

# NATURAL SELECTION IN SPATIALLY STRUCTURED POPULATIONS

ALISON ETHERIDGE

## ABSTRACT

Mathematical models play a fundamental role in theoretical population genetics and, in turn, population genetics provides a wealth of mathematical challenges. Here we illustrate this by using mathematical caricatures of the evolution of genetic types in a spatially distributed population to demonstrate the complex interplay between spatial structure, natural selection, and so-called random genetic drift (the randomness due to reproduction in a finite population). In particular, we highlight the role that the shape of the domain inhabited by the population can play in mediating the interplay between the different forces of evolution acting upon it.

## MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 92D15; Secondary 60J70, 60H30, 35K57, 35B30

## KEYWORDS

Natural selection, spatial structure, genetic drift, blocking

## 1. INTRODUCTION

Theoretical population genetics is concerned with understanding genetic differences within and between populations. It finds its origins at the beginning of the twentieth century in *the modern evolutionary synthesis*, in which Darwin's theory of evolution through natural selection and Mendel's laws of genetic inheritance were integrated. This work, pioneered by Fisher, Haldane, and Wright, provided a unified mathematical framework within which to discuss possible causes of evolution. As a result some consensus emerged about which forces influence evolution, but questions such as their relative importance remained unresolved.

The intervening century has seen a rich interplay between population genetics and the mathematical sciences: mathematical modeling has been employed to explore concepts such as adaptation, speciation, and population structure, and in the process questions arising from population genetics have stimulated the development of elegant new mathematical models and techniques, often of much wider applicability.

In population genetics, mathematical models are used both as a basis for statistical inference and as a means to validate or dismiss concepts. The purpose of the model is then not to provide detailed predictions of the fate of a particular biological population, but rather to use caricatures of the forces of evolution, and the ways in which they interact, to gain some insight into the evolutionary process. Our aim here is to illustrate this approach through models that attempt to capture some features of the interactions between natural selection, spatial structure, and the randomness due to reproduction in a finite population (so-called *genetic drift*). Rather than giving detailed proofs, for which we refer to the original papers, we shall provide informal arguments and draw out some of the lessons learned.

We shall try to minimize the use of biological jargon, but it is convenient to fix some terminology. The term *locus* is used to refer in a general way to a location on the genome. For our purposes, it will correspond to a region that codes for a gene, and it will be passed on as a single unit from parent to offspring. Genes can occur in different forms, called *alleles*, and we shall make the simplifying assumption that the gene in which we are interested has just two alleles, denoted  $a$ ,  $A$ . Evolution is fueled by mutation, the source of the genetic diversity on which natural selection acts, but we shall assume that any new mutations arising at the locus of interest are neutral, that is, do not affect fitness. Moreover, since genes are organized onto chromosomes, different genetic loci do not evolve independently of one another. However, our models will neglect this *genetic structure* and suppose that (relative) fitness is determined entirely by the alleles at the locus of interest. Although crude, such single locus models exhibit a surprisingly rich variety of behaviors.

There is a huge literature devoted to understanding the interaction between natural selection and genetic drift. In particular, in the absence of spatial structure, it is well understood that in larger populations, not only is a beneficial mutation more likely to establish and sweep to fixation (that is, increase in frequency until it is carried by every individual in the population), but it is also more likely that deleterious mutations will be expunged. Genetic drift, which drives random fluctuations in the proportions of the different alleles, is stronger

in a smaller population, and this increases the chance that a deleterious mutation is fixed just by chance [27].

The interaction between natural selection and spatial structure (ignoring genetic drift) is often investigated through reaction–diffusion equations. This was initiated by Fisher [22], who studied traveling wave solutions to the equation

$$\frac{\partial u}{\partial t} = m \frac{\partial^2 u}{\partial x^2} + su(1 - u), \quad x \in \mathbb{R}, \quad t \geq 0, \quad (1.1)$$

as a model of the spread of a favorable allele through a one-dimensional population. Here  $u(t, x) \in [0, 1]$  models the proportion of the alleles carried by the individuals at location  $x$  at time  $t$  that are of the fitter type. In [28], Kolmogorov, Petrovsky, and Piscounov considered the analogous equation in two spatial dimensions and with a general reaction term  $F(u)$  in place of  $su(1 - u)$ , although they focused on solutions that are independent of  $y$  (and thus essentially one-dimensional). Motivated by the discussion of Fisher [21], they specialized to a reaction term of the form  $\alpha u(1 - u)^2$  for their application to population genetics. Equation (1.1), with  $\partial^2 u / \partial x^2$  replaced by  $\Delta u$  in dimensions  $d \geq 2$ , is often referred to as the Fisher–KPP equation.

A special case of a result of Skorokhod [36] expresses the distribution of a branching Brownian motion in terms of the solution to (1.1). Conversely, in the particular case of a Heaviside initial condition, this allows one to express the solution to (1.1) in terms of the distribution of the rightmost particle in a binary branching Brownian motion at time  $t$ . This is often referred to as McKean’s representation [31], and underpins the remarkable work of Bramson [9], which provides much of our understanding of the traveling wave to which the solution started from a Heaviside initial condition converges. More recently, these results have been considerably extended, with a particular focus on adding a stochastic term to (1.1) to capture the effect of random genetic drift (see, for example, [33] and the references therein), resulting in a stochastic PDE:

$$du = \left( m \frac{\partial^2 u}{\partial x^2} + su(1 - u) \right) dt + \sqrt{\frac{1}{\rho} u(1 - u)} W(dt, dx), \quad (1.2)$$

with  $W$  a space-time white noise and  $\rho$  a measure of local population size. (The form of this so-called *Wright–Fisher noise* term will be motivated in Section 2.1.) The vast majority of this work is restricted to one spatial dimension. In the biologically natural setting of two spatial dimensions, although equation (1.1) generalizes in a natural way, the obvious generalization of equation (1.2) has no solution. In Section 6 we shall describe one way to circumvent this, and provide a mathematical model through which we can explore the interaction of natural selection and genetic drift in a population distributed across a spatial continuum (of any dimension). Depending on the dispersal mechanism and the local population density, an individual may be competing with its own close (and equally fit) relatives, limiting the effect of natural selection. We shall see that if the local population density is bounded, *the dimension of the space in which the population lives is important*.

Natural selection can take many forms. While equation (1.1) models the spread of an allele which is always favorable to the individual carrying it, in much of what follows

we shall be interested in populations in which individuals carry two copies of the gene and those carrying *different* alleles are selectively disadvantaged. As we explain in Section 2.1, this form of selection can be captured by replacing the reaction term in (1.1) to obtain

$$\frac{\partial u}{\partial t} = m \frac{\partial^2 u}{\partial x^2} + su(1-u)(2u-1), \quad x \in \mathbb{R}, t \geq 0. \quad (1.3)$$

Inference from genetic data typically involves using differences between the DNA sequences of a sample of individuals from the population to reconstruct information about *genealogical* ancestors of those individuals. This can then be compared to the predictions of mathematical models under different hypotheses about the forces of evolution acting on the population. The neutral mutation rate therefore dictates the scales over which we can glean meaningful information, and, since it is very small, this leads us to consider very large spatial and temporal scales. With this in mind, we apply a diffusive scaling, corresponding to modeling proportions of different alleles over spatial regions of diameter  $\mathcal{O}(1/\varepsilon)$  at times of  $\mathcal{O}(1/\varepsilon^2)$ , to obtain

$$\frac{\partial u^\varepsilon}{\partial t} = \Delta u^\varepsilon + \frac{1}{\varepsilon^2} u^\varepsilon (1 - u^\varepsilon) (2u^\varepsilon - 1), \quad (1.4)$$

where we have set  $m = 1, s = 1$ . In a sense made precise in Theorem 2.4, for suitable initial conditions, as  $\varepsilon \rightarrow 0$ ,  $u^\varepsilon$  converges to the indicator function of a set whose boundary evolves according to mean curvature flow (see Definition 2.1). We emphasize that although our main interest is in two spatial dimensions (where mean curvature flow is simply curvature flow), our mathematical results are valid in arbitrary spatial dimension  $\mathfrak{d}$ .

More generally (see Section 5.1), if there is a fitness difference between individuals carrying two  $a$  alleles and those carrying two  $A$  alleles, we consider the equation

$$\frac{\partial u}{\partial t} = m \Delta u + su(1-u)(2u - (1 - \gamma)), \quad x \in \mathbb{R}^{\mathfrak{d}}, t \geq 0. \quad (1.5)$$

As we shall explain, in order to obtain a nontrivial limit under the diffusive scaling, we also scale  $\gamma = \varepsilon v$ . The limit is then the indicator function of a set whose boundary evolves according to a mixture of “constant flow” of rate  $-v$  and mean curvature flow (for as long as this flow is defined).

Whereas mean curvature flow has no nontrivial fixed point, the spherical shell of radius  $(\mathfrak{d} - 1)/v$  (whose interior is completely occupied by the favored type) is fixed by this mixture of curvature and constant flow. In this scenario, the two components of the selection acting on the population work against one another and at this critical radius are finely balanced; for any larger radius constant flow dominates and the circle expands without bound; for a smaller radius, mean curvature flow wins out, and the circle shrinks to a point. This behavior is in sharp contrast to the situation in one spatial dimension, and it is natural then to ask about other domains; for example, what is the fate of an expanding population that must pass through an isthmus? In Section 5.2, we shall see examples of domains for which the effect of curvature flow leads to “blocking” of the expansion of the range of the selectively favored type (but in a way which will result in a stable nontrivial steady state). *The geometry of the domain in which the population lives is important.*

The main mathematical tool that we use is a representation of the solution to (1.4) in terms of a ternary branching Brownian motion which we explain in Section 3. Although reminiscent of the Skorokhod/McKean representations of the solution to the Fisher–KPP equation, it differs in using the entire tree structure of the branching process. Our approach can be seen as an adaptation of that of de Masi et al. in [13], and similar ideas have also been exploited in [29]. For us, it provides an intuitive and flexible representation of the solutions to equations like (1.3), (1.4), and (1.5), that is readily adapted to the framework of Section 6, allowing us to incorporate the effects of genetic drift.

The rest of this article is laid out as follows. In Section 2, we motivate (1.3) from a biological perspective and give a more precise statement about its limiting behavior as  $\varepsilon \rightarrow 0$ . In Section 3, we present the probabilistic representation of the solution to (1.4) and use it to provide some intuition for the emergence of mean curvature flow. In Section 4, we replace the Laplacian in (1.4) by a *fractional* Laplacian, in order to capture the corresponding behaviour in populations with long-range dispersal. In Section 5, we turn to the situation modeled by (1.5) in which there is a fitness difference between type *aa* and type *AA* individuals. In particular, we shall consider what happens when the population no longer occupies the whole Euclidean space, and we provide conditions on the geometry of its range under which the expansion of the region occupied by the fitter type is, or is not, blocked. Finally, in Section 6, we extend our models to incorporate genetic drift and explore the extent to which it breaks down the effect of natural selection. In particular, we shall see how the impact of genetic drift depends on both the local population density and the spatial dimension.

## 2. HYBRID ZONES AND CURVATURE FLOW

A hybrid zone is a narrow geographic region where two genetically distinct populations are found close together and hybridize to produce offspring of mixed ancestry. Hybrid zones are ubiquitous in nature; see, for example, [5] and [6] for an extensive catalogue and discussion. They can be maintained by a variety of mechanisms. For example, consider two populations, each of which is adapted to a different set of environmental conditions. If hybrids are less well adapted to those conditions, then an abrupt change in the environment could result in a hybrid zone. In that case the hybrid zone will not move.

The situation that we shall be trying to caricature is one in which the hybrid zone is maintained by a balance between dispersal and selection against hybrids. For instance, this might arise if two populations regain contact after a period of geographic isolation such as that imposed by the last glacial maximum (c. 18,000 years ago) when many species were forced into isolated refugia. Because they are not dependent on changes in local environmental conditions, hybrid zones maintained by this mechanism can move from place to place. In [3], Barton presented a theoretical study of the dynamics of hybrid zones. In the interests of space, we do not attempt to examine all of the influences on the motion of the zone considered by [3]; instead we present our mathematical approach and illustrate its application in three contrasting settings before adapting it to include genetic drift in Section 6.

## 2.1. Modeling selection against heterozygosity

As advertised in the introduction, we are going to focus on the case in which the hybrid zone is maintained by selection acting on a single genetic locus. We suppose that the gene at that locus has two alleles, denoted  $a$  and  $A$ , and that each individual carries two copies of the gene. One population consists of  $aa$  individuals, the other of  $AA$  individuals. Although it is possible to obtain equations like (1.1)–(1.5) as scaling limits of a variety of individual based models (see, for example, [12,23,34]), it is generally highly technical and so instead we shall motivate the models using an argument commonly found in the biological literature.

Our first aim is to understand the form of the reaction term in (1.3), and so we begin with the case in which the population is infinitely large, and has no spatial structure. We assume Hardy–Weinberg equilibrium; that is, if the proportion of  $a$ -alleles across the whole population is  $\bar{u}$ , then the proportions of individuals of types  $aa$ ,  $aA$ , and  $AA$  are given by

$$\begin{array}{c|c|c} aa & aA & AA \\ \hline \bar{u}^2 & 2\bar{u}(1-\bar{u}) & (1-\bar{u})^2 \end{array}.$$

This is expected to be a reasonable approximation if selection is not too strong (which we shall assume here). To model selection against hybrids, we assume that the three types have *relative fitnesses*

$$\begin{array}{c|c|c} aa & aA & AA \\ \hline 1 & 1-s_0 & 1 \end{array}.$$

We define relative fitness implicitly by explaining its effect. During reproduction, each individual produces a large (effectively infinite) number of germ cells, each of which carries a copy of all the genetic material of the parent (for our purposes this is just two copies of the gene in which we are interested). The germ cells then split into gametes (each containing one copy of the gene). All the gametes are put into a pool, and each individual in the next generation, independently, is created by fusing two gametes sampled at random from that pool.

The relative fitnesses above are reflected in each *heterozygote* ( $aA$ ) individual producing  $(1-s_0)$  times as many germ cells as a *homozygote* ( $aa$  or  $AA$ ) individual. The proportion of type  $a$  gametes in the pool is then

$$\begin{aligned} \bar{u}^* &= \frac{(\bar{u}^2 + \bar{u}(1-\bar{u})(1-s_0))}{(\bar{u}^2 + 2\bar{u}(1-\bar{u})(1-s_0) + (1-\bar{u})^2)} \\ &= \frac{\bar{u}^2 + \bar{u}(1-\bar{u})(1-s_0)}{1 - 2s_0\bar{u}(1-\bar{u})} \\ &= (1-s_0)\bar{u} + s_0(3\bar{u}^2 - 2\bar{u}^3) + \mathcal{O}(s_0^2) \\ &= \bar{u} + s_0\bar{u}(1-\bar{u})(2\bar{u}-1) + \mathcal{O}(s_0^2). \end{aligned}$$

In particular,

$$\bar{u}^* - \bar{u} = s_0\bar{u}(1-\bar{u})(2\bar{u}-1) + \mathcal{O}(s_0^2).$$

In an infinite population, the proportions of alleles among offspring will exactly follow those in the pool of gametes, and if  $s_0 = s/M$  (where  $M$  is large), measuring time

in units of  $M$  generations, this suggests the approximation

$$\frac{d\bar{u}}{dt} = s\bar{u}(1 - \bar{u})(2\bar{u} - 1)$$

for the dynamics of the proportion of  $a$ -alleles. The Laplacian term in (1.3) is then added to capture dispersal of offspring.

In a finite population, we must account for the randomness inherent in drawing a finite sample from the pool of gametes. We assume that the population size  $N$  is large and fixed. The number of  $a$ -alleles among offspring is  $\text{Bin}(2N, \bar{u}^*)$ , and so the *proportion* of  $a$ -alleles has mean  $\bar{u}^*$  and variance  $\frac{1}{2N}\bar{u}^*(1 - \bar{u}^*)$ . Notice in particular, that we can expect the effects of the fluctuations to be relevant over timescales of  $\mathcal{O}(N)$  generations. As before we suppose that  $s_0 = s/M$ , and measure time in units of  $M$  generations. If  $M/N = \mathcal{O}(1)$ , the dynamics of the proportion of  $a$ -alleles can then be approximated by the Wright–Fisher diffusion

$$du = su(1 - u)(2u - 1)dt + \sqrt{\frac{M}{2N}}u(1 - u)dB_t, \quad (2.1)$$

where  $B$  is a one-dimensional Brownian motion.

Replacing the reaction term in equation (1.2) by  $su(1 - u)(2u - 1)$ , in one spatial dimension we obtain what can be thought of as a spatial analogue of (2.1) in which offspring sample gametes from a pool generated by adult individuals at the location at which they were born. The Wright–Fisher noise is supposed to capture the randomness inherent in the sampling.

## 2.2. (Mean) curvature flow

We are primarily interested in two spatial dimensions, when mean curvature flow reduces to curvature flow, but our results are valid for all  $d \geq 2$ . 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  $\mathbb{R}^d$ ).

**Definition 2.1** ((Mean) curvature 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  $\mathbf{n} = \mathbf{n}_t(u)$  denote the unit (inward) normal vector to  $\Gamma_t$  at  $u$  and let  $\kappa_t(u)$  denote the curvature of  $\Gamma_t$  at  $u$ . We say that  $\Gamma$  is a *curvature flow* if

$$\frac{\partial \Gamma_t(u)}{\partial t} = \kappa_t(u)\mathbf{n}_t(u) \quad (2.2)$$

for all  $t, u$ .

In higher dimensions, we replace  $S^1$  by  $S^{d-1}$ ,  $\mathbb{R}^2$  by  $\mathbb{R}^d$ , and  $\kappa_t$  by the *mean* curvature of  $\Gamma_t$  to obtain *mean curvature flow*.

**Remark 2.2.** Perhaps the easiest way to visualize the curvature at a point  $P$  on a differentiable curve in  $\mathbb{R}^2$  is as the reciprocal of the radius of the *osculating circle* at  $P$  which (if it exists) is the circle that best approximates the curve at the point  $P$ .

The curvature tells us how quickly the tangent to the curve changes as we traverse the curve. To make this concrete, first parametrize the curve in terms of its arc length:  $\Gamma(s) =$

$(x(s), y(s))$  with  $x'(s)^2 + y'(s)^2 = 1$ . The tangent vector to the curve at  $(x(s), y(s))$ ,  $\mathbf{T}(s) = (x'(s), y'(s))$ , has norm one, and the unit normal is  $\mathbf{n}(s) = (-y'(s), x'(s))$ . If the curve is twice differentiable, then  $\mathbf{T}'(s) = \kappa(s)\mathbf{n}(s)$ , where  $\kappa(s)$  is the (signed) curvature at the point. For example, for a circle of radius  $R$ ,  $(x(s), y(s)) = (R \cos(s/R), R \sin(s/R))$ , and  $\kappa(s) \equiv 1/R$ .

The circle is, of course, a rare example for which arc length is easy to calculate, but by an application of the chain rule, this allows one to calculate  $\kappa$  in terms of an arbitrary parametrization

$$\kappa = \frac{\det(\Gamma', \Gamma'')}{\|\Gamma'\|^3}.$$

In the biologically relevant case of two dimensions, curvature flow is sometimes called the *curve-shortening* flow and its behavior is well understood. The flow has a finite lifetime  $\mathcal{T}$ . For example, if  $\Gamma_0$  is a circle of radius  $R_0$ , then  $\Gamma_t$  will be a circle with radius  $R_t$  satisfying  $dR/dt = -1/R$ , so the curve shrinks to a point in time  $R_0^2/2$ . In fact, this behavior is generic in  $\mathfrak{d} = 2$ : in [24], it was shown that if  $\Gamma_0$  is convex then so is  $\Gamma_t$  for all  $t < \mathcal{T}$ , and that as  $t \uparrow \mathcal{T}$  the asymptotic “shape” of  $\Gamma_t$  is a circle; [26] showed that any smoothly embedded closed curve becomes convex at a time  $\tau < \mathcal{T}$ .

### 2.3. The motion of hybrid zones

To state a result about the behavior of the solution to (1.4) as  $\varepsilon \rightarrow 0$ , we shall need some regularity assumptions on the initial condition.

**Assumptions 2.3** (Assumptions on  $u^\varepsilon(0, x)$ ). Let  $u^\varepsilon(0, x) = p(x)$  where  $p$  takes values in  $[0, 1]$  and set

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

We suppose that  $\Gamma$  is a smooth hypersurface which is the boundary of an open set which is topologically equivalent to a sphere. (When  $\mathfrak{d} = 2$ , this just says  $\Gamma$  is a smooth curve, topologically equivalent to a circle.) We further assume:

(C1)  $\Gamma$  is  $C^\alpha$  for some  $\alpha > 3$ ;

(C2) for  $x$  outside  $\Gamma$ ,  $p(x) < \frac{1}{2}$ ; for  $x$  inside  $\Gamma$ ,  $p(x) > \frac{1}{2}$ ;

(C3) there exists  $r, \mu > 0$  such that, for all  $x \in \mathbb{R}^2$ ,  $|p(x) - \frac{1}{2}| \geq \mu(\text{dist}(x, \Gamma) \wedge r)$ .

Condition (C1) guarantees that mean curvature flow  $(\Gamma_t(\cdot))_t$  started from  $\Gamma$  exists up to some time  $\mathcal{T} > 0$ . The second condition is just a convention; the third is to ensure that the slope of  $p$  near  $\Gamma$  is not too small, and that  $p$  is bounded away from  $1/2$  for points that are not close to  $\Gamma$ .

We write  $d(x, t)$  for the signed distance from  $x$  to  $\Gamma_t$ , chosen to be positive inside  $\Gamma_t$  and negative outside. As sets,  $\Gamma_t = \{x \in \mathbb{R}^{\mathfrak{d}} : d(x, t) = 0\}$ .



**Theorem 2.4** (Special case of Chen [11, THEOREM 3]). *Let  $u^\varepsilon(t, x)$  solve (1.4) with  $u^\varepsilon(0, x) = p(x)$  satisfying Assumptions 2.3. Fix  $T^* \in (0, \mathcal{T})$  and let  $k \in \mathbb{N}$ . There exists  $\varepsilon_d(k) > 0$ , and  $a_d(k), c_d(k) \in (0, \infty)$  such that for all  $\varepsilon \in (0, \varepsilon_d)$  and  $t$  satisfying  $a_d \varepsilon^2 |\log \varepsilon| \leq t \leq T^*$ ,*

(1) *for  $x$  such that  $d(x, t) \geq c_d \varepsilon |\log \varepsilon|$ , we have  $u^\varepsilon(t, x) \geq 1 - \varepsilon^k$ ;*

(2) *for  $x$  such that  $d(x, t) \leq -c_d \varepsilon |\log \varepsilon|$ , we have  $u^\varepsilon(t, x) \leq \varepsilon^k$ .*

### 3. A PROBABILISTIC REPRESENTATION OF SOLUTIONS TO (1.4)

In [14] an analogue of Theorem 2.4 for a model which incorporates (weak) genetic drift is proved. A large part of that paper is devoted to providing a new proof of Theorem 2.4, for which we now explain the key ideas. It is based on a probabilistic representation of the solution to (1.4), and is readily adapted to a host of other situations, some of which we describe in Sections 4, 5, and 6.

For compatibility with the literature on partial differential equations, we shall suppose that all Brownian motions run at rate 2 (and so have infinitesimal generator  $\Delta$  rather than  $\frac{1}{2}\Delta$ ). The representation is in terms of a ternary branching Brownian motion in which:

(1) each individual has an independent exponentially distributed lifetime with mean  $\varepsilon^2$  at the end of which it is replaced, at the location where it died, by three offspring; and

(2) during its lifetime, each individual follows an independent Brownian motion.

We shall only ever be interested in this process started from a single individual at time 0.

Whereas Skorokhod's representation of the solution to the Fisher–KPP equation in [36] is just in terms of the locations of the individuals in a (binary) branching Brownian motion at time  $t$ , our representation of the solution to (1.4) will also require the structure of the tree relating the individuals in the ternary branching Brownian motion. The simplest way to encode that information is to use Ulam–Harris notation. Each individual is labeled by an element of  $\mathcal{U} = \bigcup_{m=0}^{\infty} \{1, 2, 3\}^m$ . The original ancestor is labeled  $\emptyset$ . The offspring of an individual with label  $\bar{i} = (i_1, \dots, i_m)$  receive the labels  $(\bar{i}, 1)$ ,  $(\bar{i}, 2)$ , and  $(\bar{i}, 3)$ . Thus, for example,  $(1, 3)$  is the label of the third child of the first child of the original ancestor.

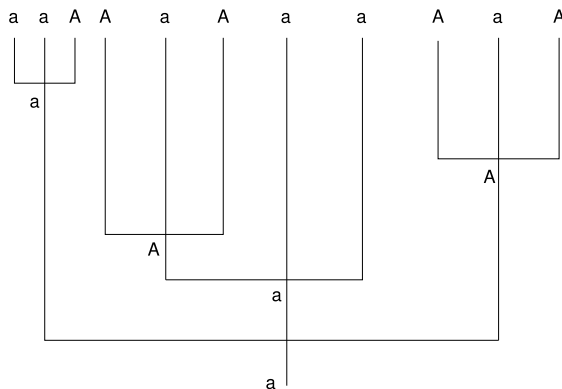
We shall use  $\mathbf{W}(t)$  to denote *historical ternary branching Brownian motion*, that is, the tree of Brownian paths traced out by ternary branching Brownian motion up until time  $t$ , and  $\mathcal{T}(\mathbf{W}(t))$  for the corresponding ternary tree, obtained by ignoring the spatial positions of individuals.

**Definition 3.1** (Majority voting in (historical) branching Brownian motion). For a fixed function  $p : \mathbb{R}^d \rightarrow [0, 1]$ , define a voting procedure on  $\mathbf{W}(t)$  as follows. We write  $\{W_i(t)\}_{i=1}^{N_t}$  for the spatial locations of the random number  $N(t)$  of individuals alive at time  $t$ . We call these individuals the leaves.

- (1) Each leaf, independently, votes  $a$  with probability  $p(W_i(t))$ ; otherwise it votes  $A$ .
- (2) At each branch point in  $\mathcal{T}(W(t))$ , the vote of the parent particle  $\bar{i}$  is the majority vote of the votes of its three children  $(\bar{i}, 1)$ ,  $(\bar{i}, 2)$ , and  $(\bar{i}, 3)$ .

This defines an iterative voting procedure, which runs inwards from the leaves of  $W(t)$  to the root. We define  $\mathbb{V}_p(W(t))$  to be the vote associated to the root.

This majority voting procedure is illustrated in Figure 1.



**FIGURE 1**

*Majority voting on a ternary tree.* Starting at the leaves, we move back to the root. At each branch point, the parent adopts the majority view of its children. In this example, the vote at the root is  $a$ .

**Lemma 3.2** (Majority voting and the Allen–Cahn equation). *Let  $W(t)$  be a historical ternary branching Brownian motion with branching rate  $1/\varepsilon^2$ , and let  $p : \mathbb{R}^d \rightarrow [0, 1]$ . The function*

$$u^\varepsilon(t, x) = \mathbb{P}_x^\varepsilon[\mathbb{V}_p(W(t)) = a] \quad (3.1)$$

*solves equation (1.4) with  $u^\varepsilon(0, x) = p(x)$ .*

The subscript  $x$  on the right hand side of equation (3.1) indicates that  $W$  starts from a single individual at  $x$  at time zero. The proof of Lemma 3.2 proceeds in a standard way by partitioning over whether or not the ancestor in the branching Brownian motion dies in the first  $\delta t$  of time, and thus calculating

$$\lim_{\delta t \rightarrow 0} \frac{(u^\varepsilon(t + \delta t, x) - u^\varepsilon(t, x))}{\delta t}. \quad (3.2)$$

To understand why majority voting gives rise to the desired nonlinearity, consider what happens if the ancestor *does* die in the first  $\delta t$  of time, which happens with probability  $\delta t/\varepsilon^2$ . Each offspring, independently, votes  $a$  with the same probability,  $u$  say. The probability that the majority of their 3 votes is  $a$  is  $u^3 + 3u^2(1 - u) = u(1 - u)(2u - 1) + u$ . Assuming

some continuity so that we can take  $u \approx u^\varepsilon(t, x)$  as  $\delta t \rightarrow 0$ , we see that the contribution to (3.2) from the event {the ancestor died before time  $\delta t$ } is  $u^\varepsilon(1 - u^\varepsilon)(2u^\varepsilon - 1)/\varepsilon^2$  as required. With probability  $1 - \delta t/\varepsilon^2$  the ancestor *did not* die before  $\delta t$ , and it followed a Brownian motion which is at position  $W_{\delta t}$  at time  $\delta t$ . Using the Markov property at time  $\delta t$ , the Laplacian term arises from this event as  $\lim_{\delta t \rightarrow 0} (\mathbb{E}_x[u^\varepsilon(t, W_{\delta t})] - u(t, x))/\delta t$ .

With the representation (3.1), the conclusion of Theorem 2.4 can be written:

- (1) for  $x$  with  $d(x, t) \geq c_d \varepsilon |\log \varepsilon|$ ,  $\mathbb{P}_x^\varepsilon[\mathbb{V}_p(\mathbf{W}(t)) = a] \geq 1 - \varepsilon^k$ ;
- (2) for  $x$  with  $d(x, t) \leq -c_d \varepsilon |\log \varepsilon|$ ,  $\mathbb{P}_x^\varepsilon[\mathbb{V}_p(\mathbf{W}(t)) = a] \leq \varepsilon^k$ .

The intuition behind the proof of these statements is very simple. First observe that majority voting increases bias: if  $p < \frac{1}{2}$ ,  $p^3 + 3p^2(1 - p) < p$ ; if  $p > \frac{1}{2}$ ,  $p^3 + 3p^2(1 - p) > p$ . Since the branching rate is  $1/\varepsilon^2$ , our branching Brownian motion sees many rounds of majority voting in a very short space of time, and so a small bias in votes at the leaves of the tree translates into a large bias at the root. As a result, a narrow interface will be generated across which there is a rapid transition from  $\mathbb{P}_x^\varepsilon[\mathbb{V}_p(\mathbf{W}(t)) = a]$  being close to zero, to it being close to one. Suppose that in fact this transition is sharp, and the solution to equation (1.4) is the indicator function of a region bounded by a surface  $\Gamma$ . Taking this solution as the new initial condition, after a small time  $h$ , we once again expect that the solution is close to a sharp interface whose position,  $\Gamma_h$ , marks the transition from a voting bias in favor of type  $a$ , to one in favor of type  $A$ . That is,  $\Gamma_h \approx \{x \in \mathbb{R}^d : T_h \mathbf{1}_\Gamma(x) = 1/2\}$  where  $T$  denotes the heat semigroup. If we replace the solution at time  $h$  by  $\mathbf{1}_{\Gamma_h}$  and repeat this process, we are actually performing the Merriman–Bence–Osher (MBO) algorithm for simulating mean curvature flow [32]. To gain some intuition for the role of mean curvature flow, consider the special case in which  $\Gamma$  is a sphere of radius  $R$  in  $\mathbb{R}^d$ . Then we are approximating  $\Gamma_h$  by  $\{x \in \mathbb{R}^d : \mathbb{P}_x[\|W_h\| > R] = 1/2\}$ , where  $W$  is  $d$ -dimensional Brownian motion. Since the radial part of a  $d$ -dimensional Brownian motion is a  $d$ -dimensional Bessel process, while it is close to  $\Gamma$ ,  $\|W\|$  will be distributed as a one-dimensional Brownian motion  $B$  with drift close to  $(d - 1)/R$  (remembering that our Brownian motions all run at rate two), and so we are approximating  $\Gamma_h$  by the set of points for which  $\mathbb{P}_{\|x\|}[B_h + h(d - 1)/R > R] = 1/2$ ; in other words, by symmetry of  $B$ , by  $\{x : \|x\| = R - h(d - 1)/R\}$ . The mean curvature of  $\Gamma$  is  $(d - 1)/R$ , and so for small  $h$ ,  $\Gamma_h$  is close to the surface obtained by evolving  $\Gamma$  according to mean curvature flow for time  $h$ .

This intuitive picture is close to the structure of the rigorous proof which has two main ingredients: an analysis of the one-dimensional solution, started from a Heaviside initial condition; and coupling (close to the interface) of  $d(W_s, t - s)$  with a one-dimensional Brownian motion.

#### 4. LONG-RANGE DISPERSAL

The Laplacian in equation (1.4) reflects an assumption that offspring remain close to their parents. However, for many organisms this may fail; see, for example, [10] for a

discussion of long range seed dispersal in plants. To incorporate this into equation (1.5), we replace the Laplacian by a fractional Laplacian,

$$\frac{\partial v}{\partial t} = (-\Delta)^{\frac{\alpha}{2}} v + sv(1-v)(2v-1), \quad (4.1)$$

where (for smooth functions  $f$  which decay sufficiently fast)

$$(-\Delta)^{\frac{\alpha}{2}} f(x) := C_{\alpha} \lim_{\delta \rightarrow 0} \int_{\mathbb{R}^d \setminus B_{\delta}(x)} \frac{f(y) - f(x)}{\|y - x\|^{d+\alpha}} dy. \quad (4.2)$$

Here  $C_{\alpha} := 2^{\alpha} \Gamma(\frac{\alpha}{2} + \frac{d}{2}) / (\pi^{d/2} |\Gamma(-\frac{\alpha}{2})|)$ , where  $\Gamma$  is the Gamma function, and  $B_{\delta}(x)$  is the ball of radius  $\delta$  about  $x$ .

The operator (4.2) is the generator of a symmetric  $\alpha$ -stable process, and the probabilistic representation of solutions to (4.1) follows by substituting a branching  $\alpha$ -stable process for the branching Brownian motion in Section 3. If we are to recover an analogue of Theorem 2.4, we expect to need to consider scales over which the spatial motion along each branch is close to a Brownian motion. We appeal to a decomposition often used in the numerical simulation of symmetric stable processes, see, for example, [2]. If we run an  $\alpha$ -stable process at rate  $I(\varepsilon)^{\alpha-2}$  (where  $I(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ ), then the process obtained by censoring jumps of size greater than  $I(\varepsilon)$  can be approximated by Brownian motion. To show that the *uncensored* process along a branch of our ternary tree is close to a Brownian motion, we need to control the number of jumps of size at least  $I(\varepsilon)$  before an exponential time with mean  $\varepsilon^2$ . Since we have time-changed the stable process by  $I(\varepsilon)^{\alpha-2}$ , the rate of such jumps is  $\mathcal{O}(I(\varepsilon)^{-2})$ , and so in order that branches on which we see jumps of size more than  $I(\varepsilon)$  be rare, we take  $I(\varepsilon)/\varepsilon \rightarrow \infty$ .

With this in mind, set  $\sigma^2 = 2C_{\alpha}/(2-\alpha)$  and rescale time and space by  $t \mapsto \varepsilon^2 t, x \mapsto \varepsilon^{2/\alpha} I(\varepsilon)^{1-2/\alpha} x$ . When  $\alpha = 2$ , we recover the diffusive scaling. Equation (4.1) becomes

$$\frac{\partial v^{\varepsilon}}{\partial t} = \frac{\sigma^{-2}}{I(\varepsilon)^{2-\alpha}} (-\Delta)^{\frac{\alpha}{2}} v^{\varepsilon} + \frac{1}{\varepsilon^2} v^{\varepsilon} (1 - v^{\varepsilon})(2v^{\varepsilon} - 1), \quad v^{\varepsilon}(0, x) = p(x). \quad (4.3)$$

In [7], an analogue of Theorem 2.4 is proved for functions  $I : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  satisfying:

$$(\mathcal{A}1) \quad \lim_{\varepsilon \rightarrow 0} I(\varepsilon) |\log(\varepsilon)|^k = 0 \quad \forall k \in \mathbb{N}.$$

$$(\mathcal{A}2) \quad \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon^2 |\log(\varepsilon)|}{I(\varepsilon)^2} = 0.$$

$$(\mathcal{A}3) \quad \lim_{\varepsilon \rightarrow 0} H(\varepsilon) := I(\varepsilon)^2 |\log(\varepsilon)|^{\frac{2d}{\alpha} - d - 1} + \frac{I(\varepsilon)^{2\alpha}}{\varepsilon^2} |\log(\varepsilon)|^{\alpha} = 0.$$

Note that Assumptions (A2) and (A3) are incompatible as soon as  $\alpha \leq 1$ .

**Theorem 4.1** ([7, THEOREM 1.5]). *Let  $\alpha \in (1, 2)$  and suppose that  $I(\varepsilon)$  satisfies Assumptions (A1)–(A3) above. Suppose  $v^{\varepsilon}$  solves equation (4.3) with initial condition  $p$  satisfying Assumptions 2.3. Let  $\mathcal{T}$  and  $d(x, t)$  be as in Section 2.3, and fix  $T^* \in (0, \mathcal{T})$ . Then there exists  $\varepsilon_d(\alpha, I)$ ,  $a_d(\alpha, I)$ ,  $c_d(\alpha, I)$ ,  $M(\alpha, I) > 0$  such that, for  $\varepsilon \in (0, \varepsilon_d)$  and  $a_d \varepsilon^2 |\log \varepsilon| \leq t \leq T^*$ ,*

$$(1) \text{ for } x \text{ with } d(x, t) \geq c_d I(\varepsilon) |\log \varepsilon|, \text{ we have } v^{\varepsilon}(t, x) \geq 1 - \frac{\varepsilon^2}{I(\varepsilon)^2} - M(H(\varepsilon) + I(\varepsilon)^{\alpha-1});$$

(2) for  $x$  with  $d(x, t) \leq -c_d I(\varepsilon) |\log \varepsilon|$ , we have  $v^\varepsilon(t, x) \leq \frac{\varepsilon^2}{I(\varepsilon)^2} + M(H(\varepsilon) + I(\varepsilon)^{\alpha-1})$ .

For example,  $I(\varepsilon) = \varepsilon |\log(\varepsilon)|$  fulfills Assumptions (A1)–(A3), and the “error”  $\varepsilon^2/I(\varepsilon)^2 + M(H(\varepsilon) + I(\varepsilon)^{\alpha-1})$  is of order  $1/(\log \varepsilon)^2$ . There are two competing effects: we want to take  $I(\varepsilon)$  as small as possible if the approximation of the small jumps of the stable process by a Brownian motion is to be good; on the other hand, we need  $I(\varepsilon)$  to be large if branches along which we see a jump of size more than  $I(\varepsilon)$  are to be rare. In contrast to the Brownian case, these cannot be balanced to obtain an error of order  $\varepsilon^k$  for arbitrary  $k$ .

## 5. ASYMMETRY AND BLOCKING

So far we have worked exclusively on the whole of Euclidean space. In this section we see that, in some scenarios, the geometry of the domain can be important.

### 5.1. An asymmetric reaction: homozygotes of different fitnesses

In our justification of equation (1.3) in Section 2, we assumed that both homozygotes were equally fit. It is natural to ask what happens if that is not the case? Suppose, for example, that we take relative fitnesses

$$\begin{array}{c|c|c} aa & aA & AA \\ \hline 1 + \gamma_1 s_1 & 1 - s_1 & 1 \end{array},$$

where  $\gamma_1$  is assumed small. Mimicking our previous approach, and setting  $(2 + \gamma_1)s_1/2 = s/M$ ,  $2/(2 + \gamma_1) = 1 - \gamma$ , we recover equation (1.5). The one-dimensional equation

$$\frac{\partial u}{\partial t} = m \frac{\partial^2 u}{\partial x^2} + su(1-u)(2u - (1 - \gamma))$$

has a traveling wave solution of the form

$$u(x, t) = \left( 1 + \exp\left(-\sqrt{\frac{s}{m}}(x + \gamma\sqrt{mst})\right) \right)^{-1}, \quad (5.1)$$

connecting 0 at  $-\infty$  to 1 at  $\infty$ , and with wave speed  $\gamma\sqrt{ms}$ . In particular, if we scale  $m$  and/or  $s$ , then we may also have to scale  $\gamma$  in order to obtain a finite wavespeed. With this in mind, [25] considers the equation

$$\frac{\partial u^\varepsilon}{\partial t} = \varepsilon^{1-\ell} \Delta u^\varepsilon + \frac{1}{\varepsilon^{1+\ell}} u^\varepsilon (1 - u^\varepsilon) (2w^\varepsilon - (1 - \gamma_\varepsilon)), \quad x \in \mathbb{R}^d, \quad t > 0, \quad (5.2)$$

where  $\gamma_\varepsilon = v\varepsilon^{\tilde{\ell}}$  for some nonnegative  $v$  and  $\tilde{\ell}$ , with the additional condition that  $v < 1$  when  $\tilde{\ell} = 0$ , and  $\ell = \min(\tilde{\ell}, 1)$ .

Notice that with these parameters, the one-dimensional wave has speed of  $\mathcal{O}(1)$  if  $\tilde{\ell} \leq 1$  and tending to zero as  $\varepsilon^{\tilde{\ell}-1}$  if  $\tilde{\ell} > 1$ . We define

$$v_\varepsilon = \begin{cases} v & \text{if } \tilde{\ell} \leq 1, \\ \gamma_\varepsilon/\varepsilon & \text{if } \tilde{\ell} \in (1, 2], \\ 0 & \text{if } \tilde{\ell} > 2. \end{cases} \quad (5.3)$$

Set  $u^\varepsilon(x, 0) = p(x)$ , take  $\Gamma = \{x \in \mathbb{R}^d : p(x) = (1 + \gamma_\varepsilon)/2\}$ , and modify Assumptions 2.3 in the obvious way (by replacing  $1/2$  by  $(1 + \gamma_\varepsilon)/2$ ).

**Theorem 5.1** (Restatement of [25, THEOREM 2.4]). *Let  $u^\varepsilon$  solve equation (5.2) with initial condition  $p$  satisfying Assumptions 2.3 (modified as described above), and let*

$$\frac{\partial \tilde{\Gamma}(s)}{\partial t} = (-v_\varepsilon + \kappa_t(s))\mathbf{n}_t(s), \quad (5.4)$$

*until the time  $\mathcal{T}$  at which  $\tilde{\Gamma}$  develops a singularity. Write  $\tilde{d}$  for the signed distance to  $\tilde{\Gamma}$  (chosen to be positive inside  $\tilde{\Gamma}$ ). Fix  $T^* \in (0, \mathcal{T})$  and  $k \in \mathbb{N}$ . There exists  $\varepsilon_d(k) > 0$ , and  $a_d(k), c_d(k) \in (0, \infty)$  such that for all  $\varepsilon \in (0, \varepsilon_d)$  and  $t$  satisfying  $a_d \varepsilon^{1+\ell} |\log \varepsilon| \leq t \leq T^*$ ,*

(1) *for  $x$  such that  $\tilde{d}(x, t) \geq c_d \varepsilon |\log \varepsilon|$ , we have  $u^\varepsilon(t, x) \geq 1 - \varepsilon^k$ ;*

(2) *for  $x$  such that  $\tilde{d}(x, t) \leq -c_d \varepsilon |\log \varepsilon|$ , we have  $u^\varepsilon(t, x) \leq \varepsilon^k$ .*

**Remark 5.2.** When  $\tilde{\ell} = 1$ , Theorem 5.1 is a special case of Theorem 1.3 of [1] in which more general “slightly unbalanced” bistable nonlinearities are considered.

For  $\tilde{\ell} \leq 2$ ,  $v_\varepsilon$  in (5.3) and (5.4) corresponds to the one-dimensional wavespeed derived above. For  $\tilde{\ell} > 2$ , the wavespeed converges to zero sufficiently quickly as  $\varepsilon \rightarrow 0$  that it is not necessary to include the corresponding small contribution from the constant flow in (5.4).

We shall focus on equation (5.2) with  $\tilde{\ell} = 1$ ,

$$\frac{\partial u^\varepsilon}{\partial t} = \Delta u^\varepsilon + \frac{1}{\varepsilon^2} u^\varepsilon (1 - u^\varepsilon) (2u^\varepsilon - (1 - \varepsilon v)) \quad x \in \mathbb{R}^d, \quad t > 0. \quad (5.5)$$

The approach of [25] is to extend the probabilistic representation to the asymmetric case.

**Lemma 5.3.** *Let  $\tilde{W}(t)$  be a historical ternary branching Brownian motion with branching rate  $(1 + \varepsilon v)/\varepsilon^2$ , and let  $p : \mathbb{R}^d \rightarrow [0, 1]$ . Define a voting procedure on  $\tilde{W}(t)$  as follows:*

- (1) *Each leaf, independently votes  $a$  with probability  $p(\tilde{W}_i(t))$ , otherwise it votes  $A$ ;*
- (2) *at a branch point, the parental vote is the majority vote of the children unless precisely one offspring vote is  $a$ , in which case the parent votes  $a$  with probability  $2\varepsilon v/(3 + 3\varepsilon v)$ .*

*Write  $\tilde{\mathbb{V}}_p(\tilde{W}(t))$  for the vote associated with the root. Then*

$$u^\varepsilon(t, x) = \mathbb{P}_x^\varepsilon[\tilde{\mathbb{V}}_p(\tilde{W}(t)) = a]$$

*solves equation (5.5) with  $u^\varepsilon(0, x) = p(x)$ .*

The proof of Theorem 5.1 closely follows the probabilistic proof of Theorem 2.4 in [14], except that the signed distance  $\tilde{d}(W_s, \tilde{\Gamma}_{t-s})$  is coupled to a one-dimensional Brownian motion with drift  $v$ .

**Remark 5.4** (Other voting schemes). The probabilistic representation above is far from unique. For example, it might seem more natural to write the reaction term in (5.5) as  $\frac{1}{\varepsilon^2}(u^\varepsilon(1-u^\varepsilon)(2u^\varepsilon-1) + \varepsilon v u^\varepsilon(1-u^\varepsilon))$ , and express the solution in terms of a branching Brownian motion with a mixture of binary branching at rate  $v/\varepsilon$ , with the rule that the parent votes  $a$  unless both offspring vote  $A$ , and ternary branching at rate  $1/\varepsilon^2$  with the majority voting rule. However, it turns out to be much more convenient to base the proof on a ternary tree. To obtain the voting mechanism above, we rewrite the quadratic term  $u(1-u)$  as the sum of two cubic terms using that  $1 = u + (1-u)$ .

Voting schemes are very general. In [35], O’Dowd showed that if  $P(u)$  is any polynomial with  $P(0) \geq 0$  and  $P(1) \leq 0$  (or vice versa), then the solution to

$$\frac{\partial u}{\partial t} = \Delta u + P(u)$$

can be represented in terms of a historical  $n$ -ary branching Brownian motion (where  $n$  is the degree of  $P$ ) and a rule for assigning votes to a parent according to the votes of its offspring.

## 5.2. Geometry matters: blocking

We now turn our attention to solutions to (5.5) on domains  $\Omega \subseteq \mathbb{R}^d$  with reflecting boundary conditions. We focus on the fate of the favored allele as it tries to expand through a semiinfinite domain. We shall consider “cylindrical” domains of the form

$$\Omega = \{(x_1, x') : x_1 \in \mathbb{R}, x' \in \phi(x_1) \subseteq \mathbb{R}^{d-1}\}. \quad (5.6)$$

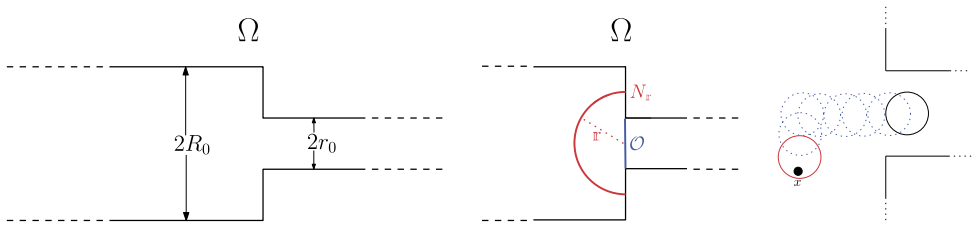
We shall always take the initial condition  $u^\varepsilon(0, x) = \mathbf{1}_{x_1 \geq 0}$ .

**Theorem 5.5** ([8], Theorems 1.4, 1.5, 1.6, 1.7, paraphrased). *Let  $u$  be the solution to equation (1.5) on  $\Omega$  with normal reflection on the boundary and initial condition  $u(x, 0) = \mathbf{1}_{x_1 \geq 0}$ . Depending on the geometry of the domain  $\Omega$  we have one of three possible asymptotic behaviors of the solution of equation (5.5):*

- (1) *there can be complete invasion, that is,  $u(x, t) \rightarrow 1$  as  $t \rightarrow \infty$  for every  $x \in \Omega$ ;*
- (2) *there can be blocking of the solution, meaning that  $u(x, t) \rightarrow u_\infty(x)$  as  $t \rightarrow \infty$ , with  $u_\infty(x) \rightarrow 0$  as  $x_1 \rightarrow -\infty$ ;*
- (3) *there can be axial partial propagation, meaning that  $u(x, t) \rightarrow u_\infty(x)$  as  $t \rightarrow \infty$ , with  $\inf_{x \in \mathbb{R} \times B_R} u_\infty(x) > c > 0$  for some  $R > 0$ , where  $B_R$  is the ball of radius  $R$  centered at 0 in  $\mathbb{R}^{d-1}$ .*

*Which behavior is observed depends on the geometry of the domain  $\Omega$ . For example, there will be complete invasion if  $\Omega$  is decreasing as  $x_1$  decreases; axial partial propagation if it contains a straight cylinder of sufficiently large cross-section; and there can be blocking if there is an abrupt change in the geometry.*

The results of [18] concerning the behavior of solutions to (5.5) complement those of [8]. (The addition of the parameter  $\varepsilon$ , which is not present in the work of [8], prevents direct



**FIGURE 2**

Left to right: (a) the domain  $\Omega$  of Theorems 5.6 and 5.7; (b) the opening  $\mathcal{O}$  and the hemispherical shell  $N_r$  used in the proof of Theorem 5.6; (c) an illustration of “chaining” used in the proof of Theorem 5.7. Image taken from [18].

comparison.) As in the previous sections, they are based on the probabilistic representation of solutions. Following [8], we begin with the very special form of  $\Omega$  depicted in Figure 2.

**Theorem 5.6 ([18, THEOREM 1.6]).** *Let  $u^\varepsilon$  denote the solution to equation (5.5) on the domain  $\Omega$  in Figure 2, with reflecting boundary condition and  $u^\varepsilon(0, x) = \mathbf{1}_{x_1 \geq 0}$ . Suppose  $r_0 < \frac{d-1}{v} \wedge R_0$ . Define  $N_r = \{x \in \Omega : \|x\| = r, x_1 < 0\}$ , where  $\frac{d-1}{v} \wedge R_0 > r > r_0$ , and let  $\hat{d}(x)$  be the signed (Euclidean) distance of any point  $x \in \Omega$  to  $N_r$  (chosen to be negative as  $x_1 \rightarrow -\infty$ ). Let  $k \in \mathbb{N}$ . Then there is  $\hat{\varepsilon}(k) > 0$  and  $M(k) > 0$  such that for all  $\varepsilon \in (0, \hat{\varepsilon})$ , and all  $t \geq 0$ ,*

$$\text{for } x = (x_1, \dots, x_d) \in \Omega \text{ such that } \hat{d}(x) \leq -M(k)\varepsilon|\log(\varepsilon)|, \text{ we have } u^\varepsilon(x, t) \leq \varepsilon^k.$$

In other words, if the aperture  $r_0$  is too small, then, for sufficiently small  $\varepsilon$ , blocking occurs. With the machinery of Section 5.1 in place, the proof is straightforward. First we check that the solution to (5.5) on  $\Omega$  is monotone in the initial condition, which allows us to compare with the solution started from an initial condition  $p$  which dominates  $\mathbf{1}_{x_1 \geq 0}$ , is radially symmetric in the left half plane, satisfies  $p(x) = (1 - \gamma_\varepsilon)/2$  on the hemispherical shell  $N_r$ , and fulfills the analogue of conditions (C2) and (C3) from Assumptions 2.3 (with  $\Gamma$  replaced by  $N_r$  and  $1/2$  by  $(1 - \gamma_\varepsilon)/2$ ). For this initial condition, it is straightforward to adapt the proof for the whole Euclidean space from [25], and indeed things are simplified considerably by the radial symmetry.

The converse of Theorem 5.6 is also true in the following sense.

**Theorem 5.7 ([18, THEOREM 1.7]).** *Let  $u^\varepsilon$  be as in Theorem 5.6. Suppose  $r_0 > \frac{d-1}{v}$ , then for all  $x \in \Omega$  and  $\delta > 0$  there is  $\hat{t} := \hat{t}(x_1, \delta, R_0, r_0) > 0$  and  $\hat{\varepsilon}$  such that, for all  $\varepsilon \in (0, \hat{\varepsilon})$  and  $t \geq \hat{t}$ , we have  $u^\varepsilon(t, x) \geq 1 - \delta$ .*

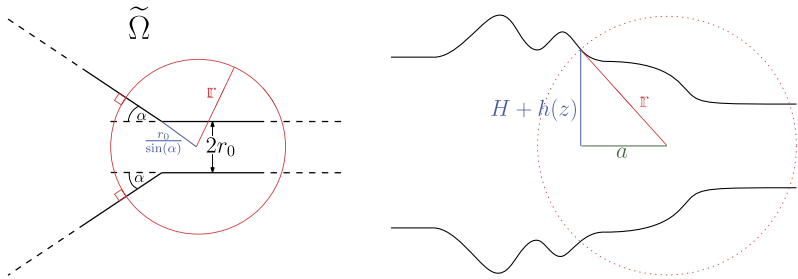
Again the proof exploits monotonicity in the initial condition. The solution dominates one started from  $(1 - \varepsilon)$  times the indicator of a ball of radius  $r > (d - 1)/v$ , with center sitting on the  $x_1$ -axis and contained in  $\Omega \cap \{x : x_1 \geq 0\}$ . This time, adapting the arguments for the solution on  $\mathbb{R}^d$  tells us that at a later time that solution dominates  $(1 - \varepsilon)$  times the indicator of a ball with larger radius  $r'$ , but the same center, strictly contained within  $\Omega$ . We now start the process again, taking as initial condition  $1 - \varepsilon$  times the indicator of a ball



of radius  $r$ , and with center shifted a distance  $r' - r$ . Continuing in this way, we can find a chain of balls connecting any point  $x \in \Omega$  to the original ball. This process of “chaining” is illustrated in Figure 2. It mirrors the use of the “sliding ball” assumption to prove complete propagation in [8].

Together, Theorems 5.6 and 5.7 say that there is a sharp transition at the critical radius  $(d - 1)/\nu$ . As described in Section 1, this is the radius of the shell at which the constant and curvature flow exactly balance on the whole of  $\mathbb{R}^d$ . However, in that case a small perturbation results in complete invasion or extinction of the favored type, here a stable interface will be maintained.

The domain  $\Omega$  of Theorem 5.6 is very special. However, the crucial step was to be able to cover the opening  $\mathcal{O}$  illustrated in Figure 2 by a hemispherical shell of less than the critical radius  $(d - 1)/\nu$ , and orthogonal to the boundary of the domain where they intersect. The same result will follow (from essentially the same argument) for any domain which can be “blocked” by a portion of such a shell in this way. As a first step, consider the domain  $\widetilde{\Omega}$ , which opens out as a truncated cone, and the shell of radius  $r$  shown in Figure 3. We can choose  $r < (d - 1)/\nu$  precisely when  $r_0 < (d - 1) \sin \alpha/\nu$ .



**FIGURE 3** (Left) The domain  $\widetilde{\Omega}$ . (See text below Theorem 5.7.) (Right) An example of a domain from Theorem 5.8. Condition (5.7) that guarantees that we can insert a portion of a spherical shell as shown with radius less than  $(d - 1)/\nu$  can be read off from that for  $\widetilde{\Omega}$  on setting  $r_0 = H + h(z)$  and  $\sin \alpha = h'(z)/\sqrt{1 + h'(z)^2}$ . Image taken from [18].

The intuition behind blocking is that if the domain opens out too rapidly, then off-spring of favored individuals are “spread too thin” and selection against hybrids will rapidly eliminate their descendants. Our approach has been to seek a shell of sufficiently small radius that intercepts the boundary of the domain orthogonally, but if the domain is opening even faster, in the sense that expanding the shell radially one stays within the domain, at least for a short time, this effect will be further amplified. This is the meaning of the condition (5.7) in the following theorem.

**Theorem 5.8 ([18, THEOREM 1.9]).** Suppose that  $u^\varepsilon$  solves (5.5) where  $\Omega \subseteq \mathbb{R}^d$  is defined as in (5.6) with

$$\phi(x_1) = \{\|x'\| \leq H + h(-x_1)\},$$

and  $h$  a nonnegative  $C^1$  function. Suppose that

$$\inf_{z>0} \left\{ H + h(z) - \left( \frac{\mathfrak{d} - 1}{\nu} \right) \frac{h'(z)}{\sqrt{1 + h'(z)^2}} \right\} < 0. \quad (5.7)$$

Fix  $k \in \mathbb{N}$ . There exist  $x_0 < 0$ ,  $\hat{\varepsilon}(k) > 0$  and  $M(k) > 0$  such that for all  $\varepsilon \in (0, \hat{\varepsilon})$  and  $t \geq 0$ , for  $x = (x_1, \dots, x_{\mathfrak{d}}) \in \Omega$  such that  $x_1 \leq x_0 - M(k)\varepsilon |\log(\varepsilon)|$  we have  $u^\varepsilon(x, t) \leq \varepsilon^k$ .

Condition (5.7) can be understood from the condition  $r_0 < (\mathfrak{d} - 1) \sin \alpha / \nu$  on  $\widetilde{\Omega}$  on setting  $r_0 = H + h(z)$  and  $\sin \alpha = h'(z) / \sqrt{1 + h'(z)^2}$ .

Conversely, if the domain does not open up sufficiently fast, we have invasion.

**Theorem 5.9** ([18, THEOREM 1.10]). Suppose that  $u^\varepsilon$  solves (5.5) where  $\Omega \subseteq \mathbb{R}^{\mathfrak{d}}$  is defined as in (5.6) with

$$\phi(x_1) = \{\|x'\| \leq H + h(-x_1)\},$$

and  $h$  a nonnegative  $C^1$  function. Suppose that

$$\inf_{z>0} \left\{ H + h(z) - \left( \frac{\mathfrak{d} - 1}{\nu} \right) \frac{h'(z)}{\sqrt{1 + h'(z)^2}} \right\} > 0.$$

Then for all  $x \in \Omega$  and  $\delta > 0$  there is  $\hat{t} := \hat{t}(x_1, \delta) > 0$  and  $\hat{\varepsilon}$  such that, for all  $\varepsilon \in (0, \hat{\varepsilon})$  and  $t \geq \hat{t}$ , we have  $u^\varepsilon(t, x) \geq 1 - \delta$ .

These results (valid for any  $\mathfrak{d} \geq 2$ ) are somewhat analogous to those of [30], which consider a plane curve evolving according to equation (5.4) in a two-dimensional cylinder with a periodic saw-toothed boundary. The authors say that such a curve is a *periodic* traveling wave with effective speed  $\delta / T_\delta$  if  $\widetilde{\Gamma}_{t+T_\delta}(s) = \widetilde{\Gamma}_t(s) + \delta$  for some  $\delta > 0$ . Setting  $h_\delta(x) = \delta h_1(x/\delta)$  and letting  $\delta \rightarrow 0$  leads to the homogenization limit of the wave, with speed  $c_0 = \lim_{\delta \rightarrow 0} c_\delta$ . They show that  $c_0 > 0$  for  $\nu H > \sin \alpha$  with  $\alpha$  determined by  $\tan \alpha = \max_x h'(x)$ , but that the wave is blocked for small enough  $\delta$  if  $\nu H < \sin \alpha$ .

## 6. ADDING NOISE

In Section 2, we motivated the noise appearing in equation (1.2) as a means of taking account of the randomness due to resampling inherent in reproduction in a finite population. Although in  $\mathfrak{d} = 1$ , where the equations are well-posed, quite a lot is known about the solutions to stochastic reaction–diffusion equations like (1.2), in  $\mathfrak{d} \geq 2$  such equations have no solution. On the other hand, we have seen in Section 5.2 that populations may behave quite differently in  $\mathfrak{d} = 1$  and  $\mathfrak{d} = 2$  and so it may be misleading to only consider the one-dimensional equation.

The Spatial  $\Lambda$ -Fleming–Viot process was introduced in [4, 17] as an alternative way to capture the effect of genetic drift in models for proportions of different allelic types in populations evolving in a spatial continuum. Although originally introduced for selectively neutral populations, it can be thought of as providing a framework for modeling, which can readily be adapted to incorporate a wealth of biologically relevant features, including natural selection.

First, we define a very special version of the Spatial  $\Lambda$ -Fleming–Viot process for a neutral population evolving in  $\mathbb{R}^d$ . As usual, we are most interested in  $d = 2$ . At each time  $t$ , the random function  $\{w_t(x) : x \in \mathbb{R}^d\}$  will model the proportion of  $a$ -alleles at spatial position  $x$  at time  $t$ . Strictly speaking, the process is only defined up to a Lebesgue-null set. The identification

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

provides a one-to-one correspondence between its state space and 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. We abuse notation and also denote the state space of the process  $(w_t)_{t \in \mathbb{R}_+}$  by  $\mathcal{M}_\lambda$ .

**Definition 6.1** (A neutral Spatial  $\Lambda$ -Fleming–Viot process (SLFV)). Fix  $u \in (0, 1]$  and  $r > 0$ . Let  $\Pi$  be a Poisson Point Process on  $\mathbb{R}_+ \times \mathbb{R}^d$  with intensity measure  $dt \otimes dx$ . The *Spatial  $\Lambda$ -Fleming–Viot process* driven by  $\Pi$ , with *event radius*  $r$  and *impact parameter*  $u$ , is the  $\mathcal{M}_\lambda$ -valued process  $(w_t)_{t \geq 0}$  with dynamics given as follows.

If  $(t, x) \in \Pi$ , a reproduction event occurs at time  $t$  within the closed ball  $B(x, r)$  of radius  $r$  centered on  $x$ :

- (1) Choose a parental location  $z$  uniformly at random in  $B(x, r)$ , 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  $y \in B(x, r)$ , set  $w_t(y) = (1 - u)w_{t-}(y) + u1_{\{\alpha_0=a\}}$ .

**Remark 6.2.** Suppose that a reproduction event affects the ball  $B(x, r)$  in which the proportion of  $a$ -alleles immediately before the event is  $w$ , and write  $w^*$  for the proportion of  $a$ -alleles immediately after the event. Then

$$\mathbb{E}[w^* - w] = 0, \quad \text{and} \quad \text{var}(w^* - w) = u^2 w(1 - w).$$

This can be compared to the (Wright–Fisher) sampling noise in Section 2.1.

This is a very special case of the SLFV, even for a neutral population. More generally, one can take both  $r$  and  $u$  to be random. See [19] for a construction of the process under very much more general conditions.

Instead of sampling a parental location, and then a parental type, we could equally have just sampled types independently and uniformly at random according to the proportions in the region affected by the event. The two-step description is convenient as we wish to trace the ancestry of a sample from the population. Things are made particularly simple as the Poisson process  $\Pi$  that dictates reproduction events is reversible (with the same distribution). We write  $\overleftarrow{\Pi}$  for the time-reversed process.

**Definition 6.3** (SLFV dual). The process  $(\mathcal{P}_t)_{t \geq 0}$  is the  $\bigcup_{l \geq 1} (\mathbb{R}^d)^l$ -valued Markov process with dynamics defined as follows.

The process starts from a finite collection of points  $\xi_1(0), \dots, \xi_{N(0)} \in \mathbb{R}^d$ . We write  $\mathcal{P}_t = (\xi_1(t), \dots, \xi_{N(t)}(t))$ , where the random number  $N(t) \in \mathbb{N}$  is the number of individuals alive at time  $t$ , and  $\{\xi_i(t)\}_{i=1}^{N(t)}$  are their locations. For each  $(t, x) \in \overleftarrow{\Pi}$ :

- (1) for each  $\xi_i(t-) \in B(x, r)$ , independently mark the corresponding individual with probability  $u$ ;
- (2) if at least one individual is marked, all marked individuals coalesce into a single individual, whose location is chosen uniformly in  $B(x, r)$ .

If no individual is marked, then nothing happens.

One can write down a formal duality between this “backwards in time” ancestral process and the SLFV. It requires a little care because the SLFV is only defined up to a Lebesgue null set. However, informally, suppose that we know  $\{w_0(x) : x \in \mathbb{R}^d\}$  and that we would like to find the type of an individual sampled from the point  $z$  at time  $t$ . Starting the dual from a single individual with  $\xi_1(0) = z$ ,  $\xi_1(t)$  is the location of the ancestor of the sampled individual at time 0, and its type is determined by sampling according to  $w_0(\xi_1(t))$ .

Each ancestral lineage evolves in a series of jumps. By translation invariance, its distribution is determined by the rate at which an ancestral lineage jumps from 0 to  $x \in \mathbb{R}^d$ . For such a jump to occur, three things must happen: first, an event has to fall that covers both 0 and  $x$ ; second, the lineage has to be among the offspring of the event; third,  $x$  has to be chosen as the location of the parent. Writing  $L_r(x) = |B_r(0) \cap B_r(x)|$  for the volume of the region in  $\mathbb{R}^d$  of possible centers for balls of radius  $r$  that cover both 0 and  $x$ , and  $V_1$  for the volume of a unit ball in  $\mathbb{R}^d$ , we see that a single ancestral lineage evolves in a series of jumps with intensity

$$dt \otimes L_r(x) u \frac{1}{V_1 r^d} dx. \quad (6.1)$$

In particular, under our assumptions, the motion of a lineage is a spatially and temporally homogeneous continuous time random walk in  $\mathbb{R}^d$ , with uniformly bounded jumps taking place at a rate proportional to  $u$ .

Note that ancestral lineages evolve independently (only) if they are far enough apart that they cannot be covered by the same event.

### 6.1. Adding (genic) selection to the SLFV

There are many ways in which to add selection to the SLFV. Perhaps the simplest is to weight the choice of parental type during a reproduction event. For example, we might weight  $A$  alleles by a factor  $1 - s$  for some small parameter  $s$ . Mimicking our approach in Section 2, if the proportion of  $a$ -alleles in  $B(x, r)$  immediately before a reproduction event is  $w$ , then the chance of choosing a type  $a$  parent is

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

We rewrite this as

$$w^* = (1 - s)w + s(1 - (1 - w)^2) + \mathcal{O}(s^2), \quad (6.2)$$

and incorporate (weak) selection into the SLFV as follows:

**Definition 6.4** (A Spatial  $\Lambda$ -Fleming–Viot process with genic selection (SLFVGS)). Fix  $u$ ,  $r$ , and  $\Pi$  as in Definition 6.1 and  $s \in (0, 1)$ . The *Spatial  $\Lambda$ -Fleming–Viot process with genic selection (SLFVGS)* driven by  $\Pi$ , with event radius  $r$ , impact parameter  $u$ , and *selection coefficient*  $s$ , is the  $\mathcal{M}_\lambda$ -valued process  $(w_t)_{t \geq 0}$  with dynamics given as follows.

If  $(t, x) \in \Pi$ , with probability  $1 - s$ , a neutral reproduction event occurs as described in Definition 6.1. With the complementary probability  $s$  the event is *selective*, in which case:

- (1) Choose two “potential” parental locations  $z_1, z_2 \in \mathbb{R}^d$  independently and uniformly at random from  $B(x, r)$ . Sample types  $\alpha_1, \alpha_2$ , according to  $w_{t-}(z_1), w_{t-}(z_2)$ , respectively.
- (2) For every  $y \in B(x, r)$ , set  $w_t(y) = (1 - u)w_{t-}(y) + u(1 - \mathbf{1}_{\{\alpha_1 = A = \alpha_2\}})$ .

Once again we define a dual process.

**Definition 6.5** (Dual to SLFVGS). The process  $(\mathcal{P}_t)_{t \geq 0}$  is the  $\bigcup_{l \geq 1} (\mathbb{R}^d)^l$ -valued Markov process with dynamics defined as follows.

For each  $(t, x) \in \overleftarrow{\Pi}$ , the corresponding event is neutral with probability  $1 - s$ , in which case proceed as in Definition 6.3. With the complementary probability  $s$ , the event is selective, in which case:

- (1) for each  $\xi_i(t-) \in B(x, r)$ , independently mark the corresponding individual with probability  $u$ ;
- (2) if at least one individual is marked, all of the marked individuals are replaced by *two* offspring, whose locations are drawn independently and uniformly in  $B(x, r)$ .

In both cases, if no individual is marked, then nothing happens.

**Remark 6.6.** From the perspective of the SLFVGS, it would be more natural to call the individuals created during a selective event in the dual process “parents” (or “potential parents”), as they are situated at the locations from which the parental alleles are sampled. We choose to call them offspring in order to emphasize that the dual process plays the role for the SLFVGS that branching Brownian motion plays for equation (1.1).

This time, to determine the type of an individual sampled from the population at time  $t$ , construct the dual as in Definition 6.5 and assign a type to each of the individuals alive at time  $t$  by sampling (independently) according to  $w_0(\xi_i(t))$ . The individual that we sampled is of the unfavored type  $A$  if and only if all of the individuals in  $\mathcal{P}_t$  are assigned type  $A$ . If there were no coalescence, this would parallel the McKean/Skorokhod representation for the Fisher–KPP equation (with Brownian motion replaced by the random walk of ancestral

lineages); genetic drift appears as coalescence. It is natural to ask what happens if we scale the SLFVGS in such a way that the random walk followed by an ancestral lineage converges to Brownian motion.

**Theorem 6.7** (Informal restatement of [20, THEOREM 1.11]). *Consider the process of Definition 6.1. Take  $\beta, \gamma, \delta > 0$ , and let the impact and selection coefficients be  $u_n = u/n^\gamma$ , and  $s_n = s/n^\delta$  (for some positive constants  $u, s$ ). Define the scaled process  $w^{(n)}(t, x) = w(nt, n^\beta x)$ . Suppose that*

$$1 - \gamma = 2\beta, \quad \text{and} \quad 1 - \delta - \gamma = 0.$$

*Then:*

- (1) *If  $\mathfrak{d} \geq 2$  and  $\beta \leq \gamma$ , or  $\mathfrak{d} = 1$  and  $\beta < \gamma$ ,  $w^{(n)}$  converges weakly to a (weak) solution of the Fisher–KPP equation.*
- (2) *If  $\beta = \gamma = 1/3$ ,  $\delta = 2/3$ , and  $\mathfrak{d} = 1$ , as  $n \rightarrow \infty$ ,  $w^{(n)}$  converges weakly to the solution of the stochastic Fisher–KPP equation (1.2).*

This result is most easily understood through the dual process of branching and coalescing lineages. Recalling (6.1), in the scaled process that is dual to  $w^{(n)}$ , a lineage jumps a distance of  $\mathcal{O}(1/n^\beta)$  at rate proportional to  $nu_n = n^{1-\gamma}$ . To obtain a nontrivial limit, we choose  $1 - \gamma = 2\beta$ , corresponding to the diffusive scaling.

Now suppose that a selective event covers a lineage. With probability  $1/n^\gamma$  the lineage is an offspring of the event, in which case two lineages are created at separation  $\mathcal{O}(1/n^\beta)$ . They may almost immediately coalesce, but with positive probability they will move apart to a distance at which they cannot be covered by the same event. In the limit as  $n \rightarrow \infty$ , we will only “see” the branching event, if the lineages move apart to distance  $\mathcal{O}(1)$  before coalescing. By comparison with simple random walk, we expect that the number of times that they will come back to a separation less than  $2r/n^\beta$  (and so have a chance to coalesce) before reaching a separation of  $\mathcal{O}(1)$  is  $\mathcal{O}(n^\beta)$  in  $\mathfrak{d} = 1$ ,  $\mathcal{O}(\log n)$  in  $\mathfrak{d} = 2$ , and  $\mathcal{O}(1)$  in  $\mathfrak{d} \geq 3$ . Now consider how many times they come back together before they coalesce. When they are overlapped by the same event, given that one of them is an offspring, the chance that the second lineage is also an offspring, and so they coalesce, is  $\mathcal{O}(1/n^\gamma)$ , from which we deduce that they must come back together  $\mathcal{O}(n^\gamma)$  times before coalescence.

Combining the above, in  $\mathfrak{d} \geq 2$ , as  $n \rightarrow \infty$ , the chance that they escape to a separation of  $\mathcal{O}(1)$  before coalescing is  $\mathcal{O}(1)$ . Since selective events happen at rate  $ns_nu_n = \mathcal{O}(n^{1-\delta-\gamma})$ , we take  $1 - \delta - \gamma = 0$  in order that branching of ancestral lineages has rate of  $\mathcal{O}(1)$ . In  $\mathfrak{d} = 1$ , if  $\beta < \gamma$  the chance of coalescing before separating is also asymptotically negligible. In all these cases, as  $n \rightarrow \infty$  the dual process converges to a branching Brownian motion, corresponding to the forwards-in-time process converging to a weak solution to the Fisher–KPP equation. If  $\mathfrak{d} = 1$  and  $\beta = \gamma$ , which combined with our other conditions requires  $\beta = \gamma = 1/3$  and  $\delta = 2/3$ , there is a positive chance of lineages separating to  $\mathcal{O}(1)$ , but they also coalesce in finite time, reflected by the Wright–Fisher noise in (1.2).

In the argument above we took  $u_n \rightarrow 0$ , corresponding to the local population density tending to infinity. In [15, 16] scaling limits of the SLFVGS are considered in which the impact  $u$  is fixed. The diffusive scaling then requires us to set  $w^{(n)}(t, x) = w(nt, \sqrt{n}x)$ . This time, when lineages are covered by the same event, they have a strictly positive chance of coalescing. Reproducing the argument above, since lineages will coalesce after coming together only a finite number of times, most branches will rapidly be lost to coalescence. In order to see *any* lineages separate to  $\mathcal{O}(1)$  requires  $s_n = \mathcal{O}(1/\sqrt{n})$  in  $d = 1$ ,  $\mathcal{O}(\log n/n)$  in  $d = 2$ , and  $\mathcal{O}(1/n)$  in  $d \geq 3$ . In contrast to the setting of [20], the local population density remains bounded as we pass to the limit and in low dimensions we see the effect of individuals competing with their own close relatives. Recall that one motivation for taking a scaling limit is that we use neutral mutations to infer information about genetic ancestry. This result says that if local population density is bounded, if selection is to be detected, the selection coefficient must be much larger in one spatial dimension than in two, and in turn larger in two dimensions than in a population without spatial structure. In particular, when local population density is bounded, *spatial dimension is important in limiting the effect of selection*.

## 6.2. The effect of genetic drift on blocking

In order to investigate the effect of genetic drift on the blocking that we saw in Section 5.2, we adapt the SLFV to incorporate the selection mechanism of Section 5.1. There is not yet any accepted way in which to incorporate boundary conditions into the SLFV. An obvious approach that can be applied to simple domains (including for example the domain  $\Omega$  of Figure 2) based on “reflected sampling” (essentially mimicking Lord Kelvin’s method of images for the heat equation) is used in [18]. The important consequence of that choice is that scaled ancestral lineages will converge to reflected Brownian motions. For brevity we shall only describe the adaptation of the SLFV on the whole Euclidean space.

**Definition 6.8** (A Spatial  $\Lambda$ -Fleming process with (asymmetric) selection against heterozygotes (SLFVSH)). Fix  $r, u$  and  $\Pi$  as in Definition 6.1. Fix  $\gamma \in (0, 1]$  and  $s \in (0, 1/(1 + \gamma))$ . In the *Spatial  $\Lambda$ -Fleming–Viot process with selection against heterozygosity (SLFVSH)*, if  $(t, x) \in \Pi$ , with probability  $1 - (1 + \gamma)s$  a neutral reproduction event occurs as described in Definition 6.1. With the complementary probability  $(1 + \gamma)s$  the event is selective, in which case:

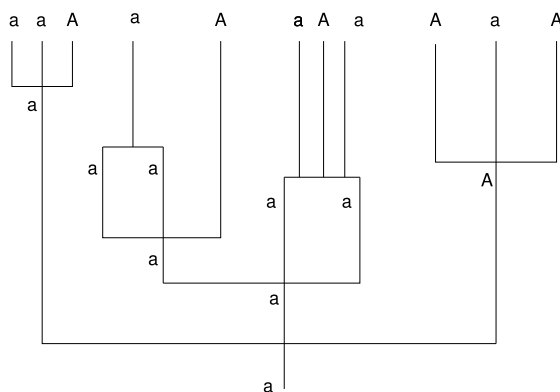
- (1) Choose *three* “potential” parental locations  $z_1, z_2, z_3 \in \mathbb{R}^d$  independently and uniformly at random from  $B(x, r)$ . Sample types  $\alpha_1, \alpha_2, \alpha_3$ , according to  $w_{t-}(z_1), w_{t-}(z_2), w_{t-}(z_3)$ , respectively. Let  $\hat{a}$  denote the most common allelic type in  $\alpha_1, \alpha_2, \alpha_3$ , except that if precisely one of  $\alpha_1, \alpha_2, \alpha_3$ , is  $a$ , with probability  $\frac{2\gamma}{3+3\gamma}$  set  $\hat{a} = a$ .
- (2) For every  $y \in B(x, r)$ , set  $w_t(y) = (1 - u)w_{t-}(y) + u1_{\{\hat{a}=a\}}$ .

The dual process mirrors the process  $(\mathcal{P}_t)_{t \geq 0}$  of Definition 6.5, except that this time, in a selective event, if at least one individual is marked then all marked individuals are

replaced by *three* offspring. Just as for the deterministic setting of Lemma 5.3, the duality relation that we exploit is between the SLFVSH and the *historical process* of branching and coalescing lineages,  $\Xi(t) := (\mathcal{P}_s)_{0 \leq s \leq t}$ , and rests on a voting scheme:

- (1) Each leaf of  $\Xi(t)$  independently votes  $a$  with probability  $p(\xi_i(t))$ , and  $A$  otherwise;
- (2) at each neutral event in  $\overleftarrow{\Pi}$ , all marked individuals adopt the vote of the offspring;
- (3) at each selective event in  $\overleftarrow{\Pi}$ , all marked individuals adopt the majority vote of the three offspring, unless precisely one vote is  $a$ , in which case they all vote  $a$  with probability  $\frac{2\gamma}{3+3\gamma}$ , otherwise they vote  $A$ .

This defines an iterative voting procedure, which runs inwards from the “leaves” of  $\Xi(t)$  to the ancestral individual  $\emptyset$  situated at the point  $x$ . The special case of majority voting, corresponding to  $\gamma = 0$ , is illustrated in Figure 4.



**FIGURE 4**

Example of majority voting on the dual to the SLFV with selection against heterozygosity. This corresponds to the duality when both homozygotes are equally fit.

**Lemma 6.9.** *With the voting procedure described above, define  $\widetilde{V}_p(\Xi(t))$  to be the vote associated to the root  $\emptyset$ . Write  $\mathbb{P}_x$  for the law of  $\Xi$  when  $\mathcal{P}_0$  is the single point  $x$ , and  $\mathbb{E}_x$  for the corresponding expectation. Then*

$$\mathbb{E}_p[w_t(x)] = \mathbb{P}_x[\widetilde{V}_p(\Xi(t)) = a].$$

To understand the influence of the genetic drift on blocking we consider two different scalings of the SLFVSH. In both cases we shall be taking a sequence  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Our results require that ancestral lineages converge to Brownian motion sufficiently quickly, compared to the rate at which  $\varepsilon_n \rightarrow 0$ , which is the purpose of the following assumption.

**Assumption 6.10.** The sequence  $\{\varepsilon_n\}_{n \in \mathbb{N}}$  is such that  $\varepsilon_n \rightarrow 0$  and  $(\log n)^{1/2} \varepsilon_n \rightarrow \infty$  as  $n \rightarrow \infty$ .



### Weak noise/selection ratio

Our first scaling is what we shall call the *weak noise/selection ratio* regime. In this regime, selection overwhelms genetic drift. It mirrors that explored in [14] and is also considered in [25]. For each  $n \in \mathbb{N}$ , and some  $\beta \in (0, 1/4)$ , set  $w^{(n)}(t, x) = w(nt, n^\beta x)$ . Let  $v > 0$ . We denote by  $u_n$  the impact parameter, and by  $s_n$  and  $\gamma_n$  the selection parameters at the  $n$ th stage of the scaling. They will be given by

$$u_n = \frac{u}{n^{1-2\beta}}, \quad s_n = \frac{1}{\varepsilon_n^2 n^{2\beta}}, \quad \gamma_n = v\varepsilon_n. \quad (6.3)$$

Adapting the proof of Theorem 1.11 in [20], and arguments in Section 3 of [14], one can show that under this scaling, for large  $n$ , the SLFVSH will be close to the solution of equation (5.5).

### Strong noise/selection ratio

We shall refer to our second scaling as the *strong noise/selection ratio* regime. In this regime, genetic drift overcomes selection. We take a sequence of impact parameters  $(u_n)_{n \in \mathbb{N}} \subseteq (0, 1)$ . Consider  $\beta \in (0, 1/2)$  and let  $\hat{u}_n := u_n n^{1-2\beta}$ . This time, we scale time by  $n/\hat{u}_n$  and space by  $n^\beta$ :  $w^{(n)}(t, x) = w(nt/\hat{u}_n, n^\beta x)$ . We consider a sequence of selection coefficients,  $(s_n)_{n \in \mathbb{N}} \subseteq (0, 1)$ , satisfying one of the following conditions:

$$\begin{cases} s_n n^{2\beta} \rightarrow 0, & \liminf_{n \rightarrow \infty} u_n \log n < \infty \text{ or } \mathfrak{d} \geq 3, \\ \frac{s_n n^{2\beta}}{u_n \log n} \rightarrow 0, & \liminf_{n \rightarrow \infty} u_n \log n = \infty \text{ and } \mathfrak{d} = 2. \end{cases} \quad (6.4)$$

The first case includes some choices of impact that were allowed in the first (weak noise/selection ratio) regime; it is the strength of drift *relative to selection* that matters. In this regime, we can take the parameters  $(\gamma_n)_{n \in \mathbb{N}}$  that dictate the asymmetry in our selection to be any sequence in  $(0, 1)$ .

**Remark 6.11.** The rationale behind these scalings is that (at least if  $u = 1$  in (6.3)) we can choose parameters in such a way that the scaled models only differ in the strength of the genetic drift (which can be thought of as the reciprocal of the impact). To see this, consider a single ancestral lineage: in the first regime, the rate at which it jumps is proportional to  $nu_n = n^{2\beta}$ ; in the second regime, it is proportional to  $nu_n/\hat{u}_n = n^{2\beta}$  (with the same constant of proportionality). In both cases we take the same spatial scaling, so the motion of ancestral lineages is the same. The rate at which a lineage “branches” as a result of being covered by a selective event in the first regime is proportional to  $nu_n s_n = n^{2\beta} s_n$ . In the second regime, it is the same,  $nu_n s_n/\hat{u}_n = n^{2\beta} s_n$ , so if we choose the same coefficients  $s_n$ , the “branching rate” is the same in both regimes. From the perspective of the dual process, the only difference between the two scalings will then be in the probability of coalescence (determined by  $u_n$ ).

**Theorem 6.12 ([18, SPECIAL CASE OF THEOREM 1.19]).** Let  $\rho_* = (\mathfrak{d} - 1)/v$  and suppose  $r_0 < \rho_*$ . Let  $(w^{(n)}(t, \cdot))_{t \geq 0}$  be the scaled SLFVSH defined above on the domain  $\Omega$  of Figure 2, with initial condition  $w^{(n)}(0, x) = \mathbf{1}_{x_1 \geq 0}$ .

- (1) *Under the weak noise/selection ratio regime, for any  $k \in \mathbb{N}$ , there exist  $n_*(k) < \infty$ , and  $a_*(k), d_*(k) \in (0, \infty)$  such that for all  $n \geq n_*$  and all  $t > 0$ ,*

$$\text{for almost every } x \text{ such that } x_1 \leq -d_*\varepsilon_n |\log \varepsilon_n|, \quad \mathbb{E}[w^{(n)}(t, x)] \leq \varepsilon_n^k.$$

- (2) *Under the strong noise/selection ratio regime, a sharp interface does not develop as  $n$  goes to infinity. Instead, there is  $\sigma^2 > 0$  such that for every  $\varepsilon > 0$  and  $t \geq 0$ , there are a reflected Brownian motion  $(W_t)_{t \geq 0}$ , and  $n_*$  such that for all  $n \geq n_*$ ,*

$$|\mathbb{E}_{w_0}[w^{(n)}(t, x)] - \mathbb{P}_x[W(\sigma^2 t) \geq 0]| \leq \varepsilon.$$

More generally, one can show that in the strong noise/selection ratio regime for  $x \neq y$ ,  $w^{(n)}(t, x)$  and  $w^{(n)}(t, y)$  decorrelate as  $n \rightarrow \infty$ .

The first statement says that in the weak noise/selection ratio regime the SLFVSH behaves approximately as the deterministic equation (5.5). The key step in the proof is to couple the dual process to a system of branching random walks in which there is no coalescence. The proof then follows the same pattern as the deterministic result with the extra twist that one must control the error arising from approximating the random walks by Brownian motions.

In the strong noise/selection ratio regime, as one can convince oneself using the argument outlined in the case of bounded neighborhood size in Section 6.1, the genetic drift is strong enough to counter the effects of selection and it breaks down the interface. We see coexistence of the populations throughout the domain. Perhaps counterintuitively, the favored type expands its range further when the population density is lower.

## 7. CONCLUSION

There is a vast body of literature that seeks to understand the interactions between natural selection, spatial structure, and genetic drift. Mathematics has provided a powerful tool; a great deal has been learned from apparently crude caricatures of the ways in which these forces interact with one another. However, as with any mathematical models, one must be cognisant of the assumptions and simplifications that are being made. In the examples presented here, we have aimed to draw out the importance of not neglecting the dimension and geometry of the domain in which a population is evolving, and of taking account of the randomness inherent in reproduction in a finite population.

## ACKNOWLEDGMENTS

My thanks to Nick Barton for introducing me to mathematical population genetics and to my graduate students, without whom very few of the results presented here would have been proved.

## REFERENCES

- [1] M. Alfaro, D. Hilhorst, and H. Matano, The singular limit of the Allen–Cahn equation and the FitzHugh–Nagumo system. *J. Differential Equations* **245** (2008), 505–565.
- [2] S. Asmussen and J. Rosiński, Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.* **38** (2001), 482–493.
- [3] N. H. Barton, The dynamics of hybrid zones. *Heredity* **43** (1979), no. 3, 341–359.
- [4] N. H. Barton, A. M. Etheridge, and A. Véber, A new model for evolution in a spatial continuum. *Electron. J. Probab.* **15** (2010), 162–216.
- [5] N. H. Barton and G. M. Hewitt, Analysis of hybrid zones. *Annu. Rev. Eol. Syst.* **16** (1985), 113–148.
- [6] N. H. Barton and G. M. Hewitt, Adaptation, speciation and hybrid zones. *Nature* **341** (1989), 497–503.
- [7] K. Becker, A. Etheridge, and I. Letter, Branching stable processes and the fractional Allen–Cahn equation. 2022, in preparation.
- [8] H. Berestycki, J. Bouhours, and G. Chapuisat, Blocking and propagation in cylinders with varying cross section. *Calc. Var. Partial Differ. Equ.* **55** (2016), no. 44.
- [9] M. Bramson, Convergence of solutions of the Kolmogorov equation to travelling waves. *Mem. Amer. Math. Soc.* **44** (1983), no. 285.
- [10] M. L. Cain, B. G. Milligan, and A. E. Strand, Long-distance seed dispersal in plant populations. *Am. J. Bot.* **87** (2000), no. 9, 1217–1227.
- [11] X. Chen, Generation and propagation of interfaces for reaction diffusion equations. *J. Differential Equations* **96** (1992), 116–141.
- [12] J. T. Cox, R. Durrett, and E. A. Perkins, Voter model perturbations and reaction diffusion equations. *Astérisque* **349** (2013).
- [13] A. de Masi, P. A. Ferrari, and J. L. Lebowitz, Reaction–diffusion equations for interacting particle systems. *J. Stat. Phys.* **44** (1986), no. 3/4, 589–644.
- [14] A. Etheridge, N. Freeman, and S. Penington, Branching Brownian motion, mean curvature flow and the motion of hybrid zones. *Electron. J. Probab.* **22** (2017), no. 103, 1–40.
- [15] A. Etheridge, N. Freeman, S. Penington, and D. Straulino, Branching Brownian motion and selection in the spatial  $\Lambda$ -Fleming–Viot process. *Ann. Appl. Probab.* **27** (2017), 2605–2645.
- [16] A. Etheridge, N. Freeman, and D. Straulino, The Brownian net and selection in the spatial  $\Lambda$ -Fleming–Viot process. *Electron. J. Probab.* **22** (2017), 1–36.
- [17] A. M. Etheridge, Drift, draft and structure: some mathematical models of evolution. *Banach Center Publ.* **80** (2008), 121–144.
- [18] A. M. Etheridge, M. D. Gooding, and I. Letter, On the effects of a wide opening in the domain of the (stochastic) Allen–Cahn equation and the motion of hybrid zones. 2022, arXiv:2204.00316.

- [19] A. M. Etheridge and T. G. Kurtz, Genealogical constructions of population models. *Ann. Probab.* **47** (2019), no. 4, 1827–1910.
- [20] A. M. Etheridge, A. Véber, and F. Yu, Rescaling limits of the spatial Lambda-Fleming–Viot process with selection. *Electron. J. Probab.* **25** (2020), no. 120, 1–89.
- [21] R. A. Fisher, *The genetical theory of natural selection*. Oxford University Press, 1930.
- [22] R. A. Fisher, The wave of advance of advantageous genes. *Annu. Eugen.* **7** (1937), 355–369.
- [23] F. Flandoli and R. Huang, The KPP equation as a scaling limit of locally interacting Brownian particles. *J. Differential Equations* **303** (2021), 608–644.
- [24] M. Gage and R. Hamilton, The heat equation shrinking convex plane curves. *J. Differential Geom.* **23** (1986), 417–491.
- [25] M. D. Gooding, *Long term behaviour of spatial population models with heterozygous or asymmetric homozygous selection*. PhD thesis, University of Oxford, 2018.
- [26] M. A. Grayson, The heat equation shrinks embedded plane curves to round points. *J. Differential Geom.* **26** (1987), 285–314–491.
- [27] M. Kimura, T. Maruyama, and J. F. Crow, The mutation load in small populations. *Genetics* **48** (1963), no. 10, 1303–1312.
- [28] A. Kolomogorov, I. Petrovsky, and N. Piscounov, Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Moscow Univ. Math. Bull.* **1** (1937), 1–25.
- [29] T. Mach, A. Sturm, and J. Swart, Recursive tree processes and the mean-field limit of stochastic flows. *Electron. J. Probab.* **25** (2020), 1–63.
- [30] H. Matano, K. I. Nakamura, and B. Lou, Periodic traveling waves in a two-dimensional cylinder with saw-toothed boundary and their homogenization limit. *Netw. Heterog. Media* **1** (2006), no. 4, 537–568.
- [31] H. P. McKean, Application of Brownian motion to the equation of Kolmogorov–Petrovski–Piskunov. *Comm. Pure Appl. Math.* **28** (1975), 323–331.
- [32] B. Merriman, J. K. Bence, and S. J. Osher, Motion of multiple junctions: A level set approach. *J. Comput. Phys.* **112** (1994), no. 2, 334–363.
- [33] C. Mueller, L. Mytnik, and L. Ryzhik, The speed of a random front for stochastic reaction–diffusion equations with strong noise. *Comm. Math. Phys.* **384** (2021), 699–732.
- [34] C. Mueller and R. Tribe, Stochastic p.d.e.'s arising from the long range contact and long range voter processes. *Probab. Theory Related Fields* **102** (1994), 519–546.
- [35] Z. O'Dowd, *Branching Brownian motion and partial differential equations*. University of Oxford MMath Dissertation, 2019.

- [36] A. V. Skorokhod, Branching diffusion processes. *Theory Probab. Appl.* **9** (1964), 492–497.

**ALISON ETHERIDGE**

Department of Statistics, University of Oxford, OX1 3LB, UK, [etheridg@stats.ox.ac.uk](mailto:etheridg@stats.ox.ac.uk)