Branching processes with spatial structure in population models

Sarah Penington

Department of Statistics
University of Oxford

This dissertation is submitted for the degree of
Doctor of Philosophy

St John's College

August 2016

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original. This thesis is submitted to the University of Oxford for the degree Doctor of Philosophy, and has not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Sarah Penington
August 2016

Acknowledgements

First and foremost I would like to thank my supervisor Alison Etheridge for her support and guidance during my DPhil, and the freedom she gave me to choose projects.

The Oxford probability group is a lovely environment to work in. I would particularly like to thank my officemates over the past three years, Dominic, Mitch, Luke, Anna, Daniel, Raphaël and Mikolaj for many amusing chats and helpful discussions.

I would also like to thank my collaborators, Louigi and Nic, for their ideas, interest and patience and also for their hospitality on trips to Bristol, Sheffield and Montreal.

I am grateful to my friends and family for their encouragement. Special thanks goes to my housemate Jan, who can always make me laugh even after a long day in the office. Also to Henry for many closely fought board games, delicious meals, attempts at choux pastry and other adventures. Finally, I'd like to thank my parents for their support and advice, and my grandmother for many cups of tea and chats.

Abstract

We consider three different settings for branching processes with spatial structure which appear in population models.

Firstly, we consider the effect of adding a competitive interaction between nearby individuals in a branching Brownian motion. Each individual has a mass which decays when other individuals are nearby. We study the front location: the location at which the local mass density drops from $\Theta(1)$ to $o(1)$. We show that there are arbitrarily large times t at which the front location is $\Theta(t^{1/3})$ behind the maximum displacement of a particle from the origin.

Secondly, we study the strength of selection in favour of a particular allele in a spatially structured population required to cause a detectable trace in the patterns of genetic variation observed in the contemporary population. We suppose that the effective local population density is small. We show that whereas in dimensions at least three, selection is barely impeded by the spatial structure, in the most relevant dimension, $d = 2$, selection must be stronger (by a factor of $\log(1/\mu)$ where μ is the neutral mutation rate) if we are to have a chance of detecting it.

Finally, we model the behaviour of what are known in population genetics as hybrid zones. These occur when two genetically distinct groups are able to reproduce, but the hybrid offspring have a lower fitness. We prove that on an appropriate time and space scale, the hybrid zone in our model evolves approximately according to mean curvature flow. We also give a probabilistic proof of a (well-known) analogous result for a special case of the Allen-Cahn equation.

In the last two cases, we use the spatial Λ-Fleming-Viot process to model the population (with different selection mechanisms), and our proofs rely on a duality with a system of branching and coalescing particles.

Table of contents

Chapter 1

Introduction

Branching processes are a central theme in the study of population models. They appear as forwards-in-time models of growing populations, and also as backwards-in-time dual processes of 'potential' ancestry in models of natural selection. This thesis contains work on branching processes in both of these contexts. The population models that we shall consider incorporate spatial structure, resulting in branching processes of particles moving in space in which nearby particles interact. This introduction covers background on branching processes and population genetics models, and gives an outline of the results in the rest of the thesis.

1.1 Branching Brownian motion

1.1.1 Definition and important properties

Branching Brownian motion is a canonical example of a branching process, and can be seen as a simple model of an evolving population. The following definition is from Etheridge (2000); a more general early definition of a branching Markov process can be found in Ikeda et al. (1965).

Definition 1.1.1 (Branching Brownian motion in $\mathbb{R}^d$)**.** *Branching Brownian motion has three ingredients:*

1. ***The spatial motion:*** *During its lifetime, each individual in the population moves around in* $\mathbb{R}^d$ *according to a d-dimensional Brownian motion.*

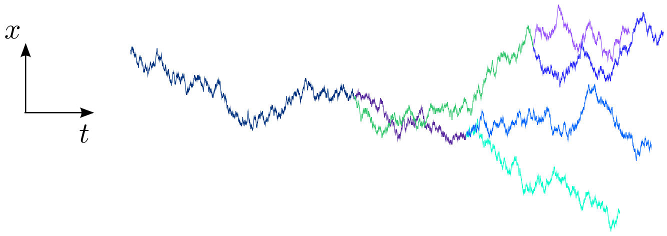

Figure 1.1: A realisation of one-dimensional branching Brownian motion.

2. ***The branching rate, V:*** *Each individual has an exponentially distributed lifetime with rate* V*. Given that the individual is alive at time* t*, the probability that it dies in the time interval* $[t, t + \delta t)$ *is* $V\delta t + o(\delta t)$*.*

3. ***The branching mechanism,*** Φ***:*** *When it dies, an individual leaves behind (at the location where it died) a random number of offspring with probability generating function* Φ*. Conditional on their time and place of birth, offspring evolve independently of each other.*

Figure 1.1 gives a realisation of branching Brownian motion (BBM) in $\mathbb{R}$. We shall write $X_1(t), X_2(t) \ldots, X_{N(t)}(t)$ to denote the locations of the individuals in a branching Brownian motion at time t, where $N(t)$ is the number of individuals alive at time t. We shall use the subscript x in $\mathbb{P}_x$, $\mathbb{E}_x$ to indicate that the BBM starts with a single individual at location x at time 0.

There are two important immediate consequences of this definition.

- **Markov property:** The exponential lifetimes of individuals are memoryless and the Brownian motions followed by each individual are Markov processes, so branching Brownian motion is a Markov process.

- **Branching property:** Conditional on their locations $X_1(t), X_2(t) \ldots, X_{N(t)}(t)$ at time t, the descendants of the individuals alive at time t evolve according to independent branching Brownian motions.

We now give two important results on branching Brownian motion; they are related to some of the main results in this thesis. The first is usually attributed to McKean (1975) (but can also be found in Skorohod (1964)) and gives a solution to partial differential equations of a particular form in terms of the law of a branching Brownian motion.

Theorem 1.1.2. *Suppose* $\psi \in C^2(\mathbb{R}^d)$ *with* $0 \leq \psi(x) \leq 1$ *for* $x \in \mathbb{R}^d$. *For* $t \geq 0$, $x \in \mathbb{R}^d$, *let*

$$v(t,x) = \mathbb{E}_x\left[\prod_{i=1}^{N(t)} \psi(X_i(t))\right],$$

where $X_1(t), \dots, X_{N(t)}(t)$ *are the locations of the individuals in a branching Brownian motion at time* t, V *is the branching rate and* Φ *is the probability generating function of the offspring distribution. Then if* Φ *is continuous,* v *solves the partial differential equation*

$$\frac{\partial v}{\partial t} = \frac{1}{2}\Delta v + V(\Phi(v) - v), \qquad v(0,x) = \psi(x). \tag{1.1}$$

Proof. We assume that at time t, $v(t,x)$ is twice continuously differentiable in the space variable x. Once we have shown under this assumption that v solves (1.1), regularity theory for the heat semigroup tells us that v is smooth as a function of x for all later times.

Throughout the proof we write E_x for the expectation under which $(B_s)_{s\geq 0}$ is a Brownian motion started at x.

We estimate $v(t+\delta t, x)$ by conditioning on the behaviour of the original ancestor during the time interval $[0, \delta t]$. Let S be the time at which the original ancestor dies, B_S its location at time S and K the number of offspring that it leaves behind when it dies. Then

$$\begin{aligned} v(t+\delta t, x) = \sum_{k=0}^{\infty} \mathbb{E}_x &\left[\prod_{j=1}^{N(t+\delta t)} \psi(X_j(t+\delta t))\middle| S \leq \delta t, K = k\right] \mathbb{P}\left[S \leq \delta t, K = k\right] \\ &+ \mathbb{E}_x\left[\prod_{j=1}^{N(t+\delta t)} \psi(X_j(t+\delta t))\middle| S > \delta t\right] \mathbb{P}\left[S > \delta t\right]. \end{aligned} \tag{1.2}$$

We now estimate each term on the right hand side of (1.2).

For each offspring $1 \leq i \leq K$ from the first branching event, let N_i denote the set of individuals at time $t+\delta t$ which are descendants of offspring i. Conditional on $S \leq \delta t$, the probability that offspring i dies before time δt is $\mathcal{O}(\delta t)$. Hence for $s \leq \delta t$ and $i \leq k$, by the Markov property at time δt,

$$\mathbb{E}_x\left[\prod_{j\in N_i}\psi(X_j(t+\delta t))\middle|(S,B_S,K)=(s,y,k)\right]=E_y\left[\upsilon(t,B_{\delta t-s})\right]+\mathcal{O}(\delta t).$$

By our assumption on the regularity of $\upsilon(t,x)$ in x, it follows that for $i\leq k$,

$$\mathbb{E}_x\left[\prod_{j\in N_i}\psi(X_j(t+\delta t))\middle|S\leq\delta t,K=k\right]=\upsilon(t,x)+\mathcal{O}(\delta t).$$

Then by the branching property,

$$\mathbb{E}_x\left[\prod_{j=1}^{N(t+\delta t)}\psi(X_j(t+\delta t))\middle|S\leq\delta t,K=k\right]=(\upsilon(t,x)+\mathcal{O}(\delta t))^k.$$

Substituting this into the first term on the right hand side of (1.2), since $S\sim\operatorname{Exp}(V)$,

$$\begin{aligned}\sum_{k=0}^{\infty}\mathbb{E}_x&\left[\prod_{j=1}^{N(t+\delta t)}\psi(X_j(t+\delta t))\middle|S\leq\delta t,K=k\right]\mathbb{P}\left[S\leq\delta t,K=k\right]\\&=\sum_{k=0}^{\infty}(\upsilon(t,x)+\mathcal{O}(\delta t))^k(V\delta t+o(\delta t))\mathbb{P}\left[K=k\right]\\&=V\delta t\Phi(\upsilon(t,x))+o(\delta t),\end{aligned}\tag{1.3}$$

where the last line follows since Φ is the probability generating function of K and Φ is continuous. To estimate the second term on the right hand side of (1.2), since $\frac{1}{2}\Delta$ is the infinitesimal generator of Brownian motion and by our assumption on the regularity of υ,

$$\begin{aligned}\mathbb{E}_x\left[\prod_{j=1}^{N(t+\delta t)}\psi(X_j(t+\delta t))\middle|S>\delta t\right]\mathbb{P}\left[S>\delta t\right]&=E_x\left[\upsilon(t,B_{\delta t})\right](1-V\delta t+o(\delta t))\\&=(\upsilon(t,x)+\tfrac{1}{2}\Delta\upsilon(t,x)\delta t+o(\delta t))(1-V\delta t+o(\delta t)).\end{aligned}\tag{1.4}$$

Substituting (1.3) and (1.4) into (1.2), as $\delta t\to 0$,

$$\frac{\upsilon(t+\delta t,x)-\upsilon(t,x)}{\delta t}=\tfrac{1}{2}\Delta\upsilon(t,x)+V(\Phi(\upsilon(t,x))-\upsilon(t,x))+o(1),$$

and the result follows. □

The second main theorem in this section is a combination of two results. The first is originally due to Bramson (1978); it gives the asymptotics for the median of the distribution of the maximum position of the individuals alive in a one-dimensional branching Brownian motion at time t. The second is a strengthening of Bramson's result from Hu and Shi (2009) (originally given in a slightly different form for branching random walks). For this theorem we take $V = 1$ and $\Phi(v) = v^2$ in the definition of branching Brownian motion, resulting in binary branching at rate 1.

Theorem 1.1.3. *Let* $m(t) = \sup\{x \in \mathbb{R} : \mathbb{P}\left[\max_{1\le i\le N(t)} X_i(t) \le x\right] \le 1/2\}$. *Then as* $t \to \infty$,

$$m(t) = \sqrt{2}t - \frac{3}{2\sqrt{2}}\log t + \mathcal{O}(1).$$

Moreover, almost surely

$$\limsup_{t\to\infty} \frac{|\max_{1\le i\le N(t)} X_i(t) - m(t)|}{\log t} < \infty.$$

A simpler version of Bramson's proof (which is then extended to cover Hu and Shi's result) is given in Roberts (2013).

One of the basic tools used to prove this and other results on BBM is the many-to-one lemma. Once again, for simplicity, we consider binary branching at rate 1.

Lemma 1.1.4 (Many-to-one lemma)**.** *For* $s \le t$, *let* $X_{i,t}(s)$ *denote the location of the ancestor of* $X_i(t)$ *at time* s. *Suppose that for* $t \ge 0$, $F_t : C[0,t] \to \mathbb{R}$ *is a measurable functional. Then*

$$\mathbb{E}_x\left[\sum_{1\le i\le N(t)} F_t((X_{i,t}(s), 0 \le s \le t))\right] = e^t E_x\left[F_t((B(s), 0 \le s \le t))\right],$$

where under E_x, B *is a Brownian motion started at* x.

It is not hard to see that since a population of k individuals increases by 1 at rate k, $\frac{d}{dt}\mathbb{E}_x[N(t)] = \mathbb{E}_x[N(t)]$ and hence $\mathbb{E}_x[N(t)] = e^t$. The proof of Lemma 1.1.4 is then a simple consequence of the fact that conditional on the set of branching events, each path $(X_{i,t}(s), 0 \le s \le t)$ is a Brownian motion.

1.1.2 Branching Brownian motion with decay of mass

In Chapter 2, (Addario-Berry and Penington (2015)), we introduce a model of branching Brownian motion with decay of mass. The idea is to give a mathematical model for the effect of competition for resources in a spatially structured population. We think of individuals as amoebae moving around in space; if no other individuals are near a particular amoeba then it has enough food, but if it is near other individuals then it must share resources and its mass decays.

We consider binary branching Brownian motion, in which individuals have an Exp(1)-distributed lifetime, after which they are replaced by two offspring. We add a competitive interaction between nearby individuals as follows. Each individual is given a mass which decays at rate proportional to the local mass density at its location times its own mass. The total mass of the system increases through branching events: when branching occurs each of the two offspring inherits the mass of its parent.

As in the previous subsection, we let $(X_i(t))_{1\le i\le N(t)}$ denote the locations of the $N(t)$ individuals at time t. For $s \le t$, let $X_{i,t}(s)$ give the location of the ancestor of $X_i(t)$ at time s. We also define $(M_i(t))_{1\le i\le N(t)}$ and call $M_i(t)$ the mass of individual i at time t. For $t\ge 0$ and $x\in\mathbb{R}$, we let

$$\zeta(t,x) = \sum_{\{1\le i\le N(t):|X_i(t)-x|\in(0,1)\}} M_i(t),$$

the local mass density at time t and location x. This can be seen as the sum of the masses of individuals that an individual at location x at time t has to compete with. Then the initial individual has mass 1 at time 0, and masses decay according to

$$\frac{d}{dt}M_i(t) = -M_i(t)\zeta(t,X_i(t)).$$

Equivalently, for each $1\le i\le N(t)$,

$$M_i(t) = \exp\left(-\int_0^t \zeta(s,X_{i,t}(s))\,ds\right).$$

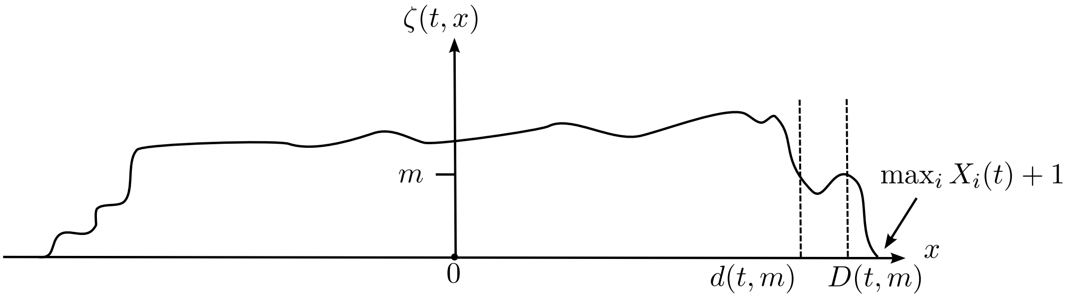

Figure 1.2: The front location for a local mass density profile.

In order to determine the effect of the competitive interaction on the branching Brownian motion, a natural first step is to look at how quickly mass spreads out in space when the system is started from a single particle at the origin. To do this we study the asymptotics of the front location: the rightmost location at which the local mass density is not $o(1)$. More formally, for $m > 0$, we write

$$d(t,m) = \min\{x > 0 : \zeta(t,x) < m\}, \quad D(t,m) = \max\{x : \zeta(t,x) > m\},$$

for the leftmost (positive) location at which the total mass of nearby particles falls below m, and the rightmost location at which it exceeds m, respectively (see Figure 1.2). Note that for $y \geq \max_{i\leq N(t)} X_i(t)+1$, $\zeta(t,y) = 0$ and so $d(t,m), D(t,m) < \max_{i\leq N(t)} X_i(t)+1$. The following theorem gives us partial results on how far d and D lag behind the rightmost particle location for large times.

Theorem 1.1.5. *Write $c^* = 3^{4/3}\pi^{2/3}/2^{7/6}$. Then almost surely, for all $m < 1$,*

$$\limsup_{t\to\infty} \frac{\sqrt{2}t - d(t,m)}{t^{1/3}} \geq c^* \quad \text{and} \quad \liminf_{t\to\infty} \frac{\sqrt{2}t - D(t,m)}{t^{1/3}} \leq c^*$$

Recalling Theorem 1.1.3, the theorem states that (1) there are arbitrarily large times t at which the first low-density region lags at least distance $c^* t^{1/3} + o(t^{1/3})$ behind the rightmost particle, and (2) there are also (potentially different) arbitrarily large times t at which there is some high-density region within distance $c^* t^{1/3} + o(t^{1/3})$ of the rightmost particle.

The proof uses results from Jaffuel (2012) and Roberts (2015) on consistent maximal displacement for branching Brownian motion. These results tell us with high probability whether or not $\{i \leq N(t) : X_{i,t}(s) \geq f(s)\,\forall s \in [0,t]\}$ is empty for large times t, where $f(s) = \sqrt{2}s - cs^{1/3} + o(s^{1/3})$ and $c > 0$ is a constant.

In the conclusion we will discuss some progress on further results relating to this model.

1.2 Population genetics models

1.2.1 The Moran model and Kingman's coalescent

We now move on to consider population genetics models and the duality between forwards-in-time models and the backwards-in-time genealogy of the population. We shall begin with a simple classical model of a population with no spatial structure, introduced by Moran (1958), to illustrate the concept of duality. Our discussion is based on a combination of Etheridge (2011) and Berestycki (2009).

Definition 1.2.1 (The neutral Moran model). *A population of N individuals evolves as follows. Let $(\pi(t), t \geq 0)$ be a Poisson process with rate $\binom{N}{2}$. At a point in π, a reproduction event occurs in which an ordered pair of individuals is sampled uniformly at random from the population, the first dies and the second splits in two.*

More formally, we label each individual in the population at time 0 by $i \in \{1, 2, \ldots, N\}$. We let $\pi_{(i,j)}$ be an independent rate 1 Poisson process for each pair (i, j) with $i \neq j$. At each point in $\pi_{(i,j)}$, one of the individuals labelled i, j (chosen at random with equal probability) dies, and the other is replaced by two offspring. The offspring are given labels i and j.

We consider the case in which each individual in the population is one of two types, a or A. In order to determine the types of a sample from the population at time t from the initial types of the population, we must trace the ancestry of the sample backwards in time. If the most recent reproduction event before time t is at time s and resulted in offspring i and j, then before time s, i and j had the same ancestor, so the ancestral lineages for i and j coalesce in the genealogy. This process continues until we have traced the ancestral lineages back to their ancestors at time 0. Each individual's type is

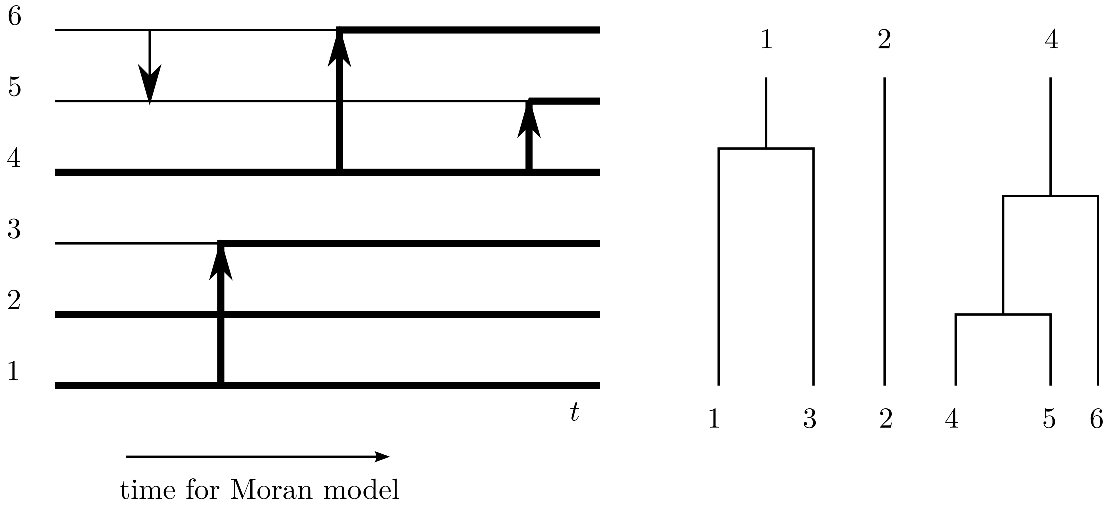

Figure 1.3: On the left is a graphical representation of the Moran model. We draw an arrow $i \to j$ at a time when i reproduces and j dies. The bold lines show the ancestors of the individuals at time t traced back to time 0. On the right is the genealogy relating individuals in the population, tracing back time from t to 0.

then the type of its ancestor at time 0. The genealogical trees from a realisation of the Moran model are shown in Figure 1.3.

For a pair i, j in the sample at time t, let $(i(s), j(s))$ denote the labels of their ancestors at time $s \leq t$. Since for each k, l, the time reversal of $\pi_{(k,l)}$ is also a Poisson process, and by the memoryless property of the exponential distribution, the length of time we must look back before time t until we see a reproduction event with offspring $(i(s), j(s))$ is exponentially distributed with rate 1.

It follows that for the Moran model run for a time $t > 0$, the genealogy of a sample of size k is given by the following continuous time Markov process $(\Pi_s)_{0 \leq s \leq t}$, taking values in $\mathscr{E}_k$, the set of equivalence classes of $[k] = \{1, 2, \ldots, k\}$.

Definition 1.2.2 (Kingman's k-coalescent)**.** $(\Pi_s)_{0 \leq s \leq t}$ *is a continuous time Markov chain on* $\mathscr{E}_k$*, started at* $\Pi_0 = (\{1\}, \{2\}, \ldots, \{k\})$ *with transition rates* $q_{\xi,\eta}$ $(\xi, \eta \in \mathscr{E}_k)$ *given by*

$$q_{\xi,\eta} = \begin{cases} 1 & \textit{if } \eta \textit{ is obtained by coalescing two of the equivalence classes of } \xi \\ 0 & \textit{otherwise.} \end{cases}$$

Each equivalence class (or ‘block’) of Π_s corresponds to an ancestor alive at time $t-s$, and consists of the labels of all the individuals in the sample descended from that ancestor.

Consider a Moran model with N individuals, each of which is either type a or A; let p_t^N denote the proportion of the population at time t which is type a. Then if a reproduction event occurs at time t, $p_t^N = p_{t-}^N + \frac{1}{N}$ if a type A individual dies and a type a reproduces; this happens with probability $\frac{N}{N-1} p_{t-}^N (1-p_{t-}^N)$. The rate of reproduction events is $\binom{N}{2}$. Hence if $p_t^N = p$,

$$p_{t+\delta t}^N = \begin{cases} p + \frac{1}{N} & \text{with probability } \frac{N^2}{2} p(1-p)\delta t + o(\delta t) \\ p - \frac{1}{N} & \text{with probability } \frac{N^2}{2} p(1-p)\delta t + o(\delta t) \\ p & \text{with probability } 1 - N^2 p(1-p)\delta t + o(\delta t). \end{cases}$$

It follows that

$$\begin{aligned} & \mathbb{E}\left[p_{t+\delta t}^N - p_t^N \middle| \sigma((p_s^N)_{0 \le s \le t}) \right] = o(\delta t) \\ \text{and} \quad & \mathrm{Var}(p_{t+\delta t}^N - p_t^N | \sigma((p_s^N)_{0 \le s \le t})) = p_t^N (1 - p_t^N)\delta t + o(\delta t). \end{aligned}$$

Hence by standard martingale arguments (see for example Chapter 4 of Ethier and Kurtz (2009)), for $p_0^N = p \in (0,1)$, $(p_t^N)_{t \ge 0}$ converges in the Skorohod topology on $\mathbb{D}[0,\infty)$ to $(p_t)_{t\ge 0}$ as $N \to \infty$, where $(p_t)_{t \ge 0}$ solves the stochastic differential equation

$$dp_t = \sqrt{p_t(1-p_t)} dW_t, \quad p_0 = p,$$

and W is a standard Brownian motion. The process $(p_t)_{t\ge 0}$ is known as the Wright-Fisher diffusion.

We now show that a simple consequence of the relationship between the Wright-Fisher diffusion and Kingman's k-coalescent is an example of what is known as a duality relation.

Definition 1.2.3 (Duality relation)**.** *Suppose* $(X_t)_{t\ge 0}$ *and* $(Y_t)_{t \ge 0}$ *are stochastic processes taking values in* $\mathcal{S}_X$ *and* $\mathcal{S}_Y$ *and with laws* $\mathbb{P}$ *and* $\mathbb{Q}$ *respectively. We say that* X

and Y *are dual to each other with respect to the function* $\psi : \mathcal{S}_X \times \mathcal{S}_Y \to \mathbb{R}$ *if for all* $x \in \mathcal{S}_X$, $y \in \mathcal{S}_Y$, $t \geq 0$,

$$\mathbb{E}_x^{\mathbb{P}} \left[\psi(X_t, y)\right] = \mathbb{E}_y^{\mathbb{Q}} \left[\psi(x, Y_t)\right], \tag{1.5}$$

where $\mathbb{E}_x^{\mathbb{P}}$ *is the law of* $(X_t)_{t\geq 0}$ *with* $X_0 = x$ *and* $\mathbb{E}_y^{\mathbb{Q}}$ *is the law of* $(Y_t)_{t\geq 0}$ *with* $Y_0 = y$.

If varying y allows us to get a sufficiently large family of functions of X_t from $\psi(X_t, y)$, the law of X_t can be characterised by the law of its dual Y_t. In the special case $\psi(x, y) = x^y$, (1.5) is known as moment duality.

Proposition 1.2.4. *Let* $(p_t)_{t\geq 0}$ *be a Wright-Fisher diffusion. Let* $(n_t)_{t\geq 0}$ *denote the number of blocks in Kingman's* n*-coalescent. Then for* $p \in [0, 1]$, $n \in \mathbb{N}$,

$$\mathbb{E}_p^{\rightarrow} \left[(p_t)^n\right] = \mathbb{E}_n^{\leftarrow} \left[p^{n_t}\right],$$

where $\mathbb{E}_p^{\rightarrow}$ *is the law of the Wright-Fisher diffusion with* $p_0 = p$ *and* $\mathbb{E}_n^{\leftarrow}$ *is the law of Kingman's* n*-coalescent.*

Proof. (Sketch based on the argument given in Berestycki (2009).) Consider a Moran model with N individuals, each of which is type a independently with probability p at time 0, and type A otherwise. Let p_t^N denote the proportion of type a at time t.

Sample n individuals uniformly at random at time t, and let E be the event that the sampled individuals are all of type a. Then conditional on the genealogy of the sample, $(\Pi_s)_{s\leq t}$, E occurs if and only if each ancestor at time 0 of the sample is type a. It follows that $\mathbb{P}\left[E\right] = \mathbb{E}_n^{\leftarrow} \left[p^{n_t}\right]$.

On the other hand, for large N, since p_t^N is the proportion of type a in the population and we sample n types at random, $\mathbb{P}\left[E\right] \approx \mathbb{E}_p^{\rightarrow} \left[(p_t^N)^n\right] \approx \mathbb{E}_p^{\rightarrow} \left[(p_t)^n\right]$. □

1.2.2 The ancestral selection graph

The Moran model is a model of *genetic drift*, the randomness due to reproduction in a finite population. We now add the effect of natural selection to our model. We suppose that the types A and a have *relative fitness* $1+s$ and 1 respectively, for some $s > 0$. This means that individuals with type A produce on average $1 + s$ times as many offspring as a individuals. We can incorporate this into the Moran model as follows.

Definition 1.2.5 (The Moran model with selection)**.** *The population of N individuals, each of type a or A, evolves as follows. Let π be a Poisson process with rate $(1-s)\binom{N}{2}$ and π^s an independent Poisson process with rate $s\binom{N}{2}$.*

At a point in π, a 'neutral' event occurs: an ordered pair of individuals is sampled uniformly at random from the population, the first dies and the second splits in two.

At a point in π^s, a 'selective' event occurs: an ordered pair of individuals is sampled uniformly at random from the population. If the chosen individuals are of the same type, the first dies and the second splits in two; if one is type a and the other is type A then the A individual splits in two and the a individual dies.

The value s is known as the selection parameter. We now motivate this definition in terms of the relative fitnesses of the types. If the proportion of a individuals in the population is x, then at a reproduction event where the type of offspring is α, since A individuals produce $1+s$ times as many offspring as a individuals we should have

$$\frac{\mathbb{P}\left[\alpha = a\right]}{\mathbb{P}\left[\alpha = A\right]} = \frac{x}{(1+s)(1-x)}.$$

It follows that

$$\mathbb{P}\left[\alpha = a\right] = \frac{x}{1+s(1-x)} = (1-s)x + sx^2 + \mathcal{O}(s^2), \tag{1.6}$$

by expanding in powers of s. We assume that s is small, and neglect the $\mathcal{O}(s^2)$ terms. Then the first term on the right hand side of (1.6) is $1-s$ times the probability that a randomly selected individual is type a, and the second term is s times the probability that a randomly selected pair of individuals are both of type a. This gives us our two types of events in Definition 1.2.5.

In order to determine the types of a sample from the population at time t from the initial types of the population, we again trace the ancestry of the sample backwards in time. If the most recent reproduction event before time t was a neutral event with offspring labelled i and j, then the parent individual was labelled either i or j with equal probability, and both i and j inherited the type of the parent. However, if the most recent reproduction event before time t was selective with offspring labelled i

and j, the probability that the parent was labelled i depends on the types of i and j before the event. In order to determine the outcome of the event we must trace back the ancestry of both 'potential' parents and determine their types from the initial types of the population.

In this way we trace backwards in time a branching and coalescing graph, known as the ancestral selection graph (ASG), introduced in Krone and Neuhauser (1997); Neuhauser and Krone (1997). At each neutral event which hits two lineages already in the graph, the lineages coalesce (this happens at rate $1-s$ for each pair in the graph); at each selective event which hits a lineage already in the graph, the lineage branches into the two potential ancestral lineages (this happens at rate $s(N-1)$ for each lineage). Once we have traced back to time 0, we determine the type of each potential ancestor from the initial types and then work back through the graph, deciding what took place at each selective event (see Figure 1.4 for an example). The type of the offspring of a selective event in the graph is a if and only if both parents are of type a.

The true genealogy of the sample is embedded in the ASG; to determine this genealogy, when defining the ASG we must specify an incoming and a continuing branch at each selective event in order to decide which is the true parent if both potential parents have the same type (see Krone and Neuhauser (1997)).

For a Moran model with selection with N individuals and selection parameter s_N, let p_t^N denote the proportion of a individuals at time t. To obtain a non-trivial limit as $N \to \infty$, we set

$$s_N = \frac{s}{N} \text{ for some constant } s > 0$$

(this is known as the weak selection regime).

If a selective event occurs at time t, $p_t^N = p_{t-}^N - \frac{1}{N}$ if the pair of parents have different types, which happens with probability $2\frac{N}{N-1}p_{t-}^N(1-p_{t-}^N)$; otherwise $p_t^N = p_{t-}^N$. Hence if $p_t^N = p$,

$$p_{t+\delta t}^N = \begin{cases} p + \frac{1}{N} & \text{with probability } (1-\frac{s}{N})\frac{N^2}{2}p(1-p)\delta t + o(\delta t) \\ p - \frac{1}{N} & \text{with probability } (1-\frac{s}{N})\frac{N^2}{2}p(1-p)\delta t + sNp(1-p)\delta t + o(\delta t) \\ p & \text{with probability } 1 - N^2 p(1-p)\delta t + o(\delta t). \end{cases}$$

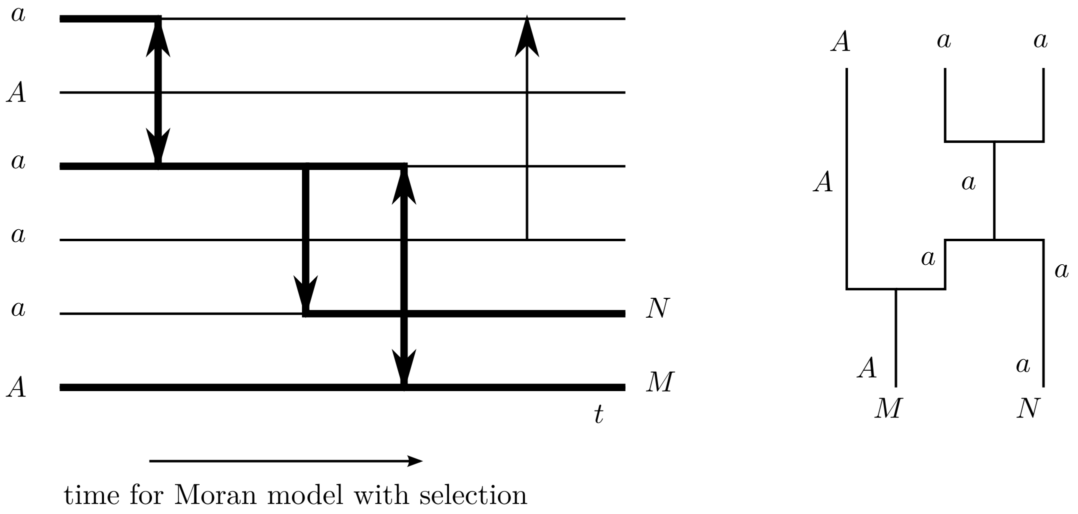

Figure 1.4: On the left is a graphical representation of the Moran model with selection. We draw an arrow $i \to j$ at a neutral event in which i reproduces and j dies. At a selective event involving i and j we draw a double headed arrow between them. We trace back the potential ancestral lineages of the sample M, N from time t to time 0; these are the bold lines. On the right is the ancestral selection graph (ASG) for M, N. The types of M and N are determined from the initial types by working down the ASG.

It follows that

$$\begin{aligned} &\mathbb{E}\left[p^N_{t+\delta t} - p^N_t | \sigma((p^N_s)_{0\le s\le t})\right] = -s p^N_t (1-p^N_t)\delta t + o(\delta t) \\ \text{and} \quad &\mathrm{Var}(p^N_{t+\delta t} - p^N_t | \sigma((p^N_s)_{0\le s\le t})) = p^N_t(1-p^N_t)\delta t + o(\delta t). \end{aligned}$$

Hence by standard martingale arguments (as in Section 1.2.1), for $p^N_0 = p \in (0,1)$, $(p^N_t)_{t\ge 0}$ converges in the Skorohod topology on $\mathbb{D}[0,\infty)$ to $(p_t)_{t\ge 0}$, where $(p_t)_{t\ge 0}$ solves the stochastic differential equation

$$dp_t = -s p_t(1-p_t)dt + \sqrt{p_t(1-p_t)}dW_t, \quad p_0 = p,$$

and W is a standard Brownian motion. The process $(p_t)_{t\ge 0}$ is known as the Wright-Fisher diffusion with selection.

As in Section 1.2.1, there is a duality relation between the Wright-Fisher diffusion with selection and the ancestral selection graph. Note that in the weak selection regime of $s_N = s/N$, for large N, the number of lineages in the ancestral selection graph for the population at time t is given approximately by the birth and death process $(n_s)_{0\le s\le t}$ in which

$$n \mapsto n-1 \text{ at rate } \binom{n}{2}, \quad n \mapsto n+1 \text{ at rate } sn. \tag{1.7}$$

Proposition 1.2.6. *Let* $(p_t)_{t\geq 0}$ *solve the stochastic differential equation*

$$dp_t = -sp_t(1-p_t)dt + \sqrt{p_t(1-p_t)}dW_t$$

where W *is a standard Brownian motion. Let* $(n_t)_{t\geq 0}$ *denote the birth and death process with rates given by* (1.7)*. Then for* $p \in [0,1]$*,* $n \in \mathbb{N}$*,*

$$\mathbb{E}_p^{\rightarrow}\left[(p_t)^n\right] = \mathbb{E}_n^{\leftarrow}\left[p^{n_t}\right],$$

where $\mathbb{E}_p^{\rightarrow}$ *is the law of* $(p_t)_{t\geq 0}$ *with* $p_0 = p$ *and* $\mathbb{E}_n^{\leftarrow}$ *is the law of* $(n_t)_{t\geq 0}$ *with* $n_0 = n$*.*

Proof. (sketch) The argument is essentially the same as the sketch proof of Proposition 1.2.4. Consider a Moran model with selection parameter s/N with N individuals, each of which is type a independently with probability p at time 0, and type A otherwise. Let p_t^N denote the proportion of type a individuals at time t.

Sample n individuals uniformly at random at time t, and let E be the event that the sampled individuals are all of type a. Then conditional on the ancestral selection graph for the sample, E occurs if and only if each potential ancestor at time 0 of the sample is type a. It follows that for large N, $\mathbb{P}[E] \approx \mathbb{E}_n^{\leftarrow}[p^{n_t}]$.

On the other hand, for large N, since p_t^N is the proportion of type a and we sample n types at random, $\mathbb{P}[E] \approx \mathbb{E}_p^{\rightarrow}\left[(p_t^N)^n\right] \approx \mathbb{E}_p^{\rightarrow}\left[(p_t)^n\right]$. □

1.2.3 Population genetics models with spatial structure

In the population genetics models that we have discussed so far in Sections 1.2.1-1.2.2, the implicit assumption is that the population is well-mixed, i.e. each new individual is equally likely to be the offspring of any of the individuals alive in the population at the time of its birth. This assumption is clearly false if we consider a population spread out over a large area.

A simple way to add the effect of spatial structure to our existing models is to subdivide the population into demes; subpopulations interact through migration between demes. The following model is related to Kimura's stepping stone model (Kimura (1953)) and Wright's island model (Wright (1943)). For simplicity we shall assume that the population at each deme has the same size.

Definition 1.2.7 (Island model)**.** *Suppose I is a connected graph (finite or infinite) with vertices labelled $1, 2, \ldots$; a population is structured into demes at each vertex of I, where the size of the population at vertex j is N for each j. The rate at which individuals in deme i migrate to deme j is denoted m_{ij}.*

Let $(\pi_i)_{i\in I}$ be an i.i.d. sequence of Poisson processes with rate $\binom{N}{2}$. Also let $(\pi_{i,j})_{i\neq j\in I}$ be a sequence of independent Poisson processes, where $\pi_{i,j}$ has rate Nm_{ij}.

For each deme i, at a point in π_i an ordered pair of individuals is sampled uniformly at random from the population in deme i, the first dies and is replaced by offspring of the second individual. Also for each $i \neq j \in I$, at a point in $\pi_{i,j}$ an individual randomly chosen from the population at j dies and is replaced by offspring of an individual randomly chosen from the population at i.

If we trace the ancestry of a sample from the population backwards in time, in order to see the rate at which lineages coalesce we must record their location. The ancestral lineages then evolve according to a spatially structured coalescent. For a pair of individuals in deme i at time t, their ancestral lineages coalesce at rate 1. Also a lineage in deme i jumps to a different deme j at rate m_{ji}. A pair of lineages in different demes, i and j, coalesce at rate $\frac{1}{N}(m_{ji} + m_{ij})$ so for large N we will not see such events.

Consider the ancestral lineages of a sample consisting of n_i individuals in deme i at time t for each $i \in I$, where $n_i = 0$ for all but finitely many $i \in I$. For large N, the genealogy is approximately given by the system of coalescing random walks described as follows.

Definition 1.2.8. *The process $(\underline{n}(t)) = (n_i(t))_{i\in I}$ is a continuous time Markov chain taking values in $\mathbb{N}_0^I$ with $n_i(0) = n_i$ $\forall i \in I$ and transition rates as follows.*

- *For each $i \in I$, $n_i \mapsto n_i - 1$ at rate $\binom{n_i}{2}$.*
- *For each $i, j \in I$ with $i \neq j$,* $\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$ *at rate $n_i m_{ji}$.*

We now add selection to the island model by letting the subpopulation within a deme evolve as in the Moran model with selection of Definition 1.2.5. We suppose that individuals are either type a or type A, and that the types a and A have relative fitness 1 and $1 + s$ respectively, where $s > 0$ is the selection parameter.

Definition 1.2.9 (Island model with selection)**.** *Suppose I is a connected graph (finite or infinite) with vertices labelled $1, 2, \ldots$; a population is structured into demes at each vertex of I, where the size of the population at vertex j is N for each j. The rate at which individuals in deme i migrate to deme j is denoted m_{ij}.*

Let $(\pi_i)_{i\in I}$ be an i.i.d. sequence of Poisson processes with rate $(1-s)\binom{N}{2}$ and let $(\pi_i^s)_{i\in I}$ be an i.i.d. sequence of Poisson processes with rate $s\binom{N}{2}$. Also let $(\pi_{i,j})_{i\neq j\in I}$ be a sequence of independent Poisson processes, where $\pi_{i,j}$ has rate Nm_{ij}.

For each deme i, at a point in π_i a neutral event occurs: an ordered pair of individuals is sampled uniformly at random from the population in deme i, the first dies and is replaced by offspring of the second individual. At a point in π_i^s a selective event occurs: an ordered pair of individuals is sampled uniformly at random from the population in deme i. If the chosen individuals are of the same type, the first dies and is replaced by offspring of the second individual; if one is type a and the other is type A then the a individual dies and is replaced by offspring of the A individual.

Also for each $i \neq j \in I$, at a point in $\pi_{i,j}$ an individual randomly chosen from the population at j dies and is replaced by offspring of an individual randomly chosen from the population at i.

As in the unstructured Moran model with selection, to determine the types of a sample from the population at time t from the initial types of the population, we must trace back the ancestry of each potential parent at a selective event, giving a branching and coalescing process with spatial structure.

We again consider the weak selection regime by taking a population with N individuals in each deme and selection parameter $s_N = \frac{s}{N}$. Each pair of lineages in deme i is hit by a neutral event and coalesces at rate $1 - s_N$ and each lineage in deme i is hit by a selective event and branches at rate $s_N(N-1)$. A lineage in deme i jumps to a different deme at rate m_{ji}. All other events in the process of potential ancestral lineages of the sample occur at rate $\mathcal{O}(\frac{1}{N})$ for large N.

For large N, the genealogy of a finite sample at time t consisting of n_i individuals in deme i for each $i \in I$ is approximately given by the system of branching and coalescing random walks described as follows.

Definition 1.2.10. *The process* $(\underline{n}(t)) = (n_i(t))_{i\in I}$ *is a continuous time Markov chain taking values in* $\mathbb{N}_0^I$ *with* $n_i(0) = n_i$ $\forall i \in I$ *and transition rates as follows.*

- *For each* $i \in I$, $n_i \mapsto n_i - 1$ *at rate* $\binom{n_i}{2}$.
- *For each* $i, j \in I$ *with* $i \neq j$, $\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$ *at rate* $n_i m_{ji}$.
- *For each* $i \in I$, $n_i \mapsto n_i + 1$ *at rate* sn_i.

As in the case of the Moran model, by considering the proportion of type a individuals in each deme and letting $N \to \infty$, we obtain a diffusion approximation of the island model.

Definition 1.2.11 (Diffusion approximation of the island model with selection)**.** *Suppose* I *is a connected graph (finite or infinite) with vertices labelled* $1, 2, \ldots$. *A population is made up of individuals of two types,* a *and* A, *and is structured into demes at each vertex of* I. *The proportion of* a *individuals in deme* i *at time* t *is denoted by* $p_i(t)$. *The proportions at each deme evolve according to the system of stochastic differential equations*

$$dp_i = \sum_{j\neq i}(m_{ji}p_j - m_{ij}p_i)dt - sp_i(1-p_i)dt + \sqrt{p_i(1-p_i)}dW_i, \tag{1.8}$$

where $(W_i)_{i\in I}$ *are independent standard Brownian motions.*

Here m_{ij} represents that rate at which individuals migrate from deme i to deme j. To see where the drift term in (1.8) comes from, note that the rate of decrease due to migration of the proportion of a individuals in deme i is $\sum_{j\neq i} m_{ij}p_i$ and the rate of increase due to migration of the proportion of a individuals to deme i is $\sum_{j\neq i} m_{ji}p_j$. The $sp_i(1-p_i)dt$ and Brownian motion terms come from reproduction within the deme, just as in the convergence of the Moran model with selection to the Wright-Fisher diffusion with selection in Section 1.2.2. We have the following duality relation.

Proposition 1.2.12. *Suppose that* $(\underline{p}(t), t \geq 0)$ *evolves according to the diffusion approximation of the island model with selection of Definition 1.2.11, and the process* $(\underline{n}(t), t \geq 0)$ *evolves according to the branching and coalescing random walk of Definition 1.2.10. Then for* $\underline{p} \in [0,1]^{\mathbb{N}}$ *and* $\underline{n} \in \mathbb{N}_0^I$*, where* $\underline{n}$ *has only finitely many non-zero components,*

$$\mathbb{E}^{\rightarrow}_{\underline{p}}\left[\prod_{i\in I} p_i(t)^{n_i}\right] = \mathbb{E}^{\leftarrow}_{\underline{n}}\left[\prod_{i\in I} p_i^{n_i(t)}\right].$$

where $\mathbb{E}^{\rightarrow}_{\underline{p}}$ *is the law of* $(\underline{p}(t), t \geq 0)$ *with* $\underline{p}(0) = \underline{p}$ *and* $\mathbb{E}^{\leftarrow}_{\underline{n}}$ *is the law of* $(\underline{n}(t), t \geq 0)$ *with* $\underline{n}(0) = \underline{n}$*.*

The heuristic proof of this result is essentially the same as the argument for Proposition 1.2.6.

1.2.4 The spatial Λ-Fleming-Viot process

For certain populations which are clearly subdivided into regions, a model such as the island model makes sense. However, many biological populations evolve in spatial continua (typically of dimension one or two), so models were developed which attempted to avoid an artificial subdivision of the population into demes. Wright (1943) and Malécot (1948) introduced models of populations evolving in $\mathbb{R}^d$ which shared the following three assumptions:

1. Individuals are distributed randomly with constant expected density everywhere.
2. Each individual, independently, produces a Poisson number of offspring with mean one.
3. Each offspring migrates independently, with the displacements being drawn from (usually) a normal distribution.

However, Felsenstein (1975) observed that these assumptions are inconsistent. A population evolving according to assumptions 2 and 3 develops clumps which gradually become larger and further apart, contradicting assumption 1. Letting the population evolve on the torus does not solve this problem as the population dies out in finite time. Conditioning the total population size to be a constant N to counteract this results in clumping of the population on the torus. Felsenstein called these problems 'the pain in the torus'. In fact Sudbury (1977) showed that clumping only occurs if the spatial motion of an ancestral lineage is recurrent, i.e. if the population evolves in one or two spatial dimensions, correcting a mistake in Felsenstein's original paper; see also Kingman (1977) for a thorough treatment of these issues.

A new general framework for modelling populations evolving in a spatial continuum was proposed in Etheridge (2008). We start by introducing an individual based model, which overcomes the so-called 'pain in the torus' by basing reproduction events on a Poisson point process. This means that the rate at which a given region is affected by reproduction events does not grow with the size of the population in the region and so prevents clumping.

Definition 1.2.13 (Individual based model)**.** *Suppose that the population is initially distributed as a Poisson point process in* $\mathbb{R}^d$*. Let* λ *be a fixed positive constant,* $\mu(dr)$ *a measure on* $(0,\infty)$ *and, for each* $r>0$*, let* $\nu_r(du)$ *be a probability measure on* $[0,1]$ *such that*

$$\int_0^\infty \int_0^1 ur^d(1+r^d)\nu_r(du)\mu(dr) < \infty. \tag{1.9}$$

Let Π *be a Poisson point process on* $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times [0,1]$ *with intensity measure* $dt \otimes dx \otimes \mu(dr)\nu_r(du)$*. The dynamics of the population are as follows.*

If (t,x,r,u) *is a point in* Π*, then at time* t *throw down a ball* $\mathcal{B}_r(x)$ *of radius* r *and centre* x*. If the ball is empty, nothing happens. If not:*

1. *Choose an individual uniformly at random from those in* $\mathcal{B}_r(x)$ *(the 'parent').*
2. *For each individual in* $\mathcal{B}_r(x)$*, kill the individual with probability* u *(independently of other individuals).*
3. *Throw down new individuals which are offspring of the parent individual according to an independent Poisson point process with intensity* $u\lambda\mathbb{1}_{\mathcal{B}_r(x)}(y)dy$*.*

Berestycki et al. (2009b) show that under condition (1.9), the process exists, and there is a critical value of λ above which the process survives and below which it dies out.

The genealogical trees relating a sample from this model are difficult to write down explicitly. The ancestor of an individual cannot be in a region of empty space, and so the locations of other lineages could give information about the possible location of an ancestor. By letting $\lambda \to \infty$, we obtain a limiting model for which we can easily write down the dual process.

We suppose that each individual in the population has a type taken from a compact space K. For each time t and location $x \in \mathbb{R}^d$, $\rho(t, x)$ is a probability measure on K. We think of $\rho(t, x)$ as the distribution of the type of an individual sampled from location x at time t.

Definition 1.2.14 (Spatial Λ-Fleming-Viot process (SLFV))**.** *The spatial Λ-Fleming-Viot process*

$$(\rho(t, x, \cdot), t \geq 0, x \in \mathbb{R}^d)$$

specifies a probability measure on the type space K for $t \geq 0$ and (almost every) $x \in \mathbb{R}^d$. Let $\mu(dr)$ be a measure on $(0, \infty)$ and, for each $r > 0$, let $\nu_r(du)$ be a probability measure on $[0, 1]$ such that

$$\int_0^\infty \int_0^1 u r^d \nu_r(du) \mu(dr) < \infty. \tag{1.10}$$

Let Π be a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times [0, 1]$ with intensity measure $dt \otimes dx \otimes \mu(dr)\nu_r(du)$. The dynamics of the process driven by Π are as follows.

At every point (t, x, r, u) of the Poisson point process Π, we select a point z uniformly at random from $\mathcal{B}_r(x)$ and a type k at random according to $\rho(t-, z, \cdot)$. We then update ρ as follows. For all $y \notin \mathcal{B}_r(x)$, $\rho(t, y, \cdot) = \rho(t-, y, \cdot)$. For all $y \in \mathcal{B}_r(x)$,

$$\rho(t, y, \cdot) = (1 - u)\rho(t-, y, \cdot) + u\delta_k.$$

In fact for each t, $(\rho(t, x), x \in \mathbb{R}^d)$ is only defined up to a Lebesgue null subset of $\mathbb{R}^d$.

The flexibility in defining the measure μ makes it easy to incorporate both small-scale reproduction events and large-scale extinction-recolonisation events in the model. For an event $(t, x, r, u) \in \Pi$, we refer to u, the proportion of types which are resampled, as the *impact* of the event; heuristically small impact corresponds to high local population density as only a small proportion of the population is replaced in a reproduction event.

For $x_1, \ldots, x_n \in \mathbb{R}^d$, in order to determine the joint distribution of $\rho(t, x_1), \ldots, \rho(t, x_n)$, for each $i \in \{1, \ldots, n\}$ we think of sampling an individual at random at location x_i at time t. We then trace back until the most recent event which covered the location x_i for some i. For each i with x_i in the region covered by the event, the probability that the chosen individual at x_i was an offspring of the event is u. Each lineage which was an offspring of the event jumps to the location of the parent of the event. If two or more lineages jump then they had the same ancestor before the event and they coalesce. Continuing this process until time 0 gives us a spatial coalescent dual process; the types of the ancestors at time 0 then determine the types of the sample at time t. Note that as we trace backwards in time, the events in the dual process will occur according to a time reversal of the Poisson point process Π; this is also a Poisson point process with the same intensity measure.

Definition 1.2.15 (SLFV dual). *Let* Π *be a Poisson point process on* $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times [0, 1]$ *with intensity measure* $dt \otimes dx \otimes \mu(dr)\nu_r(du)$*, where* μ *and* ν_r *satisfy* (1.10)*. The coalescent dual process* $(\Xi_t)_{t \geq 0}$ *driven by* Π *is the* $\bigcup_{m \geq 1} (\mathbb{R}^d)^m$*-valued Markov process with dynamics defined as follows. For each* $t \geq 0$, $\Xi_t = (\xi^i_t)_{i=1,\ldots,N_t}$ *for some* $N_t \in \mathbb{N}$*; each* ξ^i_t *is the location of an ancestral lineage at time* t *before the present. At each event* $(t, x, r, u) \in \Pi$*,*

1. *For each* $\xi^i_{t-} \in \mathcal{B}_r(x)$*, independently mark the corresponding lineage with probability* u*.*
2. *If at least one lineage is marked, all marked lineages disappear and are replaced by a single lineage, whose location at time* t *is drawn uniformly at random from within* $\mathcal{B}_r(x)$*.*

If no lineages are marked, then nothing happens.

In Etheridge and Kurtz (2014), using a lookdown construction the SLFV and its dual are constructed simultaneously on the same probability space as the limit as $\lambda \to \infty$ of the individual based model in Definition 1.2.13. The reader is referred to Barton et al. (2013c) for further background on the biological considerations and Barton et al. (2010), Berestycki et al. (2013) for mathematical results on the neutral SLFV model.

We can introduce selection to the SLFV in essentially the same way as for the Moran model. We suppose that the type space consists of just two types, A and a, with relative fitness $1+s$ and s respectively. We let $w_t(x)$ denote the proportion of type a individuals at location x at time t, i.e. $w_t(x) = \rho(t, x, \{a\})$ in the notation of Definition 1.2.14. The spatial Λ-Fleming-Viot process with selection was introduced in Etheridge et al. (2014).

Definition 1.2.16 (Spatial Λ-Fleming-Viot process with selection (SLFVS))**.** *Let* μ *be a measure on* $(0, \infty)$ *and, for each* $r > 0$*, let* ν_r *be a probability measure on* $[0, 1]$ *where* μ *and* ν_r *satisfy* (1.10)*. Let* Π *be a Poisson point process on* $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times (0, 1]$ *with intensity measure* $dt \otimes dx \otimes \mu(dr)\nu_r(du)$*. Let* $s \in (0, 1]$*.*

The spatial Λ-Fleming-Viot process with selection *(SLFVS) driven by* Π *with selection parameter* s *is the process* $(w_t)_{t\geq 0}$ *with dynamics given as follows.*

If $(t, x, r, u) \in \Pi$*, a reproduction event occurs at time* t *within the ball* $\mathcal{B}_r(x)$ *of radius* r *centred on* x*. With probability* $1 - s$ *the event is* neutral*, in which case:*

1. *Choose a parental location* z *uniformly at random within* $\mathcal{B}_r(x)$*, and a parental type,* α*, according to* $w_{t-}(z)$*, that is* $\alpha = a$ *with probability* $w_{t-}(z)$ *and* $\alpha = A$ *with probability* $1 - w_{t-}(z)$*.*
2. *For every* $y \in \mathcal{B}_r(x)$*, set* $w_t(y) = (1-u)w_{t-}(y) + u\mathbb{1}_{\{\alpha=a\}}$*.*

Otherwise the event is selective*, in which case:*

1. *Choose two 'potential' parental locations* z, z' *independently and uniformly at random within* $\mathcal{B}_r(x)$*, and at each of these sites 'potential' parental types* α*,* α'*, according to* $w_{t-}(z), w_{t-}(z')$ *respectively.*
2. *For every* $y \in \mathcal{B}_r(x)$ *set* $w_t(y) = (1-u)w_{t-}(y) + u\mathbb{1}_{\{\alpha=\alpha'=a\}}$*.*

As in the case of the diffusion approximation of the island model with selection in Section 1.2.3, to determine the types of a sample from the population at time t we must trace back the locations of the ancestors of each potential parent at a selective event, giving us a branching and coalescing dual process.

Definition 1.2.17 (Branching and coalescing dual). *Let* Π *be a Poisson point process on* $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times [0,1]$ *with intensity measure* $dt \otimes dx \otimes \mu(dr)\nu_r(du)$, *where* μ *and* ν_r *satisfy* (1.10). *Let* $s \in (0,1]$. *The branching and coalescing dual process* $(\Xi_t)_{t\geq 0}$ *driven by* Π *is the* $\bigcup_{m\geq 1}(\mathbb{R}^d)^m$*-valued Markov process with dynamics defined as follows. For each* $t \geq 0$, $\Xi_t = (\xi_t^i)_{i=1,\dots,N_t}$ *for some* $N_t \in \mathbb{N}$*; each* ξ_t^i *is the location of a potential ancestral lineage at time* t *before the present. At each event* $(t,x,r,u) \in \Pi$, *with probability* $1-s$, *the event is neutral:*

1. *For each* $\xi_{t-}^i \in \mathcal{B}_r(x)$, *independently mark the corresponding lineage with probability* u.
2. *If at least one lineage is marked, all marked lineages disappear and are replaced by a single lineage, whose location at time* t *is drawn uniformly at random from within* $\mathcal{B}_r(x)$.

Otherwise the event is selective:

1. *For each* $\xi_{t-}^i \in \mathcal{B}_r(x)$, *independently mark the corresponding lineage with probability* u.
2. *If at least one lineage is marked, all marked lineages disappear and are replaced by* two *lineages, whose locations at time* t *are drawn independently and uniformly from within* $\mathcal{B}_r(x)$.

In both cases, if no lineages are marked, then nothing happens.

See Figure 1.5 for an illustration of this dual process. We have an analogue of the moment duality from Propositions 1.2.6 and 1.2.12. However, to state a moment duality formally for the SLFVS, we would need to be able to identify $\mathbb{E}[\prod_{i=1}^n w_t(x_i)]$ for any choice of points $x_1, \dots, x_n \in \mathbb{R}^d$. The difficulty is that $w_t(x)$ is only defined at Lebesgue almost every point x and so we only have a 'weak' moment duality.

Proposition 1.2.18 (Etheridge et al. (2014)). *The spatial* Λ*-Fleming-Viot process with selection,* $(w_t)_{t\geq 0}$, *is dual to the branching and coalescing dual process* $(\Xi_t)_{t\geq 0}$ *in the*

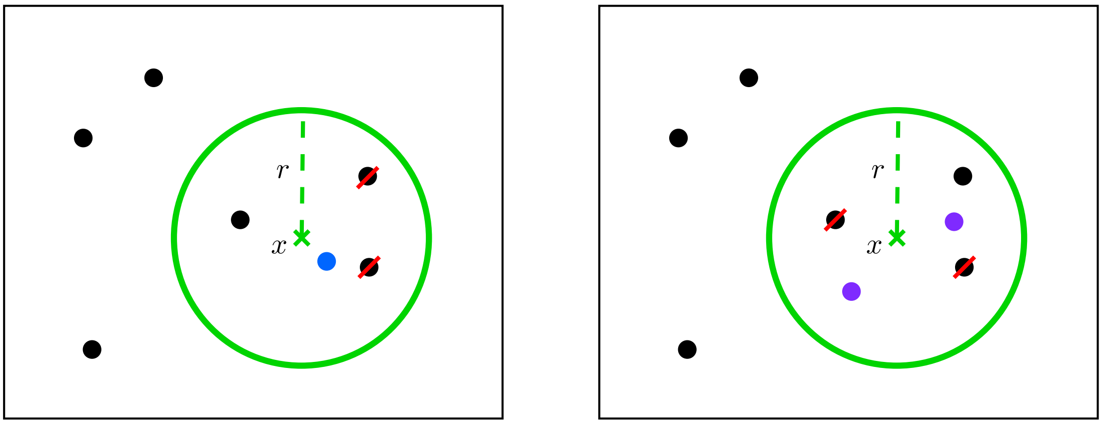

Figure 1.5: On the left is an illustration of the effect of a neutral event $(t, x, r, u) \in \Pi$ on the backwards-in-time dual process of Definition 1.2.17. The black dots show the locations of the potential ancestors at time $t-$. Of the three lineages in $\mathcal{B}_r(x)$, each is marked independently with probability u; the marked lineages are crossed out in red. The blue dot shows the location of the new lineage sampled uniformly at random from $\mathcal{B}_r(x)$. On the right is an illustration of a selective event $(t, x, r, u) \in \Pi$. Here the purple dots show the locations of the two new lineages sampled independently uniformly at random from $\mathcal{B}_r(x)$.

sense that for every $k \in \mathbb{N}$ and $\psi \in C_c((\mathbb{R}^d)^k)$, we have

$$\mathbb{E}^{\rightarrow}_{w_0}\left[\int_{(\mathbb{R}^d)^k} \psi(x_1, \dots, x_k)\left\{\prod_{j=1}^{k} w_t(x_j)\right\} dx_1 \dots dx_k\right] = \int_{(\mathbb{R}^d)^k} \psi(x_1, \dots, x_k)\mathbb{E}^{\leftarrow}_{\{x_1,\dots,x_k\}}\left[\prod_{j=1}^{N_t} w_0(\xi_t^j)\right] dx_1 \dots dx_k,$$

where $\mathbb{E}^{\rightarrow}_{w_0}$ is the law of $(w_t)_{t\geq 0}$ with initial value w_0, and $\mathbb{E}^{\leftarrow}_{\{x_1,\dots,x_k\}}$ is the law of $(\Xi_t)_{t\geq 0}$ with initial value $\Xi_0 = (x_1, \dots, x_k)$.

1.3 Selection in populations with spatial structure

In Chapters 3 and 4 of this thesis, we study the relative strength of genetic drift, natural selection and migration in a spatially structured population in two different settings. In each case we model the population using the SLFV.

1.3.1 BBM and selection in the SLFV

In Chapter 3 (Etheridge et al. (2015a)), we consider the effect of selection on a spatially structured population when the effective local population density is small. We model this situation by considering the SLFVS from Definition 1.2.16 with constant impact, i.e. $\nu_r = \delta_u \ \forall r > 0$, for some constant $u \in (0,1]$. We also assume that reproduction events have bounded radius, so μ is a finite measure on $(0, \mathcal{R}]$ for some $\mathcal{R} < \infty$.

In order to observe an effect of selection on the genealogical trees from the patterns of genetic variation in the 'present day' population, there must be a large number of neutral mutations affecting the genealogical trees. This means looking at the dual process of Defintion 1.2.17 over long timescales. Each lineage in this dual process follows a pure jump process with jumps of bounded magnitude. If we scale time by $1/n$, in order to see a non-trivial limit for the motion of a single lineage as $n \to \infty$ we must scale space by $1/\sqrt{n}$. We then find that the selection parameter s_n at the nth stage of rescaling required to obtain a non-trivial limit for the rescaled branching and coalescing dual process on $\mathbb{R}^d$ is given by

$$s_n = \begin{cases} \frac{\log n}{n} & d = 2, \\ \\ \frac{1}{n} & d \geq 3. \end{cases} \tag{1.11}$$

(The case $d = 1$ is covered in Etheridge et al. (2015b).) This is in contrast to the case where the impact at the nth stage of rescaling is $u_n \to 0$ as $n \to \infty$ (i.e. the effective local population density tends to infinity). This case was studied in Etheridge et al. (2014); here the scaling of the selection parameter required is independent of spatial dimension and the same as for the weak selection limit without spatial structure discussed in Section 1.2.2. In our results, from (1.11) we see that in dimension 2, selection must be stronger than in a population without spatial structure in order for us to detect it. This reflects strong local competition in the population.

More formally, our main result is as follows. For each $n \in \mathbb{N}$, define the measure μ^n by $\mu^n(B) = \mu(n^{1/2}B)$, for all Borel subsets B of $\mathbb{R}_+$. At the nth stage of the rescaling, our rescaled dual is driven by the Poisson point process Π^n on $\mathbb{R} \times \mathbb{R}^d \times (0, \mathcal{R}/\sqrt{n}\,] \times [0,1]$ with intensity

$$n\, dt \otimes n^{d/2}\, dx \otimes \mu^n(dr)\delta_u(dv). \tag{1.12}$$

We write $\mathcal{P}^{(n)}(p)$ for the dual process of Definition 1.2.17, rescaled as in (1.12) and (1.11), started from a single individual at the point $p \in \mathbb{R}^d$ and viewed as a collection of paths. Each path traces out a ‘potential ancestral lineage’: each time it is affected by a neutral event it jumps to the location of the parent, and if a lineage is affected by a selective event then it jumps to the location of (either) one of the ‘potential parents’. Precisely one potential ancestral lineage follows each possible route through the branching and coalescing dual process.

We denote by BBM(p, V) binary branching Brownian motion started from the point $p \in \mathbb{R}^d$, with branching rate V and diffusion constant given by

$$\sigma^2 = \tfrac{1}{d} \int_{\mathbb{R}^d} \int_0^\infty |z|^2 u \frac{V_r(0,z)}{V_r} \mu(dr)\, dz.$$

In other words, during their lifetime, which is exponentially distributed with parameter V, individuals follow d-dimensional Brownian motion with diffusion constant σ^2. We view BBM(p, V) as a set of (continuous) paths, each starting at p, with precisely one path following each possible distinct sequence of branches.

We define the events

$$\mathcal{D}_n(\varepsilon, T) = \left\{ \forall l \in \mathcal{P}^{(n)}(p),\ \exists l' \in \mathrm{BBM}(p, V) : \sup_{t\in[0,T]} |l(t) - l'(t)| \leq \varepsilon \right\},$$

$$\mathcal{D}'_n(\varepsilon, T) = \left\{ \forall l \in \mathrm{BBM}(p, V),\ \exists l' \in \mathcal{P}^{(n)}(p) : \sup_{t\in[0,T]} |l(t) - l'(t)| \leq \varepsilon \right\}.$$

Theorem 1.3.1. *Let $d \geq 2$. There exists $V \in (0, \infty)$ such that the following holds. Let $T < \infty$, $p \in \mathbb{R}^2$; then given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that, for all $n \geq N$ there is a coupling between BBM(p, V) and $\mathcal{P}^{(n)}(p)$ with $\mathbb{P}[\mathcal{D}_n(\varepsilon, T) \cap \mathcal{D}'_n(\varepsilon, T)] \geq 1 - \varepsilon$.*

The proof relies on showing that although the rate of branching in the rescaled dual process is $\log n$, almost all the new branches coalesce almost instantly and so we do not see them in the limit, giving effectively a finite, positive branching rate.

1.3.2 Hybrid zones and curvature flow

In Chapter 4 (Etheridge et al. (2016)), we study the effect of a different form of selection. The motivation for this work is the study of hybrid zones. These occur when individuals from two genetically distinct groups are able to reproduce. We suppose that the genetics is controlled by a single gene which appears in two forms (known as *alleles*) a and A.

We also suppose that individuals are *diploid* (i.e. they each have two copies of the gene); each individual is then of type aa, aA or AA. We suppose that the *heterozygous* aA individuals have a lower fitness than the *homozygous* aa and AA individuals. The region in which both alleles a and A are present in significant numbers is known as a hybrid zone.

In order to study the behaviour of the hybrid zone, we model this situation using the SLFV process. We suppose that aa, aA and AA individuals have relative fitnesses 1, $1-s$ and 1 respectively, where $s > 0$ is small. If the proportion of a-alleles in the parent population is $\bar{w}$, then we assume that the population is in Hardy-Weinberg proportions, i.e. the proportion of parent individuals of type aa, aA and AA are $\bar{w}^2$, $2\bar{w}(1-\bar{w})$ and $(1-\bar{w})^2$ respectively. These parents produce a large number of genetic copies of themselves, which split into a large pool of gametes (cells containing one copy of each gene). The gametes then fuse to produce offspring. To reflect the relative fitnesses of the parent population, a type aA parent produces $1-s$ times as many gametes as a type aa or AA parent. Then if α is the type of a randomly sampled offspring gene, we should have

$$\frac{\mathbb{P}\left[\alpha = a\right]}{\mathbb{P}\left[\alpha = A\right]} = \frac{\bar{w}^2 + \bar{w}(1-\bar{w})(1-s)}{(1-\bar{w})^2 + \bar{w}(1-\bar{w})(1-s)}.$$

It follows that

$$\begin{aligned}\mathbb{P}\left[\alpha = a\right] = \frac{\bar{w}^2 + \bar{w}(1-\bar{w})(1-s)}{1-2s\bar{w}(1-\bar{w})} &= (1-s)\bar{w} + s(3\bar{w}^2 - 2\bar{w}^3) + \mathcal{O}(s^2) \\ &= (1-s)\bar{w} + s(\bar{w}^3 + 3\bar{w}^2(1-\bar{w})) + \mathcal{O}(s^2). \quad (1.13)\end{aligned}$$

We assume that s is small and neglect the $\mathcal{O}(s^2)$ term. Note that the first term in (1.13) is $1-s$ times the probability that a randomly chosen allele from the parent population is a, and the second term is s times the probability that the majority of three independently randomly chosen alleles from the parent population are a. This gives us two types of event to incorporate into the SLFV as follows.

Definition 1.3.2 (Spatial Λ-Fleming-Viot with selection against heterozygosity)**.** *Fix* $u \in (0,1]$ *and* $\mathcal{R} \in (0,\infty)$. *Let* μ *be a finite measure on* $(0,\mathcal{R}]$. *Let* Π *be a Poisson point process on* $\mathbb{R}_+ \times \mathbb{R}^2 \times (0,\mathcal{R}]$ *with intensity measure* $dt \otimes dx \otimes \mu(dr)$.

The spatial Λ-Fleming-Viot process with selection against heterozygousity *driven by* Π *with selection parameter* s *is the process* $(w_t)_{t\geq 0}$ *with dynamics given as follows.*

If $(t,x,r)\in\Pi$*, a reproduction event occurs at time* t *within the ball* $\mathcal{B}_r(x)$ *of radius* r *centred on* x*. With probability* $1-s$ *the event is* neutral*, in which case:*

1. *Choose a parental location* z *uniformly at random within* $\mathcal{B}_r(x)$*, and a parental type,* α_0*, according to* $w_{t-}(z)$*, that is* $\alpha_0=a$ *with probability* $w_{t-}(z)$ *and* $\alpha_0=A$ *with probability* $1-w_{t-}(z)$*.*
2. *For every* $y\in\mathcal{B}_r(x)$*, set* $w_t(y)=(1-u)w_{t-}(y)+u\mathbb{1}_{\{\alpha_0=a\}}$*.*

Otherwise the event is selective*, in which case:*

1. *Choose three 'potential' parental locations* z_1,z_2,z_3 *independently and uniformly at random within* $\mathcal{B}_r(x)$*, and at each of these sites 'potential' parental types* α_1*,* α_2*,* α_3 *according to* $w_{t-}(z_1),w_{t-}(z_2),w_{t-}(z_3)$ *respectively. Let* $\widehat{\alpha}$ *denote the most common allelic type in* $\alpha_1,\alpha_2,\alpha_3$*.*
2. *For every* $y\in\mathcal{B}_r(x)$ *set* $w_t(y)=(1-u)w_{t-}(y)+u\mathbb{1}_{\{\widehat{\alpha}=a\}}$*.*

Let $\bar{w}_s(x,r)$ denote the average of $w_s(z)$ over the ball $\mathcal{B}_r(x)$. Then if $(t,x,r)\in\Pi$ is an event and $\bar{w}_{t-}(x,r)=\bar{w}$, the expected increment in $\bar{w}(x,r)$ at time t is

$$us(\bar{w}^3+3\bar{w}^2(1-\bar{w})-\bar{w})=us\bar{w}(1-\bar{w})(2\bar{w}-1).$$

We can apply a modified version of the results of Etheridge et al. (2014) to this situation. If we rescale the SLFV from Definition 1.3.2 so that at the nth stage of rescaling, for some fixed $\beta\in(0,1/3)$, the impact is given by $u_n=un^{2\beta-1}$, the selection parameter is $s_n=\rho n^{-2\beta}$ and we rescale time by $1/n$ and space by $1/n^{\beta}$, as $n\to\infty$ the rescaled process converges in a weak sense to the solution of the partial differential equation

$$\frac{\partial v}{\partial t}=\frac{m}{2}\Delta v+\alpha\rho v(1-v)(2v-1), \tag{1.14}$$

where $m,\alpha>0$ are explicit constants depending on u and μ. This is a special case of the Allen-Cahn equation; it turns out that solutions to this equation for large ρ are given

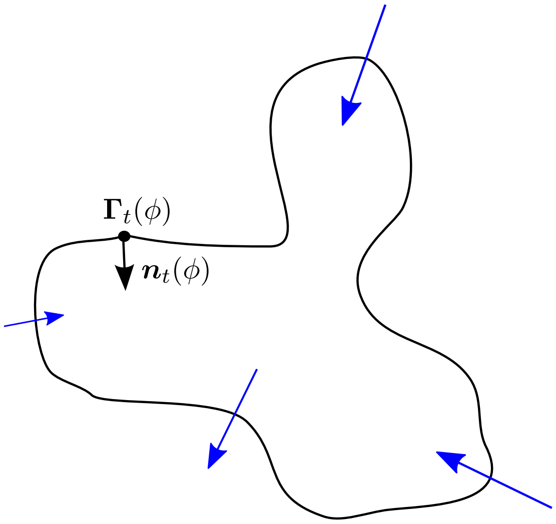

Figure 1.6: This is an illustration of the effect of curvature flow on the simple closed curve Γ_t. The rate of motion of the curve at a point is given by the inwards normal times the curvature at the point; this is indicated by the blue arrows.

approximately by the indicator function of a set whose boundary evolves according to a process known as curvature flow.

Definition 1.3.3 (Curvature flow)**.** *Let* $\mathbf{\Gamma} = (\mathbf{\Gamma}_t(\cdot))_t$ *be a family of simple closed smooth curves, indexed by* $t \in [0, \mathscr{T})$*. Let* $\boldsymbol{n} = \boldsymbol{n}_t(\phi)$ *denote the unit (inward) normal vector to* $\mathbf{\Gamma}_t$ *at* ϕ *and let* $\kappa = \kappa_t(\phi)$ *denote the curvature of* $\mathbf{\Gamma}_t$ *at* ϕ*. We say that* $\mathbf{\Gamma}$ *is a* motion by curvature flow *if*

$$\frac{\partial \mathbf{\Gamma}_t(\phi)}{\partial t} = \kappa_t(\phi)\boldsymbol{n}_t(\phi).$$

for all t, ϕ*.*

See Figure 1.6 for an illustration of curvature flow. If $\mathbf{\Gamma}_0$ is a simple closed smooth curve, then results of Gage and Hamilton (1986) and Grayson (1987) show that curvature flow starting from $\mathbf{\Gamma}_0$ is well-defined up to a finite time $\mathscr{T}$ when it shrinks to a point.

Now let $v^\varepsilon : \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}$ denote the solution of

$$\frac{\partial v^\varepsilon}{\partial t} = \Delta v^\varepsilon + \frac{1}{\varepsilon^2} v^\varepsilon (1 - v^\varepsilon)(2v^\varepsilon - 1), \qquad v^\varepsilon(0, x) = p(x). \tag{1.15}$$

We assume that the initial condition, p, of (1.15) takes values in $[0, 1]$, and that

$$\Gamma := \left\{ x \in \mathbb{R}^2 : p(x) = \frac{1}{2} \right\}$$

is a simple closed smooth curve. Then curvature flow started from Γ, which we denote $(\mathbf{\Gamma}_t(\cdot))_t$, exists up to some finite time $\mathscr{T}$; let $d(x, t)$ be the signed distance from x to

$\boldsymbol{\Gamma}_t$, chosen to be negative inside $\boldsymbol{\Gamma}_t$ and positive outside. We require some additional regularity conditions on p which we do not discuss here (see Section 4.1.1 for details). The following result is a special case of Theorem 3 in Chen (1992); the original observation was made by Allen and Cahn (1979). We provide a new probabilistic proof in Chapter 4.

Theorem 1.3.4. *Let v^ε solve (1.15) with initial condition p, and define $\mathscr{T}$, $d(x,t)$ as above. Fix $T^* \in (0, \mathscr{T})$. Let $k \in \mathbb{N}$. There exists $\varepsilon_2(k) > 0$, and $a_2(k), c_2(k) \in (0,\infty)$ such that for all $\varepsilon \in (0, \varepsilon_2)$ and t satisfying $a_2\varepsilon^2|\log\varepsilon| \leq t \leq T^*$,*

1. *for x such that $d(x,t) \geq c_2\varepsilon|\log\varepsilon|$, we have $v^\varepsilon(t,x) \geq 1 - \varepsilon^k$.*
2. *for x such that $d(x,t) \leq -c_2\varepsilon|\log\varepsilon|$, we have $v^\varepsilon(t,x) \leq \varepsilon^k$.*

Remark 1.3.5. *This result holds for solutions of* (1.15) *in d spatial dimensions if (two-dimensional) curvature flow is replaced with d-dimensional mean curvature flow. We are restricting to the two-dimensional case in this introduction because it is the biologically relevant one (and it is easier to visualise), but we shall cover the general d-dimensional results in Chapter 4.*

Recall from Theorem 1.1.2 that solutions to the partial differential equation

$$\frac{\partial v}{\partial t} = \frac{1}{2}\Delta v + V(\Phi(v) - v),$$

where $V > 0$ and Φ is the probability generating function of an offspring distribution, can be written in terms of a duality relation with a branching Brownian motion. The partial differential equation in (1.15) is not of this form, but we can find a different duality relation with a branching Brownian motion which enables us to give a purely probabilistic proof of Theorem 1.3.4.

The duality involves a ternary branching Brownian motion: a BBM in which each individual branches into three after an exponential lifetime. We write $(\boldsymbol{W}(t), t \geq 0)$ for the historical process of the BBM (which traces out the space-time trees that record the location of all the individuals alive at time s for all $s \in [0,t]$). We let $W_1(t), W_2(t), \ldots, W_{N(t)}(t)$ denote the locations of the individuals alive at time t, and write $\mathcal{T}(\boldsymbol{W}(t))$ for the tree which records the genealogy of the BBM up to time t.

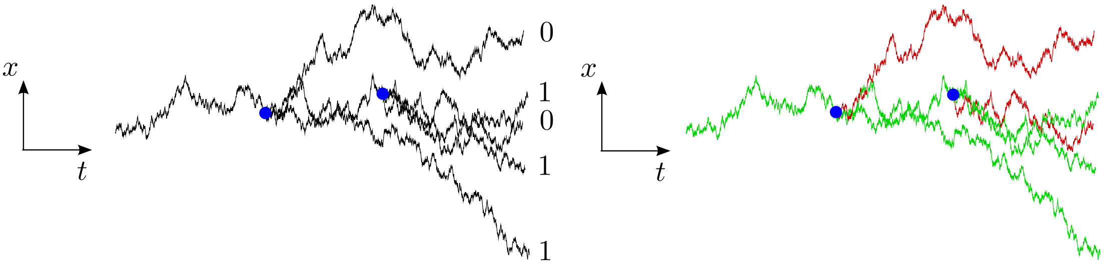

Figure 1.7: This is an illustration of the majority voting procedure on a ternary BBM. For clarity we use a one-dimensional BBM instead of a two-dimensional BBM. On the left is the ternary BBM run until some time t (with blue dots at the branching events), and the votes of each individual in the BBM at time t (sampled independently according to the function p at their locations). On the right we trace the votes inwards to the root using the majority voting procedure; vote 1 is shown in green and vote 0 is shown in red. In this case $\mathbb{V}_p(\boldsymbol{W}(t)) = 1$.

For a fixed function $p : \mathbb{R}^2 \to [0,1]$, we define a majority voting procedure on $\mathcal{T}(\boldsymbol{W}(t))$ as follows.

1. Each individual $i \in \{1, \ldots, N(t)\}$, independently, votes 1 with probability $p(W_i(t))$ and otherwise votes 0.
2. At each branch point in $\mathcal{T}(\boldsymbol{W}(t))$, the vote of the parent is the majority vote of the votes of its three offspring.

This defines an iterative voting procedure, which runs inwards from the leaves of $\mathcal{T}(\boldsymbol{W}(t))$ to the root $\emptyset$. We define $\mathbb{V}_p(\boldsymbol{W}(t))$ to be the vote associated to the root $\emptyset$. See Figure 1.7 for an illustration of this voting procedure.

For $x \in \mathbb{R}^2$, we write $\mathbb{P}_x^\varepsilon$ for the probability measure under which $(\boldsymbol{W}(t), t \geq 0)$ has the law of the historical process of ternary branching Brownian motion in $\mathbb{R}^2$ with branching rate $1/\varepsilon^2$ started from a single individual at location x at time 0.

Theorem 1.3.6. *Let* $p : \mathbb{R}^2 \to [0,1]$. *Then*

$$v^\varepsilon(t,x) = \mathbb{P}_x^\varepsilon \left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1 \right]$$

is a solution to equation (1.15) *with initial condition* $v^\varepsilon(0,x) = p(x)$.

This allows us to give a probabilistic proof of Theorem 1.3.4 by studying the majority voting procedure on branching Brownian motion. The dual process for the SLFV with selection against heterozygosity is a branching and coalescing process in which particles branch into three; instead of the weak moment duality in Proposition 1.2.18, the duality relation involves carrying out a majority voting procedure on the dual process. Our proof of Theorem 1.3.4 can be adapted to this setting. We rescale SLFV with selection against heterozygosity from Definition 1.3.2 so that at the nth stage of rescaling, for some fixed $\beta \in (0, 1/3)$, the impact is given by $u_n = un^{2\beta-1}$, the selection parameter is $s_n = \varepsilon_n^{-2} n^{-2\beta}$ and time is rescaled by $1/n$ and space by $1/n^\beta$; we call this process $(w_t^n)_{t \geq 0}$. We show that if $\varepsilon_n \to 0$ (sufficiently slowly) as $n \to \infty$ then for large n $\mathbb{E}\left[w_t^n(x)\right]$ is approximately the indicator function of a set whose boundary evolves according to curvature flow.

1.4 Outline

The rest of this thesis consists of three self-contained pieces of research: Addario-Berry and Penington (2015) in Chapter 2, Etheridge et al. (2015a) in Chapter 3 and Etheridge et al. (2016) in Chapter 4. In the conclusion we shall briefly discuss some further work and future directions.

Chapter 2

Branching Brownian motion with decay of mass

Abstract

We augment standard branching Brownian motion by adding a competitive interaction between nearby particles. Informally, when particles are in competition, the local resources are insufficient to cover the energetic cost of motion, so the particles' masses decay. In standard BBM, we may define the *front displacement* at time t as the greatest distance of a particle from the origin. For the model with masses, it makes sense to instead define the front displacement as the distance at which the local mass density drops from $\Theta(1)$ to $o(1)$. We show that one can find arbitrarily large times t for which this occurs at a distance $\Theta(t^{1/3})$ behind the front displacement for standard BBM.

2.1 Introduction

In this work, we propose a mathematical model of competition for resources within a single species, in a growing, spatially structured population, and provide an initial study of the front location in this new setting. The model is essentially standard one-dimensional branching Brownian motion (BBM), augmented with a destructive, local interaction between particles. We first briefly recall BBM: start from a single particle at a point in $\mathbb{R}$, endowed with an Exp(1) “branching clock”. The particle moves according to Brownian motion; when its clock rings, it splits in two (branches). The new particles receive independent Exp(1) clocks, and move independently (according to Brownian motion) starting from where the first particle splits, until their own clocks ring and they in turn split, *et cetera.*

Write $N(t)$ for the total number of particles at time t, and $\mathbf{X}(t) = (X_i(t), 1 \le i \le N(t))$ for the locations of such particles. We assume the particles are listed in a way that makes the vector $\mathbf{X}(t)$ exchangeable conditional on $N(t)$; one possible formalism is via the Ulam-Harris tree, with particles listed lexicographically according to their label in the tree. We refer the reader to Harris and Roberts (2015) for more details on such matters, but many different references are possible. We also write $N(t,x) = \{i : X_i(t) \ge x\}$ for the indices of particles with position greater than x at time t.

We write $(X_{i,t}(s), 0 \le s \le t)$ for the ancestral path leading to $X_i(t)$, and let $j_{i,t}(s)$ be the index of $X_{i,t}(s)$ among the time-s population, so that $X_{i,t}(s) = X_{j_{i,t}(s)}(s)$. We sometimes write $(X_i(t), i \ge 1)$, ignoring the fact that $\mathbf{X}(t)$ has finite length, for convenience. We adopt the convention that $X_k(t) = \partial$ for $k > N(t)$ (so ∂ is where new babies come from). We refer to “the particle $X_i(t)$” as shorthand for “the particle with position $X_i(t)$ at time t”; this is unambiguous at Leb-a.e. time t. We write $\mathbb{P}$ for the probability measure under which $(\mathbf{X}(t), t \ge 0)$ has the law of one-dimensional BBM with initial individual at 0, $\mathbb{E}$ for the corresponding expectation, and $(\mathcal{F}_t, t \ge 0)$ for the filtration generated by the process.

We now add destructive interaction as follows. Informally, imagine that the particles are, say, amoeba. Motion has an energetic cost, but for a single particle in isolation, this cost is exactly accounted for by the resources available in the environment. When particles are nearby, however – at distance less than one, say – they must share resources; in this case individuals do not consume enough to meet their energy expenditure, and their mass decreases. Finally, larger (more massive) individuals consume resources at a greater rate.

Formally, we define a vector $\mathbf{M}(t) = (M_i(t), i \geq 0)$, and call $M_i(t)$ the *mass* of particle $X_i(t)$. By convention, if $X_i(t) = \partial$ then $M_i(t) = 0$. Write

$$\zeta(t, x) = \sum_{\{i:|X_i(t)-x|\in(0,1)\}} M_i(t)$$

for the total mass of particles within distance one of x at time t, excluding any particles at position x. Then at time t, $M_i(t)$ decays at rate $M_i(t) \cdot \zeta(t, X_i(t))$. In other words, $\mathrm{d}M_i(t) = -M_i(t) \cdot \zeta(t, X_i(t))\mathrm{d}t$, so

$$M_i(t) = \exp(-\int_0^t \zeta(s, X_{i,t}(s))\mathrm{d}s)\,.$$

This should be viewed as defining $(\mathbf{M}(t), t \geq 0)$ to be the solution of a system of differential equations; the definition makes sense since the system has a unique solution $\mathbb{P}$-almost surely. Furthermore, the process $(\mathbf{M}(t), t \geq 0)$ is clearly $\mathcal{F}_t$-adapted.

We write $M_{i,t}(s)$ for the mass of the ancestor of $X_i(t)$ at time s (so $M_{i,t}(s) = M_{j_{i,t}(s)}(s)$). Note that along any given trajectory, mass decreases: $(M_{i,t}(s), 0 \leq s \leq t)$ is decreasing in s for each $1 \leq i \leq N(t)$. Mass enters the system through branching events, since each "child particle" inherits the mass of its parent. This is obviously physically unrealistic in some settings (e.g. for amoebae) but may be more realistic in others (e.g. in nuclear physics).

Rather than viewing $M_i(t)$ as a mass, a perspective suggested by a referee is to view $(M_{i,t}(s), 0 \leq s \leq t)$ as recording information about the local density of the environment observed along the ancestral trajectory of the particle $X_i(t)$. The interaction between the dynamics of $\mathbf{X}(t)$ and $\mathbf{M}(t)$ makes this point of view slightly complicated to interpret, but here is one possibility. Imagine adding destructive interaction to a BBM, as follows: whenever two different particles are at distance less than 1, each kills the other at rate one. Record such a killing event as a mark at the appropriate location of the BBM family tree. Particles with a mark on their ancestral trajectory are *ghosts*, which continue to move and reproduce as before, but can no longer kill other particles. Given the BBM but not the marks, one may ask for the conditional survival probabilities

$p_i(t) = \mathbb{P}[X_i(t) \text{ is alive } \mid \mathcal{F}_t]$ of the particles. The vector $\mathbf{M}(t)$ is a "linearized" version of the vector of these survival probabilities.

2.1.1 Main result

Write

$$d(t,m) = \min\{x > 0 : \zeta(t,x) < m\}, \quad D(t,m) = \max\{x : \zeta(t,x) > m\},$$

for the leftmost (positive) location at which the total mass of nearby particles falls below m, and the rightmost location at which it exceeds m, respectively. We prove the following theorem.

Theorem 2.1.1. *Write $c^* = 3^{4/3}\pi^{2/3}/2^{7/6}$. Then almost surely, for all $m < 1$,*

$$\limsup_{t\to\infty} \frac{\sqrt{2}t - d(t,m)}{t^{1/3}} \geq c^* \quad \textit{and} \quad \liminf_{t\to\infty} \frac{\sqrt{2}t - D(t,m)}{t^{1/3}} \leq c^*$$

A well-known result of Bramson (1978) states that the rightmost particle location $\max_{i\geq 1} X_i(t)$ has median $\operatorname{med}(t)$ satisfying

$$\operatorname{med}(t) = \sqrt{2}t - \frac{3}{2^{3/2}} \log t + \mathcal{O}(1).$$

Furthermore, it turns out (Hu and Shi (2009)) that $|\max_{i\geq 1} X_i(t) - \operatorname{med}(t)|$ is almost surely $\mathcal{O}(\log t)$, in that $\limsup_{t\to\infty} |\max_{i\geq 1} X_i(t) - \operatorname{med}(t)|/\log t$ is a.s. finite. In view of this, the theorem states that (1) there are arbitrarily large times t at which the first low-density region lags at least distance $c^* t^{1/3} + o(t^{1/3})$ behind the rightmost particle, and (2) there are also (potentially different) arbitrarily large times t at which there is some high-density region within distance $c^* t^{1/3} + o(t^{1/3})$ of the rightmost particle.

We believe that in fact almost surely, for all $m < 1$,

$$\lim_{t\to\infty} \frac{\sqrt{2}t - d(t,m)}{t^{1/3}} = c^* = \lim_{t\to\infty} \frac{\sqrt{2}t - D(t,m)}{t^{1/3}}.$$

If this is correct, then the front could equivalently be defined as, e.g., a median of $D(t,m)$ or $d(t,m)$ – or any other fixed quantile of one of these random variables. We provide some justification for our belief in Section 2.6. That section also contains a few open questions about the model and a discussion of various generalizations of our results (some straightforward, some conjectural), as well as describing variants of the model which have thus far resisted analysis.

2.2 Proof sketch

Here comes an outline of the key tools in our argument. The first is technical but important and also, we believe, provides important intuition when making heuristic predictions about the behaviour of the process. The remainder gives a fairly detailed overview of the proof.

Density self-correction. It is not hard to see that when $\zeta(t,x)$ is small (much less than one), and this also holds in a region around x, then $\zeta(t,\cdot)$ will exhibit exponential growth near x, at least for a short time. Indeed, we heuristically have

$$\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t,x) \approx \zeta(t,x) - \sum_{\{i:|X_i(t)-x|\in(0,1)\}} M_i(t)\cdot\zeta(t,X_i(t))\,.$$

This is not exactly correct since it ignores the effect of motion (particles may enter or leave the region near x), but it is a useful first approximation. In particular, it suggests that if $\zeta(t,y)$ is small (much less than one) for all y with $|y-x|<1$, then $\zeta(t,\cdot)$ will exhibit exponential growth near x, at least for a short time. This is indeed true; one important consequence is that if $\zeta(t,x)=\varepsilon$ and $\zeta(t,\cdot)$ is not too wild then it is very likely that $\zeta(t',x)=\Theta(1)$ for some $t'=t+\Theta(\log(1/\varepsilon))$. Similarly, when $\zeta(t,y)$ is much larger than 1 for y near x then $\zeta(t,x)$ will decrease exponentially quickly. We use the self-correcting nature of the density in several places throughout the chapter.

As an aside, we remark that if $\zeta(t,y)\approx\zeta(t,x)$ for $|y-x|<1$ then the above heuristic gives $\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t,x)\approx\zeta(t,x)(1-\zeta(t,x))$, which is suggestive of the logistic control; we briefly revisit this connection in Section 2.6.

Population + no competitors=mass. Fix $\beta>0$ and suppose that for some function $f:[0,\infty)\to\mathbb{R}$, for all $s\in[0,t]$, $D(s,\beta)<f(s)$, or in other words $\zeta(s,x)\leq\beta$ for all $x\geq f(s)$. In this case, particles that stay ahead of the moving barrier f are in

a relatively sparse environment, so do not lose mass too quickly. More precisely, if $X_i(t)$ satisfies $X_{i,t}(s) > f(s)$ for all $s \in [0,t]$ then $M_i(t) \geq e^{-\beta t}$. It follows that for any $x \geq f(t)+1$,

$$\zeta(t,x) \geq e^{-\beta t} \cdot \#\{i : |X_i(t) - x| < 1, \forall s \in [0,t], X_{i,t}(s) > f(s)\}\,.$$

For such x, if $\#\{i : |X_i(t) - x| < 1, \forall s \in [0,t], X_{i,t}(s) > f(s)\} > \beta e^{\beta t}$ then $\zeta(t,x) > \beta$, contradicting the assumption that $D(t,\beta) \leq f(t)$.

Surfing the wave. To exploit the above contradiction, we require that with high probability there are many particles staying ahead of some barrier. Such results are available: it follows fairly straightforwardly from recent studies of *consistent maximal displacement* for BBM (Roberts (2015)) that for $c > c^*$, for all large times t there are $e^{\Theta(t^{1/3})}$ particles at time t which have stayed ahead of the curve $f(s) = \sqrt{2}s - cs^{1/3}$. This allows us to take $\beta = t^{-1}$ above and obtain that there is $s \in [0,t]$ and $x \geq f(s)$ such that $\zeta(s,x) > t^{-1}$. Since the local density grows exponentially in regions with small density, we will with high probability find s' with $\zeta(s',x) > b > 0$ and $s' - s = \mathcal{O}(\log t)$. Since $x \geq f(s) \geq f(s') - \mathcal{O}(\log t)$, it follows that $D(s',b) \geq \sqrt{2}s' - c(s')^{1/3} - \mathcal{O}(\log t)$.

The lower bound is practically complete, but we must rule out the possibility that $s' = \mathcal{O}(1)$ for all t. To do so, we first establish that

$$\sup_{t>0} \frac{\max\{\zeta(t,x), x \in \mathbb{R}\}}{\log(t+2)} =: Z < \infty \quad \text{almost surely}\,.$$

Proving this is harder than might be expected; its proof, given in Section 2.4, is perhaps the most technically challenging part of the proof of the lower bound.

Once we prove that $Z < \infty$, we then reprise the above argument, but with a variable mass bound

$$\beta = \beta(s) = \begin{cases} Z\log(s+2) & \text{for } s \leq t^{1/4} \\ t^{-1} & \text{for } s \in (t^{1/4}, t]\,. \end{cases}$$

The loss of mass before time $t^{1/4}$ is insignificant compared with that which follows, so essentially the same argument as above yields that there is $s \in [0,t]$ and $x \geq f(s)$ such that $\zeta(s,x) > \beta(s)$. On the other hand, this can not happen for $s < t^{1/4}$ by the definition of Z, so it must happen later. This is enough to conclude the lower bound. The details of this argument appear in Section 2.5.

Competition implies decay. For the upper bound, given in Section 2.3 (with some technical lemmas deferred to an appendix), we invert the above argument by contradiction. In brief: if all particles to the right of a given curve have spent large amounts of time in high-mass environments, then all such individuals will have very low mass; if furthermore there are not many of them, then their total mass is also small.

More precisely, suppose that for some $t' \in (0,t)$ and some function $g : [0,\infty) \to \mathbb{R}$, for all $s \in [t',t]$, we have $d(s,m) > g(s)+1$, so $\zeta(s,x) \geq m$ for all $x \in (0,g(s))$. Then for all i,

$$M_i(t) \leq \exp(-m \cdot \mathrm{Leb}(\{s \in [t',t] : X_{i,t}(s) \in (0,g(s))\}))\,.$$

It follows that if all particles with $X_i(t) \geq g(t)$ have $\mathrm{Leb}(\{s \in [t',t] : X_{i,t}(s) \in (0,g(s))\}) \geq \ell$ then for all $x \geq g(t)+1$, recalling the notation $N(t,x)$ from the introduction,

$$\zeta(t,x) \leq e^{-m\ell} \cdot |N(t,g(t))|\,.$$

If $|N(t,g(t))| < me^{m\ell}$, this is in contradiction with the assumption that $d(t,m) > g(t)+1$.

Whitecaps are just foam. Once again using estimates related to consistent maximal displacement for BBM, we show that for $c < c^*$, with $g(s) = \sqrt{2}s - cs^{1/3}$, with high probability every particle with $X_i(t) > g(t)$ indeed spends at least a time $Ct^{1/3}$ behind the curve g. This is the content of Proposition 2.3.2. Under the assumption that $d(s,m) > g(s)+1$ for all $s \in [Ct^{1/3}/2, t]$, it follows that the particles counted by $N(t,g(t))$ are as insubstantial as sea spray; for all $x \geq g(t)+1$,

$$\zeta(t,x) \leq e^{-mCt^{1/3}/2} \cdot |N(t,g(t))|.$$

Standard and simple arguments for BBM show that $|N(t,g(t))| = e^{O(t^{1/3})}$ with high probability, so we obtain a contradiction for large t if C is sufficiently large. It follows that with high probability there is $s \in [Ct^{1/3}/2, t]$ such that $d(s,m) \leq g(s)+1$.

Consistent maximal displacement. We conclude the sketch with a brief explanation of why the threshold for consistent maximal displacement is at distance $\Theta(t^{1/3})$ from the front. We restrict ourselves to consideration of the first moment. Fix $C > 0$ and let $S = \{i \leq N(t) : \forall s \leq t,\ |X_{i,t}(s) - \sqrt{2}s| \leq C\}$. This imposes that particles stay both above a lower envelope and below an upper envelope, but the upper bound has little probability cost since the front stays behind $\sqrt{2}s + O(1)$ with high probability.

By the many-to-one lemma,

$$\mathbb{E}|S| = e^t \mathbb{P}\left[\forall s \leq t,\ |B(s) - \sqrt{2}s| \leq C\right],$$

where B is a Brownian motion. By a Girsanov transform,

$$\begin{aligned}\mathbb{P}\left[\forall s \leq t,\ |B(s) - \sqrt{2}s| \leq C\right] &= \mathbb{E}\left[e^{t-\sqrt{2}(B(t)+\sqrt{2}t)}\mathbf{1}_{[\forall s \leq t, |B(s)| \leq C]}\right] \\ &= e^{-t+\sqrt{2}C+o(C)}\mathbb{P}\left[\forall s \leq t,\ |B(s)| \leq C\right].\end{aligned}$$

By Brownian scaling, the last probability is $\mathbb{P}\left[\forall s \leq t/C^2,\ |B(s)| \leq 1\right] = e^{-\Theta(t/C^2)}$. We obtain that $\mathbb{E}|S| = e^{\sqrt{2}C+o(C)-\Theta(t/C^2)}$. Thus, $\mathbb{E}|S|$ becomes large when C is of order $t^{1/3}$.

Definitions

We sometimes need to consider the evolution of a subset of the particles starting at a time greater than zero, so it is useful to allow initial conditions other than a single mass-one particle at the origin. Generally, for $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$ and $\mathbf{m} = (m_1, \ldots, m_k) \in (0, \infty)^k$, we write $\mathbb{P}_{\mathbf{x},\mathbf{m}}$ for the probability measure corresponding to an initial condition with a particle of mass m_i at location x_i for each $1 \leq i \leq k$. We write $\mathbb{P} = \mathbb{P}_{(0),(1)}$ for the default initial condition.

We say a random variable X is *geometric with parameter* p, or is $\mathrm{Geom}(p)$-distributed, if $\mathbb{P}[X = k] = (1-p)^{k-1}p$ for positive integer k.

2.3 Upper bound

Recall from the introduction that $c^* = 3^{4/3}\pi^{2/3}2^{-7/6}$. The next proposition is a restatement of the upper bound from Theorem 2.1.1.

Proposition 2.3.1. *For any* $m > 0$, *almost surely*

$$\limsup_{t\to\infty} \frac{\sqrt{2}t - d(t,m)}{t^{1/3}} \geq c^*.$$

For the remainder of the section, we fix $c \in (0, c^*)$ and let $g(s) = \sqrt{2}s - cs^{1/3}$ for $s \geq 0$. The following is the key step of the proof.

Proposition 2.3.2. *("No one can surf* g*") For any* $C > 0$, *there exists* $\delta = \delta(c, C) > 0$ *such that for* t *sufficiently large*

$$\mathbb{P}\left[\exists i \leq N(t) \ s.t. \ \mathrm{Leb}(\{s \leq t : X_{i,t}(s) \leq g(s)\}) \leq Ct^{1/3}\right] \leq e^{-\delta t^{1/3}}.$$

The proof of Proposition 2.3.2 will take up most of this section, but we now give a brief justification of the result, and then show how it is used to prove Proposition 2.3.1. By the method used in Jaffuel (2012) for studying branching random walks, for ε sufficiently small that $c + \varepsilon < c^*$, there exists $\delta > 0$ such that

$$\mathbb{P}\left[\exists i \leq N(t) \text{ s.t. } X_{i,t}(s) \geq \sqrt{2}s - (c+\varepsilon)s^{1/3} \, \forall s \leq t\right] \leq e^{-\delta t^{1/3}}.$$

For our fixed c, we shall choose a small constant $\beta > 0$ and let $b(s) = \sqrt{2}s - c(s + \beta t)^{1/3}$ for $s \in [0, t]$. Then by adapting the method used in Jaffuel (2012), one may show that since $c < c^*$, if β is sufficiently small then for any constant K,

$$\mathbb{P}\left[\exists i \leq N(t) \text{ s.t. } X_{i,t}(s) \geq b(s) - Kt^{1/6} \, \forall s \leq t\right] \leq e^{-\delta t^{1/3}}$$

for some $\delta > 0$.

Now fix $K > 0$ large. For large t, the function b is approximately linear on intervals of length $Ct^{1/3}$. This will allow us to use Brownian scaling to show that if particle $i \leq N(t)$ only spends time $Ct^{1/3}$ time below b, then it has conditional probability at least $1/2$ of staying above $b - Kt^{1/6}$, so the probability such an $i \leq N(t)$ exists is also $\mathcal{O}(e^{-\delta t^{1/3}})$. Since $b \leq g$ this gives us Proposition 2.3.2.

Before giving the details of this argument, we prove Proposition 2.3.1 assuming Proposition 2.3.2.

Proof of Proposition 2.3.1. We continue to write $g(s) = \sqrt{2}s - cs^{1/3}$, for fixed $c \in (0, c^*)$ as above. Fix $m > 0$, let $C = 4\sqrt{2}c(1 + m^{-1})$, and let $\delta = \delta(c, C)$ be as defined in Proposition 2.3.2. It suffices to show that, as $t \to \infty$,

$$\mathbb{P}\left[\exists s \in [Ct^{1/3}/2, t] : d(s, m) \leq g(s) + 1\right] \to 1. \tag{2.1}$$

Next, fix t large. Recalling the notation $N(t, x) = \{i \leq N(t) : X_i(t) \geq x\}$, let

$$A_1 = \{i \in N(t, g(t)) : \mathrm{Leb}(\{s \leq t : X_{i,t}(s) \leq g(s)\}) \leq Ct^{1/3}\}$$

$$\text{and } A_2 = \{i \in N(t, g(t)) : \exists s \in [Ct^{1/3}/2, t] \text{ s.t. } X_{i,t}(s) < 0\}.$$

Also, let E be the event that $d(s, m) > g(s) + 1$ for all $s \in [Ct^{1/3}/2, t)$. On the event E, if $i \in N(t, g(t))$ and $i \notin A_1 \cup A_2$ then

$$M_i(t) \leq \exp(-m\mathrm{Leb}(\{Ct^{1/3}/2 \leq s \leq t : X_{i,t}(s) \in (0, g(s))\})) \leq \exp(-mCt^{1/3}/2).$$

Since all masses are at most 1, it follows that on E,

$$\sum_{i \in N(t, g(t))} M_i(t) \leq |A_1 \cup A_2| + \exp(-mCt^{1/3}/2)|N(t, g(t))|.$$

Also, for all $y \geq g(t) + 1$ we have $\zeta(t, y) \leq \sum_{i \in N(t, g(t))} M_i(t)$; we thus have

$$\begin{aligned}
&\mathbb{P}\left[\forall s \in [Ct^{1/3}/2, t], d(s, m) > g(s) + 1\right] \\
&\quad = \mathbb{P}\left[d(t, m) > g(t) + 1, E\right] \\
&\quad \leq \mathbb{P}\left[\sum_{i \in N(t, g(t))} M_i(t) > m, E\right] \\
&\quad \leq \mathbb{P}\left[A_1 \cup A_2 \neq \emptyset\right] + \mathbb{P}\left[|N(t, g(t))| \geq m \exp(mCt^{1/3}/2)\right].
\end{aligned} \tag{2.2}$$

By Proposition 2.3.2, for t sufficiently large, $\mathbb{P}\left[A_1 \neq \emptyset\right] \leq \exp(-\delta t^{1/3})$. Next, using the many-to-one lemma and then a Girsanov transform, for B a Brownian motion,

$$\mathbb{P}\left[A_2 \neq \emptyset\right] \leq \mathbb{E}\left[|A_2|\right]$$

$$
\begin{aligned}
&= e^t \mathbb{P}\left[B(t) \geq \sqrt{2}t - ct^{1/3}, \exists s \in [Ct^{1/3}/2, t] : B(s) \leq 0\right] \\
&= e^t \mathbb{E}\left[e^{t-\sqrt{2}(B(t)+\sqrt{2}t)} \mathbb{1}_{B(t) \geq -ct^{1/3}, \exists s \in [Ct^{1/3}/2,t]: B(s) \leq -\sqrt{2}s}\right] \\
&\leq e^{\sqrt{2}ct^{1/3}} \mathbb{P}\left[B(t) \geq -ct^{1/3}, \exists s \in [Ct^{1/3}/2, t] : B(s) \leq -\sqrt{2}s\right].
\end{aligned}
$$

Now partitioning according to the first interval $[j, j+1]$ in which $B(s) \leq -\sqrt{2}s$,

$$
\begin{aligned}
\mathbb{P}\left[\exists s \in [Ct^{1/3}/2, t] : B(s) \leq -\sqrt{2}s\right] &\leq \sum_{j=\lfloor Ct^{1/3}/2 \rfloor}^{\lfloor t \rfloor} \mathbb{P}\left[\sup_{s \in [j,j+1]} B(s) \geq \sqrt{2}j\right] \\
&\leq \sum_{j=\lfloor Ct^{1/3}/2 \rfloor}^{\lfloor t \rfloor} \mathbb{P}\left[\sup_{s \leq j+1} B(s) \geq \sqrt{2}j\right] \\
&= 2 \sum_{j=\lfloor Ct^{1/3}/2 \rfloor}^{\lfloor t \rfloor} \mathbb{P}\left[B(j+1) \geq \sqrt{2}j\right] \\
&\leq 2 \sum_{j=\lfloor Ct^{1/3}/2 \rfloor}^{\lfloor t \rfloor} \exp(-j^2/(j+1)) \\
&\leq 2\exp(-Ct^{1/3}/3)
\end{aligned}
$$

where the equality in the third line follows by the reflection principle, and the final inequality holds for t sufficiently large. Since $C > 4\sqrt{2}c$, it follows that

$$
\mathbb{P}[A_2 \neq \emptyset] \leq e^{\sqrt{2}ct^{1/3}} 2e^{-Ct^{1/3}/3} \leq 2e^{-\sqrt{2}ct^{1/3}/3}.
$$

Finally, by another Girsanov transform,

$$
\mathbb{P}[|N(t, g(t))| > x] \leq x^{-1}\mathbb{E}[|N(t, g(t))|] = x^{-1}\mathbb{E}\left[e^{-\sqrt{2}B(t)} \mathbf{1}_{[B(t) \geq -ct^{1/3}]}\right] \leq x^{-1}e^{\sqrt{2}ct^{1/3}}.
$$

Combining the bounds on $\mathbb{P}[A_1 \neq \emptyset]$, $\mathbb{P}[A_2 \neq \emptyset]$, and $\mathbb{P}[|N(t, g(t))| > x]$ with (2.2), we obtain that

$$
\mathbb{P}\left[\forall s \in [Ct^{1/3}/2, t], d(s, m) > g(s) + 1\right] \leq e^{-\delta t^{1/3}} + 2e^{-\sqrt{2}ct^{1/3}/3} + m^{-1}e^{(\sqrt{2}c - mC/2)t^{1/3}},
$$

which tends to 0 as $t \to \infty$ since $C > 2\sqrt{2}cm^{-1}$. This establishes (2.1) and completes the proof. □

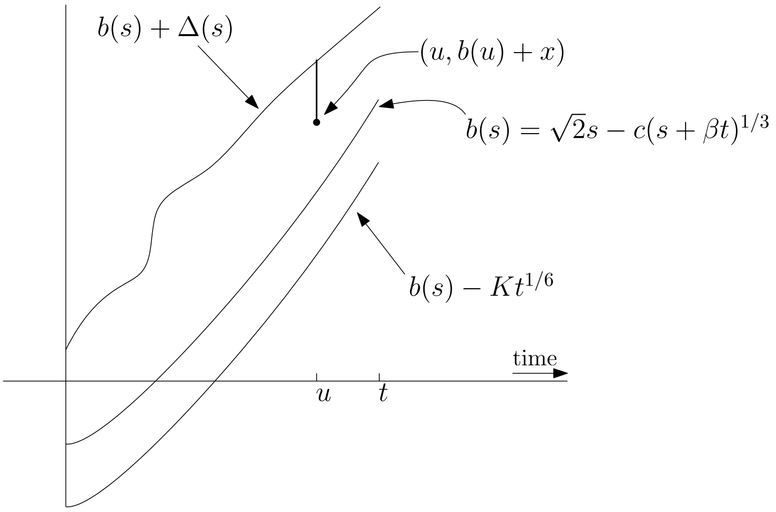

Figure 2.1: For fixed u and x, Lemma 2.3.3 bounds the probability that the Brownian motion stays between the upper and lower curves until time u and has final position on the vertical line segment. Lemma 2.3.4 compares (a) the probability that a Brownian motion stays below the upper curve until time t and spends very little time below the middle curve, with (b) the probability that the Brownian motion stays between the upper and lower curves until time t.

For the rest of this section we work towards the proof of Proposition 2.3.2. Figure 2.1 should help in interpreting the statements of the next two lemmas. Recall that we fixed $c < c^*$.

Lemma 2.3.3. *There exists* $\beta > 0$ *such that for* $b(s) = \sqrt{2}s - c(s + \beta t)^{1/3}$ *and for* $K > 0$, $t > 0$ *both sufficiently large, there exists a function* $\Delta : [0, t] \to [t^{1/4}, Kt^{1/3}]$ *with* $\Delta(t) \le Kt^{1/4}$ *and with* $|\Delta'(s)| \le 1$ *for all* $s \in [0, t]$*, such that for all* $u \in [0, t]$ *and all* $x \in [-Kt^{1/6}, \Delta(u))$*, for* $(B(s), s \ge 0)$ *a Brownian motion started at* 0*,*

$$\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\, \forall s \le u, B(u) > b(u) + x\right]$$
$$\le \exp(-u - t^{1/3}/K + \sqrt{2}(\Delta(u) - x)). \tag{2.3}$$

We prove Lemma 2.3.3 by appealing to technical lemmas from Roberts (2015), which bound the probability that a Brownian motion stays in a narrow tube of variable width. In order to verify that the results of Roberts (2015) apply for some function Δ with the above properties, we adapt a technique from Jaffuel (2012). In Jaffuel (2012), the existence of a function analogous to Δ is constructed as the solution of a certain integral equation. We defer the details of the proof to Appendix 2.A.

From this point on, we let $\beta > 0$ and $b(s)$ be as in Lemma 2.3.3. We assume that t is sufficiently large that b is increasing on $[0,\infty)$. We now show that if K is sufficiently large, a Brownian motion which spends at most $Ct^{1/3}$ time before time t below the curve b has a conditional probability of at least 1/2 of staying above the curve $b - Kt^{1/6}$ up to time t.

Lemma 2.3.4. *Let $(B(s), s \geq 0)$ be a Brownian motion started at 0. Then given $C > 0$, there is a constant $K(C) > 0$ such that for t sufficiently large, and any measurable function $\Delta : [0,t] \to (0,\infty)$,*

$$\begin{aligned}&\mathbb{P}\left[B(s) \leq b(s) + \Delta(s)\,\forall\, s \leq t, \mathrm{Leb}(\{s \leq t : B(s) \leq b(s)\}) \leq Ct^{1/3}\right]\\ &\leq 2\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\,\forall\, s \leq t\right].\end{aligned}$$

In proving Lemma 2.3.4, we will use the following auxiliary result.

Lemma 2.3.5. *Fix non-negative real numbers $(t_i, i \geq 1)$. For each $i \geq 1$ let $(X_i(u), 0 \leq u \leq t_i)$ be either a Brownian meander or a Brownian excursion of length t_i. Then writing $T = \sum_{i\geq 1} t_i$, for $x \geq 12T^{1/2}$ we have*

$$\mathbb{P}\left[\sup_{i\geq 1}\max_{u\leq t_i} X_i(u) \geq x\right] < e^{-x^2/32T}.$$

The proof of Lemma 2.3.5 is deferred to the appendix.

Proof of Lemma 2.3.4. Write

$$\begin{aligned}E &= \{B(s) \leq b(s) + \Delta(s)\,\forall\, s \leq t\}\\ A_1 &= \{B(s) \geq b(s) - Kt^{1/6}\,\forall\, s \leq t\},\\ A_2 &= \{\mathrm{Leb}(\{s \leq t : B(s) \leq b(s)\}) \leq Ct^{1/3}\}.\end{aligned}$$

To prove the lemma, it suffices to show that provided $K = K(C)$ is sufficiently large, $\mathbb{P}[A_1^c \mid A_2 \cap E] \leq 1/2$, since

$$\mathbb{P}[A_1 \cap E] \geq \mathbb{P}[A_1 \cap A_2 \cap E] = \mathbb{P}[A_2 \cap E]\,(1 - \mathbb{P}[A_1^c \mid A_2 \cap E]).$$

Fix $L \in (Ct^{1/3}, 2Ct^{1/3}]$ so that $n := t/L$ is integer; this is possible for t large enough. Then, for each $0 \leq i \leq n-2$ let $b_i : [iL, (i+2)L] \to \mathbb{R}$ be defined by

$$b_i(s) = b(iL) + \frac{s - iL}{2L}(b((i+2)L) - b(iL)) - 1. \tag{2.4}$$

By the mean value theorem, for some $s_1, s_2 \in [iL, (i+2)L]$,

$$\begin{aligned} b(iL) + \frac{s-iL}{2L}(b((i+2)L) - b(iL)) &= b(s) + (s-iL)(b'(s_2) - b'(s_1)) \\ &\leq b(s) + (2L)^2 \sup_{r \in [iL,(i+2)L]} |b''(r)| \\ &\leq b(s) + (2L)^2 \cdot \frac{2c}{9(\beta t + iL)^{5/3}} \\ &\leq b(s) + \frac{32cC^2}{9\beta^{5/3}t}, \end{aligned}$$

which is less than $b(s)+1$ for t sufficiently large. It follows that for t sufficiently large, $b_i \leq b$ on the interval $[iL, (i+2)L]$, for all $i \leq n-2$.

Next, for $i \leq n-2$ let $g_i = \inf\{s \geq iL : B(s) \geq b_i(s)\}$. Also, for $i < n-2$ let $d_i = \sup\{s \leq (i+2)L : B(s) \geq b_i(s)\}$, and let $d_{n-2} = t$. Then write

$$\mathcal{U}_i = \{s \in [g_i, d_i] : B(s) \leq b_i(s)\}.$$

For $i < n-2$, this is the set of times when B is performing an excursion below b_i which starts at or after time iL and ends at or before time $(i+2)L$. For $i = n-2$ we have $(i+2)L = t$, and in this case we include a final excursion below b_i which does not end before time t if it starts at or after time iL. The set $\mathcal{U}_i$ is a union of closed intervals, which we enumerate as $\{[l_{i,j}, r_{i,j}], j \geq 1\}$ according to a fixed rule (in decreasing order of size, say). See Figure 2.2 for an illustration of $\mathcal{U}_i$.

For all $i < n-2$, conditional on $\mathcal{U}_i$, for each $j \geq 1$ the function

$$(b_i(l_{i,j} + s) - B(l_{i,j} + s), 0 \leq s \leq r_{i,j} - l_{i,j}) \tag{2.5}$$

is a Brownian excursion of length $r_{i,j} - l_{i,j}$. The case $i = n-2$ is very slightly different, and we now describe it; for the remainder of the paragraph set $i = n-2$. If $B(t) \geq b_{n-2}(t)$

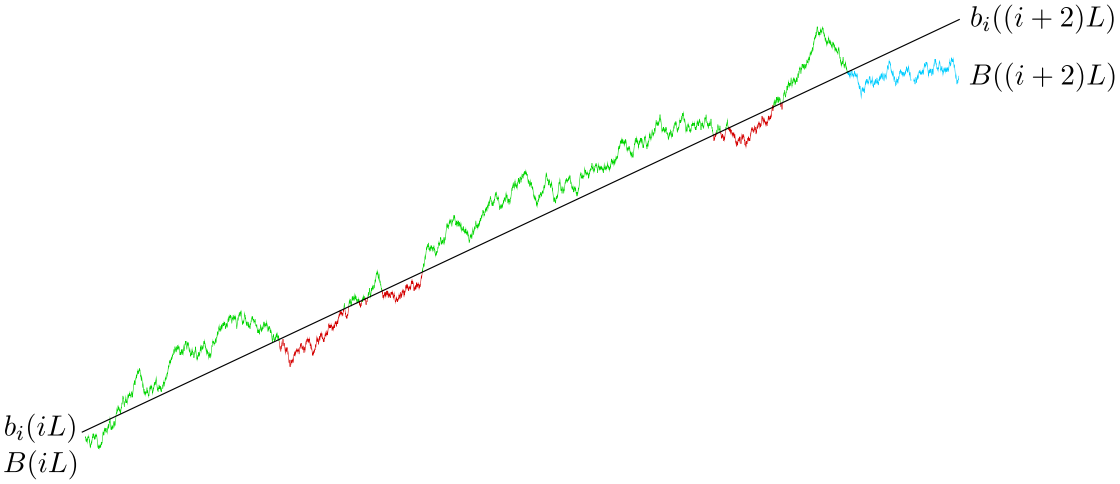

Figure 2.2: An illustration of $\mathcal{U}_i$ for a Brownian motion B. The black line shows $b_i : [iL, (i+2)L] \to \mathbb{R}$. The path of B is coloured; for times in $\mathcal{U}_i$ it is red, and during $[iL, (i+2)L] \setminus \mathcal{U}_i$ it is green. The final blue meander is in $\mathcal{U}_i$ if and only if $i = n-2$.

then there is no change. However, if $B(t) < b_{n-2}(t)$ then there is a unique integer $j \geq 1$ with $[l_{i,j}, r_{i,j}]$ with $r_{i,j} = t$; for this j the process described by (2.5) is a Brownian meander of length $r_{i,j} - l_{i,j}$; for all other j the process is a Brownian excursion.

All this is true even if we additionally condition on $A_2 \cap E$, since letting $\mathcal{U} = \cup_{i \leq n-2} \mathcal{U}_i$, the occurrence of the event $A_2 \cap E$ is determined by $\mathrm{Leb}(\mathcal{U})$ and $B|_{[0,t]\setminus\mathcal{U}}$. By Lemma 2.3.5, it follows that

$$\begin{aligned} \mathbb{P}&\left[\sup_{s\in\mathcal{U}_i}(b_i(s) - B(s)) \geq x \;\middle|\; \mathcal{U}_i, A_2 \cap E\right] \\ &= \mathbb{P}\left[\sup_{j\geq 1}\sup_{s\in[l_{i,j},r_{i,j}]}(b_i(s) - B(s)) \geq x \;\middle|\; \mathcal{U}_i, A_2 \cap E\right] \\ &\leq \exp\left(-\frac{x^2}{32\mathrm{Leb}(\mathcal{U}_i)}\right) + \mathbf{1}_{[x^2 < 144\mathrm{Leb}(\mathcal{U}_i)]}\,. \end{aligned} \tag{2.6}$$

We next analyze the event $A_1^c \cap A_2$. Note that $b_i + 1$ is the linear interpolation of b on the interval $[iL, (i+2)L]$. Since b is convex it follows that $b \leq b_i + 1$ on this interval.

If A_1^c occurs then there is $s \leq t$ such that $B(s) \leq b(s) - Kt^{1/6}$. For such s, for any i with $s \in [iL, (i+2)L]$, the preceding paragraph then implies that $B(s) \leq b(s) - Kt^{1/6} \leq b_i(s) - (Kt^{1/6} - 1)$.

Next suppose A_2 occurs, and suppose $s \leq t$ is such that $B(s) \leq b(s) - Kt^{1/6}$. Then s is in an excursion of $B(s)$ below $b(s)$. Temporarily write $[g, d]$ for the time interval during which this excursion takes place. Since A_2 occurs, $[g, \min(d, t)]$ has length at most $Ct^{1/3}$ so is strictly contained within in an interval $[iL,(i+2)L]$ for some $i \le n-2$. Since $s \in [g, \min(d,t)]$ and $B(g) = b(g) \ge b_i(g)$ and either $d \ge t$ or $B(d) = b(d) \ge b_i(d)$ but $B(s) < b_i(s)$, it follows that $s \in \mathcal{U}_i$. On the other hand, if $s \in \mathcal{U}_i$ then $B(s) \le b_i(s) \le b(s)$. Each point s lies in at most three distinct sets $\mathcal{U}_i$, so on A_2 we have

$$\sum_{i \le n-2} \mathrm{Leb}(\mathcal{U}_i) \le 3Ct^{1/3}\,.$$

Finally, suppose $A_1^c \cap A_2$ occurs. Then the observations of the preceding three paragraphs imply that there exists $i \le n-2$ and $s \in \mathcal{U}_i$ such that $B(s) \le b_i(s) - (Kt^{1/6} - 1) < b_i(s) - Kt^{1/6}/2$, the last inequality holding for t large. Combined with (2.6), this yields

$$\begin{aligned}
&\mathbb{P}\,[A_1^c \mid A_2 \cap E] \\
&\quad\le \mathbb{P}\left[\sup_{i \le n-2} \sup_{s \in \mathcal{U}_i} (b_i(s) - B(s)) \ge Kt^{1/6}/2, \sum_{i \le n-2} \mathrm{Leb}(\mathcal{U}_i) \le 3Ct^{1/3} \,\middle|\, A_2 \cap E \right] \\
&\quad\le \sup_{\substack{u_1 + \ldots + u_{n-2} \le 3Ct^{1/3} \\ u_i \ge 0}} \sum_{i=1}^{n-2} \mathbb{P}\left[\sup_{s \in \mathcal{U}_i} (b_i(s) - B(s)) \ge Kt^{1/6}/2 \,\middle|\, \mathrm{Leb}(\mathcal{U}_i) = u_i, A_2 \cap E \right] \\
&\quad\le \sup_{\substack{u_1 + \ldots + u_{n-2} \le 3Ct^{1/3} \\ u_i \ge 0}} \sum_{i=1}^{n-2} \exp(-K^2 t^{1/3}/128 u_i)\,,
\end{aligned}$$

the last bound holding provided that $K^2 > (12)^3 C$ so that $(Kt^{1/6}/2)^2 > 144(3Ct^{1/3})$. Finally, letting $x = K^2 t^{1/3}/128$, the function $f(a) = e^{-x/a}\mathbb{1}_{a>0}$ is convex for $a \in [0, x/2]$, and $f(0) = 0$, so if $K^2 > 768C$ then for each i, $f(u_i) \le (u_i/\sum u_k) f(\sum u_k)$. Hence

$$\mathbb{P}\,[A_1^c \mid A_2 \cap E] \le e^{-K^2/384C} < e^{-2} < 1/2,$$

as required. □

We next state a variant of Lemma 2.3.4 which is proved in a similar way.

Lemma 2.3.6. *Let $(B(s), s \ge 0)$ be a Brownian motion started at 0. Then given $C > 0$, there is a constant $K = K(C)$ such that for t sufficiently large, and any measurable* *function* $\Delta : [0,t] \to (0,\infty)$*,* $u \le t$ *and* $z \in [b(u), b(u)+\Delta(u))$*, we have*

$$\mathbb{P}\left[B(s) \le b(s) + \Delta(s)\,\forall\, s \le u, \mathrm{Leb}(\{s \le u : B(s) \le b(s)\}) \le Ct^{1/3}, B(u) \ge z\right]$$
$$\le 2\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\,\forall\, s \le u, B(u) \ge z\right].$$

and

$$\mathbb{P}\left[B(s) \le b(s) + \Delta(s)\,\forall\, s \le u, \mathrm{Leb}(\{s \le u : B(s) \le b(s)\}) \le Ct^{1/3}\right]$$
$$\le 2\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\,\forall\, s \le u\right].$$

Proof. These bounds are proved in the same way as Lemma 2.3.4, by only considering the times $(\mathcal{U}_i)_i$ when B is performing an excursion below b_i on the interval $[0,u]$, and using that $z \ge b(u)$, conditioning on $B(u) \ge z$ does not affect the distribution of B on $(\mathcal{U}_i)_i$ given $(\mathcal{U}_i)_i$. We omit the details. ☐

We are now in a position to complete the proof of Proposition 2.3.2, using Lemmas 2.3.3, 2.3.4 and 2.3.6.

Proof of Proposition 2.3.2. Choose β such that Lemma 2.3.3 applies, and recall that $b(s) = \sqrt{2}s - c(s+\beta t)^{1/3} \le g(s)\ \forall s \ge 0$. Then

$$\mathbb{P}\left[\exists i \le N(t) \text{ s.t. } \mathrm{Leb}(\{s \le t : X_{i,t}(s) \le g(s)\}) \le Ct^{1/3}\right]$$
$$\le \mathbb{P}\left[\exists i \le N(t) \text{ s.t. } \mathrm{Leb}(\{s \le t : X_{i,t}(s) \le b(s)\}) \le Ct^{1/3}\right].$$

We shall prove that

$$\mathbb{P}\left[\exists i \le N(t) \text{ s.t. } \mathrm{Leb}(\{s \le t : X_{i,t}(s) \le b(s)\}) \le Ct^{1/3}\right] \le e^{-\delta t^{1/3}} \tag{2.7}$$

for some $\delta > 0$ for t sufficiently large, which establishes the proposition. Take K and t sufficiently large that Lemmas 2.3.3, 2.3.4 and 2.3.6 hold. Then let $\Delta : [0,t] \to [t^{1/4}, Kt^{1/3}]$ be as in Lemma 2.3.3, and in particular satisfying that $\Delta(t) \le Kt^{1/4}$ and $|\Delta'(s)| \le 1$ for all $s \in [0,t]$.

Since $|\Delta'(s)| \leq 1$ for all $s \in [0,t]$, $\inf_{u\in[j,j+1]} \Delta(u) \geq \Delta(j) - 1$ for $j \in [0, t-1]$. Hence if for some $i \leq N(j+1)$, $X_{i,j+1}(s) \geq b(s) + \Delta(s)$ for some $s \in [j, j+1]$, then since b is increasing,

$$X_{i,j+1}(s) \geq b(j) + \inf_{u\in[j,j+1]} \Delta(u) \geq b(j) + \Delta(j) - 1. \tag{2.8}$$

Using (2.8), and partitioning the event $\{\exists i \leq N(t) \text{ s.t. } \mathrm{Leb}(\{s \leq t : X_{i,t}(s) \leq b(s)\}) \leq Ct^{1/3}\}$ according to the interval $[j, j+1]$ in which $X_{i,t}(s)$ first exceeds $b + \Delta$, we have that

$$\begin{aligned}
&\mathbb{P}\left[\exists i \leq N(t) \text{ s.t. } \mathrm{Leb}(\{s \leq t : X_{i,t}(s) \leq b(s)\}) \leq Ct^{1/3}\right] \\
&\quad \leq \mathbb{P}\left[\exists i \leq N(t) \text{ s.t. } X_{i,t}(s) < b(s) + \Delta(s)\,\forall\, s \leq t, \mathrm{Leb}(\{s \leq t : X_{i,t}(s) \leq b(s)\}) \leq Ct^{1/3}\right] \\
&\quad + \sum_{j=0}^{\lfloor t \rfloor} \mathbb{P}\Bigg[\exists i \leq N(j+1) \text{ s.t. } X_{i,j+1}(s) < b(s) + \Delta(s)\,\forall\, s \leq j, \\
&\qquad\qquad \mathrm{Leb}(\{s \leq j : X_{i,j+1}(s) \leq b(s)\}) \leq Ct^{1/3}, \sup_{s\in[j,j+1]} X_{i,j+1}(s) \geq b(j) + \Delta(j) - 1\Bigg] \\
&\quad \leq e^t \mathbb{P}\left[B(s) < b(s) + \Delta(s)\,\forall\, s \leq t, \mathrm{Leb}(\{s \leq t : B(s) \leq b(s)\}) \leq Ct^{1/3}\right] \\
&\quad + \sum_{j=0}^{\lfloor t \rfloor} e^{j+1} \mathbb{P}\Bigg[B(s) < b(s) + \Delta(s)\,\forall\, s \leq j, \\
&\qquad\qquad \mathrm{Leb}(\{s \leq j : B(s) \leq b(s)\}) \leq Ct^{1/3}, \sup_{s\in[j,j+1]} B(s) \geq b(j) + \Delta(j) - 1\Bigg],
\end{aligned}$$

where B is a Brownian motion and the last inequality follows by Markov's inequality and the many-to-one lemma. By partitioning according to the value of $B(j)$, we further have

$$\begin{aligned}
&\mathbb{P}\Bigg[B(s) < b(s) + \Delta(s)\,\forall\, s \leq j, \mathrm{Leb}(\{s \leq j : B(s) \leq b(s)\}) \leq Ct^{1/3}, \\
&\qquad\qquad \sup_{s\in[j,j+1]} B(s) \geq b(j) + \Delta(j) - 1\Bigg] \\
&\quad \leq \mathbb{P}\Bigg[B(s) < b(s) + \Delta(s)\,\forall\, s \leq j, \mathrm{Leb}(\{s \leq j : B(s) \leq b(s)\}) \leq Ct^{1/3}, \\
&\qquad\qquad B(j) \geq b(j) + \Delta(j) - \tfrac{1}{2}t^{1/4}\Bigg] \\
&\qquad + \mathbb{P}\left[\sup_{[0,1]} B(u) \geq \tfrac{1}{2}t^{1/4} - 1\right]
\end{aligned}$$

$$\mathbb{P}\left[B(s) < b(s) + \Delta(s)\,\forall\, s \le j, \mathrm{Leb}(\{s \le j : B(s) \le b(s)\}) \le Ct^{1/3}\right].$$

By the reflection principle, for N a standard normal random variable,

$$\mathbb{P}\left[\sup_{[0,1]} B(u) \ge \tfrac{1}{2}t^{1/4} - 1\right] \le 2\mathbb{P}\left[N \ge \tfrac{1}{2}t^{1/4} - 1\right] \le 2e^{-\frac{1}{9}t^{1/2}}$$

for t sufficiently large. Since $\Delta(j) - \frac{1}{2}t^{1/4} > 0$, we can now apply Lemmas 2.3.4 and 2.3.6 to conclude that for t sufficiently large

$$\begin{aligned}
&\mathbb{P}\left[\exists i \le N(t) \text{ s.t. } \mathrm{Leb}(\{s \le t : X_{i,t}(s) \le b(s)\}) \le Ct^{1/3}\right] \\
&\quad \le 2e^t\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\,\forall\, s \le t\right] \\
&\quad\quad + 2\sum_{j=0}^{\lfloor t \rfloor} e^{j+1}\Bigg(\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\,\forall\, s \le j, B(j) \ge b(j) + \Delta(j) - \tfrac{1}{2}t^{1/4}\right] \\
&\quad\quad\quad + 2e^{-\frac{1}{9}t^{1/2}}\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\,\forall\, s \le j\right]\Bigg). \qquad (2.9)
\end{aligned}$$

We can now apply Lemma 2.3.3 to each term. First, by Lemma 2.3.3 applied with $u = t$ and $x = -Kt^{1/6}$, since $\Delta(t) \le Kt^{1/4}$,

$$\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\,\forall\, s \le t\right] \le \exp(-t - t^{1/3}/K + \sqrt{2}K(t^{1/4} + t^{1/6})).$$

By Lemma 2.3.3 applied with $u = j$ and $x = \Delta(j) - \frac{1}{2}t^{1/4}$,

$$\begin{aligned}
&\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\,\forall\, s \le j, B(j) \ge b(j) + \Delta(j) - \tfrac{1}{2}t^{1/4}\right] \\
&\quad \le \exp(-j - t^{1/3}/K + \sqrt{2}\tfrac{1}{2}t^{1/4}).
\end{aligned}$$

Finally, by Lemma 2.3.3 applied with $u = j$ and $x = -Kt^{1/6}$, since $\Delta(j) \le Kt^{1/3}$,

$$\mathbb{P}\left[b(s) - Kt^{1/6} < B(s) < b(s) + \Delta(s)\,\forall\, s \le j\right] \le \exp(-j - t^{1/3}/K + \sqrt{2}K(t^{1/3} + t^{1/6})).$$

Putting everything together in (2.9),

$$\mathbb{P}\left[\exists i \le N(t) \text{ s.t. } \mathrm{Leb}(\{s \le t : X_{i,t}(s) \le b(s)\}) \le Ct^{1/3}\right]$$

$$
\begin{aligned}
&\le 2\exp(-t^{1/3}/K + \sqrt{2}K(t^{1/4}+t^{1/6})) \\
&\quad + 2e\sum_{j=0}^{\lfloor t\rfloor}\Big(\exp(-t^{1/3}/K + \sqrt{2}\tfrac{1}{2}t^{1/4}) + \exp(-t^{1/2}/9 + \mathcal{O}(t^{1/3}))\Big) \\
&\le e^{-\delta t^{1/3}}
\end{aligned}
$$

for some $\delta > 0$ for t sufficiently large, which proves (2.7). □

2.4 The greatest overall particle density

Before moving to the lower bound, we first prove logarithmic upper bounds on how the greatest particle density grows over time; these are needed to ensure that particle masses cannot *decay* too quickly. This may seem contradictory, but the point is that a particle may *a priori* quickly lose a large amount of mass if it finds itself in an extremely dense environment. The next proposition rules this out.

Proposition 2.4.1. *Let $Z = 2 \cdot 10^8$; then for all s sufficiently large,*

$$\mathbb{P}\left[\sup\{\zeta(t,x) : 0 \le t \le s, x \in \mathbb{R}\} > Z\log s\right] \le s^{-4}.$$

Proving Proposition 2.4.1 turns out to be a fair amount of work. In order that the idea is not obscured by detail, however, we set up the heart of the argument right away.

Let $z(t,x) = \sum_{\{i:|X_i(t)-x|<1/2\}} M_i(t)$. The differences between z and ζ are that z only counts mass within distance 1/2 of x, and does not ignore the mass of particles at x (should there be any).

Let $z(t) = \sup_{x\in\mathbb{R}} z(t,x)$, and define a sequence $(\tau_i, i \ge 0)$ of stopping times as follows. Fix s large and for the remainder of the section write $N = N(s) = 10^7\log s$. Let $\tau_0 = \inf\{t : z(t) \ge N-1\}$, and for $k \ge 0$ let $\tau_{k+1} = \inf\{t > \tau_k + 10^5/N : z(t) \ge N-1\}$. Then $\tau_k \ge 10^5 k/N$, so with $I = I(s) = \inf\{k : \tau_k \ge s\}$, we have $I \le \lceil Ns/10^5\rceil$ and

$$\sup\{z(t), t \le s\} \le \sup\{z(t), t < \tau_I\}.$$

Notice that the sequence of stopping times "ignores" small time intervals $[\tau_k, \tau_k + 10^5/N]$. However in any time interval $[\tau_k + 10^5/N, \tau_{k+1})$, the function z nowhere exceeds

N by the definition of the stopping time τ_{k+1}. We thus have

$$\sup\{z(t), t \leq s\} \leq \sup\{z(t), t < \tau_I\} \leq \max\left(N, \sup_{k<I} \sup_{t\in[\tau_k,\tau_k+10^5/N]} z(t)\right) \tag{2.10}$$

We prove the proposition by establishing the following facts. The first fact says that for $k < Ns/10^5$, if $z(\tau_k)$ is not too large then with high probability $z(t)$ is not too large for any $t \in [\tau_k, \tau_k + 10^5/N]$. The second says that for such k, with high probability $z(\tau_k + 10^5/N)$ is small.

Fact 2.4.2. *For s sufficiently large, for all $0 \leq k < Ns/10^5$,*

$$\mathbb{P}\left[\sup\{z(t), t \in [\tau_k, \tau_k + 10^5/N]\} > 10N, z(\tau_k) \leq N, k < I\right] < s^{-6}.$$

Fact 2.4.3. *For s sufficiently large, for all $0 \leq k < Ns/10^5$,*

$$\mathbb{P}\left[z(\tau_k + 10^5/N) \geq N - 1, k < I\right] < s^{-6}.$$

Assuming these two facts, the proposition follows easily.

Proof of Proposition 2.4.1. Fix $k \leq Ns/10^5$. Note that if $z(\tau_{k-1} + 10^5/N) < N - 1$ then $z(\tau_k-) < N - 1$. Since mass only increases by branching, it follows that almost surely a single branching event at time τ_k causes z to increase above $N - 1$. As all masses are at most 1 and branching is binary, it follows that in this case almost surely $z(\tau_k) \leq z(\tau_k-) + 1 < N$. With Fact 2.4.3, this implies that

$$\begin{aligned}\mathbb{P}\left[z(\tau_k) > N, k < I\right] &\leq \mathbb{P}\left[z(\tau_{k-1} + 10^5/N) \geq N - 1, k < I\right]\\ &\leq \mathbb{P}\left[z(\tau_{k-1} + 10^5/N) \geq N - 1, k - 1 < I\right]\\ &< s^{-6}.\end{aligned}$$

We now use that for any events A, B, C we have $\mathbb{P}[A \cap C] \leq \mathbb{P}[A \cap B \cap C] + \mathbb{P}[B^c \cap C]$. By Fact 2.4.2 and the preceding bound, we obtain that for $0 \leq k < Ns/10^5$,

$$\mathbb{P}\left[\sup\{z(t), t \in [\tau_k, \tau_k + 10^5/N]\} > 10N, k < I\right] \leq 2s^{-6}.$$

A union bound and (2.10) then yield

$$\begin{aligned}\mathbb{P}\left[\sup_{t\le s} z(t) > 10N\right] &\le \mathbb{P}\left[\sup_{k<I}\ \sup_{t\in[\tau_k,\tau_k+10^5/N]} z(t) > 10N\right]\\ &\le \sum_{k=0}^{\lfloor Ns/10^5\rfloor} \mathbb{P}\left[\sup_{t\in[\tau_k,\tau_k+10^5/N]} z(t) > 10N, k<I\right]\\ &\le \left(1+\frac{Ns}{10^5}\right)\cdot 2s^{-6}\\ &< s^{-4}\,,\end{aligned}$$

the last inequality holding for s large. Finally, it is easy to see that $\sup_{x\in\mathbb{R}}\zeta(t,x) \le 2z(t)$, so the same bound holds for $\mathbb{P}\left[\sup_{t\le s}\sup_{x\in\mathbb{R}}\zeta(t,x) > 20N\right]$, which proves the proposition. □

The reader who is willing to believe the Facts 2.4.2 and 2.4.3 without proof – or who is impatient to see how Proposition 2.4.1 is used to prove the lower bound from the main theorem – could skip directly to Section 2.5 at this point.

2.4.1 Proofs of Facts 2.4.2 and 2.4.3

We first prove a handful of technical estimates required for the proofs. The first shows that a fixed mass of particles is extremely unlikely to quickly increase its total mass. Recall the definition of $\mathbb{P}_{\mathbf{x},\mathbf{m}}$ from just before the start of Section 2.3.

Lemma 2.4.4. *Fix* $\mathbf{x} = (x_1,\dots,x_k) \in \mathbb{R}^k$ *and* $\mathbf{m} = (m_1,\dots,m_k) \in [0,1]^k$. *Under* $\mathbb{P}_{\mathbf{x},\mathbf{m}}$, *for* $1 \le j \le k$ *let* $G_j(s) = \#\{i : j_{i,s}(0) = j\}$ *be the number of time-s descendants of* x_j. *Then for any* $J \subset \{1,\dots,k\}$, *any* $x \ge \sum_{j\in J} m_j$, *for all* $t \le \log 2$ *and all* $\delta > 0$,

$$\mathbb{P}_{\mathbf{x},\mathbf{m}}\left[\sum_{j\in J} m_j G_j(t) \ge (1+\delta)x\right] \le 2\left(2^{1+\delta}(1-e^{-t})^{\delta}\right)^{x}.$$

Proof. We may clearly assume $J = \{1,\dots,k\}$. Also, adding particles to increase the mass of the starting configuration can only increase the probability we aim to bound, so we may assume that $x = \sum_{i=1}^{k} m_i$. The random variables $(G_j(s), 1 \le j \le k)$ are i.i.d. and are Geom(e^{-s})-distributed (see, e.g., McKean (1975)). Lemma 2.B.2 provides upper tail bounds for weighted sums of geometric random variables where the individual coefficients are small compared with their sum. Using that lemma (with $\varepsilon = 1 - e^{-t}$ – this is where we require that $t < \log 2$), the result follows. ☐

Since $G_j(s)$ is non-decreasing in s, we have

$$\sup_{s\in[0,t]} \sum_{\{i:j_{i,s}(0)\in J\}} M_{i,s}(0) = \sup_{s\in[0,t]} \sum_{j\in J} m_j G_j(s) = \sum_{j\in J} m_j G_j(t)\,.$$

Combining this with the preceding lemma thus also yields the following bound.

Corollary 2.4.5. *With the hypotheses and notation of Lemma 2.4.4,*

$$\mathbb{P}_{\mathbf{x},\mathbf{m}}\left[\sup_{s\in[0,t]} \sum_{\{i:j_{i,s}(0)\in J\}} M_{i,s}(0) \geq (1+\delta)x\right] \leq 2\left(2^{1+\delta}(1-e^{-t})^{\delta}\right)^{x}.$$

The next proposition says that mass does not travel far in a short time, even once branching is taken into account.

Proposition 2.4.6. *Fix* $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$ *and* $\mathbf{m} = (m_1, \ldots, m_k) \in (0,1]^k$. *Then for any* $J \subset \{1, \ldots, k\}$ *and* $x \geq \sum_{i\in J} m_i$, *for all* $t > 0$, $L > 0$ *and* $v > 0$, *we have*

$$\mathbb{P}_{\mathbf{x},\mathbf{m}}\left[\sum_{\{i:X_i(t)-X_{i,t}(0)>L, j_{i,t}(0)\in J\}} M_{i,t}(0) > vx\right] \leq \frac{\exp(t - L^2/(2t))}{v}\,.$$

Proof. We may clearly assume that $J = \{1, \ldots, k\}$. For $j \leq k$ write $S_j = \{i \leq N(t) : j_{i,t}(0) = j\}$ for the set of indices of time-t descendants of x_j. Then let $R_j = \{i \in S_j : X_i(t) - X_{i,t}(0) > L\}$, so that

$$\sum_{\{i:X_i(t)-X_{i,t}(0)>L\}} M_{i,t}(0) = \sum_{j=1}^{k} \sum_{i\in R_j} M_{i,t}(0) = \sum_{j=1}^{k} m_j |R_j|\,.$$

By the many-to-one lemma, for B a one-dimensional Brownian motion,

$$\mathbb{E}\left[|R_j|\right] = e^t \mathbb{P}\left[B(t) - B(0) > L\right] \leq \exp(t - L^2/(2t)).$$

This bound does not depend on $j \leq k$. It then follows by Markov's inequality that for $v > 0$,

$$\begin{aligned}\mathbb{P}\left[\sum_{\{i: X_i(t) - X_{i,t}(0) > L\}} M_{i,t}(0) > vx\right] &= \mathbb{P}\left[\sum_{j \leq k} m_j |R_j| > vx\right] \\ &\leq \frac{\mathbb{E}\left[\sum_{j \leq k} m_j |R_j|\right]}{vx} \\ &\leq \frac{\exp(t - L^2/(2t))}{v},\end{aligned}$$

where we have used in the last inequality that $\sum_{j \leq k} m_j \leq x$. ☐

In the sequel, we also use the following corollary, which extends Proposition 2.4.6 by considering all times in an interval $[0, t]$, rather than a fixed time $t > 0$, at the cost of a slightly weaker bound.

Corollary 2.4.7. *Under the conditions of Proposition 2.4.6, for all $t_0 > 0$, $L > 0$ and $v > 0$, and all $x \geq \sum_{i \in J} m_i$,*

$$\mathbb{P}_{\mathbf{x},\mathbf{m}}\left[\sup_{t \leq t_0} \sum_{\{i: X_i(t) - X_{i,t}(0) \geq L, j_{i,t}(0) \in J\}} M_{i,t}(0) > 2vx\right] \leq \frac{2\exp(t_0 - L^2/(2t_0))}{v}.$$

Proof. Consider the stopping time

$$\tau = \inf\left\{t : \sum_{\{i: X_i(t) - X_{i,t}(0) \geq L, j_{i,t}(0) \in J\}} M_{i,t}(0) > 2vx\right\}.$$

By symmetry,

$$\mathbb{P}\left[\sum_{\{i: X_i(t_0) - X_{i,t_0}(0) \geq L, j_{i,t_0}(0) \in J\}} M_{i,t_0}(0) > vx \;\middle|\; \tau \leq t_0\right] \geq \frac{1}{2},$$

and the corollary follows. ☐

The next lemma says that a large, concentrated mass will quickly decay; once we prove this we will have all the tools we need to establish Facts 2.4.2 and 2.4.3.

Lemma 2.4.8. *There exist $t_0 > 0$ and $C > 0$ such that the following holds. Fix $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$ and $\mathbf{m} = (m_1, \ldots, m_k) \in [0,1]^k$. Let $J = \{j : |x_j| < 1/4\}$, and* *suppose* $A = \sum_{j \in J} m_j > C$*. Then for all* $t \in [500/A, t_0]$*, setting* $I = \{i : j_{i,t}(0) \in J\}$ *we have*

$$\mathbb{P}_{\mathbf{x},\mathbf{m}}\left[\sum_{i \in I} M_i(t) > A/24\right] \leq 2e^{-200A}\,.$$

Proof. The proof is divided as follows. First, the total mass at time t of particles whose trajectory branches at least once is small. Next, among non-branching trajectories, the total mass which moves far from the origin is small. Finally, particles whose trajectories do not branch and stay near the origin will lose a large amount of mass since they are a dense environment. We now formalize this.

Write $I_b = \{i \in I : \exists i' \neq i, j_{i,t}(0) = j_{i',t}(0)\}$ for the indices of particles starting near (distance $< 1/4$) to the origin whose trajectories branch before time t. Then let $I \setminus I_b = I_f \cup I_n$, where

$$I_f = \left\{i \in I \setminus I_b : |X_{i,t}(0)| < 1/4, \sup_{s \in [0,t]} |X_{i,t}(s)| > 1/2\right\}$$

indexes non-branching trajectories that start near the origin but move far (distance $> 1/2$) from the origin before time t, and where $I_n = I \setminus (I_f \cup I_b)$ indexes non-branching trajectories that stay near the origin. Then we have

$$\sum_{i \in I} M_i(t) = \sum_{i \in I_b} M_i(t) + \sum_{i \in I_f} M_i(t) + \sum_{i \in I_n} M_i(t). \tag{2.11}$$

We shall bound each term separately.

We begin by considering branching trajectories. For each $1 \leq j \leq k$, let $G_j = \#\{i \in I : j_{i,t}(0) = j\}$. Then $i \in I_b$ precisely if $j_{i,t}(0) \in J$ and $G_{j_{i,t}(0)} > 1$. Since masses decrease with time,

$$\begin{aligned}\sum_{i \in I_b} M_i(t) &\leq \sum_{i \in I_b} m_{j_{i,t}(0)} \\ &= \sum_{j \in J} m_j G_j \mathbf{1}_{[G_j > 1]}.\end{aligned}$$

Next, since the G_j are integer-valued,

$$\sum_{j\in J} m_j G_j \mathbf{1}_{[G_j>1]} = \sum_{j\in J} m_j(G_j-1) + \sum_{j\in J} m_j \mathbf{1}_{[G_j>1]} < 2\sum_{j\in J} m_j(G_j-1),$$

which with the preceding bound gives

$$\sum_{i\in I_b} M_i(t) \le 2(\sum_{j\in J} m_j G_j - A).$$

By Lemma 2.4.4, it follows that for any fixed $\delta > 0$, if $t < \log 2$,

$$\begin{aligned}\mathbb{P}\left[\sum_{i\in I_b} M_i(t) > 2\delta A\right] &\le \mathbb{P}\left[\sum_{j\in J} m_j G_j \ge (1+\delta)A\right] \\ &\le 2(2^{1+\delta}(1-e^{-t})^{\delta})^A \\ &< 2(2^{1+\delta}t^{\delta})^A \\ &\le e^{-200A}\,, \end{aligned} \tag{2.12}$$

the last bound holding for t sufficiently small that $2^{2+\delta}t^{\delta} < e^{-200}$. We next bound $\sum_{i\in I_n} M_i(t)$, the total final mass from "typical" trajectories, which do not branch and do not move far from their starting position by time t. Fix $c \in (0,1)$ and let E be the event that for all $s \in [0,t]$, $\sum_{\{i:|X_i(s)|<1/2\}} M_i(s) > cA$. On E, if $i \in I_n$ has $j_{i,t}(0) = j$ then $M_i(t) \le m_j \cdot e^{-tcA}$. We thus have

$$\sum_{i\in I_n} M_i(t)\mathbf{1}_{[E]} \le \sum_{j\in J} m_j \cdot e^{-tcA} \cdot \mathbf{1}_{[E]} = Ae^{-tcA}\mathbf{1}_{[E]}\,.$$

Next, let $I_n(s) = \{j_{i,t}(s) : i \in I_n\}$ be the indices of time-s ancestors of individuals in I_n. Since trajectories indexed by I_n do not branch, $\sum_{i\in I_n(s)} M_i(s)$ is decreasing for $s \in [0,t]$. Necessarily $|X_i(s)| < 1/2$ for $i \in I_n(s)$, so if E^c occurs then there is $s \in [0,t]$ such that $\sum_{i\in I_n(s)} M_i(s) \le cA$. We thus have

$$\sum_{i\in I_n} M_i(t)\mathbf{1}_{[E^c]} \le cA\mathbf{1}_{[E^c]}\,,$$

and the two preceding bounds together give

$$\sum_{i\in I_n} M_i(t) \le \max\left(cA, Ae^{-tcA}\right) . \tag{2.13}$$

Finally, we turn to the final mass of non-branching trajectories that move far from the origin, counted by $\sum_{i\in I_f} M_i(t)$. For any $i \in I$, If $j_{i,t}(0) = j$ and $|x_j| < 1/4$ then in order to have $\sup_{s\in[0,t]} |X_{i,t}(s)| > 1/2$ the trajectory leading to $X_i(t)$ wanders a distance of at least $1/4$ from its starting position. Let B denote one-dimensional Brownian motion started from the origin. By the reflection principle and the fact that $\mathbb{P}\left[G > x\right] \le e^{-x^2/2}$ for G a standard normal and for all $x > 0$, we have

$$\mathbb{P}\left[\sup_{s\le t} |B(s)| > 1/4\right] \le 4\mathbb{P}\left[B(t) > \frac{1}{4}\right] \le 4\exp(-1/(32t)) .$$

Since an individual trajectory of X has the law of Brownian motion, for a particle starting at distance less than $1/4$ from the origin whose trajectory never branched, the above is a bound on the probability the trajectory attained distance $1/2$ from the origin. It follows that

$$\sum_{i\in I_f} M_i(t) \preceq_{\mathrm{st}} \sum_{j\in J} m_j \cdot \xi_j ,$$

where the terms ξ_j are i.i.d. $\mathrm{Ber}(4\exp(-1/(32t)))$. The variance of the latter sum is bounded by $A \cdot 4\exp(-1/(32t))$, so Theorem 2.B.1 yields that for any fixed $b > 0$,

$$\mathbb{P}\left[\sum_{i\in I_f} M_i(t) > (b + 4\exp(-1/(32t)))A\right] \le \left(\frac{4e^{1-1/(32t)}}{b}\right)^{bA} < e^{-200A} , \tag{2.14}$$

the final inequality for t sufficiently small.

We now combine (2.12), (2.13) and (2.14) to control $\sum_{i\in I} M_i$ using (2.11). This yields that for t sufficiently small, and in particular provided that $2^{2+\delta}t^\delta < e^{-200}$, $((4e^{1-1/(32t)})/b)^b < e^{-200}$ and that

$$2\delta + \max(c, e^{-tcA}) + b + 4\exp(-1/(32t)) < \frac{1}{24}$$

we have

$$\mathbb{P}\left[\sum_{i\in I} M_i \geq A/24\right] \leq 2e^{-200A}\,.$$

It can be checked that taking $\delta = b = c = 1/100$ does the job when $t > 100\log 100/A$ (so that $\max(c, e^{-tcA}) = 1/100$) and t is sufficiently small (it is in order to satisfy these simultaneously that we require a lower bound on A). This completes the proof. □

Proof of Fact 2.4.2. Let $\mathbb{Z}/2 = \{y/2 : y \in \mathbb{Z}\}$. Define the event

$$E = \{\max\{|X_i(r)|, i \geq 1, 0 \leq r \leq s + 10^5/N\} \leq 3s\}\,.$$

Any unit interval $[x-1/2, x+1/2]$ is covered by at most two intervals from $\{[y-1/2, y+1/2] : y \in \mathbb{Z}/2, y \in [x-1/2, x+1/2]\}$. It follows that on E, if $\tau_k < s$ but $\sup\{z(t), t \in [\tau_k, \tau_k + 10^5/N]\} > 10N$ then there is y with $y \in [-3s, 3s] \cap \mathbb{Z}/2$ such that

$$\sup_{t\in[\tau_k,\tau_k+10^5/N]} \sum_{\{i:|X_i(t)-y|<1/2\}} M_i(t) > 5N.$$

When $k < I$ we have $\tau_k < s$, so

$$\begin{aligned}&\mathbb{P}\left[\sup_{t\in[\tau_k,\tau_k+10^5/N]} z(t) > 10N, z(\tau_k) \leq N, k < I\right]\\ &\quad\leq \mathbb{P}\left[E^c\right] + \sum_{y\in[-3s,3s]\cap\mathbb{Z}/2} \mathbb{P}\left[\sup_{t\in[\tau_k,\tau_k+10^5/N]} z(t,y) > 5N, z(\tau_k) \leq N\right]. \qquad (2.15)\end{aligned}$$

Our bound on the above summands works identically for each $y \in [-3s, 3s] \cap \mathbb{Z}/2$; we explain it for $y = 0$ to avoid notational overload. So we wish to bound

$$\mathbb{P}\left[\sup_{t\in[\tau_k,\tau_k+10^5/N]} z(t,0) > 5N, z(\tau_k) \leq N\right].$$

Our strategy is as follows: we use Corollary 2.4.5 to show that with high probability, for all $t \in [\tau_k, \tau_k + 10^5/N]$ the total contribution to $z(t,0)$ from descendants of particles with $|X_i(\tau_k)| \leq 3/2$ is at most $4N$. We then use Corollary 2.4.7 to show that with high probability the contribution to $z(t,0)$ from descendants of further-off particles decreases quadratically (as a function of $|X_i(\tau_k)|$); since the quadratic series converges, this implies a bound on the total contribution from far-off particles. We now proceed to details.

For $n \in \mathbb{Z}$ let

$$Y_n = \sup_{t \in [\tau_k, \tau_k + 10^5/N]} \sum_{\{i : |X_i(t)| \leq 1/2, |X_{i,t}(\tau_k) - n| \leq 1/2\}} M_i(t);$$

Y_n counts the greatest contribution at any time $t \in [\tau_k, \tau_k + 10^5/N]$, to the mass near 0 from particles that at time τ_k are near n. We clearly have

$$\sup_{t \in [\tau_k, \tau_k + 10^5/N]} z(t, 0) \leq \sum_{n \in \mathbb{Z}} Y_n \,. \tag{2.16}$$

As sketched above, we bound the sum in two parts: the contribution from Y_{-1}, Y_0 and Y_1 is handled separately from the rest, and we do this first. Note that since masses decrease with time,

$$Y_{-1} + Y_0 + Y_1 \leq \sup_{t \in [\tau_k, \tau_k + 10^5/N]} \sum_{\{i : |X_{i,t}(\tau_k)| \leq 3/2\}} M_{i,t}(\tau_k).$$

If $z(\tau_k) \leq N$ then $\sum_{\{i : |X_i(\tau_k)| \leq 3/2\}} M_i(\tau_k) \leq 3N$ so, by Corollary 2.4.5 and the strong Markov property,

$$\begin{aligned} \mathbb{P}\left[Y_{-1} + Y_0 + Y_1 > 4N, z(\tau_k) \leq N\right] &\leq 2\left(2^{1+1/3}(1 - e^{-10^5/N})^{1/3}\right)^{3N} \\ &\leq (20^5/N)^N \,. \end{aligned}$$

Now consider $n \in \mathbb{Z}$ with $|n| \geq 2$, and assume by symmetry that $n > 0$. If $|X_i(t)| \leq 1/2$ but $|X_{i,t}(\tau_k) + n| \leq 1/2$ then $X_i(t) - X_{i,t}(\tau_k) \geq n - 1$. Assuming $z(\tau_k) \leq N$, in particular we have $z(\tau_k, -n) \leq N$. Furthermore,

$$Y_{-n} \leq \sup_{t \in [\tau_k, \tau_k + 10^5/N]} \sum_{\{i : X_i(t) - X_{i,t}(\tau_k) > n-1, X_{i,t}(\tau_k) + n \leq 1/2\}} M_{i,t}(\tau_k) \,.$$

When $n \geq 2$, applying Corollary 2.4.7 with $t_0 = 10^5/N$, $L = n-1$, $v = 1/(20(n-1)^2)$ and $x = N$, we then obtain that

$$\begin{aligned}\mathbb{P}\left[Y_{-n} > \frac{N}{10(n-1)^2}, z(\tau_k) \leq N\right] &\leq 40(n-1)^2 \exp\left(\frac{10^5}{N} - \frac{N(n-1)^2}{2\cdot 10^5}\right) \\ &< \exp\left(-\frac{N(n-1)^2}{3\cdot 10^5}\right) \\ &\leq s^{-10(n-1)^2}\,. \qquad (2.17)\end{aligned}$$

The final inequality holds since $N = N(s) = 10^7 \log s$; the second inequality holds provided N is sufficiently large. We emphasize that once N is large enough the inequality holds for *all* $n \geq 2$. Note that by symmmetry the same bound also holds for Y_n.

Using (2.16) and the two preceding probability bounds (and the fact that $(1/10)\sum_{|n|\geq 2}(n-1)^{-2} = \pi^2/30 < 1$), we thus have

$$\begin{aligned}&\mathbb{P}\left[\sup_{t\in[\tau_k,\tau_k+10^5/N]} z(t,0) > 5N, z(\tau_k) \leq N\right] \\ &\leq \mathbb{P}\left[Y_{-1}+Y_0+Y_1 > 4N, z(\tau_k) \leq N\right] + \sum_{\{n\in\mathbb{Z}:|n|\geq 2\}} \mathbb{P}\left[Y_n \geq \frac{N}{10(n-1)^2}, z(\tau_k) \leq N\right] \\ &< \left(\frac{20^5}{N}\right)^N + \sum_{|n|\geq 2} s^{-10(n-1)^2} \\ &< \left(\frac{20^5}{N}\right)^N + 4s^{-10}, \qquad (2.18)\end{aligned}$$

where the last inequality holds for s sufficiently large. The same argument yields the same bound with $z(t,y)$ in place of $z(t,0)$, and (2.15) then gives

$$\begin{aligned}&\mathbb{P}\left[\sup_{t\in[\tau_k,\tau_k+10^5/N]} z(t) > 10N, z(\tau_k) \leq N, k < I\right] \\ &\leq \mathbb{P}\left[E^c\right] + (12s+2)\cdot\left(\left(\frac{20^5}{N}\right)^N + 4s^{-10}\right) \\ &\leq \mathbb{P}\left[E^c\right] + s^{-8}\,,\end{aligned}$$

the latter bound holding for s large, since $N = 10^7 \log s$. To conclude, we use the fact that

$$\mathbb{P}\left[\max\{|X_i(s+10^5/N)|, i \geq 1\} \geq 3s \;\middle|\; E^c\right] \geq \frac{1}{2},$$

which follows by considering the stopping time $\tau = \inf\{r : \max\{|X_i(r)|, i \geq 1\} \geq 3s\}$ and using symmetry. This yields

$$\begin{aligned} \mathbb{P}\left[E^c\right] &\leq 2\mathbb{P}\left[\max\{|X_i(s+10^5/N)|, i \geq 1\} \geq 3s\right] \\ &\leq 4\mathbb{E}\left[\#\{i : X_i(s+10^5/N) \geq 3s\}\right] \\ &= 4e^{s+10^5/N}\mathbb{P}\left[(s+10^5/N)^{1/2}G \geq 3s\right] \\ &\leq e^{-3s} \qquad (2.19) \\ &< s^{-8}, \end{aligned}$$

where G is a standard normal random variable and the last two inequalities hold for s sufficiently large. □

Proof of Fact 2.4.3. The proof has aspects which will be familiar from the previous proof; we describe these first. We recycle the event E from the preceding proof. Note that on $E \cap \{k < I\}$ we have

$$z(\tau_k + 10^5/N) \leq 2 \sup_{y \in [-3s,3s] \cap \mathbb{Z}/2} z(\tau_k + 10^5/N, y),$$

so

$$\begin{aligned} &\mathbb{P}\left[z(\tau_k + 10^5/N) \geq N-1, z(\tau_k) \leq N, k < I, E\right] \\ &\quad \leq \sum_{y \in [-3s,3s] \cap \mathbb{Z}/2} \mathbb{P}\left[z(\tau_k + 10^5/N, y) > \frac{N-1}{2}, z(\tau_k) \leq N\right]. \qquad (2.20) \end{aligned}$$

We once again focus on the case $y = 0$ for notational simplicity. We write

$$Z_n = \sum_{\{i : |X_i(\tau_k + 10^5/N)| < 1/2, |X_{i,\tau_k+10^5/N}(\tau_k) - n| < 1/2\}} M_i(\tau_k + 10^5/N).$$

The indices of summation correspond to particles with position near 0 at time τ_k+10^5/N, whose time τ_k ancestor had position near n. We have

$$z(\tau_k+10^5/N,0) \leq \sum_{n\in\mathbb{Z}} Z_n.$$

Now similarly to the argument leading to (2.17), apply Proposition 2.4.6 with $t=10^5/N$, $L=n-1$, $v=1/(40(n-1)^2)$ and $x=N$ to bound Z_n for $|n|\geq 2$. We obtain that for s sufficiently large (since $(1/40)\sum_{|n|\geq 2}(n-1)^{-2}=\pi^2/120<1/4$)

$$\begin{aligned}
&\mathbb{P}\left[z(\tau_k+10^5/N,0)\geq \frac{N-1}{2}, z(\tau_k)\leq N\right] \\
&\quad\leq \mathbb{P}\left[Z_{-1}+Z_0+Z_1\geq \frac{N}{4}, z(\tau_k)\leq N\right]+\sum_{\{n\in\mathbb{Z}:|n|\geq 2\}}\mathbb{P}\left[Z_n\geq \frac{N}{40(n-1)^2}, z(\tau_k)\leq N\right] \\
&\quad\leq \mathbb{P}\left[Z_{-1}+Z_0+Z_1\geq \frac{N}{4}, z(\tau_k)\leq N\right]+4s^{-10}\,. \qquad (2.21)
\end{aligned}$$

We now bound $Z_{-1}+Z_0+Z_1$ from above by the *total mass at time* τ_k+10^5/N *of individuals whose time-*τ_k *ancestor lies in* $[-3/2,3/2]$. More precisely, recall that $X_{i,t}(s)$ is the (location of) the time-s ancestor of $X_i(t)$, and write

$$D_\ell = \sum_{\{i:X_{i,\tau_k+10^5/N}(\tau_k)\in[\ell/2,(\ell+1)/2]\}} M_i(\tau_k+10^5/N).$$

Then

$$Z_{-1}+Z_0+Z_1 \leq \sum_{\ell\in[-3,2]\cap\mathbb{Z}} D_\ell\,.$$

This holds because the time-τ_k ancestors of particles counted by $Z_{-1}+Z_0+Z_1$ all lie in $[-3/2,3/2]=\bigcup_{\ell\in[-3,2]\cap\mathbb{Z}}[\ell/2,(\ell+1)/2]$. The bound may be strict because particles counted by $Z_{-1}+Z_0+Z_1$ are additionally required to lie near 0 at time τ_k+10^5/N.

Bounding each of the summands D_ℓ by the largest summand, we then have

$$Z_{-1}+Z_0+Z_1 \leq 6\max_{\ell\in[-3,2]\cap\mathbb{Z}} D_\ell\,,$$

so

$$\mathbb{P}\left[Z_{-1}+Z_0+Z_1 \geq \frac{N}{4}, z(\tau_k) \leq N\right]$$
$$\leq 6 \max_{\ell \in [-3,2]\cap\mathbb{Z}} \mathbb{P}\left[D_\ell > \frac{N}{24}, z(\tau_k) \leq N\right]$$

The final probabilities are not hard to bound: if D_ℓ hearkens from a total time-τ_k mass which is very small then at time $\tau_k + 10^5/N$ it is still rather small by Corollary 2.4.5. On the other hand, if the aggregate mass of its time-τ_k ancestors was larger (but still at most N) then by Lemma 2.4.8, at time $\tau_k + 10^5/N$ that ancestral population has lost most of its mass.

More precisely, since $M_i(\tau_k + 10^5/N) \leq M_{i,\tau_k+10^5/N}(\tau_k)$ for each i, by Corollary 2.4.5 and the strong Markov property,

$$\mathbb{P}\left[D_\ell > \frac{N}{24} \,\middle|\, \sum_{\{j: X_j(\tau_k)\in[\ell/2,(\ell+1)/2]\}} M_j(\tau_k) \leq N/48\right] \leq 2(4(1-e^{-10^5/N}))^{N/48}$$
$$\leq 2\left(\frac{20^5}{N}\right)^{N/48}.$$

Now assume that $N = N(s) = 10^7 \log s > 48C$, where C is the constant from Lemma 2.4.8. By that lemma, with $t = 10^5/N$, since $t > 500/A$ for $A \in [N/48, N]$,

$$\mathbb{P}\left[D_\ell > \frac{N}{24} \,\middle|\, \sum_{\{j: X_j(\tau_k)\in[\ell/2,(\ell+1)/2]\}} M_j(\tau_k) \in [N/48, N]\right] \leq 2e^{-200N/48} < e^{-4N},$$

the latter inequality for $N = N(s)$ sufficiently large. This bound holds for each $\ell \in [-3,2]\cap\mathbb{Z}$. Under the assumption that $z(\tau_k) \leq N$, one of the conditions in the above conditional probabilities must occur. It follows that

$$6 \max_{\ell \in [-3,2]\cap\mathbb{Z}} \mathbb{P}\left[D_\ell > \frac{N}{24}, z(\tau_k) \leq N\right] \leq 6 \max\left(2\left(\frac{20^5}{N}\right)^{N/48}, e^{-4N}\right),$$

so for N sufficiently large

$$\mathbb{P}\left[Z_{-1}+Z_0+Z_1 \geq \frac{N}{4}, z(\tau_k) \leq N\right] \leq 6e^{-4N} = 6s^{-4\cdot 10^7}.$$

Combined with (2.21) this gives

$$\mathbb{P}\left[z(\tau_k + 10^5/N, 0) \geq \frac{N-1}{2}, z(\tau_k) \leq N\right] < 5s^{-10}.$$

The same bound holds for each $z(\tau_k + 10^5/N, y)$, so using (2.20) and the bound $\mathbb{P}\left[E^c\right] \leq e^{-3s}$ from (2.19), for s large we obtain

$$\mathbb{P}\left[z(\tau_k + 10^5/N) \geq N-1, z(\tau_k) \leq N, k < I\right] \leq (12s+2)5s^{-10} + e^{-3s} < s^{-8}\,.$$

The proof is almost complete; to finish it off we need to deal with the event $\{z(\tau_k) \leq N\}$ in the preceding probability. To do so we use induction. First, for s large, since $N = N(s) = 10^7 \log s$ and $\tau_0 = \inf\{t : z(t) \geq N-1\}$, we have $z(\tau_0-) \leq N-1$. It follows that $z(\tau_0) \leq z(\tau_0-)+1 \leq N$ (this was explained in the proof of Proposition 2.4.1), so when $k=0$ we have

$$\mathbb{P}\left[z(\tau_k + 10^5/N) \geq N-1, k < I\right] = \mathbb{P}\left[z(\tau_k + 10^5/N) \geq N-1, z(\tau_k) \leq N, k < I\right] \leq s^{-8}\,.$$

For larger k, similarly if $z(\tau_{k-1} + 10^5/N) \leq N-1$ then $z(\tau_k) \leq z(\tau_k-) + 1 \leq N$. We thus have

$$\begin{aligned}
\mathbb{P}\left[z(\tau_k + 10^5/N) \geq N-1, k < I\right] &\leq \mathbb{P}\left[z(\tau_k + 10^5/N) \geq N-1, z(\tau_k) \leq N, k < I\right] \\
&\quad + \mathbb{P}\left[z(\tau_k) > N, k < I\right] \\
&\leq s^{-8} + \mathbb{P}\left[z(\tau_{k-1} + 10^5/N) \geq N-1, k-1 < I\right],
\end{aligned}$$

so by induction and the hypothesis that $k \leq Ns/10^5$,

$$\mathbb{P}\left[z(\tau_k + 10^5/N) \geq N-1, k < I\right] \leq (k+1)\cdot s^{-8} < Ns^{-7} < s^{-6}. \qquad \square$$

2.5 Lower bound

The next proposition restates the second inequality of Theorem 2.1.1. Recall that $c^* = 3^{4/3}\pi^{2/3}/2^{7/6}$.

Proposition 2.5.1. *For any $m \in (0,1)$, almost surely*

$$\liminf_{t\to\infty} \frac{\sqrt{2}t - D(t,m)}{t^{1/3}} \le c^*.$$

Given a function $f : [0,\infty) \to \mathbb{R}$, for $t \ge 0$ let $I(t,f) = \{i \le N(t) : \forall s \in [0,t], X_{i,t}(s) \ge f(s)\}$ be the indices of particles whose ancestral trajectory stays above f up to time t. Note that $|I(t,f)|$ is decreasing in t: if a trajectory stays above f to time t then it also stays above f to time $t' < t$. It follows that $\mathbb{P}[\forall t, I(t,f) \neq \emptyset] = \lim_{t\to\infty} \mathbb{P}[I(t,f) \neq \emptyset]$, and this is a decreasing limit. We will use the following result of Roberts (2015).

Lemma 2.5.2 (Roberts (2015), Theorem 1)**.** *Let $g(t) = \sqrt{2}t - c^* t^{1/3} + c^* t^{1/3}/\log^2(t+e) - 1$. Then*

$$\lim_{t\to\infty} \mathbb{P}[I(t,g) \neq \emptyset] = p^* > 0$$

The idea of the proof of Proposition 2.5.1 is that if the density is always low beyond g then a particle staying beyond g will have reasonably large mass at time t; the lemma guarantees that such a particle has a reasonable chance p^* of existing. The next corollary implies that at the cost of a constant shift of the function g, we may increase p^* as close to one as we like. For $c \in \mathbb{R}$ write $g - c$ for the function with $(g-c)(x) = g(x) - c$.

Corollary 2.5.3. *Let $C^* = \inf\{c : \forall t, I(t, g-c) \neq \emptyset\}$. Then almost surely $C^* < \infty$.*

Proof. The proof technique is sometimes called an amplification argument. Consider the $N(t) \approx e^t$ independent copies of the BBM rooted at time-t particles, the ith copy having initial individual at position $X_i(t)$. Suppose the "translate by $X_i(t)$" of the event from Lemma 2.5.2 occurs in the kth copy; more precisely, suppose that for all $t' \ge t$ there is a descendant $X_j(t')$ of $X_k(t)$ such that for all $s \in [t,t']$,

$$X_{j,t'}(s) - X_k(t) \ge g(s-t) \ge g(s) - \sqrt{2}t - c^* t^{1/3}.$$

For $s \le t$ we also have

$$X_{j,t'}(s) \ge \inf_{i\ge1} X_i(s) \ge \inf_{s\le t}\inf_{i\ge1} X_i(s) \ge g(s) + \inf_{s\le t}\inf_{i\ge1} X_i(s) - \sup_{s\le t} g(s)\,,$$

so in this case

$$C^* \le -\inf_{s\in[0,t]}\inf_{i\ge1} X_i(s) + \sqrt{2}t + c^* t^{1/3}.$$

By the branching property (i.e. the independence of the trajectories emanating from each of the particles $(X_i(t), i \ge 1)$), it follows that

$$\mathbb{P}\left[C^* > 3t + \sqrt{2}t + c^* t^{1/3}\right] \le \mathbb{P}\left[N(t) \le 2^t\right] + \mathbb{P}\left[\inf_{s\in[0,t]}\inf_{i\ge1} X_i(s) \le -3t\right] + (1-p^*)^{2^t}, \tag{2.22}$$

where p^* is the constant from Lemma 2.5.2. Since $N(t) \sim \mathrm{Geom}(e^{-t})$ we have $\mathbb{P}\left[N(t) \le 2^t\right] \le (2/e)^t$. Finally, let $\sigma = \inf\{s : \inf_{i\ge1} X_i(s) \le -3t\}$, so $\inf_{s\in[0,t]}\inf_{i\ge1} X_i(s) \le -3t$ if and only if $\sigma < t$. Considering the descendants of the first individual to reach position $-3t$, by symmetry we have

$$\mathbb{P}\left[\inf_{i\ge1} X_i(t) \le -3t \;\middle|\; \sigma < t\right] \ge \frac{1}{2},$$

so for G a standard normal random variable,

$$\mathbb{P}\left[\sigma < t\right] \le 2\mathbb{P}\left[\inf_{i\ge1} X_i(t) \le -3t\right] \le 2e^t\mathbb{P}\left[t^{1/2}G \le -3t\right] \le 2e^{-7t/2}.$$

These bounds and (2.22) then yield

$$\mathbb{P}\left[C^* > 3t + \sqrt{2}t + c^* t^{1/3}\right] \le (2/e)^t + 2e^{-7t/2} + (1-p^*)^{2^t}.$$

This can be made arbitrarily small by taking t large. □

In order to prove Proposition 2.5.1, we require one final lemma which shows that a small mass will quickly increase to form some region of constant density within a constant distance.

Lemma 2.5.4. *For all* $\varepsilon > 0$ *and* $m \in (0,1)$ *there is* $C > 0$ *such that for all* $\mathbf{x} \in \mathbb{R}^k$ *and* $\mathbf{m} \in (0,1]^k$, *if* $z := \sum_{\{i:|x_i|<1\}} m_i > 0$ *then*

$$\mathbb{P}_{\mathbf{x},\mathbf{m}}\left[\exists t \in [0, C(1+\log(1/z))], x \in [-C, C] : \zeta(t,x) \ge m\right] \ge 1 - \varepsilon. \tag{2.23}$$

To prove the lemma we use the following fact, whose proof is left to the reader.

Fact 2.5.5. *For all* $\varepsilon > 0$*, there are* $t_0 = t_0(\varepsilon)$ *and* $c = c(\varepsilon) > 0$ *such that*

$$\mathbb{P}\left[\forall t \geq t_0, \#\{i : \forall s \in [0,t], |X_{i,t}(s)| < c\} \geq (e-\varepsilon)^t\right] > 1-\varepsilon. \tag{2.24}$$

One straightforward way to prove the fact is as follows. Let $p(c)$ denote the survival probability of branching Brownian motion with branching rate 1 and absorbing boundaries at $-c$ and c, started from the origin. By Theorems 2.1 and 2.2 in Watanabe (1965) (originally due to Sevast'yanov (1958)), $p(c) > 0$ for c sufficiently large. Then for large $R > 0$, $p(R^2)$ is the survival probability of BBM with branching rate R^2 and absorbing boundaries at $-R$ and R, so an amplification argument gives us that $p(R^2) \to 1$ as $R \to \infty$. Then use Theorems 3.2 and 3.3 in Watanabe (1965) to give the result.

As an aside, we note the very nice recent work Harris et al. (2016) on the asymptotics of this survival probability for c near the critical width $\hat{c}$ below which $p(c) = 0$.

Proof of Lemma 2.5.4. The claim is clearly true if $z \geq m$, and we hereafter assume $z \in (0,m)$. We also assume ε is small enough that $(e-\varepsilon)e^{-m}(1-\varepsilon^{1/2}) > (1+\varepsilon)$; this can only make our job harder.

By relabelling, we may assume that for some $1 \leq k' \leq k$ we have $|x_i| < 1$ for $1 \leq i \leq k'$ and $|x_i| > 1$ for $i > k'$, so that $z = \sum_{1 \leq i \leq k'} m_i$.

For $1 \leq i \leq k'$ let $J_i(t)$ index the time-t descendants of x_i whose trajectory stays fairly near the origin, i.e.,

$$J_i(t) = \{\ell \geq 1 : j_{\ell,t}(0) = i, |X_{\ell,t}(s) - x_i| < c \ \forall s \in [0,t]\},$$

where $c = c(\varepsilon)$ is chosen as in Fact 2.5.5. By that fact, we then have for $t_0 = t_0(\varepsilon)$

$$\mathbb{P}\left[\forall t \geq t_0, |J_i(t)| \geq (e-\varepsilon)^t\right] > 1-\varepsilon. \tag{2.25}$$

We hereafter assume $t \geq t_0(\varepsilon)$. Now suppose that $\zeta(s, x) < m$ for all $s \leq t$ and $|x| \leq c+1$. Then for each $1 \leq i \leq k'$, for all $j \in J_i(t)$, $M_j(t) \geq m_i \cdot e^{-mt}$, so

$$\sum_{1\leq i\leq k'} \sum_{j\in J_i(t)} M_j(t) \geq e^{-mt} \sum_{1\leq i\leq k'} m_i |J_i(t)| \geq e^{-mt}(e-\varepsilon)^t \sum_{1\leq i\leq k'} m_i \mathbf{1}_{[|J_i(t)|\geq(e-\varepsilon)^t]} \,.$$

By Markov's inequality and (2.25), since $z = \sum_{1\leq i\leq k'} m_i$,

$$\mathbb{P}\left[\sum_{1\leq i\leq k'} m_i \mathbf{1}_{[|J_i(t)|<(e-\varepsilon)^t]} \geq z\varepsilon^{1/2}\right] \leq \varepsilon^{1/2},$$

so with probability at least $1-\varepsilon^{1/2}$,

$$\sum_{1\leq i\leq k'} \sum_{j\in J_i(t)} M_j(t) \geq e^{-mt}(e-\varepsilon)^t(1-\varepsilon^{1/2}) \cdot z.$$

By our assumption on ε, we have $e^{-m}(e-\varepsilon) > (1+\varepsilon)/(1-\varepsilon^{1/2}) > 1+2\varepsilon$, so this gives

$$\sum_{\{j:|X_j(t)|<c+1\}} M_j(t) \geq (1+2\varepsilon)^{t-1} z > c+2\,,$$

the last inequality provided that $t \geq 1 + \log_{1+2\varepsilon}((c+2)/z)$. Since $[-c-1, c+1]$ can be covered by $\lfloor c+2 \rfloor$ intervals of radius 1, we see that in this case there is x with $|x| \leq c+1$ such that $\zeta(t, x) > 1$.

To sum up: with probability at least $1-\varepsilon^{1/2}$, if $t \geq t_0(\varepsilon)$ and $t \geq 1+\log_{1+2\varepsilon}((c+2)/z)$, either $\zeta(s, x) \geq m$ for some $s \leq t$ and $|x| \leq c+1$, or else $\zeta(t, x) > 1$ for some x with $|x| \leq c+1$. By choosing $C = C(\varepsilon)$ appropriately, we obtain

$$\mathbb{P}_{\mathbf{x},\mathbf{m}}\left[\exists s \in [0, C(1+\log(1/z))], x \in [-C, C] : \zeta(t, x) \geq m\right] \geq 1-\varepsilon^{1/2}. \qquad \square$$

We are now ready for the final proof of the chapter.

Proof of Proposition 2.5.1. Fix $m \in (0, 1)$. Fix $\varepsilon > 0$ and by Corollary 2.5.3 choose $L > 1$ large enough that $\mathbb{P}[C^* \geq L] < \varepsilon$. Fix t much larger than L (so that $\log\log t > L$, say). Let $Z = 2 \cdot 10^8$ and define the event

$$E = \{ \sup_{s \in [0,t]} \sup_{x \in \mathbb{R}} \zeta(s,x) \leq Z \log t\},$$

and note that for t sufficiently large, $\mathbb{P}\left[E^c\right] \leq t^{-4} \leq \varepsilon$ by Proposition 2.4.1.

Let $\sigma = \inf\{s \geq t^{1/4} : D(s, 1/t) \geq g(s) - C^* - 1\}$. We first suppose that $\sigma > t$, so that for all $s \in [t^{1/4}, t]$ we have $D(s, 1/t) < g(s) - C^* - 1$. Let i^* be such that $X_{i^*,t}(s) \geq g(s) - C^*$ for all $s \in [0,t]$; such i^* exists by the definition of C^*. If E occurs then we have

$$\begin{aligned} -\log M_{i^*}(t) &= \int_0^t \zeta(s, X_{i^*,t}(s))\mathrm{d}s \\ &\leq \int_0^{t^{1/4}} \zeta(s, X_{i^*,t}(s))\mathrm{d}s + \int_{t^{1/4}}^t \frac{1}{t}\mathrm{d}s \\ &\leq Zt^{1/4}\log t + 1\,, \end{aligned}$$

where the last bound holds since when E occurs the integrand is at most $Z \log t$.

Let $C = C(\varepsilon)$ be the constant from Lemma 2.5.4. Then given that E occurs, by that lemma (applied with $z = M_{i^*}(t) \geq \exp(-1 - Zt^{1/4}\log t)$) and the Markov property, with probability at least $1 - \varepsilon$ there is $s \in (t, t + C(2 + Zt^{1/4}\log t))$ and x with $|x| \leq C$ such that $\zeta(s, X_{i^*}(t) + x) \geq m$. If this occurs, and additionally $C^* \leq L$, we have

$$D(s,m) \geq X_{i^*}(t) - C \geq g(t) - C^* - C \geq g(s) - s^{1/4}\log^2 s\,,$$

the last bound holding for all t sufficiently large since $s - t \leq C(2 + Zt^{1/4}\log t)$, and for s and t large we have $g(s) - g(t) = \mathcal{O}(s - t)$. We thus have

$$\begin{aligned} &\mathbb{P}\left[\forall s \geq t, D(s,m) < g(s) - s^{1/4}\log^2 s \mid \sigma > t\right] \\ \leq &\mathbb{P}\left[\{\forall s \geq t, D(s,m) < g(t) - C^* - C\} \cap E \mid \sigma > t\right] \\ &+ \mathbb{P}\left[C^* \geq L \mid \sigma > t\right] + \mathbb{P}\left[E^c \mid \sigma > t\right] \\ \leq &\varepsilon + \mathbb{P}\left[C^* \geq L \mid \sigma > t\right] + \mathbb{P}\left[E^c \mid \sigma > t\right]\,. \end{aligned} \tag{2.26}$$

Next suppose that $\sigma \leq t$. Apply the strong Markov property at time σ, and apply Lemma 2.5.4 just as above (but with a starting mass in $[D(\sigma, 1/t) - 1, D(\sigma, 1/t) + 1]$ of at least $1/t = e^{-\log t}$ rather than $e^{-1-Zt^{1/4}\log t}$). We obtain that with probability at least $1-\varepsilon$ there is $s \in (\sigma, \sigma + C(1+\log t))$ such that

$$D(s,m) \geq g(\sigma) - C - C^* - 1 \geq g(s) - \log^2 s\,,$$

the last bound holding for t sufficiently large since $s - \sigma \leq C(1+\log t)$ and $\log t \leq 4\log\sigma \leq 4\log s$, and under the assumption $C^* \leq L$.

Since $\sigma \geq t^{1/4}$ and $\log^2 s < s^{1/4}\log^2 s$, it follows that

$$\begin{aligned}&\mathbb{P}\left[\forall s \geq t^{1/4}, D(s,m) < g(s) - s^{1/4}\log^2 s \mid \sigma \leq t\right] \\ &\leq \varepsilon + \mathbb{P}\left[C^* \geq L \mid \sigma \leq t\right].\end{aligned}$$

Now combine this with (2.26) using the law of total probability. We chose L and t large enough that $\mathbb{P}[C^* \geq L] \leq \varepsilon$ and $\mathbb{P}[E^c] \leq \varepsilon$, so we obtain

$$\mathbb{P}\left[\exists s \geq t^{1/4} : D(s,m) \geq g(s) - s^{1/4}\log^2 s\right] \geq 1 - 3\varepsilon\,.$$

Finally, if $D(s,m) \geq g(s) - s^{1/4}\log^2 s$ then

$$\frac{\sqrt{2}s - D(s,m)}{s^{1/3}} \leq c^* - \frac{c^*}{\log^2(s+e)} + \frac{1}{s^{1/3}} + \frac{\log^2 s}{s^{1/12}}\,,$$

which tends to c^* as $s \to \infty$. □

2.6 Discussion and questions

- The analysis of the paper should carry through fairly straightforwardly to higher dimensions $\mathbb{R}^k$, provided we redefine $d(t,m)$ and $D(t,m)$ as

$$d(t,m) = \min\{|x| : \zeta(t,x) < m\}, \quad D(t,m) = \max\{|x| : \zeta(t,x) > m\}\,.$$

At time t, the density is then at least m within the ball of radius $d(t,m)$ around 0, and less than m outside the ball of radius $D(t,m)$ around 0. The proof of the lower bound is then the same as in Sections 4 and 5. The proof of the upper bound requires ruling out the possibility that the modulus of a particle in the BBM stays ahead of a moving barrier g even though it cannot have consistent displacement more than g in any fixed direction. In order for our proof techniques to carry over, this requires sample path estimates for $\mathrm{Bes}(k)$ processes analogous to the ones derived in this chapter for Brownian motion. We expect such estimates to hold for all $k \geq 1$, though verifying this may be technical.

- We believe that Proposition 2.5.1 predicts the "true" front location, in that both $D(t,m)$ and $d(t,m)$ are typically at distance $o(t^{1/3})$ from $\sqrt{2}t - c^* t^{1/3}$ when t is large. This is our justification for the remark in the final paragraph of Section 2.1.
- In the same way as the KPP equation describes the evolution of multiplicative functionals of BBM (McKean (1975)), it seems plausible that the model proposed in this work (or a related model) should be connected to an equation of the form

$$u_t = \frac{1}{2}u_{xx} + u - u \int_{\{y:|y-x|<1\}} u(t,y)\mathrm{d}y.$$

This equation has steady states at 0 (unstable) and 1/2 (stable), and is redolent of a family of "non-local" Fisher-KPP-type equations which was introduced in Britton (1989) to model populations in which aggregation can have both a competitive advantage (safety in numbers) and disadvantage (due to competition for resources). These equations have received substantial study (Berestycki et al. (2009a); Britton (1990); Gourley (2000)); the survey Volpert and Petrovskii (2009) contains many further references, as well as perspective on the biological motivations for such study.

If a probabilistic model for such an equation were found, it could yield new results on, e.g., the front propagation speed or temporal fluctuations of solutions to the above equation. Conversely, a glance at that literature suggests new probabilistic questions: for example, what if the effect of competition is described by a kernel

κ, where $\kappa(|x-y|)$ describes the degree of competition for resources between individuals at spatial positions x and y? In our model we took $\kappa(|x-y|) = \mathbf{1}_{[|x-y|\in(0,1)]}$; a kernel which allows substantial long-range interaction might yield rather different dynamics.

- As mentioned in the introduction, one may reasonably consider the mechanism for mass growth in our model – both children inherit the mass of the parent – nonphysical. More physically realistic (at least for amoebae) is for the children to each have half the mass of the parent. One must also then change the rules to allow for mass growth; a reasonable modification is to take

$$\overline{\zeta}(t,x) = \sum_{\{i:|X_i(t)-x|\leq 1\}} M_i(s),$$

and

$$M_i(t) = \exp\Big(\int_0^t (1-\overline{\zeta}(s,X_i(s)))\mathrm{d}s\Big)\,.$$

In other words, the mass of an individual can increase, when there is little nearby competition for resources – but the larger particles get, the harder it is for them to sustain themselves. The key point is that 1 is still a universal upper bound on the greatest mass of any particle.

We conjecture that any lack of physical realism in our model is relatively insignificant for the long term behaviour, and more concretely that the front location behaves similarly in the two models. As partial evidence for this, we note that the analyses from Sections 2.4 and 2.5 carry through essentially unchanged for the model described above.

The argument from Section 2.3, however, breaks down, because a particle moving through an environment of constant density $m < 1$ will have mass which does *not* decay exponentially, even when the loss of mass due to branching is taken into account. Instead, such a particle will (at large times) have a mass which is random and typically of order $\Theta(1-m)$.

Because of this, the existing argument only establishes Proposition 2.3.1 in a highly weakened form, with the condition $m \geq 1$ rather than $m > 0$. (It is possible to do very slightly better, by considering a variable bound $m = m(t)$. One can then take $m(t) < 1$ if $1 - m(t)$ decays sufficiently quickly, but the pain-to-gain ratio in writing down such an argument in detail does not seem favourable.) But $m \in (0, 1)$ is the really meaningful region. Proving a genuine analogue of Proposition 2.3.1 for this model seems to us the only missing step to a proof of Theorem 2.1.1 for the modified dynamics.

- In the variant just described, one intriguing possibility is that there may now be particles with mass $\Theta(1)$ at large times. If there are, they will be found near the front, since that is where they can find food. Do they really exist?

- More generally, one may take

$$M_i(t) = \exp\left(\int_0^t (a - b\bar{\zeta}(s, X_i(s)))\mathrm{d}s\right).$$

  This looks, heuristically, like some sort of spatial logistic growth (Etheridge (2004); Lambert (2005)). It may be interesting to investigate what different behaviours can occur as the parameters a and b are varied.

Appendix 2.A Estimates for the upper bound

We first turn to the proof of Lemma 2.3.3. The proof relies on the following sample path estimate for Brownian motion.

Lemma 2.A.1. *Suppose $f : [0, t] \to \mathbb{R}$ and $L : [0, t] \to (0, \infty)$ are twice continuously differentiable functions, with $f(0) < 0$, $f(0) + L(0) > 0$ and f increasing. We assume that there exists a constant $Q > 0$ such that*

$$|L'(0)|L(0) + |L'(u)|L(u) + \int_0^u |L''(s)|L(s)\, ds + \int_0^u |f''(s)|L(s)\, ds - |L'(0)|f(0) \le Q$$

for all $0 \le u \le t$, which we call Assumption (A). Then there is a constant $M(Q)$ such that for $0 \le p \le 1$,

$$\mathbb{P}\left[B(s) - f(s) \in (0, L(s)) \forall s \le u, B(u) - f(u) \in (pL(u), L(u))\right]$$

$$\leq M(Q)\exp(-\tfrac{1}{2}\int_0^u f'(s)^2\,ds - \int_0^u \frac{\pi^2}{2L(s)^2}\,ds - pf'(u)L(u) - f'(0)f(0) + \tfrac{1}{2}\log L(u)).$$

This result is obtained by combining Proposition 4 and Lemma 7 in Roberts (2015) to cover the two cases $\int_0^u \frac{1}{L(s)^2}ds > 1$ and $\int_0^u \frac{1}{L(s)^2}ds \leq 1$. In order to apply Lemma 2.A.1, we exploit the existence of a solution to an integral equation; such a solution is used for a related purpose in Section 3.4 of Jaffuel (2012).

Lemma 2.A.2. *For $c < c^*$, there exists a constant $\alpha > 0$ such that the equation*

$$l(s) = \alpha + cs^{1/3} - \frac{\pi^2}{2\sqrt{2}}\int_0^s \frac{1}{l(u)^2}\,du \tag{2.27}$$

has a continuous solution on $[0,1]$ which is twice continuously differentiable on $(0,1)$ with $l(s) > 0$ for all $s \in [0,1)$ and with $l(1) = 0$.

The lemma follows from Propositions 3.2 and 3.6(iii) of Jaffuel (2012). More precisely, in those lemmas there is a variance term σ^2, and the value analogous to c^* is $a_c = \frac{3}{2}(3\pi^2\sigma^2)^{1/3}$. Taking $\sigma = 1/\sqrt{2}$ yields the above formulation.

Proof of Lemma 2.3.3. Fix $t > 0$ large. Since the integral on the RHS of (2.27) is non-negative, $l(s) \leq c + \alpha$ for all $s \in [0,1]$. Since $l(0) = \alpha > 0$, we can fix $\beta \in (0, \min(\alpha^3/8, \alpha^3/(8c^3), 1))$ sufficiently small that $l(s) \geq \alpha/2$ for $s < \beta$. Let

$$L(s) = t^{1/3}\left(\frac{1+\beta}{u_t}\right)^{1/3} l\left(\frac{(s+\beta t)u_t}{t+\beta t}\right) \tag{2.28}$$

for $0 \leq s \leq t$, where $u_t = \inf\{u \in [0,1] : l(u) \leq 2t^{-1/12}\}$. Note that $u_t \to 1$ as $t \to \infty$.

We will prove that the lemma holds for the above choice of β and with the function $\Delta(s) = L(s) - Kt^{1/6}$, provided K is sufficiently large. We must thus verify that Δ satisfies the requisite properties, and prove the bound (2.3). Write $f(s) = b(s) - Kt^{1/6} = \sqrt{2}s - c(s+\beta t)^{1/3} - Kt^{1/6}$; then to prove (2.3), it suffices to show that for $u \in [0,t]$ and all $x \in [0, L(u))$,

$$\mathbb{P}\left[B(s) \in (f(s), f(s)+L(s))\,\forall s \leq u, B(u) > f(u)+x\right] \leq \exp(-u - t^{1/3}/K + \sqrt{2}(L(u)-x)). \tag{2.29}$$

We establish this by applying Lemma 2.A.1 with the above functions f and L. We next derive the properties of f, L, L' and L'' which we require to do so.

First, note that $f(0) = -c(\beta t)^{1/3} - Kt^{1/6}$ and $L(0) \geq t^{1/3} l((\beta u_t)/(1+\beta)) \geq \alpha t^{1/3}/2$ by our choice of β, so since $\beta < (\alpha/(2c))^3$ we have $L(0) + f(0) > 0$ for t sufficiently large. For t sufficiently large, f is increasing on $[0,t]$, and f and L are twice continuously differentiable (since l is C^2 on $(0,1)$).

We assume t is sufficiently large that $(1+\beta)^{1/3} u_t^{-1/3} \leq 2$. Then for $s \in [0,t]$, by the definition of u_t, $L(s) \geq 2(1+\beta)^{1/3} u_t^{-1/3} t^{1/4} \geq 2t^{1/4}$ and since $l(r) \leq c + \alpha \ \forall r \in [0,1]$, $L(s) \leq (1+\beta)^{1/3} u_t^{-1/3} t^{1/3}(c+\alpha) \leq 2(c+\alpha)t^{1/3}$, so for all $s \in [0,t]$,

$$2t^{1/4} \leq L(s) \leq 2(c+\alpha)t^{1/3}. \tag{2.30}$$

Since l is C^2 on $(0,1)$ and $l(s) > 0$ for $s < 1$, we can differentiate both sides of (2.27) for $s \in (0,1)$ to give

$$l'(s) = \frac{1}{3} c s^{-2/3} - \frac{\pi^2}{2\sqrt{2}} \frac{1}{l(s)^2}. \tag{2.31}$$

Hence L is differentiable on $[0,t]$ with

$$L'(s) = t^{-2/3} \left(\frac{1+\beta}{u_t} \right)^{-2/3} l' \left(\frac{u_t(s+\beta t)}{t+\beta t} \right).$$

Also, for $\frac{u_t \beta}{1+\beta} \leq u \leq u_t$, by (2.31) and the definition of u_t,

$$|l'(u)| \leq \tfrac{1}{3} c(\beta u_t)^{-2/3}(1+\beta)^{2/3} + \tfrac{\pi^2}{8\sqrt{2}} t^{1/6} \leq 2t^{1/6}, \tag{2.32}$$

for t sufficiently large, so for all $s \in [0,t]$ we have

$$|L'(s)| \leq u_t^{2/3}(1+\beta)^{-2/3} t^{-2/3} 2t^{1/6} \leq 2t^{-1/2}. \tag{2.33}$$

This is a convenient moment to verify that the function $\Delta(s) = L(s) - Kt^{1/6}$ has the requisite properties. By (2.30), for t sufficiently large, $\Delta : [0,t] \to [t^{1/4}, 2(c+\alpha)t^{1/3}]$. Also $\Delta(t) = L(t) - Kt^{1/6} < 2(1+\beta)^{1/3} u_t^{-1/3} t^{1/4} \leq 4t^{1/4}$ for t sufficiently large. Finally, by (2.33), $|\Delta'(s)| \leq 2t^{-1/2} \leq 1$ for all $s \in [0,t]$, once again for t sufficiently large.

Proceeding with the proof of (2.29), we now check that Assumption (A) holds for our choice of f and L, for some constant Q which does not depend on t. For t sufficiently large, by (2.30) and (2.33) we have $\sup_{s\in[0,t]}|L'(s)L(s)| = \mathcal{O}(t^{-1/6})$, and also $|L'(0)f(0)| = \mathcal{O}(t^{-1/6})$.

By the definition of L in (2.28), for $s \in [0,t]$ we have

$$|L''(s)|L(s) = t^{-4/3}\left(\frac{1+\beta}{u_t}\right)^{-4/3}\left|l''\left(\frac{u_t(s+\beta t)}{t+\beta t}\right)\right| l\left(\frac{u_t(s+\beta t)}{t+\beta t}\right).$$

For $\frac{u_t\beta}{1+\beta} \le u \le u_t$ we also have $|l'(u)| \le 2t^{1/6}$ by (2.32) and $2t^{-1/12} \le l(u) \le c+\alpha$ by the definition of u_t. By differentiating (2.31) we obtain that, uniformly over u in the above range,

$$|l''(u)|l(u) \le \frac{2}{9}cu^{-5/3}l(u) + \frac{\pi^2}{\sqrt{2}}\frac{|l'(u)|}{l(u)^2} = \mathcal{O}(t^{1/3}),$$

and hence $\sup_{s\in[0,t]}|L''(s)|L(s) = \mathcal{O}(t^{-1})$. Finally, $\sup_{s\in[0,t]} f''(s) = \sup_{s\in[0,t]}\frac{2}{9}c(s+\beta t)^{-5/3} = \mathcal{O}(t^{-5/3})$, which with (2.30) yields $\sup_{s\in[0,t]}|f''(s)|L(s) = \mathcal{O}(t^{-4/3})$. Thus, Assumption (A) holds for some fixed constant Q not depending on t.

Having verified the conditions of Lemma 2.A.1, we now show that the bound from that lemma indeed implies (2.29).

For $0 \le u \le t$, $f'(u) = \sqrt{2} - \frac{1}{3}c(s+\beta t)^{-2/3}$ so

$$\begin{aligned}-\tfrac{1}{2}\int_0^u f'(s)^2\,ds &= -u + \sqrt{2}c(u+\beta t)^{1/3} - \sqrt{2}c(\beta t)^{1/3} + \tfrac{1}{6}c^2(u+\beta t)^{-1/3} - \tfrac{1}{6}c^2(\beta t)^{-1/3}\\ &\le -u + \sqrt{2}cu^{1/3} + \mathcal{O}(t^{-1/3}).\end{aligned}$$

Also by the definition of L in (2.28),

$$\frac{\pi^2}{2}\int_0^u \frac{1}{L(s)^2}\,du = \frac{\pi^2}{2}t^{1/3}\left(\frac{1+\beta}{u_t}\right)^{1/3}\left(\int_{\frac{\beta u_t}{1+\beta}}^{\frac{(u+\beta t)u_t}{t+\beta t}} \frac{1}{l(s)^2}\,ds\right).$$

We chose β sufficiently small that for $s \le \frac{\beta u_t}{1+\beta}$, $l(s) \ge \alpha/2$. Therefore

$$\int_0^{\frac{\beta u_t}{1+\beta}} \frac{1}{l(s)^2}\,ds \le \frac{4}{\alpha^2}\frac{\beta u_t}{1+\beta} \le \frac{\alpha}{2},$$

since we also chose $\beta < \alpha^3/8$. It follows by (2.27) that

$$\begin{aligned}\frac{\pi^2}{2}\int_0^u \frac{1}{L(s)^2}\,du &\geq \sqrt{2}t^{1/3}\left(\frac{1+\beta}{u_t}\right)^{1/3}\left(\frac{\alpha}{2}+c\left(\frac{(u+\beta t)u_t}{t+\beta t}\right)^{1/3} - l\left(\frac{(u+\beta t)u_t}{t+\beta t}\right)\right)\\ &\geq (\sqrt{2}/2)\alpha t^{1/3} + \sqrt{2}cu^{1/3} - \sqrt{2}L(u),\end{aligned}$$

where the second line follows by the definition of L in (2.28). Hence for $u \leq t$ and $p \in [0,1]$,

$$\begin{aligned}&\exp(-\tfrac{1}{2}\int_0^u f'(s)^2\,ds - \int_0^u \frac{\pi^2}{2L(s)^2}\,ds - pf'(u)L(u) - f'(0)f(0) + \tfrac{1}{2}\log L(u))\\ &\leq \exp(-u - (\sqrt{2}/2)\alpha t^{1/3} + \sqrt{2}(1-p)L(u) + \sqrt{2}c\beta^{1/3}t^{1/3} + \mathcal{O}(t^{1/6}))\\ &\leq (M(Q))^{-1}\exp(-u - t^{1/3}/K + \sqrt{2}(1-p)L(u)). \qquad (2.34)\end{aligned}$$

The last inequality holds for all large t, provided K is sufficiently large that $(\sqrt{2}/2)\alpha - \sqrt{2}c\beta^{1/3} > 1/K$; this is possible by our choice of β.

Writing $p = x/L(u)$, then

$$\begin{aligned}&\mathbb{P}\left[B(s) \in (f(s), f(s)+L(s))\,\forall s \leq u, B(u) > f(u)+x\right]\\ &\quad= \mathbb{P}\left[B(s)-f(s) \in (0, L(s))\forall s \leq u, B(u)-f(u) \in (pL(u), L(u))\right]\\ &\quad\leq \exp(-u - t^{1/3}/K + \sqrt{2}(1-p)L(u))\end{aligned}$$

for some $K > 0$ by Lemma 2.A.1 and (2.34). This establishes (2.29) and completes the proof. □

We now turn to the proof of Lemma 2.3.5, during which we will use the following fact.

Fact 2.A.3. *Let* $(W(u), 0 \leq u \leq 1)$ *be either Brownian excursion or Brownian meander, and let* N *be a standard Gaussian. Then*

$$\mathbb{P}\left[\max_{u\leq 1} W(u) \geq x\right] \leq 6\mathbb{P}\left[N \geq x/4\right].$$

Proof of Fact 2.A.3. Write B, B^{me}, B^{ex} and B^{br} for Brownian motion, meander, excursion, and bridge, all of length one. In what follows, maxima are always over $u \in [0,1]$ even if this is not explicitly written.

We have

$$\max B^{\mathrm{ex}} \stackrel{\mathrm{d}}{=} \max B^{\mathrm{br}} - \min B^{\mathrm{br}} \leq 2 \max |B^{\mathrm{br}}| \stackrel{\mathrm{d}}{=} 2 \max_{u \leq 1} |B(u) - uB(1)| \leq 4 \max |B|,$$

so by the reflection principle,

$$\mathbb{P}\left[\max B^{\mathrm{ex}} \geq x\right] \leq \mathbb{P}\left[\max |B| \geq x/4\right] \leq 2\mathbb{P}\left[\max B \geq x/4\right] = 4\mathbb{P}\left[N \geq x/4\right] .$$

Next, for $t \in [0,1]$, let $F^{\mathrm{me}}(t) = \inf_{s \in [t,1]} B^{\mathrm{me}}(s)$. By (49) in Pitman (1999) (originally due to Biane and Yor (1988)),

$$B^{\mathrm{me}} - F^{\mathrm{me}} \stackrel{\mathrm{d}}{=} |B^{\mathrm{br}}|.$$

Since $F^{\mathrm{me}}(t) \leq B^{\mathrm{me}}(1)$ $\forall t \in [0,1]$, and $B^{\mathrm{me}}(1)$ has the Rayleigh distribution with parameter 1, for X, Y i.i.d. standard Gaussian random variables,

$$\begin{aligned} \mathbb{P}\left[\max F^{\mathrm{me}} \geq x/2\right] &\leq \mathbb{P}\left[B^{\mathrm{me}}(1) \geq x/2\right] \\ &= \mathbb{P}\left[\sqrt{X^2 + Y^2} \geq x/2\right] \\ &\leq 2\mathbb{P}\left[N \geq x/(2\sqrt{2})\right] . \end{aligned}$$

Also, as above,

$$\mathbb{P}\left[\max |B^{\mathrm{br}}| \geq x/2\right] \leq \mathbb{P}\left[\max |B| \geq x/4\right] \leq 4\mathbb{P}\left[N \geq x/4\right] .$$

It follows that

$$\mathbb{P}\left[\max B^{\mathrm{me}} \geq x\right] \leq 2\mathbb{P}\left[N \geq x/(2\sqrt{2})\right] + 4\mathbb{P}\left[N \geq x/4\right] \leq 6\mathbb{P}\left[N \geq x/4\right] . \qquad \square$$

Proof of Lemma 2.3.5. First, for a standard Gaussian N we have $\mathbb{P}\left[N \geq c\right] \leq \frac{1}{\sqrt{2\pi}} \frac{1}{c} e^{-c^2/2}$, for all $c > 0$. Using this bound, Fact 2.A.3 and Brownian scaling, for each i we obtain

$$\mathbb{P}\left[\max_{u \leq t_i} X_i(u) \geq x\right] \leq 6\mathbb{P}\left[N \geq \frac{x}{4t_i^{1/2}}\right] \leq \frac{1}{\sqrt{2\pi}} \frac{24t_i^{1/2}}{x} e^{-x^2/32t_i} < \frac{12t_i^{1/2}}{x} e^{-x^2/32t_i}.$$

Provided that $x \geq 12T^{1/2}$, a union bound then yields

$$\mathbb{P}\left[\max_{i \geq 1} \max_{u \leq t_i} X_i(u) \geq x\right] \leq \sum_{i \geq 1} e^{-x^2/32t_i}.$$

Finally, the function $f(a) = e^{-x^2/a}\mathbf{1}_{[a>0]}$ is convex for $a \in [0, x^2/2]$, and $f(0) = 0$, so if $x^2 \geq 64T$ then for each i, $f(32t_i) \leq (t_i/T)f(32T)$. Hence

$$\mathbb{P}\left[\max_{i \geq 1} \max_{u \leq t_i} X_i(u) \geq x\right] \leq e^{-x^2/32T}. \qquad \square$$

Appendix 2.B Probability tail bounds

We first state a Bernstein-type inequality due to Colin McDiarmid.

Theorem 2.B.1 (McDiarmid (1998), Theorem 2.7)**.** *Let $X_1, \ldots, X_n$ be independent with $X_k - \mathbb{E}X_k \leq 1$ for each k. Write $S_n = \sum_{k=1}^n X_k$ and fix $V \geq Var(S_n) = \sum_{k=1}^n Var(X_i)$. Then for any $c \geq 0$,*

$$\mathbb{P}\left[S_n \geq \mathbb{E}S_n + c\right] \leq e^c \cdot \left(\frac{V}{V+c}\right)^{V+c} < \left(\frac{eV}{c}\right)^c.$$

The first inequality is the heart of the theorem; the second is easy and is included to simplify an application of the theorem. The next lemma provides upper tail probability estimates for weighted geometric sums.

Lemma 2.B.2. *Fix $\varepsilon < 1/2$ and let $(G_i, i \geq 1)$ be i.i.d.* Geom$(1-\varepsilon)$*. For any n and any non-negative real numbers $r_1, \ldots, r_n$ with $\max r_i / \sum r_i \leq 1/V$, for all $\delta > 0$,*

$$\mathbb{P}\left[\sum_{i=1}^n r_i G_i \geq (1+\delta)\sum_{i=1}^n r_i\right] \leq 2(2^{1+\delta}\varepsilon^\delta)^V.$$

Proof of Lemma 2.B.2. Let $\hat{G}_j = G_j - 1$ and $p_j = r_j / \sum_i r_i$. Then we must bound

$$\mathbb{P}\left[\sum_{i=1}^{n} p_i \hat{G}_i \geq \delta\right],$$

under the assumption that $\max_i p_i \leq 1/V$. First note that for c such that $\varepsilon e^c < 1$,

$$\mathbb{E}\left[\exp\left(cV \cdot \sum_{i=1}^{n} p_i \hat{G}_i\right)\right] = \prod_i \frac{1-\varepsilon}{1-\varepsilon e^{cVp_i}}.$$

For $c > 0$, the latter product is maximized (subject to the constraints that $\max_i p_i \leq 1/V$ and that $\sum_i p_i = 1$) when $p_i = 1/\lceil V \rceil$ for $\lceil V \rceil$ values of i and $p_i = 0$ otherwise. We thus obtain

$$\mathbb{E}\left[\exp(cV \cdot \sum_{i=1}^{n} p_i \hat{G}_i)\right] \leq \frac{(1-\varepsilon)^{\lceil V \rceil}}{(1-\varepsilon e^c)^{\lceil V \rceil}}.$$

For any non-negative random variable, $\mathbb{P}[X > \delta] \leq e^{-c\delta V} \mathbb{E} e^{cVX}$; taking $e^c = (2\varepsilon)^{-1}$ yields

$$\mathbb{P}\left[\sum_{i=1}^{n} p_i \hat{G}_i \geq \delta\right] \leq e^{-c\delta V} \cdot \frac{(1-\varepsilon)^V}{(1-\varepsilon e^c)^{V+1}} = 2\left(2(1-\varepsilon)(2\varepsilon)^{\delta}\right)^V < 2(2^{1+\delta}\varepsilon^{\delta})^V. \quad \square$$

Chapter 3

Branching Brownian motion and selection in the spatial Λ-Fleming-Viot process

Abstract

We ask the question “when will natural selection on a gene in a spatially structured population cause a detectable trace in the patterns of genetic variation observed in the contemporary population?”. We focus on the situation in which ‘neighbourhood size’, that is the effective local population density, is small. The genealogy relating individuals in a sample from the population is embedded in a spatial version of the ancestral selection graph and through applying a diffusive scaling to this object we show that whereas in dimensions at least three, selection is barely impeded by the spatial structure, in the most relevant dimension, $d = 2$, selection must be stronger (by a factor of $\log(1/\mu)$ where μ is the neutral mutation rate) if we are to have a chance of detecting it. The case $d = 1$ was handled in Etheridge et al. (2015b).

The mathematical interest is that although the system of branching and coalescing lineages that forms the ancestral selection graph converges to a branching Brownian motion, this reflects a delicate balance of a branching rate that grows to infinity and the instant annullation of almost all branches through coalescence caused by the strong local competition in the population.

3.1 Introduction

Our aims in this chapter are two-fold. On the one hand, we address a question of interest in population genetics: when will the action of natural selection on a gene in a spatially structured population cause a detectable trace in the patterns of genetic variation observed in the contemporary population? On the other hand, we investigate some of the rich structure underlying mathematical models for spatially evolving populations and, in particular, the systems of interacting random walks that, as dual processes (corresponding to ancestral lineages of the model), describe the genetic relationships between individuals sampled from those populations.

Since the seminal work of Fisher (1937), a large literature has developed that investigates the interaction of natural selection with the spatial structure of a population. Traditionally, the deterministic action of migration and selection is approximated by what we now call the Fisher-KPP equation and predictions from that equation are compared to data. However, many important questions depend on how selection and migration interact with a third force, the stochastic fluctuations known as random genetic drift, and this poses significant new mathematical challenges.

For the most part, random drift is modelled through Wright-Fisher noise resulting in a stochastic PDE as a model for the evolution of gene frequencies w:

$$\frac{\partial w}{\partial t} = m\Delta w - sw(1-w) + \sqrt{\gamma w(1-w)}\dot{\mathcal{W}}$$

(for suitable constants m, s and γ), where $\mathcal{W}$ is space-time white noise. This stochastic Fisher-KPP equation has been extensively studied, see, for example, Mueller et al. (2008) and references therein. However, from a modelling perspective it has two immediate shortcomings. First, it only makes sense in one spatial dimension. This is generally overcome by artificially subdividing the population, and thus replacing the stochastic PDE by a system of stochastic ordinary differential equations, coupled through migration. The second problem is that, in deriving the equation, one allows the 'neighbourhood size' to tend to infinity. We shall give a precise definition of neighbourhood size in Section 3.2.

Loosely, it is inversely proportional to the probability that two individuals sampled from sufficiently close to one another had a common parent in the previous generation and small neighbourhood size corresponds to strong genetic drift. It is understanding the implications of dropping this (usually implicit) assumption of unbounded neighbourhood size that motivated the work presented here.

Our starting point will be the spatial Λ-Fleming-Viot process with selection (SLFVS), which (along with its dual) was introduced and constructed in Etheridge et al. (2014). The dynamics of both the SLFVS and its dual are driven by a Poisson Point Process of 'events' (which model reproduction or extinction and recolonisation in the population) and will be described in detail in Section 3.2. The advantage of this model is that it circumvents the need to subdivide the population in higher dimensions. However, since our proof is based on an analysis of the branching and coalescing system of random walkers that describes the ancestry of a sample from the population, it would be straightforward to modify it to apply to, for example, an individual based model in which a fixed number of individuals reside at each point of a d-dimensional lattice.

In classical models of population genetics, in which there is no spatial structure, we generally think of population size as setting the timescale of evolution of frequencies of different genetic types. Evidently that makes no sense in our setting. However (even in the classical setting), as we explain in more detail in Section 3.3, if natural selection is to leave a distinguishable trace in contemporary patterns of genetic variation, then a sufficiency of neutral mutations must fall on the genealogical trees relating individuals in a sample. Thus, in fact, it is the neutral mutation rate which sets the timescale and, since mutation rates are very low, this leads us to consider scaling limits.

In Etheridge et al. (2014), scaling limits of the (forwards in time) SLFVS were considered in which the neighbourhood size tends to infinity. In that case, the classical Fisher-KPP equation and, in one spatial dimension, its stochastic analogue are recovered. The dual process of branching and coalescing lineages converges to branching Brownian motion, with coalescence of lineages (in one dimension) at a rate determined by the local time that they spend together. In this chapter we consider scaling limits in the (very different) regime in which neighbourhood size remains finite. In this context the interaction between genetic drift and spatial structure becomes much more important and, in contrast to Etheridge et al. (2014), it is the dual process which proves to be the more analytically tractable object.

We shall focus on the most biologically relevant case of two spatial dimensions. The case of one dimension was discussed in Etheridge et al. (2015b). The main interest there is mathematical: the dual process of branching and coalescing ancestral lineages, suitably scaled, converges to the Brownian net. However, the scaling required to obtain a non-trivial limit reveals a strong effect of the spatial structure. Here we shall identify the corresponding scalings in dimensions $d \geq 2$. Whereas in Etheridge et al. (2014), the scaling of the selection coefficient is independent of spatial dimension and, indeed, mirrors that for unstructured populations, for bounded neighbourhood size this is no longer the case. In $d = 1$ and $d = 2$ the scaling of the selection coefficient required to obtain a non-trivial limit reflects strong local competition.

Our main result, Theorem 3.2.7, is that under these (dimension-dependent) scalings, the scaled dual process converges to a branching Brownian motion. For $d \geq 3$ this is rather straightforward, but in two dimensions things are much more delicate. The mathematical interest of our result is that in $d = 2$, under our scaling, the rate of branching of ancestral lineages explodes to infinity but, crucially, all except finitely many branches are instantaneously annulled through coalescence. That this finely balanced picture produces a non-degenerate limit results from a combination of the failure of two dimensional Brownian motion to hit points and the strong (local) interactions of the approximating random walks, which cause coalescence.

From a biological perspective, the main interest is that, in contrast to the infinite neighbourhood size limit, here we see a strong effect of spatial dimension in our results. When neighbourhood size is very big, the probability of fixation for an advantageous genetic type, i.e. the probability that the genetic type establishes and sweeps through the entire population, is not affected by spatial structure. When neighbourhood size is small, in (one and) two spatial dimensions, selection has to be much stronger to leave a detectable trace than in a population with no spatial structure. Indeed, local establishment is no longer a guarantee of eventual fixation.

The rest of the chapter is laid out as follows. In Section 3.2 we describe the SLFVS and the dual process of branching and coalescing random walks, state our main result and provide a heuristic argument that explains our choice of scalings. In Section 3.3 we place our findings in the context of previous work on selective sweeps in spatially structured populations and in Section 3.4 we prove our result.

3.2 The model and main result

3.2.1 The model

To motivate the definition of the SLFVS, it is convenient to recall (a very special case of) the model without selection, introduced in Barton et al. (2010); Etheridge (2008). We shall call it the SLFV to emphasize that selection is not acting. We proceed informally, only carefully specifying the state space and conditions that are sufficient to guarantee existence of the process when we define the SLFVS itself in Definition 3.2.3. The interested reader can find much more general conditions under which the SLFV exists in Etheridge and Kurtz (2014).

We restrict ourselves to the case of just two genetic types, which we denote a and A, and we suppose that the population is evolving in $\mathbb{R}^d$. It is convenient to index time by the whole real line. At each time t, the random function $\{w_t(x),\, x \in \mathbb{R}^d\}$ is defined, up to a Lebesgue null set of $\mathbb{R}^d$, by

$$w_t(x) := \text{ proportion of type } a \text{ at spatial position } x \text{ at time } t. \tag{3.1}$$

The dynamics are driven by a Poisson point process Π on $\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}_+ \times (0,1]$. Each point $(t, x, r, u) \in \Pi$ specifies a reproduction event which will affect that part of the population at time t which lies within the closed ball $\mathcal{B}_r(x)$ of radius r centred on the point x. First the location z of the parent of the event is chosen uniformly at random from $\mathcal{B}_r(x)$. All offspring inherit the type α of the parent which is determined by $w_{t-}(z)$; that is, with probability $w_{t-}(z)$ all offspring will be type a, otherwise they will be A. A portion u of the population within the ball is then replaced by offspring so that

$$w_t(y) = (1-u)w_{t-}(y) + u\mathbb{1}_{\{\alpha=a\}}, \qquad \forall y \in \mathcal{B}_r(x).$$

The population outside the ball is unaffected by the event. We sometimes call u the impact of the event.

Under this model, the time reversal of the same Poisson Point Process of events governs the ancestry of a sample from the population. Each ancestral lineage that lies in the region affected by an event has a probability u of being among the offspring of the event, in which case, as we trace backwards in time, it jumps to the location of the parent, which is sampled uniformly from the region. In this way, ancestral lineages evolve according to (dependent) compound Poisson processes and lineages can coalesce when affected by the same event. All lineages affected by an event inherit the type of the parent of that event.

Remark 3.2.1. *In Etheridge and Kurtz (2014), the SLFV and its dual are constructed simultaneously on the same probability space, through a lookdown construction, as the limit of an individual based model, and so the dual process just described really can be interpreted as tracing the ancestry of individuals in a sample from the population.*

We are now in a position to define the neighbourhood size.

Definition 3.2.2. *Write σ^2 for the variance of the first coordinate of the location of a single ancestral lineage after one unit of time and $\eta(x)$ for the instantaneous rate of coalescence of two lineages that are currently at a separation $x \in \mathbb{R}^d$. Then the* neighbourhood size*, $\mathcal{N}$ is given by*

$$\mathcal{N} = \frac{2dC_d\sigma^2}{\int_{\mathbb{R}^d} \eta(x)dx},$$

where C_d is the volume of the unit ball in $\mathbb{R}^d$.

Neighbourhood size is used in biology to quantify the local number of breeding individuals in a continuous population; see Barton et al. (2013a) for a derivation of this formula. If we assume that the impact is the same for all events, then the impact is inversely proportional to the neighbourhood size, see Barton et al. (2013a).

There are very many different ways in which to introduce selection into the SLFV. Our approach here is a simple adaptation of that adopted in classical models of population genetics. The parental type in the SLFV is a uniform pick from the types in the region affected by the event. We can introduce a small advantage to individuals of type A by choosing the parent in a weighted way. Thus if, immediately before reproduction, the proportion of type a individuals in the region affected by the event is $\overline{w}$, then the offspring will be type a with probability $\overline{w}/(1+\boldsymbol{s}(1-\overline{w}))$. We say that the relative fitnesses of types a and A are 1 and $1+\boldsymbol{s}$ respectively and refer to $\boldsymbol{s}$ as the selection coefficient. We are interested only in small values of $\boldsymbol{s}$ and so we expand

$$\frac{\overline{w}}{1+\boldsymbol{s}(1-\overline{w})} = \overline{w}\{1-\boldsymbol{s}(1-\overline{w})\} + \mathcal{O}(\boldsymbol{s}^2) = (1-\boldsymbol{s})\overline{w} + \boldsymbol{s}\overline{w}^2 + \mathcal{O}(\boldsymbol{s}^2).$$

We shall regard $\boldsymbol{s}^2$ as being negligible. We can then think of each event, independently, as being a 'neutral' event with probability $(1-\boldsymbol{s})$ and a 'selective' event with probability $\boldsymbol{s}$. Reproduction during neutral events is exactly as before, but during selective events, we sample two *potential* parents; only if both are type a will the offspring be of type a.

Let us now give a more precise definition of the SLFVS. We retain the notation of (3.1). A construction of an appropriate state space for $x \mapsto w_t(x)$ can be found in Véber and Wakolbinger (2015). Using the identification

$$\int_{\mathbb{R}^d\times\{a,A\}} f(x,\kappa)M(dx,d\kappa) = \int_{\mathbb{R}^d} \{w(x)f(x,a) + (1-w(x))f(x,A)\}\, dx,$$

this state space is in one-to-one correspondence with the space $\mathcal{M}_\lambda$ of measures on $\mathbb{R}^d \times \{a, A\}$ with 'spatial marginal' Lebesgue measure, which we endow with the topology of vague convergence. By a slight abuse of notation, we also denote the state space of the process $(w_t)_{t\in\mathbb{R}}$ by $\mathcal{M}_\lambda$.

Definition 3.2.3 (SLFV with selection (SLFVS))**.** *Fix* $\mathcal{R} \in (0,\infty)$. *Let* μ *be a finite measure on* $(0,\mathcal{R}]$ *and, for each* $r \in (0,\mathcal{R}]$, *let* ν_r *be a probability measure on* $(0,1]$. *Further, let* Π *be a Poisson point process on* $\mathbb{R}\times\mathbb{R}^d\times(0,\mathcal{R}]\times(0,1]$ *with intensity measure*

$$dt \otimes dx \otimes \mu(dr)\nu_r(du). \tag{3.2}$$

The spatial Λ-Fleming-Viot process with selection *(SLFVS) driven by* (3.2) *is the* $\mathcal{M}_\lambda$*-valued process* $(w_t)_{t\in\mathbb{R}}$ *with dynamics given as follows.*

If $(t, x, r, u) \in \Pi$*, a reproduction event occurs at time* t *within the closed ball* $\mathcal{B}_r(x)$ *of radius* r *centred on* x*. With probability* $1 - \boldsymbol{s}$ *the event is* neutral*, in which case:*

1. *Choose a parental location* z *uniformly at random within* $\mathcal{B}_r(x)$*, and a parental type,* α*, according to* $w_{t-}(z)$*, that is* $\alpha = a$ *with probability* $w_{t-}(z)$ *and* $\alpha = A$ *with probability* $1 - w_{t-}(z)$*.*
2. *For every* $y \in \mathcal{B}_r(x)$*, set* $w_t(y) = (1-u)w_{t-}(y) + u\mathbf{1}_{\{\alpha=a\}}$*.*

Otherwise the event is selective*, in which case:*

1. *Choose two 'potential' parental locations* z, z' *independently and uniformly at random within* $\mathcal{B}_r(x)$*, and at each of these sites 'potential' parental types* α*,* α'*, according to* $w_{t-}(z), w_{t-}(z')$ *respectively.*
2. *For every* $y \in \mathcal{B}_r(x)$ *set* $w_t(y) = (1-u)w_{t-}(y) + u\mathbf{1}_{\{\alpha=\alpha'=a\}}$*. Declare the parental location to be* z *if* $\alpha = \alpha' = a$ *or* $\alpha = \alpha' = A$ *and to be* z *(resp.* z'*) if* $\alpha = A, \alpha' = a$ *(resp.* $\alpha = a, \alpha' = A$*).*

This is a very special case of the SLFVS introduced in Etheridge et al. (2014).

We are especially concerned with the dual process of the SLFVS. Whereas in the neutral case we can always identify the distribution of the location of the parent of each event, without any additional information on the distribution of types in the region, now, at a selective event, we are unable to identify which of the 'potential parents' is the true parent of the event without knowing their types. These can only be established by tracing further into the past. The resolution is to follow all *potential* ancestral lineages backwards in time. This results in a system of branching and coalescing walks.

As in the neutral case, the dynamics of the dual are driven by the same Poisson point process of events, Π, that drove the forwards in time process. The distribution of this Poisson point process is invariant under time reversal and so we shall abuse notation by reversing the direction of time when discussing the dual.

We suppose that at time 0 (which we think of as 'the present'), we sample k individuals from locations $x_1, \ldots, x_k$ and we write $\xi_s^1, \ldots, \xi_s^{N_s}$ for the locations of the N_s *potential ancestors* that make up our dual at time s before the present.

Definition 3.2.4 (Branching and coalescing dual)**.** *The branching and coalescing dual process* $(\Xi_t)_{t\geq 0}$ *driven by* Π *is the* $\bigcup_{m\geq 1}(\mathbb{R}^d)^m$*-valued Markov process with dynamics defined as follows: at each event* $(t, x, r, u) \in \Pi$*, with probability* $1 - \boldsymbol{s}$*, the event is neutral:*

1. *For each* $\xi^i_{t-} \in \mathcal{B}_r(x)$*, independently mark the corresponding lineage with probability* u*.*
2. *If at least one lineage is marked, all marked lineages disappear and are replaced by a single lineage, whose location at time* t *is drawn uniformly at random from within* $\mathcal{B}_r(x)$*.*

Otherwise the event is selective:

1. *For each* $\xi^i_{t-} \in \mathcal{B}_r(x)$*, independently mark the corresponding lineage with probability* u*.*
2. *If at least one lineage is marked, all marked lineages disappear and are replaced by* two *lineages, whose locations at time* t *are drawn independently and uniformly from within* $\mathcal{B}_r(x)$*.*

In both cases, if no lineages are marked, then nothing happens.

Since we only consider finitely many initial individuals in the sample, and the jump rate of the dual is bounded by a linear function of the number of potential ancestors, this description gives rise to a well-defined process.

This dual process is the analogue for the SLFVS of the ancestral selection graph (ASG), introduced in the companion papers Krone and Neuhauser (1997); Neuhauser and Krone (1997), which describes all the potential ancestors of a sample from a population evolving according to the Wright-Fisher diffusion with selection. Perhaps the simplest way of expressing the duality between the SLFVS and the branching and coalescing dual process is to observe that all the individuals in our sample are of type a if and only if all potential ancestral lineages are of type a at any time t in the past. This is analogous to the *moment duality* between the ASG and the Wright-Fisher diffusion with selection. However, to state this formally for the SLFVS, we would need to be able to identify $\mathbb{E}[\prod_{i=1}^n w_t(x_i)]$ for any choice of points $x_1, \dots, x_n \in \mathbb{R}^d$. The difficulty is that, just as in the neutral case, the SLFVS $w_t(x)$ is only defined at Lebesgue almost every point x and so we have to be satisfied with a 'weak' moment duality.

Proposition 3.2.5. *[Etheridge et al. (2014)] The spatial Λ-Fleming-Viot process with selection is dual to the process $(\Xi_t)_{t\geq 0}$ in the sense that for every $k \in \mathbb{N}$ and $\psi \in C_c((\mathbb{R}^d)^k)$, we have*

$$\mathbb{E}_{w_0}\left[\int_{(\mathbb{R}^d)^k} \psi(x_1, \dots, x_k)\left\{\prod_{j=1}^k w_t(x_j)\right\} dx_1 \dots dx_k\right]$$
$$= \int_{(\mathbb{R}^d)^k} \psi(x_1, \dots, x_k)\mathbb{E}_{\{x_1,\dots,x_k\}}\left[\prod_{j=1}^{N_t} w_0(\xi_t^j)\right] dx_1 \dots dx_k.$$

3.2.2 The main result

Our main result concerns a diffusive rescaling of the dual process of Definition 3.2.4 and so from now on it will be convenient if

forwards in time refers to forwards for the dual process.

We shall take the impact parameter, u, to be a fixed number in $(0, 1]$ (i.e. $\nu_r = \delta_u$ for all r). In fact, the same arguments work when u is allowed to be random, as long as $\int_{\mathcal{R}'}^{\mathcal{R}} \int_0^1 u\nu_r(du)\mu(dr) > 0$ for some $0 < \mathcal{R}' < \mathcal{R}$, but this would make our proofs notationally cumbersome.

Let us describe the scaling more precisely. Suppose that μ is a finite measure on $(0, \mathcal{R}]$. We shall assume for convenience that $\mathcal{R}$ is defined in such a way that for any $\delta > 0$, $\mu((\mathcal{R} - \delta, \mathcal{R}]) > 0$. For each $n \in \mathbb{N}$, define the measure μ^n by $\mu^n(B) = \mu(n^{1/2}B)$, for all Borel subsets B of $\mathbb{R}_+$. It will be convenient to write $\mathcal{R}_n = \mathcal{R}/\sqrt{n}$. At the nth stage of the rescaling, our rescaled dual is driven by the Poisson point process Π^n on $\mathbb{R} \times \mathbb{R}^d \times (0, \mathcal{R}_n]$ with intensity

$$n\, dt \otimes n^{d/2}\, dx \otimes \mu^n(dr). \tag{3.3}$$

This corresponds to rescaling space and time from (t, x) to $(n^{-1}t, n^{-1/2}x)$. Importantly, we do not scale the impact u. Each event of Π^n, independently, is neutral with probability

$1 - \boldsymbol{s}_n$ and selective with probability $\boldsymbol{s}_n$, where

$$\boldsymbol{s}_n = \begin{cases} \frac{\log n}{n} & d = 2, \\ \\ \frac{1}{n} & d \geq 3. \end{cases} \tag{3.4}$$

In Etheridge et al. (2015b) it was shown that in $d = 1$, one should take $\boldsymbol{s}_n = 1/\sqrt{n}$.

Although not obvious for the SLFVS itself, when considering the dual process it is not hard to understand why the scalings (3.3) and (3.4) should lead to a non-trivial limit.

If we ignore the selective events, then a single ancestral lineage evolves as a pure jump process which is homogeneous in both space and time. Write V_r for the volume of $\mathcal{B}_r(0)$. The rate at which the lineage jumps from y to $y + z$ can be written

$$m_n(dz) = nu \int_0^{\mathcal{R}_n} n^{d/2} \frac{V_r(0, z)}{V_r} \mu^n(dr)\, dz, \tag{3.5}$$

where $V_r(0, z)$ is the volume of $\mathcal{B}_r(0) \cap \mathcal{B}_r(z)$. To see this, by spatial homogeneity, we may take the lineage to be at the origin in $\mathbb{R}^d$ before the jump, and then, in order for it to jump to z, it must be affected by an event that covers both 0 and z. If the event has radius r, then the volume of possible centres, x, of such events is $V_r(0, z)$ and so the intensity with which such a centre is selected is $n\, n^{d/2} V_r(0, z) \mu^n(dr)$. The parental location is chosen uniformly from the ball $\mathcal{B}_r(x)$, so the probability that z is chosen as the parental location is dz/V_r and the probability that our lineage is actually affected by the event is u. Combining these yields (3.5).

The total rate of jumps is

$$\begin{aligned} \int_{\mathbb{R}^d} m_n(dz) &= \int_0^{\mathcal{R}_n} nu\, n^{d/2} \frac{1}{V_r} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_{|x|<r} \mathbb{1}_{|x-z|<r} dx\, dz\, \mu^n(dr) \\ &= \int_0^{\mathcal{R}_n} nu\, n^{d/2} V_r \mu^n(dr) \\ &= nuV_1 \int_0^{\mathcal{R}} r^d \mu(dr) = \Theta(n), \end{aligned} \tag{3.6}$$

and the size of each jump is $\Theta(n^{-1/2})$ and so in the limit a single lineage will evolve according to a (time-changed) Brownian motion.

Now, consider what happens at a selective event. The two new lineages are created at a separation of order $1/\sqrt{n}$. If we are to see both lineages in the limit then they must move apart to a separation of order 1 (before, possibly, coalescing back together). Ignoring possible interactions with other lineages, the probability that a pair of lineages makes such an excursion is of order 1 in $d \geq 3$, order $1/\log n$ in $d = 2$ and order $1/\sqrt{n}$ in $d = 1$. Therefore, in order to have a positive probability of seeing branching in the scaling limit, in $d \geq 3$ we only need that there are a positive number of selective events in unit (rescaled) time, and, for this, it is enough that $\boldsymbol{s}_n$ is order $1/n$. However, for $d = 2$, we need order $\log n$ branches before we expect to find one that is visible to us, hence the choice $\boldsymbol{s}_n = \log n/n$.

Remark 3.2.6. *Our scaling mirrors that described in Durrett and Zähle (2007) for a model of a* hybrid zone *(by which we mean a region in which we see both genetic types) which develops around a boundary between two regions, in one of which type a individuals are selectively favoured and in the other of which type A individuals are selectively favoured. In contrast to our continuum setting, their model is a spin system in which exactly one individual lives at each point of* $\mathbb{Z}^d$.

Before formally stating our main result, we need some notation. We shall denote by BBM(p, V) binary branching Brownian motion started from the point $p \in \mathbb{R}^d$, with branching rate V and diffusion constant given by

$$\sigma^2 = \tfrac{1}{d} \int_{\mathbb{R}^d} |z|^2 m^n(dz) = \tfrac{1}{d} \int_{\mathbb{R}^d} \int_0^\infty |z|^2 u \frac{V_r(0,z)}{V_r} \mu(dr)\, dz \tag{3.7}$$

where $m^n(dz)$ is defined in (3.5). In other words, during their lifetime, which is exponentially distributed with parameter V, individuals follow d-dimensional Brownian motion with diffusion constant σ^2, at the end of which they die, leaving behind at the location where they died exactly two offspring. We view BBM(p, V) as a set of (continuous) paths, each starting at p, with precisely one path following each possible distinct sequence of branches.

Similarly, we write $\mathcal{P}^{(n)}(p)$ for the dual process of Definition 3.2.4, rescaled as in (3.3) and (3.4), started from a single individual at the point $p \in \mathbb{R}^d$ and viewed as a collection of paths. Each path traces out a ‘potential ancestral lineage’, defined exactly as the ancestral lineages in the neutral case except that at each selective event, if a lineage is affected then it jumps to the location of (either) one of the 'potential parents'. Precisely one potential ancestral lineage follows each possible route through the branching and coalescing dual process.

We define the events

$$\begin{aligned} \mathcal{D}_n(\varepsilon, T) &= \left\{ \forall l \in \mathcal{P}^{(n)}(p),\ \exists l' \in \mathrm{BBM}(p, V) : \sup_{t \in [0,T]} |l(t) - l'(t)| \leq \varepsilon \right\}, \\ \mathcal{D}'_n(\varepsilon, T) &= \left\{ \forall l \in \mathrm{BBM}(p, V),\ \exists l' \in \mathcal{P}^{(n)}(p) : \sup_{t \in [0,T]} |l(t) - l'(t)| \leq \varepsilon \right\}. \end{aligned} \tag{3.8}$$

Theorem 3.2.7. *Let* $d \geq 2$*. There exists* $V \in (0, \infty)$ *such that the following holds. Let* $T < \infty$*,* $p \in \mathbb{R}^2$*; then given* $\varepsilon > 0$*, there exists* $N \in \mathbb{N}$ *such that, for all* $n \geq N$ *there is a coupling between BBM*(p, V) *and* $\mathcal{P}^{(n)}(p)$ *with* $\mathbb{P}\left[\mathcal{D}_n(\varepsilon, T) \cap \mathcal{D}'_n(\varepsilon, T)\right] \geq 1 - \varepsilon$*.*

We will give a proof of Theorem 3.2.7 only for $d = 2$. The case $d \geq 3$ follows from a simplified version of the 2-dimensional proof presented here.

3.2.3 Sketch of proof

Consider a pair of potential ancestral lineages, $\xi^{n,1}$ and $\xi^{n,2}$, created in some selective event which, without loss of generality, we suppose happens at time zero. Suppose that we forget about further branches and when $\xi^{n,i}$ is affected by a neutral event it jumps to the location of the parent; when it is affected by a selective event it jumps to the location of one of the potential parents (picked at random). Thus $\xi^{n,1}$ and $\xi^{n,2}$ are compound Poisson processes which interact when (and only when) $|\xi^{n,1} - \xi^{n,2}| \leq 2\mathcal{R}_n$.

We choose a large constant $c > 0$. We begin by showing that $\xi^{n,1}$ and $\xi^{n,2}$ have probability $\Theta(1/\log n)$ of reaching a distance $1/(\log n)^c$ from each other without coalescing (we then say they have diverged). We also show that the probability that $\xi^{n,1}$ and $\xi^{n,2}$ have not diverged or coalesced by time $1/(\log n)^c$ is $o(1/(\log n))$, so coalescence will be instantaneous in the limit. Moreover, once they are $1/(\log n)^c$ apart, they won't get within distance $2\mathcal{R}_n$ of each other again on a timescale of $\mathcal{O}(1)$. Hence from the point of view of our scaling they stay apart and evolve essentially independently of each other.

We exploit this observation by coupling the whole rescaled dual process with a process in which diverged lineages move independently. We use an object that we call a caterpillar which is defined in the same way as the rescaled dual process, except that selective events only result in branching if at least time $1/(\log n)^c$ has elapsed since the previous branching. We stop the caterpillar at the first time a pair of lineages has either diverged or failed to coalesce in time $1/(\log n)^c$ after branching. We then start two new independent caterpillars at the positions of the pair of lineages, and continue in the same way, giving a 'branching caterpillar'.

The branching caterpillar can be coupled with the rescaled dual process by piecing together the independent Poisson point processes of events which drive each caterpillar into a single driving Poisson point process. We show that under the coupling, the branching caterpillar and the rescaled dual process coincide with high probability, using the result that lineages at a separation of at least $1/(\log n)^c$ are unlikely to interact again. Each individual caterpillar converges in an appropriate sense to a segment of a Brownian path run for an exponentially distributed lifetime, so we can couple the branching caterpillar with the limiting branching Brownian motion.

This programme is carried out in Section 3.4.

3.3 Biological background

In this section, we shall set our work in the context of the substantial biological literature. The reader concerned only with the mathematics can safely skip to Section 3.4.

The interplay between natural selection and the spatial structure of a population is a question of longstanding interest in population genetics. Fisher (1937) studied the advance of selectively advantageous genetic types through a one-dimensional population using the deterministic differential equation now known as the Fisher-KPP equation. This equation also makes sense in higher dimensions, but ignores *genetic drift* (the randomness due to reproduction in a finite population). Work incorporating genetic drift has been restricted to either one spatial dimension (see Barton et al. (2013b) and references therein) or, more commonly, to subdivided populations. Maruyama (1970) studied the probability of *fixation* of an advantageous genetic type (the probability that eventually the whole population carries this genetic type) in a subdivided population. The assumptions made in that article are rather strong: if we think of the population as living on islands (or in colonies), then each island has constant total population size and its contribution to the next generation is in proportion to that size. Under these assumptions, the probability of fixation is not affected by the population structure: it is the same as for a gene of the same selective advantage in an unstructured population of the same total size. Much subsequent work retained Maruyama's assumptions, and so it is often assumed that spatial structure has no influence on the accumulation of favourable genes. However, Barton (1993) showed that the extra stochasticity produced by the introduction of local extinctions and colonisations could significantly change the fixation probability. This work was extended in, for example, Cherry (2003) and Whitlock (2003).

A fundamental problem in genetics is to identify which parts of the genome have been the target of natural selection. The random nature of reproduction in finite populations means that some genetic types (alleles) will be carried by everyone in the population, even though they convey no particular selective advantage. However, if a favourable mutation arises in a population and 'sweeps' to fixation (i.e. increases in frequency until everybody carries it), we expect the genealogical trees (that is the trees of ancestral lineages) relating individuals in a sample from the population to differ from those that we observe in the absence of selection. In particular, they will be more 'star-shaped'. Of course we cannot observe the genealogical trees directly, and so, instead, geneticists exploit the fact that genes are arranged on chromosomes: the ancestry at another position on the same chromosome will be correlated with that at the part of the genome that is the target of selection. In order to detect selection one therefore examines the patterns of variation at other points on the same chromosome, so-called linked loci.

In order for this approach to work, we require sufficient variability at the linked loci that we see a signal of the distortion in the genealogical tree. This means that we must consider the genealogy of a sample from the population on the timescale set by the neutral mutation rate. If selection is too strong, the genealogy will be very short and we see no mutations and so we can recover no information about the genealogical trees;

if selection is too weak, we won't be able to distinguish the patterns from those seen under neutral evolution.

Since neutral mutation rates are rather small, this means that we are interested in long timescales. Without selection, ancestral lineages in our model follow symmetric random walks with bounded variance jumps and so we expect a diffusive scaling to capture patterns of neutral variation. Since we are looking for deviations from those patterns due to the action of selection, it makes sense to consider a diffusive rescaling in the selective case too. Thus, if the neutral mutation rate is μ, then we look at the rescaled dual process with $n = 1/\mu$. If the branches produced by selection persist long enough to be visible at this scale, then there is positive probability that the pattern of (neutral) variation we see in a sample from the population will look different from the pattern we'd expect without selection.

Our results in this chapter are relevant to populations evolving in spatial continua. The question they address is 'When can we hope to detect a signal of natural selection in data?'. Whereas in the classical models of subdivided populations it is typically assumed that the population in each 'island' is large, so that neighbourhood size is big, by fixing the 'impact' parameter u in our model, we are assuming that neighbourhood size is small. As a result, reproduction events are somewhat akin to local extinction and recolonisation events, in which a significant proportion of the local population is replaced in a single event. Our main result shows that our ability to detect selection is then critically dependent on spatial dimension. For populations living in at least three spatial dimensions (of which there are very few), spatial structure has a rather weak effect. However, in two spatial dimensions, selection must be stronger and in one spatial dimension (as appropriate for example for populations living in intertidal zones) much stronger, before we can expect to be able to detect it. The explanation is that in low dimensions, it is harder for individuals carrying the favoured gene to escape the competition posed by close relatives who carry the same gene. In our mathematical work, this is reflected in the vast majority of branches in our dual process being cancelled by a coalescence event on a timescale which is negligible compared to the timescale set by the neutral mutation rate so that no evidence of these branches having occurred will be seen in the pattern of neutral mutation.

3.4 Proof of Theorem 3.2.7

Our proof is broken into two steps. First in Section 3.4.1 we consider how the pair of potential ancestral lineages created during a selective event interact with each other. In particular we find asymptotics for the probability that they diverge in a short time. This will allow us to identify the branching rate in the limiting Brownian motion. Then in Section 3.4.2 we define the caterpillar and show how to couple the dual of the SLFVS to a system of branching caterpillars. With this construction in hand, Theorem 3.2.7 follows easily.

3.4.1 Pairs of paths

In this subsection we are interested in the behaviour of a pair of potential ancestral lineages in the rescaled dual. In order that they be uniquely defined, if either is hit by a selective event then we (arbitrarily) declare that it jumps to the location of the first potential parent sampled in that event. In particular, if they are both affected by the same event, then they will necessarily coalesce. We write $\xi^{n,1}$ and $\xi^{n,2}$ for the resulting potential ancestral lineages and

$$\eta^n = \xi^{n,1} - \xi^{n,2}$$

for their separation.

Throughout this subsection, we use the notation $\mathbb{P}_{[r,r']}$ to mean that $|\eta^n_0| \in [r, r']$ and we adopt the convention that estimates of $\mathbb{P}_{[r,r']}[B]$ hold uniformly for all initial laws with mass concentrated on $[r, r']$. We extend this notation to open intervals in the obvious manner. We will also write $\mathbb{P}_r = \mathbb{P}_{[r,r]}$.

We are concerned with the behaviour of two potential ancestral lineages created during a selective event which, without loss of generality, we suppose to happen at time 0. We shall then refer to η^n as an excursion. In this case $|\eta^n_0| \leq 2\mathcal{R}_n$ and we wish to establish whether or not $|\eta^n_t|$ ever exceeds

$$\gamma_n = \frac{1}{(\log n)^c}, \tag{3.9}$$

where, in this section, we suppose that $c \geq 3$.

Remark 3.4.1. *We will, eventually, set $c = 4$, although any larger constant c would give the same result; for now we keep the dependence on c visible in our estimates.*

For reasons that will soon become apparent, it is convenient to assume that n is large enough that $7\mathcal{R}_n < \gamma_n$.

The picture of an excursion η^n that we would like to build up is, loosely speaking, as follows.

1. With probability $\kappa_n = \Theta(\frac{1}{\log n})$, $|\eta^n|$ reaches displacement γ_n within time $1/(\log n)^c$ and then $\xi^{n,1}$ and $\xi^{n,2}$ will not interact again before a fixed time $T > 0$. Consequently the displacement between them becomes macroscopic and we see two distinct paths in the limit. Moreover, $\kappa_n \log n \to \kappa \in (0, \infty)$ as $n \to \infty$.

2. With probability $1 - \Theta(\frac{1}{\log n})$, $|\eta^n|$ does not reach displacement γ_n, and $\xi^{n,1}$ and $\xi^{n,2}$ coalesce within time $1/(\log n)^c$. In this case the difference between them is microscopic and we see only one path in the limit.

3. All other outcomes have probability $\mathcal{O}(\frac{1}{(\log n)^{c-3/2}})$, which means that we won't see them in the limit.

Much of the work in making this rigorous results from the fact that $\xi^{n,1}$, $\xi^{n,2}$ only evolve independently when their separation is greater than $2\mathcal{R}_n$. Our strategy is similar to that in the proof of Lemma 4.2 in Etheridge and Véber (2012), but here we require a stronger result: rather than an estimate of the form $\kappa_n \geq C/\log n$ we need convergence of $\kappa_n \log n$.

3.4.1.1 Inner and outer excursions

We shall characterise the behaviour of η^n using several stopping times. Set $\tau_0^{out} = 0$ and define inductively, for $i \geq 0$,

$$\tau_i^{in} = \inf\{s > \tau_i^{out} \,:\, |\eta_s^n| \geq 5\mathcal{R}_n\}, \tag{3.10}$$

$$\tau_{i+1}^{out} = \inf\{s > \tau_i^{in} \,:\, |\eta_s^n| \leq 4\mathcal{R}_n\}.$$

We refer to the interval $[\tau_i^{out}, \tau_i^{in})$ (and also to the path of η^n during it) as the i^{th} inner excursion and similarly to $[\tau_{i-1}^{in}, \tau_i^{out})$ (and corresponding path) as the i^{th} outer excursion.

Since a jump of η^n has displacement at most $2\mathcal{R}_n$, although the initial (0^{th}) inner excursion starts in $(0, 2\mathcal{R}_n]$, for $i \geq 1$ we have $|\eta^n_{\tau_i^{in}}| \in [5\mathcal{R}_n, 7\mathcal{R}_n]$ and $|\eta^n_{\tau_i^{out}}| \in [2\mathcal{R}_n, 4\mathcal{R}_n]$.

Definition 3.4.2. *We define the stopping times*

$$\tau^{coal} = \inf\{s > 0 \,:\, |\eta^n_s| = 0\},$$

$$\tau^{div} = \inf\{s > 0 \,:\, |\eta^n_s| \geq \gamma_n\},$$

$$\tau^{over} = \frac{1}{(\log n)^c}.$$

We shall say that the i^{th} inner excursion coalesces if $\tau^{coal} \in [\tau_i^{out}, \tau_i^{in})$. Similarly, the i^{th} outer excursion diverges if $\tau^{div} \in [\tau_{i-1}^{in}, \tau_i^{out})$.

We define $\tau^{type} = \min(\tau^{coal}, \tau^{div}, \tau^{over})$ and say that η^n

1. *coalesces if $\tau^{type} = \tau^{coal}$,*
2. *diverges if $\tau^{type} = \tau^{div}$,*
3. *overshoots if $\tau^{type} = \tau^{over}$.*

Since almost surely η^n only jumps a finite number of times before time $(\log n)^{-c}$, almost surely τ^{type} occurs during either an inner or an outer excursion, whose index we denote by i^*.

We use ζ^n to denote the distance between the two potential parents sampled during a typical selective event. More precisely, sample a radius $r \in (0, \mathcal{R}_n]$ according to $r^2\mu^n(dr)/\int_0^{\mathcal{R}_n} s^2\mu^n(ds)$, and then sample $\xi_0^{n,1}$ and $\xi_0^{n,2}$ independently and uniformly at random from $\mathcal{B}_r(0)$; let $\zeta^n = |\xi_0^{n,1} - \xi_0^{n,2}|$.

Lemma 3.4.3. *There exists $\alpha \in (0,1)$ such that, uniformly in n, $\mathbb{P}_{\zeta^n}[i^* > m] \leq \alpha^m$.*

Lemma 3.4.4. *As $n \to \infty$, $\mathbb{P}_{\zeta^n}[\eta^n$ overshoots$] = \mathcal{O}\left(\frac{1}{(\log n)^{c-3/2}}\right)$.*

Lemma 3.4.5. *As $n \to \infty$, $\mathbb{P}_{\zeta^n}[\eta^n$ diverges$] = \Theta\left(\frac{1}{\log n}\right)$.*

Lemma 3.4.6. *As* $n \to \infty$*,* $\mathbb{P}_{\zeta^n}\left[\eta^n \text{ coalesces}\right] = 1 - \Theta\left(\frac{1}{\log n}\right)$*.*

Thus, overshoots are relatively unlikely, and typically η^n consists of a finite number of inner/outer excursions until either (1) it coalesces, with probability $1 - \Theta(\frac{1}{\log n})$, or (2) the two lineages separate to distance γ_n, with probability $\Theta(\frac{1}{\log n})$.

The remainder of this Section 3.4.1.1 is devoted to the proof of Lemmas 3.4.3-3.4.5. Lemma 3.4.6 then follows immediately, since $c \geq 3$.

We will need two more stopping times:

$$\begin{aligned} \tau_r &= \inf\{s > 0 \,:\, |\eta^n_s| \leq r\}, \\ \tau^r &= \inf\{s > 0 \,:\, |\eta^n_s| \geq r\}. \end{aligned} \tag{3.11}$$

Note that $\tau_0 = \tau^{coal}$.

Note that the random variables τ^{type}, τ^r and so on depend implicitly on n; throughout this section these random variables refer to the stopping times for the process η^n.

Proof. (Of Lemma 3.4.3.) First consider a single inner excursion of η^n. It is easily seen that there exists some $\alpha' > 0$ such that, for all n:

(†) For any $x \in (0, 5\mathcal{R}_n)$, if $|\eta^n_t| = x$ then the probability that η^n will hit 0 but not exit $\mathcal{B}_{5\mathcal{R}_n}(0)$ within its next three jumps is at least α'.

In particular, the probability that the first three jumps of an inner excursion result in a coalescence is bounded away from 0 uniformly for any $|\eta^n_{\tau_i^{out}}| \in [2\mathcal{R}_n, 4\mathcal{R}_n]$. If $i^* > m$ then at least m inner excursions must occur without a coalescence. The strong Markov property applied at the time τ_i^{out} means that, conditionally given $\eta^n_{\tau_i^{out}}$, the i^{th} inner excursion is independent of $(\eta^n_t)_{t<\tau_i^{out}}$. Repeated application of this fact, coupled with (†), shows that the probability of seeing at least m inner excursions without a single coalescence is at most $(1-\alpha')^m$. This completes the proof. □

We will shortly require a tail estimate on the supremum of the modulus of two dimensional Brownian motion W, which we record first for clarity. We write $W_t =$

(W_t^1, W_t^2) and note

$$\begin{aligned}\mathbb{P}\left[\sup_{s\in[0,t]}|W_s - W_0| \geq x\right] &\leq 2\mathbb{P}\left[\sup_{s\in[0,t]}|W_s^1 - W_0^1| \geq x/2\right]\\ &\leq 4\mathbb{P}\left[\sup_{s\in[0,t]}(W_s^1 - W_0^1) \geq x/2\right]\\ &\leq 4e^{-x^2/8t}. \qquad (3.12)\end{aligned}$$

In the first line of the above we use the triangle inequality and the fact that W^1 and W^2 have the same distribution. To deduce the second line, we note that W^1 and $-W^1$ have the same distribution. For the final line, we use the (standard) tail estimate $\mathbb{P}[\sup_{s\in[0,t]}(B_s - B_0) \geq x] \leq e^{-x^2/2t}$ for a one dimensional Brownian motion B, which can be deduced via Doob's martingale inequality applied to the submartingale $(\exp(xB_s/t))_{s\geq 0}$.

During an outer excursion, η^n is the difference between two independent walkers and so we can use Skorohod embedding to approximate its behaviour using elementary calculations for two-dimensional Brownian motion. The next lemma exploits this to bound the duration of the outer excursion and the probability that it diverges.

Lemma 3.4.7. *As $n \to \infty$,*

$$\mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}\left[\tau^{\gamma_n} \wedge \tau_{4\mathcal{R}_n} > (\log n)^{-c-1}\right] = \mathcal{O}\left(\frac{1}{(\log n)^{c-1}}\right), \qquad (3.13)$$

and

$$\mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}\left[\tau^{\gamma_n} < \tau_{4\mathcal{R}_n}\right] = \Theta\left(\frac{1}{\log n}\right). \qquad (3.14)$$

Proof. For $i = 1, 2$ let $\hat{\xi}^{n,i}$ be a pair of independent processes such that $\hat{\xi}^{n,1}$ has the same distribution as $\xi^{n,1}$ and $\hat{\xi}^{n,2}$ has the same distribution as $\xi^{n,2}$. The process $\hat{\xi}^{n,1} - \hat{\xi}^{n,2}$ is a compound Poisson process with a rotationally symmetric jump distribution and a maximum displacement of $2\mathcal{R}_n$ on each jump. Moreover (essentially by Skorohod's Embedding Theorem, see e.g. Billingsley (1995)), we can construct a process $\hat{\eta}^n$ with the same distribution as $\hat{\xi}^{n,1} - \hat{\xi}^{n,2}$ as follows.

Let $(r_m, J_m)_{m\geq 1}$ denote a sequence distributed as the jump magnitudes and jump times of $\hat{\xi}^{n,1} - \hat{\xi}^{n,2}$. Let W be a two-dimensional Brownian motion with $W_0 = \hat{\xi}_0^{n,1} - \hat{\xi}_0^{n,2}$, independent of $(r_m, J_m)_{m\geq 1}$. Now set

$$\begin{aligned} \hat{\eta}^n_t &= W_{T^{(S(t))}} \text{ where } T^{(0)} = 0, J_0 = 0, \\ T^{(m+1)} &= \inf\{s > T^{(m)} \,:\, |W_s - W_{T^{(m)}}| \geq r_m\}, \\ S(t) &= \sup\{i \geq 0 \,:\, J_i \leq t\}. \end{aligned} \tag{3.15}$$

We may then couple

$$\hat{\eta}^n = \hat{\xi}^{n,1} - \hat{\xi}^{n,2}.$$

We define $\hat{\tau}^r$ and $\hat{\tau}_r$ analogously to τ^r and τ_r, as stopping times of the process $\hat{\eta}^n$.

Note that since $(\xi^{n,1}_t, \xi^{n,2}_t)_{t\leq \tau_{4\mathcal{R}_n}}$ has the same distribution as $(\hat{\xi}^{n,1}, \hat{\xi}^{n,2})_{t\leq \tau_{4\mathcal{R}_n}}$, we may couple them so that they are almost surely equal during this time. Thus

$$\{\hat{\tau}^{\gamma_n} < \hat{\tau}_{4\mathcal{R}_n}\} = \{\tau^{\gamma_n} < \tau_{4\mathcal{R}_n}\}.$$

Let T^r and T_r be the analogues of τ^r and τ_r for W (not to be confused with $T^{(m)}$ in (3.15)). By the definition of the Skorohod embedding in (3.15) we have

$$\begin{aligned} \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}\left[\hat{\tau}^{\gamma_n} < \hat{\tau}_{4\mathcal{R}_n}\right] &\geq \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}\left[T^{\gamma_n + 2\mathcal{R}_n} < T_{4\mathcal{R}_n}\right] \\ &\geq \mathbb{P}_{5\mathcal{R}_n}\left[T^{\gamma_n + 2\mathcal{R}_n} < T_{4\mathcal{R}_n}\right]. \end{aligned} \tag{3.16}$$

The right hand side concerns only the modulus of two-dimensional Brownian motion and so can be expressed in terms of the scale function for a two-dimensional Bessel process:

$$\mathbb{P}_{5\mathcal{R}_n}\left[T^{\gamma_n + 2\mathcal{R}_n} < T_{4\mathcal{R}_n}\right] = \frac{\log(5\mathcal{R}_n) - \log(4\mathcal{R}_n)}{\log(\gamma_n + 2\mathcal{R}_n) - \log(4\mathcal{R}_n)} = \Theta\left(\frac{1}{\log n}\right), \tag{3.17}$$

which proves the lower bound in (3.14). Similarly, to see the upper bound we note that

$$\begin{aligned} \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}\left[\hat{\tau}^{\gamma_n} < \hat{\tau}_{4\mathcal{R}_n}\right] &\leq \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[T^{\gamma_n} < T_{2\mathcal{R}_n}] \\ &\leq \mathbb{P}_{7\mathcal{R}_n}\left[T^{\gamma_n} < T_{2\mathcal{R}_n}\right] \\ &= \frac{\log(7\mathcal{R}_n) - \log(2\mathcal{R}_n)}{\log(\gamma_n) - \log(2\mathcal{R}_n)} \end{aligned}$$

$$= \Theta\left(\frac{1}{\log n}\right).$$

It remains to prove (3.13). We have

$$\tau^{\gamma_n} \wedge \tau_{4\mathcal{R}_n} = \hat{\tau}^{\gamma_n} \wedge \hat{\tau}_{4\mathcal{R}_n} \leq \hat{\tau}^{\gamma_n}.$$

Remark 3.4.8. *The above inequality is a very crude estimate, but will be enough to prove* (3.13)*, which in turn will be enough to give useful bounds on the duration of excursions due to the freedom in the choice of c.*

Hence

$$\mathbb{P}_{[5\mathcal{R}_n,7\mathcal{R}_n]}\left[\tau^{\gamma_n} \wedge \tau_{4\mathcal{R}_n} > (\log n)^{-c-1}\right] \leq \mathbb{P}_{[5\mathcal{R}_n,7\mathcal{R}_n]}\left[|\hat{\eta}^n_{(\log n)^{-c-1}}| \leq \gamma_n\right]. \tag{3.18}$$

The remainder of the proof focuses on bounding the right side of (3.18). To do so, we must relate our compound Poisson process to another Brownian motion.

For $j \geq 1$, let $X_j = \hat{\eta}^n_{j/n} - \hat{\eta}^n_{(j-1)/n}$. Then $(X_j)_{j\geq 1}$ are i.i.d. and since $\hat{\xi}^{n,1}$ and $\hat{\xi}^{n,2}$ are independent, $\mathbb{E}\left[|X_1|^2\right] = 2\mathbb{E}\left[|\hat{\xi}^{n,1}_{1/n} - \hat{\xi}^{n,1}_0|^2\right]$.

Recall from (3.5) that the rate at which $\hat{\xi}^{n,1}$ jumps from y to $y+z$ is determined by the intensity measure $m^n(dz)$ so that

$$\mathbb{E}\left[|X_1|^2\right] = \frac{2}{n}\int_{\mathbb{R}^2} |z|^2 m^n(dz) = \frac{4\sigma^2}{n}, \tag{3.19}$$

where σ^2 was defined in (3.7). Now recall the definition of $S(t)$ in (3.15); the rate at which $\hat{\xi}^{n,1}$ jumps is $\int_{\mathbb{R}^2} m^n(z)dz = \Theta(n)$ by (3.6), so $S(n^{-1})$ is bounded by the sum of two Poisson($\Theta(1)$) random variables. Hence since each jump of $\hat{\eta}^n$ is bounded by $2\mathcal{R}_n$,

$$\mathbb{E}\left[|X_1|^4\right] \leq (2\mathcal{R}_n)^4\mathbb{E}\left[S(n^{-1})^4\right] = \mathcal{O}(n^{-2}). \tag{3.20}$$

Once again (since the distribution of X_1 is rotationally symmetric) we may use Skorohod's Embedding Theorem to couple $(X_i)_{i\geq 1}$ to a two-dimensional Brownian motion B started at η^n_0 and a sequence $\upsilon_1, \upsilon_2, \ldots$ of stopping times for B such that setting $\upsilon_0 = 0$,

$(\upsilon_i - \upsilon_{i-1})_{i\geq 1}$ are i.i.d. and

$$\begin{aligned} B_{\upsilon_i} - B_{\upsilon_{i-1}} &= X_i, \\ \mathbb{E}[\upsilon_i - \upsilon_{i-1}] &= \tfrac{1}{2}\mathbb{E}[|X_1|^2] = \tfrac{2\sigma^2}{n} \\ \text{and } \mathbb{E}[(\upsilon_i - \upsilon_{i-1})^2] &= \mathcal{O}(n^{-2}). \end{aligned} \tag{3.21}$$

It follows that $\mathbb{E}[\upsilon_{\lfloor tn \rfloor}] = \frac{2\sigma^2 \lfloor tn \rfloor}{n}$ and $\mathrm{Var}(\upsilon_{\lfloor tn \rfloor}) = \mathcal{O}(tn^{-1})$. Hence by Chebychev's inequality,

$$\mathbb{P}[|\upsilon_{\lfloor tn \rfloor} - 2\sigma^2 t| \geq n^{-1/3}] \leq \mathcal{O}(tn^{-1/3}).$$

Applying this result with $t = t_n := (\log n)^{-c-1}$, since $\hat{\eta}^n_{\lfloor t_n n \rfloor/n} = B_{\upsilon_{\lfloor t_n n \rfloor}}$ we have

$$\begin{aligned} &\mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}\left[|\hat{\eta}^n_{t_n}| \leq \gamma_n\right] \\ &\quad \leq \mathbb{P}\left[\inf\left\{|B_t - B_0| \,:\, t \in [2\sigma^2 t_n - n^{-1/3}, 2\sigma^2 t_n + n^{-1/3}]\right\} \leq \gamma_n + n^{-1/8} + 7\mathcal{R}_n\right] \\ &\qquad + \mathbb{P}\left[|\hat{\eta}^n_{t_n} - \hat{\eta}^n_{\lfloor t_n n \rfloor/n}| \geq n^{-1/8}\right] + \mathcal{O}(t_n n^{-1/3}). \end{aligned} \tag{3.22}$$

For the first term on the right hand side we have for n sufficiently large

$$\begin{aligned} &\mathbb{P}\left[\inf\left\{|B_t - B_0| \,:\, t \in [2\sigma^2 t_n - n^{-1/3}, 2\sigma^2 t_n + n^{-1/3}]\right\} \leq \gamma_n + n^{-1/8} + 7\mathcal{R}_n\right] \\ &\qquad \leq \mathbb{P}\left[|B_{2\sigma^2 t_n} - B_0| \leq \gamma_n + 3n^{-1/8}\right] + \mathbb{P}\left[\sup_{t\in[0,2n^{-1/3}]} |B_t - B_0| \geq \tfrac{1}{2} n^{-1/8}\right] \\ &\qquad = \mathcal{O}(\gamma_n^2 t_n^{-1}) + \mathcal{O}(e^{-\frac{1}{64} n^{1/12}}) \\ &\qquad = \mathcal{O}((\log n)^{1-c}), \end{aligned} \tag{3.23}$$

For the second inequality, we use that the density of B_t is bounded by $(2\pi t)^{-1}$ for the first term and we apply (3.12) for the second term.

Moving on to the second term on the right hand side of (3.22), since from (3.19) we have $\mathbb{E}\left[|\hat{\eta}^n_{t_n} - \hat{\eta}^n_{\lfloor t_n n \rfloor/n}|^2\right] = \mathcal{O}(n^{-1})$, by Markov's inequality

$$\mathbb{P}\left[|\hat{\eta}^n_{t_n} - \hat{\eta}^n_{\lfloor t_n n \rfloor/n}| \geq n^{-1/8}\right] = \mathcal{O}(n^{-3/4}). \tag{3.24}$$

Putting (3.23) and (3.24) into (3.22) we have

$$\mathbb{P}_{[5\mathcal{R}_n,7\mathcal{R}_n]}\left[|\hat{\eta}^n_{t_n}| \le \gamma_n\right] = \mathcal{O}((\log n)^{1-c}).$$

In view of (3.18), this completes the proof. □

Proof. (Of Lemma 3.4.4.) First consider a single inner excursion. Evidently there exists $\beta > 0$ such that, for all n:

(‡) For any $x \in (0, 5\mathcal{R}_n)$, if $|\eta^n_t| = x$ then the probability that η^n will either exit $\mathcal{B}_{5\mathcal{R}_n}(0)$ or hit 0 within its next three jumps is at least β.

Let $(J_l)_{l\ge 0}$ be the (a.s. finite) sequence of jump times of our inner excursion, and let B_k be the event that the excursion either coalesces or exits $\mathcal{B}_{5\mathcal{R}_n}(0)$ at one of $\{J_{3k+1}, J_{3k+2}, J_{3k+3}\}$. By the strong Markov property (applied at J_{3k}) and (‡), $\inf\{k \ge 0 \,:\, \mathbb{1}_{B_k} = 1\}$ is stochastically bounded above by a geometric random variable G with success probability β.

Moreover, for as long as η^n is not at 0, the rate at which it jumps is bounded below by the rate at which $\xi^{n,1}$ jumps, which is $\int_{\mathbb{R}^2} m^n(dz) = \Theta(n)$ where m^n is given by (3.5). Hence for each $l \ge 0$, $J_{l+1} - J_l$ is stochastically bounded above by E_l where the $(E_i)_{i\ge 0}$ are i.i.d. exponential random variables of this rate.

Combining these observations,

$$\begin{aligned}\mathbb{P}_{(0,5\mathcal{R}_n)}\left[\tau^{5\mathcal{R}_n} \wedge \tau_0 > n^{-1/2}\right] &\le \mathbb{P}\left[J_{\lceil 3n^{1/3}+3\rceil} \ge n^{-1/2}\right] + \mathbb{P}\left[G > n^{1/3}\right] \\ &= \mathcal{O}(n^{-1/6}) + (1-\beta)^{n^{1/3}} = \mathcal{O}(n^{-1/6})\end{aligned} \tag{3.25}$$

where the last line follows by Markov's inequality.

We are now in a position to complete the proof. Let n be sufficiently large that

$$(\log n)^{1/2}(n^{-1/2} + (\log n)^{-c-1}) \le (\log n)^{-c}. \tag{3.26}$$

Recall the definition of inner and outer excursions in (3.10) and the definitions of τ^{type} and i^* in Definition 3.4.2. Note that

$$\tau^{type} \leq \sum_{i=0}^{i^*}((\tau_i^{in} - \tau_i^{out}) + (\tau_{i+1}^{out} - \tau_i^{in})).$$

Thus, if η^n overshoots and $i^* < (\log n)^{1/2}$, then

$$(\log n)^{-c} = \tau^{type} \leq (\log n)^{1/2} \left(\max_{0 \leq i < (\log n)^{1/2}} (\tau_i^{in} - \tau_i^{out}) + \max_{0 \leq i < (\log n)^{1/2}} (\tau_{i+1}^{out} - \tau_i^{in}) \right).$$

It follows by (3.26) that either $\max_{0 \leq i < (\log n)^{1/2}}(\tau_i^{in} - \tau_i^{out}) \geq n^{-1/2}$ or $\max_{0 \leq i < (\log n)^{1/2}}(\tau_{i+1}^{out} - \tau_i^{in}) \geq (\log n)^{-c-1}$. Hence,

$$\begin{aligned}\mathbb{P}_{\zeta^n}\left[\eta^n \text{ overshoots}\right] \leq (\log n)^{1/2} \Big(& \mathbb{P}_{(0,5\mathcal{R}_n)}\left[\tau^{5\mathcal{R}_n} \wedge \tau_0 > n^{-1/2}\right] \\ & + \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}\left[\tau^{\gamma_n} \wedge \tau_{4\mathcal{R}_n} > (\log n)^{-c-1}\right] \Big) \\ & + \mathbb{P}_{\zeta^n}\left[i^* > (\log n)^{1/2}\right].\end{aligned}$$

Using (3.25), (3.13) and Lemma 3.4.3 to bound the right hand side of the above equation, we obtain

$$\begin{aligned}\mathbb{P}_{\zeta^n}\left[\eta^n \text{ overshoots}\right] &\leq (\log n)^{1/2}(\mathcal{O}(n^{-1/6}) + \mathcal{O}((\log n)^{1-c})) + \alpha^{(\log n)^{1/2}} \\ &= \mathcal{O}((\log n)^{3/2-c}),\end{aligned}$$

which completes the proof. □

Proof. (Of Lemma 3.4.5.) We note that the probability that η^n diverges is bounded above by the probability that a divergent outer excursion occurs before a coalescing inner excursion occurs. Let us write $\eta^{n,i,in}$ for the i^{th} inner excursion and $\eta^{n,i,out}$ for the i^{th} outer excursion and let us write $\tau^{r,i,in}, \tau_{r,i,in}$ and $\tau^{r,i,out}, \tau_{r,i,out}$ for the associated equivalents of τ^r and τ_r. Thus,

$$\begin{aligned}&\mathbb{P}_{\zeta^n}\left[\eta^n \text{ diverges}\right] \\ &\quad \leq \mathbb{P}_{\zeta^n}\left[\inf\left\{i \geq 1 \,:\, \tau^{\gamma_n,i,out} < \tau_{4\mathcal{R}_n,i,out}\right\} \leq \inf\{i \geq 0 \,:\, \tau_{0,i,in} < \tau^{5\mathcal{R}_n,i,in}\}\right].\end{aligned}$$

By the strong Markov property (applied successively at times τ_i^{out} and τ_i^{in}), along with (3.14) and (†), the right hand side of the above equation is bounded above by the probability that a geometric random variable with success probability $\Theta(\frac{1}{\log n})$ is smaller than an (independent) geometric random variable with success probability $\alpha' > 0$. With this in hand, an elementary calculation shows that

$$\mathbb{P}_{\zeta^n}\left[\eta^n \text{ diverges}\right] = \mathcal{O}\left(\frac{1}{\log n}\right).$$

It remains to prove a lower bound of the same order.

In similar style to (†) and (‡), it is easily seen that there exists $\delta > 0$ such that for all n:

($\star$) For any $x \in [\mathcal{R}_n, 4\mathcal{R}_n]$, if $|\eta_0^n| = x$, the probability that η^n will exit $\mathcal{B}_{5\mathcal{R}_n}(0)$ without coalescing is at least δ.

We note also that ζ^n is equal to $n^{-1/2}\zeta^1$ in distribution, so since we assumed that $\mu((\frac{3}{4}\mathcal{R}, \mathcal{R}]) > 0$, there exists $\varepsilon > 0$ such that $\mathbb{P}[\zeta^n \geq \mathcal{R}_n] \geq \varepsilon$ for all n. Thus, applying the strong Markov Property at time τ_0^{in} and using ($\star$), we obtain

$$\begin{aligned}\mathbb{P}_{\zeta^n}\left[\eta^n \text{ diverges}\right] &\geq \varepsilon\delta\mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}\left[\tau^{\gamma_n} < \tau_{4\mathcal{R}_n}\right] - \mathbb{P}_{\zeta^n}\left[\eta^n \text{ overshoots}\right] \\ &= \Theta\left(\frac{1}{\log n}\right)\end{aligned}$$

as required, where the final statement follows from Lemma 3.4.7 and Lemma 3.4.4 (since $c \geq 3$). □

3.4.1.2 Production of branches

The next step of the proof of Theorem 3.2.7 involves further analysis of pairs of potential ancestral lineages: first we need to check that once a pair has separated to a distance γ_n they won't come back together again before a fixed time K; second we need to see that $\log n$ times the divergence probability actually converges (c.f. Lemma 3.4.5) as $n \to \infty$, since this will determine the branching rate in our branching Brownian motion limit. These two statements are the object of the next two lemmas.

Lemma 3.4.9. *Fix $K \in (0,\infty)$. Then*

$$\mathbb{P}_{[(\log n)^{-c},\infty)}\left[\tau_{4\mathcal{R}_n} \le K\right] = \mathcal{O}\left(\frac{\log\log n}{\log n}\right).$$

Lemma 3.4.10. *There exists $\kappa \in (0,\infty)$ such that $(\log n)\mathbb{P}_{\zeta^n}\left[\eta^n \text{ diverges}\right] \to \kappa$ as $n \to \infty$.*

The remainder of this subsection is occupied with proving Lemmas 3.4.9 and 3.4.10.

Proof. (Of Lemma 3.4.9.) We use the Skorohod embedding of $\hat{\eta}$ into the Brownian motion W, as defined in (3.15), to reduce the claim to an equivalent statement about a two-dimensional Bessel process.

Recall that $\eta_0^n = \hat{\eta}_0^n = W_0$ and recall τ_r from (3.11), and that $\hat{\tau}_r$ and T_r are the analogues of τ_r for $\hat{\eta}$ and W respectively. We have that $\eta_s^n = \hat{\eta}_s^n$ for all $s \le \tau_{4\mathcal{R}_n}$ so

$$\begin{aligned}\mathbb{P}_{[(\log n)^{-c},\infty)}\left[\tau_{4\mathcal{R}_n} \le K\right] &= \mathbb{P}_{[(\log n)^{-c},\infty)}\left[\hat{\tau}_{4\mathcal{R}_n} \le K\right]\\ &\le \mathbb{P}_{[(\log n)^{-c},\infty)}\left[T_{4\mathcal{R}_n} \le T^{(S(K))}\right],\end{aligned} \tag{3.27}$$

where we used the Skorohod embedding given in (3.15) in the last line. For all $\tilde{K}, C > 0$, since $T^{(k)}$ is increasing in k we have

$$\mathbb{P}\left[T^{(S(K))} \ge \tilde{K}\right] \le \mathbb{P}\left[S(K) \ge Cn\right] + \mathbb{P}\left[T^{(Cn)} \ge \tilde{K}\right]. \tag{3.28}$$

By its definition in (3.15), $S(K)$ is bounded by the sum of two Poisson random variables with parameter $\chi = K\int_{\mathbb{R}^2} m^n(dz)$, where m^n is given by (3.5). In particular, $\chi = \Theta(n)$. Recall that if Z' is Poisson with parameter χ, then (using a Chernoff bound argument) for $k > \chi$,

$$\mathbb{P}[Z' > k] \le \frac{e^{-\chi}(e\chi)^k}{k^k}. \tag{3.29}$$

Hence, for C sufficiently large, there exists $\delta_1 > 0$ such that

$$\mathbb{P}\left[S(K) \ge Cn\right] \le \mathcal{O}(e^{-\delta_1 n}). \tag{3.30}$$

Now by the definition of $(T^{(m)})_{m\geq 1}$ in (3.15), and since $r_m \leq 2\mathcal{R}_n$ for each m,

$$\mathbb{P}\left[T^{(Cn)} \geq \tilde{K}\right] \leq \mathbb{P}\left[\sum_{i=1}^{Cn} R_i \geq \tilde{K}n\right],$$

where $(R_i)_{i\geq 1}$ is an i.i.d. sequence with $R_1 \overset{d}{=} \inf\{t \geq 0 : |W_t| \geq 2\mathcal{R}\}$. Since

$$\mathbb{P}\left[R_1 \geq k\right] \leq \mathbb{P}\left[R_1 \geq k-1\right]\mathbb{P}\left[|W_k - W_{k-1}| \leq 4\mathcal{R}\right] \leq \mathbb{P}\left[|W_1 - W_0| \leq 4\mathcal{R}\right]^k,$$

there exists $\lambda > 0$ such that $\mathbb{E}\left[e^{\lambda R_1}\right] < \infty$. Hence by Cramér's theorem, for $\tilde{K}$ a sufficiently large constant, there exists $\delta_2 > 0$ such that

$$\mathbb{P}\left[T^{(Cn)} \geq \tilde{K}\right] = \mathcal{O}(e^{-\delta_2 n}). \tag{3.31}$$

By (3.27) and (3.28) together with (3.30) and (3.31), we now have for $\tilde{K}$ sufficiently large

$$\mathbb{P}_{[(\log n)^{-c},\infty)}\left[\tau_{4\mathcal{R}_n} \leq K\right] \leq \mathbb{P}_{[(\log n)^{-c},\infty)}\left[T_{4\mathcal{R}_n} \leq \tilde{K}\right] + \mathcal{O}(e^{-\delta_1 n}) + \mathcal{O}(e^{-\delta_2 n}). \tag{3.32}$$

To finish, we note that

$$\begin{aligned}
&\mathbb{P}_{[(\log n)^{-c},\infty)}\left[T_{4\mathcal{R}_n} \leq \tilde{K}\right] \\
&\quad\leq \sup_{x\geq(\log n)^{-c}}\left(\mathbb{P}_x\left[T_{4\mathcal{R}_n} \leq T^{x+\log n}\right] + \mathbb{P}_x\left[T^{x+\log n} \leq \tilde{K}\right]\right) \\
&\quad\leq \sup_{x\geq(\log n)^{-c}}\left(\frac{\log(x+\log n) - \log x}{\log(x+\log n) - \log(4\mathcal{R}_n)}\right) + \mathbb{P}\left[\sup_{t\leq\tilde{K}}|W_t - W_0| \geq \log n\right] \\
&\quad= \mathcal{O}\left(\frac{\log\log n}{\log n}\right) + \mathcal{O}(e^{-(8\tilde{K})^{-1}(\log n)^2}),
\end{aligned}$$

where the second inequality uses the scale function for a two-dimensional Bessel process, and the last line uses (3.12). Substituting this into (3.32), we have the required result. □

Proof. (Of Lemma 3.4.10.) Let $p_n := \mathbb{P}_{\zeta^n}\left[\tau^{\gamma_n} < \tau_0\right]$. Note that by Lemma 3.4.4,

$$|p_n - \mathbb{P}_{\zeta^n}\left[\eta^n \text{ diverges}\right]| = \mathcal{O}\left(\frac{1}{(\log n)^{c-3/2}}\right). \tag{3.33}$$

Hence by Lemma 3.4.5, there exist $0 < d \leq D < \infty$ such that for all $n \geq 2$,

$$d \leq (\log n)p_n \leq D.$$

It follows that $(p_n)_{n\geq 1}$ has a subsequence $(p_{n_k})_{k\geq 1}$ such that $(\log n_k)p_{n_k} \to \kappa \in (0,\infty)$. Let $\varepsilon > 0$ and let $N \in \mathbb{N}$ be such that $N \geq 1/\varepsilon$ and $|(\log N)p_N - \kappa| \leq \varepsilon$. By rescaling, noting that $\zeta^n \overset{d}{=} \zeta^N(\frac{N}{n})^{1/2}$, and similarly for η^n, we have

$$p_N = \mathbb{P}_{\zeta^n}\left[\tau^{\gamma_N(Nn^{-1})^{1/2}} < \tau_0\right]. \tag{3.34}$$

Recall, for clarity, that here (as throughout this section) τ^r and τ_0 refer to the stopping times for the process η^n.

Define $X^{n,N} := |\eta^n_{\tau^{\gamma_N(Nn^{-1})^{1/2}}}|$. Increasing N, we may assume that $7\mathcal{R}_n < \gamma_N(Nn^{-1})^{1/2} \leq \gamma_n$ for $n \geq N$. Thus,

$$\begin{aligned} p_n &= \mathbb{P}_{\zeta^n}\left[\tau^{\gamma_N(Nn^{-1})^{1/2}} \leq \tau^{\gamma_n} < \tau_0\right] \\ &= \mathbb{E}_{\zeta^n}\left[\mathbb{1}_{\tau^{\gamma_N(Nn^{-1})^{1/2}} < \tau_0} \mathbb{P}_{X^{n,N}}\left[\tau^{\gamma_n} < \tau_0\right]\right]. \end{aligned} \tag{3.35}$$

Here, the first line holds since $\zeta_n < \gamma_N(Nn^{-1})^{1/2} \leq \gamma_n$, and the second line follows from the first by applying the Strong Markov Property at time $\tau^{\gamma_N(Nn^{-1})^{1/2}}$.

To estimate (3.35), note that

$$X^{n,N} \in [l^{n,N}, r^{n,N}] := [\gamma_N(Nn^{-1})^{1/2}, \gamma_N(Nn^{-1})^{1/2} + 2\mathcal{R}_n].$$

Using the Skorohod embedding defined in (3.15),

$$\begin{aligned} \mathbb{P}_{[l^{n,N}, r^{n,N}]}\left[\tau^{\gamma_n} < \tau_0\right] &\geq \inf_{x \geq \gamma_N(Nn^{-1})^{1/2}} \mathbb{P}_x\left[\tau^{\gamma_n} < \tau_{7\mathcal{R}_n}\right] \\ &\geq \inf_{x \geq \gamma_N(Nn^{-1})^{1/2}} \mathbb{P}_x\left[T^{\gamma_n + 2\mathcal{R}_n} < T_{7\mathcal{R}_n}\right] \\ &= \frac{\log(\gamma_N(Nn^{-1})^{1/2}) - \log(7\mathcal{R}_n)}{\log(\gamma_n + 2\mathcal{R}_n) - \log(7\mathcal{R}_n)} \\ &= \frac{\frac{1}{2}\log N + \mathcal{O}(\log\log N)}{\frac{1}{2}\log n + \mathcal{O}(\log\log n)}. \end{aligned} \tag{3.36}$$

Note that, in the above, we (again) use the scale function for a two-dimensional Bessel process to deduce the third line.

We require slightly more work to establish an upper bound. We have

$$\mathbb{P}_{[l^{n,N},r^{n,N}]}\left[\tau^{\gamma_n} < \tau_0\right] \le \mathbb{P}_{[l^{n,N},r^{n,N}]}\left[\tau^{\gamma_n} < \tau_{7\mathcal{R}_n}\right] + \mathbb{P}_{[l^{n,N},r^{n,N}]}\left[\tau_{7\mathcal{R}_n} < \tau^{\gamma_n} < \tau_0\right]. \quad (3.37)$$

We begin by controlling the second term on the right hand side of (3.37). By the Strong Markov Property at time $\tau_{7\mathcal{R}_n}$,

$$\begin{aligned} \mathbb{P}_{[l^{n,N},r^{n,N}]}\left[\tau_{7\mathcal{R}_n} < \tau^{\gamma_n} < \tau_0\right] &= \mathbb{E}_{[l^{n,N},r^{n,N}]}\left[\mathbb{1}_{\tau_{7\mathcal{R}_n} < \tau^{\gamma_n}} \mathbb{P}_{|\eta^n_{\tau_{7\mathcal{R}_n}}|}\left[\tau^{\gamma_n} < \tau_0\right]\right] \\ &\le \mathbb{E}_{[l^{n,N},r^{n,N}]}\left[\mathbb{P}_{|\eta^n_{\tau_{7\mathcal{R}_n}}|}\left[\tau^{\gamma_n} < \tau_0\right]\right]. \end{aligned}$$

Since $|\eta^n_{\tau_{7\mathcal{R}_n}}| \in [5\mathcal{R}_n, 7\mathcal{R}_n]$, using (3.14) in the same way as in the proof of Lemma 3.4.5,

$$\mathbb{P}_{[l^{n,N},r^{n,N}]}\left[\tau_{7\mathcal{R}_n} < \tau^{\gamma_n} < \tau_0\right] = \mathcal{O}\left(\frac{1}{\log n}\right). \quad (3.38)$$

Next, we control the first term on the right hand side of (3.37), again using the Skorohod embedding (3.15):

$$\begin{aligned} \mathbb{P}_{[l^{n,N},r^{n,N}]}\left[\tau^{\gamma_n} < \tau_{7\mathcal{R}_n}\right] &\le \mathbb{P}_{[l^{n,N},r^{n,N}]}\left[T^{\gamma_n} < T_{5\mathcal{R}_n}\right] \\ &\le \frac{\log(\gamma_N (Nn^{-1})^{1/2} + 2\mathcal{R}_n) - \log(5\mathcal{R}_n)}{\log(\gamma_n) - \log(5\mathcal{R}_n)} \\ &= \frac{\frac{1}{2}\log N + \mathcal{O}(\log\log N)}{\frac{1}{2}\log n + \mathcal{O}(\log\log n)}. \end{aligned} \quad (3.39)$$

Combining (3.36), (3.37), (3.38) and (3.39),

$$\mathbb{P}_{[l^{n,N},r^{n,N}]}\left[\tau^{\gamma_n} < \tau_0\right] = \frac{\log N + \mathcal{O}(\log\log N)}{\log n + \mathcal{O}(\log\log n)} + \mathcal{O}\left(\frac{1}{\log n}\right).$$

Hence by (3.35),

$$\begin{aligned} p_n &= \mathbb{P}_{\zeta^n}\left[\tau^{\gamma_N (Nn^{-1})^{1/2}} < \tau_0\right]\left(\frac{\log N + \mathcal{O}(\log\log N)}{\log n + \mathcal{O}(\log\log n)} + \mathcal{O}\left(\frac{1}{\log n}\right)\right) \\ &= \frac{(\log N) p_N}{\log n}\left(\frac{1 + \mathcal{O}(\frac{\log\log N}{\log N})}{1 + \mathcal{O}(\frac{\log\log n}{\log n})} + \mathcal{O}\left(\frac{1}{\log N}\right)\right), \end{aligned}$$

where we used (3.34) in the last line. Since $|(\log N)p_N - \kappa| \leq \varepsilon$ we obtain for $n \geq N$

$$\begin{aligned} (\log n)p_n &\geq (\kappa - \varepsilon)\left(\frac{1+\mathcal{O}(\frac{\log\log N}{\log N})}{1+\mathcal{O}(\frac{\log\log n}{\log n})} + \mathcal{O}\left(\frac{1}{\log N}\right)\right) \\ \text{and} \quad (\log n)p_n &\leq (\kappa + \varepsilon)\left(\frac{1+\mathcal{O}(\frac{\log\log N}{\log N})}{1+\mathcal{O}(\frac{\log\log n}{\log n})} + \mathcal{O}\left(\frac{1}{\log N}\right)\right). \end{aligned}$$

Letting $\varepsilon \to 0$ and hence $N \to \infty$, $\lim_{n\to\infty}(\log n)p_n = \kappa$. The result follows by (3.33). □

3.4.2 Convergence to branching Brownian motion

In this subsection we identify particular subsets of the dual process that we couple with objects that we call 'caterpillars'. The caterpillars play the rôle of individual branches in the limiting branching Brownian motion. Our (eventual) goal is to write down a system of 'branching caterpillars' and couple it to the SLFVS dual. Establishing these couplings is greatly simplified by viewing the branching and coalescing dual as a deterministic function of an augmented driving Poisson point process and so our first task is to recast the SLFVS dual in this way.

Recall that we have a fixed impact parameter $u \in (0, 1]$. We define, recursively, a sequence of subsets of $[0, 1]$ as follows:

$$A_u^1 = [0, u], \text{ and for } k \geq 1,\ A_u^{k+1} = uA_u^k \cup (u + (1-u)A_u^k).$$

Then if $U \sim \text{Unif}[0, 1]$, $(\mathbb{1}_{A_u^k}(U))_{k\geq 1}$ is an i.i.d. sequence of Bernoulli(u) random variables (see Lemma 3.20 in Kallenberg (2006) for a proof in the case $u = \frac{1}{2}$, where $(\mathbb{1}_{A_u^k}(U))_{k\geq 1}$ is the binary expansion of U; the general case is an easy extension of this).

Let

$$\mathscr{X} = \mathbb{R} \times \mathbb{R}^2 \times \mathbb{R}_+ \times \mathcal{B}_1(0)^2 \times [0, 1]^2.$$

Definition 3.4.11 (The dual as a deterministic function of a driving point process). *Given a simple point process* Π *on* $\mathscr{X}$*, and some* $p \in \mathbb{R}^2$*, we define* $(\mathcal{P}_t(p, \Pi))_{t\geq 0}$ *as a process on* $\cup_{k=1}^{\infty}(\mathbb{R}^2)^k$ *as follows.*

For each $t \geq 0$, $\mathcal{P}_t(p, \Pi) = (\xi_t^1, \dots, \xi_t^{N_t})$ *for some* $N_t \geq 1$. *We refer to* i *as the index of the ancestor* ξ_t^i. *We begin at time* $t = 0$ *from a single ancestor* $\mathcal{P}_0(p, \Pi) = \xi_0^1 = p$ *and proceed as follows.*

At each $(t, x, r, z_1, z_2, q, v) \in \Pi$ *with* $v \geq \boldsymbol{s_n}$, *a neutral event occurs:*

1. *Let* $\xi_{t-}^{n_1}, \dots, \xi_{t-}^{n_m}$ *denote the ancestors in* $\mathcal{B}_r(x)$ *which have not yet coalesced with an ancestor of lower index, with* $n_1 < \ldots < n_m$. *For* $1 \leq i \leq m$, *mark the ancestor* $\xi_{t-}^{n_i}$ *iff* $q \in A_u^i$. *Let* $\xi_{t-}^{r_1}, \dots \xi_{t-}^{r_l}$ *denote the marked ancestors.*

2. *If at least one ancestor is marked, we set* $\xi_t^{r_i} = x + rz_1$ *for each* i *and call this the parental location for the event. We say that the ancestor* $\xi_t^{r_i}$ *has coalesced with the ancestor* $\xi_t^{r_1}$, *for each* $i \geq 2$.

At each $(t, x, r, z_1, z_2, q, v) \in \Pi$ *with* $v < \boldsymbol{s_n}$, *a selective event occurs:*

1. *Let* $\xi_{t-}^{n_1}, \dots, \xi_{t-}^{n_m}$ *denote the ancestors in* $\mathcal{B}_r(x)$ *which have not yet coalesced with an ancestor of lower index, with* $n_1 < \ldots < n_m$. *For* $1 \leq i \leq m$, *mark the ancestor* $\xi_{t-}^{n_i}$ *iff* $q \in A_u^i$. *Let* $\xi_{t-}^{r_1}, \dots \xi_{t-}^{r_l}$ *denote the marked ancestors.*

2. *If at least one ancestor is marked, we set* $\xi_t^{r_i} = x + rz_1$ *for each* i *and add an ancestor* $\xi_t^{N_{t-}+1} = x + rz_2$. *We call* $x + rz_1$ *and* $x + rz_2$ *the parental locations of the event. We say that the ancestor* $\xi_t^{r_i}$ *has coalesced with the ancestor* $\xi_t^{r_1}$, *for each* $i \geq 2$.

For each $l \in \mathbb{N}$, *if* ξ_τ^l *has coalesced with an ancestor* ξ_τ^k *of lower index at time* τ, *we set* $\xi_t^l = \xi_t^k$ *for all* $t \geq \tau$.

In the same way as for the definition of $\mathcal{P}^{(n)}(p)$ before the statement of Theorem 3.2.7, we shall view $(\mathcal{P}_t(p, \Pi))_{t \geq 0}$ as a collection of potential ancestral lineages. Given a realization of Π, we say that a path that begins at p is a potential ancestral lineage if (1) at each neutral event that it encounters, it moves to the (single) parent and (2) at each selective event it encounters, it moves to one of the parents of that event.

Note that if Π is a Poisson point process on $\mathscr{X}$ with intensity measure

$$n\, dt \otimes n\, dx \otimes \mu^n(dr) \otimes \pi^{-1} dz_1 \otimes \pi^{-1} dz_2 \otimes dq \otimes dv \tag{3.40}$$

then as a collection of potential ancestral lineages, $(\mathcal{P}_t(p,\Pi))_{t\geq 0}$ has the same distribution as $\mathcal{P}^{(n)}(p)$.

When Π takes this form, the result is that the driving Poisson Point Process in (3.3) has been augmented by components that determine the nature of each event (neutral or selective), the parental locations of each event and which lineages in the region of the event are affected by it. We have abused notation by retaining the notation Π for this augmented process.

3.4.2.1 The caterpillar

We now introduce the notion of a caterpillar, which involves following a pair of potential ancestral lineages in the dual. We stop the caterpillar if the pair of lineages reaches displacement of $(\log n)^{-c}$, or if the pair does not coalesce within time $(\log n)^{-c}$ after last branching. While doing so, we suppress the creation of the second potential parent at any selective events that occur within time $(\log n)^{-c}$ of the previous (unsuppressed) selective event.

Let Π be a Poisson point process on $\mathscr{X}$ with intensity measure (3.40). We write $(\mathcal{P}_t(p,\Pi))_{t\geq 0} = (\xi_t^1,\dots,\xi_t^{N_t})_{t\geq 0}$ as defined in Definition 3.4.11.

Definition 3.4.12 (Caterpillar). *For $p\in\mathbb{R}^2$, we define a lifetime $h(p,\Pi)>0$, and a process $(c_t(p,\Pi))_{0\leq t\leq h(p,\Pi)}$ on $(\mathbb{R}^2)^2$, which we shall refer to as a caterpillar. For each $t\geq 0$, we write*

$$c_t(p,\Pi) = \left(c_t^1(p,\Pi), c_t^2(p,\Pi)\right),$$

dropping the dependence on (p,Π) from our notation when convenient. As part of the definition, we will also define $k^*(p,\Pi)\in\mathbb{N}$ and a sequence $(\tau_k^{br})_{k\leq k^*}$ of stopping times.

Set $\tau_0^{br}=0$ and let τ_1^{br} be the time of the first selective event after $(\log n)^{-c}$ to affect ξ^1. For $t\leq\tau_1^{br}$, let $c_t^1=c_t^2=\xi_t^1$.

Then, for $k\geq 1$, suppose we have defined $(\tau_l^{br})_{l\leq k}$; let $m(k)=N_{\tau_k^{br}}$.

For $t\in[\tau_k^{br},\tau_k^{br}+(\log n)^{-c}]$, define $c_t^1(p,\Pi)=\xi_t^1$ and $c_t^2(p,\Pi)=\xi_t^{m(k)}$.

In analogy with Definition 3.4.2, define

$$\tau_k^{div} = \inf\{t\geq\tau_k^{br} : |c_t^1-c_t^2|\geq(\log n)^{-c}\},$$

$$\begin{aligned}\tau_k^{coal} &= \inf\{t \geq \tau_k^{br} : c_t^1 = c_t^2\},\\ \tau_k^{over} &= \tau_k^{br} + (\log n)^{-c},\end{aligned} \tag{3.41}$$

and let $\tau_k^{type} = \min(\tau_k^{div}, \tau_k^{coal}, \tau_k^{over})$*. If* $\tau_k^{type} \neq \tau_k^{coal}$ *then set* $k^*(p,\Pi) = k$ *and* $h(p,\Pi) = \tau_{k^*}^{type}$*. The definition is then complete. If not, we proceed as follows.*

Let τ_{k+1}^{br} *be the time of the first selective event occurring strictly after* $\tau_k^{br} + (\log n)^{-c}$ *to affect* ξ^1*. For* $t \in [\tau_k^{br} + (\log n)^{-c}, \tau_{k+1}^{br})$*, let* $c_t^1(p,\Pi) = c_t^2(p,\Pi) = \xi_t^1$*.*

We then continue iteratively for each $k \leq k^*(p,\Pi)$*.*

We refer to $(\tau_k^{\text{br}})_{k \leq k^*}$, the times at which a selective event results in branching, as branching events. We shall abuse our previous terminology and say that a branching event diverges, coalesces or overshoots when the same is true of the excursion corresponding to the pair (c^1, c^2).

Remark 3.4.13. *Note that* $(c_t)_{t\geq 0}$ *is not a Markov process with respect to its natural filtration, since* c^1 *and* c^2 *are not allowed to branch off from each other within* $(\log n)^{-c}$ *of the previous branching event. However, for* $i = 1, 2$, $(c_t^i(p,\Pi))_{0\leq t\leq h(p,\Pi)}$ *is a Markov process with the same jump rate and jump distribution as a single potential ancestral lineage in the rescaled SLFVS dual. Moreover for each* $1 \leq k \leq k^*$, $(c_t^1, c_t^2)_{\tau_k^{br}\leq t\leq \tau_k^{type}}$ *is an excursion as defined in Section 3.4.1.*

Recall the definition of $m^n(dz)$ from (3.5) and let

$$\kappa_n = (\log n)\mathbb{P}[\tau_1^{type} \neq \tau_1^{coal}] \quad \text{and} \quad \lambda = n^{-1}\int_{\mathbb{R}^2} m^n(dz) = uV_1\int_0^{\mathcal{R}} r^d \mu(dr). \tag{3.42}$$

By combining Lemmas 3.4.10 and 3.4.4,

$$\kappa_n \to \kappa \tag{3.43}$$

as $n \to \infty$.

By the strong Markov property of Π, and since $\tau_k^{\text{type}} \leq \tau_k^{\text{br}} + (\log n)^{-c} < \tau_{k+1}^{\text{br}}$ for each k, the types of the selective events, $(\{\tau_k^{\text{type}} = \tau_k^{\text{div}}\})_{k\geq 1}$, $(\{\tau_k^{\text{type}} = \tau_k^{\text{coal}}\})_{k\geq 1}$ and

$(\{\tau_k^{\text{type}} = \tau_k^{\text{over}}\})_{k\geq 1}$ are each i.i.d. sequences. Thus,

$$k^*(p,\Pi) \sim \operatorname{Geom}(\kappa_n(\log n)^{-1}). \tag{3.44}$$

By (3.43), there exist constants $0 < a \leq A < \infty$ such that $\kappa_n \in [a, A]$ for all n sufficiently large, so

$$\mathbb{P}[k^* \geq (\log n)^{9/8}] = (1 - \tfrac{\kappa_n}{\log n})^{(\log n)^{9/8}} = \mathcal{O}(e^{-\delta(\log n)^{1/8}}) \tag{3.45}$$

for some $\delta > 0$.

Lemma 3.4.14. *We can couple $h(p,\Pi)$ with $H \sim$ Exp$(\kappa_n\lambda)$ in such a way that for some $\delta > 0$, with probability at least $1 - \mathcal{O}(e^{-\delta(\log n)^{1/8}})$*

$$|h(p,\Pi) - H| \leq 3(\log n)^{-1/4}.$$

Proof. Recall the definition of λ in (3.42). Since the total rate at which c^1 jumps is given by λn, and each jump is from a selective event independently with probability $\boldsymbol{s_n} = \frac{\log n}{n}$, by the strong Markov property of Π we have that

$$E_k := \tau_k^{\text{br}} - (\tau_{k-1}^{\text{br}} + (\log n)^{-c}) \sim \operatorname{Exp}(\lambda \log n) \tag{3.46}$$

and $(E_k, \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k\geq 1}$ is an i.i.d. sequence.

Since (for example) $\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}$ is not independent of the radius of the event at τ_k^{br}, we note that E_k and $\mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}}$ are not independent; therefore $(E_k)_{k\geq 1}$ is not independent of k^*. However, we can couple $(E_k, \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k\geq 1}$ with a sequence $(E'_k)_{k\geq 1}$ which is independent of k^* as follows.

First sample the sequence $(\mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k\geq 1}$, and then independently sample a sequence $(E'_k, A_k)_{k\geq 1}$ with the same distribution as $(E_k, \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k\geq 1}$. Then, for each $k \geq 1$, if $A_k = \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}}$ set $E_k = E'_k$, and if not sample E_k according to its conditional distribution given $\mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}}$.

We now have a coupling of $(E_k, \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k\geq 1}$ and $(E'_k)_{k\geq 1}$ such that $(E'_k)_{k\geq 1}$ is an i.i.d. sequence, independent of k^*, with $E'_1 \sim \operatorname{Exp}(\lambda \log n)$. Also, since $\mathbb{P}[\tau_k^{\text{type}} \neq$

$\tau_k^{\mathrm{coal}}] = \Theta((\log n)^{-1})$, we have that independently for each k, $E_k = E'_k$ with probability at least $1 - \Theta((\log n)^{-1})$.

We write

$$\sum_{k=1}^{k^*} E_k = \sum_{k=1}^{k^*} E'_k + \sum_{k=1}^{k^*} D_k,$$

where $D_k = E_k - E'_k$ and, by (3.44), $\sum_{k=1}^{k^*} E'_k \sim \mathrm{Exp}(\lambda\kappa_n)$.

Our next step is to bound $\sum_{k=1}^{k^*} D_k$. Firstly, applying a Chernoff bound to the binomial distribution yields

$$\begin{aligned}
&\mathbb{P}\left[\left|\{k < (\log n)^{9/8} : D_k \neq 0\}\right| \geq (\log n)^{1/4}\right] \\
&\qquad = \mathbb{P}\left[\mathrm{Bin}\big((\log n)^{9/8}, \Theta((\log n)^{-1})\big) \geq (\log n)^{1/4}\right] \\
&\qquad = \mathcal{O}\big(\exp(-\delta'(\log n)^{1/4})\big) \qquad (3.47)
\end{aligned}$$

for some $\delta' > 0$. Secondly,

$$\begin{aligned}
\mathbb{P}\left[|D_1| \geq (\log n)^{-1/2}\right] &\leq \mathbb{P}\left[E_1 \geq \tfrac{1}{2}(\log n)^{-1/2}\right] + \mathbb{P}\left[E'_1 \geq \tfrac{1}{2}(\log n)^{-1/2}\right] \\
&= 2\exp(-\lambda(\log n)^{1/2}/2). \qquad (3.48)
\end{aligned}$$

Combining (3.45), (3.47) and (3.48), we have that

$$\mathbb{P}\left[\sum_{k=1}^{k^*} D_k \geq (\log n)^{-1/4}\right] = \mathcal{O}\left(e^{-\delta''(\log n)^{1/8}}\right), \qquad (3.49)$$

for some $\delta'' \in (0, \delta)$.

Note that

$$\sum_{k=1}^{k^*} E_k = \tau_{k^*}^{\mathrm{br}} - k^*(\log n)^{-c} = h - k^*(\log n)^{-c} - (\tau_{k^*}^{\mathrm{type}} - \tau_{k^*}^{\mathrm{br}}),$$

with $0 \leq \tau_{k^*}^{\mathrm{type}} - \tau_{k^*}^{\mathrm{br}} \leq (\log n)^{-c}$. Let $H = \sum_{k=1}^{k^*} E'_k$. Then by (3.45) and (3.49), we have

$$\mathbb{P}\left[|h(p,\Pi) - H| \geq (\log n)^{9/8-c} + (\log n)^{-c} + (\log n)^{-1/4}\right] = \mathcal{O}\left(e^{-\delta''(\log n)^{1/8}}\right).$$

The result follows since $c \geq 3$. □

Our next step is to show that a caterpillar is unlikely to end with an overshooting event.

Lemma 3.4.15. *As $n \to \infty$, $\mathbb{P}\left[\tau_{k^*}^{type} = \tau_{k^*}^{over}\right] = \mathcal{O}\left((\log n)^{\frac{21}{8}-c}\right)$.*

Proof. By Lemma 3.4.4, for $k \geq 1$

$$\mathbb{P}[\tau_k^{\text{type}} = \tau_k^{\text{over}}] = \mathcal{O}((\log n)^{\frac{3}{2}-c}). \tag{3.50}$$

Moreover,

$$\{\tau_{k^*}^{\text{type}} = \tau_{k^*}^{\text{over}}\} \subset \{k^* \geq (\log n)^{9/8}\} \cup \bigcup_{k=1}^{(\log n)^{9/8}} \{\tau_k^{\text{type}} = \tau_k^{\text{over}}\}.$$

It follows, using (3.45), that

$$\mathbb{P}[\tau_{k^*}^{\text{type}} = \tau_{k^*}^{\text{over}}] = \mathcal{O}(e^{-\delta(\log n)^{1/8}}) + \mathcal{O}((\log n)^{\frac{3}{2}+\frac{9}{8}-c}) = \mathcal{O}((\log n)^{\frac{21}{8}-c}).$$

This completes the proof. □

We now show that a single caterpillar can be coupled to a Brownian motion in such a way that the caterpillar closely follows the Brownian motion during time $[0, h(p, \Pi)]$.

Recall that the rate at which ξ^1 jumps from y to $y + z$ is given by intensity measure $m^n(dz)$, defined in (3.5). Thus for $(c_t)_{t\geq 0}$ started at p, $\mathbb{E}[c_t^1] = p$ and the covariance matrix of c_t^1 is $\sigma^2 t \text{Id}$ since by (3.7),

$$\sigma^2 = \frac{1}{2}\int_{\mathbb{R}^2} |z|^2 m^n(dz).$$

Armed with this, the following lemma is no surprise.

Lemma 3.4.16. *Let $(W_t)_{t\geq 0}$ be a two-dimensional Brownian motion with $W_0 = p$. We can couple $(c_t(p, \Pi))_{t \leq h(p,\Pi)}$ with $(W_t)_{t\geq 0}$, in such a way that $(W_t)_{t\geq 0}$ is independent of $(\tau_k^{br})_{k\geq 1}$ and $k^*(p, \Pi)$, and for any $r > 0$, with probability at least $1 - \mathcal{O}((\log n)^{-r})$, for $t \leq h(p, \Pi)$,*

$$|c_t^1(p, \Pi) - W_{\sigma^2 t}| \leq (\log n)^{\frac{9}{8}-\frac{c}{3}}.$$

Remark 3.4.17. *By the definition of the caterpillar in Definition 3.4.12, for all* $t \le h(p,\Pi)$, $|c_t^2 - c_t^1| \le (\log n)^{-c}$. *Hence under the coupling of Lemma 3.4.16, with probability at least* $1 - \mathcal{O}((\log n)^{-r})$, $|c_t^2(p,\Pi) - W_{\sigma^2 t}| \le 2(\log n)^{\frac{9}{8}-\frac{c}{3}}$.

Proof. The proof is closely related to the second half of the proof of Lemma 3.4.7. Note that for $k \ge 0$, on the time interval $[\tau_k^{br} + (\log n)^{-c}, \tau_{k+1}^{br})$, c_t^1 is a pure jump process with rate of jumps from y to $y+z$ given by $(1-\boldsymbol{s_n})m^n(dz)$. Let $(\tilde{c}_t)_{t\ge 0}$ be a pure jump process with $\tilde{c}_0 = 0$ and rate of jumps from y to $y+z$ given by $(1-\boldsymbol{s_n})m^n(dz)$. For $i \ge 1$, let

$$X_i = \tilde{c}_{i/n} - \tilde{c}_{(i-1)/n}.$$

Then $(X_i)_{i\ge 1}$ are i.i.d., and as in (3.19) and (3.20), we have $\mathbb{E}[|X_1|^2] = \frac{2\sigma^2(1-\boldsymbol{s_n})}{n}$ and $\mathbb{E}[|X_1|^4] = \mathcal{O}(n^{-2})$.

By the same Skorohod embedding argument as for (3.21), there is a two-dimensional Brownian motion W started at 0 and a sequence $\upsilon_1, \upsilon_2, \dots$ of stopping times for W such that for $i \ge 1$, $W_{\upsilon_i} = \tilde{c}_{i/n}$ and

$$\mathbb{P}[|\upsilon_{\lfloor tn \rfloor} - \sigma^2(1-\boldsymbol{s_n})t| \ge n^{-1/3}] \le \mathcal{O}(tn^{-1/3}).$$

Fix $t > 0$. Since $\boldsymbol{s_n} = \frac{\log n}{n}$, for n sufficiently large,

$$\mathbb{P}[|\upsilon_{\lfloor tn \rfloor} - \sigma^2 t| \ge 2n^{-1/3}] \le \mathcal{O}(n^{-1/3}).$$

Then by a union bound over $j = 1, \dots, \lfloor n^{1/4} t \rfloor$,

$$\begin{aligned} \mathbb{P}\left[\exists j \le \lfloor n^{1/4} t \rfloor : |\upsilon_{\lfloor jn^{3/4} \rfloor} - \sigma^2 j n^{-1/4}| \ge 2n^{-1/3}\right] &\le (n^{1/4}t)\mathcal{O}(n^{-1/3}) \\ &= \mathcal{O}(n^{-1/12}). \end{aligned} \tag{3.51}$$

Again by a union bound over j,

$$\mathbb{P}\Bigg[\exists j \le \lfloor n^{1/4} t \rfloor : \sup\Big\{|W_{\sigma^2 j n^{-1/4}} - W_u| \\ : u \in [\sigma^2 j n^{-1/4} - 2n^{-1/3}, \sigma^2 (j+1) n^{-1/4} + 2n^{-1/3}]\Big\} \ge n^{-1/10}\Bigg]$$

$$\begin{aligned}&\leq (n^{1/4}t)2\mathbb{P}\left[\sup\{|W_s - W_0| : s \in [0, 5n^{-1/4}]\} \geq \tfrac{1}{2}n^{-1/10}\right]\\&\leq 4n^{1/4}t\exp(-n^{1/20}/160) = o(n^{-1/12}).\end{aligned} \tag{3.52}$$

Here, the last line follows by (3.12).

Under the complement of the event in (3.51), for all $j < \lfloor n^{1/4}t\rfloor$,

$$|\upsilon_{\lfloor jn^{3/4}\rfloor} - \sigma^2 jn^{-1/4}| \leq 2n^{-1/3} \text{ and } |\upsilon_{\lfloor (j+1)n^{3/4}\rfloor} - \sigma^2(j+1)n^{-1/4}| \leq 2n^{-1/3},$$

which implies that for i such that $jn^{-1/4} \leq in^{-1} \leq (j+1)n^{-1/4}$,

$$\upsilon_i \in \left[\sigma^2 jn^{-1/4} - 2n^{-1/3}, \sigma^2(j+1)n^{-1/4} + 2n^{-1/3}\right].$$

Hence combining (3.51) and (3.52),

$$\mathbb{P}\left[\exists i \leq \lfloor tn\rfloor : |\tilde{c}_{i/n} - W_{\sigma^2 i/n}| \geq 2n^{-1/10}\right] = \mathcal{O}(n^{-1/12}).$$

Our next step is to control $|\tilde{c}_s - \tilde{c}_{i/n}|$ during the interval $s \in [i/n, (i+1)/n]$. The distribution of the number of jumps made by $\tilde{c}$ on an interval $[i/n, (i+1)/n]$ is Poisson with parameter $(1 - \boldsymbol{s_n})\lambda$, where λ is given by (3.42), and the maximum jump size is $2\mathcal{R}_n$; using (3.29) with $\chi = (1 - \boldsymbol{s_n})\lambda$ and $k = \log n$ gives that

$$\mathbb{P}\left[\exists i \leq \lfloor tn\rfloor : \sup_{s\in[i/n,(i+1)/n]} |\tilde{c}_s - \tilde{c}_{i/n}| \geq (\log n)2\mathcal{R}_n\right] = o(n^{-1}).$$

Hence for n large enough that $(\log n)2\mathcal{R}_n \leq n^{-1/10}$, using (3.52) again to bound $|W_s - W_{\sigma^2 i/n}|$ during the interval $[\sigma^2 i/n, \sigma^2(i+1)/n]$ we have

$$\mathbb{P}\left[\sup_{s\leq t} |\tilde{c}_s - W_{\sigma^2 s}| \geq 4n^{-1/10}\right] = \mathcal{O}(n^{-1/12}). \tag{3.53}$$

We now apply this coupling to $(c^1_t)_{\tau^{br}_k + (\log n)^{-c} \leq t < \tau^{br}_{k+1}}$ for each $k \geq 0$, and let the caterpillar evolve independently of the Brownian motion on each interval $[\tau^{\text{br}}_k, \tau^{\text{br}}_k + (\log n)^{-c}]$.

More precisely, let $(\tilde{c}^k)_{k\geq 0}$ be an i.i.d. sequence of pure jump processes with $\tilde{c}^k_0 = 0$ and rate of jumps from y to $y+z$ given by $(1-\boldsymbol{s_n})m^n(dz)$. Let $(W^k)_{k\geq 0}$ be an i.i.d. sequence of 2-dimensional Brownian motions started at 0 and for each $k \geq 0$, couple W^k and $\tilde{c}^k$ in the same way as above, so that for fixed $t>0$, for each $k \geq 0$,

$$\mathbb{P}\left[\sup_{s\leq t} |\tilde{c}^k_s - W^k_{\sigma^2 s}| \geq 4n^{-1/10}\right] = \mathcal{O}(n^{-1/12}). \tag{3.54}$$

Then by the Strong Markov property for the process c^1, we can couple $(\tilde{c}^k, W^k)_{k\geq 0}$ and c^1 in such a way that for $k \geq 0$ and $s \in [0, \tau^{br}_{k+1} - (\tau^{br}_k + (\log n)^{-c}))$,

$$c^1_{s+\tau^{br}_k + (\log n)^{-c}} - c^1_{\tau^{br}_k + (\log n)^{-c}} = \tilde{c}^k_s.$$

and $(\tilde{c}^k, W^k)_{k\geq 0}$ is independent of $\big(\tau^{\text{br}}_k, (c^1_t - c^2_t)|_{[\tau^{\text{br}}_k, \tau^{\text{br}}_k + (\log n)^{-c})}\big)_{k\geq 0}$.

Let B be another independent 2-dimensional Brownian motion started at 0. We now define a single Brownian motion W by piecing together increments of B and $(W^k)_{k\geq 0}$. For $s < \sigma^2(\log n)^{-c}$, let $W_s = B_s + p$. Then for $k \geq 0$, define the increments of W on the time interval $[\sigma^2(\tau^{br}_k + (\log n)^{-c}), \sigma^2(\tau^{br}_{k+1} + (\log n)^{-c}))$ as follows. For $s \in [0, \sigma^2(\tau^{br}_{k+1} - \tau^{br}_k))$, let

$$W_{s+\sigma^2(\tau^{br}_k + (\log n)^{-c})} - W_{\sigma^2(\tau^{br}_k + (\log n)^{-c})} = W^k_s.$$

Then W is a Brownian motion independent of $\big(\tau^{\text{br}}_k, (c^1_t - c^2_t)|_{[\tau^{\text{br}}_k, \tau^{\text{br}}_k + (\log n)^{-c})}\big)_{k\geq 0}$, which implies that W is independent of both k^* and $(\tau^{\text{br}}_k)_{k\geq 1}$.

We now check that W_t is close to c^1_t for $t < h$. By (3.46),

$$\mathbb{P}\left[\tau^{\text{br}}_{k+1} - \tau^{\text{br}}_k \geq 1 + (\log n)^{-c}\right] \leq n^{-\lambda}.$$

Hence applying (3.54) with $t = 1 + (\log n)^{-c}$ for each $k \leq (\log n)^{9/8}$ and using (3.45), we have that with probability at least $1 - \mathcal{O}(e^{-\delta(\log n)^{1/8}})$, for $0 \leq k \leq k^*$ and $t \in [\tau^{\text{br}}_k + (\log n)^{-c}, \tau^{\text{br}}_{k+1})$,

$$\Big| \Big(c^1_t - c^1_{\tau^{\text{br}}_k + (\log n)^{-c}}\Big) - \Big(W_{\sigma^2 t} - W_{\sigma^2(\tau^{\text{br}}_k + (\log n)^{-c})}\Big)\Big| \leq 4n^{-1/10}. \tag{3.55}$$

For each k, by (3.12),

$$\begin{aligned}\mathbb{P}\left[\sup\left\{|W_{\sigma^2 t} - W_{\sigma^2\tau_k^{\mathrm{br}}}| : t \in [\tau_k^{\mathrm{br}}, \tau_k^{\mathrm{br}} + (\log n)^{-c}]\right\} \geq \tfrac{1}{3}(\log n)^{-c/3}\right] \\ \leq 4\exp(-(\log n)^{c/3}/72\sigma^2) \\ = o\left((\log n)^{-r-\frac{9}{8}}\right), \end{aligned} \tag{3.56}$$

for any $r > 0$. Hence, using (3.45) again,

$$\begin{aligned}\mathbb{P}\left[\sum_{k=1}^{k^*} \sup\left\{|W_{\sigma^2 t} - W_{\sigma^2\tau_k^{\mathrm{br}}}| : t \in [\tau_k^{\mathrm{br}}, \tau_k^{\mathrm{br}} + (\log n)^{-c}]\right\} \geq \tfrac{1}{3}(\log n)^{\frac{9}{8}-\frac{c}{3}}\right] \\ \leq \mathbb{P}\left[k^* \geq (\log n)^{9/8}\right] + (\log n)^{9/8} o((\log n)^{-r-\frac{9}{8}}) \\ = o((\log n)^{-r}). \end{aligned} \tag{3.57}$$

For $k \geq 0$, on the time interval $[\tau_k^{\mathrm{br}}, \tau_k^{\mathrm{br}} + (\log n)^{-c}]$ the process c_t^1 is a pure jump process with rate of jumps from y to $y + z$ given by $m^n(dz)$. Hence using the same Skorohod embedding argument as for (3.53), we can couple $(c^1_{s+\tau_k^{\mathrm{br}}} - c^1_{\tau_k^{\mathrm{br}}})_{s \leq (\log n)^{-c}}$ with a Brownian motion W' started at 0 in such a way that

$$\mathbb{P}\left[\sup_{s \leq (\log n)^{-c}} |(c^1_{s+\tau_k^{\mathrm{br}}} - c^1_{\tau_k^{\mathrm{br}}}) - W_{\sigma^2 s}| \geq 4n^{-1/10}\right] = \mathcal{O}(n^{-1/12}).$$

Applying (3.57) and (3.45), it follows that

$$\begin{aligned}\mathbb{P}\Bigg[\sum_{k=1}^{k^*} \sup\left\{|c_t^1 - c^1_{\tau_k^{\mathrm{br}}}| : t \in [\tau_k^{\mathrm{br}}, \tau_k^{\mathrm{br}} + (\log n)^{-c}]\right\} \\ \geq \tfrac{1}{3}(\log n)^{\frac{9}{8}-\frac{c}{3}} + 4n^{-1/10}(\log n)^{9/8}\Bigg] \\ = \mathcal{O}((\log n)^{-r}).\end{aligned}$$

The stated result follows by combining the above equation with (3.55), (3.45) and (3.57). □

3.4.2.2 The branching caterpillar

We now construct a branching process of caterpillars. We start from a single caterpillar and allow it to evolve until the time h. We start two independent caterpillars from the locations of c_h^1 and c_h^2. Now iterate. The independent caterpillars defined in this way will be indexed by points of $\mathcal{U} = \{\emptyset\} \cup \bigcup_{k=1}^{\infty} \{1,2\}^k$. More formally:

Definition 3.4.18 (Branching caterpillar)**.** *Let* $(\Pi_j)_{j\in\mathcal{U}}$ *be a sequence of independent Poisson point processes on* $\mathscr{X}$ *with intensity measure* (3.40)*. For* $p \in \mathbb{R}^2$*, we define* $(\mathcal{C}_t(p, (\Pi_j)_{j\in\mathcal{U}}))_{t\geq 0}$ *as a process on* $\cup_{k=1}^{\infty}(\mathbb{R}^2)^k$ *as follows. For* $s > 0$*, let*

$$\Pi_j^s = \{(t-s, x, r, z_1, z_2, q, v) : (t, x, r, z_1, z_2, q, v) \in \Pi_j\}. \tag{3.58}$$

Define (p_j, t_j, h_j) *inductively for* $j \in \mathcal{U}$ *by* $p_\emptyset = p$*,* $t_\emptyset = 0$ *and*

$$\begin{aligned} h_j &= t_j + h(p_j, \Pi_j^{t_j}) \\ t_{(j,1)} &= t_{(j,2)} = h_j \\ p_{(j,1)} &= c^1_{h_j - t_j}(p_j, \Pi_j^{t_j}) \\ p_{(j,2)} &= c^2_{h_j - t_j}(p_j, \Pi_j^{t_j}). \end{aligned}$$

Finally, define $\mathcal{U}(t) = \{j \in \mathcal{U} : t_j \leq t \leq h_j\}$ *and*

$$\mathcal{C}_t(p, (\Pi_j)_{j\in\mathcal{U}}) = (c_{t-t_j}(p_j, \Pi_j^{t_j}))_{j\in\mathcal{U}(t)}.$$

In words, $\mathcal{U}(t)$ is the set of indices of the caterpillars that are active at time t, and $\mathcal{C}_t$ is the set of (positions of) those caterpillars. Note that we translate the time coordinates in (3.58) to match our definition of a caterpillar, which began at time 0. The jumps in $\mathcal{C}_t$ occur at the time coordinates of events in $\cup_{j\in\mathcal{U}}\Pi_j$.

We now show that for any constant $a > 0$, with high probability, the longest 'chain' of caterpillars has length at most $a \log\log n + 1$. For $k \in \mathbb{N}$, let $\mathcal{U}_k = \{\emptyset\} \cup \bigcup_{j=1}^{k}\{1,2\}^j$.

Lemma 3.4.19. *Fix* $T > 0$*; then for any* $r > 0$*,* $a > 0$*,* $\mathbb{P}[\mathcal{U}(T) \not\subseteq \mathcal{U}_{\lfloor a \log\log n\rfloor}] = o((\log n)^{-r})$*.*

Proof. Fix $v \in \{1,2\}^{\lfloor a\log\log n\rfloor+1}$. Then by a union bound,

$$\mathbb{P}\left[\exists w \in \{1,2\}^{\lfloor a\log\log n\rfloor+1} \text{ s.t. } t_w \le T\right] \le 2^{\lfloor a\log\log n\rfloor+1}\mathbb{P}[t_v \le T]. \tag{3.59}$$

Note that by Lemma 3.4.14, $t_v = \sum_{i=1}^{\lfloor a\log\log n\rfloor+1} H_i + R$ where $(H_i)_{i\ge 1}$ are i.i.d. with $H_1 \sim \mathrm{Exp}(\lambda\kappa_n)$ and

$$\mathbb{P}\left[|R| \ge 3(a\log\log n+1)(\log n)^{-1/4}\right] = \mathcal{O}((\log\log n)e^{-\delta(\log n)^{1/8}}).$$

Hence (if n is sufficiently large that $3(a\log\log n+1)(\log n)^{-1/4} \le T/2$), if Z' is Poisson with parameter $3\lambda\kappa_n T/2$,

$$\mathbb{P}[t_v \le T] \le \mathbb{P}[Z' \ge a\log\log n + 1] + \mathcal{O}\left((\log\log n)e^{-\delta(\log n)^{1/8}}\right).$$

We use (3.29) and combine with (3.59) to deduce that, for any $r > 0$,

$$\mathbb{P}[\mathcal{U}(T) \not\subseteq \mathcal{U}_{\lfloor a\log\log n\rfloor}] = \mathbb{P}\left[\exists w \in \{1,2\}^{\lfloor a\log\log n\rfloor+1} \text{ s.t. } t_w \le T\right] = o((\log n)^{-r}).$$

This completes the proof. □

The next task is to couple the branching caterpillar to the rescaled dual of the SLFVS. Since we have expressed the dual as a deterministic function of the driving point process of events in Definition 3.4.11, it is enough to find an appropriate coupling of the driving events for the branching caterpillar and those of a SLFVS dual.

The idea, roughly, is as follows. Each ‘branch’ of the branching caterpillar is constructed from an independent driving process. For each of these we should like to retain those events that affected the caterpillar, but we can discard the rest. If two or more caterpillars are close enough that the events affecting them could overlap, to avoid having too many events in these regions we have to arbitrarily choose one caterpillar and discard the events affecting the others. We then supplement these with additional events, appropriately distributed to fill in the gaps and arrive at the driving Poisson point process for a SLFVS dual, with intensity as in (3.40). We will then check that the

SLFVS dual corresponding to this point process coincides with our branching caterpillar, with probability tending to one as $n \to \infty$.

To put this strategy into practice we require some notation. Let $\mathcal{U}_0 = \mathcal{U} \cup \{0\}$. For $V \subset \mathcal{U}_0$ let $\max(V)$ refer to the maximum element of V with respect to a fixed ordering in which 0 is the minimum value (it does not matter precisely which ordering we use, but we must fix one). Given a sequence $(\Pi_j)_{j \in \mathcal{U}_0}$ of independent Poisson point processes on $\mathscr{X}$ with intensity measure (3.40), define a simple point process Π as follows. Let

$$j(t,x) = \max\left(\left\{k \in \mathcal{U}(t) : \exists i \in \{1,2\} \text{ with } |c^i_{t-t_k}(p_k, \Pi^{t_k}_k) - x| \leq \mathcal{R}_n\right\} \cup \{0\}\right). \quad (3.60)$$

Note that $j(t,x) = 0$ corresponds to regions of space-time that are not near a caterpillar, so that for $(t, x, r, z_1, z_2, q, v) \in \Pi_0$, $\mathcal{B}_r(x)$ does not contain a caterpillar. Then we define

$$\Pi = \bigcup_{k \in \mathcal{U}_0} \{(t, x, r, z_1, z_2, q, v) \in \Pi_k \ : \ j(t,x) = k\}. \quad (3.61)$$

Lemma 3.4.20. Π *is a Poisson point process with intensity measure given by* (3.40).

Remark 3.4.21. *We defined the coupling* (3.61) *for each* $n \in \mathbb{N}$. *As such, in the proof of Lemma 3.4.20 we regard* n *as a constant and we will not include it inside* $\mathcal{O}(\cdot)$, *etc.*

Proof. Let $\nu(dt, dx, dr, dz_1, dz_2, dq, dv)$ be the intensity measure given in (3.40).

Let $\mathcal{B}_0$ be the set of bounded Borel subsets of $\mathbb{R}_+ \times \mathbb{R}^2 \times \mathbb{R}_+ \times \mathcal{B}_1(0)^2 \times [0,1]^2$; for $B \in \mathcal{B}_0$, let $N(B) = |\Pi \cap B|$ and for $j \in \mathcal{U}_0$, let $N_j(B) = |\Pi_j \cap B|$. Suppose $B = \cup_{i=1}^k B_i \in \mathcal{B}_0$ where for each i, $B_i = [a_i, b_i] \times D_i$ for some $a = a_1 < b_1 \leq a_2 < \ldots < b_k = b$. Let $\mathcal{B}_R \subset \mathcal{B}_0$ denote the collection of such sets B. Note that Π is a simple point process, and that therefore Π is a Poisson point process with intensity ν if and only if

$$\mathbb{P}\left[N(B) = 0\right] = e^{-\nu(B)} \quad (3.62)$$

for all $B \in \mathcal{B}_R$. (See e.g. Section 3.4 of Kingman (1992).)

For some $\delta > 0$, assume that $b_i - a_i \leq \delta$, $\forall i$ (by partitioning the B_i further if necessary). Since B is bounded, $\exists\ d < \infty$ s.t. $|x| \leq d$ for all $(t, x, r, z_1, z_2, q, v) \in B$. We can write

$$\begin{aligned}\mathbb{P}[N(B)=0] &= \mathbb{P}[\cap_{i=1}^{k}\{N(B_i)=0\}]\\ &= \mathbb{E}\left[\prod_{i=1}^{k-1}\mathbb{1}_{\{N(B_i)=0\}}\mathbb{P}\Big(N(B_k)=0\Big|(\Pi_j(a_k))_{j\in\mathcal{U}_0}\Big)\right] \end{aligned} \tag{3.63}$$

where $\Pi_j(t) := \Pi_j|_{[0,t]\times\mathbb{R}^2\times\mathbb{R}_+\times\mathcal{B}_1(0)^2\times[0,1]^2}$.

For $j\in\mathcal{U}_0$, let $D_k^j = \{(x,r,z_1,z_2,q,v)\in D_k : j(a_k,x)=j\}$ and $B_k^j = [a_k,b_k]\times D_k^j$. Also let

$$\tilde{B}_k = [a_k,b_k]\times\mathcal{B}_{d+3\mathcal{R}_n}(0)\times\mathbb{R}_+\times\mathcal{B}_1(0)^2\times[0,1]^2,$$

and let $\mathcal{V}(t) = \cup_{s\le t}\mathcal{U}(s)$.

For $t\in[a_k,b_k]$, if none of the caterpillars in $\mathcal{B}_{d+3\mathcal{R}_n}(0)$ move during the time interval $[a_k,t]$ then $j(a_k,x)=j(t,x)$ $\forall x\in\mathcal{B}_d(0)$; thus a point (t,x,r,z_1,z_2,q,v) in $\Pi\cap B_k$ must be a point in $\Pi_j\cap B_k^j$ for some j, and vice versa. We can use this observation to relate $\{N(B_k)=0\}$ and $\cap_{j\in\mathcal{U}_0}\{N_j(B_k^j)=0\}$, as follows.

If $N(B_k)=0$ and $N_j(B_k^j)\neq 0$ for some $j\in\mathcal{U}_0$, then $D_k^j\neq\emptyset$ so $j\in\mathcal{V}(a_k)\cup\{0\}$ (either $j=0$ or the caterpillar indexed by j is alive at time a_k). Also after a_k and before the point in $\Pi_j\cap B_k^j$, one of the caterpillars in $\mathcal{B}_{d+3\mathcal{R}_n}(0)$ must have moved, so there must be a point in $\Pi_l\cap\tilde{B}_k$ for some $l\in\mathcal{V}(b_k)$. Conversely, if $N_j(B_k^j)=0$ $\forall j\in\mathcal{U}_0$ and $N(B_k)\neq 0$, then there must be a point in $\Pi_l\cap\tilde{B}_k$ followed by either a point in $\Pi_0\cap B_k$ or a point in $\Pi_{l'}\cap B_k$ for some $l,l'\in\mathcal{V}(b_k)$. Hence

$$\{N(B_k)=0\}\triangle(\cap_{j\in\mathcal{U}_0}\{N_j(B_k^j)=0\})\subset\left\{N_0(B_k)+\sum_{l\in\mathcal{V}(b_k)}N_l(\tilde{B}_k)\ge 2\right\}. \tag{3.64}$$

Note that by the definition of a caterpillar in Definition 3.4.12, for each $j\in\mathcal{U}$, $h(p_j,\Pi_j^{t_j})\ge(\log n)^{-c}$. It follows that $\mathcal{V}(b_k)\subseteq\bigcup_{m=0}^{\lceil b_k(\log n)^c\rceil}\{1,2\}^m$. Also if $J\subset\mathcal{U}_0$ with $|J|=K$ then $\sum_{j\in J}N_j(\tilde{B}_k)$ has a Poisson distribution with parameter $K\nu(\tilde{B}_k)$, and since $b_k-a_k\le\delta$, $\nu(\tilde{B}_k)\le n^2\pi(d+3\mathcal{R}_n)^2\mu((0,\mathcal{R}])\delta$. Hence for Z' a Poisson random variable with parameter $(2^{2+b_k(\log n)^c}+1)\nu(\tilde{B}_k)=\mathcal{O}(\delta)$,

$$\mathbb{P}\left[N_0(B_k)+\sum_{j\in\mathcal{V}(b_k)}N_j(\tilde{B}_k)\ge 2\middle|(\Pi_j(a_k))_{j\in\mathcal{U}_0}\right]\le\mathbb{P}\left[Z'\ge 2\right]=\mathcal{O}(\delta^2).$$

By (3.64), we now have that

$$\begin{aligned}\mathbb{P}[N(B_k)=0|(\Pi_j(a_k))_{j\in\mathcal{U}_0}] &= \mathbb{P}[\cap_{j\in\mathcal{U}_0}\{N_j(B_k^j)=0\}]+\mathcal{O}(\delta^2)\\ &= \prod_{j\in\mathcal{U}_0}\exp(-\nu(B_k^j))+\mathcal{O}(\delta^2)\\ &= \exp(-\nu(B_k))+\mathcal{O}(\delta^2).\end{aligned}$$

Substituting this into (3.63) and then repeating the same argument for $k-1, k-2, \ldots, 1$,

$$\begin{aligned}\mathbb{P}[N(B)=0] &= \prod_{i=1}^{k}\exp(-\nu(B_k))+\sum_{i=1}^{k}\mathcal{O}(\delta^2)\\ &= \exp(-\nu(B))+k\mathcal{O}(\delta^2).\end{aligned}$$

By partitioning B further, we can let $\delta\to 0$ with $k=\Theta(1/\delta)$. It follows that $\mathbb{P}[N(B)=0]=\exp(-\nu(B))$. By (3.62), this completes the proof. □

It follows immediately from Lemma 3.4.20 that the collection of potential ancestral lineages in $(\mathcal{P}_t(p,\Pi))_{t\geq 0}$ has the same distribution as $\mathcal{P}^{(n)}(p)$, the rescaled SLFVS dual. We now show that under this coupling the rescaled SLFVS dual and branching caterpillar coincide with high probability.

We consider $(\mathcal{C}_t(p,(\Pi_j)_{j\in\mathcal{U}}))_{0\leq t\leq T}$ as a collection of paths as follows. The set of paths through a single caterpillar $(c_t(p,\Pi))_{t\leq h(p,\Pi)}$ with $k^*(p,\Pi)=k^*$ is given by $\{l^i\}_{i\in\{1,2\}^{k^*}}$, where $l^i(t)=c_t^1(p,\Pi)$ for $t\in[0,(\log n)^{-c}]$ and for each $1\leq k\leq k^*$, $l^i(t)=c_t^{i_k}(p,\Pi)$ for $t\in[\tau_{k-1}^{\mathrm{br}}+(\log n)^{-c},(\tau_k^{\mathrm{br}}+(\log n)^{-c})\wedge h(p,\Pi)]$. Then the collection of paths through $(\mathcal{C}_t(p,(\Pi_j)_{j\in\mathcal{U}}))_{0\leq t\leq T}$ is given by concatenating paths through the individual caterpillars, i.e. paths $l:[0,T]\to\mathbb{R}^2$ such that for some sequence $(u_m)_{m\geq 0}\subset\mathcal{U}$ with $u_{m+1}=(u_m,i_m)$ for some $i_m\in\{1,2\}$ for each m, for $t\in[t_{u_m},h_{u_m}]$, $l(t)$ follows a path through $(c_{t-t_{u_m}}(p_{u_m},\Pi_{u_m}^{t_{u_m}}))_t$ with $l(h_{u_m})=p_{u_{m+1}}$.

Lemma 3.4.22. *Fix $T>0$. Let $(\Pi_j)_{j\in\mathcal{U}_0}$ be independent Poisson point processes with intensity measure* (3.40) *and let Π be defined from $(\Pi_j)_{j\in\mathcal{U}_0}$ as in* (3.61)*. Then $(\mathcal{C}_t(p,(\Pi_j)_{j\in\mathcal{U}}))_{0\leq t\leq T}$ and $(\mathcal{P}_t(p,\Pi))_{0\leq t\leq T}$, viewed as collections of paths, are equal with probability at least $1-\mathcal{O}((\log n)^{-1/4})$.*

Proof. We shall use Lemma 3.4.19 with $a = (16 \log 2)^{-1}$. Writing, for $j \in \mathcal{U}$, $k^*(j) = k^*(p_j, \Pi_j^{t_j})$, the number of branching events in $c_{t-t_j}(p_j, \Pi_j^{t_j})$ before h_j, by a union bound over $\mathcal{U}_{\lfloor a \log\log n \rfloor}$ and (3.45),

$$\begin{aligned}\mathbb{P}[\exists j \in \mathcal{U}_{\lfloor a \log\log n \rfloor} : k^*(j) \geq (\log n)^{9/8}] &\leq 2^{2+a \log\log n} \mathcal{O}(e^{-\delta(\log n)^{1/8}}) \\ &= \mathcal{O}(e^{-\delta(\log n)^{1/8}/2}). \qquad (3.65)\end{aligned}$$

Let $(\tau_k^{\text{br}}(j))_{k \geq 1}$ denote the sequence of branching events in $c_{t-t_j}(p_j, \Pi_j^{t_j})$, and similarly define $(\tau_k^{\text{type}}(j))_{k \geq 1}$ and $(\tau_k^{\text{over}}(j))_{k \geq 1}$ as in (3.41). Note that $(\mathcal{C}_t)_{t \leq T}$ and $(\mathcal{P}_t)_{t \leq T}$ only differ as collections of paths if either a selective event affects a caterpillar during a time interval in which it ignores branching, or if two different caterpillars are simultaneously within $\mathcal{R}_n$ of some $x \in \mathbb{R}^2$ and so one of them is not driven by the pieced together Poisson point process Π. More formally, if $(\mathcal{C}_t)_{t \leq T}$ and $(\mathcal{P}_t)_{t \leq T}$ differ as collections of paths then one or more of the following events occurs.

1. $\mathcal{U}(T) \not\subseteq \mathcal{U}_{\lfloor a \log\log n \rfloor}$ or $k^*(j) \geq (\log n)^{9/8}$ for some $j \in \mathcal{U}_{\lfloor a \log\log n \rfloor}$.
2. For some $j \in \mathcal{U}_{\lfloor a \log\log n \rfloor}$ and $k \leq (\log n)^{9/8}$, the event $E_1(j, k)$ occurs: one of the lineages $c^1_{t-t_j}(p_j, \Pi_j^{t_j})$ and $c^2_{t-t_j}(p_j, \Pi_j^{t_j})$ is affected by a selective event in the time interval $[\tau_k^{\text{br}}(j), \tau_k^{\text{br}}(j) + (\log n)^{-c}]$.
3. For some $w \neq v \in \mathcal{U}_{\lfloor a \log\log n \rfloor}$, the event $E_2(v, w)$ occurs: there are $i_1, i_2 \in \{1, 2\}$ with $|c^{i_1}_{t-t_w}(p_w, \Pi_w^{t_w}) - c^{i_2}_{t-t_v}(p_v, \Pi_v^{t_v})| \leq 2\mathcal{R}_n$ for some $t \leq T$.

Recall from (3.42) and (3.5) that selective events affect a single lineage with rate $\lambda \log n$. Hence for $k \in \mathbb{N}$ and $j \in \mathcal{U}$, $\mathbb{P}[E_1(j, k)] = \mathcal{O}((\log n)^{1-c})$.

We now consider the event $E_2(v, w)$. For $w \neq v \in \mathcal{U}$, let $i = \min\{j \geq 1 : w_j \neq v_j\}$. Then let

$$w \wedge v = \begin{cases} (w_1, \ldots, w_{i-1}) & \text{if } i \geq 2 \\ \emptyset & \text{if } i = 1. \end{cases}$$

At time $h_{w \wedge v}$, either $\tau^{\text{type}}_{k^*(w\wedge v)}(w \wedge v) = \tau^{\text{over}}_{k^*(w\wedge v)}(w \wedge v)$ or $\tau^{\text{type}}_{k^*(w\wedge v)}(w \wedge v) = \tau^{\text{div}}_{k^*(w\wedge v)}(w \wedge v)$, in which case $|p_{(w\wedge v,1)} - p_{(w\wedge v,2)}| \geq (\log n)^{-c}$. Conditional on $|p_{(w\wedge v,1)} - p_{(w\wedge v,2)}| \geq$

$(\log n)^{-c}$, for i_1, $i_2 \in \{1,2\}$,

$$\left(c^{i_1}_{t-t_w}(p_w, \Pi^{t_w}_w), c^{i_2}_{t-t_v}(p_v, \Pi^{t_v}_v)\right)_{t\in[t_w,h_w]\cap[t_v,h_v]\cap[0,T]}$$

is part of the pair of potential ancestral lineages of an excursion started at time $h_{w\wedge v}$ with initial displacement at least $(\log n)^{-c}$. Hence by Lemmas 3.4.9 and 3.4.15,

$$\mathbb{P}[E_2(w,v)] = \mathcal{O}\left(\frac{\log\log n}{\log n}\right) + \mathcal{O}\left((\log n)^{\frac{21}{8}-c}\right) = \mathcal{O}\left((\log n)^{-3/8}\right)$$

since $c \geq 3$. By a union bound, and using Lemma 3.4.19 and (3.65) it follows that

$$\begin{aligned}
&\mathbb{P}\left[(\mathcal{C}_t)_{t\leq T} \neq (P_t)_{t\leq T}\right] \\
&\quad\leq o((\log n)^{-1}) + 4(\log n)^{a\log 2+9/8}\mathbb{P}[E_1(j,k)] + 16(\log n)^{2a\log 2}\mathbb{P}[E_2(w,v)] \\
&\quad= \mathcal{O}\left((\log n)^{a\log 2+\frac{9}{8}+1-c}\right) + \mathcal{O}\left((\log n)^{2a\log 2-3/8}\right) \\
&\quad= \mathcal{O}\left((\log n)^{-1/4}\right),
\end{aligned}$$

by our choice of $a = (16\log 2)^{-1}$ and since $c \geq 3$. □

We are now ready to complete the proof of Theorem 3.2.7.

Proof. (Of Theorem 3.2.7) Set $c = 4$. By Lemmas 3.4.20 and 3.4.22, we have a coupling of the rescaled SLFVS dual and the branching caterpillar under which the two processes are equal (as collections of paths) with probability at least $1 - \mathcal{O}((\log n)^{-1/4})$.

We now couple $(\mathcal{C}_t(p, (\Pi_j)_{j\in\mathcal{U}}))_{0\leq t\leq T}$ to a branching Brownian motion with branching rate $\lambda\kappa_n$. Let $((W^j_t)_{t\geq 0}, H_j)_{j\in\mathcal{U}}$ be an i.i.d. sequence, where $(W^j_t)_{t\geq 0}$ is a Brownian motion starting at 0 and $H_j \sim \mathrm{Exp}(\lambda\kappa_n)$ independent of $(W^j_t)_{t\geq 0}$. For each $j \in \mathcal{U}$, we couple $(c_{t-t_j}(p_j, \Pi^{t_j}_j))_{t\in[t_j,h_j]}$ to $((W^j_t)_{t\geq 0}, H_j)$ as in Lemmas 3.4.14 and 3.4.16.

For $j \in \mathcal{U}$, let $A_1(j)$ be the event that both $|(h_j - t_j) - H_j| \leq 3(\log n)^{-1/4}$ and for $i = 1, 2$ and $t \in [t_j, h_j]$,

$$\left|(c^i_{t-t_j}(p_j, \Pi^{t_j}_j) - p_j) - W^j_{\sigma^2(t-t_j)}\right| \leq 2(\log n)^{\frac{9}{8}-\frac{c}{3}} = 2(\log n)^{-5/24}.$$

By Lemmas 3.4.14 and 3.4.16, for any $r>0$, for each $j \in \mathcal{U}$, $\mathbb{P}[A_1(j)] \geq 1-\mathcal{O}\left((\log n)^{-r}\right)$. Hence, taking a union bound over $j \in \mathcal{U}_{\lfloor \log\log n \rfloor}$,

$$\mathbb{P}[\cap_{j\in\mathcal{U}_{\lfloor \log\log n\rfloor}} A_1(j)] \geq 1-\mathcal{O}((\log n)^{\log 2 - r}).$$

Also, let $((\tilde{W}^j_t)_{t\geq 0})_{j\in\mathcal{U}}$ be an i.i.d. sequence of Brownian motions and for $j \in \mathcal{U}$, define the event

$$\begin{aligned} A_2(j) = \Bigg\{ & \sup_{t\in[0,3(\log\log n)(\log n)^{-1/4}]} |\tilde{W}^j_{\sigma^2 t}| \\ & \qquad + \sup_{|t-H_j|\leq 3(\log\log n)(\log n)^{-1/4}} |W^j_{\sigma^2 t} - W^j_{\sigma^2 H_j}| \leq (\log n)^{-1/9} \Bigg\}. \end{aligned}$$

Then by another union bound over $\mathcal{U}_{\lfloor \log\log n\rfloor}$, since for a Brownian motion $(W_t)_{t\geq 0}$ started at 0, $\mathbb{P}\left[\sup_{t\in[0,6(\log\log n)(\log n)^{-1/4}]} |W_t| \geq \frac{1}{2}(\log n)^{-1/9}\right] = o((\log n)^{-r})$, we have that

$$\mathbb{P}[\cap_{j\in\mathcal{U}_{\lfloor \log\log n\rfloor}} A_2(j)] \geq 1-\mathcal{O}((\log n)^{\log 2 - r}).$$

By Lemma 3.4.19, $\mathbb{P}[\mathcal{U}(T) \not\subseteq \mathcal{U}_{\lfloor \log\log n\rfloor}] = o((\log n)^{-r})$.

Define a branching Brownian motion starting at p with diffusion constant σ^2 from $((W^j_t)_{t\geq 0}, (\tilde{W}^j_t)_{t\geq 0}, H_j)_{j\in\mathcal{U}}$ by letting the increments of the initial particle be given by $(W^{\emptyset}_{\sigma^2 t})_{t\geq 0}$ until time $H_\emptyset$, when it is replaced by two particles which have lifetimes H_1 and H_2 and increments given by $(\tilde{W}^1_{\sigma^2 t})_{t\geq 0}$ and $(\tilde{W}^2_{\sigma^2 t})_{t\geq 0}$ on $[H_\emptyset, H_\emptyset \vee h_\emptyset]$ and by $(W^1_{\sigma^2(t+(H_\emptyset\vee h_\emptyset)-h_\emptyset)})_{t\geq 0}$ and $(W^2_{\sigma^2(t+(H_\emptyset\vee h_\emptyset)-h_\emptyset)})_{t\geq 0}$ on $[H_\emptyset \vee h_\emptyset, H_\emptyset + H_1]$ and $[H_\emptyset \vee h_\emptyset, H_\emptyset + H_2]$ respectively and so on.

If $\mathcal{U}(T) \subseteq \mathcal{U}_{\lfloor \log\log n\rfloor}$ and $A_1(j)\cap A_2(j)$ occurs for each $j \in \mathcal{U}_{\lfloor \log\log n\rfloor}$, each path in the branching caterpillar stays within distance $2(\log\log n + 1)(\log n)^{-1/9} + 2(\log\log n + 1)(\log n)^{-5/24}$ of some path through the branching Brownian motion and vice versa.

Setting $r = \log 2 + 1/4$ gives us a coupling between the branching caterpillar and branching Brownian motion (with diffusion constant σ^2 and branching rate $\kappa_n\lambda$) such that with probability at least $1-\mathcal{O}((\log n)^{-1/4})$, up to time T each path in the rescaled SLFVS dual stays within distance $2(\log\log n)(\log n)^{-1/9} + 2(\log\log n)(\log n)^{-5/24}$ of some path through the branching Brownian motion and vice versa. Finally, we need to couple this branching Brownian motion up to time T with a branching Brownian motion with branching rate $\kappa\lambda$. By (3.43), $\kappa_n \to \kappa$ as $n \to \infty$, so this follows by straightforward bounds on the difference between the branching times and the increments of a Brownian motion during such a time. $\square$

Chapter 4

Branching Brownian motion, mean curvature flow and the motion of hybrid zones

Abstract

We provide a probabilistic proof of a well known connection between a special case of the Allen-Cahn equation and mean curvature flow. We then prove a corresponding result for scaling limits of the spatial Λ-Fleming-Viot process with selection, in which the selection mechanism is chosen to model what are known in population genetics as *hybrid zones*. Our proofs will exploit a duality with a system of branching (and coalescing) random walkers which is of some interest in its own right.

4.1 Introduction

Our central result, Theorem 4.1.8 in Section 4.1.3, is the convergence, after suitable rescaling, of a stochastic analogue of the Allen-Cahn equation to the indicator function of a region whose boundary evolves according to mean curvature flow. The main motivation for this work comes from mathematical population genetics; specifically, we are interested in the behaviour of so-called hybrid zones. These occur when genetically distinct groups of individuals meet and mate, leaving behind at least some offspring of mixed ancestry. A textbook example is the common house mouse in Denmark (Hunt and Selander, 1973) which exists in the form *Mus musculus* in the North and *M. domesticus* in the South, but hybrid zones are ubiquitous in nature, for example, Barton and Hewitt (1989) cite 170 examples. Two principal explanations have been offered for the genetic variation observed in such zones. The first is that they arise in response to spatially varying natural selection; the second is that they are formed through secondary contact of two populations that were previously genetically isolated. Whereas in the first scenario the location of the hybrid zone is determined by an environment, which is usually taken to be fixed, in the second scenario, the hybrid zone can evolve with time. It is this second scenario that interests us here.

It is usual to suppose that the underlying genetics is controlled by a single gene which occurs in two types (alleles), traditionally denoted a and A. Individuals carry two copies of the gene and while those of types aa and AA (the *homozygotes*) are equally fit, the *heterozygotes* (that is individuals of type aA) are less likely to successfully reproduce. In an infinitely dense population, provided the selection against heterozygotes is weak, when viewed over large spatial and temporal scales, the proportion of a-alleles in the population at location x at time t is modelled by the solution to

$$\frac{\partial v}{\partial t} = \Delta v + \boldsymbol{s} v(1-v)(2v-1), \tag{4.1}$$

for an appropriate initial condition, where $\boldsymbol{s} > 0$ is a scaled selection coefficient. This is a special case of the Allen-Cahn equation; we explain the origin of this particular form of nonlinearity in Section 4.1.2.

Our interest is in the behaviour of the region in which both alleles are present in substantial numbers. Because heterozygotes are less fit than homozygotes, we expect this to be a narrow band which, when viewed on large enough scales, will look like a sharp interface. More formally, we apply a diffusive scaling to (4.1) in which $t \mapsto \varepsilon^2 t$ and $x \mapsto \varepsilon x$. The Laplacian term is, of course, invariant, but the term corresponding to selection is multiplied by a factor $1/\varepsilon^2$. It is well known that for suitable initial conditions, in a sense that we make precise in Theorem 4.1.3, as $\varepsilon \downarrow 0$, the solution to the scaled equation converges to the indicator function of a set whose boundary evolves according to mean curvature flow. Thus, in the biologically relevant case of two dimensions, if we observe the population over sufficiently large spatial and temporal scales, the interface between the two populations will evolve approximately as curvature flow or *curve-shortening flow* as it is often known.

One reason for the importance of curvature flow in applications stems from an underlying variational principle: curve shortening flow decreases the length of the curve at the fastest rate possible relative to the total speed of motion (measured in the sense of the square integral of the speed of motion of points around the curve), see e.g. White (2002) for a simple explanation. In this sense, if our populations evolved deterministically, then they would minimise the boundary between them as quickly as possible. In reality this will be somewhat offset by the randomness due to reproduction, known as random genetic drift, in a population which is not infinitely dense. Indeed if genetic drift is too strong, then we can expect the random noise to obscure the nonlinear term: this is suggested by the results of Hairer et al. (2012), who consider the equation

$$dw = (\Delta w + w - w^3)dt + \sigma dW,$$

in two dimensions, where W is a mollified space-time white noise. (By considering $(1 + w)/2$, up to constants, we recover a stochastic version of (4.1).) If the mollifier is removed, then the solutions converge weakly to zero, whereas if the intensity of the noise simultaneously converges to zero sufficiently quickly, then they recover the deterministic equation. The basic question that we set out to answer is “Will hybrid zones still evolve approximately according to curvature flow in the presence of random genetic drift?”

Of course, genetic drift is not appropriately modelled by a mollified space-time white noise and so, in order to investigate this question, we must first define a model that combines selection against heterozygosity with random genetic drift. Our starting point will be the spatial Λ-Fleming-Viot process which was introduced in Barton et al. (2010); Etheridge (2008) and has been studied in a series of papers since; see e.g. Barton et al. (2013c) for a review. The advantage of this model is that it allows us to incorporate genetic drift into models of populations evolving in spatial continua, with no restriction on spatial dimension. However, since our proofs are based on a duality with a branching and coalescing random walk, we expect analogous results if we start, for example, from the classical stepping stone model in which the population is subdivided into 'islands' that sit at the vertices of $\mathbb{Z}^{\mathrm{d}}$. In what follows, we shall refer to the spatial Λ-Fleming-Viot process with selection against heterozygosity as the SLFVS. It is described carefully in Definition 4.1.5. A version of this model with selection in favour of one genetic type was constructed in Etheridge et al. (2014). There it was shown that when suitably rescaled, in two or more dimensions, the allele frequencies converge to a solution of the Fisher-KPP equation,

$$\frac{\partial v}{\partial t} = \Delta v + \boldsymbol{s} v(1-v). \tag{4.2}$$

Mimicking that result, one can obtain (4.1) as a scaling limit of the SLFVS. Combined with the known convergence of the scaled version of (4.1), this certainly suggests that there should be scalings of the SLFVS which lead to mean curvature flow. However, available proofs of Theorem 4.1.3 could not readily be adapted to our stochastic setting and so we were forced to seek an alternative approach. Our first result is therefore a new proof of Theorem 4.1.3. We then adapt this to prove convergence of the proportions of different genetic types under the SLFVS to the indicator function of a set whose boundary evolves according to mean curvature flow. The key to our proof is a probabilistic representation of solutions to (4.1) which we believe to be of interest in its own right.

Before defining the SLFVS, we recall some purely deterministic results. Although our primary interest is in two spatial dimensions, there will be no additional arguments required if we work in $\mathbb{R}^{\mathrm{d}}$ for arbitrary $\mathrm{d} > 1$.

4.1.1 The Allen-Cahn equation and mean curvature flow

The Allen-Cahn equation (Allen and Cahn, 1979) takes the form

$$\frac{\partial v^{\varepsilon}}{\partial t} = \Delta v^{\varepsilon} - \frac{1}{\varepsilon^2} f(v^{\varepsilon}), \tag{4.3}$$

where f is the derivative of a potential function F which has exactly two local minima, at v_- and v_+, say. More precisely, we insist that $f \in C^2(\mathbb{R})$ has exactly three zeros,

$v_- < v_0 < v_+$, and

$$\begin{aligned} &f(v) < 0, && \forall v \in (-\infty, v_-) \cup (v_0, v_+); \\ &f(v) > 0, && \forall v \in (v_-, v_0) \cup (v_+, \infty); \\ &f'(v_-) > 0, && f'(v_+) > 0, \quad f'(v_0) < 0. \end{aligned} \tag{4.4}$$

Although originally introduced as a model for the macroscopic motion of phase boundaries driven by surface tension, the Allen-Cahn equation has found application in many other areas. It represents a balance between two opposing tendencies: the diffusive effect of the Laplacian attempts to smooth the solution, while the potential term drives it towards the states v_- and v_+. As a result, a narrow interface between these two states develops.

Allen and Cahn observed that if the two potential wells do not have equal depth, then on the timescale $s = t/\varepsilon$, the interface will propagate at a constant speed (proportional to $F(v_-) - F(v_+)$) along its normal, towards the domain of the deeper well. On the other hand, if the potential wells have equal depth, then the interface is almost stationary on this timescale, but if we observe it over the longer timescales of (4.3), it will propagate with normal velocity equal to the mean curvature of the interface.

There is now a huge literature that makes the observation of Allen and Cahn rigorous under various regularity conditions, for example Bronsard and Kohn (1991); Evans et al. (1992); Ilmanen (1993); Sato (2008). The principal obstruction to be overcome relates to the fact that the mean curvature flow is only well-defined under some regularity conditions and, even then, only up to a finite time horizon when it either shrinks to a point or, in dimensions three and higher, develops other singularities.

Before stating a result, let us make the definition of mean curvature flow precise. We begin with the special case of two dimensions. This is the relevant dimension for our biological application and requires much less explanation. In that setting, mean curvature is just curvature and the corresponding flow is often called curve-shortening.

Recall that a function is said to be a smooth embedding if it is a diffeomorphism onto its image (which we shall implicitly assume is a subset of $\mathbb{R}^2$).

Definition 4.1.1 (Curve-shortening flow)**.** *Let* S^1 *denote the unit circle in* $\mathbb{R}^2$*. Let* $\mathbf{\Gamma} = (\mathbf{\Gamma}_t(\cdot))_t$ *be a family of smooth embeddings, indexed by* $t \in [0, \mathscr{T})$*, where for each* t*,* $\mathbf{\Gamma}_t : S^1 \to \mathbb{R}^2$*. Let* $\boldsymbol{n} = \boldsymbol{n}_t(\phi)$ *denote the unit (inward) normal vector to* $\mathbf{\Gamma}_t$ *at* ϕ *and let* $\kappa = \kappa_t(\phi)$ *denote the curvature of* $\mathbf{\Gamma}_t$ *at* ϕ*. We say that* $\mathbf{\Gamma}$ *is a* curvature flow *or* curve-shortening flow *if*

$$\frac{\partial \mathbf{\Gamma}_t(\phi)}{\partial t} = \kappa_t(\phi)\boldsymbol{n}_t(\phi). \tag{4.5}$$

for all t, ϕ*.*

Assuming that $\mathbf{\Gamma}_0$ is a smooth embedding of S^1 into $\mathbb{R}^2$, the behaviour of $\mathbf{\Gamma}_t$ under curve-shortening is completely understood. First, it has a finite lifetime which we shall denote by $\mathscr{T}$. In Gage and Hamilton (1986), it was shown that if $\mathbf{\Gamma}_0$ is convex, then so is $\mathbf{\Gamma}_t$ for all $t < \mathscr{T}$. Moreover, $\mathscr{T}$ can be chosen so that $\mathbf{\Gamma}_t$ shrinks towards a point as $t \uparrow \mathscr{T}$; in this limit the asymptotic 'shape' of $\mathbf{\Gamma}_t$ is a circle. Soon afterwards, Grayson (1987) showed that, in fact, under curve-shortening, any smoothly embedded closed curve becomes convex at a time $\tau < \mathscr{T}$, after which the results of Gage and Hamilton apply.

In higher dimensions we must replace the curvature by the *mean curvature.* Recall that to define this quantity for a $(\mathbb{d} - 1)$-dimensional hypersurface in $\mathbb{R}^{\mathbb{d}}$, we take an orthonormal basis of the tangent space and form the matrix of the second fundamental form, that is the matrix whose (i, j)th entry is the dot product of the unit normal to the hypersurface with the derivative of the ith vector in the basis in the direction of the jth. The $\mathbb{d} - 1$ principal curvatures, $\kappa_1, \ldots, \kappa_{\mathbb{d}-1}$, are the eigenvalues of the matrix and their sum, that is the trace of the matrix, is the (scalar) mean curvature. The product of the scalar mean curvature with the unit normal is called the mean curvature vector (which does not depend on the choice of normal, since reversing the direction of the normal also changes the sign of the scalar mean curvature).

Definition 4.1.2 (Mean curvature flow)**.** Mean curvature flow*, when it is defined, is obtained by replacing the curvature* κ_t *in equation (4.5) by the mean curvature.*

The behaviour of mean curvature flow in $\mathbb{d} \geq 3$ is more complex than that of curve-shortening. It was proved by Huisken (1984) that the analogue of the Gage-Hamilton

Theorem holds, that is a $(\mathbb{d}-1)$–dimensional compact convex surface must shrink to a point and its asymptotic shape is a sphere. However, the analogue of Grayson's Theorem is false. In higher dimensions singularities can develop before the enclosed volume vanishes. Since our main interest is in two dimensions, we shall not discuss this here. Instead we shall follow Chen (1992) in imposing sufficiently strong initial conditions that the solution exists for a positive time and stopping before we encounter any singularities, and we refer to Mantegazza (2011) for a detailed discussion.

Suppose that $\mathbb{d} \geq 2$. Our first result concerns the convergence as $\varepsilon \downarrow 0$, for suitable initial conditions, of the solution of

$$\frac{\partial v^{\varepsilon}}{\partial t} = \Delta v^{\varepsilon} + \frac{1}{\varepsilon^2} v^{\varepsilon}(1-v^{\varepsilon})(2v^{\varepsilon}-1), \qquad v^{\varepsilon}(0,x) = p(x), \tag{4.6}$$

to the indicator function of a set whose boundary evolves according to mean curvature flow.

The initial condition, p, of (4.6) is assumed to take values in $[0,1]$. We shall also require that it satisfies some regularity conditions. In particular, set

$$\Gamma = \left\{ x \in \mathbb{R}^{\mathbb{d}} : p(x) = \frac{1}{2} \right\}.$$

We suppose that Γ is a smooth hypersurface which is also the boundary of a bounded open set which is topologically equivalent to the sphere. We impose the following regularity conditions:

($\mathscr{C}1$) Γ is C^{α} for some $\alpha > 3$.

($\mathscr{C}2$) For x inside Γ, $p(x) < \frac{1}{2}$. For x outside Γ, $p(x) > \frac{1}{2}$.

($\mathscr{C}3$) There exist $r, \gamma > 0$ such that, for all $x \in \mathbb{R}^{\mathbb{d}}$, $|p(x) - \frac{1}{2}| \geq \gamma \left(\mathrm{dist}(x,\Gamma) \wedge r\right)$.

In particular, we can think of Γ as the image of the boundary of the unit sphere under a map f for which $|f(x) - f(y)| = \mathcal{O}(|x-y|^{\alpha})$. Condition ($\mathscr{C}3$) prevents the slope of p near the interface Γ from being too shallow, and keeps $p(x)$ bounded away from $\frac{1}{2}$ when x is not near the interface. Condition ($\mathscr{C}2$) is simply establishing a sign convention.

Under these conditions, mean curvature flow started from Γ, which we denote $(\mathbf{\Gamma}_t(\cdot))_t$, exists up to some finite time $\mathscr{T}$ (e.g. Evans and Spruck (1991)).

To give a precise statement of the result, we require some more notation. Let $d(x,t)$ be the signed distance from x to $\mathbf{\Gamma}_t$, chosen to be negative inside $\mathbf{\Gamma}_t$ and positive outside. Note that, as sets,

$$\mathbf{\Gamma}_t = \{x \in \mathbb{R}^{\mathbb{d}} : d(x,t) = 0\}.$$

Theorem 4.1.3. *Let v^ε solve (4.6) with initial condition p satisfying the conditions ($\mathscr{C}1$)-($\mathscr{C}3$), and define $\mathscr{T}$, $d(x,t)$ as above. Fix $T^* \in (0, \mathscr{T})$. Let $k \in \mathbb{N}$. There exists $\varepsilon_{\mathbb{d}}(k) > 0$, and $a_{\mathbb{d}}(k), c_{\mathbb{d}}(k) \in (0,\infty)$ such that for all $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$ and t satisfying $a_{\mathbb{d}}\varepsilon^2 |\log \varepsilon| \leq t \leq T^*$,*

1. *for x such that $d(x,t) \geq c_{\mathbb{d}}\varepsilon|\log\varepsilon|$, we have $v^\varepsilon(t,x) \geq 1 - \varepsilon^k$;*
2. *for x such that $d(x,t) \leq -c_{\mathbb{d}}\varepsilon|\log\varepsilon|$, we have $v^\varepsilon(t,x) \leq \varepsilon^k$.*

This result is not new; it is a special case of Theorem 3 of Chen (1992). Indeed, our proof will display the same key steps: first we show that an interface develops; second we show that this interface propagates according to (mean) curvature flow. To achieve the second step, we couple the distance between a $\mathbb{d}$-dimensional Brownian motion and the interface Γ_s with a one-dimensional Brownian motion. This parallels the approximation of the solution to the Allen-Cahn equation by a one-dimensional standing wave in the proof of Chen (1992) (although we remark that we achieve our coupling through a different perturbation of the potential than that used by Chen (1992)). Both steps of our proof use probabilistic arguments, exploiting a duality between solutions to (4.6) and a branching Brownian motion, which is of some interest in its own right.

4.1.2 Modelling hybrid zones

Let us now turn to our model of hybrid zones. Our starting point is the spatial Λ-Fleming-Viot process with selection. The model we consider here is a modification of that introduced for genic selection (selection in favour of just one of the alleles) in Etheridge et al. (2014), and existence of the process follows by the same arguments. Also as for genic selection, uniqueness follows from duality with a system of branching and coalescing particles, although there is a slight twist in the form that duality takes (see Section 4.3.1), mirroring our probabilistic representation of solutions to (4.1).

We suppose that there are two alleles, a and A. At each time t, the random function $\{w_t(x),\, x \in \mathbb{R}^{\mathrm{d}}\}$ is defined, up to a Lebesgue null set of $\mathbb{R}^{\mathrm{d}}$, by

$$w_t(x) := \text{ proportion of type } a \text{ at spatial position } x \text{ at time } t. \tag{4.7}$$

In other words, if we sample an allele from the point x at time t, the probability that it is of type a is $w_t(x)$.

Remark 4.1.4. *It is convenient to extend the definition of* $w_t(x)$ *to all of* $\mathbb{R}^{\mathrm{d}}$ *and so, on the Lebesgue null set on which (4.7) is not sufficient to specify* $w_t(x)$*, we shall arbitrarily impose* $w_t(x) = 0$*.*

A construction of an appropriate state space for $x \mapsto w_t(x)$ can be found in Véber and Wakolbinger (2015). Using the identification

$$\int_{\mathbb{R}^d} \{w(x)f(x,a) + (1-w(x))f(x,A)\}\, dx = \int_{\mathbb{R}^d \times \{a,A\}} f(x,\kappa) M(dx, d\kappa),$$

this state space is in one-to-one correspondence with the space $\mathcal{M}_\lambda$ of measures on $\mathbb{R}^{\mathrm{d}} \times \{a, A\}$ with 'spatial marginal' Lebesgue measure, which we endow with the topology of vague convergence. By a slight abuse of notation, we also denote the state space of the process $(w_t)_{t\in\mathbb{R}}$ by $\mathcal{M}_\lambda$.

Definition 4.1.5 (Spatial Λ-Fleming-Viot with selection against heterozygosity (SLFVS))**.** *Fix* $u \in (0, 1]$ *and* $\mathcal{R} \in (0, \infty)$*. Let* μ *be a finite measure on* $(0, \mathcal{R}]$*. Further, let* Π *be a Poisson point process on* $\mathbb{R}_+ \times \mathbb{R}^{\mathrm{d}} \times (0, \mathcal{R}]$ *with intensity measure*

$$dt \otimes dx \otimes \mu(dr). \tag{4.8}$$

The spatial Λ-Fleming-Viot process with selection *(SLFVS) driven by* Π *is the* $\mathcal{M}_\lambda$*-valued process* $(w_t)_{t\geq 0}$ *with dynamics given as follows.*

If $(t, x, r) \in \Pi$*, a reproduction event occurs at time* t *within the closed ball* $\mathcal{B}_r(x)$ *of radius* r *centred on* x*. With probability* $1 - \boldsymbol{s}$ *the event is* neutral*, in which case:*

1. *Choose a parental location z uniformly at random within $\mathcal{B}_r(x)$, and a parental type, α_0, according to $w_{t-}(z)$, that is $\alpha_0 = a$ with probability $w_{t-}(z)$ and $\alpha_0 = A$ with probability $1 - w_{t-}(z)$.*

2. *For every $y \in \mathcal{B}_r(x)$, set $w_t(y) = (1-u)w_{t-}(y) + u\mathbf{1}_{\{\alpha_0 = a\}}$.*

Otherwise the event is selective*, in which case:*

1. *Choose three 'potential' parental locations z_1, z_2, z_3 independently and uniformly at random within $\mathcal{B}_r(x)$, and at each of these sites 'potential' parental types α_1, α_2, α_3 according to $w_{t-}(z_1), w_{t-}(z_2), w_{t-}(z_3)$ respectively. Let $\widehat{\alpha}$ denote the most common allelic type in $\alpha_1, \alpha_2, \alpha_3$.*

2. *For every $y \in \mathcal{B}_r(x)$ set $w_t(y) = (1-u)w_{t-}(y) + u\mathbf{1}_{\{\widehat{\alpha} = a\}}$.*

Remark 4.1.6. *More generally, the parameter u, which we shall refer to as the* impact*, can be taken to be random. In this case, for each $r \in (0, \mathcal{R}]$, we let ν_r be a probability measure on $(0,1]$ and the driving noise, Π, is taken to be a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}^{\mathrm{d}} \times (0, \mathcal{R}] \times (0,1]$ with intensity measure*

$$dt \otimes dx \otimes \mu(dr)\nu_r(du).$$

For each point $(t, x, r, u) \in \Pi$, the corresponding reproduction event is described exactly as before.

Since $\boldsymbol{s}$ is assumed small, as one expects in a model of genetic drift, to first order the variance of the increment of the mean allele frequency in the region affected by an event is $u^2\bar{w}(1-\bar{w})$, where $\bar{w}$ is the mean of w_{t-} over the affected region. Let us try to motivate the form of the selection mechanism, which is what drives the expectation of the increments in allele frequencies. As is usual in population genetics, we have approximated a model of selection acting on a diploid population (in which each individual carries two copies of the gene) by one in which we think of selection acting on single copies of the gene, but in a way that depends on the local frequencies of the different alleles. This sort of approximation, which goes back at least to Fisher (1937), is valid when the local population size is large, corresponding in our case to the impact u being small. (In fact we are interested in limits in which the impact will tend to zero.) The idea is simple. Each individual in the population carries two copies of the gene. This subdivides the population into *homozygotes*, carrying either aa or AA and assumed equally fit, and *heterozygotes* carrying aA and assumed to have relative fitness $1 - \boldsymbol{s}$. The population is assumed to be in Hardy-Weinberg proportions, so that if the proportion of a-alleles in the parental population is $\bar{w}$, then the proportions of parents that are of type aa, aA and AA are $\bar{w}^2$, $2\bar{w}(1-\bar{w})$ and $(1-\bar{w})^2$, respectively. During reproduction, each individual produces a very large number of germ cells (cells of the same genotype). To reflect the relative fitnesses, a heterozygote produces $(1 - \boldsymbol{s})$ times as many germ cells as a homozygote. Germ cells then split into an effectively infinite pool of gametes (cells containing just one chromosome from each pair) which fuse at random to form diploid offspring. Suppose that the proportion of type a alleles in the affected region immediately before reproduction is $\bar{w}$. Then the probability that a gamete sampled from the pool is of type a is

$$\begin{aligned}\frac{\bar{w}^2 + \bar{w}(1-\bar{w})(1-\boldsymbol{s})}{1 - 2\boldsymbol{s}\bar{w}(1-\bar{w})} &= (1-\boldsymbol{s})\bar{w} + \boldsymbol{s}(3\bar{w}^2 - 2\bar{w}^3) + \mathcal{O}(\boldsymbol{s}^2) \\ &= (1-\boldsymbol{s})\bar{w} + \boldsymbol{s}(\bar{w}^3 + 3\bar{w}^2(1-\bar{w})) + \mathcal{O}(\boldsymbol{s}^2). \qquad (4.9)\end{aligned}$$

Notice that the first term in (4.9) is $1 - \boldsymbol{s}$ times the probability that an allele sampled from the parental population is of type a whereas the second is $\boldsymbol{s}$ times the probability that the majority of three alleles sampled independently from the parental population are of type a. This then motivates the two types of event in our SLFVS. In particular, if we replace a proportion u of the population by offspring, then the expected increment in $\bar{w}$ is

$$u\boldsymbol{s}(\bar{w}^3 + 3\bar{w}^2(1-\bar{w}) - \bar{w}) = u\boldsymbol{s}\bar{w}(1-\bar{w})(2\bar{w}-1),$$

which underpins the connection to (4.1).

Of course, in replacing a diploid model by one based directly on allele frequencies, we have rather muddied the notion of parent in our reproduction mechanism, so the use of the term in Definition 4.1.5 should not be interpreted too literally.

4.1.3 Convergence of the hybrid zone to mean curvature flow

To understand our main result, first we state a simple modification of a result on a rescaling of the SLFVS from Etheridge et al. (2014). To state that result, we specialise to $\mu(dr) = \delta_R(dr)$, for some fixed $R > 0$. At the nth stage of the rescaling, the impact and selection parameters are assumed to satisfy

$$u_n = \frac{u}{n^{1-2\beta}}, \qquad \text{and} \qquad \boldsymbol{s}_n = \frac{\rho}{n^{2\beta}}.$$

Next, we define the averaged process,

$$w_t^n(x) := w_{nt}(n^\beta x), \qquad \text{and} \qquad \bar{w}_t^n(x) := \frac{n^{\beta \mathrm{d}}}{V_R} \int_{B(x, n^{-\beta}R)} w_t^n(y)\, dy,$$

where V_R is the volume of the ball of radius R in $\mathbb{R}^{\mathrm{d}}$. To simplify notation, we write $\mathcal{M}$ for $\mathcal{M}_\lambda(\mathbb{R}^{\mathrm{d}} \times \{a, A\})$, and $D_{\mathcal{M}}[0,\infty)$ for the set of all càdlàg paths with values in $\mathcal{M}$. We also write $C_c^\infty(\mathbb{R}^{\mathrm{d}})$ for the set of smooth compactly supported functions on $\mathbb{R}^{\mathrm{d}}$.

Theorem 4.1.7. *[Modification of Theorem 1.3 of Etheridge et al. (2014)] Suppose that* $\beta \in (0, 1/3)$*, and that* $\bar{w}_0^n$ *converges weakly to some* $w^0 \in \mathcal{M}$*. Then, as* $n \to \infty$*, the process* $(\bar{w}_t^n)_{t \geq 0}$ *converges weakly in* $D_{\mathcal{M}}[0,\infty)$ *towards a process* $(w_t^\infty)_{t\geq 0}$ *with initial value* $w_0^\infty = w^0$*. Furthermore,* $(w_t^\infty)_{t \geq 0}$ *is the unique deterministic process for which, for every* $f \in C_c^\infty(\mathbb{R}^{\mathrm{d}})$*,*

$$\langle w_t^\infty, f\rangle = \langle w_0^\infty, f\rangle + \int_0^t \left\{ \frac{\kappa_R}{2} \langle w_s^\infty, \Delta f\rangle + u\rho V_R \langle w_s^\infty(1 - w_s^\infty)(2w_s^\infty - 1), f\rangle \right\} ds,$$

where

$$\kappa_R = \frac{u}{V_R} \int_{B(0,R)} \int_{B(x,R)} (z_1)^2 dz\, dx \tag{4.10}$$

with z_1 *the first coordinate of the vector* $z \in \mathbb{R}^{\mathrm{d}}$*. In particular,* κ_R *depends only on* R *and* d*.*

In other words, up to a change of coefficients, $(w_t^\infty)_{t \geq 0}$ is a weak solution of (4.6) with $w_0 = w^0$. Based on Theorem 4.1.7, it is natural to ask whether we can modify the scaling of $\boldsymbol{s}_n$ in such a way that $\boldsymbol{s}_n n^{2\beta} \to \infty$ as $n \to \infty$ and obtain convergence to the indicator function of a region whose boundary evolves according to mean curvature flow.

In other words, does genetic drift, which is driven by the neutral events in the SLFVS, disrupt that convergence?

To state our result, we first rescale the SLFVS as in Theorem 4.1.7. For each $n \in \mathbb{N}$, we define the finite measure μ^n on $(0, \mathcal{R}_n]$, where $\mathcal{R}_n = n^{-\beta}\mathcal{R}$, by $\mu^n(A) = \mu(n^\beta A)$ for all Borel subsets A of $(0, \infty)$. Our rescaled SLFVS will be driven by the Poisson point process Π^n on $\mathbb{R}_+ \times \mathbb{R}^{\mathrm{d}} \times (0, \infty)$ with intensity measure

$$n dt \otimes n^\beta dx \otimes \mu^n(dr). \tag{4.11}$$

Here $n^\beta dx$ denotes the scaling in which the linear dimension of the infinitesimal region dx is scaled by n^β (so that when we integrate, the volume of a region is scaled by $n^{\mathrm{d}\beta}$). Let

$$u_n = \frac{u}{n^{1-2\beta}}, \qquad \text{and} \qquad \boldsymbol{s}_n = \frac{1}{\varepsilon_n^2} \frac{1}{n^{2\beta}}. \tag{4.12}$$

It is convenient to define the constant σ^2 through

$$\sigma^2 = \frac{u}{2\mathrm{d}} \int_0^{\mathcal{R}} \int_{\mathbb{R}^{\mathrm{d}}} |z|^2 \frac{V_r(0,z)}{V_r} dz \mu(dr). \tag{4.13}$$

If $\mu(dr) = \delta_R(r)$, then we recover κ_R from (4.10).

Theorem 4.1.8. *Suppose that $\beta \in (0, 1/4)$ and let ε_n be a sequence such that $\varepsilon_n \to 0$ and $(\log n)^{1/2}\varepsilon_n \to \infty$ as $n \to \infty$. Let $(w_t^n)_{t\geq 0}$ be the SLFVS driven by Π^n and with u_n, $\boldsymbol{s}_n$ given by (4.12), and initial condition $w_0^n(x) = p(x)$. Assume that p satisfies ($\mathscr{C}1$)-($\mathscr{C}3$), and define $\mathscr{T}$, $d(x,t)$ as for Theorem 4.1.3; take $T^* < \mathscr{T}$. For $k \in \mathbb{N}$ there exist $n_*(k) < \infty$, and $a_*(k), c_*(k) \in (0,\infty)$ such that for all $n \geq n_*$ and all t satisfying $a_*\varepsilon_n^2|\log \varepsilon_n| \leq t \leq T^*$,*

1. *for almost every x such that $d(x, \sigma^2 t) \geq c_*\varepsilon_n|\log\varepsilon_n|$, we have $\mathbb{E}[w_t^n(x)] \geq 1 - \varepsilon_n^k$;*
2. *for almost every x such that $d(x, \sigma^2 t) \leq -c_*\varepsilon_n|\log\varepsilon_n|$, we have $\mathbb{E}[w_t^n(x)] \leq \varepsilon_n^k$.*

Remark 4.1.9. *In Section 4.3.1 we explain the origins of these scalings. By taking u_n to be small, we are assuming that local population density is high.*

By adapting ideas from Etheridge et al. (2015a), we expect an analogous result for values of u_n up to $\mathcal{O}(1)$, but at the expense of having to take $\varepsilon_n \to 0$ extremely slowly *(so that* $\varepsilon_n^{-1} = o(\log\log n)$*). The stronger the genetic drift, that is the bigger* u_n*, the larger the value of* n *required for the diffusive rescaling to smooth the allele frequencies under the SLFVS sufficiently for the behaviour to be close to that of the differential equation* (4.6)*.*

The rest of the chapter is laid out as follows. In Section 4.2 we establish a duality between equation (4.1) and a branching Brownian motion which we then use to prove Theorem 4.1.3. In Section 4.3 we establish an analogous duality between the SLFVS and a system of branching and coalescing particles and use it to establish Theorem 4.1.8.

4.2 Proof of Theorem 4.1.3

4.2.1 A probabilistic dual to Equation (4.6)

Our proof of Theorem 4.1.3 rests on a duality between equation (4.6) and a branching Brownian motion in which each individual, independently, follows a Brownian motion during an exponentially distributed lifetime (with mean ε^2) at the end of which it splits into *three*. Although reminiscent of the duality between the Fisher-KPP equation and binary branching Brownian motion pioneered by Skorohod (1964) and McKean (1975), here there is a slight twist. These papers allow us to deal with equations of the form

$$\frac{\partial v}{\partial t} = \frac{1}{2}\Delta v + Vf(v),$$

where V is a constant (the branching rate in the branching Brownian motion) and f is of the form $f(v) = \Phi(v) - v$ where $\Phi(v)$ is the probability generating function of a non-negative integer-valued random variable (the number of offspring of each individual in the branching Brownian motion). However, the expression for f in (4.6) is not of this form. Instead we adapt ideas from population genetics (notably from Krone and Neuhauser (1997); Neuhauser and Krone (1997)).

First, to maintain compatibility with the PDE literature, we shall adopt the convention that

$$\textit{all Brownian motions run at rate } 2. \tag{4.14}$$

That is, at time 1, Brownian motion has variance 2.

In contrast to the McKean-Skorohod setting, our representation of the solution to (4.1) is not just in terms of the spatial positions of individuals in the branching Brownian motion at a fixed time, but also depends on their genealogy. In other words, we have a duality between (4.1) and the *historical process* of the branching Brownian motion.

To write this formally, we require some notation for our ternary branching Brownian motion. We write $\boldsymbol{W}(t)$ for the historical process (which traces out the space-time trees that record the spatial position of all individuals alive at time s for all $s \in [0,t]$). This process can be constructed formally as the ternary branching Markov process in which the position of an 'individual' alive at time s is taken to be the whole Brownian path $(W_u)_{0\leq u\leq s}$ followed by its ancestors. To record the genealogy of the process we use Ulam-Harris notation to label individuals in the branching Brownian motion by elements of $\mathcal{U} = \bigcup_{m=0}^{\infty}\{1,2,3\}^m$. For example, $(3,1,2)$ is the particle which is the 2^{nd} child of the 1^{st} child of the 3^{rd} child of the initial ancestor $\emptyset$. Let $N(t) \subset \mathcal{U}$ denote the set of individuals alive at time t. We shall abuse notation slightly and write $(W_i(t))_{i\in N(t)}$ for the spatial locations of the individuals alive at time t, and $(W_i(s))_{0\leq s\leq t}$ for the unique path that connects leaf i to the root.

We say that $\mathcal{T}$ is a *time-labelled ternary tree* if $\mathcal{T}$ is a finite subtree of $\mathcal{U}$ and each internal vertex v of the tree is labelled with a time $t_v > 0$, where t_v is strictly greater than the label of the parent vertex of v. Evidently if we ignore the spatial position of individuals, each realisation of $\boldsymbol{W}(t)$ traces out a time-labelled ternary tree which records the genealogy and associates a time to each branching event. We shall use $\mathcal{T}(\boldsymbol{W}(t))$ to denote this time-labelled ternary tree.

For a fixed function $p : \mathbb{R}^d \to [0,1]$, we define a voting procedure on $\mathcal{T}(\boldsymbol{W}(t))$ as follows.

1. Each leaf i of $\mathcal{T}(\boldsymbol{W}(t))$, independently, votes 1 with probability $p(W_i(t))$ and otherwise votes 0.

2. At each branch point in $\mathcal{T}(\boldsymbol{W}(t))$, the vote of the parent particle j is the majority vote of the votes of its three children $(j,1)$, $(j,2)$ and $(j,3)$.

This defines an iterative voting procedure, which runs inwards from the leaves of $\mathcal{T}(\boldsymbol{W}(t))$ to the root $\emptyset$.

Definition 4.2.1 ($\mathbb{V}_p$)**.** *With the voting procedure described above, we define* $\mathbb{V}_p(\boldsymbol{W}(t))$ *to be the vote associated to the root* $\emptyset$*.*

For $x \in \mathbb{R}^{\mathrm{d}}$, we write $\mathbb{P}_x^\varepsilon$ for the probability measure under which $(\boldsymbol{W}(t), t \geq 0)$ has the law of the historical process of ternary branching Brownian motion in $\mathbb{R}^{\mathrm{d}}$ with branching rate $1/\varepsilon^2$ started from a single particle at location x at time 0. We write $\mathbb{E}_x^\varepsilon$ for the corresponding expectation.

Theorem 4.2.2. *Let* $p : \mathbb{R}^{\mathrm{d}} \to [0, 1]$*. Then*

$$v^\varepsilon(t, x) = \mathbb{P}_x^\varepsilon \left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1\right] \tag{4.15}$$

is a solution to equation (4.6) *with initial condition* $v^\varepsilon(0, x) = p(x)$*.*

Proof. (Sketch)

The proof mirrors that of the representation of solutions of the Fisher-KPP equation in terms of binary branching Brownian motion, and so we only sketch it. As usual the idea is to analyse the expression on the right hand side of (4.15) by partitioning on the behaviour of the branching Brownian motion in the first δt of time and then to take a limit as $\delta t \downarrow 0$.

Throughout the proof we neglect the superscript ε in $\mathbb{P}_x^\varepsilon$, $\mathbb{E}_x^\varepsilon$ and v^ε and the subscript p in $\mathbb{V}_p$. We write S for the time of the first branching event in the branching Brownian motion and W_S for the position of the ancestor at that time. It is convenient to use E for expectation when it is with respect to the law of Brownian motion $(W_\cdot)$, preserving $\mathbb{E}$ for expectation with respect to that of the historical branching Brownian motion $(\boldsymbol{W}(\cdot))$. Let V_1, V_2, V_3 denote the votes of the three offspring created at time S. By the strong Markov property of the branching Brownian motion, and the branching property, we see that the V_i are conditionally independent given (S, W_S). Moreover, since conditional on $S \leq \delta t$, the chance of a second branch before time δt is $\mathcal{O}(\delta t)$, for $s \leq \delta t$,

$$\mathbb{E}_x[V_1 | (S, W_S) = (s, y)] = E_y[v(t, W_{\delta t - s})] + \mathcal{O}(\delta t).$$

From this, if we assume enough regularity of $v(t,x)$ (which follows from that of the heat semigroup),

$$\mathbb{E}_x[V_1|S \leq \delta t] = v(t,x) + \mathcal{O}(\delta t). \tag{4.16}$$

Still conditioning on $S \leq \delta t$, in order for the vote at the root to be one, at most one of V_1, V_2, V_3 can be zero, and so using (4.16) and conditional independence of the V_i given (S, W_S),

$$\mathbb{P}_x\left[\mathbb{V}(\boldsymbol{W}(t+\delta t)) = 1|S \leq \delta t\right] = v(t,x)^3 + 3v(t,x)^2(1-v(t,x)) + \mathcal{O}(\delta t).$$

Since if $S > \delta t$ the ancestor of the branching Brownian motion simply follows a Brownian motion over $[0,\delta t]$, partitioning over the behaviour of the branching Brownian motion in the first δt of time gives

$$\begin{aligned} v(t+\delta t, x) &= \mathbb{P}_x\left[\mathbb{V}(\boldsymbol{W}(t+\delta t)) = 1|S \leq \delta t\right]\mathbb{P}\left[S \leq \delta t\right] \\ &\quad + \mathbb{P}_x\left[\mathbb{V}(\boldsymbol{W}(t+\delta t)) = 1 \,|\, S > \delta t\right]\left(1 - \mathbb{P}\left[S \leq \delta t\right]\right) \\ &= \mathbb{P}_x\left[\mathbb{V}(\boldsymbol{W}(t+\delta t)) = 1 \,|\, S \leq \delta t\right]\mathbb{P}\left[S \leq \delta t\right] \\ &\quad + E_x\left[\mathbb{P}_{W_{\delta t}}\left[\mathbb{V}(\boldsymbol{W}(t)) = 1\right]\right]\left(1 - \mathbb{P}\left[S \leq \delta t\right]\right). \end{aligned}$$

Now $\mathbb{P}[S \leq \delta t] = \varepsilon^{-2}\delta t + \mathcal{O}(\delta t^2)$ and so substituting and rearranging (and once again assuming enough regularity of $v(t,x)$) we obtain

$$\begin{aligned} \lim_{\delta t \to 0} \frac{v(t+\delta t, x) - v(t,x)}{\delta t} &= \varepsilon^{-2}\Big(v(t,x)^3 + 3v(t,x)^2(1-v(t,x)) - v(t,x)\Big) \\ &\quad + \lim_{\delta t \to 0} \frac{E_x\left[\mathbb{P}_{W_{\delta t}}\left[\mathbb{V}(\boldsymbol{W}(t)) = 1\right]\right] - v(t,x)}{\delta t} \\ &= \varepsilon^{-2}\Big(v(t,x)^3 + 3v(t,x)^2(1-v(t,x)) - v(t,x)\Big) \\ &\quad + \lim_{\delta t \to 0} \frac{E_x\left[v(t, W_{\delta t})\right] - v(t,x)}{\delta t} \\ &= \Delta v(t,x) + \varepsilon^{-2}v(t,x)(1-v(t,x))(2v(t,x)-1), \end{aligned}$$

as required. □

Armed with this representation, the proof of Theorem 4.1.3 is reduced to proving the following result about our branching Brownian motions.

Theorem 4.2.3. *Suppose* $p : \mathbb{R}^{\mathbb{d}} \to [0,1]$ *is such that* ($\mathscr{C}1$)-($\mathscr{C}3$) *hold. Define* $\mathscr{T}$, $d(x,t)$ *as for Theorem 4.1.3; fix* $T^* \in (0, \mathscr{T})$ *and let* $k \in \mathbb{N}$. *There exist* $\varepsilon_{\mathbb{d}}(k) > 0$, *and* $a_{\mathbb{d}}(k), c_{\mathbb{d}}(k) \in (0, \infty)$ *such that for all* $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$ *and* t *satisfying* $a_{\mathbb{d}}\varepsilon^2|\log \varepsilon| \leq t \leq T^*$,

1. *for* x *such that* $d(x,t) \geq c_{\mathbb{d}}\varepsilon|\log\varepsilon|$, *we have* $\mathbb{P}^{\varepsilon}_x\left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1\right] \geq 1 - \varepsilon^k$;
2. *for* x *such that* $d(x,t) \leq -c_{\mathbb{d}}\varepsilon|\log\varepsilon|$, *we have* $\mathbb{P}^{\varepsilon}_x\left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1\right] \leq \varepsilon^k$.

The proof of Theorem 4.2.3 will proceed in two steps. First, in Section 4.2.2, we prove a one-dimensional analogue of the result in the special case in which $p(x) = \mathbb{1}\{x \geq 0\}$. The proof rests on symmetry of branching Brownian motion and the monotonicity that results from the specific choice of initial condition p. The second step uses the definition of mean curvature flow and the regularity properties that follow from the conditions ($\mathscr{C}1$)-($\mathscr{C}3$). These allow us to couple the distance between the (backwards in time) mean curvature flow $(\boldsymbol{\Gamma}_{t-s})_{s\in[0,t]}$ and a (forwards in time) $\mathbb{d}$-dimensional Brownian motion W with a (forwards in time) one-dimensional Brownian motion B in such a way that $d(W_s, t-s)$ is well approximated by B_s when W_s is close to $\boldsymbol{\Gamma}_{t-s}$. This coupling is made precise in Proposition 4.2.13 in Section 4.2.3. The proof of Theorem 4.2.3, which combines these two steps by bounding the errors that occur far from the interface $\boldsymbol{\Gamma}_{t-s}$, can be found in Section 4.2.4.

Notation 4.2.4. *It is convenient to have a prominent distinction between one dimensional and multi-dimensional Brownian motion in our notation. We therefore adopt the convention that* B *will denote one dimensional Brownian motion and* $\boldsymbol{B}$ *will represent the corresponding historical branching Brownian motion and we preserve* W *and* $\boldsymbol{W}$ *for dimensions* $\mathbb{d} \geq 2$.

4.2.2 Majority voting in one dimensional BBM

In this section we consider only ternary branching Brownian motion in dimension $\mathbb{d} = 1$.

As in Section 4.2.1, for $x \in \mathbb{R}$, we write $\mathbb{P}^{\varepsilon}_x$ for the probability measure under which $(\boldsymbol{B}(t), t \geq 0)$ has the law of historical ternary branching Brownian motion in $\mathbb{R}$ with branching rate $1/\varepsilon^2$ started from a single particle at location x at time 0, and $\mathbb{E}^{\varepsilon}_x$ for the corresponding expectation. We also write P_x for the probability measure under which

$(B_t)_{t\geq 0}$ has the law of a Brownian motion started at x, and E_x for the corresponding expectation.

Throughout this section we write $\mathbb{V} := \mathbb{V}_{p_0}$ where $p_0(x) = \mathbb{1}\{x \geq 0\}$, so that a leaf votes 1 if and only if it is in the right half line. Our aim is to prove the following one-dimensional analogue of Theorem 4.2.3 for this initial condition p_0.

Theorem 4.2.5. *Let $T^* \in (0, \infty)$. For all $k \in \mathbb{N}$ there exist $c_1(k)$ and $\varepsilon_1(k) > 0$ such that, for all $t \in [0, T^*]$ and all $\varepsilon \in (0, \varepsilon_1)$,*

1. *for $z \geq c_1(k)\varepsilon|\log \varepsilon|$, we have $\mathbb{P}^\varepsilon_z\left[\mathbb{V}(\boldsymbol{B}(t)) = 1\right] \geq 1 - \varepsilon^k$*
2. *for $z \leq -c_1(k)\varepsilon|\log \varepsilon|$, we have $\mathbb{P}^\varepsilon_z\left[\mathbb{V}(\boldsymbol{B}(t)) = 1\right] \leq \varepsilon^k$.*

Remark 4.2.6. *The subscript 1 on a_1, c_1 and ε_1 is to emphasize that Theorem 4.2.5 applies in dimension 1. We shall often suppress the dependence on k in our notation.*

Note that, if $z \geq 0$, then a typical leaf of the branching Brownian motion is more likely to vote 1 than 0, and that the opposite is true for $z < 0$. Theorem 4.2.5 says that the majority voting procedure magnifies a small voting bias at the leaves into a much stronger voting bias at the root. If the votes of different leaves were independent this would be elementary, but the spatial structure of the branching Brownian motion introduces strong correlations between votes of closely related individuals. To overcome this, we first use a symmetry argument to show that the bias close to the root will be at least as strong as that at the leaves and then check that, as ε tends to zero, there is enough branching close to the root to sufficiently magnify the bias.

4.2.2.1 Proof of Theorem 4.2.5

First note that with our special choice of initial condition p_0, for any $x_1 \leq x_2 \in \mathbb{R}$,

$$\mathbb{P}^\varepsilon_{x_1}[\mathbb{V}(\boldsymbol{B}(t)) = 1] \leq \mathbb{P}^\varepsilon_{x_2}[\mathbb{V}(\boldsymbol{B}(t)) = 1]. \tag{4.17}$$

By analogy with the previous subsection, we use $\mathcal{T}(\boldsymbol{B}(t))$ to denote the time-labelled tree traced out by the branching Brownian motion up to time t, and for any time-labelled ternary tree $\mathcal{T}$ we write

$$\mathbb{P}_x^t(\mathcal{T}) = \mathbb{P}_x^\varepsilon\left[\mathbb{V}(\boldsymbol{B}(t)) = 1 \,|\, \mathcal{T}(\boldsymbol{B}(t)) = \mathcal{T}\right]. \tag{4.18}$$

By the symmetry of the Brownian motions followed by individuals in $\boldsymbol{B}(t)$ conditional on $\{\mathcal{T}(\boldsymbol{B}(t)) = \mathcal{T}\}$, applying the reflection $x \mapsto -x$ to the process, we see that for any time-labelled ternary tree $\mathcal{T}$, any time $t > 0$, and any $z \in \mathbb{R}$,

$$\mathbb{P}_z^t(\mathcal{T}) = 1 - \mathbb{P}_{-z}^t(\mathcal{T}). \tag{4.19}$$

The monotonicity in (4.17) and the symmetry in (4.19) are key to our proof of Theorem 4.2.5.

Taking $z = 0$ in (4.19) shows that $\mathbb{P}_0^t(\mathcal{T}) = \frac{1}{2}$ for all $t > 0$, and, by (4.17), for all $t > 0$ and all time-labelled ternary trees $\mathcal{T}$ we have

$$\mathbb{P}_z^t(\mathcal{T}) \geq \tfrac{1}{2} \text{ for } z > 0; \qquad \mathbb{P}_z^t(\mathcal{T}) \leq \tfrac{1}{2} \text{ for } z < 0.$$

We now introduce notation for the majority voting procedure. Let $g : [0,1]^3 \to [0,1]$ be given by

$$g(p_1, p_2, p_3) = p_1 p_2 p_3 + p_1 p_2 (1 - p_3) + p_2 p_3 (1 - p_1) + p_3 p_1 (1 - p_2). \tag{4.20}$$

This is the probability that a majority vote gives the result 1, in the special case where the three voters are independent and have probabilities p_1, p_2 and p_3 respectively of voting 1. With a slight abuse of notation, we let $g(p) = g(p, p, p)$, for $p \in [0, 1]$. Note that

$$g(1 - p_1, 1 - p_2, 1 - p_3) = 1 - g(p_1, p_2, p_3). \tag{4.21}$$

For $\mathcal{T}$ a time-labelled ternary tree with at least one branching event, suppose that the time to the first branching event in $\mathcal{T}$ is τ and that the subtrees with time labels corresponding to the (descendants of the) three offspring from the branching event are $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ (here a vertex v with time label t_v in $\mathcal{T}$ is given time label $t_v - \tau$ in $\mathcal{T}_i$).

Then, we write

$$g\left(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star)\right) = g\left(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}_1), \mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}_2), \mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}_3)\right) \tag{4.22}$$

and the identity

$$\mathbb{P}^t_z(\mathcal{T}) = E_z\left[g\left(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star)\right)\right] \tag{4.23}$$

expresses the majority voting that takes place at the first branch of $\mathcal{T}$.

Our next lemma states that the majority voting procedure cannot reduce the voting bias. In view of symmetry (4.19), when it is convenient to do so we will only state such results for the case $z \geq 0$.

Lemma 4.2.7. *For any time-labelled ternary tree $\mathcal{T}$, any time $t > 0$, and any $z \geq 0$,*

$$\mathbb{P}^t_z(\mathcal{T}) \geq P_z[B_t \geq 0].$$

Proof. The proof is by induction on the number of branching events in the tree $\mathcal{T}$. Let $\mathcal{T}_0$ denote the tree with a root and a single leaf. Then, by definition, $\mathbb{P}^t_z(\mathcal{T}_0) = P_z\left[B_t \geq 0\right].$

We now approach the inductive step. Suppose that the statement of the lemma holds for all time-labelled ternary trees with up to n internal vertices. We define $h : [0,1]^3 \to \mathbb{R}$ by

$$h(p_1, p_2, p_3) = g(p_1, p_2, p_3) - \frac{1}{3}(p_1 + p_2 + p_3),$$

and note that from (4.21) we have

$$h(1-p_1, 1-p_2, 1-p_3) = -h(p_1, p_2, p_3). \tag{4.24}$$

We can write h in the form

$$h(p_1, p_2, p_3) = \tfrac{1}{3}\sum p_{i_1}\left((1-p_{i_2})(p_{i_3} - \tfrac{1}{2}) + (1-p_{i_3})(p_{i_2} - \tfrac{1}{2})\right)$$

where the sum is over $(i_1, i_2, i_3) = (1,2,3), (2,3,1), (3,1,2)$. Hence

$$\frac{1}{2} \leq p_1, p_2, p_3 \leq 1 \;\Rightarrow\; h(p_1, p_2, p_3) \geq 0. \tag{4.25}$$

We will use the $\star$ notation defined in (4.22) for h in the same way as we use it for g.

Suppose that $\mathcal{T}$ is a time-labelled ternary tree with $n+1$ internal vertices and let τ, $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$ be as in (4.23). Using (4.23), by the definition of g and h we have

$$\begin{aligned}\mathbb{P}^t_z(\mathcal{T}) &= E_z\left[g\left(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star)\right)\right]\\ &= E_z\left[h\left(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star)\right)\right] + \frac{1}{3}\sum_{i=1}^{3} E_z\left[\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}_i)\right]. \end{aligned} \tag{4.26}$$

We begin by considering the first term of (4.26). Combining (4.24) with (4.19),

$$h(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star)) = -h(\mathbb{P}^{t-\tau}_{-B_\tau}(\mathcal{T}\star)).$$

Hence,

$$\begin{aligned}&E_z[h(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star))]\\ &\quad= E_z\left[h(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star))\mathbb{1}\{B_\tau \ge 0\}\right] + E_z\left[h(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star))\mathbb{1}\{B_\tau < 0\}\right]\\ &\quad= E_z\left[h(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star))\mathbb{1}\{B_\tau \ge 0\}\right] - E_z\left[h(\mathbb{P}^{t-\tau}_{-B_\tau}(\mathcal{T}\star))\mathbb{1}\{B_\tau < 0\}\right]\\ &\quad= \int_0^\infty h(\mathbb{P}^{t-\tau}_x(\mathcal{T}\star))(\phi_{z,2\tau}(x) - \phi_{z,2\tau}(-x))\,dx, \end{aligned} \tag{4.27}$$

where ϕ_{μ,σ^2} denotes the density of a $N(\mu,\sigma^2)$ random variable. Since $\mathbb{P}^{t-\tau}_x(\mathcal{T}_i) \ge 1/2$ for $x \ge 0$, by (4.25) we have $h(\mathbb{P}^{t-\delta t}_x(\mathcal{T}\star)) \ge 0$, and since $z \ge 0$, for all $x \ge 0$ we have

$$\phi_{z,2\tau}(x) - \phi_{z,2\tau}(-x) \ge 0,$$

which proves that (4.27) is non-negative. Using our inductive hypothesis, for $i = 1,2,3$,

$$E_z[\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}_i)] \ge E_z\left[P_{B_\tau}\left[B_{t-\tau} \ge 0\right]\right] = P_z[B_t \ge 0]$$

and so substituting into (4.26) completes the proof of Lemma 4.2.7. □

Our next task is to show that successive rounds of majority voting magnify a small bias at the leaves into a large bias at the root of a tree. Recall that for $p \in [0,1]$,

$$g(p) := g(p,p,p) = 3p^2 - 2p^3,$$

and define $g^{(n)}(p)$, inductively, by

$$g^{(1)}(p) = g(p), \qquad g^{(n+1)}(p) = g^{(n)}(g(p)).$$

Thus, $g^{(n)}(p)$ describes the probability of voting 1 at the root of an n-level regular ternary tree if the votes of the leaves are i.i.d. Bernoulli(p).

Lemma 4.2.8. *For all $k \in \mathbb{N}$ there exists $A(k) < \infty$ such that, for all $\varepsilon \in (0, \frac{1}{2}]$ and $n \geq A(k)|\log \varepsilon|$ we have*

$$g^{(n)}(\tfrac{1}{2} + \varepsilon) \geq 1 - \varepsilon^k.$$

Proof. We carry out two phases of iteration of g. First, we will show that it takes $\mathcal{O}(|\log \varepsilon|)$ iterations to obtain

$$g^{(n)}(\tfrac{1}{2} + \varepsilon) \geq \tfrac{1}{2} + \tfrac{1}{\sqrt{8}}. \tag{4.28}$$

Then we note that $\mathcal{O}(\log |k \log \varepsilon|)$ iterations are required to obtain

$$g^{(n)}(\tfrac{1}{2} + \tfrac{1}{\sqrt{8}}) \geq 1 - \varepsilon^k. \tag{4.29}$$

Since g is monotone, combining the two phases completes the proof.

For the first phase, if $\delta \in (0, 1/\sqrt{8})$ then a simple calculation shows that

$$g(\tfrac{1}{2} + \delta) = \tfrac{1}{2} + \tfrac{3}{2}\delta - 2\delta^3 \geq \tfrac{1}{2} + \tfrac{5}{4}\delta.$$

Thus if $g^{(n)}(\frac{1}{2} + \varepsilon) - \frac{1}{2} < 1/\sqrt{8}$, we have

$$g^{(n+1)}(\tfrac{1}{2} + \varepsilon) - \tfrac{1}{2} \geq \tfrac{5}{4}\left(g^{(n)}(\tfrac{1}{2} + \varepsilon) - \tfrac{1}{2}\right) \geq (\tfrac{5}{4})^n \varepsilon.$$

It follows immediately that $\mathcal{O}(|\log \varepsilon|)$ iterations are required to achieve (4.28).

For the second phase, note that $1 - g(1 - \delta) = 3\delta^2 - 2\delta^3 \leq 3\delta^2$, so that

$$1 - g^{(n+1)}(\tfrac{1}{2} + \tfrac{1}{\sqrt{8}}) \leq 3\left(1 - g^{(n)}(\tfrac{1}{2} + \tfrac{1}{\sqrt{8}})\right)^2 \leq \tfrac{1}{3}\Big(3(\tfrac{1}{2} - \tfrac{1}{\sqrt{8}})\Big)^{2^n}.$$

Noting that $3(\frac{1}{2} - \frac{1}{\sqrt{8}}) < 1$, it follows easily that the number of iterations required to obtain (4.29) is $\mathcal{O}(\log |k \log \varepsilon|)$. $\square$

We now want to see that there is a (large) regular ternary tree sitting inside $\mathcal{T}(\boldsymbol{B}(t))$. Let $\mathcal{T}_n^{reg} = \cup_{k \le n} \{1,2,3\}^k \subset \mathcal{U}$ denote the n-level regular ternary tree and, for $l \in \mathbb{R}$, let $\mathcal{T}_l^{reg} = \mathcal{T}_{\lceil l \rceil}^{reg}$. For $\mathcal{T}$ a time-labelled ternary tree, we use the relation $\mathcal{T} \supseteq \mathcal{T}_l^{reg}$ to mean that as subtrees of $\mathcal{U}$, $\mathcal{T}_l^{reg}$ is contained inside $\mathcal{T}$ (ignoring its time labels).

Lemma 4.2.9. *Let $k \in \mathbb{N}$ and let $A = A(k)$ be as in Lemma 4.2.8. Then there exist $a_1 = a_1(k)$ and $\varepsilon_1 = \varepsilon_1(k)$ such that, for all $\varepsilon \in (0, \varepsilon_1)$ and $t \ge a_1 \varepsilon^2 |\log \varepsilon|$,*

$$\mathbb{P}^\varepsilon \left[\mathcal{T}(\boldsymbol{B}(t)) \supseteq \mathcal{T}^{reg}_{A(k)|\log \varepsilon|} \right] \ge 1 - \varepsilon^k.$$

Proof. First we establish control over the tail distribution of the sum of n independent exponentially distributed (branching) times. Suppose $(X_j)_{j \ge 1}$ are i.i.d. Exp(1) random variables and let $S_n = \sum_{j=1}^n X_j$. Then

$$M_{X_1}(\lambda) = \mathbb{E}\left[e^{\lambda X_1}\right] = \begin{cases} \frac{1}{1-\lambda} & \text{if } \lambda < 1 \\ \infty & \text{if } \lambda \ge 1 \end{cases}$$

and for $a \ge 1$,

$$\Psi^*(a) := \sup_{\lambda \ge 0}(\lambda a - \log M_{X_1}(\lambda)) = \sup_{0 \le \lambda < 1} (\lambda a + \log(1-\lambda)) = a - 1 - \log a.$$

By Cramér's theorem, for $a \ge 1$,

$$\lim_{n \to \infty} \left(-\frac{1}{n} \log \mathbb{P}[S_n \ge na] \right) = \Psi^*(a) = a - 1 - \log a. \tag{4.30}$$

Suppose $a \ge 1$. For each leaf of $\mathcal{T}_l^{reg}$ we use (4.30) to estimate the probability that it is not in $\mathcal{T}(\boldsymbol{B}(t))$ and combine with a union bound (summing over leaves). For $t \ge a\varepsilon^2 \lceil A |\log \varepsilon| \rceil$ we have

$$\begin{aligned} &\mathbb{P}^\varepsilon \left[\mathcal{T}(\boldsymbol{B}(t)) \not\supseteq \mathcal{T}^{reg}_{A|\log \varepsilon|} \right] \\ &\qquad \le 3^{\lceil A|\log \varepsilon| \rceil} \mathbb{P}\left[\varepsilon^2 S_{\lceil A|\log\varepsilon|\rceil} \ge a\varepsilon^2 \lceil A|\log \varepsilon|\rceil \right] \end{aligned}$$

$$= \exp\left(\lceil A|\log\varepsilon|\rceil\left(\log 3 + \frac{1}{\lceil A|\log\varepsilon|\rceil}\log\mathbb{P}\left[S_{\lceil A|\log\varepsilon|\rceil} \geq a\lceil A|\log\varepsilon|\rceil\right]\right)\right). \tag{4.31}$$

By (4.30) (with $n = \lceil A|\log\varepsilon|\rceil$), we can choose $\varepsilon_1(k) < e^{-1}$ such that, for all $\varepsilon \in (0, \varepsilon_1)$,

$$\frac{1}{\lceil A|\log\varepsilon|\rceil}\log\mathbb{P}\left[S_{\lceil A|\log\varepsilon|\rceil} \geq a\lceil A|\log\varepsilon|\rceil\right] \leq -a + 3/2 + \log a.$$

Choose $a \geq 1$ sufficiently large that $-a + 3/2 + \log a \leq -\log 3 - k/A$. Putting this into (4.31) we obtain

$$\mathbb{P}^\varepsilon\left[\mathcal{T}(\boldsymbol{B}(t)) \not\supseteq \mathcal{T}^{reg}_{A|\log\varepsilon|}\right] \leq \exp\left(-|\log\varepsilon|k\right)$$

for $t \geq a\varepsilon^2\lceil A|\log\varepsilon|\rceil$. Letting $a_1 = a(A+1)$ completes the proof. ☐

We now control the maximal displacement of individuals in the ternary branching Brownian motion at small times. Let $N(t)$ denote the set of individuals alive in $\boldsymbol{B}(t)$.

Lemma 4.2.10. *Let $k \in \mathbb{N}$, and let $a_1(k)$ be as in Lemma 4.2.9. Then there exist $d_1(k)$, $\varepsilon_1(k)$ such that, for all $\varepsilon \in (0, \varepsilon_1(k))$ and all $s \leq a_1\varepsilon^2|\log\varepsilon|$,*

$$\mathbb{P}^\varepsilon_x\left[\exists i \in N(s) : |B_i(s) - x| \geq d_1(k)\varepsilon|\log\varepsilon|\right] \leq \varepsilon^k.$$

Proof. Write $\delta_1 = a_1\varepsilon^2|\log\varepsilon|$ and let Z be a $N(0,1)$ distributed random variable. By Markov's inequality, for $s \leq \delta_1$ we have

$$\begin{aligned}
\mathbb{P}^\varepsilon_x\left[\exists i \in N(s) : |B_i(s) - x| \geq d_1\varepsilon|\log\varepsilon|\right] &\leq \mathbb{E}^\varepsilon\left[|N(s)|\right]\mathbb{P}\left[\sqrt{2s}|Z| \geq d_1\varepsilon|\log\varepsilon|\right] \\
&\leq \mathbb{E}^\varepsilon\left[|N(\delta_1)|\right]\mathbb{P}\left[\sqrt{2\delta_1}|Z| \geq d_1\varepsilon|\log\varepsilon|\right] \\
&= e^{2\delta_1/\varepsilon^2}\mathbb{P}\left[\sqrt{2a_1}|Z| \geq d_1|\log\varepsilon|^{1/2}\right] \\
&\leq \frac{1}{\varepsilon^{2a_1}}\exp\left(-\tfrac{1}{4}\tfrac{d_1^2}{a_1}|\log\varepsilon|\right) \\
&= \varepsilon^{\frac{1}{4}\frac{d_1^2}{a_1} - 2a_1}.
\end{aligned}$$

Here the fourth line holds for $\varepsilon \in (0,1)$. The proof is completed by choosing $d_1 = d_1(k)$ large enough that $\frac{d_1^2}{4a_1} - 2a_1 \geq k$. ☐

We now have all the ingredients needed to prove Theorem 4.2.5. If $z \geq 2d_1\varepsilon|\log\varepsilon|$, then, at time $\delta_1 = a_1\varepsilon^2|\log\varepsilon|$, by Lemma 4.2.10, with high probability, all individuals in $\boldsymbol{B}(\delta_1)$ are still $\geq d_1\varepsilon|\log\varepsilon|$. Lemma 4.2.7 tells us that there is a positive voting bias at each of those points and Lemma 4.2.9 shows that this will be magnified by at least $\mathcal{O}(|\log\varepsilon|)$ rounds of majority voting as we trace back to the root. Finally, Lemma 4.2.8 gives us a lower bound on the bias at the root.

Proof of Theorem 4.2.5. We will prove the first statement of the theorem; the second then follows by symmetry.

For all $\varepsilon < 1/2$, define z_ε implicitly by the relation $\mathbb{P}\left[B_{T^*} \geq -z_\varepsilon\right] = \frac{1}{2} + \varepsilon$, and note that $z_\varepsilon \sim \varepsilon\sqrt{4\pi T^*}$ as $\varepsilon \to 0$. Let $\varepsilon_1(k) < 1/2$ be sufficiently small that Lemmas 4.2.9 and 4.2.10 hold for $\varepsilon \in (0, \varepsilon_1(k))$. Let $d_1(k)$ be given by Lemma 4.2.10 and let $c_1(k) = 2d_1(k)$ so that (by reducing ε_1 if necessary), for $\varepsilon \in (0, \varepsilon_1)$,

$$d_1(k)\varepsilon|\log\varepsilon| + z_\varepsilon \leq c_1(k)\varepsilon|\log\varepsilon|. \tag{4.32}$$

Let $a_1(k)$ be given by Lemma 4.2.9 and let

$$\delta_1 = \delta_1(k,\varepsilon) = a_1(k)\varepsilon^2|\log\varepsilon|. \tag{4.33}$$

If $t \in (0, \delta_1)$ and $z \geq c_1\varepsilon|\log\varepsilon|$, then

$$\begin{aligned}\mathbb{P}^\varepsilon_z\left[\mathbb{V}(\boldsymbol{B}(t)) = 0\right] &\leq \mathbb{P}^\varepsilon_z\left[\exists i \in N(t) \text{ such that } |B_i(t) - z| \geq d_1\varepsilon|\log\varepsilon|\right] \\ &\leq \varepsilon^k,\end{aligned}$$

where the second line follows by Lemma 4.2.10.

We now suppose that $t \in [\delta_1, T^*]$ and $z \geq c_1\varepsilon|\log\varepsilon|$. Let $\mathcal{T}_{\delta_1} = \mathcal{T}(\boldsymbol{B}(\delta_1))$ denote the time-labelled tree of the branching Brownian motion up to time δ_1. We define

$$p_{t-\delta_1}(z) = \mathbb{P}^\varepsilon_z\left[\mathbb{V}(\boldsymbol{B}(t-\delta_1)) = 1\right],$$

and

$$p^\varepsilon_{t-\delta_1}(z) = p_{t-\delta_1}(z_\varepsilon), \qquad \text{for all } z \in \mathbb{R}.$$

Finally, write $\{\boldsymbol{B}(\delta_1) > z_\varepsilon\}$ for the event $B_i(\delta_1) > z_\varepsilon$ for all $i \in N(\delta_1)$. Then,

$$\begin{aligned}\mathbb{P}_z^\varepsilon\left[\mathbb{V}(\boldsymbol{B}(t)) = 1\right] &= \mathbb{P}_z^\varepsilon\left[\mathbb{V}_{p_{t-\delta_1}(z)}(\boldsymbol{B}(\delta_1)) = 1\right] \\ &\geq \mathbb{P}_z^\varepsilon\left[\left\{\mathbb{V}_{p^\varepsilon_{t-\delta_1}(z)}(\boldsymbol{B}(\delta_1)) = 1\right\} \cap \{\boldsymbol{B}(\delta_1) > z_\varepsilon\}\right] \\ &\geq \mathbb{P}_z^\varepsilon\left[\mathbb{V}_{p^\varepsilon_{t-\delta_1}(z)}(\boldsymbol{B}(\delta_1)) = 1\right] - \varepsilon^k. \end{aligned} \tag{4.34}$$

Here, the first line follows by the Markov property of $\boldsymbol{B}$ at time δ_1. The second follows by the monotonicity property (4.17). The third line then follows by Lemma 4.2.10, using (4.32) and our hypothesis that $z \geq c_1\varepsilon|\log\varepsilon|$.

We have

$$p^\varepsilon_{t-\delta_1}(z) \geq P_{z_\varepsilon}\left[B_{t-\delta_1} \geq 0\right] \geq \tfrac{1}{2} + \varepsilon. \tag{4.35}$$

Here, the first inequality follows from Lemma 4.2.7. The second follows by the definition of z_ε, since $t - \delta_1 < T^*$.

If $p_i \geq 1/2$ for $i = 1, 2, 3$ then (4.25) implies that $g(p_1, p_2, p_3) \geq \min(p_1, p_2, p_3)$. Hence, if each leaf of $\mathcal{T}_{\delta_1}$ votes 1 independently with probability at least $\frac{1}{2} + \varepsilon$ and $\mathcal{T}_{\delta_1} \supseteq \mathcal{T}^{reg}_{A|\log\varepsilon|}$, then each of the leaves of $\mathcal{T}^{reg}_{A|\log\varepsilon|}$ votes 1 independently with probability at least $\frac{1}{2} + \varepsilon$. Therefore,

$$\mathbb{P}_z^\varepsilon\left[\mathbb{V}(\boldsymbol{B}(t)) = 1\right] \geq g^{(\lceil A|\log\varepsilon|\rceil)}(\tfrac{1}{2} + \varepsilon) - 2\varepsilon^k \geq 1 - 3\varepsilon^k.$$

Here, the first inequality follows by substituting (4.35) into (4.34) and then applying Lemma 4.2.9 and the second then follows by Lemma 4.2.8. This completes the proof. □

4.2.2.2 The slope of the interface

In proving Theorem 4.2.3 we shall also exploit a lower bound on the ‘slope’ of the interface in $\mathbb{d} = 1$ which we prove in this subsection. We obtain it as a corollary of the following result.

Proposition 4.2.11. *Suppose $x \geq 0$ and $\eta > 0$. Then for any time-labelled ternary tree $\mathcal{T}$ and any time t,*

$$\mathbb{P}^t_x(\mathcal{T}) - \mathbb{P}^t_{x-\eta}(\mathcal{T}) \geq \mathbb{P}^t_{x+\eta}(\mathcal{T}) - \mathbb{P}^t_x(\mathcal{T}).$$

Proof. The proof is by induction on the number of branching events in $\mathcal{T}$, and is similar to the proof of Lemma 4.2.7. For $\mathcal{T}_0$ a (time-labelled) tree with a root and a single leaf, we have

$$\mathbb{P}^t_x(\mathcal{T}_0) - \mathbb{P}^t_{x-\eta}(\mathcal{T}_0) = \int_{x-\eta}^{x} \phi_{0,2t}(u)\,du \geq \int_{x}^{x+\eta} \phi_{0,2t}(u)\,du = \mathbb{P}^t_{x+\eta}(\mathcal{T}_0) - \mathbb{P}^t_x(\mathcal{T}_0)$$

where ϕ_{μ,σ^2} is the density of a $\mathrm{N}(\mu, \sigma^2)$ random variable.

Now, assume that the lemma holds for all time-labelled ternary trees with at most n internal vertices. Let $\mathcal{T}$ be a time-labelled ternary tree with $n+1$ internal vertices and suppose that the time to the first branching event of $\mathcal{T}$ is τ and let $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$ denote the trees of the three offspring of that branching. Then using the notation of (4.22),

$$\begin{aligned}
&\Big(\mathbb{P}^t_x(\mathcal{T}) - \mathbb{P}^t_{x-\eta}(\mathcal{T})\Big) - \Big(\mathbb{P}^t_{x+\eta}(\mathcal{T}) - \mathbb{P}^t_x(\mathcal{T})\Big)\\
&= \Big(E_x\Big[g(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star))\Big] - E_{x-\eta}\Big[g(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star))\Big]\Big) - \Big(E_{x+\eta}\Big[g(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star))\Big] - E_x\Big[g(\mathbb{P}^{t-\tau}_{B_\tau}(\mathcal{T}\star))\Big]\Big)\\
&= \int_{-\infty}^{\infty} \Big\{\big(g(\mathbb{P}^{t-\tau}_y(\mathcal{T}\star)) - g(\mathbb{P}^{t-\tau}_{y-\eta}(\mathcal{T}\star))\big) - \big(g(\mathbb{P}^{t-\tau}_{y+\eta}(\mathcal{T}\star)) - g(\mathbb{P}^{t-\tau}_y(\mathcal{T}\star))\big)\Big\}\,\phi_{x,2\tau}(y)dy\\
&= \int_{0}^{\infty} \Big\{\big(g(\mathbb{P}^{t-\tau}_y(\mathcal{T}\star)) - g(\mathbb{P}^{t-\tau}_{y-\eta}(\mathcal{T}\star))\big) - \big(g(\mathbb{P}^{t-\tau}_{y+\eta}(\mathcal{T}\star)) - g(\mathbb{P}^{t-\tau}_y(\mathcal{T}\star))\big)\Big\}\,(\phi_{x,2\tau}(y) - \phi_{x,2\tau}(-y))\,dy.
\end{aligned} \tag{4.36}$$

Here, the second line follows by (4.23) and the last line follows from (4.21) and (4.19), which imply that $g(\mathbb{P}^t_w(\mathcal{T}\star)) = 1 - g(\mathbb{P}^t_{-w}(\mathcal{T}\star))$. Note the similarity to (4.27).

Since $x \geq 0$, we have

$$\phi_{x,2\tau}(y) - \phi_{x,2\tau}(-y) \geq 0 \tag{4.37}$$

for $y \geq 0$. In view of (4.36) we should like to check that for $y \geq 0$

$$\Big(g(\mathbb{P}^{t-\tau}_y(\mathcal{T}\star)) - g(\mathbb{P}^{t-\tau}_{y-\eta}(\mathcal{T}\star))\Big) - \Big(g(\mathbb{P}^{t-\tau}_{y+\eta}(\mathcal{T}\star)) - g(\mathbb{P}^{t-\tau}_y(\mathcal{T}\star))\Big) \geq 0. \tag{4.38}$$

By our inductive hypothesis, for $y \geq 0$ we have

$$(\mathbb{P}^{t-\tau}_y(\mathcal{T}_i) - \mathbb{P}^{t-\tau}_{y-\eta}(\mathcal{T}_i)) - (\mathbb{P}^{t-\tau}_{y+\eta}(\mathcal{T}_i) - \mathbb{P}^{t-\tau}_y(\mathcal{T}_i)) \geq 0,$$

and so by monotonicity of g, for (4.38) it is enough to check that

$$g(\mathbb{P}^{t-\tau}_{y+\eta}(\mathcal{T}\star)) - 2g(\mathbb{P}^{t-\tau}_{y}(\mathcal{T}\star)) + g\left(\mathbb{P}^{t-\tau}_{y}(\mathcal{T}\star) - (\mathbb{P}^{t-\tau}_{y+\eta}(\mathcal{T}\star) - \mathbb{P}^{t-\tau}_{y}(\mathcal{T}\star))\right) \leq 0. \tag{4.39}$$

To see that (4.39) holds, note that

$$\begin{aligned} g(p_1+\eta_1, p_2+\eta_2, p_3+\eta_3) - 2g(p_1,p_2,p_3) + g(p_1-\eta_1, p_2-\eta_2, p_3-\eta_3) \\ = 2\eta_1\eta_2(1-2p_3) + 2\eta_2\eta_3(1-2p_1) + 2\eta_3\eta_1(1-2p_2). \end{aligned}$$

and set $p_i = \mathbb{P}^{t-\tau}_{y}(\mathcal{T}_i)$ and $\eta_i = \mathbb{P}^{t-\tau}_{y+\eta}(\mathcal{T}_i) - \mathbb{P}^{t-\tau}_{y}(\mathcal{T}_i)$. Since for $y \geq 0$, $p_i \geq 1/2$, the inequality (4.39) then follows.

Putting (4.37) and (4.38) into (4.36) completes the inductive step, which in turn completes the proof. □

Corollary 4.2.12. *Take $\varepsilon_1(1)$ and $c_1(1)$ from Theorem 4.2.5. Let $\varepsilon < \min(\varepsilon_1(1), \frac{1}{24})$. Suppose that for some $t \in [0, T^*]$ and $z \in \mathbb{R}$,*

$$\left|\mathbb{P}^{\varepsilon}_{z}\left[\mathbb{V}(\boldsymbol{B}(t)) = 1\right] - \tfrac{1}{2}\right| \leq \tfrac{5}{12}, \tag{4.40}$$

and let $w \in \mathbb{R}$ with $|z-w| \leq c_1(1)\varepsilon|\log\varepsilon|$. Then

$$\left|\mathbb{P}^{\varepsilon}_{z}\left[\mathbb{V}(\boldsymbol{B}(t)) = 1\right] - \mathbb{P}^{\varepsilon}_{w}\left[\mathbb{V}(\boldsymbol{B}(t)) = 1\right]\right| \geq \frac{|z-w|}{48c_1(1)\varepsilon|\log\varepsilon|}. \tag{4.41}$$

Proof. Consider first the case $0 \leq z \leq w$. By analogy with (4.23), let $\mathbb{P}^t_y$ denote $\mathbb{P}^{\varepsilon}_{y}\left[\mathbb{V}(\boldsymbol{B}(t)) = 1\right]$. By Theorem 4.2.5 and (4.40) we have that

$$\mathbb{P}^t_{c_1(1)\varepsilon|\log\varepsilon|} - \mathbb{P}^t_z \geq 1 - \varepsilon - \tfrac{11}{12} \geq \tfrac{1}{24}. \tag{4.42}$$

Let $\eta := w - z$. For $j \in \mathbb{N}$, applying Proposition 4.2.11 j times gives that

$$\mathbb{P}^t_{(j+1)\eta+z} - \mathbb{P}^t_{j\eta+z} \leq \mathbb{P}^t_w - \mathbb{P}^t_z.$$

It follows that

$$\begin{aligned}\mathbb{P}^t_{c_1(1)\varepsilon|\log\varepsilon|} - \mathbb{P}^t_z &\leq \sum_{j=0}^{\lceil \eta^{-1}(c_1(1)\varepsilon|\log\varepsilon| - z)\rceil - 1} (\mathbb{P}^t_{(j+1)\eta+z} - \mathbb{P}^t_{j\eta+z}) \\ &\leq (\eta^{-1}(c_1(1)\varepsilon|\log\varepsilon|) + 1)(\mathbb{P}^t_w - \mathbb{P}^t_z). \end{aligned} \tag{4.43}$$

Combining (4.42) and (4.43),

$$\mathbb{P}^t_w - \mathbb{P}^t_z \geq \frac{|z-w|}{24(c_1(1)\varepsilon|\log\varepsilon| + |z-w|)} \geq \frac{|z-w|}{48c_1(1)\varepsilon|\log\varepsilon|}.$$

The corresponding result for $0 \leq w \leq z$ follows by symmetry (exchanging the roles of w and z). The case $z \leq 0$ then follows by the symmetry in (4.19). □

4.2.3 A coupling argument

The second important ingredient in our proof of Theorem 4.2.3 will be a coupling between $d(W_s, t-s)$ (the signed distance from a $\mathbb{d}$-dimensional Brownian motion W_s to Γ_{t-s}, which evolves according to (backwards in time) mean curvature flow) and a one-dimensional Brownian motion, at least when W_s is close to Γ_{t-s}. The proof requires some regularity properties of the mean curvature flow that we record in this subsection. These rest on the assumptions ($\mathscr{C}1$)-($\mathscr{C}3$).

We write $\dot{d}$ for the time derivative of d. Let $T^* \in (0, \mathscr{T})$. In this case, we have:

1. There exists $c_0 > 0$ such that for all $t \in [0, T^*]$ and $x \in \{y : |d(y,t)| \leq c_0\}$, we have

   $$|\nabla d(x,t)| = 1. \tag{4.44}$$

   Moreover, d is a $C^{\alpha, \frac{\alpha}{2}}$ function in $\{(x,t) : |d(x,t)| \leq c_0, t \leq T^*\}$.

2. Viewing $\boldsymbol{n} = \nabla d$ as the positive normal direction, for $x \in \Gamma_t$, the normal velocity of Γ_t at x is $-\dot{d}(x,t)$, and the curvature of Γ_t at x is $-\Delta d(x,t)$. Thus, (4.5) becomes

   $$\dot{d}(x,t) = \Delta d(x,t) \tag{4.45}$$

   for all x such that $d(x,t) = 0$.

3. There exists $C_0 > 0$ such that for all $t \in [0, T^*]$ and x such that $|d(x,t)| \leq c_0$,

$$\left|\nabla \left(\dot{d}(x,t) - \Delta d(x,t)\right)\right| \leq C_0. \tag{4.46}$$

4. There exist $v_0, V_0 > 0$ such that for all $t \in [0, T^* - v_0]$ and all $s \in [t, t+v_0]$,

$$|d(x,t) - d(x,s)| \leq V_0(s-t). \tag{4.47}$$

Properties 1 and 2 above come from Chen (1992) (equations (2.9), (2.10) and Proposition 2.1) and 3 and 4 follow easily from the fact that $\sup_{u \in S^1, t \leq T^*} |\Gamma_t(u)| < \infty$ and the regularity of d provided by 1.

The first property means that, for each $t \geq 0$, the region $\{x : d(x,t) \leq c_0\}$ is not self-intersecting i.e. for each x it contains, the ball $\{z : |z-x| \leq d(x,t)\}$ intersects Γ_t at precisely one point. Evidently this cannot hold, for example, as the flow collapses to a point, which is why we work up to time $T^* < \mathscr{T}$. Broadly speaking, the first two properties characterize mean curvature flow in terms of the function d.

A key ingredient of our proof of Theorem 4.2.3 is the following coupling argument.

Proposition 4.2.13. *Let* $(W_s)_{s\geq 0}$ *denote a* d*-dimensional Brownian motion started at* $x \in \mathbb{R}^{\mathrm{d}}$. *Suppose that* $t \leq T^*$, $\beta \leq c_0$ *and let*

$$T_\beta = \inf\left(\{s \in [0,t) : |d(W_s, t-s)| \geq \beta\} \cup \{t\}\right).$$

Then we can couple $(W_s)_{s\geq 0}$ *with a one-dimensional Brownian motion* $(B_s)_{s\geq 0}$ *started from* $z = d(x,t)$ *in such a way that for* $s \leq T_\beta$,

$$B_s - C_0\beta s \leq d\left(W_s, t-s\right) \leq B_s + C_0\beta s.$$

Proof. By Itô's formula, we have that for $s \leq t$

$$d\left(W_s, t-s\right) = \int_0^s A_u \, du + B_s,$$

where

$$\begin{aligned} A_u &= -\dot{d}\,(W_u, t-u) + \Delta d\,(W_u, t-u) \\ B_s &= \sum_{i=1}^{\mathbb{d}} \int_0^s \frac{\partial}{\partial x_i} d(W_u, t-u) dW_u^{(i)}. \end{aligned}$$

We will handle A_u and B_s in turn.

For each $u \in [0, T_\beta]$ there exists some $x_u \in \mathbb{R}^{\mathbb{d}}$ such that $|x_u - W_u| \leq \beta$, and $d(x_u, t-u) = 0$. By (4.45) we have $-\dot{d}(x_u, t-u) + \Delta d(x_u, t-u) = 0$. Since $\beta \leq c_0$, by (4.46) we have that, for x on the line segment connecting x_u to W_u, the gradient of $-\dot{d}(x, t-u) + \Delta d(x, t-u)$ is bounded by C_0. We thus obtain

$$|A_u| \leq C_0 \beta.$$

Since $\beta \leq c_0$, it follows by (4.44) and Lévy's characterisation (recall that our Brownian motions run at rate 2) that $(B_s)_{0 \leq s \leq T_\beta}$ is a (stopped) Brownian Motion. This completes the proof. □

Remark 4.2.14. *Proposition 4.2.13 provides a probabilistic parallel to one of the key tools used in the classical study of (mean) curvature flow; approximating the movement of the interface locally (in space and time) by a particular one dimensional standing wave.*

4.2.4 Majority voting in BBM, for $\mathbb{d} \geq 2$

Recall the notation introduced in Section 4.2.1 for ternary branching Brownian motion in dimension $\mathbb{d} \geq 2$. For $x \in \mathbb{R}^{\mathbb{d}}$, we write $\mathbb{P}_x^\varepsilon$ for the probability measure under which $(\boldsymbol{W}(t), t \geq 0)$ has the law of ternary branching Brownian motion in $\mathbb{R}^{\mathbb{d}}$ with branching rate $1/\varepsilon^2$ started from a single particle at location x at time 0. We use $\mathbb{E}_x^\varepsilon$ for the corresponding expectation. We also write P_x for the probability measure under which $(W_t)_{t \geq 0}$ has the law of a $\mathbb{d}$-dimensional Brownian motion started at x, and E_x for the corresponding expectation. As usual the notation B (resp. $\boldsymbol{B}$) refers to a one dimensional (historical branching) Brownian motion and W and $\boldsymbol{W}$ signal dimension $\mathbb{d} \geq 2$.

The proof of Theorem 4.2.3 is in two parts. First, in Section 4.2.4.1 we establish that the interface is generated in a time $\delta_{\mathbb{d}} = \mathcal{O}(\varepsilon^2|\log\varepsilon|)$. We then, in Section 4.2.4.2, use Proposition 4.2.13 and Theorem 4.2.5 to investigate how the region around the interface propagates. In order not to interrupt the flow of the proof of Theorem 4.2.3, the proof of a central lemma is deferred to Section 4.2.4.3.

Our proof rests on a comparison with the outcome $\mathbb{V}(\boldsymbol{B}(t))$ of majority voting for the one-dimensional historical branching Brownian motion. In one dimension we always implicitly take $\mathbb{V} = \mathbb{V}_{p_0}$ with $p_0(x) = \mathbb{1}\{x \geq 0\}$. We reserve the subscript p for $\mathbb{V}_p(\boldsymbol{W}(t))$ and we assume that p satisfies $(\mathscr{C}1)$-$(\mathscr{C}3)$.

4.2.4.1 Generation of the interface

In this section we prove that, as in $\mathbb{d} = 1$, in dimension $\mathbb{d} \geq 2$ an interface of width $\mathcal{O}(\varepsilon|\log\varepsilon|)$ is generated in time $\mathcal{O}(\varepsilon^2|\log\varepsilon|)$.

Proposition 4.2.15. *Let $k \in \mathbb{N}$. Then there exist $\varepsilon_{\mathbb{d}}(k), a_{\mathbb{d}}(k), b_{\mathbb{d}}(k) > 0$ such that for all $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$, if we set*

$$\delta_{\mathbb{d}}(k,\varepsilon) := a_{\mathbb{d}}(k)\varepsilon^2|\log\varepsilon| \quad \textit{and} \quad \delta'_{\mathbb{d}}(k,\varepsilon) := (a_{\mathbb{d}}(k)+k+1)\varepsilon^2|\log\varepsilon|, \tag{4.48}$$

then for $t \in [\delta_{\mathbb{d}}, \delta'_{\mathbb{d}}]$,

1. *for x such that $d(x,t) \geq b_{\mathbb{d}}\varepsilon|\log\varepsilon|$, we have $\mathbb{P}^{\varepsilon}_x\left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1\right] \geq 1 - \varepsilon^k$;*
2. *for x such that $d(x,t) \leq -b_{\mathbb{d}}\varepsilon|\log\varepsilon|$, we have $\mathbb{P}^{\varepsilon}_x\left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1\right] \leq \varepsilon^k$.*

Proof. By the same argument as for Lemma 4.2.9, given $k \in \mathbb{N}$, and taking $A(k)$ from Lemma 4.2.8, there exist $a_{\mathbb{d}}(k)$ and $\varepsilon_{\mathbb{d}}(k) > 0$ such that, for all $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$ and $t \geq a_{\mathbb{d}}\varepsilon^2|\log\varepsilon|$,

$$\mathbb{P}^{\varepsilon}\left[\mathcal{T}(\boldsymbol{W}(t)) \supseteq \mathcal{T}^{reg}_{A(k)|\log\varepsilon|}\right] \geq 1 - \varepsilon^k. \tag{4.49}$$

It is also easy to obtain a $\mathbb{d}$-dimensional equivalent of Lemma 4.2.10, with essentially the same proof (using a tail bound on a $\mathbb{d}$-dimensional normal distribution instead of one dimensional). That is, given $k \in \mathbb{N}$, there exist $d_{\mathbb{d}}(k)$, $\varepsilon_{\mathbb{d}}(k)$ such that for all

$\varepsilon \in (0, \varepsilon_{\mathrm{d}})$, for $t \in [\delta_{\mathrm{d}}, \delta'_{\mathrm{d}}]$,

$$\mathbb{P}_x^\varepsilon \left[\exists i \in N(t) : |W_i(t) - x| \geq d_{\mathrm{d}}\varepsilon|\log \varepsilon|\right] \leq \varepsilon^k. \tag{4.50}$$

We set $b_{\mathrm{d}}(k) = 2d_{\mathrm{d}}(k)$.

By (4.47) there exist $v_0, V_0 > 0$ such that for $t \leq v_0$, and any $x \in \mathbb{R}^{\mathrm{d}}$, we have $|d(x,0) - d(x,t)| \leq V_0 t$. Reducing ε_{d} if necessary, for $\varepsilon \in (0, \varepsilon_{\mathrm{d}})$ we have $\delta'_{\mathrm{d}} \leq v_0$. Thus, if $\varepsilon \in (0, \varepsilon_{\mathrm{d}})$, $t \in [\delta_{\mathrm{d}}, \delta'_{\mathrm{d}}]$ and x is such that $d(x,t) \geq b_{\mathrm{d}}\varepsilon|\log\varepsilon|$ and $|W_i(t) - x| \leq d_{\mathrm{d}}\varepsilon|\log\varepsilon|$ then combining with the triangle inequality and (4.47),

$$\begin{aligned} d\left(W_i(t), 0\right) &\geq d\left(x, t\right) - \left|d\left(x,t\right) - d\left(W_i(t), t\right)\right| - \left|d\left(W_i(t), t\right) - d\left(W_i(t), 0\right)\right| \\ &\geq b_{\mathrm{d}}\varepsilon|\log\varepsilon| - d_{\mathrm{d}}\varepsilon|\log\varepsilon| - V_0\delta'_{\mathrm{d}} \\ &= \frac{1}{2}b_{\mathrm{d}}\varepsilon|\log\varepsilon| - V_0(a_{\mathrm{d}} + k + 1)\varepsilon^2|\log\varepsilon|. \end{aligned}$$

Therefore, reducing ε_{d} if necessary, in this case we have that

$$d(W_i(t), 0) \geq \tfrac{1}{4}b_{\mathrm{d}}\varepsilon|\log\varepsilon|.$$

Applying ($\mathscr{C}2$) and ($\mathscr{C}3$),

$$\begin{aligned} p(W_i(t)) &\geq \tfrac{1}{2} + \gamma\left(\tfrac{1}{4}b_{\mathrm{d}}\varepsilon|\log\varepsilon| \wedge r\right) \\ &\geq \tfrac{1}{2} + \varepsilon, \end{aligned} \tag{4.51}$$

where we again reduce $\varepsilon_{\mathrm{d}} > 0$ (if necessary), to ensure that $\varepsilon < \gamma r$, $\varepsilon < \frac{\gamma}{4}b_{\mathrm{d}}\varepsilon|\log\varepsilon|$ for $\varepsilon \in (0, \varepsilon_{\mathrm{d}})$.

Exactly as in the proof of Theorem 4.2.5, we can now combine (4.49), (4.50) and (4.51) to deduce that for $\varepsilon \in (0, \varepsilon_{\mathrm{d}})$, $t \in [\delta_{\mathrm{d}}, \delta'_{\mathrm{d}}]$ and x such that $d(x,t) \geq b_{\mathrm{d}}\varepsilon|\log\varepsilon|$,

$$\mathbb{P}_x^\varepsilon\left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1\right] \geq 1 - 3\varepsilon^k.$$

The proof of the second statement is analogous. □

4.2.4.2 Propagation of the interface and proof of Theorem 4.2.3

We now turn to the propagation of the interface region. Our immediate goal is to establish that

$$\mathbb{P}_x^\varepsilon\left[\mathbb{V}_p(\boldsymbol{W}(t))=1\right]\approx\mathbb{P}^\varepsilon_{d(x,t)+\mathcal{O}(\varepsilon|\log\varepsilon|)}\left[\mathbb{V}(\boldsymbol{B}(t))=1\right].$$

This connection between $\boldsymbol{B}$ and $\boldsymbol{W}$ is made precise by the following result.

Proposition 4.2.16. *Let $l\in\mathbb{N}$ with $l\geq 4$. Define $a_{\mathrm{d}}(l)$ and $\delta_{\mathrm{d}}(l,\varepsilon)$ as in Proposition 4.2.15. There exist $K_1(l),K_2(l)>0$ and $\varepsilon_{\mathrm{d}}(l,K_1,K_2)>0$ such that for all $\varepsilon\in(0,\varepsilon_{\mathrm{d}})$ and $t\in[\delta_{\mathrm{d}}(l,\varepsilon),T^*]$ we have*

$$\sup_{x\in\mathbb{R}^{\mathrm{d}}}\left(\mathbb{P}_x^\varepsilon\left[\mathbb{V}_p(\boldsymbol{W}(t))=1\right]-\mathbb{P}^\varepsilon_{d(x,t)+K_1e^{K_2t}\varepsilon|\log\varepsilon|}\left[\mathbb{V}(\boldsymbol{B}(t))=1\right]\right)\leq\varepsilon^l \tag{4.52}$$

and

$$\sup_{x\in\mathbb{R}^{\mathrm{d}}}\left(\mathbb{P}_x^\varepsilon\left[\mathbb{V}_p(\boldsymbol{W}(t))=0\right]-\mathbb{P}^\varepsilon_{d(x,t)-K_1e^{K_2t}\varepsilon|\log\varepsilon|}\left[\mathbb{V}(\boldsymbol{B}(t))=0\right]\right)\leq\varepsilon^l. \tag{4.53}$$

The proof of Theorem 4.2.3, which follows easily from Proposition 4.2.16, is at the end of this subsection.

Recall that $g:[0,1]\to[0,1]$ is given by $g(p)=3p^2-2p^3$. It is convenient to extend this definition to a continuous, monotone function $g:\mathbb{R}\to[0,1]$ as follows:

$$g(p)=\begin{cases}0 & \text{if } p<0\\ 3p^2-2p^3 & \text{if } p\in[0,1]\\ 1 & \text{if } p>1.\end{cases} \tag{4.54}$$

At the heart of the proof of Proposition 4.2.16 is the following lemma, whose proof we defer to Section 4.2.4.3.

Lemma 4.2.17. *Let $l\in\mathbb{N}$ with $l\geq 4$ and $K_1>0$. There exists $K_2=K_2(K_1,l)>0$ and $\varepsilon_{\mathrm{d}}(l,K_1,K_2)>0$ such that for all $\varepsilon\in(0,\varepsilon_{\mathrm{d}})$, $x\in\mathbb{R}^{\mathrm{d}}$, $s\in[0,(l+1)\varepsilon^2|\log\varepsilon|]$ and*

$t \in [s, T^*]$,

$$E_x \left[g \left(\mathbb{P}^{\varepsilon}_{d(W_s, t-s)+K_1 e^{K_2(t-s)}\varepsilon|\log\varepsilon|} [\mathbb{V}(\boldsymbol{B}(t-s)) = 1] + \varepsilon^l \right) \right] \\ \le \tfrac{3}{4}\varepsilon^l + E_{d(x,t)} \left[g \left(\mathbb{P}^{\varepsilon}_{B_s + K_1 e^{K_2 t}\varepsilon|\log\varepsilon|} [\mathbb{V}(\boldsymbol{B}(t-s)) = 1] \right) \right] + \mathbb{1}_{s \le \varepsilon^3} \varepsilon^l \tag{4.55}$$

and

$$E_x \left[g \left(\mathbb{P}^{\varepsilon}_{d(W_s, t-s)-K_1 e^{K_2(t-s)}\varepsilon|\log\varepsilon|} [\mathbb{V}(\boldsymbol{B}(t-s)) = 0] + \varepsilon^l \right) \right] \\ \le \tfrac{3}{4}\varepsilon^l + E_{d(x,t)} \left[g \left(\mathbb{P}^{\varepsilon}_{B_s - K_1 e^{K_2 t}\varepsilon|\log\varepsilon|} [\mathbb{V}(\boldsymbol{B}(t-s)) = 0] \right) \right] + \mathbb{1}_{s \le \varepsilon^3} \varepsilon^l. \tag{4.56}$$

Proof of Proposition 4.2.16. Take $K_1 = b_{\mathrm{d}}(l) + c_1(l)$ where b_{d} is as defined in Proposition 4.2.15 and c_1 is as defined in Theorem 4.2.5. Let $K_2 = K_2(K_1, l)$, as defined in Lemma 4.2.17. Take $\varepsilon_{\mathrm{d}} > 0$ sufficiently small that Theorem 4.2.5, Proposition 4.2.15 and Lemma 4.2.17 apply for $\varepsilon \in (0, \varepsilon_{\mathrm{d}})$. We begin by observing that for $\varepsilon \in (0, \varepsilon_{\mathrm{d}})$, $t \in [\delta_{\mathrm{d}}, \delta'_{\mathrm{d}}]$ (where δ'_{d} is defined in (4.48)), and $x \in \mathbb{R}^{\mathrm{d}}$,

$$\mathbb{P}^{\varepsilon}_x [\mathbb{V}_p(\boldsymbol{W}(t)) = 1] \le \mathbb{P}^{\varepsilon}_{d(x,t)+K_1 e^{K_2 t}\varepsilon|\log\varepsilon|} [\mathbb{V}(\boldsymbol{B}(t)) = 1] + \varepsilon^l. \tag{4.57}$$

To see this, note that if $d(x,t) \le -b_{\mathrm{d}}(l)\varepsilon|\log\varepsilon|$, then by Proposition 4.2.15 we have that $\mathbb{P}^{\varepsilon}_x [\mathbb{V}_p(\boldsymbol{W}(t)) = 1] \le \varepsilon^l$. On the other hand, if $d(x,t) \ge -b_{\mathrm{d}}(l)\varepsilon|\log\varepsilon|$, then $d(x,t) + K_1 e^{K_2 t}\varepsilon|\log\varepsilon| \ge c_1(l)\varepsilon|\log\varepsilon|$, and so, by Theorem 4.2.5, (4.57) holds (since the right hand side of (4.57) is ≥ 1).

We are left with the case $t \in [\delta'_{\mathrm{d}}, T^*]$. We assume, aiming for a contradiction, that there exists $t \in [\delta'_{\mathrm{d}}, T^*]$ such that, for some $x \in \mathbb{R}^{\mathrm{d}}$,

$$\mathbb{P}^{\varepsilon}_x [\mathbb{V}_p(\boldsymbol{W}(t)) = 1] - \mathbb{P}^{\varepsilon}_{d(x,t)+K_1 e^{K_2 t}\varepsilon|\log\varepsilon|} [\mathbb{V}(\boldsymbol{B}(t)) = 1] > \varepsilon^l.$$

Let T' be the infimum of the set of such t. Choose

$$T \in [T', \min(T' + \varepsilon^{l+3}, T^*)] \tag{4.58}$$

which is in the set of such t. Hence, there exists some $x = x(l,\varepsilon) \in \mathbb{R}^{\mathrm{d}}$ such that

$$\mathbb{P}^{\varepsilon}_{x}\left[\mathbb{V}_p(\boldsymbol{W}(T)) = 1\right] - \mathbb{P}^{\varepsilon}_{d(x,T)+K_1e^{K_2T}\varepsilon|\log\varepsilon|}\left[\mathbb{V}(\boldsymbol{B}(T)) = 1\right] > \varepsilon^{l}. \tag{4.59}$$

We now seek to show that

$$\mathbb{P}^{\varepsilon}_{x}\left[\mathbb{V}_p(\boldsymbol{W}(T)) = 1\right] \le \tfrac{7}{8}\varepsilon^{l} + \mathbb{P}^{\varepsilon}_{d(x,T)+K_1e^{K_2T}\varepsilon|\log\varepsilon|}\left[\mathbb{V}(\boldsymbol{B}(T)) = 1\right]. \tag{4.60}$$

Since $\frac{7}{8}\varepsilon^{l} < \varepsilon^{l}$, once we obtain equation (4.60) we have a contradiction to (4.59), thus completing the proof.

We write S for the time of the first branching event in $\boldsymbol{W}(T)$ and W_S for the position of the initial 'ancestor' particle at that time. We note that by the strong Markov property at time $S \wedge (T - \delta_{\mathbb{d}})$,

$$\begin{aligned}\mathbb{P}^{\varepsilon}_{x}\left[\mathbb{V}_p(\boldsymbol{W}(T)) = 1\right] = \mathbb{E}^{\varepsilon}_{x}&\left[g(\mathbb{P}^{\varepsilon}_{W_S}\left[\mathbb{V}_p(\boldsymbol{W}(T-S)) = 1\right])\mathbb{1}_{S\le T-\delta_{\mathbb{d}}}\right]\\ &+ \mathbb{E}^{\varepsilon}_{x}\left[\mathbb{P}^{\varepsilon}_{W_{T-\delta_{\mathbb{d}}}}\left[\mathbb{V}_p(\boldsymbol{W}(\delta_{\mathbb{d}})) = 1\right]\mathbb{1}_{S\ge T-\delta_{\mathbb{d}}}\right].\end{aligned} \tag{4.61}$$

We begin with the second term on the right of (4.61). Since $T - \delta_{\mathbb{d}} \ge \delta'_{\mathbb{d}} - \delta_{\mathbb{d}} = (l+1)\varepsilon^2|\log\varepsilon|$ and $S \sim \mathrm{Exp}(\varepsilon^{-2})$,

$$\mathbb{E}^{\varepsilon}_{x}\left[\mathbb{P}^{\varepsilon}_{W_{T-\delta_{\mathbb{d}}}}\left[\mathbb{V}_p(\boldsymbol{W}(\delta_{\mathbb{d}})) = 1\right]\mathbb{1}_{S\ge T-\delta_{\mathbb{d}}}\right] \le \mathbb{P}^{\varepsilon}\left[S \ge (l+1)\varepsilon^2|\log\varepsilon|\right] = \varepsilon^{l+1}. \tag{4.62}$$

To bound the first term on the right of (4.61), partition on the event $\{S \le \varepsilon^{l+3}\}$ (which has probability $\le \varepsilon^{l+1}$):

$$\begin{aligned}&\mathbb{E}^{\varepsilon}_{x}\left[g(\mathbb{P}^{\varepsilon}_{W_S}\left[\mathbb{V}_p(\boldsymbol{W}(T-S)) = 1\right])\mathbb{1}_{S\le T-\delta_{\mathbb{d}}}\right]\\ &\quad\le \mathbb{P}^{\varepsilon}\left[S \le \varepsilon^{l+3}\right] + \mathbb{E}^{\varepsilon}_{x}\left[g(\mathbb{P}^{\varepsilon}_{W_S}\left[\mathbb{V}_p(\boldsymbol{W}(T-S)) = 1\right])\mathbb{1}_{S\le T-\delta_{\mathbb{d}}}\mathbb{1}_{S\ge\varepsilon^{l+3}}\right]\\ &\quad\le \varepsilon^{l+1} + \mathbb{E}^{\varepsilon}_{x}\left[g\left(\mathbb{P}^{\varepsilon}_{d(W_S,T-S)+K_1e^{K_2(T-S)}\varepsilon|\log\varepsilon|}\left[\mathbb{V}(\boldsymbol{B}(T-S)) = 1\right] + \varepsilon^{l}\right)\mathbb{1}_{S\le T-\delta_{\mathbb{d}}}\right].\end{aligned} \tag{4.63}$$

The last line follows from the minimality of T' (note that if $\varepsilon^{l+3} \le S \le T - \delta_{\mathbb{d}}$, then $T - S \in [\delta_{\mathbb{d}}, T')$ by (4.58)) and from monotonicity of g.

Conditioning on the value of S, since the path of the ancestor particle $(W_\cdot)$ is independent of S,

$$\begin{aligned}
&\mathbb{E}_x^\varepsilon \left[g\left(\mathbb{P}^\varepsilon_{d(W_S, T-S)+K_1 e^{K_2(T-S)}\varepsilon|\log\varepsilon|} \left[\mathbb{V}(\boldsymbol{B}(T-S)) = 1 \right] + \varepsilon^l \right) \mathbb{1}_{S \le T-\delta_{\mathbb{d}}} \right] \\
&\quad \le \int_0^{(l+1)\varepsilon^2|\log\varepsilon|} \varepsilon^{-2} e^{-\varepsilon^{-2}s} E_x \left[g\left(\mathbb{P}^\varepsilon_{d(W_s, T-s)+K_1 e^{K_2(T-s)}\varepsilon|\log\varepsilon|} \left[\mathbb{V}(\boldsymbol{B}(T-s)) = 1 \right] + \varepsilon^l \right) \right] ds \\
&\qquad\qquad + \mathbb{P}^\varepsilon \left[S \ge (l+1)\varepsilon^2 |\log\varepsilon| \right] \\
&\quad \le \tfrac{3}{4}\varepsilon^l + \int_0^{(l+1)\varepsilon^2|\log\varepsilon|} \varepsilon^{-2} e^{-\varepsilon^{-2}s} E_{d(x,T)} \left[g\left(\mathbb{P}^\varepsilon_{B_s + K_1 e^{K_2 T}\varepsilon|\log\varepsilon|} [\mathbb{V}(\boldsymbol{B}(T-s)) = 1] \right) \right] ds \\
&\qquad\qquad + \mathbb{P}^\varepsilon \left[S \le \varepsilon^3 \right] \varepsilon^l + \varepsilon^{l+1} \\
&\quad \le \tfrac{3}{4}\varepsilon^l + 2\varepsilon^{l+1} + \mathbb{E}^\varepsilon_{d(x,T)} \left[g\left(\mathbb{P}^\varepsilon_{B_{S'} + K_1 e^{K_2 T}\varepsilon|\log\varepsilon|} \left[\mathbb{V}(\boldsymbol{B}(T-S')) = 1 \right] \right) \mathbb{1}_{S' \le T-\delta_{\mathbb{d}}} \right].
\end{aligned} \tag{4.64}$$

Here, the second inequality follows by Lemma 4.2.17. For the final inequality, we write S' for the time of the first branching event in $(\boldsymbol{B}(s), s \ge 0)$ and $B_{S'}$ for the position of the ancestor at that time, and note that S' has the same distribution as S. The inequality follows since $T \ge \delta'_{\mathbb{d}}$ and so $T - \delta_{\mathbb{d}} \ge (l+1)\varepsilon^2|\log\varepsilon|$.

Putting (4.63), (4.64) and (4.62) into (4.61) we obtain

$$\begin{aligned}
&\mathbb{P}_x^\varepsilon \left[\mathbb{V}_p(\boldsymbol{W}(T)) = 1 \right] \\
&\qquad \le 4\varepsilon^{l+1} + \tfrac{3}{4}\varepsilon^l + \mathbb{E}^\varepsilon_{d(x,T)} \left[g\left(\mathbb{P}^\varepsilon_{B_{S'} + K_1 e^{K_2 T}\varepsilon|\log\varepsilon|} [\mathbb{V}(\boldsymbol{B}(T-S')) = 1] \right) \mathbb{1}_{S' \le T-\delta_{\mathbb{d}}} \right] \\
&\qquad \le 4\varepsilon^{l+1} + \tfrac{3}{4}\varepsilon^l + \mathbb{P}^\varepsilon_{d(x,T)+K_1 e^{K_2 T}\varepsilon|\log\varepsilon|} \left[\mathbb{V}(\boldsymbol{B}(T)) = 1 \right],
\end{aligned}$$

where the final line follows by the strong Markov Property for $(\boldsymbol{B}(\cdot))$ at time $S' \wedge (T - \delta_{\mathbb{d}})$, in similar style to (4.61). Reducing $\varepsilon_{\mathbb{d}}$, if necessary, to ensure that $\frac{3}{4}\varepsilon^l + 4\varepsilon^{l+1} \le \frac{7}{8}\varepsilon^l$ for all $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$, we obtain (4.60), which completes the proof of (4.52).

By a similar argument, using (4.56) in place of (4.55), we can also deduce (4.53). □

Proof of Theorem 4.2.3. It suffices to prove the result for sufficiently large $k \in \mathbb{N}$, and in particular we will show it for $k \ge 4$.

We choose $c_{\mathbb{d}}(k) = c_1(k) + K_1 e^{K_2 T^*}$. Thus, for any $t \in [\delta_{\mathbb{d}}, T^*]$ and $x \in \mathbb{R}^{\mathbb{d}}$ such that $d(x,t) \leq -c_{\mathbb{d}}(k)\varepsilon|\log\varepsilon|$ we have

$$d(x,t) + K_1 e^{K_2 t}\varepsilon|\log\varepsilon| \leq -c_1(k)\varepsilon|\log\varepsilon|.$$

It follows from Theorem 4.2.5 (reducing $\varepsilon_{\mathbb{d}}$ if necessary so that $\varepsilon < \varepsilon_1(k)$) and (4.52) that $\mathbb{P}_x\left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1\right] \leq 2\varepsilon^k$ for such x and t. Similarly, for x such that $d(x,t) \geq c_{\mathbb{d}}(k)\varepsilon|\log\varepsilon|$, by Theorem 4.2.5 and (4.53) we have $\mathbb{P}_x\left[\mathbb{V}_p(\boldsymbol{W}(t)) = 0\right] \leq 2\varepsilon^k$. □

4.2.4.3 Proof of Lemma 4.2.17

To complete the proof of Theorem 4.2.3, it remains to prove Lemma 4.2.17. The ideas in the proof are simple, but are easily lost in the notation, so to explain the structure we begin with an outline of the proof of the first inequality (4.55). (The proof of (4.56) goes along essentially the same lines.)

We take a large constant C and consider the cases $|d(x,t)| \geq C\varepsilon|\log\varepsilon|$ and $|d(x,t)| \leq C\varepsilon|\log\varepsilon|$ separately. Since $s = \mathcal{O}(\varepsilon^2|\log\varepsilon|)$, with high probability neither the $\mathbb{d}$-dimensional Brownian motion W nor the one-dimensional B moves a distance more than $\mathcal{O}(\varepsilon|\log\varepsilon|)$ before time s. Therefore, if C is sufficiently large and $d(x,t) \leq -C\varepsilon|\log\varepsilon|$, Theorem 4.2.5 tells us that the left-hand side of (4.55) is $\leq \varepsilon^{l+1}$; similarly, if $d(x,t) \geq C\varepsilon|\log\varepsilon|$ then the right-hand side of (4.55) is ≥ 1. This leaves the case of $|d(x,t)| \leq C\varepsilon|\log\varepsilon|$, in which we apply Proposition 4.2.13 to couple W_s with B_s in such a way that with probability $1 - \mathcal{O}(\varepsilon^{l+1})$,

$$d(W_s, t-s) \leq B_s + \mathcal{O}(\varepsilon|\log\varepsilon|)s.$$

Thus, using monotonicity (4.17), the left-hand side of (4.55) is bounded above by

$$\mathbb{E}_{d(x,t)}\left[g\left(\mathbb{P}^{\varepsilon}_{B_s + (K_1 e^{K_2(t-s)} + \mathcal{O}(s))\varepsilon|\log\varepsilon|}[\mathbb{V}(\boldsymbol{B}(t-s)) = 1] + \varepsilon^l\right)\right] + \mathcal{O}(\varepsilon^{l+1}).$$

If $|p - \frac{1}{2}| \geq \frac{7}{18}$, we can use that $|g'(p)| \leq 2/3$ to pull the ε^l outside the argument of g and then use monotonicity again to recover (4.55). The difficulty is that close to $p = \frac{1}{2}$, we have $g'(p) > 1$. In the case $\mathbb{P}_{B_s + (K_1 e^{K_2(t-s)} + \mathcal{O}(s))\varepsilon|\log\varepsilon|}[\mathbb{V}(\boldsymbol{B}(t-s)) = 1] \approx \frac{1}{2}$, we instead choose $K_2 \gg 0$, and use the lower bound on the 'slope of the interface' given by Corollary 4.2.12 to estimate the increment in $\mathbb{P}_z^\varepsilon[\mathbb{V}(\boldsymbol{B}(t-s))=1]$ when we replace $z+(K_1e^{K_2(t-s)}+\mathcal{O}(s))\varepsilon|\log\varepsilon|$ by $z+K_1e^{K_2t}\varepsilon|\log\varepsilon|$.

The remainder of this subsection contains the formal proof.

Proof of Lemma 4.2.17. We begin by proving (4.55). For the duration of the proof, for $u\geq 0$ and $z\in\mathbb{R}$ we write

$$\mathbb{Q}_z^{\varepsilon,u}=\mathbb{P}_z^\varepsilon\left[\mathbb{V}(\boldsymbol{B}(u))=1\right].$$

Recall C_0 and $c_1(k)$ from (4.46) and Theorem 4.2.5 respectively. Let

$$R=2c_1(l)+4(l+1)\mathbb{d}+1. \tag{4.65}$$

Fix K_2 such that

$$K_1(K_2-C_0)-C_0R=c_1(1). \tag{4.66}$$

Let $\varepsilon_{\mathbb{d}}=\varepsilon_1(l)$ where $\varepsilon_1(l)$ is defined in Theorem 4.2.5.

First we need an estimate for the probability that a $\mathbb{d}$-dimensional Brownian motion moves further than $\sim\varepsilon|\log\varepsilon|$ in time s (recall that $s\leq(l+1)\varepsilon^2|\log\varepsilon|$). Let

$$A_x=\left\{\sup_{u\in[0,s]}|W_u-x|\leq 2(l+1)\,\mathbb{d}\varepsilon|\log\varepsilon|\right\}.$$

Then bounding $|W_u|$ by the sum of the moduli of $\mathbb{d}$ one-dimensional Brownian motions and using the reflectional symmetry of one dimensional Brownian motion,

$$\begin{aligned}P_x\left[A_x^c\right]&\leq 2\mathbb{d}P_0\left[\sup_{u\in[0,s]}B_u>2(l+1)\varepsilon|\log\varepsilon|\right]\\&\leq 4\mathbb{d}P_0\left[B_1>2((l+1)|\log\varepsilon|)^{1/2}\right]\\&\leq 4\mathbb{d}\varepsilon^{l+1}.\end{aligned}\tag{4.67}$$

Here, since $s\leq(l+1)\varepsilon^2|\log\varepsilon|$ the second line follows by the reflection principle. The last line follows using the tail bound $\mathbb{P}[B_1\geq x]\leq e^{-x^2/4}$.

As advertised, we now consider the following three cases:

(i) $d(x,t)\leq-\left(2c_1(l)+2(l+1)\mathbb{d}+K_1e^{K_2(t-s)}\right)\varepsilon|\log\varepsilon|$,

(ii) $d(x,t) \geq \left(2c_1(l) + 2(l+1)\mathbb{d} + K_1 e^{K_2(t-s)}\right) \varepsilon |\log \varepsilon|$,

(iii) $|d(x,t)| \leq \left(2c_1(l) + 2(l+1)\mathbb{d} + K_1 e^{K_2(t-s)}\right) \varepsilon |\log \varepsilon|$.

The third case corresponds to x being close to the interface at time t. The first two cases correspond to x falling (sufficiently far) inside or outside of the interface.

Case (i): Recall that by (4.47) there exist $v_0, V_0 > 0$ such that if $s \leq v_0$ and $x \in \mathbb{R}^{\mathbb{d}}$ then

$$|d(x,t) - d(x,t-s)| \leq V_0 s. \tag{4.68}$$

We reduce $\varepsilon_{\mathbb{d}}$, if necessary, to ensure that for $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$ we have $(l+1)\varepsilon^2 |\log \varepsilon| \leq v_0$. Then if the event A_x occurs,

$$\begin{aligned}
&d(W_s, t-s) + K_1 e^{K_2(t-s)} \varepsilon |\log \varepsilon| \\
&\qquad \leq -(2c_1(l) + 2(l+1)\mathbb{d})\varepsilon |\log \varepsilon| + |d(W_s, t-s) - d(x,t)| \\
&\qquad \leq -(2c_1(l) + 2(l+1)\mathbb{d})\varepsilon |\log \varepsilon| + |d(x,t) - d(x,t-s)| + |W_s - x| \\
&\qquad \leq -2c_1(l)\varepsilon |\log \varepsilon| + V_0 (l+1)\varepsilon^2 |\log \varepsilon|.
\end{aligned} \tag{4.69}$$

Here, the second line follows from being in case (i) and the third follows from the triangle inequality. The final line then follows from (4.68) and that $s \leq (l+1)\varepsilon^2 |\log \varepsilon|$, and since A_x occurs.

Reducing $\varepsilon_{\mathbb{d}}$, if necessary, from (4.69) we have

$$d(W_s, t-s) + K_1 e^{K_2(t-s)} \varepsilon |\log \varepsilon| \leq -c_1(l) \varepsilon |\log \varepsilon|.$$

Therefore

$$\begin{aligned}
E_x \left[g\left(\mathbb{Q}^{\varepsilon, t-s}_{d(W_s, t-s) + K_1 e^{K_2(t-s)} \varepsilon |\log \varepsilon|} + \varepsilon^l \right) \right] &\leq E_x \left[g(\varepsilon^l + \varepsilon^l) \mathbb{1}_{A_x} \right] + P_x \left[A_x^c \right] \\
&\leq 6\varepsilon^{2l} + 4\mathbb{d}\varepsilon^{l+1}.
\end{aligned}$$

Here the first inequality follows by Theorem 4.2.5 and the second inequality by the definition of g in (4.54) and by (4.67). Again reducing $\varepsilon_{\mathbb{d}}$ if necessary, for $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$ we have

$$E_x[g(\mathbb{Q}^{\varepsilon,t-s}_{d(W_s,t-s)+K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|}+\varepsilon^l)]\leq\tfrac{3}{4}\varepsilon^l,$$

and so (4.55) holds in this case.

Case (ii): In this case, we have that $d(x,t)\geq(c_1(l)+2(l+1))\varepsilon|\log\varepsilon|$. A similar argument to that used for (4.67) gives us that

$$P_{d(x,t)}\left[B_s\leq c_1(l)\varepsilon|\log\varepsilon|\right]\leq\varepsilon^{l+1}. \tag{4.70}$$

It follows that in this case

$$\begin{aligned}E_{d(x,t)}\left[g\left(\mathbb{Q}^{\varepsilon,t-s}_{B_s+K_1e^{K_2t}\varepsilon|\log\varepsilon|}\right)\right]&\geq E_{d(x,t)}\left[g\left(\mathbb{Q}^{\varepsilon,t-s}_{B_s+K_1e^{K_2t}\varepsilon|\log\varepsilon|}\right)\mathbb{1}\{B_s\geq c_1(l)\varepsilon|\log\varepsilon|\}\right]\\&\geq g(1-\varepsilon^l)-\varepsilon^{l+1}\\&\geq 1-3\varepsilon^{2l}-\varepsilon^{l+1},\end{aligned}$$

where the second line follows by Theorem 4.2.5 and (4.70) and the last line by the definition of g in (4.54). Again reducing $\varepsilon_{\mathbb{d}}$ if necessary, for $\varepsilon\in(0,\varepsilon_{\mathbb{d}})$ we have

$$E_{d(x,t)}\left[g(\mathbb{Q}^{\varepsilon,t-s}_{B_s+K_1e^{K_2t}\varepsilon|\log\varepsilon|})\right]\geq 1-\tfrac{3}{4}\varepsilon^l$$

and so (4.55) holds in this case.

Case (iii): We now turn to the case in which x is close to the interface. If the event A_x occurs, for $u\in[0,s]$ we have

$$\begin{aligned}|d(W_u,t-u)|&\leq|W_u-x|+|d(x,t)|+|d(x,t)-d(x,t-u)|\\&\leq(2c_1(l)+4(l+1)\mathbb{d}+K_1e^{K_2(t-s)})\varepsilon|\log\varepsilon|+V_0(l+1)\varepsilon^2|\log\varepsilon|,\end{aligned}$$

where the second line follows by (4.68). Reducing $\varepsilon_{\mathbb{d}}$ if necessary, for $\varepsilon\in(0,\varepsilon_{\mathbb{d}})$ we have

$$|d(W_u,t-u)|\leq(R+K_1e^{K_2(t-s)})\varepsilon|\log\varepsilon|, \tag{4.71}$$

where R is defined in (4.65). We now apply Proposition 4.2.13 with

$$\beta = (R + K_1 e^{K_2(t-s)})\varepsilon|\log\varepsilon|. \tag{4.72}$$

By reducing $\varepsilon_{\mathbb{d}}$ if necessary, we have for $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$ that $\beta \le c_0$. Define

$$T_\beta = \inf(\{u \in [0,t) : |d(W_u, t-u)| \ge \beta\} \cup \{t\}).$$

Then by Proposition 4.2.13, we can couple $(W_u)_{u\ge 0}$ with $(B_u)_{u\ge 0}$, a one-dimensional Brownian motion started from $d(x,t)$, in such a way that for $u \le T_\beta$,

$$d(W_u, t-u) \le B_u + C_0\beta u. \tag{4.73}$$

Hence

$$\begin{aligned} &E_x\left[g(\mathbb{Q}^{\varepsilon,t-s}_{d(W_s,t-s)+K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|} + \varepsilon^l)\right] \\ &\quad\le E_{d(x,t)}\left[g(\mathbb{Q}^{\varepsilon,t-s}_{B_s+C_0\beta s+K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|} + \varepsilon^l)\right] + P_x\left[T_\beta \le s\right] \\ &\quad\le E_{d(x,t)}\left[g(\mathbb{Q}^{\varepsilon,t-s}_{B_s+C_0\beta s+K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|} + \varepsilon^l)\right] + 4\mathbb{d}\varepsilon^{l+1}. \end{aligned} \tag{4.74}$$

Here, the first inequlaity follows by (4.73), (4.17) and the monotonicity of g. The second inequality then follows by (4.67) (note that by (4.71), if A_x occurs then $T_\beta \ge s$).

Now let

$$E = \left\{\left|\mathbb{Q}^{\varepsilon,t-s}_{B_s+C_0\beta s+K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|} - \tfrac{1}{2}\right| \le \tfrac{5}{12}\right\}.$$

We shall consider the cases E and E^c separately to bound the right hand side of (4.74).

Consider first when the event E occurs. Note that by the definition of β in (4.72),

$$\begin{aligned} &K_1e^{K_2t}\varepsilon|\log\varepsilon| - \left(C_0\beta s + K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|\right) \\ &\quad= \left(K_1e^{K_2(t-s)}(e^{K_2s} - 1 - C_0s) - C_0Rs\right)\varepsilon|\log\varepsilon| \\ &\quad\ge (K_1(K_2 - C_0) - C_0R)\,s\varepsilon|\log\varepsilon| \\ &\quad= c_1(1)s\varepsilon|\log\varepsilon|, \end{aligned} \tag{4.75}$$

where the inequality follows since $K_2 > 0$ and the last line follows by (4.66). Reducing $\varepsilon_{\mathbb{d}}$ if necessary so that $\varepsilon_{\mathbb{d}} < \min(\varepsilon_1(1), \frac{1}{24})$, for $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$ we can apply Corollary 4.2.12 with $z = B_s + C_0\beta s + K_1 e^{K_2(t-s)}\varepsilon|\log\varepsilon|$ and $w = z + c_1(1)s\varepsilon|\log\varepsilon| \leq B_s + K_1 e^{K_2 t}\varepsilon|\log\varepsilon|$ to give that

$$\mathbb{Q}^{\varepsilon,t-s}_{B_s+C_0\beta s+K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|}\mathbb{1}_E \leq (\mathbb{Q}^{\varepsilon,t-s}_{B_s+K_1e^{K_2t}\varepsilon|\log\varepsilon|} - \tfrac{1}{48}s)\mathbb{1}_E. \tag{4.76}$$

Finally, we consider the case when the event E^c occurs. Recall that $g(p) = 3p^2 - 2p^3$ for $p \in [0,1]$, so $g'(p) = 6p(1-p)$. Hence if $p, \delta \geq 0$ with either $p + \delta \leq \frac{1}{9}$ or $p \geq \frac{8}{9}$ then

$$g(p+\delta) \leq g(p) + \tfrac{2}{3}\delta. \tag{4.77}$$

Reducing $\varepsilon_{\mathbb{d}}$ if necessary so that $\frac{1}{12} + \varepsilon^l < \frac{1}{9}$ for $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$, we have

$$\begin{aligned} g\left(\mathbb{Q}^{\varepsilon,t-s}_{B_s+C_0\beta s+K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|} + \varepsilon^l\right)\mathbb{1}_{E^c} &\leq \left(g\left(\mathbb{Q}^{\varepsilon,t-s}_{B_s+C_0\beta s+K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|}\right) + \tfrac{2}{3}\varepsilon^l\right)\mathbb{1}_{E^c} \\ &\leq \left(g\left(\mathbb{Q}^{\varepsilon,t-s}_{B_s+K_1e^{K_2t}\varepsilon|\log\varepsilon|}\right) + \tfrac{2}{3}\varepsilon^l\right)\mathbb{1}_{E^c}, \end{aligned} \tag{4.78}$$

where the first line follows by (4.77) and the last line by (4.75) and monotonicity of g.

Putting (4.76) and (4.78) into (4.74),

$$\begin{aligned} &E_x\left[g(\mathbb{Q}^{\varepsilon,t-s}_{d(W_s,t-s)+K_1e^{K_2(t-s)}\varepsilon|\log\varepsilon|} + \varepsilon^l)\right] \\ &\quad\leq E_{d(x,t)}\left[g\left(\mathbb{Q}^{\varepsilon,t-s}_{B_s+K_1e^{K_2t}\varepsilon|\log\varepsilon|} - \tfrac{1}{48}s + \varepsilon^l\right)\mathbb{1}_E\right] \\ &\qquad\quad + E_{d(x,t)}\left[\left(g\left(\mathbb{Q}^{\varepsilon,t-s}_{B_s+K_1e^{K_2t}\varepsilon|\log\varepsilon|}\right) + \tfrac{2}{3}\varepsilon^l\right)\mathbb{1}_{E^c}\right] + 4\mathbb{d}\varepsilon^{l+1} \\ &\quad\leq E_{d(x,t)}\left[g\left(\mathbb{Q}^{\varepsilon,t-s}_{B_s+K_1e^{K_2t}\varepsilon|\log\varepsilon|}\right)\right] + \tfrac{2}{3}\varepsilon^l + \varepsilon^l\mathbb{1}_{s\leq 48\varepsilon^l} + 4\mathbb{d}\varepsilon^{l+1}, \end{aligned}$$

where the last inequality follows in the case $s \leq 48\varepsilon^l$ since $|g'(p)| \leq \frac{3}{2}$ for all $p \in [0,1]$. Reducing $\varepsilon_{\mathbb{d}}$, if necessary, so that $4\mathbb{d}\varepsilon^{l+1} \leq \frac{1}{12}\varepsilon^l$ and $48\varepsilon^l \leq \varepsilon^3$ for $\varepsilon \in (0, \varepsilon_{\mathbb{d}})$ completes the proof of (4.55).

The second statement of the lemma, equation (4.56), is proved by the same argument, considering $\{\mathbb{V}(\boldsymbol{B}(u)) = 0\}$ instead of $\{\mathbb{V}(\boldsymbol{B}(u)) = 1\}$ and using $d(W_u, t-u) \geq B_u - C_0\beta u$ for $u \leq T_\beta$ in place of (4.73). □

4.3 Proof of Theorem 4.1.8

In this section we turn to the proof of our central result, Theorem 4.1.8, which provides convergence, after suitable rescaling, of the SLFVS started from an appropriate initial condition to the indicator function of a region whose boundary evolves according to mean curvature flow. The proof mimics that of Theorem 4.1.3 in exploiting a dual process. However, because of genetic drift, in addition to branching, individuals in our dual process can coalesce. The duality relation will once again be with a historical process and expressed through a majority voting procedure.

4.3.1 A branching and coalescing dual for the SLFVS

We begin by describing the dual process of branching and coalescing lineages. It is driven by the same Poisson Point Process of 'events' that drives the SLFVS. Recall from (4.11) that Π^n is a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}^{\mathrm{d}} \times (0, \infty)$ with intensity measure

$$n dt \otimes n^{\beta} dx \otimes \mu^n(dr).$$

We also let

$$u_n = \frac{u}{n^{1-2\beta}}, \qquad \text{and} \qquad \boldsymbol{s}_n = \frac{1}{\varepsilon_n^2} \frac{1}{n^{2\beta}}.$$

Definition 4.3.1 (SLFVS dual)**.** *For $n \in \mathbb{N}$, the process $(\mathcal{P}_t^n)_{t \geq 0}$ is the $\bigcup_{l \geq 1} (\mathbb{R}^{\mathrm{d}})^l$-valued Markov process with dynamics defined as follows.*

The process is started with a single individual $\mathcal{P}_0^n = x$ and for $t \geq 0$, $\mathcal{P}_t^n = (\xi_1^n(t), \ldots, \xi_{N(t)}^n(t))$ for some $N(t) \in \mathbb{N}$. At each event $(t, x, r) \in \Pi^n$, independently of all else, the event is said to be neutral with probability $1 - \boldsymbol{s}_n$. In this case:

1. *For each $\xi_i^n(t-) \in \mathcal{B}_r(x)$, independently mark the corresponding individual with probability u_n;*
2. *if at least one individual is marked, all marked individuals coalesce into a single offspring individual, whose location is drawn uniformly at random from within $\mathcal{B}_r(x)$.*

Otherwise the event is said to be selective, in which case:

1. *For each* $\xi_i^n(t-) \in \mathcal{B}_r(x)$, *independently mark the corresponding individual with probability* u_n*;*
2. *if at least one individual is marked, all of the marked individuals are replaced by* three *offspring individuals, whose locations are drawn independently and uniformly from within* $\mathcal{B}_r(x)$*.*

In both cases, if no individual is marked, then nothing happens.

Remark 4.3.2. *We have referred to the new individuals created during reproduction events as 'offspring' individuals. From a biological perspective, it would perhaps be more natural to call them 'parents' or 'potential parents', as forwards in time they correspond to the locations from which alleles from the parental generation are sampled. However, as much of our proof of Theorem 4.1.3 will carry over with minimal changes to the SLFVS setting, we wish to retain the terminology of the branching Brownian motion of the previous section.*

The duality relation that we exploit is between the SLFVS and the *historical process* of branching and coalescing lineages,

$$\Xi^n(t) := (\mathcal{P}_s^n)_{0 \le s \le t}.$$

We write $\mathbb{P}_x$ for the law of Ξ^n when $\mathcal{P}_0^n$ is the single point x and $\mathbb{E}_x$ for the corresponding expectation. For $\boldsymbol{i} \in \{1,2,3\}^{\mathbb{N}}$ with $\boldsymbol{i} = (i_1, i_2, \ldots)$, we let $(\xi_{\boldsymbol{i}}^n(\cdot))_{0 \le s \le t} \subseteq \Xi(t)$ denote the $\mathbb{R}^{\mathrm{d}}$-valued path which jumps to the location of an offspring when the individual in $\mathcal{P}_s^n$ at its location is affected by an event, jumping to the i_k^{th} offspring when it is affected by its k^{th} selective event. We shall refer to $(\xi_{\boldsymbol{i}}^n(\cdot))_{0 \le s \le t}$ as an ancestral lineage.

The voting procedure on $\Xi^n(t)$ is a minor modification of Definition 4.2.1. Let $p : \mathbb{R}^{\mathrm{d}} \to [0,1]$ be a fixed function. Recalling that the set of individuals in $\mathcal{P}_t^n$ is $\{\xi_1^n(t), \ldots, \xi_{N(t)}^n(t)\}$, for each $j \le N(t)$, the individual $\xi_j^n(t)$ votes 1 with probability $p(\xi_j^n(t))$ and otherwise votes 0; votes from different individuals are independent. As we trace backwards in time through $\Xi(t)$,

1. at each neutral event, all individuals that are marked in the event adopt the vote of the offspring individual of the event;

2. at each selective event in Π^n, all individuals that are marked in the event adopt the majority vote of the votes of the three offspring individuals of the event.

This defines an iterative voting procedure, which runs inwards from the 'leaves' of $\Xi^n(t)$ to the ancestral individual $\emptyset$.

Definition 4.3.3 ($\mathbb{V}_p$)**.** *With the voting procedure described above, we define* $\mathbb{V}_p(\Xi^n(t))$ *to be the vote associated to the root* $\emptyset$*.*

At this point the duality relation between the SLFVS and $\Xi(t)$ is easy to guess. However, in order to write it down formally, we have to overcome the fact that the SLFVS will only be defined, as a function, Lebesgue a.e. and so we cannot necessarily define $w_t^n(x)$ for a fixed point $x \in \mathbb{R}^{\mathrm{d}}$. However, if, $\psi \in C(\mathbb{R}^{\mathrm{d}}) \cap L^1(\mathbb{R}^{\mathrm{d}})$, then the function

$$\int_{\mathbb{R}^{\mathrm{d}}} \psi(x) w_t^n(x) dx,$$

is well-defined.

Theorem 4.3.4. *The spatial* Λ*-Fleming-Viot process with selection driven by* Π^n*,* $(w_t^n(x), x \in \mathbb{R}^{\mathrm{d}})_{t\geq 0}$*, is dual to the historical process* $(\Xi^n(t))_{t\geq 0}$ *in the sense that for every* $\psi \in C(\mathbb{R}^{\mathrm{d}}) \cap L^1(\mathbb{R}^{\mathrm{d}})$*, we have*

$$\mathbb{E}_p\left[\int_{\mathbb{R}^{\mathrm{d}}} \psi(x) w_t^n(x)\, dx\right] = \int_{\mathbb{R}^{\mathrm{d}}} \psi(x) \mathbb{E}_x\left[\mathbb{V}_p\left(\Xi^n(t)\right)\right] dx = \int_{\mathbb{R}^{\mathrm{d}}} \psi(x) \mathbb{P}_x\left[\mathbb{V}_p\left(\Xi^n(t)\right) = 1\right] dx. \tag{4.79}$$

Remark 4.3.5. *Of course, we are abusing notation here: the expectations on the left and right of this equation are taken with respect to different measures. The subscripts on the expectations are the initial values for the processes on each side.*

To see that the result should be true, note that (if it is defined) $w_t^n(x)$ is the probability that an allele sampled from the population at the location x at time t is of type a. In order to determine that probability, we trace back until the most recent event that covered the location x. With probability u_n, the chosen allele was an offspring of the event, in which case its type can be determined if we know the types of the potential parents of the event. If the event is neutral, the type is that of an allele (the 'parent') sampled from a point picked uniformly at random from the affected region at the time of the event; if it is selective, then the type is the 'majority vote' of three 'potential parents' sampled uniformly at random from the affected region. In order to determine the types of the potential parents, we continue to trace backwards in time, following the locations of all potential ancestors until time zero. This gives us the dual process $\Xi^n(t)$. At that time, each potential ancestor samples its type according to the initial condition w_0 at its location. We can then determine $w_t^n(x)$ by working back through $\Xi^n(t)$ using our majority voting procedure.

A formal proof of Theorem 4.3.4 using generators is a simple extension of that of the corresponding duality for the spatial Λ-Fleming-Viot process with genic selection in Etheridge et al. (2014) (and indeed can be extended to cover the more general initial conditions for the dual process considered there) and so is omitted.

The duality reduces the proof of Theorem 4.1.8 to the following analogue of Theorem 4.2.3.

Theorem 4.3.6. *Take σ^2 as in (4.13). Suppose that $\beta \in (0, 1/4)$ and let ε_n be a sequence such that $\varepsilon_n \to 0$ and $(\log n)^{1/2}\varepsilon_n \to \infty$ as $n \to \infty$. Assume p satisfies ($\mathscr{C}1$)-($\mathscr{C}3$) and define $\mathscr{T}$, $d(x, t)$ as for Theorem 4.1.3; take $T^* < \mathscr{T}$. Let $k \in \mathbb{N}$. There exist $n_*(k) \in \mathbb{N}$, and $a_*(k), c_*(k) \in (0, \infty)$ such that for all $n \geq n^*$ and all t satisfying $a_*\varepsilon_n^2|\log \varepsilon_n| \leq t \leq T^*$,*

1. *for x such that $d(x, \sigma^2 t) \geq c_*\varepsilon_n|\log \varepsilon_n|$, we have $\mathbb{P}_x [\mathbb{V}_p(\Xi^n(t)) = 1] \geq 1 - \varepsilon_n^k$.*
2. *for x such that $d(x, \sigma^2 t) \leq -c_*\varepsilon_n|\log \varepsilon_n|$, we have $\mathbb{P}_x [\mathbb{V}_p(\Xi^n(t)) = 1] \leq \varepsilon_n^k$.*

Before providing a proof of this result, let us explain why it should be true.

First consider the motion of a single ancestral lineage $\xi_i^n(\cdot)$ in $\Xi^n(t)$. It evolves as a pure jump process which is homogeneous in both space and time. Write V_r for the volume of $\mathcal{B}_r(x)$. The rate at which the lineage jumps from y to $y + z$ can be written

$$m_n(dz) = nu_n n^{\mathrm{d}\beta} \int_0^{\mathcal{R}_n} \frac{V_r(0, z)}{V_r} \mu^n(dr)\, dz, \tag{4.80}$$

where $V_r(0, z)$ is the volume of $\mathcal{B}_r(0) \cap \mathcal{B}_r(z)$. To see this, by spatial homogeneity, we may take the lineage to be at the origin in $\mathbb{R}^{\mathrm{d}}$ before the jump, and then, in order for it to jump to z, it must be affected by an event that covers both 0 and z. If the event has radius r, then the volume of possible centres, x, of such events is $V_r(0,z)$ and so the intensity with which such a centre is selected is $n\,n^{\mathbb{d}\beta}V_r(0,z)\mu^n(dr)$. The parental location is chosen uniformly from the ball $\mathcal{B}_r(x)$, so the probability that z is chosen as the parental location is dz/V_r and the probability that our lineage is actually affected by the event is u_n. Combining these yields (4.80).

The total rate of jumps is

$$\begin{aligned}
\int_{\mathbb{R}^{\mathbb{d}}} m_n(dz) &= \int_0^{\mathcal{R}_n} nu_n\, n^{\mathbb{d}\beta}\frac{1}{V_r}\int_{\mathbb{R}^{\mathbb{d}}}\int_{\mathbb{R}^{\mathbb{d}}} \mathbb{1}_{|x|<r}\mathbb{1}_{|x-z|<r}dx\,dz\,\mu^n(dr)\\
&= \int_0^{\mathcal{R}_n} nu_n\, n^{\mathbb{d}\beta}V_r\mu^n(dr)\\
&= n^{2\beta}uV_1\int_0^{\mathcal{R}} r^d\mu(dr), \qquad (4.81)
\end{aligned}$$

and the size of each jump is $\Theta(n^{-\beta})$ and so it is no surprise that in the limit a single lineage will evolve according to a (time-changed) Brownian motion. To identify the diffusion constant, we calculate:

$$\begin{aligned}
\frac{1}{2\mathbb{d}}\int_{\mathbb{R}^{\mathbb{d}}}|z|^2 m_n(dz) = \frac{1}{2\mathbb{d}}\int_{\mathbb{R}^{\mathbb{d}}}|z|^2 nu_n\int_0^{\mathcal{R}_n} n^{\mathbb{d}\beta}\frac{V_r(0,z)}{V_r}\mu^n(dr)dz&\\
= \frac{u}{2\mathbb{d}}\int_0^{\mathcal{R}}\int_{\mathbb{R}^{\mathbb{d}}}|z|^2\frac{V_r(0,z)}{V_r}dz\mu(dr),& \quad (4.82)
\end{aligned}$$

which is precisely σ^2 from (4.13).

Note also that a lineage is affected by selective events at rate

$$\left(uV_1\int_0^{\mathcal{R}} r^d\mu(dr)\right)n^{2\beta}\boldsymbol{s}_n = \eta\varepsilon_n^{-2}, \qquad (4.83)$$

where $\eta = uV_1\int_0^{\mathcal{R}} r^d\mu(dr)$. Evidently, we can bound the total number of lineages in $\Xi^n(t)$ above by the total number in a process in which each lineage, independently, branches at rate $\eta\varepsilon_n^{-2}$. Since $\varepsilon_n^{-2} = o(\log n)$, this implies that for any $\delta > 0$, with high probability, there are $o(n^\delta)$ pairs of lineages in $\Xi^n(T^*)$. Each such pair is in the region affected by some event (neutral or selective) at most $\mathcal{O}(n)$ times in $[0,T^*]$ and so the chance that we see any coalescence events is $o(nu_n^2\,n^\delta)$ for any $\delta > 0$. Since $nu_n^2 = n^{4\beta-1}$

and $\beta \in (0, 1/4)$, for large n we do not expect to see any coalescence events before time T^*.

Combining the above, the dual is well approximated by a ternary branching Brownian motion with branching rate $\Theta(\varepsilon_n^{-2})$ and so it is natural to expect that an equivalent of Theorem 4.2.3 holds.

4.3.2 Majority voting in the SLFVS, for $\mathbb{d} \geq 2$

The rigorous proof of Theorem 4.1.8 closely follows that of Theorem 4.2.3. In Section 4.3.2.3, we focus on generation of the interface, which is proved in much the same way as Proposition 4.2.15. Then, in Section 4.3.2.4, we look at the propagation of the interface. We shall see that, since it essentially focusses on a single branching event, the argument of Section 4.2.4.2 is sufficiently flexible to adapt to the SLFVS setting.

First we present the additional arguments required in the SLFVS setting. These stem from the fact that ancestral lineages in the dual of the SLFVS follow jump processes (which, when the lineages are too close together, are dependent), and from the coalescence of ancestral lineages. In Section 4.3.2.1 we show that (in between selective events) the motion of a single ancestral lineage is approximately (time-changed) Brownian motion. Then, in Section 4.3.2.2, we show that, asymptotically, the three families of descendants of offspring created during a selective event evolve independently (conditional on their locations at birth).

Remark 4.3.7. *In Sections 4.2.2 and 4.2.4 we used subscripts to distinguish variables that played the same role in each section, but had different values; e.g.* δ_1 *in* (4.33) *and* $\delta_{\mathbb{d}}$ *in* (4.48). *The corresponding quantities in this section will be denoted with a subscript* $*$, *for example* δ_* *in* (4.90).

4.3.2.1 A single lineage

We begin the proof by showing that the trajectory of a single lineage is close to that of a Brownian motion. We follow what is now a familiar argument in the context of spatial Λ-Fleming-Viot processes (see for example Etheridge et al. (2015a)).

Let $(\xi^n(t), t \geq 0)$ be a pure jump process started at $x \in \mathbb{R}^{\mathrm{d}}$ with rate of jumps from y to $y+z$ given by the intensity measure $m^n(dz)$, and let $(W(t), t \geq 0)$ be a Brownian motion in $\mathbb{R}^{\mathrm{d}}$ started at x.

Lemma 4.3.8. *For $t > 0$ fixed, there is a coupling of W and ξ^n under which*

$$\mathbb{P}\left[\left|\xi^n(t) - W(\sigma^2 t)\right| \geq n^{-\beta/6}\right] = \mathcal{O}(n^{-\beta}(t \vee 1)).$$

Proof. For $i \geq 1$, let $X_i = \xi^n_{i/n^{2\beta}} - \xi^n_{(i-1)/n^{2\beta}}$. Then $X_1, X_2, \ldots$ are i.i.d. with a rotationally symmetric distribution and, by (4.82), $\mathbb{E}[|X_1|^2] = 2\mathrm{d}\sigma^2 n^{-2\beta}$. Moreover, by (4.81), the number of jumps made by ξ^n on the time interval $[0, n^{-2\beta}]$ is Poisson, with mean $\Theta(1)$, so since each jump has magnitude at most $2\mathcal{R}_n$, $\mathbb{E}\left[|X_1|^4\right] = \mathcal{O}(n^{-4\beta})$. Then by Skorohod's second embedding Theorem, see e.g. Billingsley (1995), there is a Brownian motion W started at x and a sequence $\upsilon_1, \upsilon_2, \ldots$ of stopping times such that setting $\upsilon_0 = 0$, $(\upsilon_i - \upsilon_{i-1})_{i \geq 1}$ are i.i.d. and

$$W(\upsilon_i) = \xi^n(i/n^{2\beta}), \quad \mathbb{E}[\upsilon_i - \upsilon_{i-1}] = \tfrac{1}{2\mathrm{d}}\mathbb{E}\left[|X_1|^2\right] = \sigma^2 n^{-2\beta}, \quad \mathbb{E}[(\upsilon_i - \upsilon_{i-1})^2] = \mathcal{O}(n^{-4\beta}).$$

It follows that $\mathbb{E}[\upsilon_{\lfloor tn^{2\beta}\rfloor}] = \sigma^2 \lfloor tn^{2\beta}\rfloor n^{-2\beta}$ and $\mathrm{Var}[\upsilon_{\lfloor tn^{2\beta}\rfloor}] = \mathcal{O}(tn^{-2\beta})$. Hence by Chebychev's inequality,

$$\mathbb{P}\left[|\upsilon_{\lfloor tn^{2\beta}\rfloor} - \sigma^2 t| \geq n^{-\beta/2}\right] = \mathcal{O}(tn^{-\beta}). \tag{4.84}$$

Now we have that

$$|\xi^n(t) - W(\sigma^2 t)| \leq |\xi^n(t) - \xi^n(\lfloor tn^{2\beta}\rfloor/n^{2\beta})| + |W(\upsilon_{\lfloor tn^{2\beta}\rfloor}) - W(\sigma^2 t)|. \tag{4.85}$$

To control the first term on the right hand side, observe that

$$\mathbb{P}\left[|\xi^n(t) - \xi^n(\lfloor tn^{2\beta}\rfloor/n^{2\beta})| \geq n^{-\beta/6}/2\right] \leq \mathbb{E}\left[|X_1|^2\right](n^{-\beta/6}/2)^{-2} = \mathcal{O}(n^{-5\beta/3}). \tag{4.86}$$

To control the second term on the right hand side of (4.85), let $Z \sim N(0,1)$, then

$$\begin{aligned}
&\mathbb{P}\Big[|W(v_{\lfloor tn^{2\beta}\rfloor}) - W(\sigma^2 t)| \geq n^{-\beta/6}/2\Big] \\
&\leq \mathbb{P}\left[|v_{\lfloor tn^{2\beta}\rfloor} - \sigma^2 t| \geq n^{-\beta/2}\right] \\
&\quad + \mathbb{P}\left[|v_{\lfloor tn^{2\beta}\rfloor} - \sigma^2 t| \leq n^{-\beta/2},\, |W(v_{\lfloor tn^{2\beta}\rfloor}) - W(\sigma^2 t)| \geq n^{-\beta/6}/2\right] \\
&\leq \mathbb{P}\left[\sup_{s\in[-n^{-\beta/2}, n^{-\beta/2}]} |W(s) - W(0)| \geq n^{-\beta/6}/2\right] + \mathcal{O}(tn^{-\beta}). \\
&\leq 4\,\mathbb{d}\,\mathbb{P}\left[\sqrt{2}n^{-\beta/4}Z \geq n^{-\beta/6}/2\mathbb{d}\right] + \mathcal{O}(tn^{-\beta}). \\
&= \mathcal{O}(\exp(-\tfrac{1}{8\mathbb{d}^2}n^{\beta/6})) + \mathcal{O}(tn^{-\beta}). \qquad (4.87)
\end{aligned}$$

Here, the second inequality follows by (4.84) and the third inequality follows by bounding the modulus of a $\mathbb{d}$-dimensional Brownian motion by the sum of the moduli of $\mathbb{d}$ one-dimensional Brownian motions and then using the reflection principle. Combining (4.86) and (4.87) with (4.85) completes the proof. □

Next, we need the asymptotic distribution of an ancestral lineage and its first branch time (that is the first time that it is affected by a selective event).

Corollary 4.3.9. *Let τ be the first branch time of Ξ^n. There is a coupling of Ξ^n and W under which τ and W are independent, $\tau \sim Exp(\eta\varepsilon_n^{-2})$ where $\eta = uV_1\int_0^{\mathcal{R}} r^d\mu(dr)$, and for $i = 1,2,3$,*

$$\mathbb{P}\left[|\xi_i^n(\tau) - W(\sigma^2\tau)| \geq 3n^{-\beta/6}\right] = \mathcal{O}(n^{-\beta}).$$

Proof. The distribution of τ follows immediately from (4.83).

Now consider any ancestral lineage $\xi^n \subseteq \Xi^n$. By the thinning property of Poisson processes, at any time $t > 0$, we can write $\xi^n(t) = \xi^{n,\mathtt{sel}}(t) + \xi^{n,\mathtt{neu}}(t)$, where $\xi^{n,\mathtt{sel}}$ and $\xi^{n,\mathtt{neu}}$ are independent pure jump processes with jump intensities $\boldsymbol{s}_n m_n(dz)$ and $(1-\boldsymbol{s}_n)m_n(dz)$ respectively, and taking τ to be the first jump time of $\xi^{n,\mathtt{sel}}$, $\xi^{n,\mathtt{neu}}$ is independent of τ. Using Lemma 4.3.8 with $(1-\boldsymbol{s}_n)m_n(dz)$ in place of $m_n(dz)$, we can couple $\xi^{n,\mathtt{neu}}$ with a Brownian motion W in such a way that for any $t > 0$,

$$\mathbb{P}[|\xi^{n,\mathtt{neu}}(t) - W(\sigma^2(1-\boldsymbol{s}_n)t)| \geq n^{-\beta/6}] \leq \mathcal{O}(n^{-\beta}(t \vee 1)).$$

Since $\boldsymbol{s}_n = o(\log n/n^{2\beta})$, using Chebyshev's inequality,

$$\mathbb{P}[|W(\sigma^2 t) - W(\sigma^2(1-\boldsymbol{s}_n)t)| \geq n^{-\beta/6}] = o\left(\frac{\log n}{n^{2\beta}} n^{\beta/3}(t \vee 1)\right),$$

and so using the triangle inequality

$$\mathbb{P}\left[|\xi^n(\tau-) - W(\sigma^2\tau)| \geq 2n^{-\beta/6} \middle| \tau\right] = \mathcal{O}(n^{-\beta}(\tau \vee 1)).$$

Since $\mathbb{E}[\tau] = \Theta(\varepsilon_n^2) = o(1)$, and for $i = 1, 2, 3$, $|\xi_i^n(\tau) - \xi_1^n(\tau-)| \leq 2\mathcal{R}_n = 2n^{-\beta}\mathcal{R}$ the result follows. ☐

4.3.2.2 Independence after branching

We now define a modification of $\Xi^n(t)$ which we denote by $\Psi^n(t)$ in which lineages evolve independently after branching (so, in particular, do not coalesce) and then show that $\Xi^n(t)$ and $\Psi^n(t)$ can be coupled in such a way that they coincide with high probability.

Definition 4.3.10 (Branching jump process)**.** *For given $n \in \mathbb{N}$ and starting point $x \in \mathbb{R}^{\mathrm{d}}$, $(\Psi^n(t), t \geq 0)$ is the historical process of the branching random walk which is described as follows.*

1. *Each individual has an independent exponential lifetime with parameter $\eta\varepsilon_n^{-2}$.*
2. *During its lifetime, each individual, independently, evolves according to a pure jump process with jump intensity $(1-\boldsymbol{s}_n)m_n(dz)$.*
3. *At the end of its lifetime an individual branches into three offspring.*
4. *The locations of the offspring are determined as follows. For each branching event, independently, pick $r \in (0, \mathcal{R}_n]$ according to $r^{\mathrm{d}}\mu^n(dr)/\int_0^{\mathcal{R}_n} s^{\mathrm{d}}\mu^n(ds)$. If the parent is at the point $z \in \mathbb{R}^{\mathrm{d}}$, then each of the three offspring, independently, samples its location uniformly from $B_r(z)$.*

Remark 4.3.11. *Note that the only difference between the distributions of Ξ^n and Ψ^n is that in Ψ^n, lineages evolve independently after branching, whereas in Ξ^n, two distinct lineages may be hit by the same event in Π_n.*

We define $\mathbb{V}_p(\Psi^n(t))$ in the usual way (as in Definition 4.2.1): a leaf at location $\psi_i(t) \in \mathbb{R}^{\mathrm{d}}$ votes 1 with probability $p(\psi_i(t))$, otherwise it votes zero, and votes from different leaves are independent; working back through the tree an individual adopts the vote of the majority of its offspring and $\mathbb{V}_p(\Psi^n(t))$ is the resultant vote at the root.

Lemma 4.3.12. *Let $T^* \in (0, \infty)$, $k \in \mathbb{N}$ and $z \in \mathbb{R}^{\mathrm{d}}$. There exists $n_* \in \mathbb{N}$ such that for all $n \geq n_*$, there is a coupling of Ξ^n started from z and Ψ^n started from z such that with probability at least $1 - \varepsilon_n^k$ we have*

$$\Xi^n(T^*) = \Psi^n(T^*).$$

The remainder of this section is devoted to the proof of Lemma 4.3.12. To do so, we consider a slightly different description of the dual of the SLFVS, which will preserve the distribution of Ξ^n.

Definition 4.3.13 (Pre-emptive SLFVS dual)**.** *For $n \in \mathbb{N}$, the process $(\tilde{\mathcal{P}}_t^n)_{t \geq 0}$ is a $\bigcup_{l \geq 1} (\mathbb{R}^{\mathrm{d}})^l$-valued process of individuals, each of which may be marked. The dynamics are described as follows.*

The process is started with a single individual at the point x; we write $(\xi_1^n(t), \ldots, \xi_{N(t)}^n(t))$ for the locations of the random number $N(t)$ of individuals at time t.

At time zero, independently of all else, the individual $\xi_1^n(0)$ is marked with probability u_n.

At each event $(t, x, r) \in \Pi^n$, independently, the event is said to be neutral with probability $1 - \boldsymbol{s}_n$. In this case:

1. *if at least one individual $\xi_i^n(t-) \in \mathcal{B}_r(x)$ is marked, then all marked individuals in $\mathcal{B}_r(x)$ are replaced by a single offspring individual, whose location is drawn uniformly at random from within $\mathcal{B}_r(x)$;*
2. *for each $\xi_i^n(t) \in \mathcal{B}_r(x)$, including the offspring individual if any, independently mark the corresponding individual with probability u_n and unmark it otherwise.*

Otherwise the event is said to be selective, in which case:

1. *if at least one individual* $\xi_i^n(t-) \in \mathcal{B}_r(x)$ *is marked, the collection of marked individuals in* $\mathcal{B}_r(x)$ *is replaced by* three *offspring individuals, whose locations are drawn independently and uniformly from within* $\mathcal{B}_r(x)$*;*
2. *for each* $\xi_i^n(t) \in \mathcal{B}_r(x)$*, including the offspring individuals if any, independently mark the corresponding individual with probability* u_n *and unmark it otherwise.*

In between events in Π_n*, nothing happens. In particular, once marked, an individual remains marked until it is in the region covered by an event, and, during events, all individuals in the affected region (whether they were marked before the event of not) sample afresh from independent Bernoulli random variables to decide whether they are marked immediately after the event.*

In the same way as we defined Ξ^n, ignoring marks, we write Φ^n for the historical process corresponding to the pre-emptive dual. The distribution of Φ^n is equal to that of Ξ^n. The only difference between Definition 4.3.1 and Definition 4.3.13 is that, for each reproduction event, whether or not a individual that lies in the affected region is marked for reproduction was determined at the time of the previous reproduction event that affected a region in which it lies. Notice that for both neutral and selective events, even if no individual is marked at time $t-$, all individuals in $\mathcal{B}_r(x)$ at time t (after the reproduction has taken place), independently, renew their status as marked or unmarked.

The key observation that will allow us to couple Ξ^n (or equivalently Φ^n) and Ψ^n is that for as long as two ancestral lineages are not both marked, they evolve independently.

Lemma 4.3.14. *Let* $T^* \in (0, \infty)$*. There exists* $\alpha > 0$ *such that*

$$\mathbb{P}\left[\exists \xi_i^n \neq \xi_j^n \subseteq \Phi^n(T^*), t \in [0, T^*] \text{ s.t. } \xi_i^n \text{ and } \xi_j^n \text{ are both marked at time } t\right] = \mathcal{O}(n^{-\alpha}).$$

Proof. Write $\mathcal{T}(\Phi^n(t))$ for the genealogy of $\Phi^n(t)$. We begin by showing that for any constant $b > 0$, $\mathcal{T}(\Phi^n(T^*)) \subseteq \mathcal{T}^{\mathrm{reg}}_{b \log n}$ with high probability. Recall from (4.83) that the rate at which each lineage is affected by reproduction events is $\eta\varepsilon_n^{-2} = o(\log n)$. Let M^n be a Poisson distributed random variable with mean $T^*\eta\varepsilon_n^{-2}$. Recall that if Z' is

Poisson with parameter χ, then (using a Chernoff bound) for $k > \chi$,

$$\mathbb{P}[Z' > k] \leq \frac{e^{-\chi}(e\chi)^k}{k^k}. \tag{4.88}$$

Hence for $b > 0$ a constant, applying (4.88) with $k = b\log n$ and $\chi = T^*\eta\varepsilon_n^{-2} = o(\log n)$, taking n sufficiently large that $\frac{e\chi}{b\log n} \leq 3^{-3}$, we have

$$\mathbb{P}\left[M^n > b\log n\right] \leq 3^{-2b\log n - 1}.$$

Then by a union bound over each root to leaf ray of $\mathcal{T}^{\text{reg}}_{b\log n}$,

$$\mathbb{P}\left[\mathcal{T}(\Phi^n(T^*)) \not\subseteq \mathcal{T}^{\text{reg}}_{b\log n}\right] \leq 3^{b\log n+1}\mathbb{P}\left[M^n > b\log n\right] \leq 3^{-b\log n}. \tag{4.89}$$

Given a particular pair of lineages, $\xi_i^n, \xi_j^n \subseteq \Phi^n(t)$, we want to bound above the probability that a reproduction event occurs during $[0, T^*]$ after which both are marked. The first time that this happens, at least one of ξ_i^n and ξ_j^n must be in the region affected by the event. After the event, the probability that both lineages are marked is u_n^2 (irrespective of whether the second lineage was also in the affected region). The number of reproduction events before time T^* with region containing ξ_i^n is Poisson with mean $\Theta(n)$. Hence, the probability that a given pair ξ_i^n, ξ_j^n are both marked at some time $t \in [0, T^*]$ is $\mathcal{O}(nu_n^2) = \mathcal{O}(n^{4\beta-1})$.

Using a union bound over pairs of lineages, we have

$$\begin{aligned}
&\mathbb{P}\left[\exists \xi_i^n \neq \xi_j^n \subseteq \Phi^n(T^*) \text{ and } t \in [0, T^*] \text{ such that } \xi_i^n \text{ and } \xi_j^n \text{ are both marked at time } t\right] \\
&\qquad \leq 3^{-b\log n} + 3^{2b\log n}\mathcal{O}(n^{4\beta-1}) \\
&\qquad \leq 3^{-b\log n} + \mathcal{O}\left(\exp\left(2b(\log 3)(\log n) + (4\beta-1)\log n\right)\right).
\end{aligned}$$

Noting that $4\beta - 1 < 0$ and choosing b such that $2b(\log 3) + (4\beta - 1) < 0$ gives the required result. ☐

Proof of Lemma 4.3.12. Let

$$\tau = \inf\{t \geq 0 : \exists \xi_i^n \neq \xi_j^n \subseteq \Phi^n(T^*) \text{ such that } \xi_i^n \text{ and } \xi_j^n \text{ are both marked at time } t\}.$$

Noting that for any $k \in \mathbb{N}$ and any $\alpha > 0$ we have $n^{-\alpha} = o((\log n)^{-k/2}) = o(\varepsilon_n^k)$, by Lemma 4.3.14, for n sufficiently large, $\mathbb{P}[\tau \geq T^*] \geq 1 - \varepsilon_n^k$. For as long as ancestral lineages in Φ^n are not both marked they evolve independently, so we may couple $(\Phi^n(t))$ and $(\Psi^n(t))$ to be equal up until time τ and the result follows. $\square$

4.3.2.3 Generation of the interface

In this section we show that, in analogy to Proposition 4.2.15, the interface is generated in time of order $\varepsilon_n^2|\log \varepsilon_n|$. The proof is similar to that of Proposition 4.2.15.

Proposition 4.3.15. *Let $k \in \mathbb{N}$. Then there exist $n_*(k), a_*(k), d_*(k) > 0$ such that, for all $n \geq n_*$, if we set*

$$\delta_*(k,n) := a_*(k)\varepsilon_n^2|\log \varepsilon_n| \quad \textit{and} \quad \delta_*'(k,n) := (a_*(k) + \eta^{-1}(k+1))\varepsilon_n^2|\log \varepsilon_n|, \tag{4.90}$$

then for $t \in [\delta_*, \delta_*']$,

1. *for x such that $d(x, \sigma^2 t) \geq d_*\varepsilon|\log \varepsilon|$, we have $\mathbb{P}_x\left[\mathbb{V}_p(\Xi^n(t)) = 1\right] \geq 1 - \varepsilon_n^k$;*
2. *for x such that $d(x, \sigma^2 t) \leq -d_*\varepsilon|\log \varepsilon|$, we have $\mathbb{P}_x\left[\mathbb{V}_p(\Xi^n(t)) = 1\right] \leq \varepsilon_n^k$.*

Using the coupling from Lemma 4.3.12, it suffices to prove the result for the branching jump process $\Psi^n(t)$ in place of $\Xi^n(t)$. For this we exploit the following lemma.

Lemma 4.3.16. *Let $k \in \mathbb{N}$ and let $A(k)$ be chosen as in Lemma 4.2.8. There exist $a_*(k), B_*(k) \in (0,\infty)$, and $n_*(k) < \infty$ such that for all $n \geq n_*$ and δ_*, δ_*' as defined in* (4.90)*,*

$$\mathbb{P}\left[\mathcal{T}(\Psi^n(\delta_*)) \supseteq \mathcal{T}^{reg}_{A(k)|\log \varepsilon_n|}\right] \geq 1 - \varepsilon_n^k, \tag{4.91}$$

$$\textit{and} \qquad \mathbb{P}\left[\mathcal{T}(\Psi^n(\delta_*')) \subseteq \mathcal{T}^{reg}_{B_*(k)|\log \varepsilon_n|}\right] \geq 1 - \varepsilon_n^k. \tag{4.92}$$

Remark 4.3.17. *During the proof of Proposition 4.2.15, we deduced* (4.49)*, which is the equivalent of* (4.91)*. We did not require an equivalent of* (4.92)*. We shall use* (4.92) *here in order to prove the equivalent of* (4.50)*.*

Proof. Recall from (4.83) that a given ancestral lineage in Ψ^n branches into three after an exponential time with rate $\eta\varepsilon_n^{-2}$. Hence, (4.91) follows for a_* sufficiently large by the same proof as Lemma 4.2.9.

The proof of (4.92) is the same as that of (4.89). Let L^n be a Poisson distributed random variable with mean $\delta_*'\eta\varepsilon_n^{-2} = (a_* + \eta^{-1}(k+1))\eta|\log\varepsilon_n|$. Take $B_* = B_*(k)$ sufficiently large that $B_* \geq (a_* + \eta^{-1}(k+1))\eta$ and

$$e(a_* + \eta^{-1}(k+1))\eta B_*^{-1} < \frac{1}{3}e^{-k/B_*-1}. \tag{4.93}$$

The Chernoff bound (4.88) gives

$$\begin{aligned}\mathbb{P}\left[L^n > B_*|\log\varepsilon_n|\right] &\leq \left(e(a_* + \eta^{-1}(k+1))\eta B_*^{-1}\right)^{B_*|\log\varepsilon_n|}\\ &\leq \varepsilon^k 3^{-B_*|\log\varepsilon_n|-1},\end{aligned} \tag{4.94}$$

and, taking a union bound over each root to leaf ray of $\mathcal{T}^{\mathrm{reg}}_{B_*|\log\varepsilon_n|}$,

$$\mathbb{P}\left[\mathcal{T}(\Psi^n(\delta_*')) \not\subseteq \mathcal{T}^{\mathrm{reg}}_{B_*(k)|\log\varepsilon_n|}\right] \leq 3^{B_*|\log\varepsilon_n|+1}\mathbb{P}\left[L^n > B_*|\log\varepsilon_n|\right] \leq \varepsilon_n^k,$$

which completes the proof. □

Proof of Proposition 4.3.15. We prove this result with Ψ^n in place of Ξ^n (from which the result follows using Lemma 4.3.12). The approach closely follows that of Proposition 4.2.15 except that now we have to control the distance between the jump process followed by a lineage and Brownian motion.

Take a_* from Lemma 4.3.16, and $t \in [\delta_*, \delta_*']$. Let $(\xi^n(t), t \geq 0)$ be a pure jump process with rate of jumps from y to $y + z$ given by the intensity measure $m^n(dz)$. By Lemma 4.3.8 we can couple $(\xi^n(t), t \geq 0)$ with a $\mathbb{d}$-dimensional Brownian motion $(W(t), t \geq 0)$ in such a way that $\xi^n(0) = W(0)$ and

$$\mathbb{P}\left[|\xi^n(t) - W(\sigma^2 t)| \geq n^{-\beta/6}\right] = \mathcal{O}(n^{-\beta}).$$

For $d_*(k)$ a constant, for large enough n, since $\varepsilon_n^{-2} = o(\log n)$ we have $\frac{1}{2}d_*\varepsilon_n|\log\varepsilon_n| \geq 2n^{-\beta/6}$. Hence, for such n,

$$\begin{aligned}\mathbb{P}\left[|\xi^n(t) - \xi^n(0)| \geq \tfrac{1}{2}d_*\varepsilon_n|\log\varepsilon_n|\right] &\leq \mathbb{P}\left[|\xi^n(t) - W(\sigma^2 t)| \geq n^{-\beta/6}\right] \\ &\quad + \mathbb{P}\left[|W(\sigma^2\delta_*'(k,n))) - W(0)| \geq \tfrac{1}{4}d_*\varepsilon_n|\log\varepsilon_n|\right] \\ &\leq \mathcal{O}(n^{-\beta}) + 2\mathbb{d}\exp\left(-\frac{1}{64}\frac{d_*^2}{\sigma^2\mathbb{d}^2(a_* + \eta^{-1}(k+1))}|\log\varepsilon_n|\right) \\ &\leq 3^{-B_*|\log\varepsilon_n|}\varepsilon_n^k.\end{aligned}$$

Here the second inequality follows by bounding the modulus of a $\mathbb{d}$-dimensional Brownian motion by the sum of the moduli of $\mathbb{d}$ one-dimensional Brownian motions, and the last inequality follows for d_* sufficiently large. Using (4.92) and taking a union bound over the root to leaf rays of $\mathcal{T}_{B_*|\log\varepsilon_n|}$, for $t \in [\delta_*, \delta_*']$,

$$\begin{aligned}\mathbb{P}_x\left[\exists\xi_{\boldsymbol{i}}^n \subseteq \Psi^n(\delta_*') \text{ s.t. } |\xi_{\boldsymbol{i}}^n(t) - x| \geq \tfrac{1}{2}d_*\varepsilon_n|\log\varepsilon_n|\right] &\leq \varepsilon_n^k + 3^{B_*|\log\varepsilon_n|}3^{-B_*|\log\varepsilon_n|}\varepsilon_n^k \\ &\leq 2\varepsilon_n^k. \end{aligned} \tag{4.95}$$

Combining (4.95) with Lemma 4.3.16, we obtain that, with probability $\geq 1 - 3\varepsilon_n^k$,

1. $\mathbb{V}_p(\Psi^n(t))$ is given by independent votes at each of the leaves of $\mathcal{T}(\Psi^n(t))$.
2. $\mathcal{T}(\Psi^n(t)) \supseteq \mathcal{T}^{\text{reg}}_{A|\log\varepsilon_n|}$ and the positions of the individuals corresponding to the leaves of $\mathcal{T}(\Psi^n(t))$ are all within $\frac{1}{2}d_*\varepsilon_n|\log\varepsilon_n|$ of their starting position.

Just as in the proof of Proposition 4.2.15 we obtain Proposition 4.3.15 with Ψ^n in place of Ξ^n. An application of Lemma 4.3.12 completes the proof. ☐

4.3.2.4 Propagation of the interface

We require the following slight modification of Lemma 4.2.17.

Lemma 4.3.18. *Let $l \in \mathbb{N}$ with $l \geq 4$ and $K_1 > 0$. There exists $K_2 = K_2(K_1, l) > 0$ and $n_*(l, K_1, K_2) > 0$ such that for all $n \geq n_*$, $x \in \mathbb{R}^{\mathbb{d}}$, $s \in [\sigma^2\varepsilon_n^{l+3}, \sigma^2(l+1)\eta^{-1}\varepsilon_n^2|\log\varepsilon_n|]$ and $t \in [s, \sigma^2 T^*]$,*

$$E_x\left[g\left(\mathbb{P}^{\varepsilon_n}_{d(W_s,t-s)+K_1e^{K_2(t-s)}\varepsilon_n|\log\varepsilon_n|+3n^{-\beta/6}}[\mathbb{V}(\boldsymbol{B}(t-s)) = 1] + \varepsilon_n^l\right)\right]$$

$$\le \tfrac{3}{4}\varepsilon_n^l + E_{d(x,t)}\left[g\left(\mathbb{P}^{\varepsilon_n}_{B_s+K_1e^{K_2t}\varepsilon_n|\log\varepsilon_n|}[\mathbb{V}(\boldsymbol{B}(t-s))=1]\right)\right] + \mathbb{1}_{s\le\varepsilon_n^3}\varepsilon_n^l, \tag{4.96}$$

and

$$E_x\left[g\left(\mathbb{P}^{\varepsilon_n}_{d(W_s,t-s)-K_1e^{K_2(t-s)}\varepsilon_n|\log\varepsilon_n|-3n^{-\beta/6}}[\mathbb{V}(\boldsymbol{B}(t-s))=0]+\varepsilon_n^l\right)\right]$$
$$\le \tfrac{3}{4}\varepsilon_n^l + E_{d(x,t)}\left[g\left(\mathbb{P}^{\varepsilon_n}_{B_s-K_1e^{K_2t}\varepsilon_n|\log\varepsilon_n|}[\mathbb{V}(\boldsymbol{B}(t-s))=0]\right)\right] + \mathbb{1}_{s\le\varepsilon_n^3}\varepsilon_n^l. \tag{4.97}$$

Proof. The proof is essentially the same as that of Lemma 4.2.17. Let $R = 2c_1(l) + 4\sigma\eta^{-1/2}(l+1)\mathbb{d} + 1$ and fix K_2 such that $K_1(K_2 - C_0) - C_0R = 2c_1(1)$; let

$$A_x = \left\{\sup_{u\in[0,s]}|W_u - x| \le 2\sigma\eta^{-1/2}(l+1)\,\mathbb{d}\varepsilon|\log\varepsilon|\right\}.$$

The proof for $|d(x,t)| \ge (2c_1(l) + 2\sigma\eta^{-1/2}(l+1)\mathbb{d} + K_1e^{K_2(t-s)})\varepsilon_n|\log\varepsilon_n|$ is then the same as in the proof of Lemma 4.2.17 (since $n^{-\beta/6} = o(\varepsilon_n|\log\varepsilon_n|)$).

Since $n^{-\beta/6} = o(s\varepsilon_n|\log\varepsilon_n|)$, we have for $\beta = (R + K_1e^{K_2(t-s)})\varepsilon|\log\varepsilon|$ as in (4.72), for n sufficiently large

$$K_1e^{K_2t}\varepsilon_n|\log\varepsilon_n| - (C_0\beta s + K_1e^{K_2(t-s)}\varepsilon_n|\log\varepsilon_n| + 3n^{-\beta/6}) \ge c_1(1)s\varepsilon_n|\log\varepsilon_n|. \tag{4.98}$$

Using (4.98) in place of (4.75), the proof for $|d(x,t)| \le (2c_1(l) + 2\sigma\eta^{-1/2}(l+1)\mathbb{d} + K_1e^{K_2(t-s)})\varepsilon_n|\log\varepsilon_n|$ is the same as in the proof of Lemma 4.2.17. □

The equivalent of Proposition 4.2.16 for Ψ^n is as follows.

Proposition 4.3.19. *Let $l \in \mathbb{N}$ with $l \ge 4$. Define $a_*(l)$ and $\delta_*(l,n)$ as in Proposition 4.3.15. There exist $K_1(l), K_2(l) > 0$ and $n_*(l, K_1, K_2) > 0$ such that for all $n \ge n_*$ and $t \in [\delta_*(l,n), T^*]$ we have*

$$\sup_{x\in\mathbb{R}^{\mathbb{d}}}\left(\mathbb{P}_x\left[\mathbb{V}_p(\boldsymbol{\Psi}^n(t))=1\right] - \mathbb{P}^{\varepsilon_n}_{d(x,\sigma^2t)+K_1e^{K_2\sigma^2t}\varepsilon_n|\log\varepsilon_n|}\left[\mathbb{V}(\boldsymbol{B}(\sigma^2t))=1\right]\right) \le \varepsilon_n^l$$

and

$$\sup_{x\in\mathbb{R}^{\mathrm{d}}}\left(\mathbb{P}_x\left[\mathbb{V}_p(\boldsymbol{\Psi}^n(t))=0\right]-\mathbb{P}^{\varepsilon_n}_{d(x,\sigma^2t)-K_1e^{K_2\sigma^2t}\varepsilon_n|\log\varepsilon_n|}\left[\mathbb{V}(\boldsymbol{B}(\sigma^2t))=0\right]\right)\leq\varepsilon_n^l.$$

Proof. The proof exactly follows that of Proposition 4.2.16, with Corollary 4.3.9 and then Lemma 4.3.18 in place of Lemma 4.2.17, and Proposition 4.3.15 in place of Proposition 4.2.15. □

Proof of Theorem 4.3.6. It suffices to prove the result for sufficiently large $k\in\mathbb{N}$, and in particular we will show it for $k\geq 5$. By Lemma 4.3.12, for n sufficiently large and $t\in[0,T^*]$,

$$|\mathbb{P}_x\left[\mathbb{V}_p(\boldsymbol{\Psi}^n(t))=1\right]-\mathbb{P}_x\left[\mathbb{V}_p(\boldsymbol{\Xi}^n(t))=1\right]|\leq\varepsilon_n^{k+1}.$$

The result now follows from Proposition 4.3.19 with $l=k+1$, in the same way as in the proof of Theorem 4.2.3. □

Chapter 5

Conclusion

5.1 Asymmetric hybrid zones

In Chapter 4, to study the behaviour of hybrid zones we considered a diploid population in which individuals are of type aa, aA or AA, with relative fitness 1, $1-s$ and 1 respectively (for some small $s>0$). This assumes that the homozygote aa and AA individuals have equal fitness (at least up to first order in s; recall that we neglect $\mathcal{O}(s^2)$ terms when modelling this situation). A natural question to ask is what happens when the relative fitnesses are as follows:

$$\begin{array}{ccc} aa & aA & AA \\ 1-s_1 & 1-s_2 & 1, \end{array}$$

where $0<s_1<s_2$. We can model this using the SLFV in a similar way to Definition 4.1.5. It turns out that in addition to the neutral events and the selective events with three potential parents in Definition 4.1.5, we need to include events at which two potential parental types are chosen and the offspring are of type a if and only if both parental types are a. This results in a dual process in which particles can branch into two or three particles.

In this case, taking a rescaling limit in the same way as in Theorem 4.1.7, up to a constant time and space scaling the limit is given by

$$\frac{\partial v}{\partial t}=\frac{1}{2}\Delta v+\frac{1}{\varepsilon^2}v(1-v)(2cv-1),$$

where $c \in (1/2, 1)$. In solutions to this PDE, for small $\varepsilon > 0$, Allen and Cahn observed that an interface forms, which propagates on the timescale $s = t/\varepsilon$ at a constant rate along its normal towards the domain in which $v \approx 0$ (this is made rigorous in Chen (1992)).

We would like to prove an analogous result to Theorem 4.1.8 on the motion of hybrid zones in this regime, and also to study the fluctuations around the limiting deterministic process in both the symmetric and asymmetric regimes; this will be the subject of future joint work with Alison Etheridge and Nic Freeman.

5.2 Drift load with spatial structure

In joint work with Raphael Forien, we considered the case of *overdominance* in a diploid population for a gene with two alleles a and A; overdominance occurs when heterozygote aA individuals have a higher fitness than homozygote aa and AA individuals. We suppose that the relative fitnesses are as follows:

$$\begin{array}{ccc} aa & aA & AA \\ 1-s_1 & 1 & 1-s_2, \end{array} \tag{5.1}$$

where s_1, $s_2 > 0$. In this situation, in an infinite population we expect stable non-zero frequencies of a-alleles and A-alleles to be maintained in the population.

In a well-mixed population, if $\bar{w}$ is the frequency of a-alleles, then the proportion of aa, aA and AA individuals will be $\bar{w}^2$, $2\bar{w}(1-\bar{w})$ and $(1-\bar{w})^2$ respectively, so the mean fitness of the population must be lower than the fitness of aA individuals. In particular, the mean fitness is given by

$$1 - s_1\bar{w}^2 - s_2(1-\bar{w})^2 = 1 - \frac{s_1 s_2}{s_1+s_2} - (s_1+s_2)\left(\bar{w} - \frac{s_2}{s_1+s_2}\right)^2. \tag{5.2}$$

The term $\frac{s_1 s_2}{s_1+s_2}$ which gives the minimum possible reduction in mean fitness is called the segregation load.

Robertson (1970) studied the expected mean fitness of a population of N individuals with no spatial structure and relative fitnesses given by (5.1). If $p^N(t)$ denotes the proportion of a-alleles in the population at time t, the expectation of the second term on the right hand side of (5.2) is given by

$$\Delta^N(t) = (s_1 + s_2)\mathbb{E}\left[\left(p^N(t) - \frac{s_2}{s_1 + s_2}\right)^2\right],$$

which is known as the drift load. Robertson found that irrespective of s_1 and s_2, $\Delta^N(t)$ converges to $1/4N$ as $t \to \infty$. This is due to a trade-off between genetic drift and natural selection. The stronger selection is, the quicker the allele frequency is pushed back to the equilibrium, but at the same time even a small step away from the optimal frequency is very costly in terms of mean fitness. On the other hand, if natural selection is relatively weak, the allele frequency can wander off more easily, but the mean fitness of the population decreases more slowly.

In Forien and Penington (2015), we study the same effect in a spatially structured population using the SLFV. We find that spatial structure significantly reduces the drift load, and that the drift load in a spatially structured population depends on the strength of selection. Moreover, we find that populations living in a space with a higher dimension have a reduced drift load compared to populations evolving in smaller dimensions. This result illustrates the fact that migration prevents the allele frequencies from straying too far away from the equilibrium frequency, because incoming migrants are on average close to this equilibrium; in higher dimensions, this effect is stronger.

In contrast to the proofs in Chapters 3-4, here the proofs do not use the SLFV dual process. Instead we prove a central limit theorem for the SLFV with a general class of selection mechanisms using martingale problem arguments. In the special case of the SLFV with genic selection as in Definition 1.2.16, Etheridge et al. (2014) shows that in dimension $d \geq 2$, in a regime in which both the impact parameter and the selection parameter tend to zero, by rescaling time and space we obtain a deterministic limiting process.

When the reproduction events have bounded radius, the limiting process is given by the Fisher KPP equation,

$$\frac{\partial f_t}{\partial t} = \frac{1}{2}\Delta f_t - s f_t(1 - f_t). \tag{5.3}$$

We prove a slightly stronger form of convergence to this deterministic rescaling limit. We also study the fluctuations of the allele frequency; since the convergence to the deterministic limit is too slow we study fluctuations about a deterministic approximation of $(f_t)_{t\geq 0}$. We find that if the impact parameter is sufficiently small compared to the selection parameter and the fluctuations are rescaled in the right way then in the limit they solve the following stochastic partial differential equation,

$$dz_t = \left[\frac{1}{2}\Delta z_t - s(1-2f_t)z_t\right] dt + \sqrt{f_t(1-f_t)}\, dW_t, \tag{5.4}$$

where W is space-time white noise, and f is the solution of (5.3).

We also consider a second regime for the SLFVS to allow large scale extinction-recolonization events; we let the radius of reproduction events follow an α-stable distribution truncated at zero. For this regime, as in Etheridge et al. (2014), we find the Fisher KPP equation with non-local diffusion as a rescaling limit (i.e. with a fractional Laplacian instead of the usual Laplacian). The Laplacian is also replaced by a fractional Laplacian in (5.4), the equation satisfied by the limiting fluctuations, and the noise W becomes a coloured noise with spatial correlations of order $|x-y|^{-\alpha}$.

5.3 Further results on BBM with decay of mass

Recall the definition of BBM with decay of mass in Chapter 2. In joint work with Louigi Addario-Berry, we have proved some further results on BBM with decay of mass. The results rely on the following bound on the largest particle mass at time t.

Proposition 5.3.1. *For any $\alpha < 1$, there exists a constant $a > 0$ such that for t sufficiently large,*

$$\mathbb{P}\left[\max_{1\leq i\leq N(t)} M_i(t) \geq at^{-\alpha}\right] \leq 2t^{-4}.$$

Recall that in Proposition 2.4.1 we showed that for t sufficiently large, with high probability, $\sup_{s\leq t}\sup_{x\in\mathbb{R}} \zeta(s,x) \leq Z \log t$ where Z is a constant. Using Proposition 5.3.1, we can strengthen the results in Section 2.4 and show that the maximum local mass density is bounded.

Proposition 5.3.2. *There exists a constant $C < \infty$ such that for t sufficiently large,*

$$\mathbb{P}\left[\sup_{s \geq t} \sup_{x \in \mathbb{R}} \zeta(s,x) \geq C\right] \leq Ct^{-3}.$$

By Borel-Cantelli, the following is a direct corollary.

Corollary 5.3.3. *Almost surely,* $\limsup_{t \to \infty} \sup_{x \in \mathbb{R}} \zeta(t,x) \leq C$.

We can also use Proposition 5.3.1 to show that it takes time at most $\mathcal{O}(\log t)$ for $d(t,m)$ to catch up with $D(t,m)$.

Theorem 5.3.4. *There exists $m^* > 0$ such that for $m \leq m^*$, there is $R(m) < \infty$ a constant and a random time $T(m) < \infty$ a.s. such that for $t \geq T$,*

$$\inf_{s \geq 0} d(t + R\log t + s, m) \geq D(t,m).$$

By combining Theorem 2.1.1 and Theorem 5.3.4, we get the following result on the front locations d and D.

Theorem 5.3.5. *Write $c^* = 3^{4/3}\pi^{2/3}/2^{7/6}$. Then for $m \leq m^*$, almost surely*

$$\limsup_{t \to \infty} \frac{\sqrt{2}t - d(t,m)}{t^{1/3}} \geq c^*, \qquad \liminf_{t \to \infty} \frac{\sqrt{2}t - d(t,m)}{t^{1/3}} \leq c^*,$$

$$\limsup_{t \to \infty} \frac{\sqrt{2}t - D(t,m)}{t^{1/3}} \geq c^* \quad \textit{and} \quad \liminf_{t \to \infty} \frac{\sqrt{2}t - D(t,m)}{t^{1/3}} \leq c^*.$$

Theorem 5.3.5 says that for $m \leq m^*$, (1) a.s. there are arbitrarily large times t at which $d(t,m) = \sqrt{2}t - c^* t^{1/3} + o(t^{1/3})$ and (2) a.s. there are arbitrarily large times t at which $D(t,m) = \sqrt{2}t - c^* t^{1/3} + o(t^{1/3})$.

In ongoing joint work with Louigi Addario-Berry and Julien Berestycki, we aim to show that at large times the mass density evolves approximately according to a non-local Fisher KPP equation

$$\frac{\partial u}{\partial t} = \tfrac{1}{2}\Delta u + u - u \int_{|y|<1} u(t, x+y)dy.$$

Bibliography

L Addario-Berry and S E Penington. The front location in BBM with decay of mass. *arXiv preprint arXiv:1503.01455*, 2015.

A Allen and J Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coursening. *Acta Metall.*, 27:1084–1095, 1979.

N H Barton. The probability of fixation of a favoured allele in a subdivided population. *Genetical Research*, 62(02):149–157, 1993.

N H Barton and G M Hewitt. Adaptation, speciation and hybrid zones. *Nature*, 341: 497–503, 1989.

N H Barton, A M Etheridge, and A Véber. A new model for evolution in a spatial continuum. *Electron. J. Probab*, 15(7):162–216, 2010.

N H Barton, A M Etheridge, J Kelleher, and A Véber. Inference in two dimensions: allele frequencies versus lengths of shared sequence blocks. *Theoretical population biology*, 87:105–119, 2013a.

N H Barton, A M Etheridge, J Kelleher, and A Véber. Genetic hitchhiking in spatially extended populations. *Theoretical population biology*, 87:75–89, 2013b.

N H Barton, A M Etheridge, and A Véber. Modelling evolution in a spatial continuum. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01002, 2013c.

H Berestycki, G Nadin, B Perthame, and L Ryzhik. The non-local Fisher–KPP equation: travelling waves and steady states. *Nonlinearity*, 22(12):2813, 2009a.

N Berestycki. Recent progress in coalescent theory. *Ensaios Matematicos*, 16(1):1–193, 2009.

N Berestycki, A M Etheridge, and M Hutzenthaler. Survival, extinction and ergodicity in a spatially continuous population model. In *Markov Proc. Rel. Fields*, 2009b.

N Berestycki, A M Etheridge, and A Véber. Large scale behaviour of the spatial Λ-Fleming–Viot process. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 49, pages 374–401, 2013.

P Biane and M Yor. Quelques précisions sur le méandre brownien. *Bulletin des sciences mathématiques*, 112(1):101–109, 1988.

P Billingsley. *Probability and Measure.* Wiley, 1995.

M D Bramson. Maximal displacement of branching Brownian motion. *Communications on Pure and Applied Mathematics*, 31(5):531–581, 1978.

N F Britton. Aggregation and the competitive exclusion principle. *Journal of Theoretical Biology*, 136(1):57–66, 1989.

N F Britton. Spatial structures and periodic travelling waves in an integro-differential reaction-diffusion population model. *SIAM Journal on Applied Mathematics*, 50(6): 1663–1688, 1990.

L Bronsard and R Kohn. Motion by mean curvature as the singular limit of Ginzburg-Landau dynamics. *J. Diff. Eq.*, 90:211–217, 1991.

X Chen. Generation and propagation of interfaces for reaction-diffusion equations. *J. Diff. Eq.*, 96:116–141, 1992.

J L Cherry. Selection in a subdivided population with local extinction and recolonization. *Genetics*, 164(2):789–795, 2003.

R Durrett and I Zähle. On the width of hybrid zones. *Stochastic Processes and their Applications*, 117(12):1751–1763, 2007.

A M Etheridge. *An introduction to superprocesses*, volume 20. American Mathematical Society, 2000.

A M Etheridge. Survival and extinction in a locally regulated population. *Annals of Applied Probability*, pages 188–214, 2004.

A M Etheridge. Drift, draft and structure: some mathematical models of evolution. *Banach Center Publ*, 80:121–144, 2008.

A M Etheridge. *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint-Flour XXXIX-2009*, volume 39. Springer Science & Business Media, 2011.

A M Etheridge and T G Kurtz. Genealogical constructions of population models. *arXiv preprint arXiv:1402.6724*, 2014.

A M Etheridge and A Véber. The spatial Λ-Fleming-Viot process on a large torus: genealogies in the presence of recombination. *Ann. Appl. Probab.*, pages 2165–2209, 2012.

A M Etheridge, A Véber, and F Yu. Rescaling limits of the spatial Λ-Fleming-Viot process with selection. *arXiv preprint arXiv:1406.5884*, 2014.

A M Etheridge, N Freeman, S E Penington, and D Straulino. Branching Brownian motion and Selection in the spatial Λ-Fleming-Viot process. *arXiv preprint arXiv:1512.03766*, 2015a.

A M Etheridge, N Freeman, and D Straulino. The Brownian Net and Selection in the spatial Λ-Fleming-Viot process. *arXiv preprint arXiv:1506.01158*, 2015b.

A M Etheridge, N Freeman, and S E Penington. Branching Brownian motion, mean curvature flow and the motion of hybrid zones. *arXiv preprint arXiv:1607.07563*, 2016.

S N Ethier and T G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

L C Evans and J Spruck. Motion of level sets by mean curvature I. *J. Diff. Geom.*, 33: 635–681, 1991.

L C Evans, H M Soner, and P E Souganidis. Phase transitions and generalized motion by mean curvature. *Comm. Pure Appl. Math.*, 45:1097–1123, 1992.

J Felsenstein. A pain in the torus: some difficulties with models of isolation by distance. *American Naturalist*, pages 359–368, 1975.

R A Fisher. The wave of advance of advantageous genes. *Ann. Eugenics*, 7:355–369, 1937.

R Forien and S E Penington. A central limit theorem for the spatial Λ- Fleming-Viot process with selection. *arXiv preprint arXiv:1512.06163*, 2015.

M Gage and R Hamilton. The heat equation shrinking convex plane curves. *J. Diff. Geom.*, 23:417–491, 1986.

S A Gourley. Travelling front solutions of a nonlocal Fisher equation. *Journal of mathematical biology*, 41(3):272–284, 2000.

M A Grayson. The heat equation shrinks embedded plane curves to round points. *J. Diff. Geom.*, 26:285–314, 1987.

M Hairer, M D Ryser, and H Weber. On the triviality of the 2D stochastic Allen-Cahn equation. *Elect. J. Prob.*, 17(39):1–14, 2012.

S C Harris and M I Roberts. The many-to-few lemma and multiple spines. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 2015.

S C Harris, M Hesse, A E Kyprianou, et al. Branching Brownian motion in a strip: Survival near criticality. *The Annals of Probability*, 44(1):235–275, 2016.

Y Hu and Z Shi. Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees. *The Annals of Probability*, 37(2):742–789, 2009.

G Huisken. Flow by mean curvature of convex surfaces into spheres. *J. Diff. Geom.*, 20: 237–266, 1984.

W G Hunt and R K Selander. Biochemical genetics of hybridization in European house mice. *Heredity*, 31:11–33, 1973.

N Ikeda, M Nagasawa, and S Watanabe. On branching Markov processes. *Proceedings of the Japan Academy*, 41(9):816–821, 1965.

T Ilmanen. Convergence of the Allen-Cahn equation to Brakke's motion by mean curvature. *J. Diff. Geom.*, 38:417–461, 1993.

B Jaffuel. The critical barrier for the survival of branching random walk with absorption. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 48, pages 989–1009, 2012.

O Kallenberg. *Foundations of modern probability.* Springer Science & Business Media, 2006.

M Kimura. Stepping stone model of population. *Ann. Rep. Nat. Inst. Genetics Japan*, 3:62–63, 1953.

J F C Kingman. Remarks on the spatial distribution of a reproducing population. *Journal of applied probability*, pages 577–583, 1977.

J F C Kingman. *Poisson processes.* Oxford university press, 1992.

S M Krone and C Neuhauser. Ancestral processes with selection. *Theor. Pop. Biol.*, 51: 210–237, 1997.

A Lambert. The branching process with logistic growth. *The Annals of Applied Probability*, 15(2):1506–1535, 2005.

G Malécot. *Les mathématiques de l'hérédité.* Masson et Cie, 1948.

C Mantegazza. *Lecture notes on mean curvature flow*, volume 290 of *Progress in Mathematics*. Birkhäuser, 2011.

T Maruyama. On the fixation probability of mutant genes in a subdivided population. *Genetical research*, 15(02):221–225, 1970.

C McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, pages 195–248. Springer, 1998.

H P McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. *Communications on Pure and Applied Mathematics*, 28(3):323–331, 1975.

P A P Moran. Random processes in genetics. In *Proceedings of the Cambridge Philosophical Society*, volume 54, page 60, 1958.

P Mörters and Y Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

C Mueller, L Mytnik, and J Quastel. Small noise asymptotics of traveling waves. *Markov Processes and Related Fields*, 14:333–342, 2008.

C Neuhauser and S M Krone. Genealogies of samples in models with selection. *Genetics*, 145:519–534, 1997.

J Pitman. Brownian motion, bridge, excursion, and meander characterized by sampling at independent uniform times. *Electronic Journal of Probability*, 4, 1999.

M I Roberts. A simple path to asymptotics for the frontier of a branching Brownian motion. *The Annals of Probability*, 41(5):3518–3541, 2013.

M I Roberts. Fine asymptotics for the consistent maximal displacement of branching Brownian motion. *Electronic Journal of Probability*, 20, 2015.

A Robertson. The reduction in fitness from genetic drift at heterotic loci in small populations. *Genetical research*, 15(02):257–259, 1970.

N Sato. A simple proof of convergence of the Allen-Cahn equation to Brakke's motion by mean curvature. *Indiana Univ. Math. J.*, 57:1743–1752, 2008.

B A Sevast'yanov. Branching stochastic processes for particles diffusing in a bounded domain with absorbing boundaries. *Theory of Probability & Its Applications*, 3(2): 111–126, 1958.

A V Skorohod. Branching diffusion processes. *Th. Prob. Appl.*, 9:492–497, 1964.

A Sudbury. Clumping effects in models of isolation by distance. *Journal of Applied Probability*, pages 391–395, 1977.

A Véber and A Wakolbinger. The spatial Lambda-Fleming-Viot process: an event-based construction and a lookdown representation. *Ann. Inst. H. Poincaré Probab. Statist.*, 51:570–598, 2015.

V Volpert and S Petrovskii. Reaction–diffusion waves in biology. *Physics of life reviews*, 6(4):267–310, 2009.

S Watanabe. On the branching process for brownian particles with an absorbing boundary. *Journal of Mathematics of Kyoto University*, 4(2):385–398, 1965.

B White. Evolution of curves and surfaces by mean curvature. *Proceedings of the International Congress of Mathematicians, Beijing 2002*, I:525–538, 2002.

M C Whitlock. Fixation probability and time in subdivided populations. *Genetics*, 164 (2):767–779, 2003.

S Wright. Isolation by distance. *Genetics*, 28(2):114, 1943.