^2_\xi)(d(u',v'))}_{=(y^1_\xi([0,u])-y^2_\xi([0,u]))^-}\Big)du\Big] \\
&= \mathbf{E}\Big[\int_{I^*}\Big(1 - 2\cdot 1(y^2_0([0,u]) > y^1_0([0,u]))(y^1_\xi([0,u]) - y^2_\xi([0,u]))^+ \\
&\qquad + 1 - 2\cdot 1(y^1_0([0,u]) > y^2_0([0,u]))(y^1_\xi([0,u]) - y^2_\xi([0,u]))^-\Big)du\Big] \\
&= \mathbf{E}\Big[\int_{I^*}\mathrm{sgn}(y^1_0([0,u]) - y^2_0([0,u]))(y^1_\xi([0,u]) - y^2_\xi([0,u]))^+ \\
&\quad - \mathrm{sgn}(y^1_0([0,u]) - y^2_0([0,u]))(y^1_\xi([0,u]) - y^2_\xi([0,u]))^- \\
&\quad + 1(y^1_0([0,u]) = y^2_0([0,u]))|y^1_\xi([0,u]) - y^2_\xi([0,u])|du\Big] \\
&= \mathbf{E}\Big[\int_{I^*}\mathrm{sgn}(\Delta_0(u))\Delta_\xi(u) + 1(\Delta_0(u)=0)|\Delta_\xi(u)|du\Big]
\end{aligned} \tag{3.85}$$

which is exactly (3.84).

For (3.82) we have $\Delta_\xi(u) = \mathrm{sgn}(\Delta_\xi(u))|\Delta_\xi(u)|$ and furthermore because of space-shift invariance

$$\mathbf{E}[|\Delta_0(u)|] = \mathbf{E}[|\Delta_\xi(u)|], \tag{3.86}$$

so

$$\begin{aligned}
\dot{K}^{p,\epsilon}_{\mathrm{mig}}(t) &= \sum_{\xi\in\Lambda} a(0,\xi)\int_{I^*}\mathbf{E}\big[\big(\mathrm{sgn}(\Delta_0(u))\mathrm{sgn}(\Delta_\xi(u)) - 1\big)|\Delta_\xi(u)| \\
&\qquad\qquad + 1(\Delta_0(u)=0)|\Delta_\xi(u)|\big]du \\
&= \sum_{\xi\in\Lambda} a(0,\xi)\int_{I^*}\mathbf{E}[-2\cdot 1(\mathrm{sgn}(\Delta_0(u)) \ne \mathrm{sgn}(\Delta_\xi(u)))1(\Delta_0(u)\ne 0)|\Delta_\xi(u)|]du.
\end{aligned} \tag{3.87}$$

□

2. For $\bar{k}$ we again consider $(\overline{y}^1\delta_1, \overline{y}^2\delta_1)$ instead of (y^1, y^2). Then using (3.82) and by space-shift invariance

$$\dot{\bar{K}}^{p,\epsilon}_{\mathrm{mig}}(t) = -2\sum_{\xi\in\Lambda} a(0,\xi)\mathbf{E}[1(\mathrm{sgn}(\bar{y}^1_0 - \bar{y}^2_0) \ne \mathrm{sgn}(\bar{y}^1_\xi - \bar{y}^2_\xi))1(\bar{y}^1_0 \ne \bar{y}^2_0)|\bar{y}^1_\xi - \bar{y}^2_\xi|] \tag{3.88}$$

as $\Delta(u) \neq 0$ is only possible for $u \in I^\star \setminus I$. This is the same as equation (3.64) in Dawson and Greven (2003).

Conclusion. If we now put together the results for the branching and the migration mechanism we derive the results of (3.59) and (3.60) for any ϵ-particle process and therefore also for the diffusion limit.

1. For $\dot{K}^{p,\epsilon}(t)$ the result follows from

$$\dot{K}^{p,\epsilon}(t) = \dot{K}^{p,\epsilon}_{\text{br},1}(t) + \dot{K}^{p,\epsilon}_{\text{br},2}(t) + \dot{K}^{p,\epsilon}_{\text{mig}}(t) \tag{3.89}$$

   with $\dot{K}^{p,\epsilon}_{\text{br},1}$ from (3.64), $\dot{K}^{p,\epsilon}_{\text{br},2}$ from (3.66) and $\dot{K}^{p,\epsilon}_{\text{mig}}$ from (3.82).
2. For $\dot{\bar{K}}^{p,\epsilon}(t)$ the branching has no effect, as was seen in (3.77), so

$$\dot{\bar{K}}^{p,\epsilon}(t) = \dot{\bar{K}}^{p,\epsilon}_{\text{mig}}(t) \tag{3.90}$$

   and (3.60) follows from (3.88).

□

3.3. *Coupling of a finite and an infinite system.* When we want to couple a finite with an infinite system the only new thing we have to do is taking care of the migration as it differs in the finite and infinite system. The way we do our coupling will rely on the same mechanisms as the coupling of two infinite systems introduced in the last section. Given a finite system of size N and an infinite system we will again have to describe the dynamics of the coupled process. The state space of the coupled process will be $\mathcal{E} \times \mathcal{E}^N$. We will denote the coupled process by $Y^N(\cdot) = (y_\xi(\cdot), y^N_\xi(\cdot))_{\xi\in\Lambda}$. We use the notation of the last section.

The construction is very similar to the case of two infinite systems. We will define the pregenerator of the process for functions F having the same form as in (3.26) except that $F : \mathcal{E}^p \times \mathcal{E}^{N,p} \to \mathbb{R}$.

The branching mechanism only deals with one site. So it is identical with the mechanism for two infinite systems.

For the migration mechanisms two sites ξ, η are involved and things differ in the finite versus infinite system on the border of Λ^N. Assume $\xi, \eta \in \Lambda^N$ and suppose in the infinite system a migration from ξ to η occurs. Then via the measures $\psi(y, y^N)$ also in the finite system a corresponding transition occurs. But if $\xi \in \Lambda^N, \eta \notin \Lambda^N$ things differ in the finite versus infinite system. The dynamics suggests that in the infinite system a transition from ξ to η occurs but in the finite system from ξ to η mod N. If $\xi \notin \Lambda^N$ it is clear that only in the infinite system a change can occur.

This leads to the following pregenerator for the diffusion limit of the coupling of an infinite and a finite process. For F of the form (3.37)

$$\begin{aligned}
G_{\text{coup}}^{N,d}F(Y) = &\sum_{\xi\in\Lambda} h(\overline{y}_\xi) \wedge h(\overline{y}_\xi^N) \int_{(I^\star)^2} \tfrac{1}{2}\frac{\partial^2 F(Y)}{\partial y_\xi^2}(u,v)\psi(y_\xi, y_\xi^N)(d(u,v)) \\
&+ (h(\overline{y}_\xi) - h(\overline{y}_\xi^N))^+ \int_I \tfrac{1}{2}\frac{\partial^2 F(Y)}{\partial y_\xi^2}(u,\star) y_\xi(du) \\
&\quad + (h(\overline{y}_\xi) - h(\overline{y}_\xi^N))^- \int_I \tfrac{1}{2}\frac{\partial^2 F(Y)}{\partial y_\xi^2}(\star,v) y_\xi^N(dv) \\
&+ \sum_{\xi,\eta\in\Lambda^N} a(\eta,\xi) \int_{(I^\star)^2} \Big(\frac{\partial F(Y)}{\partial y_\eta}(u,v) - \frac{\partial F(Y)}{\partial y_\xi}(u,v)\Big)\psi(y_\xi, y_\xi^N)(d(u,v)) \\
&+ \sum_{\xi\in\Lambda^N,\eta\notin\Lambda^N} a(\eta,\xi) \int_{(I^\star)^2} \Big(\frac{\partial F(Y)}{\partial y_{\eta \bmod N}}(\star,v) + \frac{\partial F(Y)}{\partial y_\eta}(u,\star) \\
&\qquad\qquad - \frac{\partial F(Y)}{\partial y_\xi}(u,v)\Big)\psi(y_\xi, y_\xi^N)(d(u,v)) \\
&+ \sum_{\xi\notin\Lambda^N} a(\eta,\xi) \int_{I^\star} \Big(\frac{\partial F(Y)}{\partial y_\eta}(u,\star) - \frac{\partial F(Y)}{\partial y_\xi}(u,\star)\Big) y_\xi(du).
\end{aligned} \tag{3.91}$$

Then the existence of a process with this pregenerator is given by:

Proposition 3.13. *Let $\mu = \mathcal{P}(\mathcal{E} \times \mathcal{E}^N)$ with $\pi_i\mu \in \mathcal{L}_2(\mathcal{E}), i = 1,2$. Then the $(\mu, G_{coup}^{N,d})$ martingale problem has a solution. Any solution Y^N of this martingale problem is a coupling for our processes X and X^N, i.e. $\mathcal{L}(\pi_1 Y^N) = \mathcal{L}(X)$ and $\mathcal{L}(\pi_2 Y) = \mathcal{L}(X^N)$ given $\mathcal{L}(X(0)) = \pi_1\mu$ and $\mathcal{L}(X^N(0)) = \pi_2\mu$.*

The proof works analogously as for the coupling of two infinite systems and is omitted.

Functionals of the coupled process. Again we want to apply our coupling to the Wasserstein metric of the two components of our process, i.e.

$$K^{N,d}(t) := \mathbf{E}[(k \circ \pi_\xi)(Y^N(t))]. \tag{3.92}$$

As we are dealing with shift invariant initial measures (see Lemma 3.6) we take $\xi = 0$. We derive the following proposition.

Proposition 3.14. *Let $\mu \in \mathcal{T}_2(\mathcal{E} \times \mathcal{E}^N)$. Then for the mean effect of the dynamics on the Wasserstein metric*

$$\begin{aligned}
\frac{d}{dt}K^{N,d}(t) &= \int_{I^\star} \mathbf{E}[|h(\overline{y}_0(t)) - h(\overline{y}_0^N(t))| y_0(t)([0,u]), \Delta_0(u)(t) = 0] du \\
&- 2 \sum_{\xi \in \Lambda^N} a(0,\xi) \int_{I^\star} \mathbf{E}[1(sgn(\Delta_0(u)(t)) \neq sgn(\Delta_\xi(u)(t))) \\
&\qquad\qquad 1(\Delta_0(u)(t) \neq 0)|\Delta_\xi(u)(t)|] du \\
&+ \sum_{\xi \notin \Lambda^N} a(0,\xi) \int_{I^\star} \mathbf{E}[(sgn(\Delta_0(u)(t)) + 1(\Delta_0(u)(t) = 0))\overline{y}_\xi(t) - |\Delta_0(u)(t)|] du \\
&+ \sum_{\xi \in \Lambda^N} (a^N(0,\xi) - a(0,\xi)) \Big(2\mathbf{E}[\overline{y}_\xi^N(t)] \\
&\quad - 2\mathbf{E}\Big[\int_I v y_\xi^N(t)(v) dv + \int_{I^\star} 1(y_0(t)([0,v]) > y_0^N(t)([0,v])) y_\xi^N([0,v]) dv \Big] \Big).
\end{aligned} \tag{3.93}$$

Proof. Our reasoning is the same as in the last section. We would have to construct approximating particle processes and prove that the above equation holds for all of them and therefore carries over to the diffusion limit. Branching only affects one site and so its effect is the same as for the coupling of two infinite systems. The migration is not so simple. We have three cases here. Assume the transition in the infinite system is from ξ to η and we study the effect in 0.

The first case is $\xi \in \Lambda^N, \eta = 0$ which makes the same effect as for two infinite systems. For $\xi \notin \Lambda^N, \eta = 0$ only the infinite system is affected. Using (3.81) this gives the second migration term in (3.93). The third case is $\xi \in \Lambda^N, \eta = 0 \mod N$. Here the effect is the same as a transition of $(\star, v)$ from ξ to 0 directly. Altogether we arrive at (3.93). Some more details are found in Pfaffelhuber (2003). □

4. Proof of Theorem *A*

In Cox and Greven (1994b) a certain way of proving the finite system scheme was proposed. We will differ from the program of this paper. That is because we cannot prove with our coupling that the expected distance of two infinite systems will decrease over time. (That is needed for the proof of A8 of the above paper.) Therefore we will alter the proof to adapt it to our needs.

In the abstract theorem of Cox and Greven (1994b) two time scales are needed. That are a macroscopic time scale, β_N, and a microscopic one, l_N. The idea was that a macroscopic variable of the finite system, the density Θ^N will change only in the macroscopic time scale but for times $t_N = o(\beta_N)$ the density is constant. This is only an asymptotic statement because for any finite system the density will according to local fluctuations change a little for any time t. The other time scale, l_N, was then needed to guarantee that the system relaxes into a local equilibrium.

We will also need these two time scales but in our proof the definition of the small time scale, which we will call p_N is more involved. In contrast to the program of Cox and Greven (1994b) we need two more time scales (called l_N and m_N) in order to define p_N and obtain the desired convergence result. All these time scales will be defined in Section 4.4.

We begin this chapter by sketching the whole proof in Section 4.1, mainly according to the figure on the next page. We propose a four step program of the proof which is carried out in the proceeding sections.

4.1. *Sketch of the proof.* In Theorem A we had the choice between two assumptions. Here we assume $t_N := t\beta_N$ and $t > 0$. For the other cases see the remark at the end of this section.

The proof of Theorem A will be broken down to several steps. The figure on the next page shows how we want to proceed. Here we use the notation $\mathcal{J} := \mathcal{M}_f(I)$ and $S(\cdot)$ and $S^N(\cdot)$ are the semigroups of the infinite and the finite system. First we explain our steps according to the diagram.

In *Step 1* we prove tightness for the sequence of laws

$$(\mathcal{L}(Z^N(\cdot)))_{N\in\mathbb{N}} = (\mathcal{L}(\Theta^N(X^N(\cdot\beta_N))))_{N\in\mathbb{N}} \tag{4.1}$$

in Proposition 4.1. By this tightness and Prohorov's theorem we pick a convergent subsequence in Corollary 4.2 such that

$$(Z^{N_k}(t))_{t\geq 0} = (\Theta^{N_k}(X^{N_k}(t\beta_{N_k})))_{t\geq 0} \overset{k\to\infty}{\Rightarrow} (\breve{Z}(t))_{t\geq 0} \tag{4.2}$$

for some process $(\breve{Z}(t))_{t\geq 0}$. We even have by Corollary 4.3 that for any sequence $(p_N)_{N\in\mathbb{N}}$ with $p_N = o(\beta_N)$

$$\Theta^{N_k}(X^{N_k}(t\beta_{N_k} - p_{N_k})) \overset{k\to\infty}{\Rightarrow} \breve{Z}(t) \tag{4.3}$$

what will be used in Step 2. Recall that because of the assumptions of Theorem A we have $\breve{Z}(0) = \theta$ for any subsequential weak limit point.

In *Step 2* we take an arbitrary sequence $(p_N)_{N\in\mathbb{N}}$ with $p_N = o(\beta_N)$ and consider the sequence $X^{N_k}(t\beta_{N_k} - p_{N_k})$, i.e. the coordinate processes of the finite systems at time $t\beta_{N_k} - p_{N_k}$. The sequence $(\mathcal{L}(X^{N_k}(t\beta_{N_k} - p_{N_k})))_{k\in\mathbb{N}}$ is tight by Proposition 4.5 since we started in the right state space so we can pick a convergent subsequence

$$\mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k})) \overset{k\to\infty}{\Rightarrow} \mu(t). \tag{4.4}$$

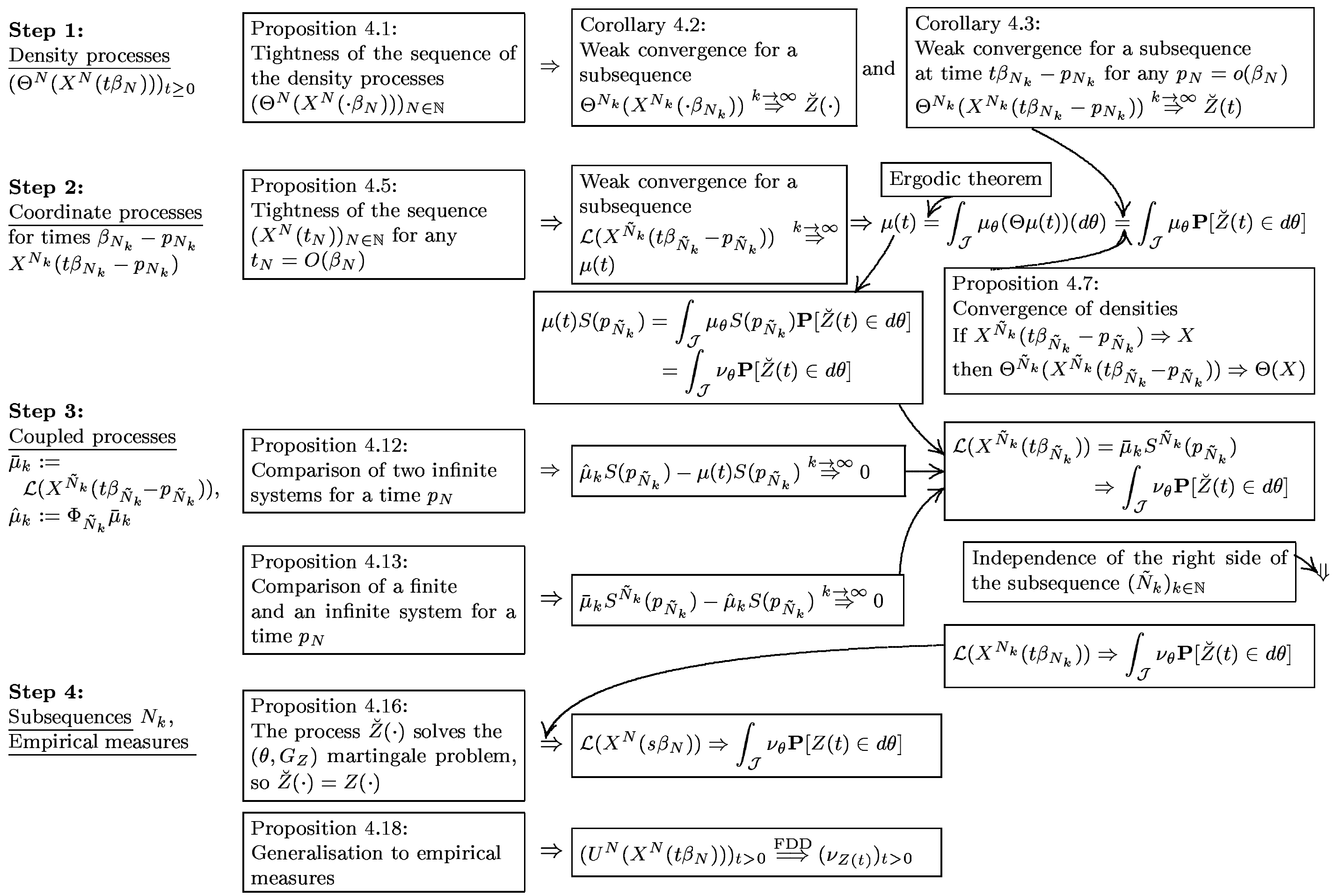
Now we use the ergodic theorem and bring Proposition 4.7 into play. We will obtain

$$\mu(t) = \int_{\mathcal{M}_f(I)} \mu_\theta(\Theta\mu(t))(d\theta) = \int_{\mathcal{M}_f(I)} \mu_\theta \mathbf{P}[\breve{Z}(t) \in d\theta]. \tag{4.5}$$

Here $\Theta\mu(t)$ is the measure of the map Θ under $\mu(t)$ and $\mu_\theta \in \mathcal{M}_\theta$. The first '=' is derived by the ergodic theorem for stationary random fields as we can make the ergodic decomposition of $\mu(t)$.

The second '=' comes from the fact that along a convergent sequence of the coordinate process the sequence of densities also converges by Proposition 4.7 and the limit point was identified as $\breve{Z}(t)$ in Step 1. So $X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k})$ converges weakly to some mixture of μ_θ's that corresponds to the probabilities of the densities along that subsequence $\tilde{N}_k$.

In *Step 3* we want to find a special sequence $(p_N)_{N\in\mathbb{N}} = o(\beta_N)$. When we have found it we will let the system $X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k})$ evolve for some time $p_{\tilde{N}_k}$. Since we know that the density is asymptotically constant over time spans $o(\beta_N)$ (otherwise the sequence of rescaled density processes could not be tight) we will later on justify



Step 1: Density processes $(\Theta^N(X^N(t\beta_N)))_{t\geq 0}$
Proposition 4.1: Tightness of the sequence of the density processes $(\Theta^N(X^N(\cdot\beta_N)))_{N\in\mathbb{N}}$
$\Rightarrow$
Corollary 4.2: Weak convergence for a subsequence $\Theta^{N_k}(X^{N_k}(\cdot\beta_{N_k})) \overset{k\to\infty}{\Rightarrow} \breve{Z}(\cdot)$
and
Corollary 4.3: Weak convergence for a subsequence at time $t\beta_{N_k} - p_{N_k}$ for any $p_N = o(\beta_N)$ $\Theta^{N_k}(X^{N_k}(t\beta_{N_k} - p_{N_k})) \overset{k\to\infty}{\Rightarrow} \breve{Z}(t)$
Step 2: Coordinate processes for times $\beta_{N_k} - p_{N_k}$ $X^{N_k}(t\beta_{N_k} - p_{N_k})$
Proposition 4.5: Tightness of the sequence $(X^N(t_N))_{N\in\mathbb{N}}$ for any $t_N = O(\beta_N)$
$\Rightarrow$
Weak convergence for a subsequence $\mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k})) \overset{k\to\infty}{\Rightarrow} \mu(t)$
Ergodic theorem
$\Rightarrow \mu(t) = \int_{\mathcal{J}} \mu_\theta(\Theta\mu(t))(d\theta) = \int_{\mathcal{J}} \mu_\theta \mathbf{P}[\breve{Z}(t) \in d\theta]$
$\mu(t)S(p_{\tilde{N}_k}) = \int_{\mathcal{J}} \mu_\theta S(p_{\tilde{N}_k})\mathbf{P}[\breve{Z}(t) \in d\theta] = \int_{\mathcal{J}} \nu_\theta \mathbf{P}[\breve{Z}(t) \in d\theta]$
Proposition 4.7: Convergence of densities If $X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k}) \Rightarrow X$ then $\Theta^{\tilde{N}_k}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k})) \Rightarrow \Theta(X)$
Step 3: Coupled processes $\bar{\mu}_k := \mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k}))$, $\hat{\mu}_k := \Phi_{\tilde{N}_k}\bar{\mu}_k$
Proposition 4.12: Comparison of two infinite systems for a time p_N
$\Rightarrow$
$\hat{\mu}_k S(p_{\tilde{N}_k}) - \mu(t)S(p_{\tilde{N}_k}) \overset{k\to\infty}{\Rightarrow} 0$
$\mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k})) = \bar{\mu}_k S^{\tilde{N}_k}(p_{\tilde{N}_k}) \Rightarrow \int_{\mathcal{J}} \nu_\theta \mathbf{P}[\breve{Z}(t) \in d\theta]$
Proposition 4.13: Comparison of a finite and an infinite system for a time p_N
$\Rightarrow$
$\bar{\mu}_k S^{\tilde{N}_k}(p_{\tilde{N}_k}) - \hat{\mu}_k S(p_{\tilde{N}_k}) \overset{k\to\infty}{\Rightarrow} 0$
Independence of the right side of the subsequence $(\tilde{N}_k)_{k\in\mathbb{N}}$ $\Downarrow$
$\mathcal{L}(X^{N_k}(t\beta_{N_k})) \Rightarrow \int_{\mathcal{J}} \nu_\theta \mathbf{P}[\breve{Z}(t) \in d\theta]$
Step 4: Subsequences N_k, Empirical measures
Proposition 4.16: The process $\breve{Z}(\cdot)$ solves the (θ, G_Z) martingale problem, so $\breve{Z}(\cdot) = Z(\cdot)$
$\Rightarrow$
$\mathcal{L}(X^N(s\beta_N)) \Rightarrow \int_{\mathcal{J}} \nu_\theta \mathbf{P}[Z(t) \in d\theta]$
Proposition 4.18: Generalisation to empirical measures
$\Rightarrow$
$(U^N(X^N(t\beta_N)))_{t>0} \overset{\text{FDD}}{\Longrightarrow} (\nu_{Z(t)})_{t>0}$

the following sequence of asymptotic equivalences

$$\begin{aligned}\mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k})) &= \mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k}))S^{\tilde{N}_k}(p_{\tilde{N}_k}) \approx \mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k}))S(p_{\tilde{N}_k}) \\ &\approx \int_{\mathcal{M}_f(I)} \mu_\theta S(p_{\tilde{N}_k})\mathbf{P}[\breve{Z}(t) \in d\theta] \approx \int_{\mathcal{M}_f(I)} \nu_\theta \mathbf{P}[\breve{Z}(t) \in d\theta]\end{aligned} \tag{4.6}$$

for large k. Here $S(\cdot)$ and $S^N(\cdot)$ denote the semigroups of the infinite and finite systems respectively. Now the right side is no longer dependent on the subsequence $\tilde{N}_k$, so the calculation holds for $(N_k)_{k\in\mathbb{N}}$. To control the three '$\approx$' in the above equations we have to choose the right sequence $(p_N)_{N\in\mathbb{N}}$; how we do this exactly is explained in the beginning of Step 3. Then we prove two things.

For the first '$\approx$' we prove that for initial measures $\mu^N = X^N(t\beta_N - p_N)$ and $\mu = \Phi_N\mu^N$, i.e. for equal initial distributions at the start, the evolution on Λ^N and on Λ are asymptotically equal by Proposition 4.13,

$$\mu^N S^N(p_N) - \mu S(p_N) \overset{N\to\infty}{\Rightarrow} 0. \tag{4.7}$$

We will need the coupling of infinite and finite systems of Section 3.2.

For the second '$\approx$' we already have that for any sequence $(p_N)_{N\in\mathbb{N}}$ we can take a subsequence $(p_{\tilde{N}_k})_{k\in\mathbb{N}}$ with $\mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k})) \Rightarrow \mu(t)$ for some $\mu(t)$. So if we could simply replace $\mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k}))$ by $\mu(t)$ the '$\approx$' would be clear from Step 2. We will prove in Proposition 4.12 that the above '$\approx$' holds. To derive this result several steps are needed that all use the coupling techniques from Chapter 3. As we saw there the increase in the Wasserstein distance can be controlled by the process of total masses which can be handled more easily.

The third '$\approx$' is a consequence of the long-time behaviour of our system which was summarised in Theorem 1.4.

This Step 3 is not standard and here is where we differ from the proofs of the finite system scheme for other models, e.g. interacting diffusions in Cox et al. (1995). Namely the sequence p_N will depend on several other things and can not be computed in one step. Also we constructed the coupling from Chapter 3 mainly for this purpose. In Cox et al. (1995) when calculating the difference of two coupled processes you derive a Lyapunov function, i.e. a function decreasing over time. This is not necessarily the case for our coupling as we have seen e.g. in (3.59).

In *Step 4* we prove that by Proposition 4.16 no matter which sequence $(N_k)_{k\in\mathbb{N}}$ we pick in Step 1 the process $(\breve{Z}(t))_{t\geq 0}$ is the same. We do this as in the corresponding proof in Cox et al. (1995) with some minor modifications. Namely we introduce empirical measures of our process and prove that this process solves the (θ, G_Z) local martingale problem with θ from the assumptions of our theorem and G_Z from (2.8). As this martingale problem is well posed the process $\breve{Z}(\cdot)$ is independent of the subsequence N_k and equal to $Z(\cdot)$. In this step we will as well see that not only (2.11) holds but the stronger result (2.12). This is done in Lemma 4.17 and Proposition 4.18.

Remark.

1. Here we only sketched the proof of the theorem for t_N and μ^N fulfilling the first assumption of Theorem A with the special choice $t_N = t\beta_N$ and $t > 0$.

The generalisation to general sequences that fulfil the first assumption of Theorem A is straightforward and is obtained by using the same techniques as in the proof here.

If the second assumption of Theorem A is fulfilled then we are in the case $t = 0$. Here we need an additional assumption for the theorem to be true. That is due to Proposition 4.7 as it is only valid for the additional assumption $\mu^N \Rightarrow \mu \in \mathcal{M}_\theta$.

2. Concerning the existence of second moments for the initial distributions μ^N we first make in our proof the stronger assumption of uniform existence of qth moments for some $q > 2$. Then the restriction to second moments follows by a truncation argument for the μ^N and coupling techniques. Again it is Proposition 4.7 where these qth moments are needed.

4.2. *Step 1: Tightness of the density process.* In this step we prove tightness of the sequence of rescaled density processes in Proposition 4.1. We prove this to be able to pick a convergent subsequence. That can be done by Corollary 4.2. Then along this subsequence we even have convergence at a fixed time for a slightly different time scale in Corollary 4.3.

Proposition 4.1. *The sequence* $(Z^N(\cdot))_{N\in\mathbb{N}}$ *is tight.*

Proof. We proceed in two steps. Recall that the processes $Z^N(\cdot)$ have state space $\mathcal{M}_f(I)$. By Jakubowski's criterion for tightness (cf. Dawson (1993), Theorem 3.6.4.) we state in the first step that it is enough to prove tightness for certain real-valued processes. In the second step we prove tightness for these processes using Aldous' criterion for real-valued processes (see Dawson (1993), Theorem 3.6.5).

The space $\mathcal{M}_f(I)$, endowed with the topology of weak convergence, is Polish (see e.g. Dawson (1993) 3.1.1). We set

$$\mathcal{F} := \{F : \mathcal{M}_f(I) \to \mathbb{R}_+ : \exists f_F \in \mathcal{C}(I) : F(x) = \langle f_F, x\rangle, f_F \geq 0\}. \tag{4.8}$$

Clearly $\mathcal{F}$ separates points in $\mathcal{M}_f(I)$ and is closed under addition. For tightness of $(Z^N(\cdot))_{N\in\mathbb{N}}$ we have to verify two things:

- For each $\epsilon, T > 0$ there is a compact $K_{\epsilon,T} \subseteq \mathcal{M}_f(I)$:
$$\inf_{N\in\mathbb{N}} \mathbf{P}[Z^N(t) \in K_{\epsilon,T} \ \ (0 \leq t \leq T)] \geq 1 - \epsilon. \qquad (n \in \mathbb{N}) \tag{4.9}$$
- For each $F \in \mathcal{F}$ the sequence $(F(Z^N(\cdot)))_{N\in\mathbb{N}}$ is tight.

Step 1: Paths of $\mathbf{Z^N}(\cdot)$ *are in a compact set with high probability.* For 1., we already know, that the sequence $(\overline{Z}^N(\cdot))_{N\in\mathbb{N}}$ with $\overline{Z}^N := \langle Z^N, 1\rangle$ is tight since $(\overline{Z}^N(\cdot))_{N\in\mathbb{N}}$ is the same sequence as the sequence of density processes in Cox et al. (1995) and tightness for this process was shown in (3.1) of this paper. That means by the definition of tightness that

$$\forall \epsilon > 0\ \exists K_\epsilon \subseteq D([0,\infty), \mathbb{R}_+) \text{ compact : } \inf_{N\in\mathbb{N}} \mathbf{P}[\overline{Z}^N(\cdot) \in K_\epsilon] \geq 1 - \epsilon \tag{4.10}$$

where $D([0,\infty), \mathbb{R}_+)$ is equipped with the Skorohod topology. By Theorem 3.6.3 and Remark 3.6.4 of Ethier and Kurtz (1986) it is necessary for the set K_ϵ that

$$\forall \epsilon, T > 0\ \exists \Gamma_{\epsilon,T} \subseteq \mathbb{R}_+ \text{ compact : } Z^N(t) \in \Gamma_{\epsilon,T} \quad (x \in K_\epsilon, 0 \leq t \leq T). \tag{4.11}$$

Therefore

$$\inf_{N\in\mathbb{N}} \mathbf{P}[\overline{Z}^N(t) \in \Gamma_{\epsilon,T} \ \ (0 \le t \le T)] \ge \inf_{N\in\mathbb{N}} \mathbf{P}[\overline{Z}^N(\cdot) \in K_\epsilon]. \tag{4.12}$$

Now define

$$K_{\epsilon,T} := \{x \in \mathcal{M}_f(I) : \langle x, 1\rangle \in \Gamma_{\epsilon,T}\} \qquad (\epsilon, T > 0), \tag{4.13}$$

which is compact. Then we have

$$\inf_{N\in\mathbb{N}} \mathbf{P}[Z^N(t) \in K_{\epsilon,T} \ \ (0 \le t \le T)] = \inf_{N\in\mathbb{N}} \mathbf{P}[\overline{Z}^N(t) \in \Gamma_{\epsilon,T} \ \ (0 \le t \le T)] \ge 1 - \epsilon. \tag{4.14}$$

Step 2: Tightness of $(\boldsymbol{F}(\boldsymbol{Z^N}(\cdot)))_{\boldsymbol{N}\in\mathbb{N}}$. Next, we want to prove the second assumption of the tightness criterion by Jakubowski's in our special case. We do this by using a theorem about tightness from Aldous. In the sequel we fix $F \in \mathcal{F}$ and the corresponding $f := f_F$ respectively such that $F(Z^N) = \langle Z^N, f\rangle$. To calculate the infinitesimal characteristics of $(\langle \Theta^N(X^N(t)), f\rangle)_{t\ge 0}$ we use the pregenerator G^N from (1.22) of the underlying process $X^N(\cdot)$. Take any $g \in \mathcal{C}_b^2(\mathbb{R})$. Then

$$\begin{aligned} \frac{\partial g(\langle \Theta^N(X), f\rangle)}{\partial x_\xi^N}(u) &= \frac{1}{|\Lambda^N|} g'(\langle \Theta^N(X), f\rangle) f(u), \\ \frac{\partial^2 g(\langle \Theta^N(X), f\rangle)}{\partial (x_\xi^N)^2}(u,v) &= \frac{1}{|\Lambda^N|^2} g''(\langle \Theta^N(X), f\rangle) f(u) f(v). \end{aligned} \tag{4.15}$$

Hence we have by (1.22)

$$\begin{aligned} G^N g(\langle \Theta^N, f\rangle) &= G^N g\Big(\langle \frac{1}{|\Lambda^N|} \sum_{\xi\in\Lambda^N} x_\xi^N, f\rangle\Big) \\ &= \sum_{\xi\in\Lambda^N} \sum_{\eta\in\Lambda^N} \frac{1}{|\Lambda^N|} g'(\langle \Theta^N, f\rangle)\big(a^N(\xi,\eta)\langle x_\eta^N, f\rangle - \langle x_\xi^N, f\rangle\big) \\ &\qquad + \sum_{\xi\in\Lambda^N} \frac{1}{|\Lambda^N|^2} g''(\langle x_\xi^N, f\rangle) \tfrac{1}{2} h(\overline{x}_\xi^N) \langle x_\xi^N, f^2\rangle \\ &= \sum_{\xi\in\Lambda^N} \frac{1}{|\Lambda^N|^2} g''(\langle x_\xi^N, f\rangle) \tfrac{1}{2} h(\overline{x}_\xi^N) \langle x_\xi^N, f^2\rangle. \end{aligned} \tag{4.16}$$

So we have for the infinitesimal characteristics of $(\langle Z^N(\cdot), f\rangle)_{t\ge 0}$

$$G^N g(\langle Z^N, f\rangle) = \beta_N G^N g(\langle \Theta^N(X), f\rangle) = \frac{1}{|\Lambda^N|} \sum_{\xi\in\Lambda^N} g''(\langle x_\xi^N, f\rangle) \tfrac{1}{2} h(\overline{x}_\xi^N) \langle x_\xi^N, f^2\rangle. \tag{4.17}$$

We know from Cox et al. (1995), Lemma 2.2.c
that

$$\begin{aligned} \sup_{N\in\mathbb{N}} \sup_{0\le t\le T\beta_N} \mathbf{E}[h(\overline{x}_\xi^N(t))\langle x_\xi^N(t), f^2\rangle] &\le ||f||_\infty^2 \sup_{N\in\mathbb{N}} \sup_{0\le t\le T\beta_N} \mathbf{E}[h(\overline{x}_\xi^N(t))\overline{x}_\xi^N(t)] \\ &\le \sup_{N\in\mathbb{N}} \sup_{0\le t\le T\beta_N} \mathbf{E}[(\overline{x}_\xi^N(t))^2] =: C_{T,f} < \infty. \end{aligned} \tag{4.18}$$

Therefore $(\langle Z^N(t), f\rangle)_{t\geq 0}$ is a square integrable martingale with increasing process

$$[\langle Z^N, f\rangle](t) = \frac{1}{|\Lambda^N|}\int_0^t \sum_{\xi\in\Lambda^N} h(\overline{x}_\xi^N(s\beta_N))\langle x_\xi^N(s\beta_N), f^2\rangle ds. \tag{4.19}$$

Now we use the tightness criterion of Aldous. We have to proof two assertions:

a) for each $t \geq 0$ the sequence $(\langle Z^N(t), f\rangle)_{N\in\mathbb{N}}$ is tight,

b) for each $\epsilon, T > 0$, any sequence of stopping times $(\tau_N)_{N\in\mathbb{N}}$, bounded by T and any sequence $(\delta_N)_{N\in\mathbb{N}}$ with $\delta_N \downarrow 0$

$$\lim_{N\to\infty} \mathbf{P}[|\langle Z^N(\tau_N + \delta_N), f\rangle - \langle Z^N(\tau_N), f\rangle| > \epsilon] = 0. \tag{4.20}$$

For a) we already know from 1. that $(\langle Z^N(t), 1\rangle)_{N\in\mathbb{N}}$ is tight and so

$$\begin{aligned}\forall\epsilon > 0\ \exists K_\epsilon:\ \inf_{N\in\mathbb{N}} \mathbf{P}[\langle Z^N(t), f\rangle \in ||f||_\infty K_\epsilon] &\geq \mathbf{P}[\langle Z^N(t), ||f||_\infty\rangle \in ||f||_\infty K_\epsilon]\\ &= \mathbf{P}[\langle Z^N(t), 1\rangle \in K_\epsilon] \geq 1-\epsilon.\end{aligned} \tag{4.21}$$

For b) we know that $(\langle Z^N(\cdot), f\rangle)_{t\geq 0}$ is a square integrable martingale and we know

that the underlying processes $(X^N(t))_{t\geq 0}$ are strong Markov processes. The reason for that is that the martingale problem for G^N from (1.22) has a unique solution and uniqueness of the martingale problem implies the strong Markov property (see Ethier and Kurtz (1986), Theorem 4.4.2). So we have

$$\begin{aligned}
&\lim_{N\to\infty} \mathbf{P}[|\langle Z^N(\tau_N + \delta_N), f\rangle - \langle Z^N(\tau_N), f\rangle| > \epsilon]\\
&\quad\leq \lim_{N\to\infty} \frac{1}{\epsilon^2}\mathbf{E}[(\langle Z^N(\tau_N+\delta_N), f\rangle - \langle Z^N(\tau_N), f\rangle)^2]\\
&\quad= \lim_{N\to\infty} \frac{1}{\epsilon^2}\mathbf{E}\big[\mathbf{E}^{X^N_{\tau_N}}[(\langle Z^N(\delta_N), f\rangle - \langle Z^N(0), f\rangle)^2]\big]\\
&\quad= \lim_{N\to\infty} \frac{1}{\epsilon^2}\mathbf{E}\big[\mathbf{E}^{X^N_{\tau_N}}[[\langle Z^N, f\rangle](\delta_N) - [\langle Z^N, f\rangle](0)]\big]\\
&\quad= \lim_{N\to\infty} \frac{1}{\epsilon^2}\mathbf{E}\Big[\mathbf{E}^{X^N_{\tau_N}}\Big[\frac{1}{|\Lambda^N|}\int_0^{\delta_N}\sum_{\xi\in\Lambda^N} h(\overline{x}_\xi^N(s))\langle x_\xi^N(s), f^2\rangle ds\Big]\Big]\\
&\quad= \lim_{N\to\infty} \frac{1}{\epsilon^2}\mathbf{E}\Big[\int_{\tau_N}^{\tau_N+\delta_N}\frac{1}{|\Lambda^N|}\sum_{\xi\in\Lambda^N} h(\overline{x}_\xi^N(s))\langle x_\xi^N(s), f^2\rangle ds\Big]\\
&\quad= \lim_{N\to\infty} \frac{1}{\epsilon^2}\int_0^{\delta_N}\frac{1}{|\Lambda^N|}\sum_{\xi\in\Lambda^N} \mathbf{E}\big[h(\overline{x}_\xi^N(\tau_N+s))\langle x_\xi^N(\tau_N+s), f^2\rangle\big] ds\\
&\quad\leq \lim_{N\to\infty} \frac{1}{\epsilon^2}\delta_N C_{T,f} = 0
\end{aligned} \tag{4.22}$$

and we are done. □

Corollary 4.2. *There is a subsequence $(N_k)_{k\in\mathbb{N}}$ such that*

$$(Z^{N_k}(t))_{t\geq 0} \overset{k\to\infty}{\Rightarrow} (\breve{Z}(t))_{t\geq 0} \tag{4.23}$$

for some process $(\breve{Z}(t))_{t\geq 0}$ with paths in $D([0,\infty), \mathcal{M}_f(I))$.

Proof. Since $\mathcal{M}_f(I)$ is Polish, $D([0,\infty),\mathcal{M}_f(I))$ is separable and complete with respect to the Skorohod metric, (see Ethier and Kurtz (1986), 3.5.6). Since $\mathcal{L}(Z^N(\cdot)) \in \mathcal{P}(D([0,\infty),\mathcal{M}_f(I)))$ $(N \in \mathbb{N})$, we have that the sequence of laws is relatively compact in $\mathcal{P}(D([0,\infty),\mathcal{M}_f(I)))$ by Prohorov's theorem (see Ethier and Kurtz (1986), 3.2.2). So there is a convergent subsequence, say $(\mathcal{L}(Z^{N_k}(\cdot)))_{k\in\mathbb{N}}$, i.e. $Z^{N_k}(\cdot)$ converges weakly. □

Corollary 4.3. *Let $(p_N)_{N\in\mathbb{N}}$ be a sequence with $p_N = o(\beta_N)$. Take N_k and $(\breve{Z}(t))_{t\geq 0}$ from Corollary 4.2. Then we have*

$$\Theta^{N_k}(X^{N_k}(t\beta_{N_k} - p_{N_k})) \overset{k\to\infty}{\Rightarrow} \breve{Z}(t) \qquad (t \geq 0). \tag{4.24}$$

Proof. For fixed $t \geq 0$ and $f \in \mathcal{C}(I)$ we calculate

$$\begin{aligned}
&\lim_{k\to\infty} \mathbf{E}\big[(\langle\Theta^{N_k}(X^{N_k}(t\beta_{N_k} - p_{N_k})), f\rangle - \langle\Theta^{N_k}(X^{N_k}(t\beta_{N_k})), f\rangle)^2\big] \\
&\quad = \lim_{k\to\infty} \mathbf{E}\big[(\langle Z^{N_k}(t - \tfrac{p_{N_k}}{\beta_{N_k}}), f\rangle - \langle Z^{N_k}(t), f\rangle)^2\big] \\
&\quad = \lim_{k\to\infty} \mathbf{E}\big[[\langle Z^{N_k}, f\rangle](t - \tfrac{p_{N_k}}{\beta_{N_k}}) - [\langle Z^{N_k}, f\rangle](t)\big] \\
&\quad = \lim_{k\to\infty} \mathbf{E}\Big[\int_{t-\frac{p_{N_k}}{\beta_{N_k}}}^{t} \frac{1}{|\Lambda^{N_k}|} \sum_{\xi\in\Lambda^{N_k}} h(\overline{x}_\xi^{N_k}(s))\langle x_\xi^{N_k}(s), f^2\rangle ds\Big] \\
&\quad \leq \lim_{k\to\infty} \frac{p_{N_k}}{\beta_{N_k}} C_{t,f} = 0
\end{aligned} \tag{4.25}$$

as $p_N = o(\beta_N)$ and $\beta_N = |\Lambda^N|$.

By this we see that also

$$|\langle\Theta^{N_k}(X^{N_k}(t\beta_{N_k} - p_{N_k})), f\rangle - \langle\Theta^{N_k}(X^{N_k}(t\beta_{N_k})), f\rangle| \xrightarrow{N\to\infty} 0 \tag{4.26}$$

in probability. So we can apply Slutzky's Theorem (Corollary 3.3.2 in Ethier and Kurtz (1986)) to obtain

$$\langle\Theta^{N_k}(X^{N_k}(t\beta_{N_k})), f\rangle \overset{k\to\infty}{\Rightarrow} \langle\breve{Z}(t), f\rangle \qquad (t \geq 0, f \in \mathcal{C}_b(I)). \tag{4.27}$$

So the corollary follows by Theorem 4.2 of Kallenberg (1983).

□

4.3. *Step 2: Representation of the coordinate processes.* Here we have to prove two facts. First, $(X^N(t_N))_{N\in\mathbb{N}}$ is tight for any sequence $(t_N)_{N\in\mathbb{N}}$ with $t_N = O(\beta_N)$ (see Proposition 4.5) and second the density process converges if the coordinate process does (see Proposition 4.7).

We start by recalling a basic proposition from Cox et al. (1995).

Proposition 4.4. *Assume $d \geq 3$, i.e. the random walk kernel $\hat{a}(.,.)$ is transient.*

1. *The existence of second moments is preserved by the evolution for the infinite system, i.e.*

$$\mu \in \mathcal{T}_2(\mathcal{E}) \quad \Rightarrow \quad \sup_{t\geq 0} \mathbf{E}^\mu[(\overline{x}_0(t))^2] =: M < \infty. \tag{4.28}$$

Moreover if $C = \langle \mu, \overline{x}_0^2 \rangle$ then M depends on μ only through C and is increasing in C.

2. *Also qth moments are preserved under the evolution of the infinite system, i.e. there is a $q > 2$ such that for any $2 \leq q' \leq q$:*

$$\mu \in \mathcal{T}_{q'}(\mathcal{E}) \quad \Rightarrow \quad \sup_{t\geq 0} \mathbf{E}^{\mu}[|\overline{x}_0(t)|^{q'}] < \infty \tag{4.29}$$

3. *The existence of second moments is preserved by the evolution for the finite system for times up to $\mathcal{O}(\beta_N)$, i.e. for all $T < \infty$*

$$\mu^N \in \mathcal{T}_2(\mathcal{E}^N), \quad \sup_{N\in\mathbb{N}} \langle\!\langle \mu^N, \overline{x}_0^2 \rangle\!\rangle < \infty \quad \Rightarrow \quad \sup_{N\in\mathbb{N}} \sup_{0\leq t\leq T\beta_N} \mathbf{E}^{\mu^N}[|\overline{x}_0^N(t)|^2] < \infty \tag{4.30}$$

4. *Again qth moments are preserved for times up to $\mathcal{O}(\beta_N)$, i.e. there is a $q > 2$ such that for any $2 \leq q' \leq q$ and all $T < \infty$:*

$$\mu_N \in \mathcal{T}_{q'}(\mathcal{E}_N), \quad \sup_{N\in\mathbb{N}} \langle\!\langle \mu^N, \overline{x}_0^{q'} \rangle\!\rangle < \infty \quad \Rightarrow \quad \sup_{N\in\mathbb{N}} \sup_{0\leq t\leq T\beta_N} \mathbf{E}^{\mu^N}[|\overline{x}_0^N(t)|^{q'}] < \infty. \tag{4.31}$$

Proof. Since

$$((\overline{x}_\xi(t))_{\xi\in\Lambda})_{t\geq 0} = ((\langle x_\xi(t), 1\rangle)_{\xi\in\Lambda})_{t\geq 0} \text{ and } ((\overline{x}_\xi^N(t))_{\xi\in\Lambda})_{t\geq 0} = ((\langle x_\xi^N(t), 1\rangle)_{\xi\in\Lambda})_{t\geq 0} \tag{4.32}$$

are the same processes as considered in Cox et al. (1995) the proposition follows from Lemma 2.2 therein.

Note that 2. and 4., corresponding to b) and d) of Cox et al. (1995) are slightly stronger, because the result is given for a fixed q but for all $2 < q' \leq q$. But in fact in Cox et al. (1995) the above result is proved. □

Using this result it is now easy to prove tightness for the finite systems at timescales of order $\mathcal{O}(\beta_N)$.

Proposition 4.5. *Let $t_N = \mathcal{O}(\beta_N)$. If $\mathcal{L}(X^N(0)) \in \mathcal{M}^N \quad (N \in \mathbb{N})$ then the sequence $(X^N(t_N))_{N\in\mathbb{N}}$ is tight and all weak limit points belong to $\mathcal{T}_2(\mathcal{E})$.*

Proof. This is an easy consequence of Proposition 4.4. First by the uniform boundedness of the second moments of the total mass process we know that $(\overline{x}_0^N(t_N))_{N\in\mathbb{N}}$ is tight and by Fatou's Lemma any weak limit point also has second moments. Because of the space-shift invariance $(\overline{X}^N(t_N))_{N\in\mathbb{N}}$ is also tight. To see this take $\epsilon, K > 0$ such that

$$\mathbf{P}[\overline{x}_\xi^N(t_N) \geq K] \leq \frac{1}{K^2}\mathbf{E}[(\overline{x}_\xi^N(t_N))^2] \leq \frac{C}{K^2} \leq \epsilon \quad (\xi \in \Lambda) \tag{4.33}$$

with C from Proposition 4.4. So for any $\xi \in \Lambda$ we obtain that $(\overline{x}_\xi(t_N))_{N\in\mathbb{N}}$ is tight. Then also $(\overline{X}^N(t_N))_{N\in\mathbb{N}}$ is tight by Proposition 3.2.4 of Ethier and Kurtz (1986).

Since for any compact set $K \subseteq \mathbb{R}$ the set

$$\{x \in \mathcal{M}_f(I) : \langle x, 1\rangle \in K\} \tag{4.34}$$

is compact in the weak topology of finite measures we obtain tightness of $(X^N(t_N))$. □

For our next proposition we need L^2-theory. This mostly carries over from Cox et al. (1995). Before we start we need a lemma.

Lemma 4.6. *For $\mu \in \mathcal{T}_2(\mathcal{E})$ there exists a random measure Θ such that*

$$\frac{1}{|\Lambda_N|}\sum_{\xi\in\Lambda^N}\langle x_\xi, f\rangle \xrightarrow{N\to\infty} \langle\Theta(X), f\rangle \qquad (f\in\mathcal{C}(I)) \tag{4.35}$$

μ-almost surely and in $L^2(\mu)$.

Proof. Define $\mu^f \in \mathcal{T}_2(\overline{\mathcal{E}})$ as the image measure of μ under the map

$$\langle .,f\rangle : \begin{cases}\mathcal{E} & \to \overline{\mathcal{E}},\\ X & \mapsto (\langle x_\xi, f\rangle)_{\xi\in\Lambda}.\end{cases} \tag{4.36}$$

The L^2-ergodic theorem applied to μ^f implies that for any measurable $f\in\mathcal{C}(I)$ there exists a random variable Θ^f such that

$$\frac{1}{|\Lambda_N|}\sum_{\xi\in\Lambda^N}\langle x_\xi, f\rangle \xrightarrow{N\to\infty} \Theta^f(X) \tag{4.37}$$

μ-almost surely and in $L^2(\mu)$. So it remains to check if Θ, defined by $\langle\Theta, f\rangle = \Theta^f$ is a measure. To see this take indicator functions $f_1, f_2, \ldots$. It is clear that $\Theta^{f_1+f_2} = \Theta^{f_1}+\Theta^{f_2}$. So by standard criteria of construction of measures it remains to check if for $f_n\downarrow 0$ it is $\inf_n \Theta^{f_n} = 0$ almost surely. But by monotone convergence and the ergodic theorem we have that

$$\langle\!\langle \mu, \inf_n \Theta^{f_n}\rangle\!\rangle = \inf_n \langle\!\langle \mu, \Theta^{f_n}\rangle\!\rangle = \inf_n \langle\!\langle \mu, \langle x_0, f\rangle\rangle\!\rangle = 0 \tag{4.38}$$

and we are done. □

Proposition 4.7. *Let $\nu^N := \mathcal{L}(X^N(0)) \in \mathcal{M}^N$ and $t_N\to\infty$, $\frac{t_N}{\beta_N}\to t\in[0,\infty)$. Assume one of the following:*

- $\frac{t_N}{N^2}\xrightarrow{N\to\infty}\infty$ *and* $\Theta^N(X^N(0)) \Rightarrow \theta \in \mathcal{M}_f(I)$,
- $t_N = \mathcal{O}(N^2)$ *and* $\nu^N \Rightarrow \nu \in \mathcal{M}_\theta$.

Furthermore let $\mu_N := \Phi_N\mathcal{L}(X^N(t_N))$ and assume $\mu_{N_k}\Rightarrow\mu$ for some subsequence. Let $\mathcal{L}(X) = \mu$. Then

$$\Theta^{N_k}(X^{N_k}(t_{N_k})) \Rightarrow \Theta(X). \tag{4.39}$$

Proof. We omit the dependence on the subsequence and assume $\mu_N\Rightarrow\mu$.

From general L^2-Theory (see for example Koopmans (1974)) for any $\mu\in\mathcal{T}_2(\mathcal{E})$ with intensity measure θ there exists a finite measure $\lambda^f\in\mathcal{M}_f([-\pi,\pi)^d)$, called the spectral measure of μ^f (recall this from (4.36)), such that

$$\langle\!\langle \mu, (\langle x_\xi, f\rangle - \langle\theta, f\rangle)(\langle x_\eta, f\rangle - \langle\theta, f\rangle)\rangle\!\rangle = \int_{[-\pi,\pi)^d}\exp(iu(\xi-\eta))\lambda^f(du). \tag{4.40}$$

Furthermore as the calculation on page 180 of Cox et al. (1995) reveals

$$\mu\in\mathcal{M}_\theta \quad\Longleftrightarrow\quad \mu\in\mathcal{T}_2(\mathcal{E}) \text{ with intensity } \theta,\ \lambda^f(\{0\}) = 0\ (f\in\mathcal{C}(I)). \tag{4.41}$$

We first prove that $\mu^f, \mu_1^f, \mu_2^f \ldots$ and the corresponding spectral measures $\lambda^f, \lambda_1^f, \lambda_2^f, \ldots$ fulfil the assumptions of Lemma 2.5 of Cox et al. (1995). We proceed as in the proof of Lemma 2.9 in Cox et al. (1995). As we assumed qth moments to be uniformly bounded for some $q>2$ (see Remark 4.1),

$$\sup_{N\in\mathbb{N}}\langle\!\langle \mu_N, \langle x_0, f\rangle^q\rangle\!\rangle < \infty, \tag{4.42}$$

we also have by weak convergence (see Lemma 3.9) and Proposition 4.4

$$\lim_{N\to\infty}\mathbf{E}[\langle x_0^N(t_N),f\rangle\langle x_\xi^N(t_N),f\rangle] = \lim_{N\to\infty}\langle\!\langle \mu_N,\langle x_0,f\rangle\langle x_\xi,f\rangle\rangle\!\rangle = \langle\!\langle \mu,\langle x_0,f\rangle\langle x_\xi,f\rangle\rangle\!\rangle \tag{4.43}$$

and

$$\lim_{N\to\infty}\mathbf{E}[\langle\Theta^N(X^N(0)),f\rangle^2] = \langle\theta,f\rangle^2. \tag{4.44}$$

When the process has intensity measure $\tilde{\theta}$, we have, as shown in Cox et al. (1995)

$$\begin{aligned}\lambda_N^f(\{0\}) &= \langle\!\langle \mu_N,(\langle\Theta^N(X),f\rangle-\langle\tilde{\theta},f\rangle)^2\rangle\!\rangle = \mathbf{E}[(\langle\Theta^N(X^N(t_N)),f\rangle-\langle\tilde{\theta},f\rangle)^2],\\ \lambda^f(\{0\}) &= \langle\!\langle \mu,(\langle\Theta(X),f\rangle-\langle\tilde{\theta},f\rangle)^2\rangle\!\rangle.\end{aligned} \tag{4.45}$$

So we have to prove

$$\lim_{N\to\infty}\mathbf{E}[\langle\Theta^N(X^N(t_N)),f\rangle^2] = \langle\!\langle \mu,\langle\Theta(X),f\rangle^2\rangle\!\rangle. \tag{4.46}$$

Because of space-shift invariance this is the same as

$$\lim_{N\to\infty}\mathbf{E}[\langle x_0^N(t_N),f\rangle\langle\Theta^N(X^N(t_N)),f\rangle] = \langle\!\langle \mu,\langle x_0,f\rangle\langle\Theta(X),f\rangle\rangle\!\rangle. \tag{4.47}$$

To prove this equation we make a moment calculation which is the same as equation (3.107) in Dawson and Greven (2003) for the finite case

$$\begin{aligned}\mathbf{E}[\langle x_0^N(t_N),f\rangle\langle x_\xi^N(t_N),f\rangle] &= \sum_{\xi',\eta'\in\Lambda^N} a_{t_N}^N(0,\xi')a_{t_N}^N(\xi,\eta')\mathbf{E}[\langle x_{\xi'}^N(0),f\rangle\langle x_{\eta'}^N(0),f\rangle]\\ &\quad+\int_0^{t_N}\hat{a}_{2s}^N(0,\xi)\mathbf{E}[h(\overline{x}_0^N(t_N-s))\langle x_0^N(t_N-s),f^2\rangle]ds.\end{aligned} \tag{4.48}$$

With this equation the proof of (4.47) is just analogous to the proof of Lemma 2.9 of Cox et al. (1995). We leave out the straight-forward details.

Let us now apply Lemma 2.5 of Cox et al. (1995). Write $F\nu$ for the image measure of ν under F. We have with the use of Lemma 4.6

$$\langle\Theta,f\rangle\mu_N = \Theta\mu_N^f \Rightarrow \Theta\mu^f = \langle\Theta,f\rangle\mu \tag{4.49}$$

So we have for $g\in\mathcal{C}_b(\mathbb{R})$

$$\lim_{n\to\infty}\langle\!\langle \mu_N,g(\langle\Theta,f\rangle)\rangle\!\rangle = \langle\!\langle \mu,g(\langle\Theta,f\rangle)\rangle\!\rangle. \tag{4.50}$$

Functions $F(x)=g(\langle x,f\rangle)$ $(f\in\mathcal{C}(I),g\in\mathcal{C}_b(\mathbb{R}),\ x\in\mathcal{M}_f(I))$ are convergence determining in $\mathcal{P}(\mathcal{M}_F(I))$ (i.e. these functions can guarantee weak convergence, see Dawson (1993), p. 35f. and Ethier and Kurtz (1986), p.113f.). So the result follows. □

4.4. *Step 3: Coupled processes.* This section is the one where the proof differs mostly from the one given in Cox et al. (1995). There the following program was carried out:

First prove for the infinite system that if initial states converge then also the states after a time $t_N\uparrow\infty$ converge, i.e.

$$\mu_N\Rightarrow\mu \qquad \Rightarrow \qquad (\Phi_N\mu_N)S(t_N)-\mu S(t_N)\Rightarrow 0. \tag{4.51}$$

Secondly identify a sequence $(l_N)_{N\in\mathbb{N}}$ with $l_N = o(\beta_N)$ for which the infinite and finite systems stay comparable, i.e.

$$(\Phi_N \mu^N)S(l_N) - \Phi(\mu^N S^N(l_N)) \Rightarrow 0. \tag{4.52}$$

But here we have to take another approach because we cannot prove the first point of the program. That is because in our coupling the expected distance between the two processes can increase. Instead we prove the existence of a universal time scale for which both statements hold. Exactly we prove that there is a time p_N having various properties. If we have $\mu^N := \mathcal{L}(X^N(t\beta_N - p_N))$ and μ is a weak limit point of a subsequence μ^{N_k} of μ^N then both two infinite systems converge after time p_N, i.e.

$$(\Phi_{N_k}\mu^{N_k})S(p_{N_k}) - \mu S(p_{N_k}) \Rightarrow 0 \tag{4.53}$$

and an infinite and a finite systems converge after time p_N, i.e.

$$(\Phi_N \mu^N)S(p_N) - \Phi_N(\mu^N S^N(p_N)) \Rightarrow 0. \tag{4.54}$$

These two statements are given in Propositions 4.12 and 4.13.

But before coming to these two main propositions we have to prove some lemmata and derive some general comparison estimates. We therefore use our couplings of Chapter 3, both the coupling of two infinite systems and the coupling of a finite and an infinite system.

To carry out this program it will be necessary to introduce two more time scales (l_N and m_N) in order to define p_N and prove Propositions 4.12 and 4.13. We must do this because of technical properties of our system that are in the way when proving the finite system scheme in the same manner as in the proof for the total masses from Cox et al. (1995). Consider the coupling defined in Chapter 3. We calculated the change of the expected Wasserstein distance between our two components with the help of a generator calculation. We came up to (3.59) and saw that the expected distance can increase when time evolves. Now we have to see if this increase is still small when the system size grows. Therefore we first prove the existence of a time scale m_N for which (4.52) holds even if we make m_N smaller. To derive this time scale we need the time scale l_N from Cox et al. (1995).

This was a time scale for which (see Cox et al. (1995), Proposition 2.4)

$$(\Phi_N \overline{\mu}^N)\overline{S}(l_N) - \phi_N(\overline{\mu}^N \overline{S}^N(l_N)) \Rightarrow 0. \tag{4.55}$$

So it plays the same role for the process of total masses than the time scale m_N for the multitype system. Then by making m_N possibly smaller we derive the desired times scale p_N.

The relation between these time scales will be

$$\beta_N > l_N \geq m_N \geq p_N \qquad \text{and} \qquad p_N \to \infty. \tag{4.56}$$

The time scales β_N and l_N are now already given, $\beta_N = (2N)^d$ and l_N as constructed in Cox et al. (1995).

Recall $\mathcal{T}_2(\mathcal{E}^N)$ and $\mathcal{T}_2(\mathcal{E}\times\mathcal{E})$ from (3.21). It is also important here the fact from Corollary 3.6 that $\Psi(\cdot,\cdot)$ and $\Psi^N(\cdot,\cdot)$ are measurable and $\Psi(\mu^1,\mu^2)$ and $\Psi^N(\mu,\mu^N)$ are space-shift invariant.

We have to make some definitions.

Definition 4.8. *1. Define for $C < \infty$*

$$\mathcal{V}_C^N := \big\{(\mu,\mu^N) \in \mathcal{T}_2(\mathcal{E}) \times \mathcal{T}_2(\mathcal{E}^N) : \Phi_N \mu^N = \mu, \langle\!\langle \mu, \overline{y}_0 \rangle\!\rangle \leq C\big\}. \tag{4.57}$$

where $y_0 := \pi_0 Y$ for $Y \in \mathcal{E}$.

2. Define for $\epsilon > 0$

$$\mathcal{W}_\epsilon := \big\{(\mu_1,\mu_2) \in \mathcal{T}_2(\mathcal{E}) \times \mathcal{T}_2(\mathcal{E}) : \langle\!\langle \Psi(\mu_1,\mu_2), |\overline{y}_0^1 - \overline{y}_0^2| \rangle\!\rangle < \epsilon\big\} \tag{4.58}$$

and

$$\mathcal{W}_\epsilon^N := \big\{(\mu,\mu^N) \in \mathcal{T}_2(\mathcal{E}) \times \mathcal{T}_2(\mathcal{E}^N) : \langle\!\langle \Psi^N(\mu,\mu^N), |\overline{y}_0 - \overline{y}_0^N| \rangle\!\rangle < \epsilon\big\}. \tag{4.59}$$

where $y_0^i := \pi_0 Y^i$ for $(Y^1, Y^2) \in \mathcal{E} \times \mathcal{E}$ and $y_0 := \pi_0 Y$, $y_0^N := \pi_0 Y^N$ for $(Y, Y^N) \in \mathcal{E} \times \mathcal{E}^N$. For simplicity we write

$$\begin{aligned} \mathbf{E}^{(\mu_1,\mu_2)}[.] &:= \mathbf{E}^{\Psi(\mu_1,\mu_2)}[.] \quad \text{for } (\mu_1,\mu_2) \in \mathcal{T}_2(\mathcal{E}) \times \mathcal{T}_2(\mathcal{E}), \\ \mathbf{E}^{(\mu,\mu^N)}[.] &:= \mathbf{E}^{\Psi^N(\mu,\mu^N)}[.] \quad \text{for } (\mu,\mu^N) \in \mathcal{T}_2(\mathcal{E}) \times \mathcal{T}_2(\mathcal{E}^N). \end{aligned} \tag{4.60}$$

The main difficulty we have to deal with is that the Wasserstein distance can increase over time. Nevertheless we can control the way it increases. This is done in the next lemmata. It is helpful that we already know a lot about the process of total masses.

Lemma 4.9. *1. Let $\epsilon_N \downarrow 0$. If $(\mu,\mu_N) \in \mathcal{W}_{\epsilon_N}$ then there is a sequence $\delta_N \downarrow 0$ depending only on ϵ_N such that*

$$\begin{aligned} \sup_{0 \leq t < \infty} \int_I \mathbf{E}^{(\mu,\mu_N)}[|h(\overline{y}_0^1(t)) - h(\overline{y}_0^2(t))| y_0^1([0,u])(t), \\ y_0^1([0,u])(t) = y_0^2([0,u])(t)] du \leq \delta_N. \end{aligned} \tag{4.61}$$

2. If $(\mu,\mu^N) \in \mathcal{V}_C^N$ then there is a sequence $\delta_N \downarrow 0$ depending only on C such that

$$\begin{aligned} \sup_{0 \leq t \leq l_N} \int_I \mathbf{E}^{(\mu,\mu^N)}[|h(\overline{y}_0(t)) - h(\overline{y}_0^N(t))| y_0([0,u])(t), \\ y_0([0,u])(t) = y_0^N([0,u])(t)] du \leq \delta_N. \end{aligned} \tag{4.62}$$

Proof. We prove both statements simultaneously and omit the dependencies on t. For the projections on the first and second component of one site ξ we use in both cases y_ξ and y_ξ^N respectively. Since h is locally Lipschitz continuous and $o(x)$ there is for any $L < \infty$ a $M < \infty$ and functions h_1 and h_2 such that

$$h(x) = h_1(x) + h_2(x), \tag{4.63}$$

where $h_1 \leq h$ is globally Lipschitz with constant L and $h_2 \leq h(x)1_{\{x \geq M\}}$. Furthermore M can be chosen such that $M \to \infty$ as $L \to \infty$. We define

$$\tau_M := \sup_{x \geq M} \frac{h(x)}{x} \xrightarrow{M\to\infty} 0. \tag{4.64}$$

Furthermore by the uniform existence of second moments by Proposition 4.4 we know that there exists C'' with

$$\sup_{N\in\mathbb{N}} \sup_{(\mu,\mu^N)\in\mathcal{W}_{\epsilon_N}} \sup_{0 \leq t < l_N} \mathbf{E}^{(\mu,\mu^N)}[\overline{y}_0^2 + (\overline{y}_0^N)^2 + \overline{y}_0 + \overline{y}_0^N] \leq C'' < \infty, \tag{4.65}$$

$$\sup_{N\in\mathbb{N}} \sup_{(\mu,\mu^N)\in\mathcal{V}_C^N} \sup_{0 \leq t \leq l_N} \mathbf{E}^{(\mu,\mu^N)}[\overline{y}_0^2 + (\overline{y}_0^N)^2 + \overline{y}_0 + \overline{y}_0^N] \leq C'' < \infty. \tag{4.66}$$

For 1. we take $(\mu,\mu^N)\in\mathcal{W}_{\epsilon_N}$ and calculate with $\epsilon'_N := \sqrt{\epsilon_N}$

$$\begin{aligned}\mathbf{P}^{(\mu,\mu^N)}[|\overline{y}_0-\overline{y}_0^N|\geq\epsilon'_N] &\leq \frac{1}{\epsilon'_N}\mathbf{E}^{(\mu,\mu^N)}[|\overline{y}_0-\overline{y}_0^N|] \\ &\leq \frac{1}{\epsilon'_N}\langle\!\langle\Psi(\mu,\mu^N),|\overline{y}_0-\overline{y}_0^N|\rangle\!\rangle \leq \epsilon'_N\downarrow 0.\end{aligned} \tag{4.67}$$

That is because by Proposition 2.3 of Cox et al. (1995) the expected difference for the total masses is non-increasing over time and therefore never greater than at time 0.

For 2. with $(\mu,\mu^N)\in\mathcal{V}_C^N$ we define

$$\epsilon_N := \sup_{(\mu,\mu^N)\in\mathcal{V}_C^N}\ \sup_{0\leq t\leq l_N}\mathbf{E}^{(\mu,\mu^N)}[|\overline{y}_0-\overline{y}_0^N|]\downarrow 0 \tag{4.68}$$

which was proved in Cox et al. (1995), Proposition 2.4.a. Moreover, $\epsilon'_N := \sqrt{\epsilon_N}$. We obtain

$$\begin{aligned}\mathbf{P}^{(\mu,\mu^N)}[|\overline{y}_0-\overline{y}_0^N|\geq\epsilon'_N] &\leq \frac{1}{\epsilon'_N}\mathbf{E}^{(\mu,\mu^N)}[|\overline{y}_0-\overline{y}_0^N|] \\ &\leq \frac{1}{\epsilon'_N}\sup_{0\leq t\leq l_N}\mathbf{E}^{(\mu,\mu^N)}[|\overline{y}_0-\overline{y}_0^N|]\leq\epsilon'_N\downarrow 0\end{aligned} \tag{4.69}$$

as before.

Define $A_N := \{|\overline{y}_0-\overline{y}_0^N|\geq\epsilon'_N\}$ and $c_N := \frac{1}{\sqrt{\epsilon'_N}}$. Then writing $\mathbf{E}$ for the expectation with respect to processes that started in $(\mu,\mu^N)\in\mathcal{W}_\epsilon$ and $(\mu,\mu^N)\in\mathcal{V}_C^N$ respectively

$$\begin{aligned}&\int_I\mathbf{E}[|h(\overline{y}_0)-h(\overline{y}_0^N)|y_0([0,u]),y_0([0,u])=y_0^N([0,u])]du \\ &\quad\leq\mathbf{E}[|h(\overline{y}_0)-h(\overline{y}_0^N)|\overline{y}_0\wedge\overline{y}_0^N] \\ &\quad\leq\mathbf{E}[|h_1(\overline{y}_0)-h_1(\overline{y}_0^N)|\overline{y}_0\wedge\overline{y}_0^N,|\overline{y}_0-\overline{y}_0^N|<\epsilon'_N] \\ &\qquad+\mathbf{E}[|h_1(\overline{y}_0)-h_1(\overline{y}_0^N)|\overline{y}_0\wedge\overline{y}_0^N,|\overline{y}_0-\overline{y}_0^N|\geq\epsilon'_N,\overline{y}_0\wedge\overline{y}_0^N<c_N] \\ &\qquad+\mathbf{E}[|h_1(\overline{y}_0)-h_1(\overline{y}_0^N)|\overline{y}_0\wedge\overline{y}_0^N,|\overline{y}_0-\overline{y}_0^N|\geq\epsilon'_N,\overline{y}_0\wedge\overline{y}_0^N\geq c_N] \\ &\qquad+\mathbf{E}[h(\overline{y}_0)\overline{y}_0,\overline{y}_0\geq M]+\mathbf{E}[h(\overline{y}_0^N)\overline{y}_0^N,\overline{y}_0^N\geq M] \\ &\quad\leq L\epsilon'_N\mathbf{E}[\overline{y}_0\wedge\overline{y}_0^N]+Lc_N\mathbf{E}[|\overline{y}_0-\overline{y}_0^N|]+\mathbf{E}[h_1(\overline{y}_0)\overline{y}_0,\overline{y}_0\geq c_N] \\ &\quad+\mathbf{E}[h_1(\overline{y}_0^N)\overline{y}_0^N,\overline{y}_0^N\geq c_N]\quad+\tau_M\left(\mathbf{E}[\overline{y}_0^2]+\mathbf{E}[(\overline{y}_0^N)^2]\right) \\ &\quad\leq L\epsilon'_NC''+Lc_N\epsilon_N+\tau_{c_N}\left(\mathbf{E}[\overline{y}_0^2]+\mathbf{E}[(\overline{y}_0^N)^2]\right)+\tau_MC'' \\ &\quad\leq L(C''+1)\epsilon'_N+C''(\tau_{c_N}+\tau_M).\end{aligned} \tag{4.70}$$

Therefore if we define

$$L_N := \frac{1}{\sqrt{\epsilon'_N}}\to\infty \tag{4.71}$$

and take the corresponding sequence $M_N\to\infty$ we obtain our result for

$$\delta_N := (C''+1)\sqrt{\epsilon'_N}+C''(\tau_{c_N}+\tau_{M_N})\xrightarrow{N\to\infty}0. \tag{4.72}$$

□

Lemma 4.10. *Let $Y^N(\cdot)$ be a solution of the $G^{N,d}_{coup}$ martingale problem of Proposition 3.13. Then there is a sequence $m_N \leq l_N$ with*

$$\sup_{(\mu,\mu^N)\in\mathcal{V}_C^N}\ \sup_{0\leq t\leq m_N} \mathbf{E}^{(\mu,\mu^N)}[\rho_W(y_0(t), y_0^N(t))] \xrightarrow{N\to\infty} 0. \tag{4.73}$$

Proof. For $(\mu,\mu^N)\in\mathcal{T}_2(\mathcal{E})\times\mathcal{T}_2(\mathcal{E}^N)$ we have from Proposition 3.14

$$\frac{d}{dt}K^{N,d}(t) = \mathbf{E}[G^{N,d}_{\text{coup}}(k\circ\pi_0)(Y^N(t))] = \dot{K}^{N,p}_{\text{br}}(t) + \dot{K}^{N,p}_{\text{mig}}(t). \tag{4.74}$$

For the branching term we have just shown

$$\dot{K}^{N,p}_{\text{br}}(t) \leq \delta_N \to 0 \tag{4.75}$$

with δ_N from Lemma 4.9.

The migration terms yield

$$\dot{K}^{N,p}_{\text{mig}}(t) \leq \sum_{\xi\notin\Lambda} 2a(0,\xi)\mathbf{E}[\overline{y}_\xi] + \sum_{\xi\in\Lambda^N} 2(a^N(0,\xi) - a(0,\xi))\mathbf{E}[\overline{y}_0^N]. \tag{4.76}$$

By Proposition 4.4 there exists some constant $C' < \infty$ depending only on C such that

$$\sup_{t\geq 0}\mathbf{E}^\mu[\overline{y}_0(t)] \leq C', \qquad \sup_{0\leq t\leq T\beta_N}\mathbf{E}[\overline{y}_0^N(t)] \leq C' \tag{4.77}$$

and therefore

$$\dot{K}^{N,p}_{\text{mig}}(t) \leq 4C'\sum_{\xi\notin\Lambda^N} a(0,\xi) =: \tau_N \to 0. \tag{4.78}$$

As we have assumed $K^N(0) = 0$ we obtain

$$K^{N,d}(t) = K^{N,d}(0) + \int_0^t \dot{K}^{N,d}(s)ds \leq t(\tau_N + \delta_N), \tag{4.79}$$

so taking

$$m_N := (\tau_N + \delta_N)^{-\frac{1}{2}} \wedge l_N \to \infty \tag{4.80}$$

we arrive at (4.73). □

Next we want to prove comparison estimates for the coupling of two processes that are close together at the beginning. We deal with two cases, the coupling of two infinite systems and the coupling of an infinite with a finite system.

Lemma 4.11. *Take m_N from Lemma 4.10 and l_N as in* (4.55).

1. *For any sequence $\epsilon_N \downarrow 0$ there is a sequence $\epsilon'_N \downarrow 0$ depending only on ϵ_N and l_N such that for any $p_N \leq l_N$*

$$(\mu,\mu^N)\in\mathcal{W}^N_{\epsilon_N} \quad\Rightarrow\quad (\mu S(p_N), \mu^N S^N(p_N)) \in \mathcal{W}^N_{\epsilon'_N}. \tag{4.81}$$

2. *Let $\epsilon_N \downarrow 0$. There is a sequence $p_N \leq m_N$, $p_N \uparrow \infty$ depending only on ϵ_N and m_N such that*

$$(\mu,\mu_N)\in\mathcal{W}_{\epsilon_N} \text{ with } \mu_N \Rightarrow \mu \quad\Rightarrow\quad \sup_{0\leq t\leq p_N}\mathbf{E}^{(\mu,\mu_N)}[\rho_W(y_0^1(t), y_0^2(t))] \to 0. \tag{4.82}$$

Proof. 1. The proof of Proposition 2.4 in Cox et al. (1995) reveals that l_N is chosen such that for
$M := \sup_{t\geq 0,\xi\in\Lambda} \mathbf{E}[\overline{x}_\xi(t)] < \infty$ we have

$$\begin{aligned}\mathbf{E}^{(\mu,\mu^N)}&[|\overline{y}_0(p_N) - \overline{y}_0^N(p_N)|] \\ &\leq \mathbf{E}^{(\mu,\mu^N)}[|\overline{y}_0(0) - \overline{y}_0^N(0)|] + 2M \int_0^{p_N} \sum_{j\notin\Lambda_{N/2}} (a_s(0,j) + a(0,j))ds \\ &\leq \epsilon_N + 2M \int_0^{l_N} \sum_{j\notin\Lambda_{N/2}} (a_s(0,j) + a(0,j))ds =: \epsilon'_N \downarrow 0\end{aligned} \tag{4.83}$$

2. Take δ_N from Lemma 4.9.1 only depending on ϵ_N. Then we have from Proposition 3.10

$$\begin{aligned}&\mathbf{E}^{(\mu,\mu_N)}[\rho_W(y_0^1(t), y_0^2(t))] \leq \mathbf{E}^{(\mu,\mu^N)}[\rho_W(y_0^1(0), y_0^2(0))] \\ &+ \int_0^t \int_I \mathbf{E}^{(\mu,\mu_N)}[|h(\overline{y}_0^1(s)) - h(\overline{y}_0^2(s))| y_0^1([0,u])(s), y_0^1([0,u])(s) = y_0^2([0,u])(s)]duds \\ &\quad \leq \langle\!\langle \Psi(\mu,\mu^N), \rho_W(y_0^1, y_0^2)\rangle\!\rangle + t\delta_N \to 0.\end{aligned} \tag{4.84}$$

Now take $p_N := \delta_N^{-1/2} \wedge m_N$.

□

We began with introducing four time scales, β_N, l_N, m_N and p_N. Let a sequence ϵ_N be given and take ϵ'_N as in the last lemma. Then the last lemma says that we only have to deal with two time scales, β_N and p_N from now on where p_N is defined according to ϵ'_N. We will not need l_N and m_N any more. We only needed these two time scales to define our p_N. Therefore our plan of the proof for Propositions 4.12 and 4.13 is essentially the same as in Cox and Greven (1994b). For these proofs we will consider the time scale p_N as given by the last lemmata.

Proposition 4.12. *There is a sequence $p_N \leq m_N$, $p_N \to \infty$ such that with $\mu^N := \mathcal{L}(X^N(t\beta_N - p_N))$ and $\mu^{N_k} \Rightarrow \mu$ for some subsequence then*

$$\mathbf{E}^{(\mu,\Phi_{N_k}\mu^{N_k})}[\rho(y_0^1(p_{N_k}), y_0^2(p_{N_k}))] \to 0, \tag{4.85}$$

i.e.

$$(\Phi_{N_k}\mu^{N_k})S(p_{N_k}) - \mu S(p_{N_k}) \Rightarrow 0. \tag{4.86}$$

Proof. Define for any $t_N = o(\beta_N)$

$$\mu^{t_N} := \mathcal{L}(X^N(t\beta_N - t_N)). \tag{4.87}$$

We make the special choice of $t_N = m_N$ with m_N from Proposition 4.10. From Theorem 2 of Cox et al. (1995) we know that

$$\overline{\mu}^{m_N} \Rightarrow \overline{\mu} := \int \overline{\nu}_\theta \mathbf{P}[Z(t) \in d\theta]. \tag{4.88}$$

Therefore

$$\epsilon_N := \langle\!\langle \Psi^N(\mu,\mu^{m_N}), |\overline{y}_0 - \overline{y}_0^N| \rangle\!\rangle \downarrow 0 \tag{4.89}$$

and so $(\mu, \mu^{m_N}) \in \mathcal{W}^N_{\epsilon_N}$. From Lemma 4.11.1 take the corresponding ϵ'_N and from 4.11.2 the corresponding p_N. Then

$$(\mu S(m_N - p_N), \mu^{m_N} S^N(m_N - p_N)) = (\mu S(m_N - p_N), \mu^{p_N}) \in \mathcal{W}^N_{\epsilon'_N} \tag{4.90}$$

and as $\overline{\mu} S(m_N - p_N) = \overline{\mu}$ also $(\mu, \mu^{p_N}) \in \mathcal{W}^N_{\epsilon'_N}$, so $(\mu, \Phi_N \mu^{p_N}) \in \mathcal{W}_{\epsilon'_N}$. As we chose p_N corresponding to ϵ'_N we are done by Lemma 4.11.2. □

Proposition 4.13. *Take p_N from Proposition 4.12 and $\mu^N := \mathcal{L}(X^N(t\beta_N - p_N))$. Then*

$$\mathbf{E}^{(\Phi_N \mu^N, \mu^N)}[\rho(y_0(p_N), y_0^N(p_N))] \to 0, \tag{4.91}$$

i.e.

$$(\Phi_N \mu^N) S(p_N) - \Phi^N(\mu^N S^N(p_N)) \Rightarrow 0. \tag{4.92}$$

Proof. This is a direct consequence of Lemma 4.10. □

4.5. *Step 4: Independence of subsequences, Generalisation to empirical measures.* First of all we prove that the map $\theta \mapsto \nu_\theta$ is continuous as we will need that result in this step.

Lemma 4.14. *The map $\theta \mapsto \nu_\theta$ is continuous.*

Proof. Take a sequence $(\theta_N)_{N\in\mathbb{N}}$ with $\theta_N \to \theta$ (in the weak topology of $\mathcal{M}_f(I)$). As second moments exist uniformly for $(\nu_{\theta_N})_{N\in\mathbb{N}}$ the sequence $(\nu_{\theta_N})_{N\in\mathbb{N}}$ is tight and so there is a convergent subsequence $(\nu_{\theta_{N_k}})_{k\in\mathbb{N}}$ with $\nu_{\theta_{N_k}} \Rightarrow \nu$.

Furthermore we have by an application of Theorem 4.2 of Kallenberg (1983) that

$$\langle\!\langle \nu, x_0 \rangle\!\rangle = \lim_{k\to\infty} \langle\!\langle \nu_{\theta_{N_k}}, x_0 \rangle\!\rangle = \lim_{k\to\infty} \theta_{N_k} = \theta. \tag{4.93}$$

Recall the definition of $\mathcal{T}_1$, $\mathcal{M}_\theta$ and $\mathcal{I}$ from (1.28)-(1.30). As $\nu_{\theta_{N_k}} \Rightarrow \nu$ there is a sequence ϵ_{N_k} with $(\nu, \nu_{\theta_{N_k}}) \in \mathcal{W}_{\epsilon_{N_k}}$. By Lemma 4.11.2 we have for any t

$$\mathbf{E}^{(\nu, \nu_{\theta_{N_k}})}[\rho_W(y_0^1(t), y_0^2(t))] \to 0, \tag{4.94}$$

so by the space-shift invariance of the measures $\nu, \nu_{\theta_{N_1}}, \nu_{\theta_{N_2}}, \ldots$

$$\nu_{\theta_{N_k}} S(t) \Rightarrow \nu S(t), \tag{4.95}$$

but then by the stationarity of ν_θ

$$\nu = \lim_{k\to\infty} \nu_{\theta_{N_k}} = \lim_{k\to\infty} \nu_{\theta_{N_k}} S(t) = \nu S(t), \tag{4.96}$$

so $\nu \in \mathcal{I}$ and also $\nu \in \mathcal{T}_1$ as a weak limit of space-shift invariant measures. It remains to show that ν is an extremal point of $\mathcal{I} \cap \mathcal{T}^1$ since then it must be according to Theorem 1.4 an element of $\{\nu_\theta : \theta \in \mathcal{M}_f(I)\}$. As the intensity is θ we obtain $\nu = \nu_\theta$ and so $\nu_{\theta_n} \Rightarrow \nu_\theta$.

First we make a moment calculation.

We know that

$$\lim_{t\to\infty} \mathbf{Cov}^{\delta_{\underline{\theta}}}[\langle x_\xi(t), f\rangle, \langle x_\eta(t), f\rangle] = \mathbf{Cov}^{\nu_\theta}[\langle x_\xi, f\rangle, \langle x_\eta, f\rangle], \tag{4.97}$$

and so by Lemma 3.9 for any non-negative $f \in \mathcal{C}(I)$ by (3.107) of Dawson and Greven (2003) and $\theta \in \mathcal{M}_f(I)$.

$$\begin{aligned}\mathbf{Cov}^{\nu_\theta}[\langle x_\xi, f\rangle, \langle x_\eta, f\rangle] &= \lim_{t\to\infty} \mathbf{Cov}^{\delta_{\underline{\theta}}}[\langle x_\xi(t), f\rangle, \langle x_\eta(t), f\rangle] \\ &= \lim_{t\to\infty} \int_0^t \hat{a}_{2(t-s)}(\xi,\eta)\mathbf{E}^{\delta_{\underline{\theta}}}[h(\overline{x}_0(t))\langle x_0(t), f^2\rangle]ds\end{aligned} \tag{4.98}$$

Especially

$$\mathbf{Cov}^{\nu_\theta}[\langle x_\xi, f\rangle, \langle x_\eta, f\rangle] \geq 0 \tag{4.99}$$

and by uniform existence of second moments by Proposition 4.4 and $h(x) \leq C' + C''x$ for constants $C', C'' < \infty$

$$\begin{aligned}\mathbf{Cov}^{\nu_\theta}[\langle x_\xi, f\rangle, \langle x_\eta, f\rangle] &\leq \underbrace{\sup_{t\geq 0} \mathbf{E}^{\delta_{\underline{\theta}}}[||f||_\infty^2(C'\overline{x}_0(t) + C''\overline{x}_0(t)^2)]}_{=:M_\theta<\infty} \int_0^\infty \hat{a}_s(\xi,\eta)ds \\ &= M_\theta \hat{A}(\xi,\eta) \xrightarrow{\xi-\eta\to\infty} 0\end{aligned} \tag{4.100}$$

for transient $\hat{a}(\cdot,\cdot)$.

Equipped with these calculations we assume

$$\nu = \int_{\mathcal{M}_f(I)} \nu_\theta \mathbf{P}[Z \in d\theta] \tag{4.101}$$

for some random variable Z. As we want to show that ν is an extreme point of $\mathcal{I} \cap \mathcal{T}_1$ we have to show that Z is almost surely constant. First we calculate

$$\begin{aligned}\mathbf{Cov}^{\nu}[\langle x_\xi, f\rangle, \langle x_\eta, f\rangle] &= \langle\!\langle \nu, \langle x_\xi, f\rangle\langle x_\eta, f\rangle\rangle\!\rangle - \langle\!\langle \nu, \langle x_\xi, f\rangle\rangle\!\rangle\langle\!\langle \nu, \langle x_\eta, f\rangle\rangle\!\rangle \\ &= \lim_{N\to\infty} \langle\!\langle \nu, \langle x_\xi, f\rangle\langle x_\eta, f\rangle \wedge N\rangle\!\rangle - \langle\!\langle \nu, \langle x_\xi, f\rangle\rangle\!\rangle\langle\!\langle \nu, \langle x_\eta, f\rangle\rangle\!\rangle \\ &= \lim_{N\to\infty}\lim_{k\to\infty} \langle\!\langle \nu_{\theta_{N_k}}, \langle x_\xi, f\rangle\langle x_\eta, f\rangle \wedge N\rangle\!\rangle - \langle\!\langle \nu_{\theta_{N_k}}, \langle x_\xi, f\rangle\rangle\!\rangle\langle\!\langle \nu, \langle x_\eta, f\rangle\rangle\!\rangle \\ &\leq \lim_{k\to\infty} \mathbf{Cov}^{\nu_{\theta_{N_k}}}[\langle x_\xi, f\rangle, \langle x_\eta, f\rangle] \leq \hat{A}(\xi,\eta) \lim_{k\to\infty} M_{\theta_{N_k}} \xrightarrow{\xi-\eta\to\infty} 0\end{aligned} \tag{4.102}$$

as $M_{\theta_{N_k}}$ is bounded. On the other hand we obtain

$$\begin{aligned}&\mathbf{Cov}^{\nu}[\langle x_\xi, f\rangle, \langle x_\eta, f\rangle] \\ &= \int_{\mathcal{M}_f(I)} \langle\!\langle \nu_\theta, \langle x_\xi, f\rangle\langle x_\eta, f\rangle\rangle\!\rangle \mathbf{P}[Z \in d\theta] - \Big(\int_{\mathcal{M}_f(I)} \langle\!\langle \nu_\theta, \langle x_0, f\rangle\rangle\!\rangle \mathbf{P}[Z \in d\theta]\Big)^2 \\ &\geq \int_{\mathcal{M}_f(I)} \langle\theta, f\rangle^2 \mathbf{P}[Z \in d\theta] - \Big(\int_{\mathcal{M}_f(I)} \langle\theta, f\rangle \mathbf{P}[Z \in d\theta]\Big)^2 \\ &= \int_{\mathcal{M}_f(I)} \Big(\langle\theta, f\rangle - \int_{\mathcal{M}_f(I)} \langle\theta, f\rangle \mathbf{P}[Z \in d\theta]\Big)^2 \mathbf{P}[Z \in d\theta] = \mathrm{Var}[\langle Z, f\rangle]\end{aligned} \tag{4.103}$$

independent of ξ and η. This is only possible if

$$\mathrm{Var}[\langle Z, f\rangle] = 0 \qquad (f \in \mathcal{C}(I), f \geq 0). \tag{4.104}$$

and that means that there is a $\theta \in \mathcal{M}_f(I)$ such that $Z = \theta$ almost surely. So ν is extremal and we are done. □

Now we will prove the independence of $(Z(t))_{t\geq 0}$ of the choice of $(N_k)_{k\in\mathbb{N}}$ in Step 1. Before we can do this we need a different formulation for the branching term in the pregenerator of the process Z from (2.8).

Lemma 4.15. *For the branching term in* (2.8)

$$\frac{1}{\theta}\langle\!\langle \nu_{\overline{\theta}}, h(\overline{x}_0)\overline{x}_0\rangle\!\rangle\theta = \langle\!\langle \nu_\theta, h(\overline{x}_0)x_0\rangle\!\rangle. \tag{4.105}$$

Proof. We need equation (1.90) from Dawson and Greven (2003). We prove our statement by separating the effects of the total masses and relative weights. Let $A \in \mathcal{B}(I)$. Then by Theorem 1.4

$$\begin{aligned}\langle\!\langle \nu_\theta, h(\overline{x}_0)x_0\rangle\!\rangle(A) &= \langle\!\langle \nu_\theta, h(\overline{x}_0)\langle x_0, A\rangle\rangle\!\rangle \\ &= \lim_{t\to\infty} \mathbf{E}^{\delta_{\underline{\theta}}}[h(\overline{x}_0(t))\langle x_0(t), A\rangle] \\ &= \lim_{t\to\infty} \mathbf{E}^{\delta_{\underline{\theta}}}\big[h(\overline{x}_0(t))\overline{x}_0(t)\mathbf{E}^{\delta_{\underline{\theta}}}[\langle \hat{x}_0(t), A\rangle | (\overline{X}_s)_{s\geq 0}]\big] \\ &= \lim_{t\to\infty} \mathbf{E}^{\delta_{\underline{\theta}}}\Big[h(\overline{x}_0(t))\overline{x}_0(t)\sum_{\eta\in\Lambda} A_t^{\overline{X}}(0,\eta)\langle\hat{\theta}, A\rangle\Big] \\ &= \langle\hat{\theta}, A\rangle\langle\!\langle \nu_\theta, h(\overline{x}_0)\overline{x}_0\rangle\!\rangle = \frac{1}{\theta}\langle\!\langle \nu_{\overline{\theta}}, h(\overline{x}_0)\overline{x}_0\rangle\!\rangle\langle\theta, A\rangle\end{aligned} \tag{4.106}$$

as for the last equality when only considering total masses ν_θ becomes $\nu_{\overline{\theta}}$. □

The next proposition gives the desired independence of $(Z(t))_{t\geq 0}$ of the choice of $(N_k)_{k\in\mathbb{N}}$ in Step 1.

Proposition 4.16. *The process $(\breve{Z}(t))_{t\geq 0}$ solves the (θ, G_Z) local martingale problem for θ from the assumptions of Theorem* A *and G_Z given in* (2.8)*. Especially as this local martingale problem is well posed $(\breve{Z}(t))_{t\geq 0}$ is independent of the subsequence N_k.*

4.5.1. *Remark.* Let $(\breve{Z}(t))_{t\geq 0}$ be any accumulation point of $(Z^N(t))_{t\geq 0}$. As both $(Z^N(t))_{t\geq 0}$ and $X^N(t\beta_N - p_N)$ (with p_N from Step 3) are tight sequences there is a subsequence $\tilde{N}_k$ with

$$(Z^{\tilde{N}_k}(t))_{t\geq 0} \Rightarrow (\breve{Z}(t))_{t\geq 0}, \quad \mathcal{L}(X^{\tilde{N}_k}(t\beta_{\tilde{N}_k} - p_{\tilde{N}_k})) \Rightarrow \mu = \int_{\mathcal{M}_f(I)} \mu_\theta \mathbf{P}[\breve{Z}(t) \in d\theta] \tag{4.107}$$

with space-shift ergodic distributions μ_θ with intensity θ.

Then we will prove using empirical measures

$$\mathbf{E}\Big[\Big|[\langle Z^{\tilde{N}_k}, f\rangle](t) - \int_0^t \langle\!\langle \nu_{Z^{\tilde{N}_k}(s)}, h(\overline{x_0})\langle x_0, f^2\rangle\rangle\!\rangle ds\Big|\Big] \xrightarrow{k\to\infty} 0 \tag{4.108}$$

for any $f \in \mathcal{C}(I)$. As we have just proved that the map $\theta \mapsto \nu_\theta$ is continuous the process $(\langle \breve{Z}(t), f\rangle)_{t\geq 0}$ must then be a square integrable martingale with increasing process (recall Lemma 4.15)

$$\langle\!\langle \nu_{\breve{Z}(t)}, h(\overline{x}_0)\langle x_0, f^2\rangle\rangle\!\rangle = \frac{1}{\theta}\langle\!\langle \nu_{\overline{\breve{Z}(t)}}, h(\overline{x}_0)x_0\rangle\!\rangle\langle \breve{Z}(t), f^2\rangle, \tag{4.109}$$

or, alternatively, $(\breve{Z}(t))_{t\geq 0}$ solves the (θ, G_Z) local martingale problem for the pregenerator given in (2.8). But this local martingale problem has a unique solution by Lemma 2.1 and therefore must not depend on the subsequence.

To prove (4.108) we calculate (recall (4.19))

$$\begin{aligned}
&\mathbf{E}\Big[\Big|[\langle Z^{\tilde N_k}, f\rangle](t) - \int_0^t \langle\!\langle \nu_{Z^{\tilde N_k}(s)}, h(\overline{x_0})\langle x_0, f^2\rangle\rangle\!\rangle ds\Big|\Big] \\
&\leq \mathbf{E}\Big[\Big|\int_0^t \frac{1}{|\Lambda^{\tilde N_k}|}\sum_{\xi\in\Lambda^{\tilde N_k}} h(\overline{x}_\xi^{\tilde N_k}(s\beta_{\tilde N_k}))\langle x_\xi^{\tilde N_k}(s\beta_{\tilde N_k}), f^2\rangle \\
&\qquad\qquad - \langle\!\langle \nu_{Z^{\tilde N_k}(s)}, h(\overline{x}_0)\langle x_0, f^2\rangle\rangle\!\rangle ds\Big|\Big] \\
&\leq \mathbf{E}\Big[\int_0^t |\langle\!\langle U^{\tilde N_k}(X^{\tilde N_k}(s\beta_{\tilde N_k})), h(\overline{x}_0)\langle x_0, f^2\rangle\rangle\!\rangle - \langle\!\langle \nu_{Z^{\tilde N_k}(s)}, h(\overline{x}_0)\langle x_0, f^2\rangle\rangle\!\rangle|ds\Big],
\end{aligned} \tag{4.110}$$

so the only thing we will have to do is to set $F(X) = h(\overline{x}_0)\langle x_0, f^2\rangle$ which only depends on one variable and which is not necessarily bounded or Lipschitz and prove Lemma 4.17 for that special function F. The proof of Proposition 4.16 comes after this lemma.

Lemma 4.17. *Let $F : \mathcal{E} \to \mathbb{R}$ be a bounded function that depends only on finitely many variables $\xi \in A_F$ and which is Lipschitz with constant C_F, i.e.*

$$|F(X) - F(Y)| \leq C_F \sum_{\xi\in A_F} \rho_W(x_\xi, y_\xi) \qquad (X, Y \in \mathcal{E}). \tag{4.111}$$

Furthermore let $\tilde N_k$ be a subsequence with

$$(Z^{\tilde N_k}(t))_{t\geq 0} \Rightarrow (\breve Z(t))_{t\geq 0}, \quad \mathcal{L}(X^{\tilde N_k}(t\beta_{\tilde N_k} - p_{\tilde N_k})) \Rightarrow \mu = \int_{\mathcal{M}_f(I)} \mu_\theta \mathbf{P}[\breve Z(t) \in d\theta] \tag{4.112}$$

with space-shift ergodic distributions μ_θ with intensity θ. Then

$$\mathbf{E}[|\langle\!\langle U^{\tilde N_k}(X^{\tilde N_k}(t\beta_{\tilde N_k})), F\rangle\!\rangle - \langle\!\langle \nu_{Z^{\tilde N_k}(t)}, F\rangle\!\rangle|] \xrightarrow{k\to\infty} 0. \tag{4.113}$$

Proof. As the sequence $\tilde N_k$ is fixed we will simplify the notation and write simply N instead of $\tilde N_k$.

First we introduce some notation. The F-mean of size N of an infinite system X is denoted by

$$D_N(F, X) = \frac{1}{|\Lambda^N|}\sum_{\xi\in\Lambda^N} F(\sigma_\xi X) \tag{4.114}$$

with the space-shift operators σ_i on the lattice. We introduce three processes:

$$\begin{aligned}
\tilde X^N &: \text{finite system with ini. dist.} & \mathcal{L}(\tilde X^N(0)) &= \mathcal{L}(X^N(t\beta_N - p_N)) \\
\tilde X &: \text{infinite system with ini. dist.} & \mathcal{L}(\tilde X(0)) &= \mu = \int_{\mathcal{M}_f(I)} \mu_\theta \mathbf{P}[\breve Z(t) \in d\theta], \\
\tilde{\tilde X} &: \text{infinite system with ini. dist.} & \mathcal{L}(\tilde{\tilde X}(0)) &= \int_{\mathcal{M}_f(I)} \nu_\theta \mathbf{P}[\breve Z(t) \in d\theta].
\end{aligned} \tag{4.115}$$

These processes are coupled according to the couplings defined in Chapter 3. Then we estimate

$$
\begin{aligned}
\mathbf{E}[|\langle\!\langle U^N(X^N(t\beta_N)), F\rangle\!\rangle - \langle\!\langle \nu_{Z^N(t)}, F\rangle\!\rangle|]
&\leq \underbrace{\mathbf{E}[|\langle\!\langle U^N(\tilde{X}^N(p_N)), F\rangle\!\rangle - D_N(F, \tilde{X}(p_N))|]}_{A} \\
&+ \underbrace{\mathbf{E}[|D_N(F, \tilde{X}(p_N)) - D_N(F, \tilde{\tilde{X}}(p_N))|]}_{B} \\
&+ \underbrace{\mathbf{E}[|D_N(F, \tilde{\tilde{X}}(p_N)) - \langle\!\langle \nu_{\breve{Z}(t)}, F\rangle\!\rangle|]}_{C} \\
&+ \underbrace{\mathbf{E}[|\langle\!\langle \nu_{\breve{Z}(t)}, F\rangle\!\rangle - \langle\!\langle \nu_{Z^N(t)}, F\rangle\!\rangle|]}_{D}.
\end{aligned}
\tag{4.116}
$$

Now we can estimate the terms $A - D$. For A, we can take for large N a number $K \leq N$ such that $\xi + \eta \in \Lambda^N$ for $\xi \in \Lambda^{N-K}, \eta \in A_F$. Then for some constant C'

$$
\begin{aligned}
A &= \mathbf{E}[|\langle\!\langle U^N(\tilde{X}^N(p_N)), F\rangle\!\rangle - D_N(F, \tilde{X}(l_N))|] \\
&\leq \frac{1}{|\Lambda^N|}\Big(\sum_{\xi\in\Lambda^{N-K}} + \sum_{\xi\in\Lambda^N\setminus\Lambda^{N-K}}\Big)\mathbf{E}\big[|F(\sigma_\xi^N \tilde{X}^N(p_N)) - F(\sigma_\xi \tilde{X}(p_N))|\big] \\
&\leq C_F|A_F|\mathbf{E}[\rho_W(\tilde{x}_0^N(p_N), \tilde{x}_0(p_N))] + \frac{|\Lambda^N \setminus \Lambda^{N-K}|}{|\Lambda^N|}C'|A_F| \xrightarrow{N\to\infty} 0
\end{aligned}
\tag{4.117}
$$

by Propositions 4.12 and 4.13. Next,

$$
\begin{aligned}
B &= \mathbf{E}[|D_N(F, \tilde{X}(p_N)) - D_N(F, \tilde{\tilde{X}}(p_N))|] \\
&\leq \frac{1}{|\Lambda^N|}\sum_{\xi\in\Lambda^N}\mathbf{E}[|F(\sigma_\xi \tilde{X}(p_N)) - F(\sigma_\xi \tilde{\tilde{X}}(p_N))|] \\
&\leq C_F|A_F|\int_{\mathcal{M}_f(I)}\mathbf{E}^{\mu_\theta\otimes\nu_\theta}[\rho_W(\tilde{x}_0(p_N), \tilde{\tilde{x}}_0(p_N))]\mathbf{P}[\breve{Z}(t)\in d\theta] \xrightarrow{N\to\infty} 0
\end{aligned}
\tag{4.118}
$$

by Theorem 1.4. Then,

$$
\begin{aligned}
C &= \mathbf{E}[|D_N(F, \tilde{\tilde{X}}(p_N))| - \langle\!\langle \nu_{\breve{Z}(t)}, F\rangle\!\rangle|] \\
&\leq \mathbf{E}\Big[\Big|\frac{1}{|\Lambda^N|}\sum_{\xi\in\Lambda^N}F(\sigma_\xi, \tilde{\tilde{X}}(p_N)) - \langle\!\langle \nu_{\breve{Z}(t)}, F\rangle\!\rangle\Big|\Big] \\
&= \int_{\mathcal{M}_f(I)}\Big|\frac{1}{|\Lambda^N|}\sum_{\xi\in\Lambda^N}\langle\!\langle \nu_\theta, F\circ\sigma_\xi\rangle\!\rangle - \langle\!\langle \nu_\theta, F\rangle\!\rangle\Big|\mathbf{P}[\breve{Z}(t)\in d\theta] \xrightarrow{N\to\infty} 0
\end{aligned}
\tag{4.119}
$$

as the ν_θ are mixing (hence space-shift ergodic) by Theorem 1.4 and the ergodic theorem. Last, for D as $Z^N(t) \Rightarrow \breve{Z}(t)$ we can chose a common probability space (also denoted by $\mathbf{P}$) such that $Z^N(t) \to \breve{Z}(t)$ almost surely. Then we have

$$\begin{aligned}
D &= \mathbf{E}[\langle\!\langle \nu_{\breve{Z}(t)}, F\rangle\!\rangle - \langle\!\langle \nu_{Z^N(t)}, F\rangle\!\rangle|] \\
&= \Big| \int_{\mathcal{M}_f(I)} \int_{\mathcal{E}} F(X)\nu_\theta(dX)\mathbf{P}[\breve{Z}(t) \in d\theta] \\
&\qquad\qquad - \int_{\mathcal{M}_f(I)} \int_{\mathcal{E}} F(Y)\nu_{\theta^N}(dY)\mathbf{P}[Z^N(t) \in d\theta^N] \Big| \\
&\le \int_{(\mathcal{M}_f(I))^2} \int_{\mathcal{E}^2} |F(X) - F(Y)|\Psi(\nu_\theta, \nu_{\theta^N})(dX, dY)\mathbf{P}[\breve{Z}(t) \in d\theta, Z^N(t) \in d\theta^N] \\
&\le C_F|A_F| \int_{(\mathcal{M}_f(I))^2} \int_{(\mathcal{M}_f(I))^2} \rho_W(x_0, y_0)\Psi(\nu_\theta, \nu_{\theta^N})(dx_0, dy_0) \\
&\qquad\qquad \mathbf{P}[\breve{Z}(t) \in d\theta, Z^N(t) \in d\theta^N] \\
&\xrightarrow{N\to\infty} 0
\end{aligned} \tag{4.120}$$

by the continuity of $\theta \mapsto \nu_\theta$ from Lemma 4.14 and dominated convergence.

Proof of Proposition 4.16. Now we come to the proof of Proposition 4.16 via the remark after the proposition, i.e. we want to prove that the right hand side of (4.110) converges to 0 for $N \to \infty$. The calculation is based on the proof of (3.2) in Cox et al. (1995). However we have only assumed second moments and $h(x) = o(x)$ and not, as in that paper, qth moments for some $q > 2$ and $\limsup_{x\to\infty} \frac{h(x)}{x} < \hat{A}(0,0)^{-1}$. On page 195 of that paper it is mentioned that our assumptions are sufficient to prove their equation (3.2). We will carry this out here.

To do this take $\epsilon > 0$ and find $M < \infty$ such that

$$\sup_{x\ge M} \frac{h(x)}{x} \le \epsilon. \tag{4.121}$$

We point out that for any distribution $\mu \in \mathcal{P}(\mathcal{E})$ or $\mu \in \mathcal{P}(\mathcal{E}^N)$ with $\langle\!\langle \mu, \overline{x}_0^2\rangle\!\rangle \le K$ we find

$$\langle\!\langle \mu, h(\overline{x}_0)\overline{x}_0 1_{\{\overline{x}_0\ge M\}}\rangle\!\rangle = \langle\!\langle \mu, \frac{h(\overline{x}_0)}{\overline{x}_0}\overline{x}_0^2 1_{\{\overline{x}_0\ge M\}}\rangle\!\rangle \le \epsilon\langle\!\langle \mu, \overline{x}_0^2\rangle\!\rangle \le \epsilon K. \tag{4.122}$$

Then write

$$h(x) = h_1(x) + h_2(x) \tag{4.123}$$

where h_1 is a bounded function which is globally Lipschitz and $h_2(x) \le h(x)1_{\{x\ge M\}}$. Clearly, by Lemma 4.17 we find

$$\mathbf{E}[|\langle\!\langle U^N(X^N(t\beta_N)), h_1(\overline{x}_0)\langle x_0, f^2\rangle\rangle\!\rangle - \langle\!\langle \nu_{Z^N(t)}, h_1(\overline{x}_0), \langle x_0, f^2\rangle\rangle\!\rangle|] \xrightarrow{N\to\infty} 0 \tag{4.124}$$

so all we have to do is to prove that there is a constant C with

$$\mathbf{E}[\langle\!\langle U^N(X^N(t\beta_N)), h_2(\overline{x}_0)\langle x_0, f^2\rangle\rangle\!\rangle] \le C\epsilon, \quad \mathbf{E}[\langle\!\langle \nu_{Z^N(t)}, h_2(\overline{x}_0), \langle x_0, f^2\rangle\rangle\!\rangle|] \le C\epsilon. \tag{4.125}$$

To prove the first inequality recall that f is bounded and therefore we find a constant C_f with

$$\mathbf{E}[\langle\!\langle U^N(X^N(t\beta_N)), h_2(\overline{x}_0)\langle x_0, f^2\rangle\rangle\!\rangle] \leq C_f \mathbf{E}[h_2(\overline{x}_0^N(t\beta_N))\overline{x}_0^N(t\beta_N)] \leq C_f \epsilon K \tag{4.126}$$

as $\mathbf{E}[(\overline{x}_0^N(t\beta_N))^2] \leq K$ for some K by Proposition 4.4.

For the second inequality consider the distributions μ_N given by

$$\mu_N = \int_{\mathcal{M}_f(I)} \nu_\theta \mathbf{P}[Z^N(t) \in d\theta]. \tag{4.127}$$

Then we know from the proof of Lemma 2.12d of Cox et al. (1995) that there are constants C' and C'' with $\langle\!\langle \nu_\theta, \overline{x}_0^2\rangle\!\rangle \leq C' + C''\overline{\theta}^2$ and so

$$\begin{aligned}\langle\!\langle \mu_N, \overline{x}_0^2\rangle\!\rangle &= \int_{\mathcal{M}_f(I)} \langle\!\langle \nu_\theta, \overline{x}_0^2\rangle\!\rangle \mathbf{P}[Z^N(t) \in d\theta] = \int_0^\infty \langle\!\langle \nu_{\overline{\theta}}, x_0^2\rangle\!\rangle \mathbf{P}[\overline{Z}^N(t) \in d\theta] \\ &\leq \int_0^\infty (C' + C''\theta^2)\mathbf{P}[\overline{Z}^N(t) \in d\theta] = C' + C''\mathbf{E}[(\overline{Z}^N(t))^2] \leq C' + C''K\end{aligned} \tag{4.128}$$

again by Proposition 4.4. So,

$$\mathbf{E}[\langle\!\langle \nu_{Z^N(t)}, h_2(\overline{x}_0), \langle x_0, f^2\rangle\rangle\!\rangle|] \leq C_f \langle\!\langle \mu_N, h(\overline{x}_0)\overline{x}_0 1_{\overline{x}_0 \geq M}\rangle\!\rangle \leq C_f(C' + C''K)\epsilon \tag{4.129}$$

and we are done. □

We have proved until now that

$$(Z^N(t))_{t\geq 0} \Rightarrow (Z(t))_{t\geq 0} \quad \text{and} \quad \mathcal{L}(X^N(t\beta_N)) \Rightarrow \int_{\mathcal{M}_f(I)} \nu_\theta \mathbf{P}[Z(t) \in d\theta] \tag{4.130}$$

what can be seen from the sketch of the proof at the beginning of this chapter. So we do not have to deal any more with subsequences.

We will now come to a generalisation concerning empirical measures which have already been used in Lemma 4.17 to prove (4.130).

Proposition 4.18. *It is*

$$\mathcal{L}((U^N(X^N(t\beta_N)))_{t>0}) \underset{FDD}{\overset{N\to\infty}{\Longrightarrow}} \mathcal{L}((\nu_{Z(t)})_{t>0}). \tag{4.131}$$

Proof. We have to show that Lemma 4.17 can guarantee the weak convergence result of our proposition.

The algebra generated by functions

$$\Phi : \begin{cases} \mathcal{P}(\mathcal{E}) & \to \mathbb{R} \\ \mu & \mapsto g(\langle\!\langle \mu, F\rangle\!\rangle) \end{cases} \tag{4.132}$$

with a Lipschitz continuous function g with Lipschitz constant C_g and F as in (4.111) is strongly separating in $\mathcal{P}(\mathcal{P}(\mathcal{E}))$. But then

$$\begin{aligned}&\mathbf{E}\big[|\Phi(U^N(X^N(t\beta_N))) - \Phi(\nu_{Z(t)})|\big] \\ &\qquad \leq C_g \mathbf{E}\big[|\langle\!\langle U^N(X^N(t\beta_N)), F\rangle\!\rangle - \langle\!\langle \nu_{Z(t)}, F\rangle\!\rangle|\big] \to 0\end{aligned} \tag{4.133}$$

by Lemma 4.17.

This result can easily be lifted to the convergence of finite dimensional distributions by taking suitable test functions $\Phi : (\mathcal{P}(\mathcal{E}))^n \to \mathbb{R}$. □

Acknowledgements. This is part of my PhD Thesis. I thank my PhD supervisor, Andreas Greven, for various enlightening discussions.

References

J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander. On the attracting orbit of a non-linear transformation arising from renormalization of hierarchically interacting diffusions. I. The compact case. *Canad. J. Math.* **47** (1), 3–27 (1995).

J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander. On the attracting orbit of a nonlinear transformation arising from renormalization of hierarchically interacting diffusions. II. The non-compact case. *J. Funct. Anal.* **146** (1), 236–298 (1997).

P. Billingsley. *Convergence of Probability Measures.* Wiley, New York (1999).

J. T. Cox, D. A. Dawson and A. Greven. Mutually catalytic super branching random walks: large finite systems and renormalization analysis. *Memoirs of the AMS* **171** (2004).

J. T. Cox, K. Fleischmann and A. Greven. Comparison of interacting diffusions and an application to their ergodic theory. *Probab. Theor. Relat. Fields* **105** (4), 513–528 (1996).

J. T. Cox and A. Greven. On the long term behavior of some finite particle systems. *Probab. Theory Related Fields* **85** (2), 195–237 (1990).

J. T. Cox and A. Greven. Ergodic theorems for infinite systems of locally interacting diffusions. *Ann. Probab.* **22** (2), 833–853 (1994a).

J. T. Cox and A. Greven. The finite systems scheme: an abstract theorem and a new example. In *Measure-valued processes, stochastic partial differential equations, and interacting systems*, volume 5 of *CRM Proceedings and Lecture Notes*, pages 55–67. Amer. Math. Soc., Providence (1994b).

J. T. Cox, A. Greven and T. Shiga. Finite and infinite systems of interacting diffusions. *Probab. Theor. Relat. Fields* **103** (2), 165–197 (1995).

J. T. Cox, A. Greven and T. Shiga. Finite and infinite systems of interacting diffusions: cluster formation and universality properties. *Math. Nachr.* **192**, 105–124 (1998).

G. Dall'Aglio. Sugli estremi dei momenti delle funzioni di ripartizone doppie. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, Series III* **10**, 33–74 (1956).

D. A. Dawson. Measure-valued Markov processes. In P.L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XXI—1991*, volume 1541 of *Lecture Notes in Mathematics*, pages 1–260. Springer, Berlin (1993).

D. A. Dawson and K. Fleischmann. Longtime behavior of a branching process controlled by branching catalysts. *Stoch. Process. Appl.* **71** (2), 241–257 (1997).

D. A. Dawson and A. Greven. Multiple space-time scale analysis for interacting branching models. *Electron. J. Probab.* **1**, no. 14, 84 pp. (1996).

D. A. Dawson and A. Greven. State dependent multitype spatial branching processes and their longtime behavior. *Electron. J. Probab.* **8** (4), 93pp. (2003).

D. A. Dawson and A. Greven. Renormalization analysis of certain state dependent branching systems. *Manuscript* (2006).

D. A. Dawson, A. Greven and J. Vaillancourt. Equilibria and quasi-equilibria for infinite collections of interacting Fleming-Viot processes. *Trans. Amer. Math. Soc.* **347** (7), 2277–2360 (1995).

R. L. Dobrushin. Markov processes with a large number of locally interacting components: existence of a limit process and its ergodicity. *Problems Inform. Transmission* **7**, 149–164 (1971).

P. Donnelly and Th. Kurtz. A countable representation of the Fleming-Viot processes. *Ann. Probab.* **27**, 698–742 (1996).

S. N. Ethier and T. G. Kurtz. *Markov processes, Characterization and Convergence.* Wiley, New York (1986).

A. Gibbs and F. E. Su. On choosing and bounding probability metrics. *Manuscript* page 20 pp. (2001).

A. Greven, V. Limic and A. Winter. Representation theorems for interacting Moran models, interacting Fisher-Wright diffusions and applications. *Electron. J. Probab.* **10(39)**, 1286–1358 (2005).

O. Kallenberg. *Random Measures.* Akademie-Verlag, Berlin (1983).

I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus.* Springer-Verlag, New York (1991).

L. H. Koopmans. *The spectral analysis of time series.* Academic Press, New York and London (1974).

T. M. Liggett and F. Spitzer. Ergodic theorems for coupled random walks and other systems with locally interacting components. *Z. Wahrsch. Verw. Gebiete* **56** (4), 443–468 (1981).

L. Mytnik. Uniqueness for a mutually catalytic branching model. *Prob. Theor. Relat. Fields* **112**, 245–253 (1998).

P. Pfaffelhuber. *State-dependent interacting multitype branching systems.* Ph.D. thesis, University of Erlangen (2003).

S. T. Rachev. *Probability Metrics and the stability of stochastic models.* Wiley (1991).

T. Shiga. Ergodic theorems and exponential decay of sample paths for certain interacting diffusion systems. *Osaka J. Math.* **29** (4), 789–807 (1992).

T. Shiga and A. Shimizu. Infinite-dimensional stochastic differential equations and their applications. *J. Math. Kyoto Univ.* **20** (3), 395–416 (1980).