

## Branching Brownian motion, mean curvature flow and the motion of hybrid zones

Alison Etheridge\*    Nic Freeman†    Sarah Penington‡

### 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  $\Lambda$ -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.

**Keywords:** branching Brownian motion; mean curvature flow; hybrid zones; spatial  $\Lambda$ -Fleming-Viot; population genetics.

**AMS MSC 2010:** 60J85; 92D15.

Submitted to EJP on July 7, 2017, final version accepted on November 17, 2017.

Supersedes arXiv:1607.07563.

## 1 Introduction

Our central result, Theorem 1.8 in Section 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

---

\*University of Oxford. E-mail: etheridg@stats.ox.ac.uk

†University of Sheffield. E-mail: n.p.freeman@sheffield.ac.uk

‡University of Oxford. E-mail: penington@maths.ox.ac.uk.

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

$$\frac{\partial v}{\partial t} = \Delta v + sv(1-v)(2v-1), \quad (1.1)$$

for an appropriate initial condition, where  $s > 0$  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 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 (1.1) in which  $t \mapsto \epsilon^2 t$  and  $x \mapsto \epsilon x$ . The Laplacian term is, of course, invariant, but the term corresponding to selection is multiplied by a factor  $1/\epsilon^2$ . It is well known that for suitable initial conditions, in a sense that we make precise in Theorem 1.3, as  $\epsilon \downarrow 0$ , 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

$$dw = (\Delta w + w - w^3)dt + \sigma dW, \quad (1.2)$$

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 (1.1).) 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  $\Lambda$ -Fleming-Viot process which was introduced in Etheridge [2008], Barton et al. [2010] and has been studied in a series of papers since; see e.g. Barton et al. [2013] 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  $\mathbb{Z}^d$ . In what follows, we shall refer to the spatial  $\Lambda$ -Fleming-Viot process with selection against heterozygosity as the SAFVS. It is described carefully in Definition 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,

$$\frac{\partial v}{\partial t} = \Delta v + sv(1 - v). \quad (1.3)$$

Mimicking that result, one can obtain (1.1) as a scaling limit of the SAFVS. Combined with the known convergence of the scaled version of (1.1), this certainly suggests that there should be scalings of the SAFVS which lead to mean curvature flow. However, available proofs of Theorem 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 1.3. We then adapt this to prove convergence of the proportions of different genetic types under the SAFVS 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 (1.1) which we believe to be of interest in its own right.

The probabilistic representation that we exploit closely resembles that used by De Masi et al. [1986]. They start from a spin system on a cubic lattice under the combined influence of spin flip (Glauber) dynamics and simple spin exchange (Kawasaki) dynamics. They prove that if the simple exchange occurs fast enough (on a timescale  $\delta^{-2}$ ) then the macroscopic density (measured on the scale  $\delta^{-1}$ ) evolves according to a reaction-diffusion equation. Their proofs are presented in one spatial dimension and for the special case in which the limiting equation is the deterministic analogue of (1.2), but they can readily be extended to higher dimensions. A further rescaling could then be applied to recover mean curvature flow.

Our analysis differs from that of De Masi et al. [1986], not only in replacing the spin system by the (continuum) SLFVS, in which the stochasticity is specifically tailored to model the random genetic drift in which we are interested, but also in passing directly from the stochastic system to the mean curvature flow in a single scaling, thereby ensuring that if genetic drift is sufficiently weak it won’t completely obscure the motion of the interface. There are also examples in which mean curvature flow is recovered directly from a scaling of a spin system. For example, Funaki and Spohn [1997] obtain it from the Ginzburg-Landau  $\nabla\phi$  interface model. The noise is an additive white noise at each lattice point, and, observing that the deterministic part of equation (1.2) is the  $L^2$  gradient flow of the Ginzburg-Landau free energy functional with the potential energy given by  $V(w) = (w^2 - 1)^2/4$ , we see that their result can be thought of as a spin-system analogue of the work of Hairer et al. [2012].

Before defining the SAFVS, 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  $\mathbb{R}^d$  for arbitrary  $d > 1$ .

### 1.1 The Allen-Cahn equation and mean curvature flow

The Allen-Cahn equation [Allen and Cahn, 1979] takes the form

$$\frac{\partial v^\epsilon}{\partial t} = \Delta v^\epsilon - \frac{1}{\epsilon^2} f(v^\epsilon), \quad (1.4)$$

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

$$\begin{aligned} f(v) &< 0, & \forall v \in (-\infty, v_-) \cup (v_0, v_+); \\ f(v) &> 0, & \forall v \in (v_-, v_0) \cup (v_+, \infty); \\ f'(v_-) &> 0, & f'(v_+) > 0, \quad f'(v_0) < 0. \end{aligned} \tag{1.5}$$

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  $s = t/\epsilon$ , 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 (1.4), 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 1.1** (Curve-shortening flow). *Let  $S^1$  denote the unit circle in  $\mathbb{R}^2$ . Let  $\Gamma = (\Gamma_t(\cdot))_t$  be a family of smooth embeddings, indexed by  $t \in [0, \mathcal{T})$ , where for each  $t$ ,  $\Gamma_t : S^1 \rightarrow \mathbb{R}^2$ . Let  $\mathbf{n} = \mathbf{n}_t(\phi)$  denote the unit (inward) normal vector to  $\Gamma_t$  at  $\phi$  and let  $\kappa = \kappa_t(\phi)$  denote the curvature of  $\Gamma_t$  at  $\phi$ . We say that  $\Gamma$  is a curvature flow or curve-shortening flow if*

$$\frac{\partial \Gamma_t(\phi)}{\partial t} = \kappa_t(\phi) \mathbf{n}_t(\phi). \tag{1.6}$$

for all  $t, \phi$ .

Assuming that  $\Gamma_0$  is a smooth embedding of  $S^1$  into  $\mathbb{R}^2$ , the behaviour of  $\Gamma_t$  under curve-shortening is completely understood. First, it has a finite lifetime which we shall denote by  $\mathcal{T}$ . In Gage and Hamilton [1986], it was shown that if  $\Gamma_0$  is convex, then so is  $\Gamma_t$  for all  $t < \mathcal{T}$ . Moreover,  $\mathcal{T}$  can be chosen so that  $\Gamma_t$  shrinks towards a point as  $t \uparrow \mathcal{T}$ ; in this limit the asymptotic ‘shape’ of  $\Gamma_t$  is a circle. Soon afterwards, Grayson [1987] showed that, in fact, under curve-shortening, any smoothly embedded closed curve becomes convex at a time  $\tau < \mathcal{T}$ , 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  $(d - 1)$ -dimensional hypersurface in  $\mathbb{R}^d$ , 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  $i$ th vector in the basis in the direction of the  $j$ th. The  $d - 1$  principal curvatures,  $\kappa_1, \dots, \kappa_{d-1}$ , 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 1.2** (Mean curvature flow). Mean curvature flow, when it is defined, is obtained by replacing the curvature  $\kappa_t$  in equation (1.6) by the mean curvature.

The behaviour of mean curvature flow in  $d \geq 3$  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  $(d - 1)$ -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  $d \geq 2$ . Our first result concerns the convergence as  $\epsilon \downarrow 0$ , for suitable initial conditions, of the solution of

$$\frac{\partial v^\epsilon}{\partial t} = \Delta v^\epsilon + \frac{1}{\epsilon^2} v^\epsilon (1 - v^\epsilon)(2v^\epsilon - 1), \quad v^\epsilon(0, x) = p(x), \tag{1.7}$$

to the indicator function of a set whose boundary evolves according to mean curvature flow.

The initial condition,  $p$ , of (1.7) is assumed to take values in  $[0, 1]$ . We shall also require that it satisfies some regularity conditions. In particular, set

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

We suppose that  $\Gamma$  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:

- (C1)  $\Gamma$  is  $C^\alpha$  for some  $\alpha > 3$ .
- (C2) For  $x$  inside  $\Gamma$ ,  $p(x) < \frac{1}{2}$ . For  $x$  outside  $\Gamma$ ,  $p(x) > \frac{1}{2}$ .
- (C3) There exist  $r, \gamma > 0$  such that  $|p(x) - \frac{1}{2}| \geq \gamma (\text{dist}(x, \Gamma) \wedge r)$  for all  $x \in \mathbb{R}^d$ .

Condition (C3) prevents the slope of  $p$  near the interface  $\Gamma$  from being too shallow, and keeps  $p(x)$  bounded away from  $\frac{1}{2}$  when  $x$  is not near the interface. Condition (C2) is simply establishing a sign convention. Under these conditions, mean curvature flow started from  $\Gamma$ , which we denote  $(\Gamma_t(\cdot))_t$ , exists up to some finite time  $\mathcal{T}$  (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  $\Gamma_t$ , chosen to be negative inside  $\Gamma_t$  and positive outside. Note that, as sets,

$$\Gamma_t = \{x \in \mathbb{R}^d : d(x, t) = 0\}.$$

**Theorem 1.3.** Let  $v^\epsilon$  solve (1.7) with initial condition  $p$  satisfying the conditions (C1)-(C3), and define  $\mathcal{T}$ ,  $d(x, t)$  as above. Fix  $T^* \in (0, \mathcal{T})$ . Let  $k \in \mathbb{N}$ . There exists  $\epsilon_d(k) > 0$ , and  $a_d(k), c_d(k) \in (0, \infty)$  such that for all  $\epsilon \in (0, \epsilon_d)$  and  $t$  satisfying  $a_d \epsilon^2 |\log \epsilon| \leq t \leq T^*$ ,

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

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  $d$ -dimensional Brownian motion and the interface  $\Gamma_s$  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 (1.7) and a branching Brownian motion, which is of some interest in its own right.

### 1.2 Modelling hybrid zones

Let us now turn to our model of hybrid zones. Our starting point is the spatial  $\Lambda$ -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 3.1), mirroring our probabilistic representation of solutions to (1.1).

We suppose that there are two alleles,  $a$  and  $A$ . At each time  $t$ , the random function  $\{w_t(x), x \in \mathbb{R}^d\}$  is defined, up to a Lebesgue null set of  $\mathbb{R}^d$ , by

$$w_t(x) := \text{proportion of type } a \text{ at spatial position } x \text{ at time } t. \tag{1.8}$$

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 1.4.** It is convenient to extend the definition of  $w_t(x)$  to all of  $\mathbb{R}^d$  and so, on the Lebesgue null set on which (1.8) 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  $\mathbb{R}^d \times \{a, A\}$  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 1.5** (Spatial  $\Lambda$ -Fleming-Viot with selection against heterozygosity (SLFVS)). *Fix  $u \in (0, 1]$  and  $\mathcal{R} \in (0, \infty)$ . Let  $\mu$  be a finite measure on  $(0, \mathcal{R}]$ . Further, let  $\Pi$  be a Poisson point process on  $\mathbb{R}_+ \times \mathbb{R}^d \times (0, \mathcal{R}]$  with intensity measure*

$$dt \otimes dx \otimes \mu(dr). \tag{1.9}$$

The spatial  $\Lambda$ -Fleming-Viot process with selection (SLFVS) driven by  $\Pi$  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  $1 - s$  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,  $\alpha_0$ , 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  $w_t(y) = (1 - u)w_{t-}(y) + u\mathbf{1}_{\{\alpha_0=a\}}$ .

With the complementary probability  $s$  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  $\alpha_1, \alpha_2, \alpha_3$  according to  $w_{t-}(z_1), w_{t-}(z_2), w_{t-}(z_3)$  respectively. Let  $\hat{\alpha}$  denote the most common allelic type in  $\alpha_1, \alpha_2, \alpha_3$ .
2. For every  $y \in \mathcal{B}_r(x)$  set  $w_t(y) = (1 - u)w_{t-}(y) + u\mathbf{1}_{\{\hat{\alpha}=a\}}$ .

**Remark 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  $\nu_r$  be a probability measure on  $(0, 1]$  and the driving noise,  $\Pi$ , is taken to be a Poisson point process on  $\mathbb{R}_+ \times \mathbb{R}^d \times (0, \mathcal{R}] \times (0, 1]$  with intensity measure

$$dt \otimes dx \otimes \mu(dr)\nu_r(du).$$

For each point  $(t, x, r, u) \in \Pi$ , the corresponding reproduction event is described exactly as before.

Since  $s$  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  $1 - s$ . 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  $(1 - s)$  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

$$\begin{aligned} \frac{\bar{w}^2 + \bar{w}(1 - \bar{w})(1 - s)}{1 - 2s\bar{w}(1 - \bar{w})} &= (1 - s)\bar{w} + s(3\bar{w}^2 - 2\bar{w}^3) + \mathcal{O}(s^2) \\ &= (1 - s)\bar{w} + s(\bar{w}^3 + 3\bar{w}^2(1 - \bar{w})) + \mathcal{O}(s^2). \end{aligned} \quad (1.10)$$

Notice that the first term in (1.10) is  $1 - s$  times the probability that an allele sampled from the parental population is of type  $a$  whereas the second is  $s$  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

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

which underpins the connection to (1.1).

Of course, in replacing a diploid model by one based directly on allele frequencies, we have rather muddled the notion of parent in our reproduction mechanism, so the use of the term in Definition 1.5 should not be interpreted too literally.

### 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  $n$ th stage of the rescaling, the impact and selection parameters are assumed to satisfy

$$u_n = \frac{u}{n^{1-2\beta}}, \quad \text{and} \quad s_n = \frac{\rho}{n^{2\beta}}.$$

Next, we define the averaged process,

$$w_t^n(x) := w_{nt}(n^\beta x), \quad \text{and} \quad \bar{w}_t^n(x) := \frac{n^{\beta d}}{V_R} \int_{B(x, n^{-\beta} R)} w_t^n(y) dy,$$

where  $V_R$  is the volume of the ball of radius  $R$  in  $\mathbb{R}^d$ . To simplify notation, we write  $\mathcal{M}$  for  $\mathcal{M}_\lambda(\mathbb{R}^d \times \{a, A\})$ , and  $D_{\mathcal{M}}[0, \infty)$  for the set of all càdlàg paths with values in  $\mathcal{M}$ . We also write  $C_c^\infty(\mathbb{R}^d)$  for the set of smooth compactly supported functions on  $\mathbb{R}^d$ .

**Theorem 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  $n \rightarrow \infty$ , 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  $f \in C_c^\infty(\mathbb{R}^d)$ ,*

$$\langle w_t^\infty, f \rangle = \langle w_0^\infty, f \rangle + \int_0^t \left\{ \frac{\kappa_R}{2} \langle w_s^\infty, \Delta f \rangle + u\rho V_R \langle w_s^\infty(1 - w_s^\infty)(2w_s^\infty - 1), f \rangle \right\} ds,$$

where

$$\kappa_R = \frac{u}{V_R} \int_{B(0, R)} \int_{B(x, R)} (z_1)^2 dz dx \tag{1.11}$$

with  $z_1$  the first coordinate of the vector  $z \in \mathbb{R}^d$ . In particular,  $\kappa_R$  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 (1.7) with  $w_0 = w^0$ . Based on Theorem 1.7, it is natural to ask whether we can modify the scaling of  $s_n$  in such a way that  $s_n n^{2\beta} \rightarrow \infty$  as  $n \rightarrow \infty$  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 1.7. For each  $n \in \mathbb{N}$ , we define the finite measure  $\mu^n$  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  $\Pi^n$  on  $\mathbb{R}_+ \times \mathbb{R}^d \times (0, \infty)$  with intensity measure

$$ndt \otimes n^\beta dx \otimes \mu^n(dr). \tag{1.12}$$



Here  $n^\beta dx$  denotes the scaling in which the linear dimension of the infinitesimal region  $dx$  is scaled by  $n^\beta$  (so that when we integrate, the volume of a region is scaled by  $n^{d\beta}$ ). Let

$$u_n = \frac{u}{n^{1-2\beta}}, \quad \text{and} \quad s_n = \frac{1}{\epsilon_n^2} \frac{1}{n^{2\beta}}. \tag{1.13}$$

It is convenient to define the constant  $\sigma^2$  through

$$\sigma^2 = \frac{u}{2d} \int_0^{\mathcal{R}} \int_{\mathbb{R}^d} |z|^2 \frac{V_r(0, z)}{V_r} dz \mu(dr). \tag{1.14}$$

If  $\mu(dr) = \delta_R(r)$ , then we recover  $\kappa_R$  from (1.11).

**Theorem 1.8.** *Suppose that  $\beta \in (0, 1/4)$  and let  $\epsilon_n$  be a sequence such that  $\epsilon_n \rightarrow 0$  and  $(\log n)^{1/2} \epsilon_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Let  $(w_t^n)_{t \geq 0}$  be the SAFVS driven by  $\Pi^n$  and with  $u_n, s_n$  given by (1.13), and initial condition  $w_0^n(x) = p(x)$ . Assume that  $p$  satisfies (C1)-(C3), and define  $\mathcal{T}, d(x, t)$  as for Theorem 1.3; take  $T^* < \mathcal{T}$ . For  $k \in \mathbb{N}$  there exist  $n_*(k) < \infty$ , and  $a_*(k), d_*(k) \in (0, \infty)$  such that for all  $n \geq n_*$  and all  $t$  satisfying  $a_* \epsilon_n^2 |\log \epsilon_n| \leq t \leq T^*$ ,*

1. *for almost every  $x$  such that  $d(x, \sigma^2 t) \geq d_* \epsilon_n |\log \epsilon_n|$ , we have  $\mathbb{E}[w_t^n(x)] \geq 1 - \epsilon_n^k$ ;*
2. *for almost every  $x$  such that  $d(x, \sigma^2 t) \leq -d_* \epsilon_n |\log \epsilon_n|$ , we have  $\mathbb{E}[w_t^n(x)] \leq \epsilon_n^k$ .*

**Remark 1.9.** Since  $w_t^n(x) \in [0, 1]$ , it follows by Markov's inequality that for all  $n \geq n_*(k+1)$  and all  $t$  satisfying  $a_*(k+1) \epsilon_n^2 |\log \epsilon_n| \leq t \leq T^*$ ,

1. *for almost every  $x$  such that  $d(x, \sigma^2 t) \geq d_*(k+1) \epsilon_n |\log \epsilon_n|$ , we have  $\mathbb{P}[w_t^n(x) \leq 1 - \epsilon_n] \leq \epsilon_n^k$ ;*
2. *for almost every  $x$  such that  $d(x, \sigma^2 t) \leq -d_*(k+1) \epsilon_n |\log \epsilon_n|$ , we have  $\mathbb{P}[w_t^n(x) \geq \epsilon_n] \leq \epsilon_n^k$ .*

**Remark 1.10.** In Section 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. [2015], we expect an analogous result for values of  $u_n$  up to  $\mathcal{O}(1)$ , but at the expense of having to take  $\epsilon_n \rightarrow 0$  extremely slowly (so that  $\epsilon_n^{-1} = o(\log \log n)$ ). 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 (1.7).

The rest of the paper is laid out as follows. In Section 2 we establish a duality between equation (1.1) and a branching Brownian motion which we then use to prove Theorem 1.3. In Section 3 we establish an analogous duality between the SLFVS and a system of branching and coalescing particles and use it to establish Theorem 1.8.

## 2 Proof of Theorem 1.3

### 2.1 A probabilistic dual to Equation (1.7)

Our proof of Theorem 1.3 rests on a duality between equation (1.7) and a branching Brownian motion in which each individual, independently, follows a Brownian motion during an exponentially distributed lifetime (with mean  $\epsilon^2$ ) 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 (1.7) is not of this form. Instead we adapt the dual of De Masi et al. [1986], in a manner that is familiar 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

$$\text{all Brownian motions run at rate 2.} \tag{2.1}$$

That is, at time 1, Brownian motion has variance 2.

In contrast to the McKean-Skorohod setting, the representation of the solution to (1.1) 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 (1.1) 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  $\mathbf{W}(t)$  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<sup>nd</sup> child of the 1<sup>st</sup> child of the 3<sup>rd</sup> 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  $\mathbf{W}(t)$  traces out a time-labelled ternary tree which records the genealogy and associates a time to each branching event. We shall use  $\mathcal{T}(\mathbf{W}(t))$  to denote this time-labelled ternary tree.

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

1. Each leaf  $i$  of  $\mathcal{T}(\mathbf{W}(t))$ , independently, votes 1 with probability  $p(W_i(t))$  and otherwise votes 0.
2. At each branch point in  $\mathcal{T}(\mathbf{W}(t))$ , 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  $\mathcal{T}(\mathbf{W}(t))$  to the root  $\emptyset$ .

**Definition 2.1** ( $\mathbb{V}_p$ ). *With the voting procedure described above, we define  $\mathbb{V}_p(\mathbf{W}(t))$  to be the vote associated to the root  $\emptyset$ .*

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

**Theorem 2.2.** Let  $p : \mathbb{R}^d \rightarrow [0, 1]$ . Then

$$v^\epsilon(t, x) = \mathbb{P}_x^\epsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] \tag{2.2}$$

is a solution to equation (1.7) with initial condition  $v^\epsilon(0, x) = p(x)$ .

*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 (2.2) by partitioning on the behaviour of the branching Brownian motion in the first  $\delta t$  of time and then to take a limit as  $\delta t \downarrow 0$ .

Throughout the proof we neglect the superscript  $\epsilon$  in  $\mathbb{P}_x^\epsilon$ ,  $\mathbb{E}_x^\epsilon$  and  $v^\epsilon$  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 ( $W$ ), preserving  $\mathbb{E}$  for expectation with respect to that of the historical branching Brownian motion ( $\mathbf{W}(\cdot)$ ). 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  $\delta t$  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),

$$\mathbb{E}_x[V_1 | S \leq \delta t] = v(t, x) + \mathcal{O}(\delta t). \tag{2.3}$$

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 (2.3) and conditional independence of the  $V_i$  given  $(S, W_S)$ ,

$$\mathbb{P}_x [\mathbb{V}(\mathbf{W}(t + \delta t)) = 1 | S \leq \delta t] = v(t, x)^3 + 3v(t, x)^2(1 - v(t, x)) + \mathcal{O}(\delta t).$$

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  $\delta t$  of time gives

$$\begin{aligned} v(t + \delta t, x) &= \mathbb{P}_x [\mathbb{V}(\mathbf{W}(t + \delta t)) = 1 | S \leq \delta t] \mathbb{P} [S \leq \delta t] \\ &\quad + \mathbb{P}_x [\mathbb{V}(\mathbf{W}(t + \delta t)) = 1 | S > \delta t] (1 - \mathbb{P} [S \leq \delta t]) \\ &= \mathbb{P}_x [\mathbb{V}(\mathbf{W}(t + \delta t)) = 1 | S \leq \delta t] \mathbb{P} [S \leq \delta t] \\ &\quad + E_x [\mathbb{P}_{W_{\delta t}} [\mathbb{V}(\mathbf{W}(t)) = 1]] (1 - \mathbb{P} [S \leq \delta t]). \end{aligned}$$

Now  $\mathbb{P}[S \leq \delta t] = \epsilon^{-2}\delta t + \mathcal{O}(\delta t^2)$  and so substituting and rearranging (and once again assuming enough regularity of  $v(t, x)$ ) we obtain

$$\begin{aligned} \lim_{\delta t \rightarrow 0} \frac{v(t + \delta t, x) - v(t, x)}{\delta t} &= \epsilon^{-2} (v(t, x)^3 + 3v(t, x)^2(1 - v(t, x)) - v(t, x)) \\ &\quad + \lim_{\delta t \rightarrow 0} \frac{E_x [\mathbb{P}_{W_{\delta t}} [\mathbb{V}(\mathbf{W}(t)) = 1]] - v(t, x)}{\delta t} \\ &= \epsilon^{-2} (v(t, x)^3 + 3v(t, x)^2(1 - v(t, x)) - v(t, x)) \\ &\quad + \lim_{\delta t \rightarrow 0} \frac{E_x [v(t, W_{\delta t})] - v(t, x)}{\delta t} \\ &= \Delta v(t, x) + \epsilon^{-2} v(t, x)(1 - v(t, x))(2v(t, x) - 1), \end{aligned}$$

as required. ■

**Remark 2.3.** The same method can be used to establish analogues of (2.2) with more general voting systems. For any  $m \in \mathbb{N}$  and  $\alpha : \{0, 1\}^m \rightarrow [0, 1]$ , we can define an iterative voting procedure on  $m$ -ary trees as follows. At any branch point, if the votes of the  $m$  offspring are given by  $V_1, \dots, V_m$ , then the vote of their parent is 1 with probability  $\alpha(V_1, \dots, V_m)$ . Such a voting system results in solutions to equations of the form  $\frac{\partial v}{\partial t} = \Delta v + V(g(v) - v)$ , where  $g(v) = \mathbb{E}[\alpha(B_1(v), \dots, B_m(v))]$  for  $B_1(v), B_2(v), \dots$  a sequence of i.i.d. Bernoulli( $v$ ) random variables.

For example,  $\frac{\partial v}{\partial t} = \Delta v + V(v^4 + 4v^3(1 - v) - v)$  corresponds to branching Brownian motion, with branching into 4 occurring at rate  $V$ , in which, at each branch point, the parent is given vote 1 if and only if at least three of the offspring vote 1.

Armed with this representation, the proof of Theorem 1.3 is reduced to proving the following result about our branching Brownian motions.

**Theorem 2.4.** *Suppose  $p : \mathbb{R}^d \rightarrow [0, 1]$  is such that (C1)-(C3) hold. Define  $\mathcal{I}$ ,  $d(x, t)$  as for Theorem 1.3; fix  $T^* \in (0, \mathcal{I})$  and let  $k \in \mathbb{N}$ . There exist  $\epsilon_d(k) > 0$ , and  $a_d(k), c_d(k) \in (0, \infty)$  such that for all  $\epsilon \in (0, \epsilon_d)$  and  $t$  satisfying  $a_d \epsilon^2 |\log \epsilon| \leq t \leq T^*$ ,*

1. *for  $x$  such that  $d(x, t) \geq c_d \epsilon |\log \epsilon|$ , we have  $\mathbb{P}_x^\epsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] \geq 1 - \epsilon^k$ ;*
2. *for  $x$  such that  $d(x, t) \leq -c_d \epsilon |\log \epsilon|$ , we have  $\mathbb{P}_x^\epsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] \leq \epsilon^k$ .*

The proof of Theorem 2.4 will proceed in two steps. First, in Section 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 (C1)-(C3). These allow us to couple the distance between the (backwards in time) mean curvature flow  $(\Gamma_{t-s})_{s \in [0, t]}$  and a (forwards in time)  $d$ -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  $\Gamma_{t-s}$ . This coupling is made precise in Proposition 2.14 in Section 2.3. The proof of Theorem 2.4, which combines these two steps by bounding the errors that occur far from the interface  $\Gamma_{t-s}$ , can be found in Section 2.4.

**Notation 2.5.** *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  $\mathbf{B}$  will represent the corresponding historical branching Brownian motion and we preserve  $W$  and  $\mathbf{W}$  for dimensions  $d \geq 2$ .*

## 2.2 Majority voting in one-dimensional BBM

In this section we consider only ternary branching Brownian motion in dimension  $d = 1$ .

As in Section 2.1, for  $x \in \mathbb{R}$ , we write  $\mathbb{P}_x^\epsilon$  for the probability measure under which  $(\mathbf{B}(t), t \geq 0)$  has the law of historical ternary branching Brownian motion in  $\mathbb{R}$  with branching rate  $1/\epsilon^2$  started from a single particle at location  $x$  at time 0, and  $\mathbb{E}_x^\epsilon$  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 2.4 for this initial condition  $p_0$ .

**Theorem 2.6.** *Let  $T^* \in (0, \infty)$ . For all  $k \in \mathbb{N}$  there exist  $c_1(k)$  and  $\epsilon_1(k) > 0$  such that, for all  $t \in [0, T^*]$  and all  $\epsilon \in (0, \epsilon_1)$ ,*

1. for  $z \geq c_1(k)\epsilon|\log \epsilon|$ , we have  $\mathbb{P}_z^\epsilon[\mathbb{V}(\mathbf{B}(t)) = 1] \geq 1 - \epsilon^k$
2. for  $z \leq -c_1(k)\epsilon|\log \epsilon|$ , we have  $\mathbb{P}_z^\epsilon[\mathbb{V}(\mathbf{B}(t)) = 1] \leq \epsilon^k$ .

**Remark 2.7.** The subscript 1 on  $a_1, c_1$  and  $\epsilon_1$  is to emphasize that Theorem 2.6 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 2.6 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  $\epsilon$  tends to zero, there is enough branching close to the root to sufficiently magnify the bias.

### 2.2.1 Proof of Theorem 2.6

First note that with our special choice of initial condition  $p_0$ , for any  $x_1 \leq x_2 \in \mathbb{R}$ ,

$$\mathbb{P}_{x_1}^\epsilon[\mathbb{V}(\mathbf{B}(t)) = 1] \leq \mathbb{P}_{x_2}^\epsilon[\mathbb{V}(\mathbf{B}(t)) = 1]. \tag{2.4}$$

By analogy with the previous subsection, we use  $\mathcal{T}(\mathbf{B}(t))$  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

$$\mathbb{P}_x^t(\mathcal{T}) = \mathbb{P}_x^\epsilon[\mathbb{V}(\mathbf{B}(t)) = 1 \mid \mathcal{T}(\mathbf{B}(t)) = \mathcal{T}]. \tag{2.5}$$

By the symmetry of the Brownian motions followed by individuals in  $\mathbf{B}(t)$  conditional on  $\{\mathcal{T}(\mathbf{B}(t)) = \mathcal{T}\}$ , 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  $z \in \mathbb{R}$ ,

$$\mathbb{P}_z^t(\mathcal{T}) = 1 - \mathbb{P}_{-z}^t(\mathcal{T}). \tag{2.6}$$

The monotonicity in (2.4) and the symmetry in (2.6) are key to our proof of Theorem 2.6.

Taking  $z = 0$  in (2.6) shows that  $\mathbb{P}_0^t(\mathcal{T}) = \frac{1}{2}$  for all  $t > 0$ , and, by (2.4), for all  $t > 0$  and all time-labelled ternary trees  $\mathcal{T}$  we have

$$\mathbb{P}_z^t(\mathcal{T}) \geq \frac{1}{2} \text{ for } z > 0; \quad \mathbb{P}_z^t(\mathcal{T}) \leq \frac{1}{2} \text{ for } z < 0.$$

We now introduce notation for the majority voting procedure. Let  $g : [0, 1]^3 \rightarrow [0, 1]$  be given by

$$g(p_1, p_2, p_3) = p_1p_2p_3 + p_1p_2(1 - p_3) + p_2p_3(1 - p_1) + p_3p_1(1 - p_2). \tag{2.7}$$

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

$$g(1 - p_1, 1 - p_2, 1 - p_3) = 1 - g(p_1, p_2, p_3). \tag{2.8}$$

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  $\tau$  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

$$g(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star)) = g(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}_1), \mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}_2), \mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}_3)) \tag{2.9}$$

and the identity

$$\mathbb{P}_z^t(\mathcal{T}) = E_z [g(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star))] \tag{2.10}$$

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 (2.6), when it is convenient to do so we will only state such results for the case  $z \geq 0$ .

**Lemma 2.8.** *For any time-labelled ternary tree  $\mathcal{T}$ , any time  $t > 0$ , and any  $z \geq 0$ ,*

$$\mathbb{P}_z^t(\mathcal{T}) \geq P_z[B_t \geq 0].$$

*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,  $\mathbb{P}_z^t(\mathcal{T}_0) = P_z[B_t \geq 0]$ .

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  $h : [0, 1]^3 \rightarrow \mathbb{R}$  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 (2.8) we have

$$h(1 - p_1, 1 - p_2, 1 - p_3) = -h(p_1, p_2, p_3). \tag{2.11}$$

We can write  $h$  in the form

$$h(p_1, p_2, p_3) = \frac{1}{3} \sum p_{i_1} \left( (1 - p_{i_2})(p_{i_3} - \frac{1}{2}) + (1 - p_{i_3})(p_{i_2} - \frac{1}{2}) \right)$$

where the sum is over  $(i_1, i_2, i_3) = (1, 2, 3), (2, 3, 1), (3, 1, 2)$ . Hence

$$\frac{1}{2} \leq p_1, p_2, p_3 \leq 1 \Rightarrow h(p_1, p_2, p_3) \geq 0. \tag{2.12}$$

We will use the  $\star$  notation defined in (2.9) 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  $\tau, \mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$  be as in (2.10). Using (2.10), by the definition of  $g$  and  $h$  we have

$$\begin{aligned} \mathbb{P}_z^t(\mathcal{T}) &= E_z [g(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star))] \\ &= E_z [h(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star))] + \frac{1}{3} \sum_{i=1}^3 E_z [\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}_i)]. \end{aligned} \tag{2.13}$$

We will show that the first term of (2.13) is non-negative. Combining (2.11) with (2.6),

$$h(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star)) = -h(\mathbb{P}_{-B_\tau}^{t-\tau}(\mathcal{T}^\star)).$$

Hence,

$$\begin{aligned} E_z[h(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star))] &= E_z [h(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star)) \mathbb{1}\{B_\tau \geq 0\}] + E_z [h(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star)) \mathbb{1}\{B_\tau < 0\}] \\ &= E_z [h(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star)) \mathbb{1}\{B_\tau \geq 0\}] - E_z [h(\mathbb{P}_{-B_\tau}^{t-\tau}(\mathcal{T}^\star)) \mathbb{1}\{B_\tau < 0\}] \\ &= \int_0^\infty h(\mathbb{P}_x^{t-\tau}(\mathcal{T}^\star))(\phi_{z,2\tau}(x) - \phi_{z,2\tau}(-x)) dx, \end{aligned} \tag{2.14}$$

where  $\phi_{\mu, \sigma^2}$  denotes the density of a  $N(\mu, \sigma^2)$  random variable. Since  $\mathbb{P}_x^{t-\tau}(\mathcal{T}_i) \geq 1/2$  for  $x \geq 0$ , by (2.12) we have  $h(\mathbb{P}_x^{t-\delta t}(\mathcal{T}_\star)) \geq 0$ , and since  $z \geq 0$ , for all  $x \geq 0$  we have

$$\phi_{z, 2\tau}(x) - \phi_{z, 2\tau}(-x) \geq 0,$$

which proves that (2.14) is non-negative. This shows that the first term of (2.13) is non-negative and we now move on to the second term.

Using our inductive hypothesis, for  $i = 1, 2, 3$ ,

$$E_z[\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}_i)] \geq E_z[P_{B_\tau}[B_{t-\tau} \geq 0]] = P_z[B_t \geq 0]$$

and so substituting into (2.13) completes the proof of Lemma 2.8. ■

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

$$g^{(1)}(p) = g(p), \quad g^{(n+1)}(p) = g^{(n)}(g(p)).$$

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 2.9.** *For all  $k \in \mathbb{N}$  there exists  $A(k) < \infty$  such that, for all  $\epsilon \in (0, \frac{1}{2}]$  and  $n \geq A(k) \lceil \log \epsilon \rceil$  we have*

$$g^{(n)}(\frac{1}{2} + \epsilon) \geq 1 - \epsilon^k.$$

*Proof.* We carry out two phases of iteration of  $g$ . First, we will show that it takes  $\mathcal{O}(\lceil \log \epsilon \rceil)$  iterations to obtain

$$g^{(n)}(\frac{1}{2} + \epsilon) \geq \frac{1}{2} + \frac{1}{\sqrt{8}}. \tag{2.15}$$

Then we note that  $\mathcal{O}(\log |k \log \epsilon|)$  iterations are required to obtain

$$g^{(n)}(\frac{1}{2} + \frac{1}{\sqrt{8}}) \geq 1 - \epsilon^k. \tag{2.16}$$

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

$$g(\frac{1}{2} + \delta) = \frac{1}{2} + \frac{3}{2}\delta - 2\delta^3 \geq \frac{1}{2} + \frac{5}{4}\delta.$$

Thus if  $g^{(n)}(\frac{1}{2} + \epsilon) - \frac{1}{2} < 1/\sqrt{8}$ , we have

$$g^{(n+1)}(\frac{1}{2} + \epsilon) - \frac{1}{2} \geq \frac{5}{4} \left( g^{(n)}(\frac{1}{2} + \epsilon) - \frac{1}{2} \right) \geq \left( \frac{5}{4} \right)^n \epsilon.$$

It follows immediately that  $\mathcal{O}(\lceil \log \epsilon \rceil)$  iterations are required to achieve (2.15).

For the second phase, note that  $1 - g(1 - \delta) = 3\delta^2 - 2\delta^3 \leq 3\delta^2$ , so that  $1 - g(\frac{1}{2} + \frac{1}{\sqrt{8}}) \leq 3(\frac{1}{2} - \frac{1}{\sqrt{8}})^2$ . Also for  $n \geq 1$ , if  $1 - g^{(n)}(\frac{1}{2} + \frac{1}{\sqrt{8}}) \leq \frac{1}{3} \left( 3(\frac{1}{2} - \frac{1}{\sqrt{8}}) \right)^{2^n}$ , then

$$1 - g^{(n+1)}(\frac{1}{2} + \frac{1}{\sqrt{8}}) \leq 3 \left( 1 - g^{(n)}(\frac{1}{2} + \frac{1}{\sqrt{8}}) \right)^2 \leq \frac{1}{3} \left( 3(\frac{1}{2} - \frac{1}{\sqrt{8}}) \right)^{2^{n+1}}.$$

Hence, by induction, for all  $n \geq 1$  we have  $1 - g^{(n)}(\frac{1}{2} + \frac{1}{\sqrt{8}}) \leq \frac{1}{3} \left( 3(\frac{1}{2} - \frac{1}{\sqrt{8}}) \right)^{2^n}$ . Noting that  $3(\frac{1}{2} - \frac{1}{\sqrt{8}}) < 1$ , it follows easily that the number of iterations required to obtain (2.16) is  $\mathcal{O}(\log |k \log \epsilon|)$ . ■

We now want to see that there is a (large) regular ternary tree sitting inside  $\mathcal{T}(\mathbf{B}(t))$ . Let  $\mathcal{T}_n^{reg} = \cup_{k \leq n} \{1, 2, 3\}^k \subset \mathcal{U}$  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 2.10.** *Let  $k \in \mathbb{N}$  and let  $A = A(k)$  be as in Lemma 2.9. Then there exist  $a_1 = a_1(k)$  and  $\epsilon_1 = \epsilon_1(k)$  such that, for all  $\epsilon \in (0, \epsilon_1)$  and  $t \geq a_1 \epsilon^2 \lceil \log \epsilon \rceil$ ,*

$$\mathbb{P}^\epsilon \left[ \mathcal{T}(\mathbf{B}(t)) \supseteq \mathcal{T}_{A(k) \lceil \log \epsilon \rceil}^{reg} \right] \geq 1 - \epsilon^k.$$

*Proof.* First we establish control over the tail distribution of the sum of  $n$  independent exponentially distributed (branching) times. Suppose  $(X_j)_{j \geq 1}$  are i.i.d. Exp(1) random variables and let  $S_n = \sum_{j=1}^n X_j$ . Then

$$M_{X_1}(\lambda) = \mathbb{E} \left[ e^{\lambda X_1} \right] = \begin{cases} \frac{1}{1-\lambda} & \text{if } \lambda < 1 \\ \infty & \text{if } \lambda \geq 1 \end{cases}$$

and for  $a \geq 1$ ,

$$\Psi^*(a) := \sup_{\lambda \geq 0} (\lambda a - \log M_{X_1}(\lambda)) = \sup_{0 \leq \lambda < 1} (\lambda a + \log(1 - \lambda)) = a - 1 - \log a.$$

By Cramér’s theorem, for  $a \geq 1$ ,

$$\lim_{n \rightarrow \infty} \left( -\frac{1}{n} \log \mathbb{P}[S_n \geq na] \right) = \Psi^*(a) = a - 1 - \log a. \tag{2.17}$$

Suppose  $a \geq 1$ . For each leaf of  $\mathcal{T}_l^{reg}$  we use (2.17) to estimate the probability that it is not in  $\mathcal{T}(\mathbf{B}(t))$  and combine with a union bound (summing over leaves). For  $t \geq a \epsilon^2 \lceil A \log \epsilon \rceil$  we have

$$\begin{aligned} & \mathbb{P}^\epsilon \left[ \mathcal{T}(\mathbf{B}(t)) \not\supseteq \mathcal{T}_{\lceil A \log \epsilon \rceil}^{reg} \right] \\ & \leq 3^{\lceil A \log \epsilon \rceil} \mathbb{P} \left[ \epsilon^2 S_{\lceil A \log \epsilon \rceil} \geq a \epsilon^2 \lceil A \log \epsilon \rceil \right] \\ & = \exp \left( \lceil A \log \epsilon \rceil \left( \log 3 + \frac{\log \mathbb{P} \left[ S_{\lceil A \log \epsilon \rceil} \geq a \lceil A \log \epsilon \rceil \right]}{\lceil A \log \epsilon \rceil} \right) \right). \end{aligned} \tag{2.18}$$

By (2.17) (with  $n = \lceil A \log \epsilon \rceil$ ), we can choose  $\epsilon_1(k) < e^{-1}$  such that, for all  $\epsilon \in (0, \epsilon_1)$ ,

$$\frac{\log \mathbb{P} \left[ S_{\lceil A \log \epsilon \rceil} \geq a \lceil A \log \epsilon \rceil \right]}{\lceil A \log \epsilon \rceil} \leq -a + 3/2 + \log a.$$

Choose  $a \geq 1$  sufficiently large that  $-a + 3/2 + \log a \leq -\log 3 - k/A$ . Putting this into (2.18) we obtain

$$\mathbb{P}^\epsilon \left[ \mathcal{T}(\mathbf{B}(t)) \not\supseteq \mathcal{T}_{\lceil A \log \epsilon \rceil}^{reg} \right] \leq \exp(-\lceil \log \epsilon \rceil k)$$

for  $t \geq a \epsilon^2 \lceil A \log \epsilon \rceil$ . 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  $\mathbf{B}(t)$ .

**Lemma 2.11.** *Let  $k \in \mathbb{N}$ , and let  $a_1(k)$  be as in Lemma 2.10. Then there exist  $d_1(k), \epsilon_1(k)$  such that, for all  $\epsilon \in (0, \epsilon_1(k))$  and all  $s \leq a_1 \epsilon^2 \lceil \log \epsilon \rceil$ ,*

$$\mathbb{P}_x^\epsilon \left[ \exists i \in N(s) : |B_i(s) - x| \geq d_1(k) \epsilon \lceil \log \epsilon \rceil \right] \leq \epsilon^k.$$



*Proof.* By rescaling time by  $\epsilon^{-2}$  and space by  $\epsilon^{-1}$ , we have that for  $s \geq 0$ ,

$$\begin{aligned} \mathbb{P}_x^\epsilon [\exists i \in N(s) : |B_i(s) - x| \geq d_1(k)\epsilon |\log \epsilon|] \\ = \mathbb{P}_x^1 [\exists i \in N(\epsilon^{-2}s) : |B_i(\epsilon^{-2}s) - x| \geq d_1(k) |\log \epsilon|]. \end{aligned}$$

Therefore it suffices to show that for  $\epsilon$  sufficiently small, for all  $s \leq a_1 |\log \epsilon|$ ,

$$\mathbb{P}_x^1 [\exists i \in N(s) : |B_i(s) - x| \geq d_1(k) |\log \epsilon|] \leq \epsilon^k.$$

Note that since the maximal displacement of branching Brownian motion at a large time  $t$  is linear in  $t$  rather than  $\sqrt{t}$ , we shall have to take  $d_1$  as a large constant.

Let  $Z$  be a  $N(0, 1)$  distributed random variable. By Markov's inequality, for  $s \leq a_1 |\log \epsilon|$  we have

$$\begin{aligned} \mathbb{P}_x^1 [\exists i \in N(s) : |B_i(s) - x| \geq d_1 |\log \epsilon|] &\leq \mathbb{E}^1 [|N(s)|] \mathbb{P} [\sqrt{2s}|Z| \geq d_1 |\log \epsilon|] \\ &\leq \mathbb{E}^1 [|N(a_1 |\log \epsilon|)|] \mathbb{P} [\sqrt{2a_1 |\log \epsilon|}|Z| \geq d_1 |\log \epsilon|] \\ &= e^{2a_1 |\log \epsilon|} \mathbb{P} [\sqrt{2a_1 |\log \epsilon|}|Z| \geq d_1 |\log \epsilon|^{1/2}] \\ &\leq \frac{1}{\epsilon^{2a_1}} \exp\left(-\frac{1}{4} \frac{d_1^2}{a_1} |\log \epsilon|\right) \\ &= \epsilon^{\frac{1}{4} \frac{d_1^2}{a_1} - 2a_1}. \end{aligned}$$

Here the penultimate line holds for  $\epsilon > 0$  sufficiently small. 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 2.6. If  $z \geq 2d_1\epsilon |\log \epsilon|$ , then, at time  $\delta_1 = a_1\epsilon^2 |\log \epsilon|$ , by Lemma 2.11, with high probability, all individuals in  $\mathbf{B}(\delta_1)$  are still  $\geq d_1\epsilon |\log \epsilon|$ . Lemma 2.8 tells us that there is a positive voting bias at each of those points and Lemma 2.10 shows that this will be magnified by at least  $\mathcal{O}(|\log \epsilon|)$  rounds of majority voting as we trace back to the root. Finally, Lemma 2.9 gives us a lower bound on the bias at the root.

*Proof of Theorem 2.6.* We will prove the first statement of the theorem; the second then follows by symmetry.

For all  $\epsilon < 1/2$ , define  $z_\epsilon$  implicitly by the relation  $\mathbb{P} [B_{T^*} \geq -z_\epsilon] = \frac{1}{2} + \epsilon$ , and note that  $z_\epsilon \sim \epsilon\sqrt{4\pi T^*}$  as  $\epsilon \rightarrow 0$ . Let  $\epsilon_1(k) < 1/2$  be sufficiently small that Lemmas 2.10 and 2.11 hold for  $\epsilon \in (0, \epsilon_1(k))$ . Let  $d_1(k)$  be given by Lemma 2.11 and let  $c_1(k) = 2d_1(k)$  so that (by reducing  $\epsilon_1$  if necessary), for  $\epsilon \in (0, \epsilon_1)$ ,

$$d_1(k)\epsilon |\log \epsilon| + z_\epsilon \leq c_1(k)\epsilon |\log \epsilon|. \tag{2.19}$$

Let  $a_1(k)$  be given by Lemma 2.10 and let

$$\delta_1 = \delta_1(k, \epsilon) = a_1(k)\epsilon^2 |\log \epsilon|. \tag{2.20}$$

If  $t \in (0, \delta_1)$  and  $z \geq c_1\epsilon |\log \epsilon|$ , then

$$\begin{aligned} \mathbb{P}_z^\epsilon [\mathbf{V}(\mathbf{B}(t)) = 0] &\leq \mathbb{P}_z^\epsilon [\exists i \in N(t) \text{ such that } |B_i(t) - z| \geq d_1\epsilon |\log \epsilon|] \\ &\leq \epsilon^k, \end{aligned}$$

where the second line follows by Lemma 2.11.

We now suppose that  $t \in [\delta_1, T^*]$  and  $z \geq c_1 \epsilon |\log \epsilon|$ . Let  $\mathcal{T}_{\delta_1} = \mathcal{T}(\mathbf{B}(\delta_1))$  denote the time-labelled tree of the branching Brownian motion up to time  $\delta_1$ . We define

$$p_{t-\delta_1}(z) = \mathbb{P}_z^\epsilon [\mathbb{V}(\mathbf{B}(t - \delta_1)) = 1],$$

and

$$p_{t-\delta_1}^\epsilon(z) = p_{t-\delta_1}(z_\epsilon), \quad \text{for all } z \in \mathbb{R}.$$

Finally, write  $\{\mathbf{B}(\delta_1) > z_\epsilon\}$  for the event  $B_i(\delta_1) > z_\epsilon$  for all  $i \in N(\delta_1)$ . Then,

$$\begin{aligned} \mathbb{P}_z^\epsilon [\mathbb{V}(\mathbf{B}(t)) = 1] &= \mathbb{P}_z^\epsilon \left[ \mathbb{V}_{p_{t-\delta_1}(z)}(\mathbf{B}(\delta_1)) = 1 \right] \\ &\geq \mathbb{P}_z^\epsilon \left[ \left\{ \mathbb{V}_{p_{t-\delta_1}^\epsilon(z)}(\mathbf{B}(\delta_1)) = 1 \right\} \cap \{\mathbf{B}(\delta_1) > z_\epsilon\} \right] \\ &\geq \mathbb{P}_z^\epsilon \left[ \mathbb{V}_{p_{t-\delta_1}^\epsilon(z)}(\mathbf{B}(\delta_1)) = 1 \right] - \epsilon^k. \end{aligned} \tag{2.21}$$

Here, the first line follows by the Markov property of  $\mathbf{B}$  at time  $\delta_1$ . The second follows by the monotonicity property (2.4). The third line then follows by Lemma 2.11, using (2.19) and our hypothesis that  $z \geq c_1 \epsilon |\log \epsilon|$ .

We have

$$p_{t-\delta_1}^\epsilon(z) \geq P_{z_\epsilon} [B_{t-\delta_1} \geq 0] \geq \frac{1}{2} + \epsilon. \tag{2.22}$$

Here, the first inequality follows from Lemma 2.8. The second follows by the definition of  $z_\epsilon$ , since  $t - \delta_1 < T^*$ .

If  $p_i \geq 1/2$  for  $i = 1, 2, 3$  then (2.12) 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  $\frac{1}{2} + \epsilon$  and  $\mathcal{T}_{\delta_1} \supseteq \mathcal{T}_{A|\log \epsilon}^{reg}$ , then each of the leaves of  $\mathcal{T}_{A|\log \epsilon}^{reg}$  votes 1 independently with probability at least  $\frac{1}{2} + \epsilon$ . Therefore,

$$\mathbb{P}_z^\epsilon [\mathbb{V}(\mathbf{B}(t)) = 1] \geq g^{(\lceil A|\log \epsilon \rceil)}(\frac{1}{2} + \epsilon) - 2\epsilon^k \geq 1 - 3\epsilon^k.$$

Here, the first inequality follows by substituting (2.22) into (2.21) and then applying Lemma 2.10 and the second then follows by Lemma 2.9. This completes the proof. ■

### 2.2.2 The slope of the interface

In proving Theorem 2.4 we shall also exploit a lower bound on the ‘slope’ of the interface in  $d = 1$  which we prove in this subsection. We obtain it as a corollary of the following result.

**Proposition 2.12.** *Suppose  $x \geq 0$  and  $\eta > 0$ . Then for any time-labelled ternary tree  $\mathcal{T}$  and any time  $t$ ,*

$$\mathbb{P}_x^t(\mathcal{T}) - \mathbb{P}_{x-\eta}^t(\mathcal{T}) \geq \mathbb{P}_{x+\eta}^t(\mathcal{T}) - \mathbb{P}_x^t(\mathcal{T}).$$

*Proof.* The proof is by induction on the number of branching events in  $\mathcal{T}$ , and is similar to the proof of Lemma 2.8. For  $\mathcal{T}_0$  a (time-labelled) tree with a root and a single leaf, we have

$$\mathbb{P}_x^t(\mathcal{T}_0) - \mathbb{P}_{x-\eta}^t(\mathcal{T}_0) = \int_{x-\eta}^x \phi_{0,2t}(u) du \geq \int_x^{x+\eta} \phi_{0,2t}(u) du = \mathbb{P}_{x+\eta}^t(\mathcal{T}_0) - \mathbb{P}_x^t(\mathcal{T}_0)$$

where  $\phi_{\mu, \sigma^2}$  is the density of a  $N(\mu, \sigma^2)$  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  $\tau$  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 (2.9),

$$\begin{aligned}
 & (\mathbb{P}_x^t(\mathcal{T}) - \mathbb{P}_{x-\eta}^t(\mathcal{T})) - (\mathbb{P}_{x+\eta}^t(\mathcal{T}) - \mathbb{P}_x^t(\mathcal{T})) \\
 &= (E_x [g(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star))] - E_{x-\eta} [g(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star))]) \\
 &\quad - (E_{x+\eta} [g(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star))] - E_x [g(\mathbb{P}_{B_\tau}^{t-\tau}(\mathcal{T}^\star))]) \\
 &= \int_{-\infty}^{\infty} \left\{ (g(\mathbb{P}_y^{t-\tau}(\mathcal{T}^\star)) - g(\mathbb{P}_{y-\eta}^{t-\tau}(\mathcal{T}^\star))) - (g(\mathbb{P}_{y+\eta}^{t-\tau}(\mathcal{T}^\star)) - g(\mathbb{P}_y^{t-\tau}(\mathcal{T}^\star))) \right\} \phi_{x,2\tau}(y) dy \\
 &= \int_0^{\infty} \left\{ (g(\mathbb{P}_y^{t-\tau}(\mathcal{T}^\star)) - g(\mathbb{P}_{y-\eta}^{t-\tau}(\mathcal{T}^\star))) \right. \\
 &\quad \left. - (g(\mathbb{P}_{y+\eta}^{t-\tau}(\mathcal{T}^\star)) - g(\mathbb{P}_y^{t-\tau}(\mathcal{T}^\star))) \right\} (\phi_{x,2\tau}(y) - \phi_{x,2\tau}(-y)) dy.
 \end{aligned} \tag{2.23}$$

Here, the second line follows by (2.10) and the last line follows from (2.8) and (2.6), which imply that  $g(\mathbb{P}_w^t(\mathcal{T}^\star)) = 1 - g(\mathbb{P}_{-w}^t(\mathcal{T}^\star))$ . Note the similarity to (2.14).

Since  $x \geq 0$ , we have

$$\phi_{x,2\tau}(y) - \phi_{x,2\tau}(-y) \geq 0 \tag{2.24}$$

for  $y \geq 0$ . In view of (2.23) we should like to check that for  $y \geq 0$

$$(g(\mathbb{P}_y^{t-\tau}(\mathcal{T}^\star)) - g(\mathbb{P}_{y-\eta}^{t-\tau}(\mathcal{T}^\star))) - (g(\mathbb{P}_{y+\eta}^{t-\tau}(\mathcal{T}^\star)) - g(\mathbb{P}_y^{t-\tau}(\mathcal{T}^\star))) \geq 0. \tag{2.25}$$

By our inductive hypothesis, for  $y \geq 0$  we have

$$(\mathbb{P}_y^{t-\tau}(\mathcal{T}_i) - \mathbb{P}_{y-\eta}^{t-\tau}(\mathcal{T}_i)) - (\mathbb{P}_{y+\eta}^{t-\tau}(\mathcal{T}_i) - \mathbb{P}_y^{t-\tau}(\mathcal{T}_i)) \geq 0,$$

and so by monotonicity of  $g$ , for (2.25) it is enough to check that

$$g(\mathbb{P}_{y+\eta}^{t-\tau}(\mathcal{T}^\star)) - 2g(\mathbb{P}_y^{t-\tau}(\mathcal{T}^\star)) + g(\mathbb{P}_{y-\eta}^{t-\tau}(\mathcal{T}^\star)) \leq 0. \tag{2.26}$$

To see that (2.26) 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  $p_i = \mathbb{P}_y^{t-\tau}(\mathcal{T}_i)$  and  $\eta_i = \mathbb{P}_{y+\eta}^{t-\tau}(\mathcal{T}_i) - \mathbb{P}_y^{t-\tau}(\mathcal{T}_i)$ . Since for  $y \geq 0$ ,  $p_i \geq 1/2$ , the inequality (2.26) then follows.

Putting (2.24) and (2.25) into (2.23) completes the inductive step, which in turn completes the proof. ■

**Corollary 2.13.** Take  $\epsilon_1(1)$  and  $c_1(1)$  from Theorem 2.6. Let  $\epsilon < \min(\epsilon_1(1), \frac{1}{24})$ . Suppose that for some  $t \in [0, T^*]$  and  $z \in \mathbb{R}$ ,

$$|\mathbb{P}_z^\epsilon [\mathbb{V}(\mathbf{B}(t)) = 1] - \frac{1}{2}| \leq \frac{5}{12}, \tag{2.27}$$

and let  $w \in \mathbb{R}$  with  $|z - w| \leq c_1(1)\epsilon |\log \epsilon|$ . Then

$$|\mathbb{P}_z^\epsilon [\mathbb{V}(\mathbf{B}(t)) = 1] - \mathbb{P}_w^\epsilon [\mathbb{V}(\mathbf{B}(t)) = 1]| \geq \frac{|z - w|}{48c_1(1)\epsilon |\log \epsilon|}. \tag{2.28}$$

*Proof.* Consider first the case  $0 \leq z \leq w$ . By analogy with (2.10), let  $\mathbb{P}_y^t$  denote  $\mathbb{P}_y^\epsilon [\mathbb{V}(\mathbf{B}(t)) = 1]$ . By Theorem 2.6 and (2.27) we have that

$$\mathbb{P}_{c_1(1)\epsilon |\log \epsilon|}^t - \mathbb{P}_z^t \geq 1 - \epsilon - \frac{11}{12} \geq \frac{1}{24}. \tag{2.29}$$

Let  $\eta := w - z$ . For  $j \in \mathbb{N}$ , applying Proposition 2.12  $j$  times gives that

$$\mathbb{P}_{(j+1)\eta+z}^t - \mathbb{P}_{j\eta+z}^t \leq \mathbb{P}_w^t - \mathbb{P}_z^t.$$

It follows that

$$\begin{aligned} \mathbb{P}_{c_1(1)\epsilon|\log \epsilon|}^t - \mathbb{P}_z^t &\leq \sum_{j=0}^{\lceil \eta^{-1}(c_1(1)\epsilon|\log \epsilon|-z) \rceil - 1} (\mathbb{P}_{(j+1)\eta+z}^t - \mathbb{P}_{j\eta+z}^t) \\ &\leq (\eta^{-1}(c_1(1)\epsilon|\log \epsilon|) + 1)(\mathbb{P}_w^t - \mathbb{P}_z^t). \end{aligned} \tag{2.30}$$

Combining (2.29) and (2.30),

$$\mathbb{P}_w^t - \mathbb{P}_z^t \geq \frac{|z - w|}{24(c_1(1)\epsilon|\log \epsilon| + |z - w|)} \geq \frac{|z - w|}{48c_1(1)\epsilon|\log \epsilon|}.$$

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 (2.6). ■

### 2.3 A coupling argument

The second important ingredient in our proof of Theorem 2.4 will be a coupling between  $d(W_s, t - s)$  (the signed distance from a  $d$ -dimensional Brownian motion  $W_s$  to  $\Gamma_{t-s}$ , which evolves according to (backwards in time) mean curvature flow) and a one-dimensional Brownian motion, at least when  $W_s$  is close to  $\Gamma_{t-s}$ . The proof requires some regularity properties of the mean curvature flow that we record in this subsection. These rest on the assumptions (C1)-(C3).

We write  $\dot{d}$  for the time derivative of  $d$ . Let  $T^* \in (0, \mathcal{T})$ . 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

$$|\nabla d(x, t)| = 1. \tag{2.31}$$

Moreover,  $d$  is a  $C^{\alpha, \frac{\alpha}{2}}$  function in  $\{(x, t) : |d(x, t)| \leq c_0, t \leq T^*\}$ .

2. Viewing  $\mathbf{n} = \nabla d$  as the positive normal direction, for  $x \in \Gamma_t$ , the normal velocity of  $\Gamma_t$  at  $x$  is  $-\dot{d}(x, t)$ , and the curvature of  $\Gamma_t$  at  $x$  is  $-\Delta d(x, t)$ . Thus, (1.6) becomes

$$\dot{d}(x, t) = \Delta d(x, t) \tag{2.32}$$

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  $|d(x, t)| \leq c_0$ ,

$$\left| \nabla \left( \dot{d}(x, t) - \Delta d(x, t) \right) \right| \leq C_0. \tag{2.33}$$

4. There exist  $v_0, V_0 > 0$  such that for all  $t \in [0, T^* - v_0]$  and all  $s \in [t, t + v_0]$ ,

$$|d(x, t) - d(x, s)| \leq V_0(s - t). \tag{2.34}$$

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  $\Gamma_t$  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  $T^* < \mathcal{T}$ . Broadly speaking, the first two properties characterize mean curvature flow in terms of the function  $d$ .

A key ingredient of our proof of Theorem 2.4 is the following coupling argument.

**Proposition 2.14.** *Let  $(W_s)_{s \geq 0}$  denote a  $\mathfrak{d}$ -dimensional Brownian motion started at  $x \in \mathbb{R}^{\mathfrak{d}}$ . Suppose that  $t \leq T^*$ ,  $\beta \leq c_0$  and let*

$$T_\beta = \inf (\{s \in [0, t) : |d(W_s, t - s)| \geq \beta\} \cup \{t\}).$$

*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  $s \leq T_\beta$ ,*

$$B_s - C_0\beta s \leq d(W_s, t - s) \leq B_s + C_0\beta s.$$

*Proof.* By Itô's formula, we have that for  $s \leq t$

$$d(W_s, t - s) = \int_0^s A_u du + B_s,$$

where

$$A_u = -\dot{d}(W_u, t - u) + \Delta d(W_u, t - u)$$

$$B_s = \sum_{i=1}^{\mathfrak{d}} \int_0^s \frac{\partial}{\partial x_i} d(W_u, t - u) dW_u^{(i)}.$$

We will handle  $A_u$  and  $B_s$  in turn.

For each  $u \in [0, T_\beta]$  there exists some  $x_u \in \mathbb{R}^{\mathfrak{d}}$  such that  $|x_u - W_u| \leq \beta$ , and  $d(x_u, t - u) = 0$ . By (2.32) we have  $-\dot{d}(x_u, t - u) + \Delta d(x_u, t - u) = 0$ . Since  $\beta \leq c_0$ , by (2.33) 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 (2.31) 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 2.15.** Proposition 2.14 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.

### 2.4 Majority voting in BBM, for $\mathfrak{d} \geq 2$

Recall the notation introduced in Section 2.1 for ternary branching Brownian motion in dimension  $\mathfrak{d} \geq 2$ . For  $x \in \mathbb{R}^{\mathfrak{d}}$ , we write  $\mathbb{P}_x^\epsilon$  for the probability measure under which  $(\mathbf{W}(t), t \geq 0)$  has the law of ternary branching Brownian motion in  $\mathbb{R}^{\mathfrak{d}}$  with branching rate  $1/\epsilon^2$  started from a single particle at location  $x$  at time 0. We use  $\mathbb{E}_x^\epsilon$  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  $\mathfrak{d}$ -dimensional Brownian motion started at  $x$ , and  $E_x$  for the corresponding expectation. As usual the notation  $B$  (resp.  $\mathbf{B}$ ) refers to a one dimensional (historical branching) Brownian motion and  $W$  and  $\mathbf{W}$  signal dimension  $\mathfrak{d} \geq 2$ .

The proof of Theorem 2.4 is in two parts. First, in Section 2.4.1 we establish that the interface is generated in a time  $\delta_{\mathfrak{d}} = \mathcal{O}(\epsilon^2 |\log \epsilon|)$ . We then, in Section 2.4.2, use

Proposition 2.14 and Theorem 2.6 to investigate how the region around the interface propagates. In order not to interrupt the flow of the proof of Theorem 2.4, the proof of a central lemma is deferred to Section 2.4.3.

Our proof rests on a comparison with the outcome  $\mathbb{V}(\mathbf{B}(t))$  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  $\mathbb{V}_p(\mathbf{W}(t))$  and we assume that  $p$  satisfies (C1)-(C3).

### 2.4.1 Generation of the interface

In this section we prove that, as in  $d = 1$ , in dimension  $d \geq 2$  an interface of width  $\mathcal{O}(\epsilon |\log \epsilon|)$  is generated in time  $\mathcal{O}(\epsilon^2 |\log \epsilon|)$ .

**Proposition 2.16.** *Let  $k \in \mathbb{N}$ . Then there exist  $\epsilon_d(k), a_d(k), b_d(k) > 0$  such that for all  $\epsilon \in (0, \epsilon_d)$ , if we set*

$$\delta_d(k, \epsilon) := a_d(k)\epsilon^2 |\log \epsilon| \quad \text{and} \quad \delta'_d(k, \epsilon) := (a_d(k) + k + 1)\epsilon^2 |\log \epsilon|, \tag{2.35}$$

then for  $t \in [\delta_d, \delta'_d]$ ,

1. for  $x$  such that  $d(x, t) \geq b_d \epsilon |\log \epsilon|$ , we have  $\mathbb{P}_x^\epsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] \geq 1 - \epsilon^k$ ;
2. for  $x$  such that  $d(x, t) \leq -b_d \epsilon |\log \epsilon|$ , we have  $\mathbb{P}_x^\epsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] \leq \epsilon^k$ .

*Proof.* By the same argument as for Lemma 2.10, given  $k \in \mathbb{N}$ , and taking  $A(k)$  from Lemma 2.9, there exist  $a_d(k)$  and  $\epsilon_d(k) > 0$  such that, for all  $\epsilon \in (0, \epsilon_d)$  and  $t \geq a_d \epsilon^2 |\log \epsilon|$ ,

$$\mathbb{P}^\epsilon \left[ \mathcal{T}(\mathbf{W}(t)) \supseteq \mathcal{T}_{A(k)|\log \epsilon}^{reg} \right] \geq 1 - \epsilon^k. \tag{2.36}$$

It is also easy to obtain a  $d$ -dimensional equivalent of Lemma 2.11, with essentially the same proof (using a tail bound on a  $d$ -dimensional normal distribution instead of one dimensional). That is, given  $k \in \mathbb{N}$ , there exist  $d_d(k), \epsilon_d(k)$  such that for all  $\epsilon \in (0, \epsilon_d)$ , for  $t \in [\delta_d, \delta'_d]$ ,

$$\mathbb{P}_x^\epsilon [\exists i \in N(t) : |W_i(t) - x| \geq d_d \epsilon |\log \epsilon|] \leq \epsilon^k. \tag{2.37}$$

We set  $b_d(k) = 2d_d(k)$ .

By (2.34) there exist  $v_0, V_0 > 0$  such that for  $t \leq v_0$ , and any  $x \in \mathbb{R}^d$ , we have  $|d(x, 0) - d(x, t)| \leq V_0 t$ . Reducing  $\epsilon_d$  if necessary, for  $\epsilon \in (0, \epsilon_d)$  we have  $\delta'_d \leq v_0$ . Thus, if  $\epsilon \in (0, \epsilon_d)$ ,  $t \in [\delta_d, \delta'_d]$  and  $x$  is such that  $d(x, t) \geq b_d \epsilon |\log \epsilon|$  and  $|W_i(t) - x| \leq d_d \epsilon |\log \epsilon|$  then combining with the triangle inequality and (2.34),

$$\begin{aligned} d(W_i(t), 0) &\geq d(x, t) - |d(x, t) - d(W_i(t), t)| - |d(W_i(t), t) - d(W_i(t), 0)| \\ &\geq b_d \epsilon |\log \epsilon| - d_d \epsilon |\log \epsilon| - V_0 \delta'_d \\ &= \frac{1}{2} b_d \epsilon |\log \epsilon| - V_0 (a_d + k + 1) \epsilon^2 |\log \epsilon|. \end{aligned}$$

Therefore, reducing  $\epsilon_d$  if necessary, in this case we have that

$$d(W_i(t), 0) \geq \frac{1}{4} b_d \epsilon |\log \epsilon|.$$

Applying (C2) and (C3),

$$\begin{aligned} p(W_i(t)) &\geq \frac{1}{2} + \gamma \left( \frac{1}{4} b_d \epsilon |\log \epsilon| \wedge r \right) \\ &\geq \frac{1}{2} + \epsilon, \end{aligned} \tag{2.38}$$

where we again reduce  $\epsilon_{\text{d}} > 0$  (if necessary), to ensure that  $\epsilon < \gamma r$ ,  $\epsilon < \frac{\gamma}{4} b_{\text{d}} \epsilon |\log \epsilon|$  for  $\epsilon \in (0, \epsilon_{\text{d}})$ .

Exactly as in the proof of Theorem 2.6, we can now combine (2.36), (2.37) and (2.38) to deduce that for  $\epsilon \in (0, \epsilon_{\text{d}})$ ,  $t \in [\delta_{\text{d}}, \delta'_{\text{d}}]$  and  $x$  such that  $d(x, t) \geq b_{\text{d}} \epsilon |\log \epsilon|$ ,

$$\mathbb{P}_x^\epsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] \geq 1 - 3\epsilon^k.$$

The proof of the second statement is analogous. ■

### 2.4.2 Propagation of the interface and proof of Theorem 2.4

We now turn to the propagation of the interface region.

**Proposition 2.17.** *Let  $l \in \mathbb{N}$  with  $l \geq 4$ . Define  $a_{\text{d}}(l)$  and  $\delta_{\text{d}}(l, \epsilon)$  as in Proposition 2.16. There exist  $K_1(l), K_2(l) > 0$  and  $\epsilon_{\text{d}}(l, K_1, K_2) > 0$  such that for all  $\epsilon \in (0, \epsilon_{\text{d}})$  and  $t \in [\delta_{\text{d}}(l, \epsilon), T^*]$  we have*

$$\sup_{x \in \mathbb{R}^{\text{d}}} \left( \mathbb{P}_x^\epsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] - \mathbb{P}_{d(x,t)+K_1 e^{K_2 t} \epsilon |\log \epsilon|}^\epsilon [\mathbb{V}(\mathbf{B}(t)) = 1] \right) \leq \epsilon^l \tag{2.39}$$

and

$$\sup_{x \in \mathbb{R}^{\text{d}}} \left( \mathbb{P}_x^\epsilon [\mathbb{V}_p(\mathbf{W}(t)) = 0] - \mathbb{P}_{d(x,t)-K_1 e^{K_2 t} \epsilon |\log \epsilon|}^\epsilon [\mathbb{V}(\mathbf{B}(t)) = 0] \right) \leq \epsilon^l. \tag{2.40}$$

The proof of Theorem 2.4, which follows easily from Proposition 2.17, is at the end of this subsection.

Recall that  $g : [0, 1] \rightarrow [0, 1]$  is given by  $g(p) = 3p^2 - 2p^3$ . It is convenient to extend this definition to a continuous, monotone function  $g : \mathbb{R} \rightarrow [0, 1]$  as follows:

$$g(p) = \begin{cases} 0 & \text{if } p < 0 \\ 3p^2 - 2p^3 & \text{if } p \in [0, 1] \\ 1 & \text{if } p > 1. \end{cases} \tag{2.41}$$

At the heart of the proof of Proposition 2.17 is the following lemma, whose proof we defer to Section 2.4.3.

**Lemma 2.18.** *Let  $l \in \mathbb{N}$  with  $l \geq 4$  and  $K_1 > 0$ . There exists  $K_2 = K_2(K_1, l) > 0$  and  $\epsilon_{\text{d}}(l, K_1, K_2) > 0$  such that for all  $\epsilon \in (0, \epsilon_{\text{d}})$ ,  $x \in \mathbb{R}^{\text{d}}$ ,  $s \in [0, (l + 1)\epsilon^2 |\log \epsilon|]$  and  $t \in [s, T^*]$ ,*

$$\begin{aligned} E_x \left[ g \left( \mathbb{P}_{d(W_s, t-s)+K_1 e^{K_2(t-s)} \epsilon |\log \epsilon|}^\epsilon [\mathbb{V}(\mathbf{B}(t-s)) = 1] + \epsilon^l \right) \right] \\ \leq \frac{3}{4} \epsilon^l + E_{d(x,t)} \left[ g \left( \mathbb{P}_{B_s+K_1 e^{K_2 t} \epsilon |\log \epsilon|}^\epsilon [\mathbb{V}(\mathbf{B}(t-s)) = 1] \right) \right] + \mathbb{1}_{s \leq \epsilon^3} \epsilon^l \end{aligned} \tag{2.42}$$

and

$$\begin{aligned} E_x \left[ g \left( \mathbb{P}_{d(W_s, t-s)-K_1 e^{K_2(t-s)} \epsilon |\log \epsilon|}^\epsilon [\mathbb{V}(\mathbf{B}(t-s)) = 0] + \epsilon^l \right) \right] \\ \leq \frac{3}{4} \epsilon^l + E_{d(x,t)} \left[ g \left( \mathbb{P}_{B_s-K_1 e^{K_2 t} \epsilon |\log \epsilon|}^\epsilon [\mathbb{V}(\mathbf{B}(t-s)) = 0] \right) \right] + \mathbb{1}_{s \leq \epsilon^3} \epsilon^l. \end{aligned} \tag{2.43}$$

*Proof of Proposition 2.17.* Take  $K_1 = b_{\text{d}}(l) + c_1(l)$  where  $b_{\text{d}}$  is as defined in Proposition 2.16 and  $c_1$  is as defined in Theorem 2.6. Let  $K_2 = K_2(K_1, l)$ , as defined in Lemma 2.18. Take  $\epsilon_{\text{d}} > 0$  sufficiently small that Theorem 2.6, Proposition 2.16 and Lemma 2.18 apply for  $\epsilon \in (0, \epsilon_{\text{d}})$ . We begin by observing that for  $\epsilon \in (0, \epsilon_{\text{d}})$ ,  $t \in [\delta_{\text{d}}, \delta'_{\text{d}}]$  (where  $\delta'_{\text{d}}$  is defined in (2.35)), and  $x \in \mathbb{R}^{\text{d}}$ ,

$$\mathbb{P}_x^\epsilon [\mathbb{V}_p(\mathbf{W}(t)) = 1] \leq \mathbb{P}_{d(x,t)+K_1 e^{K_2 t} \epsilon |\log \epsilon|}^\epsilon [\mathbb{V}(\mathbf{B}(t)) = 1] + \epsilon^l. \tag{2.44}$$

To see this, note that if  $d(x, t) \leq -b_{\text{d}}(l)\epsilon|\log \epsilon|$ , then by Proposition 2.16,  $\mathbb{P}_x^\epsilon[\mathbb{V}_p(\mathbf{W}(t)) = 1] \leq \epsilon^l$ . On the other hand, if  $d(x, t) \geq -b_{\text{d}}(l)\epsilon|\log \epsilon|$ , then  $d(x, t) + K_1 e^{K_2 t} \epsilon |\log \epsilon| \geq c_1(l)\epsilon|\log \epsilon|$ , and so, by Theorem 2.6, (2.44) holds (since the right hand side of (2.44) is  $\geq 1$ ).

We are left with the case  $t \in [\delta'_{\text{d}}, T^*]$ . We assume, aiming for a contradiction, that there exists  $t \in [\delta'_{\text{d}}, T^*]$  such that, for some  $x \in \mathbb{R}^{\text{d}}$ ,

$$\mathbb{P}_x^\epsilon[\mathbb{V}_p(\mathbf{W}(t)) = 1] - \mathbb{P}_{d(x,t)+K_1 e^{K_2 t} \epsilon |\log \epsilon|}^\epsilon[\mathbb{V}(\mathbf{B}(t)) = 1] > \epsilon^l.$$

Let  $T'$  be the infimum of the set of such  $t$ . Choose

$$T \in [T', \min(T' + \epsilon^{l+3}, T^*)] \tag{2.45}$$

which is in the set of such  $t$ . Hence, there exists some  $x = x(l, \epsilon) \in \mathbb{R}^{\text{d}}$  such that

$$\mathbb{P}_x^\epsilon[\mathbb{V}_p(\mathbf{W}(T)) = 1] - \mathbb{P}_{d(x,T)+K_1 e^{K_2 T} \epsilon |\log \epsilon|}^\epsilon[\mathbb{V}(\mathbf{B}(T)) = 1] > \epsilon^l. \tag{2.46}$$

We now seek to show that

$$\mathbb{P}_x^\epsilon[\mathbb{V}_p(\mathbf{W}(T)) = 1] \leq \frac{7}{8}\epsilon^l + \mathbb{P}_{d(x,T)+K_1 e^{K_2 T} \epsilon |\log \epsilon|}^\epsilon[\mathbb{V}(\mathbf{B}(T)) = 1]. \tag{2.47}$$

Since  $\frac{7}{8}\epsilon^l < \epsilon^l$ , once we obtain equation (2.47) we have a contradiction to (2.46), thus completing the proof.

We write  $S$  for the time of the first branching event in  $\mathbf{W}(T)$  and  $W_S$  for the position of the initial ‘ancestor’ particle at that time. We note that by the strong Markov property at time  $S \wedge (T - \delta_{\text{d}})$ ,

$$\begin{aligned} \mathbb{P}_x^\epsilon[\mathbb{V}_p(\mathbf{W}(T)) = 1] &= \mathbb{E}_x^\epsilon[g(\mathbb{P}_{W_S}^\epsilon[\mathbb{V}_p(\mathbf{W}(T - S)) = 1])\mathbb{1}_{S \leq T - \delta_{\text{d}}}] \\ &\quad + \mathbb{E}_x^\epsilon\left[\mathbb{P}_{W_{T - \delta_{\text{d}}}}^\epsilon[\mathbb{V}_p(\mathbf{W}(\delta_{\text{d}})) = 1]\mathbb{1}_{S \geq T - \delta_{\text{d}}}\right]. \end{aligned} \tag{2.48}$$

We begin with the second term on the right of (2.48). Since  $T - \delta_{\text{d}} \geq \delta'_{\text{d}} - \delta_{\text{d}} = (l + 1)\epsilon^2|\log \epsilon|$  and  $S \sim \text{Exp}(\epsilon^{-2})$ ,

$$\mathbb{E}_x^\epsilon\left[\mathbb{P}_{W_{T - \delta_{\text{d}}}}^\epsilon[\mathbb{V}_p(\mathbf{W}(\delta_{\text{d}})) = 1]\mathbb{1}_{S \geq T - \delta_{\text{d}}}\right] \leq \mathbb{P}^\epsilon[S \geq (l + 1)\epsilon^2|\log \epsilon|] = \epsilon^{l+1}. \tag{2.49}$$

To bound the first term on the right of (2.48), partition on the event  $\{S \leq \epsilon^{l+3}\}$  (which has probability  $\leq \epsilon^{l+1}$ ):

$$\begin{aligned} &\mathbb{E}_x^\epsilon[g(\mathbb{P}_{W_S}^\epsilon[\mathbb{V}_p(\mathbf{W}(T - S)) = 1])\mathbb{1}_{S \leq T - \delta_{\text{d}}}] \\ &\leq \mathbb{P}^\epsilon[S \leq \epsilon^{l+3}] + \mathbb{E}_x^\epsilon[g(\mathbb{P}_{W_S}^\epsilon[\mathbb{V}_p(\mathbf{W}(T - S)) = 1])\mathbb{1}_{S \leq T - \delta_{\text{d}}}\mathbb{1}_{S \geq \epsilon^{l+3}}] \\ &\leq \epsilon^{l+1} + \mathbb{E}_x^\epsilon\left[g\left(\mathbb{P}_{d(W_S, T - S)+K_1 e^{K_2(T - S)} \epsilon |\log \epsilon|}^\epsilon[\mathbb{V}(\mathbf{B}(T - S)) = 1] + \epsilon^l\right)\mathbb{1}_{S \leq T - \delta_{\text{d}}}\right]. \end{aligned} \tag{2.50}$$

The last inequality follows from the minimality of  $T'$  (note that if  $\epsilon^{l+3} \leq S \leq T - \delta_{\text{d}}$ , then  $T - S \in [\delta_{\text{d}}, T')$  by (2.45)) and from monotonicity of  $g$ .

Conditioning on the value of  $S$ , since the path of the ancestor particle ( $W$ .) is independent of  $S$ ,



$$\begin{aligned}
 & \mathbb{E}_x^\epsilon \left[ g \left( \mathbb{P}_{d(W_S, T-S) + K_1 e^{K_2(T-S)} \epsilon |\log \epsilon|}^\epsilon [\mathbf{V}(\mathbf{B}(T-S)) = 1] + \epsilon^l \right) \mathbb{1}_{S \leq T - \delta_d} \right] \\
 & \leq \int_0^{(l+1)\epsilon^2 |\log \epsilon|} \epsilon^{-2} e^{-\epsilon^{-2}s} \\
 & \quad E_x \left[ g \left( \mathbb{P}_{d(W_s, T-s) + K_1 e^{K_2(T-s)} \epsilon |\log \epsilon|}^\epsilon [\mathbf{V}(\mathbf{B}(T-s)) = 1] + \epsilon^l \right) \right] ds \\
 & \quad + \mathbb{P}^\epsilon [S \geq (l+1)\epsilon^2 |\log \epsilon|] \\
 & \leq \int_0^{(l+1)\epsilon^2 |\log \epsilon|} \epsilon^{-2} e^{-\epsilon^{-2}s} E_{d(x, T)} \left[ g \left( \mathbb{P}_{B_s + K_1 e^{K_2 T} \epsilon |\log \epsilon|}^\epsilon [\mathbf{V}(\mathbf{B}(T-s)) = 1] \right) \right] ds \\
 & \quad + \frac{3}{4} \epsilon^l + \mathbb{P}^\epsilon [S \leq \epsilon^3] \epsilon^l + \epsilon^{l+1} \\
 & \leq \frac{3}{4} \epsilon^l + 2\epsilon^{l+1} + \mathbb{E}_{d(x, T)}^\epsilon \left[ g \left( \mathbb{P}_{B_{S'} + K_1 e^{K_2 T} \epsilon |\log \epsilon|}^\epsilon [\mathbf{V}(\mathbf{B}(T-S')) = 1] \right) \mathbb{1}_{S' \leq T - \delta_d} \right]. \quad (2.51)
 \end{aligned}$$

Here, the second inequality follows by Lemma 2.18. For the final inequality, we write  $S'$  for the time of the first branching event in  $(\mathbf{B}(s), s \geq 0)$  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  $T \geq \delta'_d$  and so  $T - \delta_d \geq (l+1)\epsilon^2 |\log \epsilon|$ .

Putting (2.50), (2.51) and (2.49) into (2.48) we obtain

$$\begin{aligned}
 & \mathbb{P}_x^\epsilon [\mathbf{V}_p(\mathbf{W}(T)) = 1] \\
 & \leq 4\epsilon^{l+1} + \frac{3}{4} \epsilon^l + \mathbb{E}_{d(x, T)}^\epsilon \left[ g \left( \mathbb{P}_{B_{S'} + K_1 e^{K_2 T} \epsilon |\log \epsilon|}^\epsilon [\mathbf{V}(\mathbf{B}(T-S')) = 1] \right) \mathbb{1}_{S' \leq T - \delta_d} \right] \\
 & \leq 4\epsilon^{l+1} + \frac{3}{4} \epsilon^l + \mathbb{P}_{d(x, T) + K_1 e^{K_2 T} \epsilon |\log \epsilon|}^\epsilon [\mathbf{V}(\mathbf{B}(T)) = 1],
 \end{aligned}$$

where the second inequality follows by the strong Markov Property for  $(\mathbf{B}(\cdot))$  at time  $S' \wedge (T - \delta_d)$ , in similar style to (2.48). Reducing  $\epsilon_d$ , if necessary, to ensure that  $\frac{3}{4} \epsilon^l + 4\epsilon^{l+1} \leq \frac{7}{8} \epsilon^l$  for all  $\epsilon \in (0, \epsilon_d)$ , we obtain (2.47), which completes the proof of (2.39).

By a similar argument, using (2.43) in place of (2.42), we can also deduce (2.40). ■

*Proof of Theorem 2.4.* It suffices to prove the result for sufficiently large  $k \in \mathbb{N}$ , and in particular we will show it for  $k \geq 4$ .

We choose  $c_d(k) = c_1(k) + K_1 e^{K_2 T^*}$ . Thus, for any  $t \in [\delta_d, T^*]$  and  $x \in \mathbb{R}^d$  such that  $d(x, t) \leq -c_d(k)\epsilon |\log \epsilon|$  we have

$$d(x, t) + K_1 e^{K_2 t} \epsilon |\log \epsilon| \leq -c_1(k)\epsilon |\log \epsilon|.$$

It follows from Theorem 2.6 (reducing  $\epsilon_d$  if necessary so that  $\epsilon < \epsilon_1(k)$ ) and (2.39) that  $\mathbb{P}_x [\mathbf{V}_p(\mathbf{W}(t)) = 1] \leq 2\epsilon^k$  for such  $x$  and  $t$ . Similarly, for  $x$  such that  $d(x, t) \geq c_d(k)\epsilon |\log \epsilon|$ , by Theorem 2.6 and (2.40) we have  $\mathbb{P}_x [\mathbf{V}_p(\mathbf{W}(t)) = 0] \leq 2\epsilon^k$ . ■

### 2.4.3 Proof of Lemma 2.18

To complete the proof of Theorem 2.4, it remains to prove Lemma 2.18. 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 (2.42). (The proof of (2.43) goes along essentially the same lines.)

We take a large constant  $C$  and consider the cases  $|d(x, t)| \geq C\epsilon |\log \epsilon|$  and  $|d(x, t)| \leq C\epsilon |\log \epsilon|$  separately. Since  $s = \mathcal{O}(\epsilon^2 |\log \epsilon|)$ , with high probability neither the  $d$ -dimensional Brownian motion  $W$  nor the one-dimensional  $B$  moves a distance more than  $\mathcal{O}(\epsilon |\log \epsilon|)$  before time  $s$ . Therefore, if  $C$  is sufficiently large and  $d(x, t) \leq -C\epsilon |\log \epsilon|$ , Theorem 2.6 tells us that the left-hand side of (2.42) is  $\leq \epsilon^{l+1}$ ; similarly, if  $d(x, t) \geq C\epsilon |\log \epsilon|$  then the right-hand side of (2.42) is  $\geq 1$ . This leaves the case of  $|d(x, t)| \leq C\epsilon |\log \epsilon|$ , in

which we apply Proposition 2.14 to couple  $W_s$  with  $B_s$  in such a way that with probability  $1 - \mathcal{O}(\epsilon^{l+1})$ ,

$$d(W_s, t - s) \leq B_s + \mathcal{O}(\epsilon |\log \epsilon|)s.$$

Thus, using monotonicity (2.4), the left-hand side of (2.42) is bounded above by

$$\mathbb{E}_{d(x,t)} \left[ g \left( \mathbb{P}_{B_s + (K_1 e^{K_2(t-s)} + \mathcal{O}(s))\epsilon |\log \epsilon|}^\epsilon [\mathbb{V}(\mathbf{B}(t-s)) = 1] + \epsilon^l \right) \right] + \mathcal{O}(\epsilon^{l+1}).$$

If  $|p - \frac{1}{2}| \geq \frac{7}{18}$ , we can use that  $|g'(p)| \leq 2/3$  to pull the  $\epsilon^l$  outside the argument of  $g$  and then use monotonicity again to recover (2.42). The difficulty is that close to  $p = \frac{1}{2}$ , we have  $g'(p) > 1$ . In the case

$$\mathbb{P}_{B_s + (K_1 e^{K_2(t-s)} + \mathcal{O}(s))\epsilon |\log \epsilon|}^\epsilon [\mathbb{V}(\mathbf{B}(t-s)) = 1] \approx \frac{1}{2},$$

we instead choose  $K_2 \gg 0$ , and use the lower bound on the ‘slope of the interface’ given by Corollary 2.13 to estimate the increment in  $\mathbb{P}_z^\epsilon [\mathbb{V}(\mathbf{B}(t-s)) = 1]$  when we replace  $z + (K_1 e^{K_2(t-s)} + \mathcal{O}(s))\epsilon |\log \epsilon|$  by  $z + K_1 e^{K_2 t} \epsilon |\log \epsilon|$ .

The remainder of this subsection contains the formal proof.

*Proof of Lemma 2.18.* We begin by proving (2.42). For the duration of the proof, for  $u \geq 0$  and  $z \in \mathbb{R}$  we write

$$\mathbb{Q}_z^{\epsilon,u} = \mathbb{P}_z^\epsilon [\mathbb{V}(\mathbf{B}(u)) = 1].$$

Recall  $C_0$  and  $c_1(k)$  from (2.33) and Theorem 2.6 respectively. Let

$$R = 2c_1(l) + 4(l+1)d + 1. \tag{2.52}$$

Fix  $K_2$  such that

$$K_1(K_2 - C_0) - C_0 R = c_1(1). \tag{2.53}$$

Let  $\epsilon_d = \epsilon_1(l)$  where  $\epsilon_1(l)$  is defined in Theorem 2.6.

First we need an estimate for the probability that a  $d$ -dimensional Brownian motion moves further than  $\sim \epsilon |\log \epsilon|$  in time  $s$  (recall that  $s \leq (l+1)\epsilon^2 |\log \epsilon|$ ). Let

$$A_x = \left\{ \sup_{u \in [0,s]} |W_u - x| \leq 2(l+1)d \epsilon |\log \epsilon| \right\}.$$

Then bounding  $|W_u|$  by the sum of the moduli of  $d$  one-dimensional Brownian motions and using the reflectional symmetry of one dimensional Brownian motion,

$$\begin{aligned} P_x [A_x^c] &\leq 2dP_0 \left[ \sup_{u \in [0,s]} B_u > 2(l+1)\epsilon |\log \epsilon| \right] \\ &\leq 4dP_0 \left[ B_1 > 2((l+1)|\log \epsilon|)^{1/2} \right] \\ &\leq 4d\epsilon^{l+1}. \end{aligned} \tag{2.54}$$

Here, since  $s \leq (l+1)\epsilon^2 |\log \epsilon|$  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)  $d(x, t) \leq -(2c_1(l) + 2(l+1)d + K_1 e^{K_2(t-s)}) \epsilon |\log \epsilon|$ ,
- (ii)  $d(x, t) \geq (2c_1(l) + 2(l+1)d + K_1 e^{K_2(t-s)}) \epsilon |\log \epsilon|$ ,
- (iii)  $|d(x, t)| \leq (2c_1(l) + 2(l+1)d + K_1 e^{K_2(t-s)}) \epsilon |\log \epsilon|$ .

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 (2.34) there exist  $v_0, V_0 > 0$  such that if  $s \leq v_0$  and  $x \in \mathbb{R}^d$  then

$$|d(x, t) - d(x, t - s)| \leq V_0 s. \tag{2.55}$$

We reduce  $\epsilon_{\text{d}}$ , if necessary, to ensure that for  $\epsilon \in (0, \epsilon_{\text{d}})$  we have  $(l + 1)\epsilon^2 |\log \epsilon| \leq v_0$ . Then if the event  $A_x$  occurs,

$$\begin{aligned} d(W_s, t - s) + K_1 e^{K_2(t-s)} \epsilon |\log \epsilon| &\leq -(2c_1(l) + 2(l + 1)\text{d})\epsilon |\log \epsilon| + |d(W_s, t - s) - d(x, t)| \\ &\leq -(2c_1(l) + 2(l + 1)\text{d})\epsilon |\log \epsilon| + |d(x, t) - d(x, t - s)| + |W_s - x| \\ &\leq -2c_1(l)\epsilon |\log \epsilon| + V_0(l + 1)\epsilon^2 |\log \epsilon|. \end{aligned} \tag{2.56}$$

Here, the second line follows from being in case (i) and the third follows from the triangle inequality. The final line then follows from (2.55) and that  $s \leq (l + 1)\epsilon^2 |\log \epsilon|$ , and since  $A_x$  occurs.

Reducing  $\epsilon_{\text{d}}$ , if necessary, from (2.56) we have

$$d(W_s, t - s) + K_1 e^{K_2(t-s)} \epsilon |\log \epsilon| \leq -c_1(l)\epsilon |\log \epsilon|.$$

Therefore

$$\begin{aligned} E_x \left[ g \left( \mathbb{Q}_{d(W_s, t-s) + K_1 e^{K_2(t-s)} \epsilon |\log \epsilon|}^{\epsilon, t-s} + \epsilon^l \right) \right] &\leq E_x [g(\epsilon^l + \epsilon^l) \mathbb{1}_{A_x}] + P_x [A_x^c] \\ &\leq 6\epsilon^{2l} + 4\text{d}\epsilon^{l+1}. \end{aligned}$$

Here the first inequality follows by Theorem 2.6 and the second inequality by the definition of  $g$  in (2.41) and by (2.54). Again reducing  $\epsilon_{\text{d}}$  if necessary, for  $\epsilon \in (0, \epsilon_{\text{d}})$  we have

$$E_x [g(\mathbb{Q}_{d(W_s, t-s) + K_1 e^{K_2(t-s)} \epsilon |\log \epsilon|}^{\epsilon, t-s} + \epsilon^l)] \leq \frac{3}{4}\epsilon^l,$$

and so (2.42) holds in this case.

Case (ii): In this case, we have that  $d(x, t) \geq (c_1(l) + 2(l + 1))\epsilon |\log \epsilon|$ . A similar argument to that used for (2.54) gives us that

$$P_{d(x,t)} [B_s \leq c_1(l)\epsilon |\log \epsilon|] \leq \epsilon^{l+1}. \tag{2.57}$$

It follows that in this case

$$\begin{aligned} E_{d(x,t)} \left[ g \left( \mathbb{Q}_{B_s + K_1 e^{K_2 t} \epsilon |\log \epsilon|}^{\epsilon, t-s} \right) \right] &\geq E_{d(x,t)} \left[ g \left( \mathbb{Q}_{B_s + K_1 e^{K_2 t} \epsilon |\log \epsilon|}^{\epsilon, t-s} \right) \mathbb{1}_{\{B_s \geq c_1(l)\epsilon |\log \epsilon|\}} \right] \\ &\geq g(1 - \epsilon^l) - \epsilon^{l+1} \\ &\geq 1 - 3\epsilon^{2l} - \epsilon^{l+1}, \end{aligned}$$

where the second inequality follows by Theorem 2.6 and (2.57) and the last inequality by the definition of  $g$  in (2.41). Again reducing  $\epsilon_{\text{d}}$  if necessary, for  $\epsilon \in (0, \epsilon_{\text{d}})$  we have

$$E_{d(x,t)} \left[ g(\mathbb{Q}_{B_s + K_1 e^{K_2 t} \epsilon |\log \epsilon|}^{\epsilon, t-s}) \right] \geq 1 - \frac{3}{4}\epsilon^l$$

and so (2.42) 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

$$\begin{aligned} |d(W_u, t - u)| &\leq |W_u - x| + |d(x, t)| + |d(x, t) - d(x, t - u)| \\ &\leq (2c_1(l) + 4(l + 1)\text{d})\epsilon |\log \epsilon| + K_1 e^{K_2(t-s)} \epsilon |\log \epsilon| + V_0(l + 1)\epsilon^2 |\log \epsilon|, \end{aligned}$$

where the second line follows by (2.55). Reducing  $\epsilon_{\text{d}}$  if necessary, for  $\epsilon \in (0, \epsilon_{\text{d}})$  we have

$$|d(W_u, t - u)| \leq (R + K_1 e^{K_2(t-s)})\epsilon |\log \epsilon|, \tag{2.58}$$

where  $R$  is defined in (2.52). We now apply Proposition 2.14 with

$$\beta = (R + K_1 e^{K_2(t-s)})\epsilon |\log \epsilon|. \tag{2.59}$$

By reducing  $\epsilon_{\text{d}}$  if necessary, we have for  $\epsilon \in (0, \epsilon_{\text{d}})$  that  $\beta \leq c_0$ . Define

$$T_\beta = \inf(\{u \in [0, t) : |d(W_u, t - u)| \geq \beta\} \cup \{t\}).$$

Then by Proposition 2.14, we can couple  $(W_u)_{u \geq 0}$  with  $(B_u)_{u \geq 0}$ , a one-dimensional Brownian motion started from  $d(x, t)$ , in such a way that for  $u \leq T_\beta$ ,

$$d(W_u, t - u) \leq B_u + C_0 \beta u. \tag{2.60}$$

Hence

$$\begin{aligned} E_x \left[ g\left(\mathbb{Q}_{d(W_s, t-s) + K_1 e^{K_2(t-s)}\epsilon |\log \epsilon| + \epsilon^l}\right) \right] \\ \leq E_{d(x,t)} \left[ g\left(\mathbb{Q}_{B_s + C_0 \beta s + K_1 e^{K_2(t-s)}\epsilon |\log \epsilon| + \epsilon^l}\right) \right] + P_x [T_\beta \leq s] \\ \leq E_{d(x,t)} \left[ g\left(\mathbb{Q}_{B_s + C_0 \beta s + K_1 e^{K_2(t-s)}\epsilon |\log \epsilon| + \epsilon^l}\right) \right] + 4\text{d}\epsilon^{l+1}. \end{aligned} \tag{2.61}$$

Here, the first inequality follows by (2.60), (2.4) and the monotonicity of  $g$ . The second inequality then follows by (2.54) (note that by (2.58), if  $A_x$  occurs then  $T_\beta \geq s$ ).

Now let

$$E = \left\{ \left| \mathbb{Q}_{B_s + C_0 \beta s + K_1 e^{K_2(t-s)}\epsilon |\log \epsilon| - \frac{1}{2}} \right| \leq \frac{5}{12} \right\}.$$

We shall consider the cases  $E$  and  $E^c$  separately to bound the right hand side of (2.61).

Consider first when the event  $E$  occurs. Note that by the definition of  $\beta$  in (2.59),

$$\begin{aligned} K_1 e^{K_2 t} \epsilon |\log \epsilon| - \left( C_0 \beta s + K_1 e^{K_2(t-s)} \epsilon |\log \epsilon| \right) \\ = \left( K_1 e^{K_2(t-s)} (e^{K_2 s} - 1 - C_0 s) - C_0 R s \right) \epsilon |\log \epsilon| \\ \geq (K_1(K_2 - C_0) - C_0 R) s \epsilon |\log \epsilon| \\ = c_1(1) s \epsilon |\log \epsilon|, \end{aligned} \tag{2.62}$$

where the inequality follows since  $K_2 > 0$  and the last line follows by (2.53). Reducing  $\epsilon_{\text{d}}$  if necessary so that  $\epsilon_{\text{d}} < \min(\epsilon_1(1), \frac{1}{24})$ , for  $\epsilon \in (0, \epsilon_{\text{d}})$  we can apply Corollary 2.13 with  $z = B_s + C_0 \beta s + K_1 e^{K_2(t-s)}\epsilon |\log \epsilon|$  and  $w = z + c_1(1) s \epsilon |\log \epsilon| \leq B_s + K_1 e^{K_2 t} \epsilon |\log \epsilon|$  to give that

$$\mathbb{Q}_{B_s + C_0 \beta s + K_1 e^{K_2(t-s)}\epsilon |\log \epsilon|}^{\epsilon, t-s} \mathbb{1}_E \leq \left( \mathbb{Q}_{B_s + K_1 e^{K_2 t} \epsilon |\log \epsilon|}^{\epsilon, t-s} - \frac{1}{48} s \right) \mathbb{1}_E. \tag{2.63}$$

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

$$g(p + \delta) \leq g(p) + \frac{2}{3} \delta. \tag{2.64}$$

Reducing  $\epsilon_{\text{d}}$  if necessary so that  $\frac{1}{12} + \epsilon^l < \frac{1}{9}$  for  $\epsilon \in (0, \epsilon_{\text{d}})$ , we have

$$\begin{aligned} g\left(\mathbb{Q}_{B_s + C_0 \beta s + K_1 e^{K_2(t-s)}\epsilon |\log \epsilon| + \epsilon^l}\right) \mathbb{1}_{E^c} &\leq \left( g\left(\mathbb{Q}_{B_s + C_0 \beta s + K_1 e^{K_2(t-s)}\epsilon |\log \epsilon|}\right) + \frac{2}{3} \epsilon^l \right) \mathbb{1}_{E^c} \\ &\leq \left( g\left(\mathbb{Q}_{B_s + K_1 e^{K_2 t} \epsilon |\log \epsilon|}\right) + \frac{2}{3} \epsilon^l \right) \mathbb{1}_{E^c}, \end{aligned} \tag{2.65}$$

where the first inequality follows by (2.64) and the last inequality by (2.62) and monotonicity of  $g$ .

Putting (2.63) and (2.65) into (2.61),

$$\begin{aligned} & E_x \left[ g \left( \mathbb{Q}_{d(W_s, t-s) + K_1 e^{K_2(t-s)\epsilon} |\log \epsilon|}^{\epsilon, t-s} + \epsilon^l \right) \right] \\ & \leq E_{d(x,t)} \left[ g \left( \mathbb{Q}_{B_s + K_1 e^{K_2 t \epsilon} |\log \epsilon|}^{\epsilon, t-s} - \frac{1}{48} s + \epsilon^l \right) \mathbb{1}_E \right] \\ & \quad + E_{d(x,t)} \left[ \left( g \left( \mathbb{Q}_{B_s + K_1 e^{K_2 t \epsilon} |\log \epsilon|}^{\epsilon, t-s} \right) + \frac{2}{3} \epsilon^l \right) \mathbb{1}_{E^c} \right] + 4\mathfrak{d}\epsilon^{l+1} \\ & \leq E_{d(x,t)} \left[ g \left( \mathbb{Q}_{B_s + K_1 e^{K_2 t \epsilon} |\log \epsilon|}^{\epsilon, t-s} \right) \right] + \frac{2}{3} \epsilon^l + \epsilon^l \mathbb{1}_{s \leq 48\epsilon^l} + 4\mathfrak{d}\epsilon^{l+1}, \end{aligned}$$

where the last inequality follows in the case  $s \leq 48\epsilon^l$  since  $|g'(p)| \leq \frac{3}{2}$  for all  $p \in [0, 1]$ . Reducing  $\epsilon_{\mathfrak{d}}$ , if necessary, so that  $4\mathfrak{d}\epsilon^{l+1} \leq \frac{1}{12}\epsilon^l$  and  $48\epsilon^l \leq \epsilon^3$  for  $\epsilon \in (0, \epsilon_{\mathfrak{d}})$  completes the proof of (2.42).

The second statement of the lemma, equation (2.43), is proved by the same argument, considering  $\{\mathbb{V}(\mathbf{B}(u)) = 0\}$  instead of  $\{\mathbb{V}(\mathbf{B}(u)) = 1\}$  and using  $d(W_u, t-u) \geq B_u - C_0\beta u$  for  $u \leq T_\beta$  in place of (2.60). ■

### 3 Proof of Theorem 1.8

In this section we turn to the proof of our central result, Theorem 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 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.

#### 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 (1.12) that  $\Pi^n$  is a Poisson point process on  $\mathbb{R}_+ \times \mathbb{R}^{\mathfrak{d}} \times (0, \infty)$  with intensity measure

$$n dt \otimes n^\beta dx \otimes \mu^n(dr).$$

We also let

$$u_n = \frac{u}{n^{1-2\beta}}, \quad \text{and} \quad \mathfrak{s}_n = \frac{1}{\epsilon^2} \frac{1}{n^{2\beta}}.$$

**Definition 3.1** (SLFVS dual). *For  $n \in \mathbb{N}$ , the process  $(\mathcal{P}_t^n)_{t \geq 0}$  is the  $\bigcup_{l \geq 1} (\mathbb{R}^{\mathfrak{d}})^l$ -valued Markov process with dynamics defined as follows.*

*The process is started with a single individual  $\mathcal{P}_0^n = x$  and for  $t \geq 0$ ,  $\mathcal{P}_t^n = (\xi_1^n(t), \dots, \xi_{N(t)}^n(t))$  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  $1 - \mathfrak{s}_n$ . 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  $\mathcal{B}_r(x)$ .*

*With the complementary probability  $\mathfrak{s}_n$ , 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 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 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,

$$\Xi^n(t) := (\mathcal{P}_s^n)_{0 \leq s \leq t}.$$

We write  $\mathbb{P}_x$  for the law of  $\Xi^n$  when  $\mathcal{P}_0^n$  is the single point  $x$  and  $\mathbb{E}_x$  for the corresponding expectation. For  $\mathbf{i} \in \{1, 2, 3\}^{\mathbb{N}}$  with  $\mathbf{i} = (i_1, i_2, \dots)$ , we let  $(\xi_{\mathbf{i}}^n(s), 0 \leq s \leq t) \subseteq \Xi^n(t)$  denote the  $\mathbb{R}^d$ -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  $i_k^{\text{th}}$  offspring when it is affected by its  $k^{\text{th}}$  selective event. We shall refer to  $(\xi_{\mathbf{i}}^n(s), 0 \leq s \leq t)$  as an ancestral lineage.

The voting procedure on  $\Xi^n(t)$  is a minor modification of Definition 2.1. Let  $p : \mathbb{R}^d \rightarrow [0, 1]$  be a fixed function. Recalling that the set of individuals in  $\mathcal{P}_t^n$  is  $\{\xi_1^n(t), \dots, \xi_{N(t)}^n(t)\}$ , for each  $j \leq N(t)$ , 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  $\Xi^n(t)$ ,

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  $\Pi^n$ , 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 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  $\Xi^n(t)$  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  $x \in \mathbb{R}^d$ . However, if,  $\psi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ , then the function

$$\int_{\mathbb{R}^d} \psi(x) w_t^n(x) dx,$$

is well-defined.

**Theorem 3.4.** *The spatial  $\Lambda$ -Fleming-Viot process with selection driven by  $\Pi^n$ ,  $(w_t^n(x), x \in \mathbb{R}^d)_{t \geq 0}$ , is dual to the historical process  $(\Xi^n(t), t \geq 0)$  in the sense that for every  $\psi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ , we have*

$$\begin{aligned} \mathbb{E}_p \left[ \int_{\mathbb{R}^d} \psi(x) w_t^n(x) dx \right] &= \int_{\mathbb{R}^d} \psi(x) \mathbb{E}_x \left[ \mathbb{V}_p(\Xi^n(t)) \right] dx \\ &= \int_{\mathbb{R}^d} \psi(x) \mathbb{P}_x \left[ \mathbb{V}_p(\Xi^n(t)) = 1 \right] dx. \end{aligned} \tag{3.1}$$

**Remark 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 3.4 using generators is a simple extension of that of the corresponding duality for the spatial  $\Lambda$ -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 1.8 to the following analogue of Theorem 2.4.

**Theorem 3.6.** *Take  $\sigma^2$  as in (1.14). Suppose that  $\beta \in (0, 1/4)$  and let  $\epsilon_n$  be a sequence such that  $\epsilon_n \rightarrow 0$  and  $(\log n)^{1/2}\epsilon_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Assume  $p$  satisfies (C1)–(C3) and define  $\mathcal{T}$ ,  $d(x, t)$  as for Theorem 1.3; take  $T^* < \mathcal{T}$ . Let  $k \in \mathbb{N}$ . There exist  $n_*(k) \in \mathbb{N}$ , and  $a_*(k), d_*(k) \in (0, \infty)$  such that for all  $n \geq n^*$  and all  $t$  satisfying  $a_*\epsilon_n^2 |\log \epsilon_n| \leq t \leq T^*$ ,*

1. *for  $x$  such that  $d(x, \sigma^2 t) \geq d_*\epsilon_n |\log \epsilon_n|$ , we have  $\mathbb{P}_x [\mathbb{V}_p(\Xi^n(t)) = 1] \geq 1 - \epsilon_n^k$ .*
2. *for  $x$  such that  $d(x, \sigma^2 t) \leq -d_*\epsilon_n |\log \epsilon_n|$ , we have  $\mathbb{P}_x [\mathbb{V}_p(\Xi^n(t)) = 1] \leq \epsilon_n^k$ .*

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

$$m_n(dz) = nu_n n^{d\beta} \int_0^{\mathcal{R}_n} \frac{V_r(0, z)}{V_r} \mu^n(dr) dz, \tag{3.2}$$

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  $n n^{d\beta} V_r(0, z) \mu^n(dr)$ . The offspring location is chosen uniformly from the ball  $\mathcal{B}_r(x)$ , so the probability that  $z$  is chosen as the offspring location is  $dz/V_r$  and the probability that our lineage is actually affected by the event is  $u_n$ . Combining these yields (3.2).

The total rate of jumps is

$$\begin{aligned} \int_{\mathbb{R}^d} m_n(dz) &= \int_0^{\mathcal{R}_n} nu_n n^{d\beta} \frac{1}{V_r} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_{|x|<r} \mathbb{1}_{|x-z|<r} dx dz \mu^n(dr) \\ &= \int_0^{\mathcal{R}_n} nu_n n^{d\beta} V_r \mu^n(dr) \\ &= n^{2\beta} u V_1 \int_0^{\mathcal{R}} r^d \mu(dr), \end{aligned} \tag{3.3}$$

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:

$$\begin{aligned} \frac{1}{2d} \int_{\mathbb{R}^d} |z|^2 m_n(dz) &= \frac{1}{2d} \int_{\mathbb{R}^d} |z|^2 nu_n \int_0^{\mathcal{R}_n} n^{d\beta} \frac{V_r(0, z)}{V_r} \mu^n(dr) dz \\ &= \frac{u}{2d} \int_0^{\mathcal{R}} \int_{\mathbb{R}^d} |z|^2 \frac{V_r(0, z)}{V_r} dz \mu(dr), \end{aligned} \tag{3.4}$$

which is precisely  $\sigma^2$  from (1.14).

Note also that a lineage is affected by selective events at rate

$$\left( u V_1 \int_0^{\mathcal{R}} r^d \mu(dr) \right) n^{2\beta} s_n = \eta \epsilon_n^{-2}, \tag{3.5}$$

where  $\eta = u V_1 \int_0^{\mathcal{R}} r^d \mu(dr)$ . 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  $\eta \epsilon_n^{-2}$ . Since  $\epsilon_n^{-2} = o(\log n)$ , 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  $\Theta(\epsilon_n^{-2})$  and so it is natural to expect that an equivalent of Theorem 2.4 holds.

### 3.2 Majority voting in the SLFVS, for $d \geq 2$

The rigorous proof of Theorem 1.8 closely follows that of Theorem 2.4. In Section 3.2.3, we focus on generation of the interface, which is proved in much the same way as Proposition 2.16. Then, in Section 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 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 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 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 3.7.** In Sections 2.2 and 2.4 we used subscripts to distinguish variables that played the same role in each section, but had different values; e.g.  $\delta_1$  in (2.20) and  $\delta_d$  in (2.35). The corresponding quantities in this section will be denoted with a subscript \*, for example  $\delta_*$  in (3.12).



**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  $\Lambda$ -Fleming-Viot processes (see for example Etheridge et al. [2015]).

Let  $(\xi^n(t), t \geq 0)$  be a pure jump process started at  $x \in \mathbb{R}^d$  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  $\mathbb{R}^d$  started at  $x$ .

**Lemma 3.8.** *For  $t > 0$  fixed, there is a coupling of  $W$  and  $\xi^n$  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  $X_1, X_2, \dots$  are i.i.d. with a rotationally symmetric distribution and, by (3.4),  $\mathbb{E}[|X_1|^2] = 2d\sigma^2 n^{-2\beta}$ . Moreover, by (3.3), the number of jumps made by  $\xi^n$  on the time interval  $[0, n^{-2\beta}]$  is Poisson, with mean  $\Theta(1)$ , so since each jump has magnitude at most  $2\mathcal{R}_n$ ,  $\mathbb{E}[|X_1|^4] = \mathcal{O}(n^{-4\beta})$ . Then by Skorohod’s second embedding Theorem, see e.g. Billingsley [1995], there is a Brownian motion  $W$  started at  $x$  and a sequence  $v_1, v_2, \dots$  of stopping times such that setting  $v_0 = 0$ ,  $(v_i - v_{i-1})_{i \geq 1}$  are i.i.d. and

$$\begin{aligned} W(v_i) &= \xi^n(i/n^{2\beta}), \\ \mathbb{E}[v_i - v_{i-1}] &= \frac{1}{2d} \mathbb{E}[|X_1|^2] = \sigma^2 n^{-2\beta}, \\ \mathbb{E}[(v_i - v_{i-1})^2] &= \mathcal{O}(n^{-4\beta}). \end{aligned}$$

It follows that  $\mathbb{E}[v_{\lfloor tn^{2\beta} \rfloor}] = \sigma^2 \lfloor tn^{2\beta} \rfloor n^{-2\beta}$  and  $\text{Var}[v_{\lfloor tn^{2\beta} \rfloor}] = \mathcal{O}(tn^{-2\beta})$ . Hence by Chebyshev’s inequality,

$$\mathbb{P} \left[ \left| v_{\lfloor tn^{2\beta} \rfloor} - \sigma^2 t \right| \geq n^{-\beta/2} \right] = \mathcal{O}(tn^{-\beta}). \tag{3.6}$$

Now we have that

$$\left| \xi^n(t) - W(\sigma^2 t) \right| \leq \left| \xi^n(t) - \xi^n(\lfloor tn^{2\beta} \rfloor / n^{2\beta}) \right| + \left| W(v_{\lfloor tn^{2\beta} \rfloor}) - W(\sigma^2 t) \right|. \tag{3.7}$$

To control the first term on the right hand side, observe that

$$\mathbb{P} \left[ \left| \xi^n(t) - \xi^n(\lfloor tn^{2\beta} \rfloor / n^{2\beta}) \right| \geq n^{-\beta/6} / 2 \right] \leq \mathbb{E}[|X_1|^2] (n^{-\beta/6} / 2)^{-2} = \mathcal{O}(n^{-5\beta/3}). \tag{3.8}$$

To control the second term on the right hand side of (3.7), let  $Z \sim N(0, 1)$ , then

$$\begin{aligned} &\mathbb{P} \left[ \left| W(v_{\lfloor tn^{2\beta} \rfloor}) - W(\sigma^2 t) \right| \geq n^{-\beta/6} / 2 \right] \\ &\leq \mathbb{P} \left[ \left| v_{\lfloor tn^{2\beta} \rfloor} - \sigma^2 t \right| \geq n^{-\beta/2} \right] \\ &\quad + \mathbb{P} \left[ \left| v_{\lfloor tn^{2\beta} \rfloor} - \sigma^2 t \right| \leq n^{-\beta/2}, \left| W(v_{\lfloor tn^{2\beta} \rfloor}) - W(\sigma^2 t) \right| \geq n^{-\beta/6} / 2 \right] \\ &\leq \mathbb{P} \left[ \sup_{s \in [-n^{-\beta/2}, n^{-\beta/2}]} |W(s) - W(0)| \geq n^{-\beta/6} / 2 \right] + \mathcal{O}(tn^{-\beta}) \\ &\leq 4d \mathbb{P} \left[ \sqrt{2} n^{-\beta/4} Z \geq n^{-\beta/6} / 2d \right] + \mathcal{O}(tn^{-\beta}) \\ &= \mathcal{O}(\exp(-\frac{1}{8d^2} n^{\beta/6})) + \mathcal{O}(tn^{-\beta}). \end{aligned} \tag{3.9}$$

Here, the second inequality follows by (3.6) and the third inequality follows by bounding the modulus of a  $d$ -dimensional Brownian motion by the sum of the moduli of  $d$  one-dimensional Brownian motions and then using the reflection principle. Combining (3.8) and (3.9) with (3.7) 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 3.9.** *Let  $\tau$  be the first branch time of  $\Xi^n$ . There is a coupling of  $\Xi^n$  and  $W$  under which  $\tau$  and  $W$  are independent,  $\tau \sim \text{Exp}(\eta\epsilon_n^{-2})$  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  $\tau$  follows immediately from (3.5).

Now consider any ancestral lineage  $\xi^n(\cdot) \subseteq \Xi^n$ . By the thinning property of Poisson processes, at any time  $t > 0$ , we can write  $\xi^n(t) = \xi^{n,\text{sel}}(t) + \xi^{n,\text{neu}}(t)$ , where  $\xi^{n,\text{sel}}(\cdot)$  and  $\xi^{n,\text{neu}}(\cdot)$  are independent pure jump processes with jump intensities  $s_n m_n(dz)$  and  $(1 - s_n)m_n(dz)$  respectively, and taking  $\tau$  to be the first jump time of  $\xi^{n,\text{sel}}$ ,  $\xi^{n,\text{neu}}$  is independent of  $\tau$ . Using Lemma 3.8 with  $(1 - s_n)m_n(dz)$  in place of  $m_n(dz)$ , we can couple  $\xi^{n,\text{neu}}$  with a Brownian motion  $W$  in such a way that for any  $t > 0$ ,

$$\mathbb{P}[|\xi^{n,\text{neu}}(t) - W(\sigma^2(1 - s_n)t)| \geq n^{-\beta/6}] \leq \mathcal{O}(n^{-\beta}(t \vee 1)).$$

Since  $s_n = o(\log n/n^{2\beta})$ , using Chebyshev's inequality,

$$\mathbb{P}[|W(\sigma^2 t) - W(\sigma^2(1 - s_n)t)| \geq n^{-\beta/6}] = o\left(\frac{\log n}{n^{2\beta}} n^{\beta/3}(t \vee 1)\right),$$

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  $\mathbb{E}[\tau] = \Theta(\epsilon_n^2) = o(1)$ , and for  $i = 1, 2, 3$ ,  $|\xi_i^n(\tau) - \xi_i^n(\tau-)| \leq 2\mathcal{R}_n = 2n^{-\beta}\mathcal{R}$  the result follows. ■

### 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 3.10** (Branching jump process). *For given  $n \in \mathbb{N}$  and starting point  $x \in \mathbb{R}^d$ ,  $(\Psi^n(t), t \geq 0)$  is the historical process of the branching random walk which is described as follows.*

1. Each individual has an independent exponential lifetime with parameter  $\eta\epsilon_n^{-2}$ .
2. During its lifetime, each individual, independently, evolves according to a pure jump process with jump intensity  $(1 - s_n)m_n(dz)$ .
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

$$\frac{r^d \mu^n(dr)}{\int_0^{\mathcal{R}_n} s^d \mu^n(ds)}.$$

*If the parent is at the point  $z \in \mathbb{R}^d$ , then each of the three offspring, independently, samples its location uniformly from  $B_r(z)$ .*

**Remark 3.11.** Note that the only difference between the distributions of  $\Xi^n$  and  $\Psi^n$  is that in  $\Psi^n$ , lineages evolve independently after branching, whereas in  $\Xi^n$ , two distinct lineages may be hit by the same event in  $\Pi_n$ .

We define  $\mathbb{V}_p(\Psi^n(t))$  in the usual way (as in Definition 2.1): a leaf at location  $\psi_i(t) \in \mathbb{R}^d$  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 3.12.** *Let  $T^* \in (0, \infty)$ ,  $k \in \mathbb{N}$  and  $z \in \mathbb{R}^d$ . There exists  $n_* \in \mathbb{N}$  such that for all  $n \geq n_*$ , there is a coupling of  $\Xi^n$  started from  $z$  and  $\Psi^n$  started from  $z$  such that with probability at least  $1 - \epsilon_n^k$  we have*

$$\Xi^n(T^*) = \Psi^n(T^*).$$

The remainder of this section is devoted to the proof of Lemma 3.12. To do so, we consider a slightly different description of the dual of the SAFVS, which will preserve the distribution of  $\Xi^n$ .

**Definition 3.13** (Pre-emptive SLFVS dual). *For  $n \in \mathbb{N}$ , the process  $(\tilde{\mathcal{P}}_t^n)_{t \geq 0}$  is a  $\bigcup_{l \geq 1} (\mathbb{R}^d)^l$ -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  $x$  and we write*

$$(\xi_1^n(t), \dots, \xi_{N(t)}^n(t))$$

*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  $1 - s_n$ . 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.*

*With the complementary probability  $s_n$ , 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  $\Pi_n$ , 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 or 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  $\Xi^n$ , ignoring marks, we write  $\Phi^n$  for the historical process corresponding to the pre-emptive dual. The distribution of  $\Phi^n$  is equal to that of  $\Xi^n$ . The only difference between Definition 3.1 and Definition 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  $\Xi^n$  (or equivalently  $\Phi^n$ ) and  $\Psi^n$  is that for as long as two ancestral lineages are not both marked, they evolve independently.

**Lemma 3.14.** *Let  $T^* \in (0, \infty)$ . There exists  $\alpha > 0$  such that*

$$\mathbb{P}\left[\exists \xi_i^n \neq \xi_j^n \subseteq \Phi^n(T^*), t \in [0, T^*] \text{ such that } \xi_i^n \text{ and } \xi_j^n \text{ are both marked at time } t\right] = \mathcal{O}(n^{-\alpha}).$$

*Proof.* Write  $\mathcal{T}(\Phi^n(t))$  for the genealogy of  $\Phi^n(t)$ . We begin by showing that for any constant  $b > 0$ ,  $\mathcal{T}(\Phi^n(T^*)) \subseteq \mathcal{T}_{b \log n}^{\text{reg}}$  with high probability. Recall from (3.5) that the rate at which each lineage is affected by reproduction events is  $\eta \epsilon_n^{-2} = o(\log n)$ . Let  $M^n$  be a Poisson distributed random variable with mean  $T^* \eta \epsilon_n^{-2}$ . Recall that if  $Z'$  is Poisson with parameter  $\chi$ , then (using a Chernoff bound) for  $k > \chi$ ,

$$\mathbb{P}[Z' > k] \leq \frac{e^{-\chi} (e\chi)^k}{k^k}. \tag{3.10}$$

Hence for  $b > 0$  a constant, applying (3.10) with  $k = b \log n$  and  $\chi = T^* \eta \epsilon_n^{-2} = o(\log n)$ , taking  $n$  sufficiently large that  $\frac{e\chi}{b \log n} \leq 3^{-2}$ , we have

$$\mathbb{P}[M^n > b \log n] \leq 3^{-2b \log n}.$$

Then by a union bound over each root to leaf ray of  $\mathcal{T}_{b \log n}^{\text{reg}}$ ,

$$\mathbb{P}\left[\mathcal{T}(\Phi^n(T^*)) \not\subseteq \mathcal{T}_{b \log n}^{\text{reg}}\right] \leq 3^{b \log n} \mathbb{P}[M^n > b \log n] \leq 3^{-b \log n}. \tag{3.11}$$

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  $\xi_i^n$  and  $\xi_j^n$  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  $\xi_i^n$  is Poisson with mean  $\Theta(n)$ . Hence, the probability that a given pair  $\xi_i^n, \xi_j^n$  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

$$\begin{aligned} \mathbb{P}\left[\exists \xi_i^n \neq \xi_j^n \subseteq \Phi^n(T^*) \text{ and } t \in [0, T^*] \text{ such that } \xi_i^n \text{ and } \xi_j^n \text{ are both marked at time } t\right] \\ \leq 3^{-b \log n} + 3^{2b \log n} \mathcal{O}(n^{4\beta-1}) \\ \leq 3^{-b \log n} + \mathcal{O}\left(\exp\left(2b(\log 3)(\log n) + (4\beta - 1) \log n\right)\right). \end{aligned}$$

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 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  $n^{-\alpha} = o((\log n)^{-k/2}) = o(\epsilon_n^k)$ , by Lemma 3.14, for  $n$  sufficiently large  $\mathbb{P}[\tau \geq T^*] \geq 1 - \epsilon_n^k$ . For as long as ancestral lineages in  $\Phi^n$  are not both marked they evolve independently, so we may couple  $(\Phi^n(t))$  and  $(\Psi^n(t))$  to be equal up until time  $\tau$  and the result follows. ■

**3.2.3 Generation of the interface**

In this section we show that, in analogy to Proposition 2.16, the interface is generated in time of order  $\epsilon_n^2 |\log \epsilon_n|$ . The proof is similar to that of Proposition 2.16.

**Proposition 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*

$$\delta_*(k, n) := a_*(k)\epsilon_n^2 |\log \epsilon_n| \text{ and } \delta'_*(k, n) := (a_*(k) + \eta^{-1}(k + 1))\epsilon_n^2 |\log \epsilon_n|, \tag{3.12}$$

then for  $t \in [\delta_*, \delta'_*]$ ,

1. for  $x$  such that  $d(x, \sigma^2 t) \geq d_* \epsilon |\log \epsilon|$ , we have  $\mathbb{P}_x [\mathbb{V}_p(\Xi^n(t)) = 1] \geq 1 - \epsilon_n^k$ ;
2. for  $x$  such that  $d(x, \sigma^2 t) \leq -d_* \epsilon |\log \epsilon|$ , we have  $\mathbb{P}_x [\mathbb{V}_p(\Xi^n(t)) = 1] \leq \epsilon_n^k$ .

Using the coupling from Lemma 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 3.16.** *Let  $k \in \mathbb{N}$  and let  $A(k)$  be chosen as in Lemma 2.9. There exist  $a_*(k), B_*(k) \in (0, \infty)$ , and  $n_*(k) < \infty$  such that for all  $n \geq n_*$  and  $\delta_*, \delta'_*$  as defined in (3.12),*

$$\mathbb{P} \left[ \mathcal{T}(\Psi^n(\delta_*)) \supseteq \mathcal{T}_{A(k)|\log \epsilon_n}^{\text{reg}} \right] \geq 1 - \epsilon_n^k, \tag{3.13}$$

and 
$$\mathbb{P} \left[ \mathcal{T}(\Psi^n(\delta'_*)) \subseteq \mathcal{T}_{B_*(k)|\log \epsilon_n}^{\text{reg}} \right] \geq 1 - \epsilon_n^k. \tag{3.14}$$

**Remark 3.17.** During the proof of Proposition 2.16, we deduced (2.36), which is the equivalent of (3.13). We did not require an equivalent of (3.14). We shall use (3.14) here in order to prove the equivalent of (2.37).

*Proof.* Recall from (3.5) that a given ancestral lineage in  $\Psi^n$  branches into three after an exponential time with rate  $\eta \epsilon_n^{-2}$ . Hence, (3.13) follows for  $a_*$  sufficiently large by the same proof as Lemma 2.10.

The proof of (3.14) is the same as that of (3.11). Let  $L^n$  be a Poisson distributed random variable with mean  $\delta'_* \eta \epsilon_n^{-2} = (a_* + \eta^{-1}(k + 1))\eta |\log \epsilon_n|$ . Take  $B_* = B_*(k)$  sufficiently large that  $B_* \geq (a_* + \eta^{-1}(k + 1))\eta$  and

$$e(a_* + \eta^{-1}(k + 1))\eta B_*^{-1} < \frac{1}{3} e^{-k/B_*}. \tag{3.15}$$

The Chernoff bound (3.10) gives

$$\begin{aligned} \mathbb{P} [L^n > B_* |\log \epsilon_n|] &\leq (e(a_* + \eta^{-1}(k + 1))\eta B_*^{-1})^{B_* |\log \epsilon_n|} \\ &\leq \epsilon^k 3^{-B_* |\log \epsilon_n|}, \end{aligned} \tag{3.16}$$

and, taking a union bound over each root to leaf ray of  $\mathcal{T}_{B_*|\log \epsilon_n}^{\text{reg}}$ ,

$$\mathbb{P} \left[ \mathcal{T}(\Psi^n(\delta'_*)) \not\subseteq \mathcal{T}_{B_*(k)|\log \epsilon_n}^{\text{reg}} \right] \leq 3^{B_* |\log \epsilon_n|} \mathbb{P} [L^n > B_* |\log \epsilon_n|] \leq \epsilon_n^k,$$

which completes the proof. ■

*Proof of Proposition 3.15.* We prove this result with  $\Psi^n$  in place of  $\Xi^n$  (from which the result follows using Lemma 3.12). The approach closely follows that of Proposition 2.16 except that now we have to control the distance between the jump process followed by a lineage and Brownian motion.

Take  $a_*$  from Lemma 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 3.8

we can couple  $(\xi^n(t), t \geq 0)$  with a  $d$ -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  $\epsilon_n^{-2} = o(\log n)$  we have  $\frac{1}{2}d_*\epsilon_n |\log \epsilon_n| \geq 2n^{-\beta/6}$ . Hence, for such  $n$ ,

$$\begin{aligned} & \mathbb{P} \left[ |\xi^n(t) - \xi^n(0)| \geq \frac{1}{2}d_*\epsilon_n |\log \epsilon_n| \right] \\ & \leq \mathbb{P} \left[ |\xi^n(t) - W(\sigma^2 t)| \geq n^{-\beta/6} \right] \\ & \quad + \mathbb{P} \left[ |W(\sigma^2 \delta'_*(k, n)) - W(0)| \geq \frac{1}{4}d_*\epsilon_n |\log \epsilon_n| \right] \\ & \leq \mathcal{O}(n^{-\beta}) + 2d \exp \left( -\frac{1}{64} \frac{d_*^2}{d^2 \sigma^2 (d_* + \eta^{-1}(k+1))} |\log \epsilon_n| \right) \\ & \leq 3^{-B_* |\log \epsilon_n|} \epsilon_n^k. \end{aligned}$$

Here the second inequality follows by bounding the modulus of a  $d$ -dimensional Brownian motion by the sum of the moduli of  $d$  one-dimensional Brownian motions, and the last inequality follows for  $d_*$  sufficiently large. Using (3.14) and taking a union bound over the root to leaf rays of  $\mathcal{T}_{B_* |\log \epsilon_n|}^{\text{reg}}$ , for  $t \in [\delta_*, \delta'_*]$ ,

$$\begin{aligned} \mathbb{P}_x \left[ \exists \xi_i^n \subseteq \Psi^n(\delta'_*) \text{ s.t. } |\xi_i^n(t) - x| \geq \frac{1}{2}d_*\epsilon_n |\log \epsilon_n| \right] & \leq \epsilon_n^k + 3^{B_* |\log \epsilon_n|} 3^{-B_* |\log \epsilon_n|} \epsilon_n^k \\ & \leq 2\epsilon_n^k. \end{aligned} \tag{3.17}$$

Combining (3.17) with Lemma 3.16, we obtain that, with probability  $\geq 1 - 3\epsilon_n^k$ ,

1.  $\mathbb{V}_p(\Psi^n(t))$  is given by independent votes at each of the leaves of  $\mathcal{T}(\Psi^n(t))$ .
2.  $\mathcal{T}(\Psi^n(t)) \supseteq \mathcal{T}_{A |\log \epsilon_n|}^{\text{reg}}$  and the positions of the individuals corresponding to the leaves of  $\mathcal{T}(\Psi^n(t))$  are all within  $\frac{1}{2}d_*\epsilon_n |\log \epsilon_n|$  of their starting position.

Just as in the proof of Proposition 2.16 we obtain Proposition 3.15 with  $\Psi^n$  in place of  $\Xi^n$ . An application of Lemma 3.12 completes the proof. ■

### 3.2.4 Propagation of the interface

We require the following slight modification of Lemma 2.18.

**Lemma 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  $n \geq n_*$ ,  $x \in \mathbb{R}^d$ ,  $s \in [\sigma^2 \epsilon_n^{l+3}, \sigma^2(l+1)\eta^{-1}\epsilon_n^2 |\log \epsilon_n|]$  and  $t \in [s, \sigma^2 T^*]$ ,*

$$\begin{aligned} E_x \left[ g \left( \mathbb{P}_{d(W_s, t-s) + K_1 e^{K_2(t-s)} \epsilon_n |\log \epsilon_n| + 3n^{-\beta/6}}^{\epsilon_n} [\mathbb{V}(\mathbf{B}(t-s)) = 1] + \epsilon_n^l \right) \right] \\ \leq \frac{3}{4}\epsilon_n^l + E_{d(x,t)} \left[ g \left( \mathbb{P}_{B_s + K_1 e^{K_2 t} \epsilon_n |\log \epsilon_n|}^{\epsilon_n} [\mathbb{V}(\mathbf{B}(t-s)) = 1] \right) \right] + \mathbb{1}_{s \leq \epsilon_n^3} \epsilon_n^l, \end{aligned} \tag{3.18}$$

and

$$\begin{aligned} E_x \left[ g \left( \mathbb{P}_{d(W_s, t-s) - K_1 e^{K_2(t-s)} \epsilon_n |\log \epsilon_n| - 3n^{-\beta/6}}^{\epsilon_n} [\mathbb{V}(\mathbf{B}(t-s)) = 0] + \epsilon_n^l \right) \right] \\ \leq \frac{3}{4}\epsilon_n^l + E_{d(x,t)} \left[ g \left( \mathbb{P}_{B_s - K_1 e^{K_2 t} \epsilon_n |\log \epsilon_n|}^{\epsilon_n} [\mathbb{V}(\mathbf{B}(t-s)) = 0] \right) \right] + \mathbb{1}_{s \leq \epsilon_n^3} \epsilon_n^l. \end{aligned} \tag{3.19}$$

*Proof.* The proof is essentially the same as that of Lemma 2.18. Let  $R = 2c_1(l) + 4\sigma\eta^{-1/2}(l+1)d + 1$  and fix  $K_2$  such that  $K_1(K_2 - C_0) - C_0R = 2c_1(1)$ ; let

$$A_x = \left\{ \sup_{u \in [0, s]} |W_u - x| \leq 2\sigma\eta^{-1/2}(l+1) d\epsilon |\log \epsilon| \right\}.$$

The proof for  $|d(x, t)| \geq (2c_1(l) + 2\sigma\eta^{-1/2}(l+1)d + K_1e^{K_2(t-s)})\epsilon_n |\log \epsilon_n|$  is then the same as in the proof of Lemma 2.18 (since  $n^{-\beta/6} = o(\epsilon_n |\log \epsilon_n|)$ ).

Since  $n^{-\beta/6} = o(s\epsilon_n |\log \epsilon_n|)$ , we have for  $\beta = (R + K_1e^{K_2(t-s)})\epsilon |\log \epsilon|$  as in (2.59), for  $n$  sufficiently large

$$K_1e^{K_2t}\epsilon_n |\log \epsilon_n| - (C_0\beta s + K_1e^{K_2(t-s)})\epsilon_n |\log \epsilon_n| + 3n^{-\beta/6} \geq c_1(1)s\epsilon_n |\log \epsilon_n|. \quad (3.20)$$

Using (3.20) in place of (2.62), the proof for  $|d(x, t)| \leq (2c_1(l) + 2\sigma\eta^{-1/2}(l+1)d + K_1e^{K_2(t-s)})\epsilon_n |\log \epsilon_n|$  is the same as in the proof of Lemma 2.18. ■

The equivalent of Proposition 2.17 for  $\Psi^n$  is as follows.

**Proposition 3.19.** *Let  $l \in \mathbb{N}$  with  $l \geq 4$ . Define  $a_*(l)$  and  $\delta_*(l, n)$  as in Proposition 3.15. There exist  $K_1(l), K_2(l) > 0$  and  $n_*(l, K_1, K_2) > 0$  such that for all  $n \geq n_*$  and  $t \in [\delta_*(l, n), T^*]$  we have*

$$\sup_{x \in \mathbb{R}^d} \left( \mathbb{P}_x [\mathbb{V}_p(\Psi^n(t)) = 1] - \mathbb{P}_{d(x, \sigma^2 t) + K_1e^{K_2\sigma^2 t}\epsilon_n |\log \epsilon_n|}^{\epsilon_n} [\mathbb{V}(\mathbf{B}(\sigma^2 t)) = 1] \right) \leq \epsilon_n^l \quad (3.21)$$

and

$$\sup_{x \in \mathbb{R}^d} \left( \mathbb{P}_x [\mathbb{V}_p(\Psi^n(t)) = 0] - \mathbb{P}_{d(x, \sigma^2 t) - K_1e^{K_2\sigma^2 t}\epsilon_n |\log \epsilon_n|}^{\epsilon_n} [\mathbb{V}(\mathbf{B}(\sigma^2 t)) = 0] \right) \leq \epsilon_n^l. \quad (3.22)$$

*Proof.* The proof exactly follows that of Proposition 2.17, with Corollary 3.9 and then Lemma 3.18 in place of Lemma 2.18, and Proposition 3.15 in place of Proposition 2.16. ■

*Proof of Theorem 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 3.12, for  $n$  sufficiently large and  $t \in [0, T^*]$ ,

$$|\mathbb{P}_x [\mathbb{V}_p(\Psi^n(t)) = 1] - \mathbb{P}_x [\mathbb{V}_p(\Xi^n(t)) = 1]| \leq \epsilon_n^{k+1}.$$

The result now follows from Proposition 3.19 with  $l = k + 1$ , in the same way as in the proof of Theorem 2.4. ■

## References

- A Allen and J Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metall.*, 27:1084–1095, 1979.
- 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.*, 15:162–216, 2010. MR-2594876
- 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, 2013. MR-3036210
- P Billingsley. *Probability and Measure*. Wiley, 1995. MR-1324786
- L Bronsard and R Kohn. Motion by mean curvature as the singular limit of Ginzburg-Landau dynamics. *J. Diff. Eq.*, 90:211–217, 1991. MR-1101239
- X Chen. Generation and propagation of interfaces for reaction-diffusion equations. *J. Diff. Eq.*, 96: 116–141, 1992. MR-1153311

- A De Masi, P A Ferrari, and J L Lebowitz. Reaction-diffusion equations for interacting particle systems. *J. Stat. Phys.*, 44(3/4):589–644, 1986. MR-0857069
- A M Etheridge. Drift, draft and structure: some mathematical models of evolution. *Banach Center Publ.*, 80:121–144, 2008. MR-2433141
- A M Etheridge, A Véber, and F Yu. Rescaling limits of the spatial Lambda-Fleming-Viot process with selection. *arXiv preprint arXiv:1406.5884*, 2014. MR-3024966
- A M Etheridge, N Freeman, S Penington, and D Straulino. Branching Brownian motion and selection in the Spatial Lambda-Fleming-Viot process. *arXiv preprint arXiv:1512.03766*, 2015. MR-3719942
- L C Evans and J Spruck. Motion of level sets by mean curvature I. *J. Diff. Geom.*, 33:635–681, 1991. MR-1100206
- 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. MR-1177477
- R A Fisher. The wave of advance of advantageous genes. *Ann. Eugenics*, 7:355–369, 1937.
- T Funaki and H Spohn. Motion by mean curvature from the Ginzburg-Landau  $|\nabla\phi|$  interface model. *Comm. Math. Phys.*, 185:1–36, 1997. MR-1463032
- M Gage and R Hamilton. The heat equation shrinking convex plane curves. *J. Diff. Geom.*, 23:417–491, 1986. MR-0840401
- M A Grayson. The heat equation shrinks embedded plane curves to round points. *J. Diff. Geom.*, 26:285–314, 1987. MR-0906392
- M Hairer, M D Ryser, and H Weber. On the triviality of the 2D stochastic Allen-Cahn equation. *Elect. J. Probab.*, 17(39):1–14, 2012. MR-2928722
- G Huisken. Flow by mean curvature of convex surfaces into spheres. *J. Diff. Geom.*, 20:237–266, 1984. MR-0772132
- W G Hunt and R K Selander. Biochemical genetics of hybridization in European house mice. *Heredity*, 31:11–33, 1973.
- T Ilmanen. Convergence of the Allen-Cahn equation to Brakke’s motion by mean curvature. *J. Diff. Geom.*, 38:417–461, 1993.
- S M Krone and C Neuhauser. Ancestral processes with selection. *Theor. Pop. Biol.*, 51:210–237, 1997.
- C Mantegazza. *Lecture notes on mean curvature flow*, volume 290 of *Progress in Mathematics*. Birkhäuser, 2011.
- H P McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovski-Piskunov. *Comm. Pure Appl. Math.*, 28:323–331, 1975.
- C Neuhauser and S M Krone. Genealogies of samples in models with selection. *Genetics*, 145:519–534, 1997.
- 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.
- A V Skorohod. Branching diffusion processes. *Th. Prob. Appl.*, 9:492–497, 1964.
- 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.
- B White. Evolution of curves and surfaces by mean curvature. *Proceedings of the International Congress of Mathematicians, Beijing 2002*, I:525–538, 2002.