

BRANCHING BROWNIAN MOTION AND SELECTION IN THE SPATIAL Λ -FLEMING–VIOT PROCESS

BY ALISON ETHERIDGE^{*,1}, NIC FREEMAN[†],
SARAH PENINGTON^{*,2} AND DANIEL STRAULINO^{*,3}

University of Oxford^{} and University of Sheffield[†]*

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

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

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

Since the seminal work of Fisher (1937), a large literature has developed that investigates the interaction of natural selection with the spatial structure of a pop-

Received December 2015; revised August 2016.

¹Supported in part by EPSRC Grant EP/I01361X/1.

²Supported by EPSRC DTG EP/K503113/1.

³Supported by CONACYT.

MSC2010 subject classifications. Primary 60G99; secondary 92B05.

Key words and phrases. Spatial Lambda–Fleming–Viot process, branching, coalescing, natural selection, branching Brownian motion, population genetics.

ulation. Traditionally, the deterministic action of migration and selection is approximated by what we now call the Fisher–KPP equation and predictions from that equation are compared to data. However, many important questions depend on how selection and migration interact with a third force, the stochastic fluctuations known as random genetic drift and this poses significant new mathematical challenges.

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

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

(for suitable constants m , s and γ), where \mathcal{W} is space-time white noise. This stochastic Fisher–KPP equation has been extensively studied; see, for example, [Mueller, Mytnik and Quastel \(2008\)](#) and references therein. However, from a modelling perspective, it has two immediate shortcomings. First, it only makes sense in one spatial dimension. This is generally overcome by artificially subdividing the population, and thus replacing the stochastic PDE by a system of stochastic ordinary differential equations, coupled through migration. The second problem is that, in deriving the equation, one allows the “neighbourhood size” to tend to infinity. We shall give a precise definition of neighbourhood size in Section 2. Loosely, it is inversely proportional to the probability that two individuals sampled from sufficiently close to one another had a common parent in the previous generation and small neighbourhood size corresponds to strong genetic drift. It is understanding the implications of dropping this (usually implicit) assumption of unbounded neighbourhood size that motivated the work presented here.

Our starting point will be the Spatial Λ -Fleming–Viot process with selection (SAFVS), which (along with its dual) was introduced and constructed in [Etheridge, Véber and Yu \(2014\)](#). The dynamics of both the SAFVS and its dual are driven by a Poisson Point Process of “events” (which model reproduction or extinction and recolonisation in the population) and will be described in detail in Section 2. The advantage of this model is that it circumvents the need to subdivide the population in higher dimensions. However, since our proof is based on an analysis of the branching and coalescing system of random walkers that describes the ancestry of a sample from the population, it would be straightforward to modify it to apply to, for example, an individual based model in which a fixed number of individuals reside at each point of a d -dimensional lattice.

In classical models of population genetics, in which there is no spatial structure, we generally think of population size as setting the timescale of evolution of frequencies of different genetic types. Evidently, that makes no sense in our setting. However (even in the classical setting), as we explain in more detail in Section 3, if natural selection is to leave a distinguishable trace in contemporary patterns of genetic variation, then a sufficiency of neutral mutations must fall on the genealogical trees relating individuals in a sample. Thus, in fact, it is the neutral mutation

rate which sets the timescale and, since mutation rates are very low, this leads us to consider scaling limits.

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

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

Our main result, Theorem 2.7, is that under these (dimension-dependent) scalings, the scaled dual process converges to a branching Brownian motion. For $d \geq 3$, this is rather straightforward, but in two dimensions things are much more delicate. The mathematical interest of our result is that in $d = 2$, under our scaling, the rate of branching of ancestral lineages explodes to infinity but, crucially, all except finitely many branches are instantaneously annulled through coalescence. That this finely balanced picture produces a nondegenerate limit results from a combination of the failure of two-dimensional Brownian motion to hit points and the strong (local) interactions of the approximating random walks, which cause coalescence.

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

The rest of the paper is laid out as follows. In Section 2, we describe the SAFVS and the dual process of branching and coalescing random walks, state our main result and provide a heuristic argument that explains our choice of scalings. In Section 3, we place our findings in the context of previous work on selective sweeps in spatially structured populations, and in Section 4 we prove our result.

2. The model and main result.

2.1. *The model.* To motivate the definition of the SAFVS, it is convenient to recall (a very special case of) the model without selection, introduced in Barton, Etheridge and Véber (2010), Drift (2008). We shall call it the SAFV to emphasize that selection is not acting. We proceed informally, only carefully specifying the state space and conditions that are sufficient to guarantee existence of the process when we define the SAFVS itself in Definition 2.3. The interested reader can find much more general conditions under which the SAFV exists in Etheridge and Kurtz (2014).

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

$$(2.1) \quad w_t(x) := \text{proportion of type } a \text{ at spatial position } x \text{ at time } t.$$

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

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

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

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

REMARK 2.1. In Etheridge and Kurtz (2014), the SAFV and its dual are constructed simultaneously on the same probability space, through a lockdown construction, as the limit of an individual based model, and so the dual process just described really can be interpreted as tracing the ancestry of individuals in a sample from the population.

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

DEFINITION 2.2. Write σ^2 for the variance of the first coordinate of the location of a single ancestral lineage after one unit of time and $\eta(x)$ for the instantaneous rate of coalescence of two lineages that are currently at a separation $x \in \mathbb{R}^d$. Then the neighbourhood size, \mathcal{N} , is given by

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

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

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

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

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

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

Let us now give a more precise definition of the SAFVS. We retain the notation of (2.1). A construction of an appropriate state space for $x \mapsto w_t(x)$ can be found in Véber and Wakolbinger (2015). Using the identification

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

this state space is in one-to-one correspondence with the space \mathcal{M}_λ 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}_λ .

DEFINITION 2.3 (SAFV with selection (SAFVS)). Fix $\mathcal{R} \in (0, \infty)$. Let μ be a finite measure on $(0, \mathcal{R}]$ and, for each $r \in (0, \mathcal{R}]$, let ν_r be a probability measure on $(0, 1]$. Further, let Π be a Poisson point process on $\mathbb{R} \times \mathbb{R}^d \times (0, \mathcal{R}] \times (0, 1]$ with intensity measure

$$(2.2) \quad dt \otimes dx \otimes \mu(dr) \nu_r(du).$$

The *spatial Λ -Fleming–Viot process with selection* (SAFVS) driven by (2.2) is the \mathcal{M}_λ -valued process $(w_t)_{t \in \mathbb{R}}$ with dynamics given as follows.

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

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

With the complementary probability s the event is *selective*, in which case:

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

This is a very special case of the SAFVS introduced in Etheridge, Véber and Yu (2014).

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

As in the neutral case, the dynamics of the dual are driven by the same Poisson point process of events, Π , that drove the forwards in time process. The distribution

of this Poisson point process is invariant under time reversal and so we shall abuse notation by reversing the direction of time when discussing the dual.

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

DEFINITION 2.4 (Branching and coalescing dual). The branching and coalescing dual process $(\Xi_t)_{t \geq 0}$ driven by Π is the $\bigcup_{m \geq 1} (\mathbb{R}^d)^m$ -valued Markov process with dynamics defined as follows: at each event $(t, x, r, u) \in \Pi$, with probability $1 - s$, the event is neutral:

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

With the complementary probability s , the event is selective:

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

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

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

This dual process is the analogue for the SAFVS of the Ancestral Selection Graph (ASG), introduced in the companion papers Krone and Neuhauser (1997), Neuhauser and Krone (1997), which describes all the potential ancestors of a sample from a population evolving according to the Wright–Fisher diffusion with selection. Perhaps the simplest way of expressing the duality between the SAFVS and the branching and coalescing dual process is to observe that all the individuals in our sample are of type a if and only if all potential ancestral lineages are of type a at any time t in the past. This is analogous to the *moment duality* between the ASG and the Wright–Fisher diffusion with selection. However, to state this formally for the SAFVS, we would need to be able to identify $\mathbb{E}[\prod_{i=1}^n w_t(x_i)]$ for any choice of points $x_1, \dots, x_n \in \mathbb{R}^d$. The difficulty is that, just as in the neutral case, the SAFVS $w_t(x)$ is only defined at Lebesgue almost every point x and so we have to be satisfied with a “weak” moment duality.

PROPOSITION 2.5. [Etheridge, Véber and Yu (2014)] *The spatial Λ -Fleming–Viot process with selection is dual to the process $(\Xi_t)_{t \geq 0}$ in the sense that for every $k \in \mathbb{N}$ and $\psi \in C_c((\mathbb{R}^d)^k)$, we have*

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

2.2. *The main result.* Our main result concerns a diffusive rescaling of the dual process of Definition 2.4 and so from now on it will be convenient if

forwards in time refers to forwards for the dual process.

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

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

$$(2.4) \quad n dt \otimes n^{d/2} dx \otimes \mu^n(dr).$$

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

$$(2.5) \quad s_n = \begin{cases} \frac{\log n}{n} & d = 2, \\ \frac{1}{n} & d \geq 3. \end{cases}$$

In Etheridge, Freeman and Straulino (2015), it was shown that in $d = 1$, one should take $s_n = 1/\sqrt{n}$.

Although not obvious for the SAFVS itself, when considering the dual process it is not hard to understand why the scalings (2.4) and (2.5) should lead to a nontrivial limit.

If we ignore the selective events, then a single ancestral lineage evolves as a pure jump process which is homogeneous in both space and time. Write V_r for

the volume of $\mathcal{B}_r(0)$. The rate at which the lineage jumps from y to $y + z$ can be written

$$(2.6) \quad m_n(dz) = nu \int_0^{\mathcal{R}_n} n^{d/2} \frac{V_r(0, z)}{V_r} \mu^n(dr) dz,$$

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 $nn^{d/2} V_r(0, z) \mu^n(dr)$. The parental location is chosen uniformly from the ball $\mathcal{B}_r(x)$, so the probability that z is chosen as the parental location is dz/V_r and the probability that our lineage is actually affected by the event is u . Combining these yields (2.6).

The total rate of jumps is

$$(2.7) \quad \begin{aligned} \int_{\mathbb{R}^d} m_n(dz) &= \int_0^{\mathcal{R}_n} nun^{d/2} \frac{1}{V_r} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_{|x|<r} \mathbb{1}_{|x-z|<r} dx dz \mu^n(dr) \\ &= \int_0^{\mathcal{R}_n} nun^{d/2} V_r \mu^n(dr) \\ &= nuV_1 \int_0^{\mathcal{R}} r^d \mu(dr) = \Theta(n), \end{aligned}$$

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

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

REMARK 2.6. Our scaling mirrors that described in Durrett and Zähle (2007) for a model of a *hybrid zone* (by which we mean a region in which we see both genetic types) which develops around a boundary between two regions, in one of which type a individuals are selectively favoured and in the other of which type A individuals are selectively favoured. In contrast to our continuum setting, their model is a spin system in which exactly one individual lives at each point of \mathbb{Z}^d .

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

$$(2.8) \quad \sigma^2 = \frac{1}{d} \int_{\mathbb{R}^d} |z|^2 m^n(dz) = \frac{1}{d} \int_{\mathbb{R}^d} \int_0^\infty |z|^2 u \frac{V_r(0, z)}{V_r} \mu(dr) dz,$$

where $m^n(dz)$ is defined in (2.6). In other words, during their lifetime, which is exponentially distributed with parameter V , individuals follow d -dimensional Brownian motion with diffusion constant σ^2 , at the end of which they die, leaving behind at the location where they died exactly two offspring. We view $\text{BBM}(p, V)$ as a set of (continuous) paths, each starting at p , with precisely one path following each possible distinct sequence of branches.

Similarly, we write $\mathcal{P}^{(n)}(p)$ for the dual process of Definition 2.4, rescaled as in (2.4) and (2.5), started from a single individual at the point $p \in \mathbb{R}^d$ and viewed as a collection of paths. Each path traces out a “potential ancestral lineage”, defined exactly as the ancestral lineages in the neutral case except that at each selective event, if a lineage is affected then it jumps to the location of (either) one of the “potential parents”. Precisely one potential ancestral lineage follows each possible route through the branching and coalescing dual process.

We define the events:

$$(2.9) \quad \begin{aligned} \mathcal{D}_n(\varepsilon, T) &= \left\{ \forall l \in \mathcal{P}^{(n)}(p), \exists l' \in \text{BBM}(p, V) : \sup_{t \in [0, T]} |l(t) - l'(t)| \leq \varepsilon \right\}, \\ \mathcal{D}'_n(\varepsilon, T) &= \left\{ \forall l \in \text{BBM}(p, V), \exists l' \in \mathcal{P}^{(n)}(p) : \sup_{t \in [0, T]} |l(t) - l'(t)| \leq \varepsilon \right\}. \end{aligned}$$

THEOREM 2.7. *Let $d \geq 2$. There exists $V \in (0, \infty)$ such that the following holds. Let $T < \infty, p \in \mathbb{R}^2$; then given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that, for all $n \geq N$ there is a coupling between $\text{BBM}(p, V)$ and $\mathcal{P}^{(n)}(p)$ with $\mathbb{P}[\mathcal{D}_n(\varepsilon, T) \cap \mathcal{D}'_n(\varepsilon, T)] \geq 1 - \varepsilon$.*

We will give a proof of Theorem 2.7 only for $d = 2$. The case $d \geq 3$ follows from a simplified version of the 2-dimensional proof presented here.

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

We choose a large constant $c > 0$. We begin by showing that $\xi^{n,1}$ and $\xi^{n,2}$ have probability $\Theta(1/\log n)$ of reaching a distance $1/(\log n)^c$ from each other without

coalescing (we then say they have diverged). We also show that the probability that $\xi^{n,1}$ and $\xi^{n,2}$ have not diverged or coalesced by time $1/(\log n)^c$ is $o(1/(\log n))$, so coalescence will be instantaneous in the limit. Moreover, once they are $1/(\log n)^c$ apart, they won't get within distance $2\mathcal{R}_n$ of each other again on a timescale of $\mathcal{O}(1)$. Hence, from the point of view of our scaling they stay apart and evolve essentially independently of each other.

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

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

This programme is carried out in Section 4.

3. Biological background. In this section, we shall set our work in the context of the substantial biological literature. The reader concerned only with the mathematics can safely skip to Section 4.

The interplay between natural selection and the spatial structure of a population is a question of longstanding interest in population genetics. Fisher (1937) studied the advance of selectively advantageous genetic types through a one-dimensional population using the deterministic differential equation now known as the Fisher-KPP equation. This equation also makes sense in higher dimensions, but ignores *genetic drift* (the randomness due to reproduction in a finite population). Work incorporating genetic drift has been restricted to either one spatial dimension [see Barton et al. (2013a) and references therein] or, more commonly, to subdivided populations. Maruyama (1970) studied the probability of *fixation* of an advantageous genetic type (the probability that eventually the whole population carries this genetic type) in a subdivided population. The assumptions made in that article are rather strong: if we think of the population as living on islands (or in colonies), then each island has constant total population size and its contribution to the next generation is in proportion to that size. Under these assumptions, the probability of fixation is not affected by the population structure: it is the same as for a gene

of the same selective advantage in an unstructured population of the same total size. Much subsequent work retained Maruyama's assumptions, and so it is often assumed that spatial structure has no influence on the accumulation of favourable genes. However, Barton (1993) showed that the extra stochasticity produced by the Introduction of local extinctions and colonisations could significantly change the fixation probability. This work was extended in, for example, Cherry (2003) and Whitlock (2003).

A fundamental problem in genetics is to identify which parts of the genome have been the target of natural selection. The random nature of reproduction in finite populations means that some genetic types (alleles) will be carried by everyone in the population, even though they convey no particular selective advantage. However, if a favourable mutation arises in a population and "sweeps" to fixation (i.e., increases in frequency until everybody carries it), we expect the genealogical trees (i.e., the trees of ancestral lineages) relating individuals in a sample from the population to differ from those that we observe in the absence of selection. In particular, they will be more "star-shaped". Of course, we cannot observe the genealogical trees directly, and so, instead, geneticists exploit the fact that genes are arranged on chromosomes: the ancestry at another position on the same chromosome will be correlated with that at the part of the genome that is the target of selection. In order to detect selection, one therefore examines the patterns of variation at other points on the same chromosome, so-called linked loci.

In order for this approach to work, we require sufficient variability at the linked loci that we see a signal of the distortion in the genealogical tree. This means that we must consider the genealogy of a sample from the population on the timescale set by the neutral mutation rate. If selection is too strong, the genealogy will be very short and we see no mutations and so we can recover no information about the genealogical trees; if selection is too weak, we will not be able to distinguish the patterns from those seen under neutral evolution.

Since neutral mutation rates are rather small, this means that we are interested in long timescales. Without selection, ancestral lineages in our model follow symmetric random walks with bounded variance jumps and so we expect a diffusive scaling to capture patterns of neutral variation. Since we are looking for deviations from those patterns due to the action of selection, it makes sense to consider a diffusive rescaling in the selective case also. Thus, if the neutral mutation rate is μ , then we look at the rescaled dual process with $n = 1/\mu$. If the branches produced by selection persist long enough to be visible at this scale, then there is positive probability that the pattern of (neutral) variation we see in a sample from the population will look different from the pattern we would expect without selection.

Our results in this paper are relevant to populations evolving in spatial continua. The question they address is "When can we hope to detect a signal of natural selection in data?" Whereas in the classical models of subdivided populations, it

is typically assumed that the population in each “island” is large, so that neighbourhood size is big, by fixing the “impact” parameter u in our model, we are assuming that neighbourhood size is small. As a result, reproduction events are somewhat akin to local extinction and recolonisation events, in which a significant proportion of the local population is replaced in a single event. Our main result shows that our ability to detect selection is then critically dependent on spatial dimension. For populations living in at least three spatial dimensions (of which there are very few), spatial structure has a rather weak effect. However, in two spatial dimensions, selection must be stronger and in one spatial dimension (as appropriate, e.g., for populations living in intertidal zones) much stronger, before we can expect to be able to detect it. The explanation is that in low dimensions, it is harder for individuals carrying the favoured gene to escape the competition posed by close relatives who carry the same gene. In our mathematical work, this is reflected in the vast majority of branches in our dual process being cancelled by a coalescence event on a timescale which is negligible compared to the timescale set by the neutral mutation rate so that no evidence of these branches having occurred will be seen in the pattern of neutral mutation.

4. Proof of Theorem 2.7. Our proof is broken into two steps. First, in Section 4.1 we consider how the pair of potential ancestral lineages created during a selective event interact with each other. In particular, we find asymptotics for the probability that they diverge in a short time. This will allow us to identify the branching rate in the limiting Brownian motion. Then in Section 4.2, we define the caterpillar and show how to couple the dual of the SAFVS to a system of branching caterpillars. With this construction in hand, Theorem 2.7 follows easily.

4.1. *Pairs of paths.* In this subsection, we are interested in the behaviour of a pair of potential ancestral lineages in the rescaled dual. In order that they be uniquely defined, if either is hit by a selective event then we (arbitrarily) declare that it jumps to the location of the first potential parent sampled in that event. In particular, if they are both affected by the same event, then they will necessarily coalesce. We write $\xi^{n,1}$ and $\xi^{n,2}$ for the resulting potential ancestral lineages and

$$\eta^n = \xi^{n,1} - \xi^{n,2}$$

for their separation.

Throughout this subsection, we use the notation $\mathbb{P}_{[r,r']}$ to mean that $|\eta_0^n| \in [r, r']$ and we adopt the convention that estimates of $\mathbb{P}_{[r,r']}[B]$ hold uniformly for all initial laws with mass concentrated on $[r, r']$. We extend this notation to open intervals in the obvious manner. We will also write $\mathbb{P}_r = \mathbb{P}_{[r,r]}$.

We are concerned with the behaviour of two potential ancestral lineages created during a selective event which, without loss of generality, we suppose to happen at

time 0. We shall then refer to η^n as an excursion. In this case $|\eta_0^n| \leq 2\mathcal{R}_n$ and we wish to establish whether or not $|\eta_t^n|$ ever exceeds

$$(4.1) \quad \gamma_n = \frac{1}{(\log n)^c},$$

where, in this section, we suppose that $c \geq 3$.

REMARK 4.1. We will, eventually, set $c = 4$, although any larger constant c would give the same result; for now we keep the dependence on c visible in our estimates.

For reasons that will soon become apparent, it is convenient to assume that n is large enough that $7\mathcal{R}_n < \gamma_n$.

The picture of an excursion η^n that we would like to build up is, loosely speaking, as follows:

1. With probability $\kappa_n = \Theta(\frac{1}{\log n})$, $|\eta^n|$ reaches displacement γ_n within time $1/(\log n)^c$ and then $\xi^{n,1}$ and $\xi^{n,2}$ will not interact again before a fixed time $T > 0$. Consequently, the displacement between them becomes macroscopic and we see two distinct paths in the limit. Moreover, $\kappa_n \log n \rightarrow \kappa \in (0, \infty)$ as $n \rightarrow \infty$.
2. With probability $1 - \Theta(\frac{1}{\log n})$, $|\eta^n|$ does not reach displacement γ_n , and $\xi^{n,1}$ and $\xi^{n,2}$ coalesce within time $1/(\log n)^c$. In this case, the difference between them is microscopic and we see only one path in the limit.
3. All other outcomes have probability $\mathcal{O}(\frac{1}{(\log n)^{c-3/2}})$, which means that we will not see them in the limit.

Much of the work in making this rigorous results from the fact that $\xi^{n,1}, \xi^{n,2}$ only evolve independently when their separation is greater than $2\mathcal{R}_n$. Our strategy is similar to that in the proof of Lemma 4.2 in Etheridge and Véber (2012), but here we require a stronger result: rather than an estimate of the form $\kappa_n \geq C/\log n$ we need convergence of $\kappa_n \log n$.

4.1.1. *Inner and outer excursions.* We shall characterise the behaviour of η^n using several stopping times. Set $\tau_0^{\text{out}} = 0$ and define inductively, for $i \geq 0$,

$$(4.2) \quad \begin{aligned} \tau_i^{\text{in}} &= \inf\{s > \tau_i^{\text{out}} : |\eta_s^n| \geq 5\mathcal{R}_n\}, \\ \tau_{i+1}^{\text{out}} &= \inf\{s > \tau_i^{\text{in}} : |\eta_s^n| \leq 4\mathcal{R}_n\}. \end{aligned}$$

We refer to the interval $[\tau_i^{\text{out}}, \tau_i^{\text{in}})$ (and also to the path of η^n during it) as the i th inner excursion and similar to $[\tau_{i-1}^{\text{in}}, \tau_i^{\text{out}})$ (and corresponding path) as the i th outer excursion.

Since a jump of η^n has displacement at most $2\mathcal{R}_n$, although the initial (0th) inner excursion starts in $(0, 2\mathcal{R}_n]$, for $i \geq 1$ we have $|\eta_{\tau_i^{\text{in}}}^n| \in [5\mathcal{R}_n, 7\mathcal{R}_n]$ and $|\eta_{\tau_i^{\text{out}}}^n| \in [2\mathcal{R}_n, 4\mathcal{R}_n]$.

DEFINITION 4.2. We define the stopping times

$$\begin{aligned} \tau^{\text{coal}} &= \inf\{s > 0 : |\eta_s^n| = 0\}, \\ \tau^{\text{div}} &= \inf\{s > 0 : |\eta_s^n| \geq \gamma_n\}, \\ \tau^{\text{over}} &= \frac{1}{(\log n)^c}. \end{aligned}$$

We shall say that the i th inner excursion coalesces if $\tau^{\text{coal}} \in [\tau_i^{\text{out}}, \tau_i^{\text{in}})$. Similarly, the i th outer excursion diverges if $\tau^{\text{div}} \in [\tau_{i-1}^{\text{in}}, \tau_i^{\text{out}})$.

We define $\tau^{\text{type}} = \min(\tau^{\text{coal}}, \tau^{\text{div}}, \tau^{\text{over}})$ and say that η^n :

1. coalesces if $\tau^{\text{type}} = \tau^{\text{coal}}$,
2. diverges if $\tau^{\text{type}} = \tau^{\text{div}}$,
3. overshoots if $\tau^{\text{type}} = \tau^{\text{over}}$.

Since almost surely η^n only jumps a finite number of times before time $(\log n)^{-c}$, almost surely τ^{type} occurs during either an inner or an outer excursion, whose index we denote by i^* .

We use ζ^n to denote the distribution of the distance between the two potential parents sampled during a selective event.

LEMMA 4.3. *There exists $\alpha \in (0, 1)$ such that, uniformly in n , $\mathbb{P}_{\zeta^n}[i^* > m] \leq \alpha^m$.*

LEMMA 4.4. *As $n \rightarrow \infty$, $\mathbb{P}_{\zeta^n}[\eta^n \text{ overshoots}] = \mathcal{O}(\frac{1}{(\log n)^{c-3/2}})$.*

LEMMA 4.5. *As $n \rightarrow \infty$, $\mathbb{P}_{\zeta^n}[\eta^n \text{ diverges}] = \Theta(\frac{1}{\log n})$.*

LEMMA 4.6. *As $n \rightarrow \infty$, $\mathbb{P}_{\zeta^n}[\eta^n \text{ coalesces}] = 1 - \Theta(\frac{1}{\log n})$.*

Thus, overshoots are relatively unlikely, and typically η^n consists of a finite number of inner/outer excursions until either (1) it coalesces, with probability $1 - \Theta(\frac{1}{\log n})$, or (2) the two lineages separate to distance γ_n , with probability $\Theta(\frac{1}{\log n})$.

The remainder of this Section 4.1.1 is devoted to the proof of Lemmas 4.3–4.5. Lemma 4.6 then follows immediately, since $c \geq 3$.

We will need two more stopping times:

$$(4.3) \quad \begin{aligned} \tau_r &= \inf\{s > 0 : |\eta_s^n| \leq r\}, \\ \tau^r &= \inf\{s > 0 : |\eta_s^n| \geq r\}. \end{aligned}$$

Note that $\tau_0 = \tau^{\text{coal}}$.

Note that the random variables τ^{type} , τ^r and so on depend implicitly on n ; throughout this section these random variables refer to the stopping times for the process η^n .

PROOF OF LEMMA 4.3. First, consider a single inner excursion of η^n . It is easily seen that there exists some $\alpha' > 0$ such that, for all n :

(\dagger) For any $x \in (0, 5\mathcal{R}_n)$, if $|\eta^n_t| = x$ then the probability that η^n will hit 0 but not exit $\mathcal{B}_{5\mathcal{R}_n}(0)$ within its next three jumps is at least α' .

In particular, the probability that the first three jumps of an inner excursion result in a coalescence is bounded away from 0 uniformly for any $|\eta^n_{\tau_i^{\text{out}}}| \in [2\mathcal{R}_n, 4\mathcal{R}_n]$. If $i^* > m$, then at least m inner excursions must occur without a coalescence. The strong Markov property applied at the time τ_i^{out} means that, conditionally given $\eta^n_{\tau_i^{\text{out}}}$, the i th inner excursion is independent of $(\eta^n_t)_{t < \tau_i^{\text{out}}}$. Repeated application of this fact, coupled with (\dagger), shows that the probability of seeing at least m inner excursions without a single coalescence is at most $(1 - \alpha')^m$. This completes the proof. \square

We will shortly require a tail estimate on the supremum of the modulus of two-dimensional Brownian motion W , which we record first for clarity. We write $W_t = (W_t^1, W_t^2)$ and note

$$\begin{aligned}
 \mathbb{P}\left[\sup_{s \in [0, t]} |W_s - W_0| \geq x\right] &\leq 2\mathbb{P}\left[\sup_{s \in [0, t]} |W_s^1 - W_0^1| \geq x/2\right] \\
 (4.4) \qquad \qquad \qquad &\leq 4\mathbb{P}\left[\sup_{s \in [0, t]} (W_s^1 - W_0^1) \geq x/2\right] \\
 &\leq 4e^{-x^2/8t}.
 \end{aligned}$$

In the first line of the above, we use the triangle inequality and the fact that W^1 and W^2 have the same distribution. To deduce the second line, we note that W^1 and $-W^1$ have the same distribution. For the final line, we use the (standard) tail estimate $\mathbb{P}[\sup_{s \in [0, t]} (B_s - B_0) \geq x] \leq e^{-x^2/2t}$ for a one-dimensional Brownian motion B , which can be deduced via Doob’s martingale inequality applied to the submartingale $(\exp(xB_s/t))_{s \geq 0}$.

During an outer excursion, η^n is the difference between two independent walkers and so we can use Skorohod embedding to approximate its behaviour using elementary calculations for two-dimensional Brownian motion. The next lemma exploits this to bound the duration of the outer excursion and the probability that it diverges.

LEMMA 4.7. As $n \rightarrow \infty$,

$$(4.5) \qquad \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[\tau^{\gamma_n} \wedge \tau_{4\mathcal{R}_n} > (\log n)^{-c-1}] = \mathcal{O}\left(\frac{1}{(\log n)^{c-1}}\right)$$

and

$$(4.6) \quad \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[\tau^{\gamma_n} < \tau_{4\mathcal{R}_n}] = \Theta\left(\frac{1}{\log n}\right).$$

PROOF. For $i = 1, 2$, let $\hat{\xi}^{n,i}$ be a pair of independent processes such that $\hat{\xi}^{n,1}$ has the same distribution as $\xi^{n,1}$ and $\hat{\xi}^{n,2}$ has the same distribution as $\xi^{n,2}$. The process $\hat{\xi}^{n,1} - \hat{\xi}^{n,2}$ is a compound Poisson process with a rotationally symmetric jump distribution and a maximum displacement of $2\mathcal{R}_n$ on each jump. Moreover [essentially by Skorohod’s embedding theorem see, e.g., Billingsley (1995)], we can construct a process $\hat{\eta}^n$ with the same distribution as $\hat{\xi}^{n,1} - \hat{\xi}^{n,2}$ as follows.

Let $(r_m, J_m)_{m \geq 1}$ denote a sequence distributed as the jump magnitudes and jump times of $\hat{\xi}^{n,1} - \hat{\xi}^{n,2}$. Let W be a two-dimensional Brownian motion with $W_0 = \hat{\xi}_0^{n,1} - \hat{\xi}_0^{n,2}$, independent of $(r_m, J_m)_{m \geq 1}$. Now set

$$(4.7) \quad \begin{aligned} \hat{\eta}_t^n &= W_{T(s(t))} \quad \text{where } T^{(0)} = 0, J_0 = 0, \\ T^{(m+1)} &= \inf\{s > T^{(m)} : |W_s - W_{T^{(m)}}| \geq r_m\}, \\ S(t) &= \sup\{i \geq 0 : J_i \leq t\}. \end{aligned}$$

We may then couple

$$\hat{\eta}^n = \hat{\xi}^{n,1} - \hat{\xi}^{n,2}.$$

We define $\hat{\tau}^r$ and $\hat{\tau}_r$ analogously to τ^r and τ_r , as stopping times of the process $\hat{\eta}^n$.

Note that since $(\xi_t^{n,1}, \xi_t^{n,2})_{t \leq \tau_{4\mathcal{R}_n}}$ has the same distribution as $(\hat{\xi}_t^{n,1}, \hat{\xi}_t^{n,2})_{t \leq \tau_{4\mathcal{R}_n}}$, we may couple them so that they are almost surely equal during this time. Thus,

$$\{\hat{\tau}^{\gamma_n} < \hat{\tau}_{4\mathcal{R}_n}\} = \{\tau^{\gamma_n} < \tau_{4\mathcal{R}_n}\}.$$

Let T^r and T_r be the analogues of τ^r and τ_r for W [not to be confused with $T^{(m)}$ in (4.7)]. By the definition of the Skorohod embedding in (4.7), we have

$$(4.8) \quad \begin{aligned} \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[\hat{\tau}^{\gamma_n} < \hat{\tau}_{4\mathcal{R}_n}] &\geq \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[T^{\gamma_n+2\mathcal{R}_n} < T_{4\mathcal{R}_n}] \\ &\geq \mathbb{P}_{5\mathcal{R}_n}[T^{\gamma_n+2\mathcal{R}_n} < T_{4\mathcal{R}_n}]. \end{aligned}$$

The right-hand side concerns only the modulus of two-dimensional Brownian motion and so can be expressed in terms of the scale function for a two-dimensional Bessel process:

$$(4.9) \quad \begin{aligned} \mathbb{P}_{5\mathcal{R}_n}[T^{\gamma_n+2\mathcal{R}_n} < T_{4\mathcal{R}_n}] &= \frac{\log(5\mathcal{R}_n) - \log(4\mathcal{R}_n)}{\log(\gamma_n + 2\mathcal{R}_n) - \log(4\mathcal{R}_n)} \\ &= \Theta\left(\frac{1}{\log n}\right), \end{aligned}$$

which proves the lower bound in (4.6). Similar to the upper bound, we note that

$$\begin{aligned} \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[\hat{\tau}^{\gamma_n} < \hat{\tau}_{4\mathcal{R}_n}] &\leq \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[T^{\gamma_n} < T_{2\mathcal{R}_n}] \\ &\leq \mathbb{P}_{7\mathcal{R}_n}[T^{\gamma_n} < T_{2\mathcal{R}_n}] \\ &= \frac{\log(7\mathcal{R}_n) - \log(2\mathcal{R}_n)}{\log(\gamma_n) - \log(2\mathcal{R}_n)} \\ &= \Theta\left(\frac{1}{\log n}\right). \end{aligned}$$

It remains to prove (4.5). We have

$$\tau^{\gamma_n} \wedge \tau_{4\mathcal{R}_n} = \hat{\tau}^{\gamma_n} \wedge \hat{\tau}_{4\mathcal{R}_n} \leq \hat{\tau}^{\gamma_n}.$$

REMARK 4.8. The above inequality is a very crude estimate, but will be enough to prove (4.5), which in turn will be enough to give useful bounds on the duration of excursions due to the freedom in the choice of c .

Hence,

$$(4.10) \quad \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[\tau^{\gamma_n} \wedge \tau_{4\mathcal{R}_n} > (\log n)^{-c-1}] \leq \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[|\hat{\eta}_{(\log n)^{-c-1}}^n| \leq \gamma_n].$$

The remainder of the proof focuses on bounding the right-hand side of (4.10). To do so, we must relate our compound Poisson process to another Brownian motion.

For $j \geq 1$, let $X_j = \hat{\eta}_{j/n}^n - \hat{\eta}_{(j-1)/n}^n$. Then $(X_j)_{j \geq 1}$ are i.i.d. and since $\hat{\xi}^{n,1}$ and $\hat{\xi}^{n,2}$ are independent, $\mathbb{E}[|X_1|^2] = 2\mathbb{E}[|\hat{\xi}_{1/n}^{n,1} - \hat{\xi}_0^{n,1}|^2]$.

Recall from (2.6) that the rate at which $\hat{\xi}^{n,1}$ jumps from y to $y + z$ is determined by the intensity measure $m^n(dz)$ so that

$$(4.11) \quad \mathbb{E}[|X_1|^2] = \frac{2}{n} \int_{\mathbb{R}^2} |z|^2 m^n(dz) = \frac{4\sigma^2}{n},$$

where σ^2 was defined in (2.8). Now recall the definition of $S(t)$ in (4.7); the rate at which $\hat{\xi}^{n,1}$ jumps is $\int_{\mathbb{R}^2} m^n(z) dz = \Theta(n)$ by (2.7), so $S(n^{-1})$ is bounded by the sum of two $\text{Poisson}(\Theta(1))$ random variables. Hence, since each jump of $\hat{\eta}^n$ is bounded by $2\mathcal{R}_n$,

$$(4.12) \quad \mathbb{E}[|X_1|^4] \leq (2\mathcal{R}_n)^4 \mathbb{E}[S(n^{-1})^4] = \mathcal{O}(n^{-2}).$$

Once again (since the distribution of X_1 is rotationally symmetric), we may use Skorohod’s embedding theorem to couple $(X_i)_{i \geq 1}$ to a two-dimensional Brownian motion B started at η_0^n and a sequence ν_1, ν_2, \dots of stopping times for B such that

setting $v_0 = 0$, $(v_i - v_{i-1})_{i \geq 1}$ are i.i.d. and

$$\begin{aligned}
 & B_{v_i} - B_{v_{i-1}} = X_i, \\
 (4.13) \quad & \mathbb{E}[v_i - v_{i-1}] = \frac{1}{2} \mathbb{E}[|X_1|^2] = \frac{2\sigma^2}{n} \\
 & \text{and } \mathbb{E}[(v_i - v_{i-1})^2] = \mathcal{O}(n^{-2}).
 \end{aligned}$$

It follows that $\mathbb{E}[v_{\lfloor tn \rfloor}] = \frac{2\sigma^2 \lfloor tn \rfloor}{n}$ and $\text{Var}(v_{\lfloor tn \rfloor}) = \mathcal{O}(tn^{-1})$. Hence, by Chebyshev's inequality,

$$\mathbb{P}[|v_{\lfloor tn \rfloor} - 2\sigma^2 t| \geq n^{-1/3}] \leq \mathcal{O}(tn^{-1/3}).$$

Applying this result with $t = t_n := (\log n)^{-c-1}$, since $\hat{\eta}_{\lfloor tn n \rfloor / n}^n = B_{v_{\lfloor tn n \rfloor}}$ we have

$$\begin{aligned}
 (4.14) \quad & \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[|\hat{\eta}_{t_n}^n| \leq \gamma_n] \\
 & \leq \mathbb{P}[\inf\{|B_t - B_0| : t \in [2\sigma^2 t_n - n^{-1/3}, 2\sigma^2 t_n + n^{-1/3}]\}] \\
 (4.15) \quad & \leq \gamma_n + n^{-1/8} + 7\mathcal{R}_n + \mathbb{P}[|\hat{\eta}_{t_n}^n - \hat{\eta}_{\lfloor tn n \rfloor / n}^n| \geq n^{-1/8}] + \mathcal{O}(t_n n^{-1/3}).
 \end{aligned}$$

For the first term on the right-hand side, we have for n sufficiently large

$$\begin{aligned}
 & \mathbb{P}[\inf\{|B_t - B_0| : t \in [2\sigma^2 t_n - n^{-1/3}, 2\sigma^2 t_n + n^{-1/3}]\}] \leq \gamma_n + n^{-1/8} + 7\mathcal{R}_n \\
 & \leq \mathbb{P}[|B_{2\sigma^2 t_n} - B_0| \leq \gamma_n + 3n^{-1/8}] + \mathbb{P}\left[\sup_{t \in [0, 2n^{-1/3}]} |B_t - B_0| \geq \frac{1}{2} n^{-1/8}\right] \\
 (4.16) \quad & = \mathcal{O}(\gamma_n^2 t_n^{-1}) + \mathcal{O}(e^{-\frac{1}{64} n^{1/12}}) \\
 & = \mathcal{O}((\log n)^{1-c}).
 \end{aligned}$$

For the second inequality, we use that the density of B_t is bounded by $(2\pi t)^{-1}$ for the first term and we apply (4.4) for the second term.

Moving on to the second term on the right-hand side of (4.14), since from (4.11) we have $\mathbb{E}[|\hat{\eta}_{t_n}^n - \hat{\eta}_{\lfloor tn n \rfloor / n}^n|^2] = \mathcal{O}(n^{-1})$, by Markov's inequality

$$(4.17) \quad \mathbb{P}[|\hat{\eta}_{t_n}^n - \hat{\eta}_{\lfloor tn n \rfloor / n}^n| \geq n^{-1/8}] = \mathcal{O}(n^{-3/4}).$$

Putting (4.16) and (4.17) into (4.14), we have

$$\mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[|\hat{\eta}_{t_n}^n| \leq \gamma_n] = \mathcal{O}((\log n)^{1-c}).$$

In view of (4.10), this completes the proof. \square

PROOF OF LEMMA 4.4. First, consider a single inner excursion. Evidently, there exists $\beta > 0$ such that, for all n :

(\ddagger) For any $x \in (0, 5\mathcal{R}_n)$, if $|\eta_t^n| = x$ then the probability that η^n will either exit $\mathcal{B}_{5\mathcal{R}_n}(0)$ or hit 0 within its next three jumps is at least β .

Let $(J_l)_{l \geq 0}$ be the (a.s. finite) sequence of jump times of our inner excursion, and let B_k be the event that the excursion either coalesces or exits $\mathcal{B}_{5\mathcal{R}_n}(0)$ at one of $\{J_{3k+1}, J_{3k+2}, J_{3k+3}\}$. By the strong Markov property (applied at J_{3k}) and (\ddagger) , $\inf\{k \geq 0 : \mathbb{1}_{B_k} = 1\}$ is stochastically bounded above by a geometric random variable G with success probability β .

Moreover, for as long as η^n is not at 0, the rate at which it jumps is bounded below by the rate at which $\xi^{n,1}$ jumps, which is $\int_{\mathbb{R}^2} m^n(dz) = \Theta(n)$ where m^n is given by (2.6). Hence, for each $l \geq 0$, $J_{l+1} - J_l$ is stochastically bounded above by E_l where the $(E_i)_{i \geq 0}$ are i.i.d. exponential random variables of this rate.

Combining these observations,

$$\begin{aligned}
 \mathbb{P}_{(0,5\mathcal{R}_n)}[\tau^{5\mathcal{R}_n} \wedge \tau_0 > n^{-1/2}] &\leq \mathbb{P}[J_{\lceil 3n^{1/3}+3 \rceil} \geq n^{-1/2}] + \mathbb{P}[G > n^{1/3}] \\
 (4.18) \qquad \qquad \qquad &= \mathcal{O}(n^{-1/6}) + (1 - \beta)^{n^{1/3}} = \mathcal{O}(n^{-1/6}),
 \end{aligned}$$

where the last line follows by Markov’s inequality.

We are now in a position to complete the proof. Recall that η^n overshoots if it has neither coalesced nor diverged by time $(\log n)^{-c}$. Let n be sufficiently large that

$$(\log n)^{1/2}(n^{-1/2} + (\log n)^{-c-1}) \leq (\log n)^{-c}.$$

Thus, if η^n overshoots and $i^* < (\log n)^{1/2}$, then at least one inner excursion must have lasted longer than $n^{-1/2}$ or at least one outer excursion must have lasted longer than $(\log n)^{-c}$. Hence,

$$\begin{aligned}
 \mathbb{P}_{\zeta^n}[\eta^n \text{ overshoots}] &\leq (\log n)^{1/2}(\mathbb{P}_{(0,5\mathcal{R}_n)}[\tau^{5\mathcal{R}_n} \wedge \tau_0 > n^{-1/2}] \\
 &\quad + \mathbb{P}_{[5\mathcal{R}_n,7\mathcal{R}_n]}[\tau^{\gamma_n} \wedge \tau_{4\mathcal{R}_n} > (\log n)^{-c-1}]) \\
 &\quad + \mathbb{P}_{\zeta^n}[i^* > (\log n)^{1/2}].
 \end{aligned}$$

Using (4.18), (4.5) and Lemma 4.3 to bound the right-hand side of the above equation, we obtain

$$\begin{aligned}
 \mathbb{P}_{\zeta^n}[\eta^n \text{ overshoots}] &\leq (\log n)^{1/2}(\mathcal{O}(n^{-1/6}) + \mathcal{O}((\log n)^{1-c})) + \alpha^{(\log n)^{1/2}} \\
 &= \mathcal{O}((\log n)^{3/2-c}),
 \end{aligned}$$

which completes the proof. \square

PROOF OF LEMMA 4.5. We note that the probability that η^n diverges is bounded above by the probability that a divergent outer excursion occurs before a coalescing inner excursion occurs. Let us write $\eta^{n,i,in}$ for the i th inner excursion and $\eta^{n,i,out}$ for the i th outer excursion and let us write $\tau^{r,i,in}$, $\tau_{r,i,in}$ and $\tau^{r,i,out}$, $\tau_{r,i,out}$ for the associated equivalents of τ^r and τ_r . Thus,

$$\begin{aligned}
 &\mathbb{P}_{\zeta^n}[\eta^n \text{ diverges}] \\
 &\leq \mathbb{P}_{\zeta^n}[\inf\{i \geq 1 : \tau^{\gamma_n,i,out} < \tau_{4\mathcal{R}_n,i,out}\} \leq \inf\{i \geq 0 : \tau_{0,i,in} < \tau^{5\mathcal{R}_n,i,in}\}].
 \end{aligned}$$

By the strong Markov property (applied successively at times τ_i^{out} and τ_i^{in}), along with (4.6) and (\dagger) , the right-hand side of the above equation is bounded above by the probability that a geometric random variable with success probability $\Theta(\frac{1}{\log n})$ is smaller than an (independent) geometric random variable with success probability $\alpha' > 0$. With this in hand, an elementary calculation shows that

$$\mathbb{P}_{\zeta^n}[\eta^n \text{ diverges}] = \mathcal{O}\left(\frac{1}{\log n}\right).$$

It remains to prove a lower bound of the same order.

In similar style to (\dagger) and (\ddagger) , it is easily seen that there exists $\delta > 0$ such that for all n :

(\star) For any $x \in [\mathcal{R}_n, 4\mathcal{R}_n]$, if $|\eta_0^n| = x$, the probability that η^n will exit $\mathcal{B}_{5\mathcal{R}_n}(0)$ without coalescing is at least δ .

We note also that ζ^n is equal to $n^{-1/2}\zeta^1$ in distribution, so since we assumed that $\mu((\frac{3}{4}\mathcal{R}, \mathcal{R}]) > 0$, there exists $\varepsilon > 0$ such that $\mathbb{P}[\zeta^n \geq \mathcal{R}_n] \geq \varepsilon$ for all n . Thus, applying the strong Markov property at time τ_0^{in} and using (\star) , we obtain

$$\begin{aligned} \mathbb{P}_{\zeta^n}[\eta^n \text{ diverges}] &\geq \varepsilon \delta \mathbb{P}_{[5\mathcal{R}_n, 7\mathcal{R}_n]}[\tau^{\gamma^n} < \tau_{4\mathcal{R}_n}] - \mathbb{P}_{\zeta^n}[\eta^n \text{ overshoots}] \\ &= \Theta\left(\frac{1}{\log n}\right) \end{aligned}$$

as required, where the final statement follows from Lemma 4.7 and Lemma 4.4 (since $c \geq 3$). \square

4.1.2. *Production of branches.* The next step of the proof of Theorem 2.7 involves further analysis of pairs of potential ancestral lineages: first, we need to check that once a pair has separated to a distance γ_n they will not come back together again before a fixed time K ; second, we need to see that $\log n$ times the divergence probability actually converges (cf. Lemma 4.5) as $n \rightarrow \infty$, since this will determine the branching rate in our branching Brownian motion limit. These two statements are the object of the next two lemmas.

LEMMA 4.9. *Fix $K \in (0, \infty)$. Then*

$$\mathbb{P}_{[(\log n)^{-c}, \infty)}[\tau_{4\mathcal{R}_n} \leq K] = \mathcal{O}\left(\frac{\log \log n}{\log n}\right).$$

LEMMA 4.10. *There exists $\kappa \in (0, \infty)$ such that $(\log n)\mathbb{P}_{\zeta^n}[\eta^n \text{ diverges}] \rightarrow \kappa$ as $n \rightarrow \infty$.*

The remainder of this subsection is occupied with proving Lemmas 4.9 and 4.10.

PROOF OF LEMMA 4.9. We use the Skorohod embedding of $\hat{\eta}$ into the Brownian motion W , as defined in (4.7), to reduce the claim to an equivalent statement about a two-dimensional Bessel process.

Recall that $\eta_0^n = \hat{\eta}_0^n = W_0$ and recall τ_r from (4.3), and that $\hat{\tau}_r$ and T_r are the analogues of τ_r for $\hat{\eta}$ and W respectively. We have that $\eta_s^n = \hat{\eta}_s^n$ for all $s \leq \tau_{4\mathcal{R}_n}$ so

$$(4.19) \quad \begin{aligned} \mathbb{P}_{[(\log n)^{-c}, \infty)}[\tau_{4\mathcal{R}_n} \leq K] &= \mathbb{P}_{[(\log n)^{-c}, \infty)}[\hat{\tau}_{4\mathcal{R}_n} \leq K] \\ &\leq \mathbb{P}_{[(\log n)^{-c}, \infty)}[T_{4\mathcal{R}_n} \leq T^{(S(K))}], \end{aligned}$$

where we used the Skorohod embedding given in (4.7) in the last line. For all $\tilde{K}, C > 0$, since $T^{(k)}$ is increasing in k we have

$$(4.20) \quad \mathbb{P}[T^{(S(K))} \geq \tilde{K}] \leq \mathbb{P}[S(K) \geq Cn] + \mathbb{P}[T^{(Cn)} \geq \tilde{K}].$$

By its definition in (4.7), $S(K)$ is bounded by the sum of two Poisson random variables with parameter $\chi = K \int_{\mathbb{R}^2} m^n(dz)$, where m^n is given by (2.6). In particular, $\chi = \Theta(n)$. Recall that if Z' is Poisson with parameter χ , then (using a Chernoff bound argument) for $k > \chi$,

$$(4.21) \quad \mathbb{P}[Z' > k] \leq \frac{e^{-\chi} (e\chi)^k}{k^k}.$$

Hence, for C sufficiently large, there exists $\delta_1 > 0$ such that

$$(4.22) \quad \mathbb{P}[S(K) \geq Cn] \leq \mathcal{O}(e^{-\delta_1 n}).$$

Now by the definition of $(T^{(m)})_{m \geq 1}$ in (4.7), and since $r_m \leq 2\mathcal{R}_n$ for each m ,

$$\mathbb{P}[T^{(Cn)} \geq \tilde{K}] \leq \mathbb{P}\left[\sum_{i=1}^{Cn} R_i \geq \tilde{K}n\right],$$

where $(R_i)_{i \geq 1}$ is an i.i.d. sequence with $R_1 \stackrel{d}{=} \inf\{t \geq 0 : |W_t| \geq 2\mathcal{R}\}$. Since

$$\mathbb{P}[R_1 \geq k] \leq \mathbb{P}[R_1 \geq k - 1] \mathbb{P}[|W_k - W_{k-1}| \leq 4\mathcal{R}] \leq \mathbb{P}[|W_1 - W_0| \leq 4\mathcal{R}]^k,$$

there exists $\lambda > 0$ such that $\mathbb{E}[e^{\lambda R_1}] < \infty$. Hence by Cramér’s theorem, for \tilde{K} a sufficiently large constant, there exists $\delta_2 > 0$ such that

$$(4.23) \quad \mathbb{P}[T^{(Cn)} \geq \tilde{K}] = \mathcal{O}(e^{-\delta_2 n}).$$

By (4.19) and (4.20) together with (4.22) and (4.23), we now have for \tilde{K} sufficiently large

$$(4.24) \quad \begin{aligned} \mathbb{P}_{[(\log n)^{-c}, \infty)}[\tau_{4\mathcal{R}_n} \leq K] \\ \leq \mathbb{P}_{[(\log n)^{-c}, \infty)}[T_{4\mathcal{R}_n} \leq \tilde{K}] + \mathcal{O}(e^{-\delta_1 n}) + \mathcal{O}(e^{-\delta_2 n}). \end{aligned}$$

To finish, we note that

$$\begin{aligned} & \mathbb{P}_{[(\log n)^{-c}, \infty)}[T_{4\mathcal{R}_n} \leq \tilde{K}] \\ & \leq \sup_{x \geq (\log n)^{-c}} (\mathbb{P}_x[T_{4\mathcal{R}_n} \leq T^{x+\log n}] + \mathbb{P}_x[T^{x+\log n} \leq \tilde{K}]) \\ & \leq \sup_{x \geq (\log n)^{-c}} \left(\frac{\log(x + \log n) - \log x}{\log(x + \log n) - \log(4\mathcal{R}_n)} \right) + \mathbb{P}\left[\sup_{t \leq \tilde{K}} |W_t - W_0| \geq \log n\right] \\ & = \mathcal{O}\left(\frac{\log \log n}{\log n}\right) + \mathcal{O}(e^{-(8\tilde{K})^{-1}(\log n)^2}), \end{aligned}$$

where the second line uses the scale function for a two-dimensional Bessel process, and the third line uses (4.4). Substituting this into (4.24), we have the required result. \square

PROOF OF LEMMA 4.10. Let $p_n := \mathbb{P}_{\zeta^n}[\tau^{\gamma_n} < \tau_0]$. Note that by Lemma 4.4,

$$(4.25) \quad |p_n - \mathbb{P}_{\zeta^n}[\eta^n \text{ diverges}]| = \mathcal{O}\left(\frac{1}{(\log n)^{c-3/2}}\right).$$

Hence, by Lemma 4.5, there exist $0 < d \leq D < \infty$ such that, for all $n \geq 2$,

$$d \leq (\log n)p_n \leq D.$$

It follows that $(p_n)_{n \geq 1}$ has a subsequence $(p_{n_k})_{k \geq 1}$ such that $(\log n_k)p_{n_k} \rightarrow \kappa \in (0, \infty)$. Let $\varepsilon > 0$ and let $N \in \mathbb{N}$ be such that $N \geq 1/\varepsilon$ and $|(\log N)p_N - \kappa| \leq \varepsilon$. By rescaling, noting that $\zeta^n \stackrel{d}{=} \zeta^N (\frac{N}{n})^{1/2}$, and similarly for η^n , we have

$$(4.26) \quad p_N = \mathbb{P}_{\zeta^N}[\tau^{\gamma_N(Nn^{-1})^{1/2}} < \tau_0].$$

Recall, for clarity, that here (as throughout this section) τ^r and τ_0 refer to the stopping times for the process η^n .

Define $X^{n,N} := |\eta^n_{\tau^{\gamma_N(Nn^{-1})^{1/2}}}|$. Increasing N , we may assume that $7\mathcal{R}_n < \gamma_N(Nn^{-1})^{1/2} \leq \gamma_n$ for $n \geq N$. Thus,

$$(4.27) \quad \begin{aligned} p_n &= \mathbb{P}_{\zeta^n}[\tau^{\gamma_N(Nn^{-1})^{1/2}} \leq \tau^{\gamma_n} < \tau_0] \\ &= \mathbb{E}_{\zeta^n}[\mathbb{1}_{\tau^{\gamma_N(Nn^{-1})^{1/2}} < \tau_0} \mathbb{P}_{X^{n,N}}[\tau^{\gamma_n} < \tau_0]]. \end{aligned}$$

Here, the first line holds since $\zeta_n < \gamma_N(Nn^{-1})^{1/2} \leq \gamma_n$, and the second line follows from the first by applying the strong Markov property at time $\tau^{\gamma_N(Nn^{-1})^{1/2}}$.

To estimate (4.27), note that

$$X^{n,N} \in [l^{n,N}, r^{n,N}] := [\gamma_N(Nn^{-1})^{1/2}, \gamma_N(Nn^{-1})^{1/2} + 2\mathcal{R}_n].$$

Using the Skorohod embedding defined in (4.7),

$$\begin{aligned}
 \mathbb{P}_{[l^{n,N}, r^{n,N}]}[\tau^{\gamma_n} < \tau_0] &\geq \inf_{x \geq \gamma_N(Nn^{-1})^{1/2}} \mathbb{P}_x[\tau^{\gamma_n} < \tau_{7\mathcal{R}_n}] \\
 &\geq \inf_{x \geq \gamma_N(Nn^{-1})^{1/2}} \mathbb{P}_x[T^{\gamma_n+2\mathcal{R}_n} < T_{7\mathcal{R}_n}] \\
 (4.28) \qquad &= \frac{\log(\gamma_N(Nn^{-1})^{1/2}) - \log(7\mathcal{R}_n)}{\log(\gamma_n + 2\mathcal{R}_n) - \log(7\mathcal{R}_n)} \\
 &= \frac{\frac{1}{2} \log N + \mathcal{O}(\log \log N)}{\frac{1}{2} \log n + \mathcal{O}(\log \log n)}.
 \end{aligned}$$

Note that, in the above, we (again) use the scale function for a two-dimensional Bessel process to deduce the third line.

We require slightly more work to establish an upper bound. We have

$$\begin{aligned}
 (4.29) \qquad \mathbb{P}_{[l^{n,N}, r^{n,N}]}[\tau^{\gamma_n} < \tau_0] \\
 \leq \mathbb{P}_{[l^{n,N}, r^{n,N}]}[\tau^{\gamma_n} < \tau_{7\mathcal{R}_n}] + \mathbb{P}_{[l^{n,N}, r^{n,N}]}[\tau_{7\mathcal{R}_n} < \tau^{\gamma_n} < \tau_0].
 \end{aligned}$$

We begin by controlling the second term on the right-hand side of (4.29). By the strong Markov property at time $\tau_{7\mathcal{R}_n}$,

$$\begin{aligned}
 \mathbb{P}_{[l^{n,N}, r^{n,N}]}[\tau_{7\mathcal{R}_n} < \tau^{\gamma_n} < \tau_0] &= \mathbb{E}_{[l^{n,N}, r^{n,N}]}[\mathbb{1}_{\tau_{7\mathcal{R}_n} < \tau^{\gamma_n}} \mathbb{P}_{|\eta_{\tau_{7\mathcal{R}_n}}^n|}[\tau^{\gamma_n} < \tau_0]] \\
 &\leq \mathbb{E}_{[l^{n,N}, r^{n,N}]}[\mathbb{P}_{|\eta_{\tau_{7\mathcal{R}_n}}^n|}[\tau^{\gamma_n} < \tau_0]].
 \end{aligned}$$

Since $|\eta_{\tau_{7\mathcal{R}_n}}^n| \in [5\mathcal{R}_n, 7\mathcal{R}_n]$, using (4.6) in the same way as in the proof of Lemma 4.5,

$$(4.30) \qquad \mathbb{P}_{[l^{n,N}, r^{n,N}]}[\tau_{7\mathcal{R}_n} < \tau^{\gamma_n} < \tau_0] = \mathcal{O}\left(\frac{1}{\log n}\right).$$

Next, we control the first term on the right-hand side of (4.29), again using the Skorohod embedding (4.7):

$$\begin{aligned}
 \mathbb{P}_{[l^{n,N}, r^{n,N}]}[\tau^{\gamma_n} < \tau_{7\mathcal{R}_n}] &\leq \mathbb{P}_{[l^{n,N}, r^{n,N}]}[T^{\gamma_n} < T_{5\mathcal{R}_n}] \\
 (4.31) \qquad &\leq \frac{\log(\gamma_N(Nn^{-1})^{1/2} + 2\mathcal{R}_n) - \log(5\mathcal{R}_n)}{\log(\gamma_n) - \log(5\mathcal{R}_n)} \\
 &= \frac{\frac{1}{2} \log N + \mathcal{O}(\log \log N)}{\frac{1}{2} \log n + \mathcal{O}(\log \log n)}.
 \end{aligned}$$

Combining (4.28), (4.29), (4.30) and (4.31),

$$\mathbb{P}_{[l^{n,N}, r^{n,N}]}[\tau^{\gamma_n} < \tau_0] = \frac{\log N + \mathcal{O}(\log \log N)}{\log n + \mathcal{O}(\log \log n)} + \mathcal{O}\left(\frac{1}{\log n}\right).$$

Hence, by (4.27),

$$\begin{aligned}
 p_n &= \mathbb{P}_{\zeta^n}[\tau^{\gamma_N(Nn^{-1})^{1/2}} < \tau_0] \left(\frac{\log N + \mathcal{O}(\log \log N)}{\log n + \mathcal{O}(\log \log n)} + \mathcal{O}\left(\frac{1}{\log n}\right) \right) \\
 &= \frac{(\log N)p_N}{\log n} \left(\frac{1 + \mathcal{O}\left(\frac{\log \log N}{\log N}\right)}{1 + \mathcal{O}\left(\frac{\log \log n}{\log n}\right)} + \mathcal{O}\left(\frac{1}{\log N}\right) \right),
 \end{aligned}$$

where we used (4.26) in the last line. Since $|(\log N)p_N - \kappa| \leq \varepsilon$, we obtain for $n \geq N$

$$(\log n)p_n \geq (\kappa - \varepsilon) \left(\frac{1 + \mathcal{O}\left(\frac{\log \log N}{\log N}\right)}{1 + \mathcal{O}\left(\frac{\log \log n}{\log n}\right)} + \mathcal{O}\left(\frac{1}{\log N}\right) \right)$$

and

$$(\log n)p_n \leq (\kappa + \varepsilon) \left(\frac{1 + \mathcal{O}\left(\frac{\log \log N}{\log N}\right)}{1 + \mathcal{O}\left(\frac{\log \log n}{\log n}\right)} + \mathcal{O}\left(\frac{1}{\log N}\right) \right).$$

Letting $\varepsilon \rightarrow 0$, and hence $N \rightarrow \infty$, $\lim_{n \rightarrow \infty} (\log n)p_n = \kappa$. The result follows by (4.25). \square

4.2. *Convergence to branching Brownian motion.* In this subsection, we identify particular subsets of the dual process that we couple with objects that we call ‘‘caterpillars’’. The caterpillars play the role of individual branches in the limiting branching Brownian motion. Our (eventual) goal is to write down a system of ‘‘branching caterpillars’’ and couple it to the SAFVS dual. Establishing these couplings is greatly simplified by viewing the branching and coalescing dual as a deterministic function of an augmented driving Poisson point process and so our first task is to recast the SAFVS dual in this way.

Recall that we have a fixed impact parameter $u \in (0, 1]$. We define, recursively, a sequence of subsets of $[0, 1]$ as follows:

$$A_u^1 = [0, u], \quad \text{and for } k \geq 1, A_u^{k+1} = uA_u^k \cup (u + (1 - u)A_u^k).$$

Then if $U \sim \text{Unif}[0, 1]$, $(\mathbb{1}_{A_u^k}(U))_{k \geq 1}$ is an i.i.d. sequence of Bernoulli(u) random variables [see Lemma 3.20 in [Kallenberg \(2006\)](#) for a proof in the case $u = \frac{1}{2}$, where $(\mathbb{1}_{A_u^k}(U))_{k \geq 1}$ is the binary expansion of U ; the general case is an easy extension of this].

Let

$$\mathcal{X} = \mathbb{R} \times \mathbb{R}^2 \times \mathbb{R}_+ \times \mathcal{B}_1(0)^2 \times [0, 1]^2.$$

DEFINITION 4.11 (The dual as a deterministic function of a driving point process). Given a simple point process Π on \mathcal{X} , and some $p \in \mathbb{R}^2$, we define $(\mathcal{P}_t(p, \Pi))_{t \geq 0}$ as a process on $\bigcup_{k=1}^\infty (\mathbb{R}^2)^k$ as follows.

For each $t \geq 0$, $\mathcal{P}_t(p, \Pi) = (\xi_t^1, \dots, \xi_t^{N_t})$ for some $N_t \geq 1$. We refer to i as the index of the ancestor ξ_t^i . We begin at time $t = 0$ from a single ancestor $\mathcal{P}_0(p, \Pi) = \xi_0^1 = p$ and proceed as follows.

At each $(t, x, r, z_1, z_2, q, v) \in \Pi$ with $v \geq \mathbf{s}_n$, a neutral event occurs:

1. Let $\xi_{t-}^{n_1}, \dots, \xi_{t-}^{n_m}$ denote the ancestors in $\mathcal{B}_r(x)$ which have not yet coalesced with an ancestor of lower index, with $n_1 < \dots < n_m$. For $1 \leq i \leq m$, mark the ancestor $\xi_{t-}^{n_i}$ iff $q \in A_u^i$. Let $\xi_{t-}^{r_1}, \dots, \xi_{t-}^{r_l}$ denote the marked ancestors.

2. If at least one ancestor is marked, we set $\xi_t^{r_i} = x + rz_1$ for each i and call this the parental location for the event. We say that the ancestor $\xi_t^{r_i}$ has coalesced with the ancestor $\xi_t^{r_1}$, for each $i \geq 2$.

At each $(t, x, r, z_1, z_2, q, v) \in \Pi$ with $v < \mathbf{s}_n$, a selective event occurs:

1. Let $\xi_{t-}^{n_1}, \dots, \xi_{t-}^{n_m}$ denote the ancestors in $\mathcal{B}_r(x)$ which have not yet coalesced with an ancestor of lower index, with $n_1 < \dots < n_m$. For $1 \leq i \leq m$, mark the ancestor $\xi_{t-}^{n_i}$ iff $q \in A_u^i$. Let $\xi_{t-}^{r_1}, \dots, \xi_{t-}^{r_l}$ denote the marked ancestors.

2. If at least one ancestor is marked, we set $\xi_t^{r_i} = x + rz_1$ for each i and add an ancestor $\xi_t^{N_t+1} = x + rz_2$. We call $x + rz_1$ and $x + rz_2$ the parental locations of the event. We say that the ancestor $\xi_t^{r_i}$ has coalesced with the ancestor $\xi_t^{r_1}$, for each $i \geq 2$.

For each $l \in \mathbb{N}$, if ξ_τ^l has coalesced with an ancestor ξ_τ^k of lower index at time τ , we set $\xi_t^l = \xi_t^k$ for all $t \geq \tau$.

In the same way as for the definition of $\mathcal{P}^{(n)}(p)$ before the statement of Theorem 2.7, we shall view $(\mathcal{P}_t(p, \Pi))_{t \geq 0}$ as a collection of potential ancestral lineages. Given a realization of Π , we say that a path that begins at p is a potential ancestral lineage if (1) at each neutral event that it encounters, it moves to the (single) parent and (2) at each selective event it encounters, it moves to one of the parents of that event.

Note that if Π is a Poisson point process on \mathcal{X} with intensity measure

$$(4.32) \quad n dt \otimes n dx \otimes \mu^n(dr) \otimes \pi^{-1} dz_1 \otimes \pi^{-1} dz_2 \otimes dq \otimes dv$$

then as a collection of potential ancestral lineages, $(\mathcal{P}_t(p, \Pi))_{t \geq 0}$ has the same distribution as $\mathcal{P}^{(n)}(p)$.

When Π takes this form, the result is that the driving Poisson point process in (2.4) has been augmented by components that determine the nature of each event (neutral or selective), the parental locations of each event and which lineages in the region of the event are affected by it. We have abused notation by retaining the notation Π for this augmented process.

4.2.1. *The caterpillar.* We now introduce the notion of a caterpillar, which involves following a pair of potential ancestral lineages in the dual. We stop the caterpillar if the pair of lineages reaches displacement of $(\log n)^{-c}$, or if the pair does not coalesce within time $(\log n)^{-c}$ after last branching. While doing so, we suppress the creation of the second potential parent at any selective events that occur within time $(\log n)^{-c}$ of the previous (unsuppressed) selective event.

Let Π be a Poisson point process on \mathcal{X} with intensity measure (4.32). We write $(\mathcal{P}_t(p, \Pi))_{t \geq 0} = (\xi_t^1, \dots, \xi_t^{N_t})_{t \geq 0}$ as defined in Definition 4.11.

DEFINITION 4.12 (Caterpillar). For $p \in \mathbb{R}^2$, we define a lifetime $h(p, \Pi) > 0$, and a process $(c_t(p, \Pi))_{0 \leq t \leq h(p, \Pi)}$ on $(\mathbb{R}^2)^2$, which we shall refer to as a caterpillar. For each $t \geq 0$, we write

$$c_t(p, \Pi) = (c_t^1(p, \Pi), c_t^2(p, \Pi)),$$

dropping the dependence on (p, Π) from our notation, when convenient. As part of the definition, we will also define $k^*(p, \Pi) \in \mathbb{N}$ and a sequence $(\tau_k^{\text{br}})_{k \leq k^*}$ of stopping times.

Set $\tau_0^{\text{br}} = 0$ and let τ_1^{br} be the time of the first selective event after $(\log n)^{-c}$ to affect ξ^1 . For $t \leq \tau_1^{\text{br}}$, let $c_t^1 = c_t^2 = \xi_t^1$.

Then, for $k \geq 1$, suppose we have defined $(\tau_l^{\text{br}})_{l \leq k}$; let $m(k) = N_{\tau_k^{\text{br}}}$.

For $t \in [\tau_k^{\text{br}}, \tau_k^{\text{br}} + (\log n)^{-c}]$, define $c_t^1(p, \Pi) = \xi_t^1$ and $c_t^2(p, \Pi) = \xi_t^{m(k)}$.

In analogy with Definition 4.2, define

$$\begin{aligned} \tau_k^{\text{div}} &= \inf\{t \geq \tau_k^{\text{br}} : |c_t^1 - c_t^2| \geq (\log n)^{-c}\}, \\ \tau_k^{\text{coal}} &= \inf\{t \geq \tau_k^{\text{br}} : c_t^1 = c_t^2\}, \\ \tau_k^{\text{over}} &= \tau_k^{\text{br}} + (\log n)^{-c}, \end{aligned} \tag{4.33}$$

and let $\tau_k^{\text{type}} = \min(\tau_k^{\text{div}}, \tau_k^{\text{coal}}, \tau_k^{\text{over}})$. If $\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}$ then set $k^*(p, \Pi) = k$ and $h(p, \Pi) = \tau_{k^*}^{\text{type}}$. The definition is then complete. If not, we proceed as follows.

Let τ_{k+1}^{br} be the time of the first selective event occurring strictly after $\tau_k^{\text{br}} + (\log n)^{-c}$ to affect ξ^1 . For $t \in [\tau_k^{\text{br}} + (\log n)^{-c}, \tau_{k+1}^{\text{br}}]$, let $c_t^1(p, \Pi) = c_t^2(p, \Pi) = \xi_t^1$.

We then continue iteratively for each $k \leq k^*(p, \Pi)$.

We refer to $(\tau_k^{\text{br}})_{k \leq k^*}$, the times at which a selective event results in branching, as branching events. We shall abuse our previous terminology and say that a branching event diverges, coalesces or overshoots when the same is true of the excursion corresponding to the pair (c^1, c^2) .

REMARK 4.13. Note that $(c_t)_{t \geq 0}$ is not a Markov process with respect to its natural filtration, since c^1 and c^2 are not allowed to branch off from each

other within $(\log n)^{-c}$ of the previous branching event. However, for $i = 1, 2$, $(c_t^i(p, \Pi))_{0 \leq t \leq h(p, \Pi)}$ is a Markov process with the same jump rate and jump distribution as a single potential ancestral lineage in the rescaled SAFVS dual. Moreover, for each $1 \leq k \leq k^*$, $(c_t^1, c_t^2)_{\tau_k^{\text{br}} \leq t \leq \tau_k^{\text{type}}}$ is an excursion as defined in Section 4.1.

Recall the definition of $m^n(dz)$ from (2.6) and let

$$(4.34) \quad \kappa_n = (\log n) \mathbb{P}[\tau_1^{\text{type}} \neq \tau_1^{\text{coal}}] \quad \text{and} \quad \lambda = n^{-1} \int_{\mathbb{R}^2} m^n(dz) = \Theta(1).$$

By combining Lemma 4.10 and Lemma 4.4,

$$(4.35) \quad \kappa_n \rightarrow \kappa$$

as $n \rightarrow \infty$.

By the strong Markov property of Π , and since $\tau_k^{\text{type}} \leq \tau_k^{\text{br}} + (\log n)^{-c} \leq \tau_{k+1}^{\text{br}}$ for each k , the types of the selective events, $(\{\tau_k^{\text{type}} = \tau_k^{\text{div}}\})_{k \geq 1}$, $(\{\tau_k^{\text{type}} = \tau_k^{\text{coal}}\})_{k \geq 1}$ and $(\{\tau_k^{\text{type}} = \tau_k^{\text{over}}\})_{k \geq 1}$ are each i.i.d. sequences. Thus,

$$(4.36) \quad k^*(p, \Pi) \sim \text{Geom}(\kappa_n (\log n)^{-1}).$$

By (4.35), there exist constants $0 < a \leq A < \infty$ such that $\kappa_n \in [a, A]$ for all n sufficiently large, so

$$(4.37) \quad \mathbb{P}[k^* \geq (\log n)^{9/8}] = \left(1 - \frac{\kappa_n}{\log n}\right)^{(\log n)^{9/8}} = \mathcal{O}(e^{-\delta (\log n)^{1/8}})$$

for some $\delta > 0$.

LEMMA 4.14. *We can couple $h(p, \Pi)$ with $H \sim \text{Exp}(\kappa_n \lambda)$ in such a way that for some $\delta > 0$, with probability at least $1 - \mathcal{O}(e^{-\delta (\log n)^{1/8}})$*

$$|h(p, \Pi) - H| \leq 3(\log n)^{-1/4}.$$

PROOF. Recall the definition of λ in (4.34). Since the total rate at which c^1 jumps is given by λn , and each jump is from a selective event independently with probability $\mathbf{s}_n = \frac{\log n}{n}$, by the strong Markov property of Π we have that

$$(4.38) \quad E_k := \tau_k^{\text{br}} - (\tau_{k-1}^{\text{br}} + (\log n)^{-c}) \sim \text{Exp}(\lambda \log n)$$

and $(E_k, \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k \geq 1}$ is an i.i.d. sequence.

Since (e.g.) $\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}$ is not independent of the radius of the event at τ_k^{br} , we note that E_k and $\mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}}$ are not independent; therefore, $(E_k)_{k \geq 1}$ is not independent of k^* . However, we can couple $(E_k, \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k \geq 1}$ with a sequence $(E'_k)_{k \geq 1}$ which is independent of k^* as follows.

First, sample the sequence $(\mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k \geq 1}$, and then independently sample a sequence $(E'_k, A_k)_{k \geq 1}$ with the same distribution as $(E_k, \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k \geq 1}$. Then, for each $k \geq 1$, if $A_k = \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}}$ set $E_k = E'_k$, and if not sample E_k according to its conditional distribution given $\mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}}$.

We now have a coupling of $(E_k, \mathbb{1}_{\{\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}\}})_{k \geq 1}$ and $(E'_k)_{k \geq 1}$ such that $(E'_k)_{k \geq 1}$ is an i.i.d. sequence, independent of k^* , with $E'_1 \sim \text{Exp}(\lambda \log n)$. Also, since $\mathbb{P}[\tau_k^{\text{type}} \neq \tau_k^{\text{coal}}] = \Theta((\log n)^{-1})$, we have that independently for each k , $E_k = E'_k$ with probability at least $1 - \Theta((\log n)^{-1})$.

We write

$$\sum_{k=1}^{k^*} E_k = \sum_{k=1}^{k^*} E'_k + \sum_{k=1}^{k^*} D_k,$$

where $D_k = E_k - E'_k$ and, by (4.36), $\sum_{k=1}^{k^*} E'_k \sim \text{Exp}(\lambda \kappa_n)$.

Our next step is to bound $\sum_{k=1}^{k^*} D_k$. Firstly, applying a Chernoff bound to the binomial distribution yields

$$\begin{aligned} \mathbb{P}[|k < (\log n)^{9/8} : D_k \neq 0|] &\geq (\log n)^{1/4} \\ (4.39) \quad &= \mathbb{P}[\text{Bin}((\log n)^{9/8}, \Theta((\log n)^{-1})) \geq (\log n)^{1/4}] \\ &= \mathcal{O}(\exp(-\delta' (\log n)^{1/4})) \end{aligned}$$

for some $\delta' > 0$. Second,

$$\begin{aligned} \mathbb{P}[|D_1| \geq (\log n)^{-1/2}] &\leq \mathbb{P}\left[E_1 \geq \frac{1}{2}(\log n)^{-1/2}\right] + \mathbb{P}\left[E'_1 \geq \frac{1}{2}(\log n)^{-1/2}\right] \\ (4.40) \quad &= 2 \exp(-\lambda (\log n)^{1/2} / 2). \end{aligned}$$

Combining (4.37), (4.39) and (4.40), we have that

$$(4.41) \quad \mathbb{P}\left[\sum_{k=1}^{k^*} D_k \geq (\log n)^{-1/4}\right] = \mathcal{O}(e^{-\delta'' (\log n)^{1/8}}),$$

for some $\delta'' \in (0, \delta)$.

Note that

$$\sum_{k=1}^{k^*} E_k = \tau_{k^*}^{\text{br}} - k^* (\log n)^{-c} = h - k^* (\log n)^{-c} - (\tau_{k^*}^{\text{type}} - \tau_{k^*}^{\text{br}}),$$

with $0 \leq \tau_{k^*}^{\text{type}} - \tau_{k^*}^{\text{br}} \leq (\log n)^{-c}$. Let $H = \sum_{k=1}^{k^*} E'_k$. Then by (4.37) and (4.41), we have

$$\mathbb{P}[|h(p, \Pi) - H| \geq (\log n)^{9/8-c} + (\log n)^{-c} + (\log n)^{-1/4}] = \mathcal{O}(e^{-\delta'' (\log n)^{1/8}}).$$

The result follows since $c \geq 3$. \square

Our next step is to show that a caterpillar is unlikely to end with an overshooting event.

LEMMA 4.15. *As $n \rightarrow \infty$, $\mathbb{P}[\tau_{k^*}^{\text{type}} = \tau_{k^*}^{\text{over}}] = \mathcal{O}((\log n)^{\frac{21}{8}-c})$.*

PROOF. By Lemma 4.4, for $k \geq 1$,

$$(4.42) \quad \mathbb{P}[\tau_k^{\text{type}} = \tau_k^{\text{over}}] = \mathcal{O}((\log n)^{\frac{3}{2}-c}).$$

Moreover,

$$\{\tau_{k^*}^{\text{type}} = \tau_{k^*}^{\text{over}}\} \subset \{k^* \geq (\log n)^{9/8}\} \cup \bigcup_{k=1}^{(\log n)^{9/8}} \{\tau_k^{\text{type}} = \tau_k^{\text{over}}\}.$$

It follows, using (4.37), that

$$\mathbb{P}[\tau_{k^*}^{\text{type}} = \tau_{k^*}^{\text{over}}] = \mathcal{O}(e^{-\delta(\log n)^{1/8}}) + \mathcal{O}((\log n)^{\frac{3}{2}+\frac{9}{8}-c}) = \mathcal{O}((\log n)^{\frac{21}{8}-c}).$$

This completes the proof. \square

We now show that a single caterpillar can be coupled to a Brownian motion in such a way that the caterpillar closely follows the Brownian motion, during time $[0, h(p, \Pi)]$.

Recall that the rate at which ξ^1 jumps from y to $y + z$ is given by intensity measure $m^n(dz)$, defined in (2.6). Thus, for $(c_t)_{t \geq 0}$ started at p , $\mathbb{E}[c_t^1] = p$ and the covariance matrix of c_t^1 is $\sigma^2 t \text{Id}$ since by (2.8),

$$\sigma^2 = \frac{1}{2} \int_{\mathbb{R}^2} |z|^2 m^n(dz).$$

Armed with this, the following lemma is no surprise.

LEMMA 4.16. *Let $(W_t)_{t \geq 0}$ be a two-dimensional Brownian motion with $W_0 = p$. We can couple $(c_t(p, \Pi))_{t \leq h(p, \Pi)}$ with $(W_t)_{t \geq 0}$, in such a way that $(W_t)_{t \geq 0}$ is independent of $(\tau_k^{\text{br}})_{k \geq 1}$ and $k^*(p, \Pi)$, and for any $r > 0$, with probability at least $1 - \mathcal{O}((\log n)^{-r})$, for $t \leq h(p, \Pi)$,*

$$|c_t^1(p, \Pi) - W_{\sigma^2 t}| \leq (\log n)^{\frac{9}{8}-\frac{c}{3}}.$$

REMARK 4.17. By the definition of the caterpillar in Definition 4.12, for all $t \leq h(p, \Pi)$, $|c_t^2 - c_t^1| \leq (\log n)^{-c}$. Hence, under the coupling of Lemma 4.16, with probability at least $1 - \mathcal{O}((\log n)^{-r})$, $|c_t^2(p, \Pi) - W_{\sigma^2 t}| \leq 2(\log n)^{\frac{9}{8}-\frac{c}{3}}$.

PROOF OF LEMMA 4.16. The proof is closely related to the second half of the proof of Lemma 4.7. Note for $k \geq 0$, on the time interval $[\tau_k^{\text{br}} + (\log n)^{-c}, \tau_{k+1}^{\text{br}})$, c_t^1

is a pure jump process with rate of jumps from y to $y + z$ given by $(1 - \mathbf{s}_n)m^n(dz)$. Let $(\tilde{c}_t)_{t \geq 0}$ be a pure jump process with $\tilde{c}_0 = 0$ and rate of jumps from y to $y + z$ given by $(1 - \mathbf{s}_n)m^n(dz)$. For $i \geq 1$, let

$$X_i = \tilde{c}_{i/n} - \tilde{c}_{(i-1)/n}.$$

Then $(X_i)_{i \geq 1}$ are i.i.d., and as in (4.11) and (4.12), we have $\mathbb{E}[|X_1|^2] = \frac{2\sigma^2(1-\mathbf{s}_n)}{n}$ and $\mathbb{E}[|X_1|^4] = \mathcal{O}(n^{-2})$.

By the same Skorohod embedding argument as for (4.13), there is a two-dimensional Brownian motion W started at 0 and a sequence v_1, v_2, \dots of stopping times for W such that for $i \geq 1$, $W_{v_i} = \tilde{c}_{i/n}$ and

$$\mathbb{P}[|v_{\lfloor tn \rfloor} - \sigma^2(1 - \mathbf{s}_n)t| \geq n^{-1/3}] \leq \mathcal{O}(tn^{-1/3}).$$

Fix $t > 0$. Since $\mathbf{s}_n = \frac{\log n}{n}$, for n sufficiently large,

$$\mathbb{P}[|v_{\lfloor tn \rfloor} - \sigma^2 t| \geq 2n^{-1/3}] \leq \mathcal{O}(n^{-1/3}).$$

Then by a union bound over $j = 1, \dots, \lfloor n^{1/4}t \rfloor$,

$$\begin{aligned} \mathbb{P}[\exists j \leq \lfloor n^{1/4}t \rfloor : |v_{\lfloor jn^{3/4} \rfloor} - \sigma^2 jn^{-1/4}| \geq 2n^{-1/3}] &\leq (n^{1/4}t)\mathcal{O}(n^{-1/3}) \\ (4.43) \qquad \qquad \qquad &= \mathcal{O}(n^{-1/12}). \end{aligned}$$

Again by a union bound over j ,

$$\begin{aligned} \mathbb{P}[\exists j \leq \lfloor n^{1/4}t \rfloor : \sup\{|W_{\sigma^2 jn^{-1/4}} - W_u| : \\ (4.44) \quad u \in [\sigma^2 jn^{-1/4} - 2n^{-1/3}, \sigma^2(j+1)n^{-1/4} + 2n^{-1/3}]\} \geq n^{-1/10}] \\ \leq (n^{1/4}t)2\mathbb{P}\left[\sup\{|W_s - W_0| : s \in [0, 4n^{-1/3}]\} \geq \frac{1}{2}n^{-1/10}\right] \\ \leq 4n^{1/4}t \exp(-n^{2/15}/128) = o(n^{-1/12}). \end{aligned}$$

Here, the last line follows by (4.4).

Under the complement of the event of (4.43), for all $j < \lfloor n^{1/4}t \rfloor$, $|v_{\lfloor jn^{3/4} \rfloor} - \sigma^2 jn^{-1/4}| \leq 2n^{-1/3}$ and $|v_{\lfloor (j+1)n^{3/4} \rfloor} - \sigma^2(j+1)n^{-1/4}| \leq 2n^{-1/3}$,

which implies that for i such that $jn^{-1/4} \leq in^{-1} \leq (j+1)n^{-1/4}$,

$$v_i \in [\sigma^2 jn^{-1/4} - 2n^{-1/3}, \sigma^2(j+1)n^{-1/4} + 2n^{-1/3}].$$

Hence, combining (4.43) and (4.44),

$$\mathbb{P}[\exists i \leq \lfloor tn \rfloor : |\tilde{c}_{i/n} - W_{\sigma^2 i/n}| \geq 2n^{-1/10}] = \mathcal{O}(n^{-1/12}).$$

Our next step is to control $|\tilde{c}_s - \tilde{c}_{i/n}|$ during the interval $s \in [i/n, (i+1)/n]$. The distribution of the number of jumps made by \tilde{c} on an interval $[i/n, (i+1)/n]$

is Poisson with parameter $(1 - \mathbf{s}_n)\lambda$, where λ is given by (4.34), and the maximum jump size is $2\mathcal{R}_n$; using (4.21) with $\chi = (1 - \mathbf{s}_n)\lambda$ and $k = \log n$ gives that

$$\mathbb{P}\left[\exists i \leq \lfloor tn \rfloor : \sup_{s \in [i/n, (i+1)/n]} |\tilde{c}_s - \tilde{c}_{i/n}| \geq (\log n)2\mathcal{R}_n\right] = o(n^{-1}).$$

Hence, for n large enough that $(\log n)2\mathcal{R}_n \leq n^{-1/10}$, using (4.44) again to bound $|W_s - W_{\sigma^2 i/n}|$ during the interval $[\sigma^2 i/n, \sigma^2(i+1)/n]$ we have

$$(4.45) \quad \mathbb{P}\left[\sup_{s \leq t} |\tilde{c}_s - W_{\sigma^2 s}| \geq 4n^{-1/10}\right] = \mathcal{O}(n^{-1/12}).$$

We now apply this coupling to $(c_t^1)_{\tau_k^{\text{br}} + (\log n)^{-c} \leq t \leq \tau_{k+1}^{\text{br}}$ for each $k \geq 0$, and let the caterpillar evolve independently of the Brownian motion on each interval $[\tau_k^{\text{br}}, \tau_{k+1}^{\text{br}} + (\log n)^{-c}]$.

More precisely, let $(\tilde{c}^k)_{k \geq 0}$ be an i.i.d. sequence of pure jump processes with $\tilde{c}_0^k = 0$ and rate of jumps from y to $y + z$ given by $(1 - \mathbf{s}_n)m^n(dz)$. Let $(W^k)_{k \geq 0}$ be an i.i.d. sequence of 2-dimensional Brownian motions started at 0 and for each $k \geq 0$, couple W^k and \tilde{c}^k in the same way as above, so that for fixed $t > 0$, for each $k \geq 0$,

$$(4.46) \quad \mathbb{P}\left[\sup_{s \leq t} |\tilde{c}_s^k - W_{\sigma^2 s}^k| \geq 4n^{-1/10}\right] = \mathcal{O}(n^{-1/12}).$$

Then by the strong Markov property for the process c^1 , we can couple $(\tilde{c}^k, W^k)_{k \geq 0}$ and c^1 in such a way that for $k \geq 0$ and $s \in [0, \tau_{k+1}^{\text{br}} - (\tau_k^{\text{br}} + (\log n)^{-c})]$,

$$c_{s + \tau_k^{\text{br}} + (\log n)^{-c}}^1 - c_{\tau_k^{\text{br}} + (\log n)^{-c}}^1 = \tilde{c}_s^k,$$

and $(\tilde{c}^k, W^k)_{k \geq 0}$ is independent of $(\tau_k^{\text{br}}, (c_t^1 - c_t^2)|_{[\tau_k^{\text{br}}, \tau_k^{\text{br}} + (\log n)^{-c}]})_{k \geq 0}$.

Let B be another independent 2-dimensional Brownian motion started at 0. We now define a single Brownian motion W by piecing together increments of B and $(W^k)_{k \geq 0}$. For $s < \sigma^2(\log n)^{-c}$, let $W_s = B_s + p$. Then for $k \geq 0$, define the increments of W on the time interval $[\sigma^2(\tau_k^{\text{br}} + (\log n)^{-c}), \sigma^2(\tau_{k+1}^{\text{br}} + (\log n)^{-c})]$ as follows. For $s \in [0, \sigma^2(\tau_{k+1}^{\text{br}} - \tau_k^{\text{br}})]$, let

$$W_{s + \sigma^2(\tau_k^{\text{br}} + (\log n)^{-c})} - W_{\sigma^2(\tau_k^{\text{br}} + (\log n)^{-c})} = W_s^k.$$

Then W is a Brownian motion independent of $(\tau_k^{\text{br}}, (c_t^1 - c_t^2)|_{[\tau_k^{\text{br}}, \tau_k^{\text{br}} + (\log n)^{-c}]})_{k \geq 0}$, which implies that W is independent of both k^* and $(\tau_k^{\text{br}})_{k \geq 1}$.

We now check that W_t is close to c_t^1 for $t < h$. By (4.38),

$$\mathbb{P}[\tau_{k+1}^{\text{br}} - \tau_k^{\text{br}} \geq 1 + (\log n)^{-c}] \leq n^{-\lambda}.$$

Hence, applying (4.46) with $t = 1 + (\log n)^{-c}$ for each $k \leq (\log n)^{9/8}$ and using (4.37), we have that with probability at least $1 - \mathcal{O}(e^{-\delta(\log n)^{1/8}})$, for $0 \leq k \leq k^*$

and $t \in [\tau_k^{\text{br}} + (\log n)^{-c}, \tau_{k+1}^{\text{br}})$,

$$(4.47) \quad |(c_t^1 - c_{\tau_k^{\text{br}} + (\log n)^{-c}}^1) - (W_{\sigma^2 t} - W_{\sigma^2(\tau_k^{\text{br}} + (\log n)^{-c})})| \leq 4n^{-1/10}.$$

For each k , by (4.4),

$$(4.48) \quad \begin{aligned} & \mathbb{P} \left[\sup \{ |W_{\sigma^2 t} - W_{\sigma^2 \tau_k^{\text{br}}}| : t \in [\tau_k^{\text{br}}, \tau_k^{\text{br}} + (\log n)^{-c}] \} \geq \frac{1}{3} (\log n)^{-c/3} \right] \\ & \leq 4 \exp(-(\log n)^{c/3} / 72 \sigma^2) \\ & = o((\log n)^{-r - \frac{9}{8}}), \end{aligned}$$

for any $r > 0$. Hence, using (4.37) again,

$$(4.49) \quad \begin{aligned} & \mathbb{P} \left[\sum_{k=1}^{k^*} \sup \{ |W_{\sigma^2 t} - W_{\sigma^2 \tau_k^{\text{br}}}| : t \in [\tau_k^{\text{br}}, \tau_k^{\text{br}} + (\log n)^{-c}] \} \geq \frac{1}{3} (\log n)^{\frac{9}{8} - \frac{c}{3}} \right] \\ & \leq \mathbb{P}[k^* \geq (\log n)^{9/8}] + (\log n)^{9/8} o((\log n)^{-r - \frac{9}{8}}) \\ & = o((\log n)^{-r}). \end{aligned}$$

For $k \geq 0$, on the time interval $[\tau_k^{\text{br}}, \tau_k^{\text{br}} + (\log n)^{-c}]$ the process c_t^1 is a pure jump process with rate of jumps from y to $y + z$ given by $m^n(dz)$. Hence, using the same Skorohod embedding argument as for (4.45), we can couple $(c_{s+\tau_k^{\text{br}}}^1 - c_{\tau_k^{\text{br}}}^1)_{s \leq (\log n)^{-c}}$ with a Brownian motion W' started at 0 in such a way that

$$\mathbb{P} \left[\sup_{s \leq (\log n)^{-c}} |(c_{s+\tau_k^{\text{br}}}^1 - c_{\tau_k^{\text{br}}}^1) - W_{\sigma^2 s}'| \geq 4n^{-1/10} \right] = \mathcal{O}(n^{-1/12}).$$

Applying (4.49) and (4.37), it follows that

$$\begin{aligned} & \mathbb{P} \left[\sum_{k=1}^{k^*} \sup \{ |c_t^1 - c_{\tau_k^{\text{br}}}^1| : t \in [\tau_k^{\text{br}}, \tau_k^{\text{br}} + (\log n)^{-c}] \} \right. \\ & \quad \left. \geq \frac{1}{3} (\log n)^{\frac{9}{8} - \frac{c}{3}} + 4n^{-1/10} (\log n)^{9/8} \right] \\ & = \mathcal{O}((\log n)^{-r}). \end{aligned}$$

The stated result follows by combining the above equation with (4.47), (4.37) and (4.49). \square

4.2.2. *The branching caterpillar.* We now construct a branching process of caterpillars. We start from a single caterpillar and allow it to evolve until the time h . We start two independent caterpillars from the locations of c_h^1 and c_h^2 . Now iterate. The independent caterpillars defined in this way will be indexed by points of $\mathcal{U} = \{\emptyset\} \cup \bigcup_{k=1}^\infty \{1, 2\}^k$. More formally, we have the following.

DEFINITION 4.18 (Branching caterpillar). Let $(\Pi_j)_{j \in \mathcal{U}}$ be a sequence of independent Poisson point processes on \mathcal{X} with intensity measure (4.32). For $p \in \mathbb{R}^2$, we define $(\mathcal{C}_t(p, (\Pi_j)_{j \in \mathcal{U}}))_{t \geq 0}$ as a process on $\bigcup_{k=1}^\infty (\mathbb{R}^2)^k$ as follows. For $s > 0$, let

$$(4.50) \quad \Pi_j^s = \{(t - s, x, r, z_1, z_2, q, v) : (t, x, r, z_1, z_2, q, v) \in \Pi_j\}.$$

Define (p_j, t_j, h_j) inductively for $j \in \mathcal{U}$ by $p_\emptyset = p, t_\emptyset = 0$ and

$$\begin{aligned} h_j &= t_j + h(p_j, \Pi_j^{t_j}), \\ t_{(j,1)} &= t_{(j,2)} = h_j, \\ p_{(j,1)} &= c_{h_j - t_j}^1(p_j, \Pi_j^{t_j}), \\ p_{(j,2)} &= c_{h_j - t_j}^2(p_j, \Pi_j^{t_j}). \end{aligned}$$

Finally, define $\mathcal{U}(t) = \{j \in \mathcal{U} : t_j \leq t \leq h_j\}$ and

$$\mathcal{C}_t(p, (\Pi_j)_{j \in \mathcal{U}}) = (c_{t-t_j}(p_j, \Pi_j^{t_j}))_{j \in \mathcal{U}(t)}.$$

In words, $\mathcal{U}(t)$ is the set of indices of the caterpillars that are active at time t , and \mathcal{C}_t is the set of (positions of) those caterpillars. Note that we translate the time coordinates in (4.50) to match our definition of a caterpillar, which began at time 0. The jumps in \mathcal{C}_t occur at the time coordinates of events in $\bigcup_{j \in \mathcal{U}} \Pi_j$.

We now show that for any constant $a > 0$, with high probability, the longest “chain” of caterpillars has length at most $a \log \log n + 1$. For $k \in \mathbb{N}$, let $\mathcal{U}_k = \{\emptyset\} \cup \bigcup_{j=1}^k \{1, 2\}^j$.

LEMMA 4.19. Fix $T > 0$; then for any $r > 0, a > 0, \mathbb{P}[\mathcal{U}(T) \not\subseteq \mathcal{U}_{\lfloor a \log \log n \rfloor}] = o((\log n)^{-r})$.

PROOF. Fix $v \in \{1, 2\}^{\lfloor a \log \log n \rfloor + 1}$. Then by a union bound,

$$(4.51) \quad \mathbb{P}[\exists w \in \{1, 2\}^{\lfloor a \log \log n \rfloor + 1} \text{ s.t. } t_w \leq T] \leq 2^{\lfloor a \log \log n \rfloor + 1} \mathbb{P}[t_v \leq T].$$

Note that by Lemma 4.14, $t_v = \sum_{i=1}^{\lfloor a \log \log n \rfloor + 1} H_i + R$ where $(H_i)_{i \geq 1}$ are i.i.d. with $H_1 \sim \text{Exp}(\lambda \kappa_n)$ and

$$\mathbb{P}[R \geq 3(a \log \log n + 1)(\log n)^{-1/4}] = \mathcal{O}((\log \log n)e^{-\delta(\log n)^{1/8}}).$$

Hence [if n is sufficiently large that $3(a \log \log n + 1)(\log n)^{-1/4} \leq T/2$], if Z' is Poisson with parameter $\lambda \kappa_n T/2$,

$$\mathbb{P}[t_v \leq T] \leq \mathbb{P}[Z' \geq a \log \log n + 1] + \mathcal{O}((\log \log n)e^{-\delta(\log n)^{1/8}}).$$

We use (4.21) and combine with (4.51) to deduce that, for any $r > 0$,

$$\mathbb{P}[\mathcal{U}(T) \not\subseteq \mathcal{U}_{\lfloor a \log \log n \rfloor}] = \mathbb{P}[\exists w \in \{1, 2\}^{\lfloor a \log \log n \rfloor + 1} \text{ s.t. } t_w \leq T] = o((\log n)^{-r}).$$

This completes the proof. \square

The next task is to couple the branching caterpillar to the rescaled dual of the SAFVS. Since we have expressed the dual as a deterministic function of the driving point process of events in Definition 4.11, it is enough to find an appropriate coupling of the driving events for the branching caterpillar and those of a SAFVS dual.

The idea, roughly, is as follows. Each “branch” of the branching caterpillar is constructed from an independent driving process. For each of these, we should like to retain those events that affected the caterpillar, but we can discard the rest. If two or more caterpillars are close enough that the events affecting them could overlap, to avoid having too many events in these regions we have to arbitrarily choose one caterpillar and discard the events affecting the others. We then supplement these with additional events, appropriately distributed to fill in the gaps and arrive at the driving Poisson point process for a SAFVS dual, with intensity as in (4.32). We will then check that the SAFVS dual corresponding to this point process coincides with our branching caterpillar, with probability tending to one as $n \rightarrow \infty$.

To put this strategy into practice, we require some notation. Let $\mathcal{U}_0 = \mathcal{U} \cup \{0\}$. For $V \subset \mathcal{U}_0$ let $\max(V)$ refer to the maximum element of V with respect to a fixed ordering in which 0 is the minimum value (it does not matter precisely which ordering we use, but we must fix one). Given a sequence $(\Pi_j)_{j \in \mathcal{U}_0}$ of independent Poisson point processes on \mathcal{X} with intensity measure (4.32), define a simple point process Π as follows. Let

$$(4.52) \quad \begin{aligned} j(t, x) = \max(\{ &k \in \mathcal{U}(t) : \exists i \in \{1, 2\} \\ &\text{with } |c_{t-t_k}^i(p_k, \Pi_k^{t_k}) - x| \leq \mathcal{R}_n \} \cup \{0\}). \end{aligned}$$

Note that $j(t, x) = 0$ corresponds to regions of space-time that are not near a caterpillar, so that for $(t, x, r, z_1, z_2, q, v) \in \Pi_0, \mathcal{B}_r(x)$ does not contain a caterpillar. Then we define

$$(4.53) \quad \Pi = \bigcup_{k \in \mathcal{U}_0} \{(t, x, r, z_1, z_2, q, v) \in \Pi_k : j(t, x) = k\}.$$

LEMMA 4.20. Π is a Poisson point process with intensity measure given by (4.32).

REMARK 4.21. We defined the coupling (4.53) for each $n \in \mathbb{N}$. As such, in the proof of Lemma 4.20 we regard n as a constant and we will not include it inside $\mathcal{O}(\cdot)$, etc.

PROOF OF LEMMA 4.20. Let $\nu(dt, dx, dr, dz_1, dz_2, dq, dv)$ be the intensity measure given in (4.32).

Let \mathcal{B}_0 be the set of bounded Borel subsets of $\mathbb{R}_+ \times \mathbb{R}^2 \times \mathbb{R}_+ \times \mathcal{B}_1(0)^2 \times [0, 1]^2$; for $B \in \mathcal{B}_0$, let $N(B) = |\Pi \cap B|$ and for $j \in \mathcal{U}_0$, let $N_j(B) = |\Pi_j \cap B|$. Suppose

$B = \bigcup_{i=1}^k B_i \in \mathcal{B}_0$ where for each i , $B_i = [a_i, b_i] \times D_i$ for some $a = a_1 < b_1 \leq a_2 < \dots < b_k = b$. Let $\mathcal{B}_R \subset \mathcal{B}_0$ denote the collection of such sets B . Note that Π is a simple point process, and that therefore Π is a Poisson point process with intensity ν if and only if

$$(4.54) \quad \mathbb{P}[N(B) = 0] = e^{-\nu(B)}$$

for all $B \in \mathcal{B}_R$. [See, e.g., Section 3.4 of Kingman (1993).]

For some $\delta > 0$, assume that $b_i - a_i \leq \delta, \forall i$ (by partitioning the B_i further if necessary). Since B is bounded, $\exists d < \infty$ s.t. $|x| \leq d$ for all $(t, x, r, z_1, z_2, q, v) \in B$. We can write

$$(4.55) \quad \begin{aligned} \mathbb{P}[N(B) = 0] &= \mathbb{P}\left[\bigcap_{i=1}^k \{N(B_i) = 0\}\right] \\ &= \mathbb{E}\left[\prod_{i=1}^{k-1} \mathbb{1}_{\{N(B_i)=0\}} \mathbb{P}(N(B_k) = 0 | (\Pi_j(a_k))_{j \in \mathcal{U}_0})\right], \end{aligned}$$

where $\Pi_j(t) := \Pi_j|_{[0,t] \times \mathbb{R}^2 \times \mathbb{R}_+ \times \mathcal{B}_1(0)^2 \times [0,1]^2}$.

For $j \in \mathcal{U}_0$, let $D_k^j = \{(x, r, z_1, z_2, q, v) \in D_k : j(a_k, x) = j\}$ and $B_k^j = [a_k, b_k] \times D_k^j$. Also let

$$\tilde{B}_k = [a_k, b_k] \times \mathcal{B}_{d+3\mathcal{R}_n}(0) \times \mathbb{R}_+ \times \mathcal{B}_1(0)^2 \times [0, 1]^2,$$

and let $\mathcal{V}(t) = \bigcup_{s \leq t} \mathcal{U}(s)$.

For $t \in [a_k, b_k]$, if none of the caterpillars in $\mathcal{B}_{d+3\mathcal{R}_n}(0)$ move during the time interval $[a_k, t]$ then $j(a_k, x) = j(t, x) \forall x \in \mathcal{B}_d(0)$; thus, a point $(t, x, r, z_1, z_2, q, v)$ in $\Pi \cap B_k$ must be a point in $\Pi_j \cap B_k^j$ for some j , and vice versa. We can use this observation to relate $\{N(B_k) = 0\}$ and $\bigcap_{j \in \mathcal{U}_0} \{N_j(B_k^j) = 0\}$, as follows.

If $N(B_k) = 0$ and $N_j(B_k^j) \neq 0$ for some $j \in \mathcal{U}_0$, then $D_k^j \neq \emptyset$ so $j \in \mathcal{V}(a_k) \cup \{0\}$ (either $j = 0$ or the caterpillar indexed by j is alive at time a_k). Also after a_k and before the point in $\Pi_j \cap B_k^j$, one of the caterpillars in $\mathcal{B}_{d+3\mathcal{R}_n}(0)$ must have moved, so there must be a point in $\Pi_l \cap \tilde{B}_k$ for some $l \in \mathcal{V}(b_k)$. Conversely, if $N_j(B_k^j) = 0 \forall j \in \mathcal{U}_0$ and $N(B_k) \neq 0$, then there must be a point in $\Pi_l \cap \tilde{B}_k$ followed by either a point in $\Pi_0 \cap B_k$ or a point in $\Pi_{l'} \cap B_k$ for some $l, l' \in \mathcal{V}(b_k)$. Hence,

$$(4.56) \quad \begin{aligned} &\{N(B_k) = 0\} \Delta \left(\bigcap_{j \in \mathcal{U}_0} \{N_j(B_k^j) = 0\} \right) \\ &\subset \left\{ N_0(B_k) + \sum_{l \in \mathcal{V}(b_k)} N_l(\tilde{B}_k) \geq 2 \right\}. \end{aligned}$$

Note that by the definition of a caterpillar in Definition 4.12, for each $j \in \mathcal{U}$, $h(p_j, \Pi_j^{t_j}) \geq (\log n)^{-c}$. It follows that $\mathcal{V}(b_k) \subseteq \bigcup_{m=0}^{\lceil b_k(\log n)^c \rceil} \{1, 2\}^m$. Also if $J \subset$

\mathcal{U}_0 with $|J| = K$, then $\sum_{j \in J} N_j(\tilde{B}_k)$ has a Poisson distribution with parameter $K\nu(\tilde{B}_k)$, and since $b_k - a_k \leq \delta$, $\nu(\tilde{B}_k) \leq n^2\pi(d + 3\mathcal{R}_n)^2\mu((0, \mathcal{R}])\delta$. Hence, for Z' a Poisson random variable with parameter $(2^{2+b_k(\log n)^c} + 1)\nu(\tilde{B}_k) = \mathcal{O}(\delta)$,

$$\mathbb{P}\left[N_0(B_k) + \sum_{j \in \mathcal{V}(b_k)} N_j(\tilde{B}_k) \geq 2 | (\Pi_j(a_k))_{j \in \mathcal{U}_0}\right] \leq \mathbb{P}[Z' \geq 2] = \mathcal{O}(\delta^2).$$

By (4.56), we now have that

$$\begin{aligned} \mathbb{P}[N(B_k) = 0 | (\Pi_j(a_k))_{j \in \mathcal{U}_0}] &= \mathbb{P}\left[\bigcap_{j \in \mathcal{U}_0} \{N_j(B_k^j) = 0\}\right] + \mathcal{O}(\delta^2) \\ &= \prod_{j \in \mathcal{U}_0} \exp(-\nu(B_k^j)) + \mathcal{O}(\delta^2) \\ &= \exp(-\nu(B_k)) + \mathcal{O}(\delta^2). \end{aligned}$$

Substituting this into (4.55) and then repeating the same argument for $k - 1, k - 2, \dots, 1$,

$$\begin{aligned} \mathbb{P}[N(B) = 0] &= \prod_{i=1}^k \exp(-\nu(B_k)) + \sum_{i=1}^k \mathcal{O}(\delta^2) \\ &= \exp(-\nu(B)) + k\mathcal{O}(\delta^2). \end{aligned}$$

By partitioning B further, we can let $\delta t \rightarrow 0$ with $k = \Theta(1/\delta)$. It follows that $\mathbb{P}[N(B) = 0] = \exp(-\nu(B))$. By (4.54), this completes the proof. \square

It follows immediately from Lemma 4.20 that the collection of potential ancestral lineages in $(\mathcal{P}_t(p, \Pi))_{t \geq 0}$ has the same distribution as $\mathcal{P}^{(n)}(p)$, the rescaled SAFVS dual. We now show that under this coupling the rescaled SAFVS dual and branching caterpillar coincide with high probability.

We consider $(\mathcal{C}_t(p, (\Pi_j)_{j \in \mathcal{U}}))_{0 \leq t \leq T}$ as a collection of paths as follows. The set of paths through a single caterpillar $(c_t(p, \Pi))_{t \leq h(p, \Pi)}$ with $k^*(p, \Pi) = k^*$ is given by $\{l^i\}_{i \in \{1, 2\}^{k^*}}$, where $l^i(t) = c_t^1(p, \Pi)$ for $t \in [0, (\log n)^{-c}]$ and for each $1 \leq k \leq k^*$, $l^i(t) = c_t^{ik}(p, \Pi)$ for $t \in [\tau_{k-1}^{\text{br}} + (\log n)^{-c}, (\tau_k^{\text{br}} + (\log n)^{-c}) \wedge h(p, \Pi)]$. Then the collection of paths through $(\mathcal{C}_t(p, (\Pi_j)_{j \in \mathcal{U}}))_{0 \leq t \leq T}$ is given by concatenating paths through the individual caterpillars, that is, paths $l : [0, T] \rightarrow \mathbb{R}^2$ such that for some sequence $(u_m)_{m \geq 0} \subset \mathcal{U}$ with $u_{m+1} = (u_m, i_m)$ for some $i_m \in \{1, 2\}$ for each m , for $t \in [t_{u_m}, h_{u_m}]$, $l(t)$ follows a path through $(c_{t-t_{u_m}}(p_{u_m}, \Pi_{u_m}^{t_{u_m}}))_t$ with $l(h_{u_m}) = p_{u_{m+1}}$.

LEMMA 4.22. Fix $T > 0$. Let $(\Pi_j)_{j \in \mathcal{U}_0}$ be independent Poisson point processes with intensity measure (4.32) and let Π be defined from $(\Pi_j)_{j \in \mathcal{U}_0}$ as in (4.53). Then $(\mathcal{C}_t(p, (\Pi_j)_{j \in \mathcal{U}}))_{0 \leq t \leq T}$ and $(\mathcal{P}_t(p, \Pi))_{0 \leq t \leq T}$, viewed as collections of paths, are equal with probability at least $1 - \mathcal{O}((\log n)^{-1/4})$.

PROOF. We shall use Lemma 4.19 with $a = (16 \log 2)^{-1}$. Writing, for $j \in \mathcal{U}$, $k^*(j) = k^*(p_j, \Pi_j^{t_j})$, the number of branching events in $c_{t-t_j}(p_j, \Pi_j^{t_j})$ before h_j , by a union bound over $\mathcal{U}_{\lfloor a \log \log n \rfloor}$ and (4.37),

$$(4.57) \quad \begin{aligned} \mathbb{P}[\exists j \in \mathcal{U}_{\lfloor a \log \log n \rfloor} : k^*(j) \geq (\log n)^{9/8}] &\leq 2^{2+a \log \log n} \mathcal{O}(e^{-\delta(\log n)^{1/8}}) \\ &= \mathcal{O}(e^{-\delta(\log n)^{1/8}/2}). \end{aligned}$$

Let $(\tau_k^{\text{br}}(j))_{k \geq 1}$ denote the sequence of branching events in $c_{t-t_j}(p_j, \Pi_j^{t_j})$, and similarly define $(\tau_k^{\text{type}}(j))_{k \geq 1}$ and $(\tau_k^{\text{over}}(j))_{k \geq 1}$ as in (4.33). Note that $(\mathcal{C}_t)_{t \leq T}$ and $(\mathcal{P}_t)_{t \leq T}$ only differ as collections of paths if either a selective event affects a caterpillar during a time interval in which it ignores branching, or if two different caterpillars are simultaneously within \mathcal{R}_n of some $x \in \mathbb{R}^2$ and so one of them is not driven by the pieced together Poisson point process Π . More formally, if $(\mathcal{C}_t)_{t \leq T}$ and $(\mathcal{P}_t)_{t \leq T}$ differ as collections of paths then one or more of the following events occurs:

1. $\mathcal{U}(T) \not\subseteq \mathcal{U}_{\lfloor a \log \log n \rfloor}$ or $k^*(j) \geq (\log n)^{9/8}$ for some $j \in \mathcal{U}_{\lfloor a \log \log n \rfloor}$.
2. For some $j \in \mathcal{U}_{\lfloor a \log \log n \rfloor}$ and $k \leq (\log n)^{9/8}$, the event $E_1(j, k)$ occurs: one of the lineages $c_{t-t_j}^1(p_j, \Pi_j^{t_j})$ and $c_{t-t_j}^2(p_j, \Pi_j^{t_j})$ is affected by a selective event in the time interval $[\tau_k^{\text{br}}(j), \tau_k^{\text{br}}(j) + (\log n)^{-c}]$.
3. For some $w \neq v \in \mathcal{U}_{\lfloor a \log \log n \rfloor}$, the event $E_2(v, w)$ occurs: there are $i_1, i_2 \in \{1, 2\}$ with $|c_{t-t_w}^{i_1}(p_w, \Pi_w^{t_w}) - c_{t-t_v}^{i_2}(p_v, \Pi_v^{t_v})| \leq 2\mathcal{R}_n$ for some $t \leq T$.

Recall from (4.34) and (2.6) that selective events affect a single lineage with rate $\lambda \log n$. Hence, for $k \in \mathbb{N}$ and $j \in \mathcal{U}$, $\mathbb{P}[E_1(j, k)] = \mathcal{O}((\log n)^{1-c})$.

We now consider the event $E_2(v, w)$. For $w \neq v \in \mathcal{U}$, let $i = \min\{j \geq 1 : w_j \neq v_j\}$. Then let

$$w \wedge v = \begin{cases} (w_1, \dots, w_{i-1}) & \text{if } i \geq 2, \\ \emptyset & \text{if } i = 1. \end{cases}$$

At time $h_{w \wedge v}$, either $\tau_{k^*(w \wedge v)}^{\text{type}}(w \wedge v) = \tau_{k^*(w \wedge v)}^{\text{over}}(w \wedge v)$ or $\tau_{k^*(w \wedge v)}^{\text{type}}(w \wedge v) = \tau_{k^*(w \wedge v)}^{\text{div}}(w \wedge v)$, in which case $|p_{(w \wedge v, 1)} - p_{(w \wedge v, 2)}| \geq (\log n)^{-c}$. Conditional on $|p_{(w \wedge v, 1)} - p_{(w \wedge v, 2)}| \geq (\log n)^{-c}$, for $i_1, i_2 \in \{1, 2\}$,

$$(c_{t-t_w}^{i_1}(p_w, \Pi_w^{t_w}), c_{t-t_v}^{i_2}(p_v, \Pi_v^{t_v}))_{t \in [t_w, h_w] \cap [t_v, h_v] \cap [0, T]}$$

is part of the pair of potential ancestral lineages of an excursion started at time $h_{w \wedge v}$ with initial displacement at least $(\log n)^{-c}$. Hence, by Lemmas 4.9 and 4.15,

$$\mathbb{P}[E_2(w, v)] = \mathcal{O}\left(\frac{\log \log n}{\log n}\right) + \mathcal{O}((\log n)^{\frac{21}{8}-c}) = \mathcal{O}((\log n)^{-3/8})$$

since $c \geq 3$. By a union bound, and using Lemma 4.19 and (4.57) it follows that

$$\begin{aligned} & \mathbb{P}[(C_t)_{t \leq T} \neq (P_t)_{t \leq T}] \\ & \leq o((\log n)^{-1}) + 4(\log n)^{a \log 2 + 9/8} \mathbb{P}[E_1(j, k)] \\ & \quad + 16(\log n)^{2a \log 2} \mathbb{P}[E_2(w, v)] \\ & = \mathcal{O}((\log n)^{a \log 2 + \frac{9}{8} + 1 - c}) + \mathcal{O}((\log n)^{2a \log 2 - 3/8}) \\ & = \mathcal{O}((\log n)^{-1/4}), \end{aligned}$$

by our choice of $a = (16 \log 2)^{-1}$ and since $c \geq 3$. \square

We are now ready to complete the proof of Theorem 2.7.

PROOF OF THEOREM 2.7. Set $c = 4$. By Lemmas 4.20 and 4.22, we have a coupling of the rescaled SAFV dual and the branching caterpillar under which the two processes are equal (as collections of paths) with probability at least $1 - \mathcal{O}((\log n)^{-1/4})$.

We now couple $(C_t(p, (\Pi_j)_{j \in \mathcal{U}}))_{0 \leq t \leq T}$ to a branching Brownian motion with branching rate λ_{κ_n} . Let $((W_t^j)_{t \geq 0}, H_j)_{j \in \mathcal{U}}$ be an i.i.d. sequence, where $(W_t^j)_{t \geq 0}$ is a Brownian motion starting at 0 and $H_j \sim \text{Exp}(\lambda_{\kappa_n})$ independent of $(W_t^j)_{t \geq 0}$. For each $j \in \mathcal{U}$, we couple $(c_{t-t_j}(p_j, \Pi_j^{t_j}))_{t \in [t_j, h_j]}$ to $((W_t^j)_{t \geq 0}, H_j)$ as in Lemmas 4.14 and 4.16.

For $j \in \mathcal{U}$, let $A_1(j)$ be the event that both $|(h_j - t_j) - H_j| \leq 3(\log n)^{-1/4}$ and for $i = 1, 2$ and $t \in [t_j, h_j]$,

$$|(c_{t-t_j}^i(p_j, \Pi_j^{t_j}) - p_j) - W_{\sigma^2(t-t_j)}^j| \leq 2(\log n)^{\frac{9}{8} - \frac{c}{3}} = 2(\log n)^{-5/24}.$$

By Lemmas 4.14 and 4.16, for any $r > 0$, for each $j \in \mathcal{U}$, $\mathbb{P}[A_1(j)] \geq 1 - \mathcal{O}((\log n)^{-r})$. Hence, taking a union bound over $j \in \mathcal{U}_{\lfloor \log \log n \rfloor}$,

$$\mathbb{P}\left[\bigcap_{j \in \mathcal{U}_{\lfloor \log \log n \rfloor}} A_1(j)\right] \geq 1 - \mathcal{O}((\log n)^{\log 2 - r}).$$

Also, for $j \in \mathcal{U}$, define the event

$$\begin{aligned} A_2(j) = & \left\{ \sup_{t \in [0, 3(\log n)^{-1/4}]} |W_{\sigma^2 t}^j| \right. \\ & \left. + \sup_{t \in [H_j - 3(\log n)^{-1/4}, H_j]} |W_{\sigma^2 t}^j - W_{\sigma^2 H_j}^j| \leq (\log n)^{-1/9} \right\}. \end{aligned}$$

Then by another union bound over $\mathcal{U}_{\lfloor \log \log n \rfloor}$, since for a Brownian motion $(W_t)_{t \geq 0}$ started at 0, $\mathbb{P}[\sup_{t \in [0, 3(\log n)^{-1/4}]} |W_t| \geq \frac{1}{2}(\log n)^{-1/9}] = o((\log n)^{-r})$, we have

that

$$\mathbb{P}\left[\bigcap_{j \in \mathcal{U}_{\lfloor \log \log n \rfloor}} A_2(j)\right] \geq 1 - \mathcal{O}((\log n)^{\log 2 - r}).$$

By Lemma 4.19, $\mathbb{P}[\mathcal{U}(T) \not\subseteq \mathcal{U}_{\lfloor \log \log n \rfloor}] = o((\log n)^{-r})$.

Define a branching Brownian motion starting at p with diffusion constant σ^2 from $((W_t^j)_{t \geq 0}, H_j)_{j \in \mathcal{U}}$ by letting the increments of the initial particle be given by $(W_{\sigma^2 t}^\emptyset)_{t \geq 0}$ until time H_\emptyset , when it is replaced by two particles which have lifetimes H_1 and H_2 and increments given by $(W_{\sigma^2 t}^1)_{t \geq 0}$, $(W_{\sigma^2 t}^2)_{t \geq 0}$ and so on.

If $\mathcal{U}(T) \subseteq \mathcal{U}_{\lfloor \log \log n \rfloor}$ and $A_1(j) \cap A_2(j)$ occurs for each $j \in \mathcal{U}_{\lfloor \log \log n \rfloor}$, each path in the branching caterpillar stays within distance $2(\log \log n + 1)(\log n)^{-1/9} + 2(\log \log n + 1)(\log n)^{-5/24}$ of some path through the branching Brownian motion and vice versa.

Setting $r = \log 2 + 1/4$ gives us a coupling between the branching caterpillar and branching Brownian motion (with diffusion constant σ^2 and branching rate $\kappa_n \lambda$) such that with probability at least $1 - \mathcal{O}((\log n)^{-1/4})$, up to time T each path in the rescaled SAFVS dual stays within distance $2(\log \log n)(\log n)^{-1/9} + 2(\log \log n)(\log n)^{-5/24}$ of some path through the branching Brownian motion and vice versa. Finally, we need to couple this branching Brownian motion up to time T with a branching Brownian motion with branching rate $\kappa \lambda$. By (4.35), $\kappa_n \rightarrow \kappa$ as $n \rightarrow \infty$, so this follows by straightforward bounds on the difference between the branching times and the increments of a Brownian motion during such a time. \square

Acknowledgements. Our results (with different proofs) form part of the DPhil thesis of the last author. We would like to thank the examiners, Christina Goldschmidt and Anton Wakolbinger, for their careful reading of the thesis and detailed feedback. We would also like to thank the two anonymous referees for their careful reading of the paper and valuable comments.

REFERENCES

- BARTON, N. H. (1993). The probability of fixation of a favoured allele in a subdivided population. *Genet. Res.* **62** 149–157.
- BARTON, N. H., ETHERIDGE, A. M. and VÉBER, A. (2010). A new model for evolution in a spatial continuum. *Electron. J. Probab.* **15** 162–216. [MR2594876](#)
- BARTON, N. H., ETHERIDGE, A. M., KELLEHER, J. and VÉBER, A. (2013a). Genetic hitchhiking in spatially extended populations. *Theor. Popul. Biol.* **87** 75–89.
- BARTON, N. H., ETHERIDGE, A. M., KELLEHER, J. and VÉBER, A. (2013b). Inference in two dimensions: Allele frequencies versus lengths of shared sequence blocks. *Theor. Popul. Biol.* **87** 105–119.
- BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. [MR1324786](#)
- CHERRY, J. L. (2003). Selection in a subdivided population with local extinction and recolonization. *Genetics* **164** 789–795.

- DURRETT, R. and ZÄHLE, I. (2007). On the width of hybrid zones. *Stochastic Process. Appl.* **117** 1751–1763. [MR2437727](#)
- ETHERIDGE, A. M. (2008). Drift, draft and structure: Some mathematical models of evolution. *Banach Center Publ.* **80** 121–144.
- ETHERIDGE, A., FREEMAN, N. and STRAULINO, D. (2015). The Brownian net and selection in the spatial Lambda-Fleming-Viot Process. Preprint. Available at [arXiv:1506.01158](#).
- ETHERIDGE, A. M. and KURTZ, T. G. (2014). Genealogical constructions of population models. Preprint. Available at [arXiv:1402.6724](#).
- ETHERIDGE, A. M. and VÉBER, A. (2012). The spatial Λ -Fleming-Viot process on a large torus: Genealogies in the presence of recombination. *Ann. Appl. Probab.* **22** 2165–2209. [MR3024966](#)
- ETHERIDGE, A. M., VÉBER, A. and YU, F. (2014). Rescaling limits of the spatial Lambda-Fleming-Viot process with selection. Preprint. Available at [arXiv:1406.5884](#).
- FISHER, R. A. (1937). The wave of advance of advantageous genes. *Ann. Eugenics* **7** 355–369.
- KALLENBERG, O. (2006). *Foundations of Modern Probability*. Springer, Berlin.
- KINGMAN, J. F. C. (1993). *Poisson Processes*. Oxford Univ. Press, New York. [MR1207584](#)
- KRONE, S. M. and NEUHAUSER, C. (1997). Ancestral processes with selection. *Theor. Popul. Biol.* **51** 210–237.
- MARUYAMA, T. (1970). On the fixation probability of mutant genes in a subdivided population. *Genet. Res.* **15** 221–225.
- MUELLER, C., MYTNIK, L. and QUASTEL, J. (2008). Small noise asymptotics of traveling waves. *Markov Process. Related Fields* **14** 333–342.
- NEUHAUSER, C. and KRONE, S. M. (1997). Genealogies of samples in models with selection. *Genetics* **145** 519–534.
- VÉBER, A. and WAKOLBINGER, A. (2015). The spatial Lambda-Fleming-Viot process: An event-based construction and a lockdown representation. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 570–598.
- WHITLOCK, M. C. (2003). Fixation probability and time in subdivided populations. *Genetics* **164** 767–779.

A. ETHERIDGE
S. PENINGTON
D. STRAULINO
DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
24-29 ST GILES
OXFORD
UNITED KINGDOM
E-MAIL: etheridg@stats.ox.ac.uk
sarah.penington@sjc.ox.ac.uk
danielstraulino@gmail.com

N. FREEMAN
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF SHEFFIELD
HOUNSFIELD ROAD
SHEFFIELD
UNITED KINGDOM
E-MAIL: n.p.freeman@sheffield.ac.uk