The Annals of Applied Probability
2018, Vol. 28, No. 2, 651–687
https://doi.org/10.1214/14-AAP1062

THE SHAPE OF MULTIDIMENSIONAL BRUNET–DERRIDA PARTICLE SYSTEMS

BY NATHANAËL BERESTYCKI[1] AND LEE ZHUO ZHAO

University of Cambridge

We introduce particle systems in one or more dimensions in which particles perform branching Brownian motion and the population size is kept constant equal to $N > 1$, through the following selection mechanism: at all times only the N fittest particles survive, while all the other particles are removed. Fitness is measured with respect to some given score function $s : \mathbb{R}^d \to \mathbb{R}$. For some choices of the function s, it is proved that the cloud of particles travels at positive speed in some possibly random direction. In the case where s is linear, we show under some mild assumptions that the shape of the cloud scales like $\log N$ in the direction parallel to motion but at least $(\log N)^{3/2}$ in the orthogonal direction. We conjecture that the exponent $3/2$ is sharp. In order to prove this, we obtain the following result of independent interest: in one-dimensional systems, the genealogical time is greater than $c(\log N)^3$. We discuss several open problems and explain how our results can be viewed as a rigorous justification in our setting of empirical observations made by Burt [*Evolution* **54** (2000) 337–351] in support of Weismann's arguments for the role of recombination in population genetics.

1. Introduction.

1.1. *Main results.* Let $d \geq 1$ and let $s : \mathbb{R}^d \mapsto \mathbb{R}$ denote a fixed function, which we will refer to as the *score* or *fitness* function in what follows. We consider the following system of N particles in $\mathbb{R}^d$, $(X_1(t), \dots, X_N(t))$ defined informally by the following two rules:

- Each particle X_i follows the trajectory of an independent Brownian motion.
- In addition, each particle undergoes binary branching at rate 1. After each branching event, we remove from the population the particle i with minimal score, that is, $\min_{1 \leq i \leq n} s(X_i(t))$.

Note in particular that the population size stays constant (equal to N) throughout time. Unless otherwise specified, we will always order particles $X_1(t), \dots, X_N(t)$ by decreasing fitness, that is, so that

$$s\bigl(X_1(t)\bigr) \geq \cdots \geq s\bigl(X_N(t)\bigr) \tag{1}$$

with arbitrary choice in case of a tie.

Received May 2013; revised August 2014.

[1]Supported in part by EPSRC Grants EP/GO55068/1 and EP/I03372X/1.

MSC2010 subject classifications. 60K35, 92B05.

Key words and phrases. Brunet–Derrida particle systems, branching Brownian motion, random travelling wave, recombination.

This process can be seen as a multidimensional generalisation of the model of branching Brownian motion with selection in $\mathbb{R}$ introduced by Brunet, Derrida, Mueller and Munier [10, 11]. The latter is the model which arises as a particular case of the above description with $d = 1$ and $s(x) = x$.

The motivation for this process was the study of the effect of natural selection on the genealogy of a population. Using nonrigorous methods, Brunet et al. made several striking predictions, which we summarise below. Ordering the particles from right to left [so $X_1(t) \geq \cdots \geq X_N(t)$]:

(i) Then for fixed N, $\lim_{t\to\infty}(X_1(t)/t) = \lim_{t\to\infty}(X_N(t)/t) \stackrel{\text{def}}{=} v_N$, almost surely, where v_N is a deterministic constant.

(ii) As $N \to \infty$, $v_N = v_\infty - c/(\log N)^2 + o((\log N)^{-2})$, where v_∞ is the speed of the rightmost particle in a free branching Brownian motion (or free branching random walk if time is discrete), and c is an explicit constant.

(iii) Finally, the genealogical time scale for this population is $(\log N)^3$. More precisely, the genealogy of an arbitrary sample of the population, resealed by $(\log N)^3$, converges to the Bolthausen–Sznitman coalescent (see, for instance, [6] for definitions and more discussion about this problem).

The arguments of Brunet et al. [10, 11] relied on a nonrigorous analogy with noisy Fisher–Kolmogorov–Petrovskii–Piskounov (FKPP) equation

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2} + u(1-u), \tag{2}$$

and relied strongly on ideas developed earlier by Brunet and Derrida [7–9] on the effect of noise on such an equation. For this reason, this process is sometimes known as the Brunet–Derrida particle system. From a rigorous point of view, proofs of (i) and (ii) can be found in the paper of Bérard and Gouéré [2], while a rigorous proof of (iii) can be found in [4] for a closely related model. However, (iii) remains open for the original Brunet–Derrida process, though exciting progress in this direction has been achieved recently by Maillard [20].

The main goal of this paper is to study geometric properties of the d-dimensional systems and to partly resolve prediction 3 above in the case $d = 1$. We start with our results in d dimensions. Our results are valid in two particular cases:

(Case A) Euclidean case: $s(x_1, \dots, x_d) = \sqrt{x_1^2 + \cdots + x_d^2}$.

(Case B) Linear case: for some vector $\lambda \in \mathbb{R}^d$, $s(x) = \langle \lambda, x\rangle$.

See Figure 1 for two realisations of the process in the Euclidean case (case A). The linear case (case B) is particularly relevant from the point of view of applications, since it is reasonable to assume that when an individual's total fitness depends on $d \geq 2$ loci (i.e., d specific locations along the DNA sequence of the individual), the total fitness of that individual is a linear combination of the fitnesses of each of the loci. In this interpretation, we thus view each coordinate as

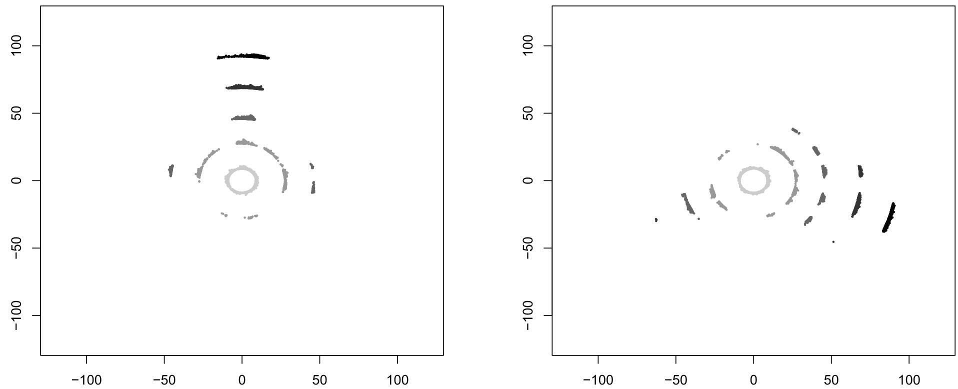

FIG. 1. *Two realisations of the particle system with* $N = 1000$, $d = 2$, $s(x, y) = x^2 + y^2$ *and jump distribution uniform in the unit disk. The particles are plotted after* 20, 60, 100, 150 *and* 200 *generations with decreasing brightness.*

the fitness of the allele on the corresponding locus, and so the "spatial" position has nothing to do with the geographical position of that individual in space. See below for further discussion about the biological relevance of our results.

Simulations in the Euclidean case (case A) suggest that after an initial phase where the particles live in fragmented clusters on a circle of a given radius (which increases at linear speed), particles eventually aggregate in one clump, which travels at that speed in a random direction. A similar phenomenon is observed in simulations for the linear case (case B). Our first result makes this observation rigorous. In order to state it, it is convenient to introduce some notation. If $t > 0$ and $1 \leq n \leq N$, write $X_n(t) = R_n(t)\Theta_n(t)$, where $R_n(t) > 0$ and $\Theta_n(t) \in \mathbb{S}^{d-1}$ is chosen so that following a given ancestral line yields a continuous function $\Theta_n(t)$. [Note that for $d \geq 2$, almost surely $X_n(t) \neq 0$ for all $t > 0$ and $1 \leq n \leq N$.]

THEOREM 1.1. *Let* $N > 1$ *and consider a Brunet–Derrida process in* $\mathbb{R}^d$ *with* N *particles, driven by the Euclidean score function* $s(x) = \|x\|$ (*case A*). *Then*

$$\max_{1 \leq n, m \leq N} \frac{\|X_n(t) - X_m(t)\|}{t} \to 0, \tag{3}$$

as $t \to \infty$ *almost surely.*

Moreover,

$$\frac{R_1(t)}{t} \to v_N, \qquad \Theta_1(t) \to \Theta, \tag{4}$$

where $v_N > 0$ *is a deterministic constant and* Θ *is a random variable on* $\mathbb{S}^{d-1}$. *Both these convergences hold almost surely.*

REMARK 1.2. In the above theorem, (3) says that the particles eventually aggregate in one clump. On the other hand, (4) says that the clump travels at linear speed v_N, in a randomly chosen direction Θ. See Remark 1.6 for the identification of v_N. We will discuss below more precisely the diameter of the cloud of particles, which (for a fixed N, as $t \to \infty$) stays of order one. Note that if the distribution of the initial configuration is rotationally symmetric then Θ is uniformly distributed on $\mathbb{S}^{d-1}$.

REMARK 1.3. This theorem is actually true for a more general class of Brunet–Derrida systems than the ones discussed in this Introduction and, indeed, in much of the paper. See Remark 2.14 for a discussion of the class of processes to which our proofs apply.

We are also able to obtain a lower bound for the correct genealogical time for the one-dimensional process up to some mild conditions on the initial position of the particles.

A similar result holds in the linear case.

THEOREM 1.4. *Let $N > 1$ and consider a Brunet–Derrida process in $\mathbb{R}^d$ with N particles, driven by the linear score function $s(x) = \langle \lambda, x \rangle$ for some $\lambda \in \mathbb{S}^{d-1}$ (case B). Then*

$$\max_{1 \le n,m \le N} \frac{\|X_n(t) - X_m(t)\|}{t} \to 0, \tag{5}$$

as $t \to \infty$ almost surely.

Moreover,

$$\frac{X_1(t)}{t} \to \lambda v_N, \tag{6}$$

almost surely, where $v_N > 0$ is a deterministic constant.

REMARK 1.5. The theorem above says that in this case, the direction of the cloud of particles is deterministic and is simply λ.

REMARK 1.6. It is not hard to see that the v_N appearing in Theorems 1.1 and 1.4 are both equal to the asymptotic speed of a one-dimensional (standard) Brunet–Derrida system. Hence, adapting a result of Bérard and Gouéré [2] for branching Brownian motion, we get

$$v_N = \sqrt{2} - \frac{\pi^2}{\sqrt{2}(\log N)^2} + o\big((\log N)^{-2}\big), \tag{7}$$

as $N \to \infty$. (The asymptotic behaviour of v_N is not needed anywhere in the paper.) See Lemma 2.7 for the proof.

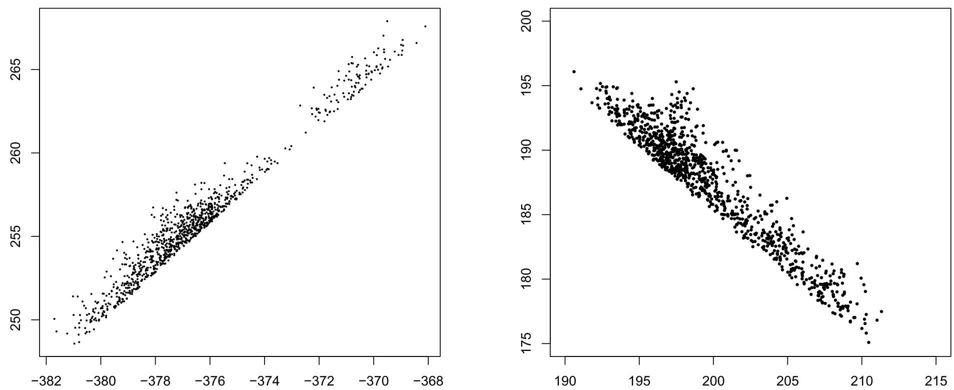

FIG. 2. *Close-up on the cloud of particles with* $N = 1000$. (Left) $s(x, y) = \|x\|$, $t = 1000$. (Right) $s(x, y) = x + y$, $t = 200$.

Our next results concern the dimensions of the cloud of particles. The simulations above suggest, somewhat counter-intuitively, that the cloud of particles is more elongated in the direction orthogonal to the fitness gradient (and the limiting direction of the cloud). This is corroborated by a close-up view of the cloud of particles (see Figure 2).

We are able to establish this phenomenon under some reasonable assumptions on the initial condition, in case B. Fix $\lambda \in \mathbb{S}^{d-1}$ and let $H = \lambda^\perp$ be the orthogonal hyperplane. Let p_H denote orthogonal projection onto H. Define

$$\mathrm{diam}_t = \max_{1\le m,n\le N} \left|\langle X_n(t) - X_m(t), \lambda\rangle\right|$$

and

$$\mathrm{diam}_t^\perp = \max_{1\le m,n\le N} \left\| p_H\big(X_n(t)\big) - p_H\big(X_m(t)\big)\right\|.$$

Finally, for all $x \in \mathbb{R}^d$ let $\widehat{x} = \langle \lambda, x\rangle$.

We introduce an assumption on the initial condition which will be used in several results below. Let $X_1(t), \dots, X_N(t)$ denote the particles of a Brunet–Derrida system driven by the linear score function $s(x) = \widehat{x}$. Let $\widehat{X}_n(t) = \langle X_n(t), \lambda\rangle$, and label the particles by decreasing fitness $\widehat{X}_1(t) \ge \cdots \ge \widehat{X}_N(t)$. Let $x_0 \in \mathbb{R}^d$ be arbitrary, $\mu \le \sqrt{2}$, and suppose that for some $\delta < 1$,

$$\sum_{n=1}^{N} e^{\mu(\widehat{X}_n(0) - \widehat{x}_0)} \le N^\delta. \tag{8}$$

THEOREM 1.7. *Suppose that* (8) *holds with* $\mu = \sqrt{2}$ *and that there is initially a particle with fitness greater than* $\widehat{x}_0$. *Then there exists* $c_\delta > 0$ (*depending only on* δ) *such that for* $t = c_\delta(\log N)^3$, *there exists* $a > 0$ *such that*

$$\liminf_{\eta\to 0} \liminf_{N\to\infty} \mathbb{P}\big(\mathrm{diam}_t \le a \log N, \mathrm{diam}_t^\perp \ge \eta(\log N)^{3/2}\big) = 1. \tag{9}$$

The phenomenon above has consequences in population genetics which are discussed below. Note that in fact we expect that under the same initial condition, the order of magnitude of diam_t really is $\log N$, in the sense that we also have $\mathrm{diam}_t \geq a' \log N$ with probability tending to 1 as $N \to \infty$, for some constant $a' < a$.

We now make a series of comments on the meaning of the initial condition (8).

REMARK 1.8. Intuitively, condition (8) is likeliest to hold when x_0 is the position of a maximal particle. Then (8) says that, after projecting onto $\mathrm{Span}(\lambda)$, for some $\xi < 1$, only $O(N^{\xi})$ particles lie within distance $O(\log N)$ of the maximal particle. More precisely, (8) holds as soon as there exists $c > 0$ and $\xi < 1$ such that at most N^{ξ} particles lie in the interval $[\widehat{X}_1(0) - c\log N, \widehat{X}_1(0)]$.

REMARK 1.9. An example of an initial condition which satisfies (8) with high probability for any $\mu > 0$ and x_0 being the position of the maximal particleis as follows: sample $X_1, \dots, X_N$ in $\mathbb{R}^d$ independently according to a fixed distribution such that if $\widehat{X} = \langle X, \lambda \rangle$, then for all $x > 0$,

$$c_1 e^{-\alpha_1 x} \leq \mathbb{P}(\widehat{X} > x) \leq c_2 e^{-\alpha_2 x} \tag{10}$$

for some positive constants c_1, c_2 and α_1, α_2.

REMARK 1.10. We believe, but have been unable to prove, that if the initial condition is as in the above remark then (8) will in fact be satisfied at arbitrary large times with $\mu = \sqrt{2}$. Indeed, comparing with results in [4], we expect indeed that, at "equilibrium" (see Section 1.2 for definition), $\widehat{X}_1(0) = (1/\sqrt{2}) \log N$ and

$$Y_N = \sum_n e^{\sqrt{2}\widehat{X}_n(0)} \approx NL \int_0^L e^{\sqrt{2}x} \cdot e^{-\sqrt{2}x} \sin\left(\frac{\pi x}{L}\right) dx \sim cNL^2,$$

where $L = (1/\sqrt{2})(\log N + 3 \log\log N)$. Hence taking $x_0 = X_1(0)$, the right-hand side of (8) should be of order L^2, and thus (8) should be satisfied at equilibrium, with lots of room to spare. Thus condition (8) can be thought of as a very weak condition specifying that the population is in an approximately "metastable" state, as in [4].

As we will see, the result in Theorem 1.7 is closely related to estimates about the genealogical timescale (or, more precisely, the time of the most recent common ancestor) in the population. In fact, Theorem 1.7 will follow easily from the following result.

THEOREM 1.11. *Let $N > 1$ and consider a Brunet–Derrida system with N particles driven by the linear score function $s(x) = \widehat{x} = \langle x, \lambda \rangle$. Assume that* (8) *holds with $\mu = \sqrt{2}$ and some $x_0 \in \mathbb{R}^d$. Then there exists $c_\delta > 0$* (*depending only on δ*) *such that any particle with fitness greater than $\widehat{x}_0$ at time 0 has descendants alive at time $c_\delta (\log N)^3$ with probability tending to 1 as $N \to \infty$.*

By projecting the particle system onto Span(λ), we obtain a one-dimensional (standard) Brunet–Derrida system. Thus Theorem 1.11 applies *verbatim* to such systems, which partly confirms a prediction of [10, 11] [see item (iii) at the start of the Introduction].

The heart of the proof relies on delicate quantitative estimates concerning the displacement of the minimal position in one-dimensional (standard) Brunet–Derrida systems. This is a difficult quantity to study rigorously, as the evolution of the minimum depends on all the particles nearby, which make up all but a negligible fraction of the population. In particular, as a process it is non-Markovian in its own filtration, and not continuous, though in the limit $N \to \infty$ it should become deterministic and continuous. Our result is as follows.

PROPOSITION 1.12. *Consider a (standard) one-dimensional Brunet–Derrida system with N particles, ordered by decreasing fitness $X_1(t) \geq \cdots \geq X_N(t)$. Assume that the initial condition satisfies (8) for some arbitrary $x_0 \in \mathbb{R}$, and with μ given by*

$$\mu = \sqrt{2 - \frac{2\pi^2}{(\log N)^2}}. \tag{11}$$

Then there exists constants $c_\delta > 0$, c, and $\kappa > 0$ (depending only on δ) such that

$$\mathbb{P}\big(X_N(t) \leq x_0 + \mu t, \forall t \leq c_\delta (\log N)^3\big) \geq 1 - cN^{-\kappa}. \tag{12}$$

We note that a corresponding lower bound for the progression of the minimal position can be established from an intermediate result of Bérard and Gouéré [2], with their proof adapted for branching Brownian motion.

PROPOSITION 1.13. *Consider a (standard) one-dimensional Brunet–Derrida system with N particles, ordered by decreasing fitness $X_1(t) \geq \cdots \geq X_N(t)$. For all $\eta > 0$, there exists $c_\eta > 0$ such that for any initial condition as $N \to \infty$,*

$$\mathbb{P}\left(X_N(t) - X_N(0) \leq \left(\sqrt{2} - \frac{(1+\eta)\pi^2}{\sqrt{2}(\log N)^2}\right)t, \forall t \leq c_\eta (\log N)^3\right) \to 0. \tag{13}$$

1.2. *Discussion and open problems*. *Long term behaviour for general fitness functions.* Theorems 1.1 and 1.4 establish the long-term behaviour for the cloud of particles for the two special cases where the function s is either the Euclidean norm or a linear function. In both cases, the cloud escapes to ∞ at positive speed in a possibly random direction. It would be interesting to see how general a phenomenon this is. For instance, assume that $s : \mathbb{R}^d \to \mathbb{R}$ is a smooth, unbounded convex function. What can be said about the long-term behaviour then? One first observation is that the cloud of particles should essentially stay concentrated on level sets of the function s.

Genealogy. In both cases studied here (Euclidean case or case A, and linear case or case B), we observe that the population lines up on an essentially one-dimensional subspace of $\mathbb{R}^d$. For truly one-dimensional systems, it is predicted that the Bolthausen–Sznitman coalescent describes the genealogy of a finite sample of size k from the population, after rescaling time by $(\log N)^3$ (where k is fixed and $N \to \infty$). It is therefore reasonable to predict that the same property will hold in higher dimensions as well, at least in cases A and B and perhaps more generally as well, suggesting that the Bolthausen–Sznitman coalescent is a universal scaling limit in all dimensions, subject to assumptions on the function s.

It is also natural to ask about the time $\tau_N(t)$ to the most recent common ancestor for the entire population at some large time t, and not just for a subset of it. If we extrapolate the Bolthausen–Sznitman prediction to the entire population, and using a result of Goldschmidt and Martin [15] (see also Theorem 6.5 in [6]), we conjecture that

$$\lim_{N\to\infty}\lim_{t\to\infty}\frac{\tau_N(t)}{(\log N)^3\log\log N}=c \tag{14}$$

exists in probability.

Equilibrium shape in one dimension. Consider the empirical distribution of a (standard) one-dimensional Brunet–Derrida particle system:

$$\nu_t^N=\frac{1}{N}\sum_{n=1}^N\delta_{X_n^N(t)},$$

and the associated càdlàg empirical tail distribution

$$F^N(t,x)=\frac{1}{N}\sum_{n=1}^N\mathbf{1}\{X_n^N(t)>x\}=\nu_t^N\big((x,\infty)\big).$$

It is not hard to see that the system of particles, viewed from the minimum position at time t, has regeneration times where the whole population descends from the leader in a bounded interval of time. This is a highly atypical event, but which nonetheless happens with positive frequency for a fixed N as $t\to\infty$; see Lemma 2.8. It can be shown that this implies that $(X_1-X_N(t),\dots,X_N-X_N(t);t\ge 0)$ is a Harris positive recurrent chain; see Proposition 3.1 in [13] for a proof in the discrete case of branching random walk. (It is not hard to adapt these arguments to our branching Brownian motion setup. However, we do not include the details here as it would unnecessarily burden the exposition and we will not use this fact anywhere in this paper.) Therefore, for all fixed $N>1$, $F^N(t,x+X_N(t))$ converges pointwise to some limit distribution $F_{\mathrm{eq}}^N(x)$ as $t\to\infty$, wherever F_{eq}^N is continuous. It is natural to start the particle system in some initial condition distributed according to F_{eq}^N and ask for its properties. We believe, but have been unable to prove, that F_{eq}^N satisfies (8). In fact, we make the following conjecture about F_{eq}^N.

Reasoning by analogy with the results of Durrett and Remenik [13], and using the martingale problem for the empirical distributions of a free branching Brownian motion (see, e.g., Lemma 1.10 in Etheridge [14]), we expect $F^N(t,x)$ to converge in distribution to $F(t,x)$, the solution to the free boundary problem

$$\begin{cases} \dfrac{\partial F}{\partial t} = \dfrac{1}{2}\dfrac{\partial^2 F}{\partial x^2} + F(t,x) & \forall x > \gamma(t), \\ F(t,x) = 1 & \forall x \le \gamma(t), \end{cases} \tag{15}$$

where $\gamma : [0,\infty) \to \mathbb{R}$ is a continuous, increasing function starting from 0, which is part of the unknown in (15). (Note that Durrett and Remenik's argument breaks down for particles that perform Brownian motion, as it is essential in their coupling that particles sit still in between branching events. It is unclear how to adapt their argument to the case of Brownian motion.) The first equation is simply the linearised FKPP equation (2), which is satisfied asymptotically as $x \to \infty$ by the distribution tail of the position of the rightmost particle in a (free) branching Brownian motion. The second equation on the other hand represents the effect of selection, and $\gamma(t)$ then describes the limiting position of the minimal particle. Durrett and Remenik (2011) show the existence of a family of travelling wave solutions for a class of problems similar to (15). Here, the traveling wave solutions can be found explicitly: if $F(t,x) = W(x-ct)$ solves (15), we find

$$-cW' = \frac{1}{2}W'' + W.$$

This is a second-order differential equation which, as is well known, has positive solutions only if the speed c of the traveling wave satisfies $c \ge \sqrt{2}$. For $c = \sqrt{2}$, the solution is

$$W_*(x) = (\sqrt{2}x + 1)e^{-\sqrt{2}x}. \tag{16}$$

Turning back to F^N_{eq}, stationarity suggests that F^N_{eq} is in the limit as $N \to \infty$ a traveling wave solution of (15). But by Proposition 1.12 if F^N_{eq} is a travelling wave solution the speed would have to be at most $\sqrt{2}$, and so equal to $\sqrt{2}$. Therefore, we conjecture that

$$F^N_{\mathrm{eq}}(x) \to W_*(x) \tag{17}$$

uniformly on compact sets as $N \to \infty$. A referee has noted that this conjecture independently already appeared in the work of Maillard [19], p. 19, and Groisman and Jonckheere [16], Conjecture 3.1, which appeared on the arxiv immediately before this article.

Equilibrium shape in high dimensions. Let $d \ge 1$ and fix an arbitrary smooth selection function s. For reasons similar to above, it is possible to define a notion of limiting equilibrium shape of the system as $t \to \infty$. Theorem 1.7 gives information about the dimensions (width and length) of the limiting shape in case B.

However, an inspection of the simulations suggests that particles are far from uniformly distributed within that shape. In the direction λ, we expect the density of particles to be close to $W_*(x)$ for the same reasons as above. In the transverse direction $\lambda^\perp$, however, particles appear somewhat "clustered". Indeed, this is to be expected given the hierarchical structure of the Bolthausen–Sznitman coalescent. Clusters of particles represent groups of particles coming from a close common ancestor. However, clusters are also intertwined because of heat kernel smoothing. It is an interesting question to identify the density of particles at equilibrium.

1.3. *Biological applications*: *The effect of recombination*. As alluded to in earlier parts of this Introduction, our Brunet–Derrida system in more than one dimension can be thought of as a model for the effect of selection on multiple linked loci. In this interpretation, we track the fitness of not one but d loci in a population of size N. Each particle corresponds to one-half of an individual's genetic material, and each of the d coordinates of that particle represents the fitness at the corresponding locus. Her total fitness will then be a function of these d values, typically just the sum. In this interpretation, we are assuming that the total fitness of each particle evolves like independent Brownian motions and branch independently of one another, which is a simplification because in reality, two particles—making up one individual—will branch simultaneously. For the same reasons, whereas in our model we only remove one particle at a time, it would make more sense to remove two particles at once (also making up an individual). But we choose to ignore the correlations between an individual's two genetic halves, and still believe that the model captures some important features of reproduction. Note that, as specified above, the model ignores the possibility of recombination. We will now explain the effect of adding (a small amount of) recombination to the model and show that it leads to an increase in overall fitness through an interesting and indirect mechanism.

It has been a longstanding problem in evolutionary biology to explain the ubiquitous nature of diploid populations over haploid populations. We very briefly summarise what is a highly complex issue below. In diploid populations, the chance of a particular gene being transmitted to an offspring is only 50%, whereas it is 100% in haploid populations. This would suggest that haploid populations are far more advantageous from the point of view of a particular gene. This paradox was in fact raised soon after the introduction of Darwin's theory of natural selection and evolution.

As early as 1889, Weismann [22] advocated that sex functions to provide variation for natural selection to act upon. However, it is fair to say that no real consensus was achieved in the population genetics community, especially after influential arguments by Williams [23] raised doubts on Weismann's theory. The controversy reached the point where understanding the advantage of sexual reproduction became the "queen of problems in evolutionary biology" [1]. We refer to Burt [12] for an excellent and highly readable survey of this question.

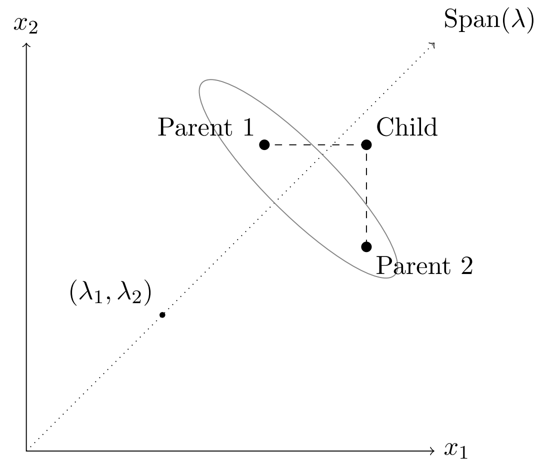

FIG. 3. *In the presence of recombination, the offspring of two individuals with positions* (x_1, y_1) *and* (x_2, y_2) *is either* (x_1, y_2) *or* (x_2, y_1). *If the fitness across loci is negatively correlated* (*i.e., if the shape of the cloud is elongated in the direction transverse to fitness gradient*), *this leads to an overall increase in the variance of the fitness distribution, even though the mean in unchanged. In turn, this results in increased response to natural selection.*

In his study of the problem, Burt [12] observed empirically that his models led to a negative correlation between the fitness on the two chromosomes, which is equivalent to a cloud of particles being spread out in the direction orthogonal to the fitness gradient (see Figure 1D of [12]). He then reasoned that a small amount of recombination would lead to a reduction in this correlation and greater variance in the overall fitness, ultimately leading to a fitter population, as can be seen on Figure 3. Here, the implicit assumption is that recombination is sufficiently rare that we can consider these events one at a time (and in between two such events the population has the time to come back to equilibrium). Note that after recombination the average fitness of all particles is the same as immediately before recombination. However, the variance is now greater. In particular, in our Brunet–Derrida setting the top-right particle in Figure 3 will typically generate the whole population, relatively quickly. Indeed, by Theorem 1.7 the top-right particle will be at a distance of order $(\log N)^{3/2}$ from the minimal particle, after projection onto $\mathrm{Span}(1, 1)$. This is much more than is needed to regenerate the population (in fact, an advance of $(1/\sqrt{2})(\log N + 3 \log\log N + A)$ is presumably enough, for A a large constant. See [4] for a rigorous proof in the slightly different but related context of branching Brownian motion with absorption). A single recombination event thus eventually generates a shift in the overall fitness of the population of size roughly $(\log N)^{3/2}$.

It is not our purpose here to explain the models considered by Burt [12] and in what way the Brunet–Derrida systems of this paper are (or are not) related. We will simply observe that the Brunet–Derrida systems are also models of natural selection and our Theorem 1.7, in the case $s(x, y) = x + y$, can be seen as a rigorous justification in our setting of the empirical observations made in [12] to justify the Weissmannian proposal. Naturally, this deserves further critical investigation.

2. Asymptotic direction: Proof of Theorems 1.1 and 1.4. We are now ready to give a proof of Theorems 1.1 and 1.4. In this section $N > 1$ is fixed.

We first begin with a formal construction of the Brunet–Derrida particle system. Let $(J_i)_{i\geq 0}$ be the jump times of a Poisson process with rate N with $J_0 = 0$, and let $(K_i)_{i\geq 1}$ be an independent sequence of i.i.d. uniform random variables on $\{1, \dots, N\}$. The process is started in some given initial condition. Then inductively, for each $i \geq 1$, assuming that the system is defined up to time J_{i-1} with $s(X_1(J_{i-1})) \geq \cdots \geq s(X_N(J_{i-1}))$, we define

$$X_n(t) = X_n(J_{i-1}) + Z_n^i(t - J_{i-1}), \qquad t \in [J_{i-1}, J_i), \tag{18}$$

where $(Z_n^i(t), t \geq 0)$ are independent Brownian motions in $\mathbb{R}^d$, independent from (K_i) and (J_i). At time J_i, we duplicate particle $X_{K_i}(J_i^-)$ and remove the particle $\min_{1\leq n\leq N} s(X_n(J_i^-))$. Note that if the duplicated particle is the particle of minimal score, the net effect is that nothing happens. We now relabel the particles over this interval in the usual convention of descending fitness so

$$s\big(X_1(t)\big) \geq \cdots \geq s\big(X_N(t)\big), \qquad t \in [J_{i-1}, J_i].$$

2.1. *Proof of Theorem* 1.4. We start with a few elementary facts about (free) branching Brownian motion $\overline{X}_1(t), \dots, \overline{X}_{\overline{N}(t)}(t)$ in $\mathbb{R}$, where $\overline{N}(t)$ is the number of particles at time t. In keeping with our convention for this article, we order particles from right to left. We assume that initially there is one particle at the origin.

The following lemma is a trivial but useful result to relate the statistics for all the particles alive in a free branching Brownian motion to a single Brownian motion and is sometimes known in the literature as the many-to-one lemma (see, e.g., [17]).

LEMMA 2.1. *Let T be a random stopping time of the filtration $\overline{\mathcal{F}}_t = \sigma(\overline{X}_i(s), i \leq \overline{N}(s), s \leq t)$, and assume that T is almost surely finite. For $s < T$ and each $i \leq \overline{N}(T)$, let $\overline{Y}_i(s)$ be the position of the unique ancestor of $\overline{X}_i(T)$. Then for any bounded measurable functional g on the path space $C([0, \infty))$,*

$$\mathbb{E}\Big[\sum_{i\leq \overline{N}(T)} g\big((\overline{Y}_i(s))_{s\leq T}\big)\Big] = \mathbb{E}\big[e^T g\big((B_s)_{s\leq T}\big)\big],$$

where $(B_s)_{s\geq 0}$ is a standard Brownian motion.

With the many-to-one lemma, we can obtain a naive bound for the maximum displacement of a particle at time t from its parent at time 0, as well as the running maximum.

LEMMA 2.2. *For any* $K > 0$,

$$\mathbb{P}(\overline{X}_1(t) \geq \sqrt{2}t + K) \leq e^{-\sqrt{2}K}.$$

Moreover,

$$\mathbb{P}\Big(\sup_{s \leq t} \overline{X}_1(s) \geq \sqrt{2}t + K\Big) \leq 2e^{-\sqrt{2}K}.$$

PROOF. By Lemma 2.1,

$$\begin{aligned}\mathbb{P}(\overline{X}_1(t) \geq \sqrt{2}t + K) &\leq \mathbb{E}\Big[\sum_{i \leq \overline{N}(t)} \mathbf{1}\{\overline{X}_i(t) \geq \sqrt{2}t + K\}\Big] \\ &= e^t \mathbb{P}(B_t \geq \sqrt{2}t + K) \\ &\leq e^t e^{-\frac{1}{2t}(\sqrt{2}t+K)^2} \leq e^{-\sqrt{2}K},\end{aligned}$$

where we use the well-known tail bound for a standard normal random variable X, and $a > 0$,

$$\mathbb{P}(X \geq a) \leq e^{-\frac{a^2}{2}}, \tag{19}$$

which will be used throughout. For the historic maximum, a similar argument shows

$$\mathbb{P}\Big(\sup_{s \leq t} \overline{X}_1(s) \geq \sqrt{2}t + K\Big) \leq e^t \mathbb{P}\Big(\sup_{s \leq t} B_s \geq \sqrt{2}t + K\Big).$$

Using the reflection principle,

$$\mathbb{P}\Big(\sup_{s \leq t} B_s \geq \sqrt{2}t + K\Big) = 2\mathbb{P}(B_t \geq \sqrt{2}t + K),$$

and the result follows. $\square$

When a Brunet–Derrida system is driven by a score function $s(x) = \langle x, \lambda \rangle$, where $\lambda \in \mathbb{S}^{d-1}$, we have already noted that after projecting the particle system onto $\mathrm{Span}(\lambda)$, we recover a standard one-dimensional Brunet–Derrida system. For such systems, we have an easy but useful coupling used by Bérard and Gouéré (Lemma 1 of [2]) in the discrete setup of branching random walk.

LEMMA 2.3. *Consider two* (*standard*) *one-dimensional Brunet–Derrida systems*, $(X_n(t), 1 \leq n \leq N)_{t \geq 0}$ *and* $(Y_n(t), 1 \leq n \leq N')_{t \geq 0}$, $N \leq N'$, *which are initially ordered* $X(0) \prec Y(0)$ *in the sense of stochastic domination*: *that is*, *there is a coupling of* $X(0)$ *and* $Y(0)$ *such that*

$$Y_1(0) \geq X_1(0); \dots; Y_N(0) \geq X_N(0).$$

Then we can couple $X(t)$ *and* $Y(t)$ *for all times* $t \geq 0$ *in such a way that* $X_i(t) \leq Y_i(t)$ *for all times* $t \geq 0$ *and all* $1 \leq i \leq N$.

PROOF. Given the initial coupling of $X(0)$ and $Y(0)$ as above, we couple $X(t)$ and $Y(t)$ for all times using (18), as follows. We use the same family $(Z_n^i(t), t \geq 0)_{n \leq N', i \geq 1}$ of independent Brownian motions in $\mathbb{R}$, we fix J_i' the jump times of a Poisson process with rate N', and K_i' which are uniform on $\{1, \dots, N'\}$ and independent. At a jump time J_i', if $K_i' = k \in \{1, \dots, N\}$ then we split the kth particle in both $X(t)$ and $Y(t)$, while if $K_i' = k > N$ we split the kth particle only in $Y(t)$ and nothing happens for $X(t)$. It is easy to check that no particle of $X(t)$ can overtake a particle of smaller label in $Y(t)$ at any time, by construction, and hence $X_i(t) \leq Y_i(t)$ still holds for all time and for all $1 \leq i \leq N$. $\square$

Adapting the (easy) proof of Proposition 2 of [2], one obtains the following.

LEMMA 2.4. *Consider a one-dimensional Brunet–Derrida system initially with* $X_1(0) \geq \cdots \geq X_N(0) = 0$. *Then*

$$\frac{X_1(t)}{t} \to v_N,$$

almost surely and in L^1, *where* $v_N > 0$ *is a deterministic constant.*

The argument is based on the monotonicity of Lemma 2.3 and Kingman's subadditive ergodic theorem. To see that $v_N > 0$ for $N > 1$, we observe that $v_N \geq v_2$ by Lemma 2.3 by monotonicity in N. Moreover, v_2 is easily shown to be strictly positive. Indeed, if J_n is the nth jump time of the system, then $X_1(J_n) = \sum_{i=1}^n Z_i$, where the random variables Z_i are i.i.d. and are distributed according $Z_i =_d \max(B_1(T), B_2(T))$, for two independent standard Brownian motions B_1, B_2 and an independent exponential random variable T with rate 2. Since $\mathbb{E}(Z_i) > 0$, it follows that $v_2 > 0$, and hence $v_N > 0$ for $N \geq 2$.

The same argument also applies to $X_N(t)$, but a priori the limiting velocity v_N' might be distinct from v_N. In fact, the following lemma, which can be proved in the same fashion as Proposition 1 of [2], shows that $v_N = v_N'$.

LEMMA 2.5. *Let* $(X_n(s), 1 \leq n \leq N)_{s \geq 0}$ *be a* (*standard*) *one-dimensional Brunet–Derrida system. Then for all* $\varepsilon > 0$ *and* $t > (1 + \kappa) \log N$ *for some* $\kappa > 0$,

$$\lim_{N \to \infty} \mathbb{P}\bigl(X_1(t) - X_N(t) \geq (3\sqrt{2} + \varepsilon) \log N\bigr) = 0.$$

COROLLARY 2.6. *For all* $N > 1$ *and* $\varepsilon > 0$,

$$\lim_{s \to \infty} \mathbb{P}\biggl(\frac{X_1(s) - X_N(s)}{s} \geq \varepsilon\biggr) = 0$$

and as $t \to \infty$, $X_N(t)/t \to v_N$ *almost surely and in* L^1.

We now return to Brunet–Derrida systems in $\mathbb{R}^d$, and start the proof of Theorem 1.4. Let $H = \{x \in \mathbb{R}^d : \langle x, \lambda \rangle = 0\}$ be the orthogonal hyperplane to λ and let p_H be the orthogonal projection onto H. In the rest of the article, we will sometimes, with a slight abuse of language, refer to a particle by its position, say $X_n(t)$. It should be clear from the context what is meant.

Referring back to the construction of the system via (18), conditional on $\mathcal{F}_{J_{i-1}}$ (where $\mathcal{F}_t$ is the filtration generated by the whole system up to time t), particles perform $(d-1)$-dimensional Brownian motion on H independent of the motion in $\mathrm{Span}(\lambda)$ up to time J_i for every $i \geq 1$. Moreover, since $s(x) = \langle x, \lambda \rangle$, $p_H(X_m(J_i))$ is independent of the event that the particle X_m survives a branching event at time J_i. Together, these two properties imply by induction that the path of a particle conditioned to survive until time t when projected onto H has the law of a standard $(d-1)$-dimensional Brownian motion. In other words, if $X_n(t)$ is a surviving particle at time t and $Y_n(s)$ is the ancestor of $X_n(t)$ at time $s \leq t$, then $(p_H(Y_n(s)), s \leq t)$ is a standard $(d-1)$-dimensional Brownian motion in H. Hence, for some constant c depending only on the dimension, using (19),

$$\mathbb{P}\Big(\sup_{1 \leq n \leq N} \sup_{s \leq t} \|Y_n(s)\|_H \geq \delta t\Big) \leq cNe^{-(\delta^2/2)t},$$

where $\|x\|_H = \|p_H(x)\|$, and so the same inequality holds with $Y_n(s)$ replaced by $X_n(s)$. The right-hand side of the inequality is summable when $t = 1, 2, \ldots$; hence by the Borel–Cantelli lemma,

$$\sup_{1 \leq n,m \leq N} \frac{\|X_n(t) - X_m(t)\|_H}{t} \leq 2 \sup_{1 \leq n \leq N} \frac{\|X_n(t)\|_H}{t} \to 0,$$

almost surely as $t \to \infty$. Together with Lemma 2.4, this completes the proof of Theorem 1.4.

2.2. *Proof of Theorem* 1.1. Assume now $s(x) = \|x\|$. Recall in this setting, for $X_n(t) \neq 0$, we write $X_n(t) = R_n(t)\Theta_n(t)$ where $R_n(t) > 0$ and $\Theta_n(t) \in \mathbb{S}^{d-1}$ is continuous whenever $X_n(t)$ is continuous. Note also for $d \geq 2$, d-dimensional Brownian motion almost surely never hits 0. Hence, except for particles initially at 0, this decomposition is always well defined. We can work around particles starting from 0 by instead taking the system at time $t > 0$ as its initial state without altering the proofs. Therefore, without loss of generality, we shall assume from here on that $R_N(0) > 0$ and we need not worry about any particles at 0.

When considering $(R_n(t), 1 \leq n \leq N)$, we can work in a one-dimensional setting and construct the system in a similar manner as before, except now the displacement step (18) becomes

$$R_n(t) = R_n(J_{i-1}) + S_n^i\big(R_n(i-1), t - J_{i-1}\big), \qquad t \in [J_{i-1}, J_i), \tag{20}$$

where $(S_n^i(r,t), 1 \leq n \leq N)_{r \geq 0}$ are independent Bessel flows: that is, for each fixed $r \geq 0$, $i \geq 1$ and $1 \leq n \leq N$, $S_n^i(r,t)$ is the pathwise unique strong solution of the stochastic differential equation

$$dS_n^i(r,t) = dB_n^i(t) + \frac{d-1}{2S_n^i(r,t)}\,dt, \tag{21}$$

where $B_n^i(t)$ is a standard one-dimensional Brownian motion, and $S_n^i(r,t)$ is a solution starting from $S_n^i(r,0) = r$. Using monotonicity of the solution at any fixed time with respect to r (or the easily established continuity of this solution with respect to r), this gives a well-defined Bessel flow for all $r \geq 0$ simultaneously.

Recall that the default ordering is in descending fitness, and so $R_1(t) \geq \cdots \geq R_N(t)$.

LEMMA 2.7. *For all $N > 1$,*

$$\frac{R_1(t) - R_N(t)}{t} \to 0,$$

as $t \to \infty$ almost surely. Moreover,

$$\frac{R_1(t)}{t} \to v_N,$$

almost surely as $t \to \infty$, where v_N is the speed of a one-dimensional Brunet–Derrida particle system (defined in Lemma 2.4).

PROOF. Given $(R_n(t), 1 \leq n \leq N)$ constructed in the usual manner and with (20), we define a family of one-dimensional Brunet–Derrida systems $(Y_n^{i,\varepsilon}(t), 1 \leq n \leq N)$ constructed in the same manner as $(R_n(t), 1 \leq n \leq N)$ with the same (J_i), (K_i), but with the displacement step

$$Y_n^{i,\varepsilon}(t) = Y_n^{i,\varepsilon}(J_{i-1}) + W_n^{i,\varepsilon}(t - J_{i-1}), \qquad t \in [J_{i-1}, J_i), \tag{22}$$

where $(W_n^{i,\varepsilon}(t), 1 \leq n \leq N)$ are independent Brownian motions in $\mathbb{R}$ with drift ε. These processes satisfy the stochastic differential equation

$$dW_n^{i,\varepsilon}(t) = dB_n^i(t) + \varepsilon\,dt. \tag{23}$$

Suppose we couple the family $(Y_n^{\varepsilon}(t), 1 \leq n \leq N)$ to $(R_n(t), 1 \leq n \leq N)$ by using the same underlying $B_n^i(t)$ to drive the solutions to (21) and (23) for each i and n. Then we see that under this coupling

$$\liminf_{t\to\infty} \frac{R_N(t)}{t} \geq \liminf_{t\to\infty} \frac{Y_N^0(t)}{t}.$$

But $(Y_n^0(t), 1 \leq n \leq N)$ is a standard one-dimensional Brunet–Derrida system and by Lemma 2.4,

$$\lim_{t\to\infty} \frac{Y_N^0(t)}{t} = v_N$$

almost surely, where $v_N > 0$. Therefore, almost surely, $R_N(t) \to \infty$. It follows that, for all $\varepsilon > 0$, the drift terms in (21), namely $(d-1)/(2S_n^i(t))$, are uniformly bounded above by ε for t sufficiently large. Hence, in this coupling for every $\varepsilon > 0$,

$$\limsup_{t\to\infty} \frac{R_1(t)}{t} \leq \limsup_{t\to\infty} \frac{Y_1^\varepsilon(t)}{t}$$

almost surely. However, we note that if the drift is a constant equal to ε, then $Y_i^\varepsilon(t) = Y_i^0(t) + \varepsilon t$ for all $1 \leq i \leq N$ and all $t \geq 0$, hence by Lemma 2.4,

$$\lim_{t\to\infty} \frac{Y_1^\varepsilon(t)}{t} = v_N + \varepsilon,$$

almost surely. Since $\varepsilon > 0$ was arbitrary, we have that almost surely

$$\lim_{t\to\infty} \frac{R_1(t)}{t} = \lim_{t\to\infty} \frac{R_N(t)}{t} = v_N. \qquad \square$$

Having established the asymptotic behaviour of $R(t)$, we now turn our attention to $\Theta(t)$. The main idea here is that the time to the most recent common ancestor for all N particles, can be naively dominated uniformly over all time. We shall formalise this statement with the following lemma. Let $\tau(t)$ be the time to the most recent common ancestor for $X_1(t), \dots, X_N(t)$, that is,

$$\begin{aligned}\tau(t) = \inf\{s \geq 0 :\ &\text{all particles at time } t \text{ descend from} \\ &\text{a single ancestor at time } t-s\}.\end{aligned} \tag{24}$$

If the above set is empty, then by convention we define $\tau(t) = t$. Occasionally, in what follows we will drop the dependence on t, and simply write τ instead of $\tau(t)$ to ease readability.

LEMMA 2.8. *For all t sufficiently large, $\tau(t) - 1$ is stochastically dominated by a geometric random variable of parameter p, where $p > 0$.*

REMARK 2.9. The number p which we obtain from the proof is extremely small. In reality, $\tau(t)$ is likely to be much smaller than suggested by this lemma. Indeed, we believe that for large t and large N, $\tau(t)$ is of order $(\log N)^3 \log\log N$ [see (14) for more details].

PROOF. Let $s \geq 0$, and consider the system $X_1(s), \dots, X_N(s)$ at time s. We assume that the Brunet–Derrida system $(X_n(t), 1 \leq n \leq N)_{t \geq s}$ is obtained from a free branching Brownian motion $(\overline{X}_n(t), 1 \leq n \leq \overline{N}(t))_{t\geq 0}$ in the obvious manner, that is, by enforcing selection at all times. Let A_s be the event that, for this free process, when the particle located at $X_1(s)$ at time s first branches after time s, its score is $\geq R_1(s) + 1$, and it subsequently produces at least N offsprings by time $s+1$ whose scores stay above $R_1(s) + 1/2$ throughout the interval $[s, s+$

1]. Let B_s be the event that for the free process, the particles initially located at $X_2(s), \dots, X_N(s)$ do not branch between times s and $s+1$, and that

$$\sup_{2 \le n \le N} \sup_{t \in [s, s+1]} \|Y_n(t) - Y_n(s)\| \le 1/2, \tag{25}$$

where $Y_n(t)$ is the location at time t of the descendant of the particle located at $X_n(s)$ at time s. Note that $Y_n(s)$ is well defined since $X_n(s)$ has a unique descendant at all times in $[s, s+1]$ for $2 \le n \le N$.

Note that A_s and B_s are independent events. Moreover, B_s is independent of $(X_n(s))_{1 \le n \le N}$, so there exists $p_2 > 0$ such that

$$\mathbb{P}(B_s | \mathcal{F}_s) = \mathbb{P}(B_s) \ge p_2$$

almost surely for all s, where $\mathcal{F}_s$ denotes the filtration generated by the entire process up to time s. Likewise, A_s given $R_1(s)$ is independent of $\mathcal{F}_s$. To lose the dependence on $R_1(s)$, we use an analogous coupling as in the proof of Lemma 2.7 where we stochastically bound $(R_n(t), 1 \le n \le N)$ from below by a standard one-dimensional Brunet–Derrida $(Y_n^0(t), 1 \le n \le N)$. We define the event A'_s that, for a *one-dimensional* free branching Brownian motion, a particle located at $R_1(s)$ first branches after time $s > 0$, its score is $\ge R_1(s) + 1$, and it subsequently produces at least N offspring by time $s+1$ whose score always stays above $R_1(s) + 1/2$ throughout the interval $[s, s+1]$. We now have that A'_s is independent of $R_1(s)$ and $\mathcal{F}_s$ and

$$\mathbb{P}(A_s | \mathcal{F}_s) \ge \mathbb{P}(A'_s | \mathcal{F}_s) = \mathbb{P}(A'_s).$$

So there exists $p_1 > 0$ such that

$$\mathbb{P}(A_s | \mathcal{F}_s) \ge p_1$$

almost surely for all s. We call $p = p_1 p_2 > 0$, and deduce from the above that if $G_s = A_s \cap B_s$,

$$\mathbb{P}(G_s | \mathcal{F}_s) \ge p$$

almost surely for all s. Note that when $A_s \cap B_s$ occurs, all the particles at time $s+1$ in the Brunet–Derrida system necessarily descend from the maximum particle at time s. Hence $\tau(s+1) \le 1$.

Applying this argument iteratively, we deduce that

$$\mathbb{P}(\tau(t) > k) \le \mathbb{P}(G_{t-k}^{\complement} \cap G_{t-k+1}^{\complement} \cap \cdots \cap G_{t-1}^{\complement})$$

from which the result follows. $\square$

With Lemma 2.8, we are now in a position to complete the proof of (3) with the following lemma. Endow $\mathbb{S}^{d-1}$ with the usual spherical (angular) metric D: for $\Theta_1, \Theta_2 \in \mathbb{S}^{d-1}$, let $D(\Theta_1, \Theta_2)$ be the distance on the sphere. In $\mathbb{R}^d$,

$$D(\Theta_1, \Theta_2) = \cos^{-1} \langle \Theta_1, \Theta_2 \rangle.$$

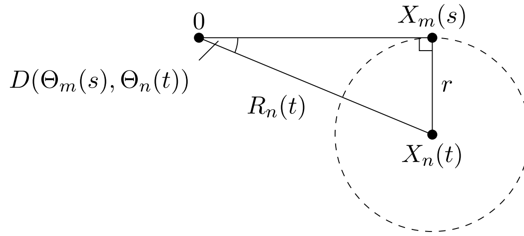

FIG. 4. *Proof of* (26). *The angle is maximised when the triangle formed by* 0, $X_m(s)$ *and* $X_n(t)$ *is rectilinear.*

LEMMA 2.10. *For all* $N > 1$,

$$\max_{1 \le m,n \le N} D\big(\Theta_m(t), \Theta_n(t)\big) \to 0$$

as $t \to \infty$ *almost surely.*

PROOF. Given two particles $X_m(s), X_n(t) \in \mathbb{R}^d$, let $r = \|X_m(s) - X_n(t)\|$ and assume for now that $r \le R_n(t) = \|X_n(t)\|$. Then a simple geometric argument (see Figure 4) shows that the distance $D(\Theta_m(s), \Theta_n(t))$ is biggest if $X_m(s)$ is perpendicular to $X_m(s) - X_n(t)$. Hence for $r \le R_n(t)$,

$$D\big(\Theta_m(s), \Theta_n(t)\big) \le \sin^{-1}\left(\frac{r}{R_n(t)}\right) \le \frac{\pi r}{2R_n(t)}, \tag{26}$$

since $\sin^{-1}(x) \le \frac{\pi}{2}x$ for all $0 \le x \le 1$.

Given $0 < \Delta < t$ and let $\tau = \tau(t)$ be the time to the most recent common ancestor of all the surviving particles at time t, as in (24). We first note that

$$\mathbb{P}(\tau \ge \Delta) \le (1-p)^{\Delta},$$

where p is as in Lemma 2.8. Hence picking $\Delta = C_1 \log t$ for some sufficiently large $C_1 > 0$, and applying the first Borel–Cantelli lemma shows that there exists $T_1 > 0$, possibly random, such that almost surely, $\tau \le \Delta$ for all $t > T_1$.

On the event $\{\tau \le \Delta\}$, let $X_k(t-\tau)$ be the position of the most recent common ancestor of all the surviving particles at time t. Since $\sup D \le \pi$, using (26), we have

$$\begin{aligned} D\big(\Theta_m(t), \Theta_n(t)\big) &\le D\big(\Theta_n(t), \Theta_k(t-\tau)\big) + D\big(\Theta_m(t), \Theta_k(t-\tau)\big) \\ &\le \frac{\pi\rho}{R_k(t-\tau)} \mathbf{1}\big\{\rho \le R_k(t-\tau)\big\} + \pi \mathbf{1}\big\{\rho > R_k(t-\tau)\big\} \qquad (27) \\ &\le \frac{\pi\rho}{R_k(t-\tau)}, \end{aligned}$$

where $\rho = \rho(t) = \sup_{u \le \tau} \max_{1 \le n \le N} \|X_n(t-\tau+u) - X_k(t-\tau)\|$. [As with $\tau = \tau(t)$, we drop the dependence on t in order to ease readability.] Note that on the event $\{\tau \le \Delta\}$, ρ can be stochastically dominated by

$$\rho \preceq \sup_{s\le\Delta} \sup_n \|\overline{Z}_n(s)\|,$$

where $(\overline{Z}_n(u), 1 \le n \le \overline{N}(u))$ is a d-dimensional branching Brownian motion started from one particle at 0 (and as before $\preceq$ denotes stochastic domination). Writing for all n and u,

$$\overline{Z}_n(u) = \big(\overline{Z}_n^{(1)}(u), \dots, \overline{Z}_n^{(d)}(u)\big) \in \mathbb{R}^d,$$

we have that $\overline{Z}_n^{(1)}(s), \dots, \overline{Z}_n^{(d)}(s)$ are one-dimensional branching Brownian motions and so a simple union bound yields, summing over all d coordinates, and using symmetry,

$$\begin{aligned}
&\mathbb{P}(\rho > \sqrt{2}d\Delta + C_2 d \log t, \tau \le \Delta) \\
&\quad \le d\mathbb{P}\Big(\sup_{s\le\Delta} \sup_n \big|\overline{Z}_n^{(1)}(s)\big| > \sqrt{2}\Delta + C_2 \log t\Big) \\
&\quad \le 2d\mathbb{P}\Big(\sup_{s\le\Delta} \sup_n \overline{Z}_n^{(1)}(s) > \sqrt{2}\Delta + C_2 \log t\Big) \qquad (28)\\
&\quad \le 4de^{-\sqrt{2}C_2 \log t},
\end{aligned}$$

where (28) follows by Lemma 2.2 with $t = \Delta$ and $K = C_2 \log t$. Restricting to integer values of t, this is also summable for sufficiently large C_2, hence we deduce that almost surely there exists $T_2 > 0$, possibly random, such that if $t > T_2$ is an integer then $\tau \le C_1 \log t$ and $\rho \le C_3 \log t$ where $C_3 = \sqrt{2}dC_1 + dC_2$.

Then for $t > T_2$ an integer, applying (27) and Lemma 2.7, there exists some $C_4 > 0$ such that

$$D\big(\Theta_m(t), \Theta_n(t)\big) \le \frac{\pi C_3 \log t}{R_k(t-\tau)} \le \frac{C_4 \log t}{v_N(t - C_1 \log t)}.$$

The right-hand side tends to 0 as $t \to \infty$ uniformly over m, n, so almost surely

$$\sup_{1\le m,n\le N} D\big(\Theta_m(t), \Theta_n(t)\big) \to 0, \qquad (29)$$

along integers, almost surely. By comparing $\Theta_m(t)$ to $\Theta_m(\lfloor t\rfloor)$, where here $\Theta_m(u)$ denotes the angle of the ancestor of $X_m(t)$ at time u, and using a similar argument as above, we obtain that

$$\sup_{t\in[j,j+1)} \sup_{1\le m\le N} D\big(\Theta_m(t), \Theta_m(\lfloor t\rfloor)\big) \to 0$$

almost surely as $j \to \infty$. Combining with (29), we obtain the desired result. $\square$

At this stage, (3) is proved and we turn to the proof of (4). We already know that $R_1(t)/t \to v_N$ almost surely so it remains to prove that $\Theta_1(t) \to \Theta$ almost surely. We now introduce an important notion of *spine* for this process, as follows. Note that by Lemma 2.8, the system eventually has a unique most recent common ancestor. We also observed that almost surely, the last time $s(t) = t - \tau(t)$ such that all particles alive at time t descend from a single ancestor at time $s(t)$ satisfies $s(t) \to \infty$ as $t \to \infty$, almost surely. Hence if we consider the genealogical path of the most recent common ancestor, we see that there is a unique immortal genealogical path in the system, or the "spine," from which all the particles that are eventually ever alive in the system descend from.

Let $X_*(t)$ be the particle of the spine at time t. This is the location of the ancestor at time t of the most recent common ancestor of the population at time u for all u sufficiently large. Then $X_*(t)$ is a continuous function of t, almost surely. We consider the usual angular decomposition: $X_*(t) = R_*(t)\Theta_*(t)$ where $R_*(t) > 0$ and $\Theta_*(t) \in \mathbb{S}^{d-1}$ is continuous. We now complete the proof of Theorem 1.1 by showing the angular part of the spine converges. By Lemma 2.10, this will imply that $\Theta_1(t)$ also converges almost surely.

PROPOSITION 2.11. *For all $N > 1$, $\Theta_*(t)$ converges almost surely as $t \to \infty$.*

We offer two proofs of this proposition. One is shorter but relies explicitly on stochastic calculus, and hence works only for the exact situation described in this paper. On the other hand, the second proof is a bit longer but more robust; in particular it carries over to slightly more general Brunet–Derrida particle systems than the ones we consider in this paper; see Remark 2.14 at the end of this section.

FIRST PROOF OF PROPOSITION 2.11. We start by recalling the classical *skew-product decomposition* of Brownian motion (see, e.g., Section 7.15 of [18]). The version we present here is Theorem 1.1(d) of [21].

Let $(X(t), t \geq 0)$ be a d-dimension Brownian motion, and write $X(t) = R(t)\Theta(t)$ with $R(t) > 0$ and $\Theta(t) \in \mathbb{S}^{d-1}$ and R, Θ continuous. Let

$$H_t = \int_0^t R(s)^{-2}\, ds. \tag{30}$$

Then:

(i) $(R(t), t \geq 0)$ is a Bessel process of order d.

(ii) Under the time change $\Phi(H_t) = \Theta(t)$, $(\Phi(t), t \geq 0)$ is a Brownian motion on $\mathbb{S}^{d-1}$.

(iii) $(\Phi(t), t \geq 0)$ is independent of $(R(t), t \geq 0)$.

In the special case of $d = 2$, we can write $\Theta(t)$ as $e^{iB(H_t)}$ where $(B(t), t \geq 0)$ is a standard Brownian motion in $\mathbb{R}$ independent of $(R(t), t \geq 0)$. We note here that in this case $\Phi(t) = e^{iB(t)}$.

Now consider the system $(\overline{X}(t), t \geq 0)$ that results from not enforcing selection: the underlying free d-dimensional branching Brownian motions started from

N particles at $X_1(0), \dots, X_N(0)$ coupled to the Brunet–Derrida system. It is clear that this can be constructed by considering the skew product decomposition of every Brownian path in the system $\overline{X}_i(t) = \overline{R}_i(t)\overline{\Theta}_i(t) = \overline{R}_i(t)\overline{\Phi}_i(H_i(t))$. This is a bit cumbersome, but we include the details to emphasise that the angular structure of the whole free branching Brownian motion can be incorporated to the process after its radial part is constructed. It will follow that any information on the radial behaviour of *any* particle in this process translates into information about its angular part.

Let $\mathcal{T}$ be the underlying branching tree (which by assumption is just an ordinary Yule process). We use Neveu's formalism for binary trees, that is, $\mathcal{T}$ is a set of vertices given by $\mathcal{T} = \bigcup_{n=0}^{\infty} \{0, 1\}^n$ and each vertex v has attached to it an independent exponential random variable of mean 1, X_v, representing the lifetime of this individual. We call $[s_v, t_v]$ the interval of time over which this particle is alive, thus $t_v - s_v = X_v$ and so $s_v = \sum_{w \preceq v} X_w$ (with $w \preceq v$ means w is ancestor of v). We also attach to each v a Bessel process $R_v(t)$ defined over the interval of time $[s_v, t_v]$ in the natural way, by solving the SDE

$$dR_v(t) = dB_v(t) + \frac{d-1}{2R_v(t)}\,dt, \qquad t \in [s_v, t_v],$$

where the Brownian motions B_v are independent for different vertices v, and by requiring continuity of the resulting Bessel process when we move up along the branches of the tree. We extend the definition of $R_v(t)$ to the entire interval $[0, t_v]$ simply by defining $R_v(s) = R_w(s)$ where w is the unique ancestor of v alive at time s (i.e., such that $s \in [s_w, t_w]$).

We further enrich this structure by associating to each vertex v an angle process $\Theta_v(t)$, also defined over the interval of time $[s_v, t_v]$, which is defined by applying the construction (30) in between two successive branching events. More precisely, let

$$H_v(t) = \int_0^t R_v(s)^{-2}\,ds,$$

let $s'_v = H_v(s_v)$ and $t'_v = H_v(t_v)$. Consider a family of Brownian motions on $(\Phi_v(t), t \in [s'_v, t'_v], v \in \mathcal{T})$ on $\mathbb{S}^{d-1}$ such that the evolution of Φ_v over $[s'_v, t'_v]$ are independent for different vertices $v \in \mathcal{T}$. As above, we extend $\Phi_v(t)$ to the interval $[0, t'_v]$ by defining $\Phi_v(t) = \Phi_w(t)$ where w is the unique ancestor of v such that $t \in [s'_v, t'_v]$, and we have chosen Φ_v so that $\Phi_v(t)$ is a continuous function of t over $[0, t'_v]$ for all $v \in \mathcal{T}$. We now define Θ_v by the formula

$$\Theta_v(t) = \Phi_v\bigl(H_v(t)\bigr),$$

for $s_v \le t \le t_v$.

Let $\overline{S}(t)$ be the set of particles alive at time t, that is, the set of vertices $v \in \mathcal{T}$ such that $t \in [s_v, t_v]$. Let $\overline{N}(t) = |\overline{S}(t)|$ and order the vertices in $\overline{S}(t)$ by $v_1, \dots, v_{\overline{N}(t)}$ in such a way that $\overline{R}_1(t) \ge \overline{R}_2(t) \ge \cdots$, where $\overline{R}_i(t) = R_{v_i}(t)$. Also let $\overline{\Theta}_i(t) = \Theta_{v_i}(t)$ for $1 \le i \le N(t)$. Then our system of branching Brownian motions consists of

$$\overline{X}_i(t) = \overline{R}_i(t)\overline{\Theta}_i(t), \qquad 1 \le i \le N(t), t \ge 0.$$

Having described the skew product decomposition of a free branching Brownian motion $(\overline{X}_i(t), t \ge 0, 1 \le i \le \overline{N}(t))$, we proceed with the proof of Proposition 2.11. By a *ray*, we mean a sequence $V = \{v_1, v_2, \dots\}$ such that v_n is in generation n of the tree and $v_n \preceq v_{n+1}$ for all $n \ge 0$. For each given ray V, we can follow the trajectory $X_V(t)$ of the Brownian motion associated with V, that is, $X_V(t) = X_v(t)$ for the a.s. unique $v \in V$ such that $t \in [s_v, t_v]$. We can also consider $R_V(t) = R_v(t)$ its radial part and $\Theta_V(t) = \Theta_v(t)$ its angular part. Observe then that we have, by construction, $\Theta_V(t) = \Phi_V(H_V(t))$ where Φ_V is a Brownian motion on $\mathbb{S}^{d-1}$ and $H_V(t) = \int_0^t R_V(s)^{-2}\,ds$.

Now, consider the spine defined before and the associated ray V in the tree. Then by definition we have $X_V(t) = X_*(t)$ for all $t \ge 0$. We deduce from Lemma 2.7 that

$$\frac{\|X_V(t)\|}{t} \ge v_N/2$$

for all sufficiently large t, almost surely. It follows that $H_V(t) = \int_0^t R_V(s)^{-2}\,ds$ converges almost surely as $t \to \infty$ to a limit $H_V(\infty)$. Hence $\Theta_V(t)$ also converges almost surely as $t \to \infty$ to a limit, namely $\Phi_V(H_V(\infty))$. In other words, $\Theta_*(t)$ almost surely converges as $t \to \infty$ to a limit. $\square$

SECOND PROOF OF PROPOSITION 2.11. Our second proof relies on a suitable martingale argument rather than stochastic calculus, and hence is more robust. See Remark 2.14 for a discussion of the setups to which it carries. Consider a free branching Brownian motion $\overline{X} = (\overline{X}_i(t), 1 \le i \le \overline{N}(t), t \ge 0)$, and write $\overline{X}_i(t) = \overline{R}_i(t)\overline{\Theta}_i(t)$ for $t \ge 0$ and $1 \le i \le \overline{N}(t)$. Let $\mathcal{F}_t^R = \sigma(\overline{R}_i(u), 1 \le i \le \overline{N}(u), u \le t)$ and let $\mathcal{F}_t^\Theta = \sigma(\overline{\Theta}_i(u), 1 \le i \le \overline{N}(u), u \le t)$. Let $\mathcal{G}_t = \sigma(\mathcal{F}_\infty^R \cup \mathcal{F}_t^\Theta)$, and note that $(\theta_*(t), t \ge 0)$ is adapted to the filtration $(\mathcal{G}_s, s \ge 0)$.

We start by explaining the argument in the case $d = 2$, which is a bit simpler to describe. Recall in the case $d = 2$, we can write $X_*(t) = R_*(t)e^{i\theta_*(t)}$, where $R_*(t) > 0$ and $\theta_*(t)$ is a continuous function. This way of writing $X_*(t)$ is unique modulo a global constant multiple of 2π in $\theta_*(t)$, which we fix once and for all at time 0.

LEMMA 2.12. $(\theta_*(t), t \ge 0)$ *is a martingale with respect to* $(\mathcal{G}_t, t \ge 0)$.

PROOF. It is easy to check that $\theta_*(s)$ is integrable. Indeed we can bound $|\theta_*(s)|$ by the sum of the changes in argument between any individual present at time 0 and any of its descendants at time s, ignoring selection. Conditioning first on the number $N(s)$ of descendants at time s, and using the many-to-one lemma, we see that

$$\mathbb{E}\big(|\theta^*(s)|\,|N(s)\big) \le N(s)c,$$

where c is the corresponding expectation for a single Brownian motion. In particular $c < \infty$. Since $\mathbb{E}(N(s)) = Ne^s < \infty$, we deduce $\mathbb{E}(|\theta^*(s)|) < \infty$.

Fix $s > 0$ and $t > s$. We wish to show that $\mathbb{E}(\theta^*(t)|\mathcal{G}_s) = \theta^*(s)$. To do this, we first condition on $\mathcal{F}_s = \sigma(\mathcal{F}_s^R \cup \mathcal{F}_s^\Theta)$, which is the σ-field generated by the process up to time s. We consider the set of locations of all N particles present at time s. Let z be such a location, so that $X_i(s) = z$ for some i. Consider the transformation $T = T_z$ which is a reflection in the line $\mathbb{R}z$:

$$T_z(x) = 2\langle x, z'\rangle z' - x, \qquad x \in \mathbb{R}^d,$$

where $z' = z/\|z\|$. Note that T_z is an orthogonal transformation, and hence leaves the Wiener measure invariant. Consider the free branching Brownian motion (without selection) started from the locations of all N particles at time s. We apply T_z to every descendant of the particle which is at z at time s. Likewise we also apply the corresponding transformation T_z to all the descendants of each particle present at time s, where z ranges over the set of locations of al N particles present at time s. We call $T(\overline{X}(t)) = (T(\overline{X}_i(t)), 1 \le i \le \overline{N}(t))$ for $t \ge s$ the resulting transformation of all the particles in the branching Brownian motion. We note that since each T_z leaves Brownian motion invariant, given $\mathcal{F}_s$, $\{T(\overline{X}(t)), t \ge s\}$ has also the law of a free branching Brownian motion starting from the configuration of particles present in the system at time s. Moreover, for a fixed z, T_z is an isometry so we have $\|T(\overline{X}_i(t))\| = \overline{R}_i(t)$ for all $t \ge s$ and all $1 \le i \le \overline{N}(t)$. In particular, a particle $T(\overline{X}_i(t))$ survives the selection procedure if and only if its mirror image $\overline{X}_i(t)$ does. In particular, the branching times and tree structure of the system are invariant under T.

These two properties imply that when we turn to the system X with selection, conditional on $\mathcal{F}_s$, $T(X)$ has the same distribution as X. On the other hand, observe that if the population $X(t)$ at time t has a particle at x descending from a particle at z at time s, then

$$\arg T(x) = \arg T_z(x) = 2\arg z - \arg x.$$

Applying this to $z = X_*(s)$ and $x = X_*(t)$ shows that if $\Delta\theta_*(t) = \theta_*(t) - \theta_*(s)$ and if F, G are two bounded continuous functions on $C([t,\infty))$ and $C([t,\infty)^N)$, respectively, then

$$\begin{aligned}
&\mathbb{E}\big(F\big(\Delta\theta_*(t), t \ge s\big)G\big(R_i(t), 1 \le i \le N, t \ge s\big)|\mathcal{F}_s\big)\\
&\quad = \mathbb{E}\big(F\big(-\Delta\theta_*(t), t \ge s\big)G\big(R_i(t), 1 \le i \le N, t \ge s\big)|\mathcal{F}_s\big)
\end{aligned}$$

and hence $\Delta\theta_*(t)$ has the same distribution as $-\Delta\theta_*(t)$ given $\mathcal{G}_s$, since F and G are arbitrary. In particular,

$$\mathbb{E}\big[\Delta\theta_*(t)|\mathcal{G}_s\big] = \mathbb{E}\big[-\Delta\theta_*(t)|\mathcal{G}_s\big] = 0,$$

as desired, so θ^* is a $(\mathcal{G}_s, s \ge 0)$-martingale. $\square$

When $d \geq 3$, we reason similarly but it is necessary to first project onto a two-dimensional subspace Π before applying a similar reasoning. Let Π be a given such plane and let p_Π be the orthogonal projection onto Π. For some fixed $e \in \Pi$ and $x \in \mathbb{R}^d$, define $\arg_\Pi(x)$ to be the continuous directed angle between $p_\Pi(x)$ and e.

LEMMA 2.13. *$(\arg_\Pi(X_*(t)), t \geq 0)$ is a martingale with respect to $(\mathcal{G}_t, t \geq 0)$.*

For $z \in \mathbb{R}^d$, let $T_z(x)$ be defined by

$$T_z(x) = x - 2\big(p_\Pi(x) - \langle p_\Pi(x), z'\rangle z'\big), \qquad x \in \mathbb{R}^d,$$

where $z' = p_\Pi(z)/\|p_\Pi(z)\|$. More descriptively, if $x = u + v$ where $u \in \Pi$ and v is orthogonal to Π, then $T_z(x) = T_z(u) + T_z(v)$, where $T_z(v) = v$ and $T_z(u)$ is the reflection of u in the line $\mathbb{R}P_\Pi(z)$ within the plane Π. As before, applying this transformation to each descendant of a particle located at z at time s yields a transformation T of the branching Brownian motion, which leaves the modulus of particles $\|T(\overline{X}_i(t))\| = \|\overline{X}_i(t)\|$ unchanged, and leaves the law of branching Brownian motion also unchanged. But the choice of T gives $\arg_\Pi(T(x)) = 2\theta_\Pi(z) - \arg_\Pi(x)$ if a particle at x descends from a particle at z. Thus

$$\mathbb{E}\big[\arg_\Pi\big(X_*(t)\big) - \arg_\Pi\big(X_*(s)\big)|\mathcal{G}_s\big] = \mathbb{E}\big[\arg_\Pi\big(X_*(s)\big) - \arg_\Pi\big(X_*(t)\big)|\mathcal{G}_s\big] = 0,$$

as above. This concludes the proof of the lemma.

We are now ready to conclude the second proof of Proposition 2.11. It suffices to prove that $\theta_*(t)$ converges as $t \to \infty$. We can assume without loss of generality that $d = 2$, as it suffices to show that $\arg_\Pi(X_*(t))$ converges as $t \to \infty$ for any fixed arbitrary two-dimensional subspace Π. Thus we will assume $d = 2$.

Let $q > 0$ be a fixed positive number which will be chosen sufficiently large below. Let $s > 0$ and assume that s is an integer multiple of q. Define $\rho_*(s) = \sup_{u \in [s-q,s]} \|X_*(u) - X_*(s)\|$ and define a stopping time T of the filtration $(\mathcal{G}_s)_{s \geq 0}$ to be the first time s which is an integer multiple of q such that $\rho_*(s) \geq R_*(s)$. Let $\theta_*^T(s) = \theta_*(s \wedge T)$ be the martingale $\theta_*(s)$ stopped at T. The reason for stopping at T is to ensure a bound similar to (26) holds for $\theta_*^T(t)$. The precise bound is, for $s > 0$ a multiple of q,

$$\big|\theta_*^T(s) - \theta_*^T(s-q)\big| \leq \frac{\pi \rho_*(s)}{2R_*(s-q)} \leq \frac{\pi}{2}. \tag{31}$$

Since θ_*^T is a martingale, we deduce that for t an integer multiple of q,

$$\begin{aligned} \mathbb{E}\big[\theta_*^T(t)^2\big] &= \sum_{k=0}^{t/q-1} \mathbb{E}\big[\big(\theta_*^T\big((k+1)q\big) - \theta_*^T(kq)\big)^2\big] \\ &\leq \sum_{k=0}^{t/q-1} \frac{\pi^2}{4} \mathbb{E}\bigg[1 \wedge \frac{\rho_*((k+1)q)^2}{R_*(kq)^2}\bigg]. \end{aligned} \tag{32}$$

We want to get an upper bound for the terms in the right-hand side. We observe that $R_*(s) \geq \|X_N(s)\|$ [since the spine $X_*(s)$ is the position of some particle at time s, and all particles have modulus greater than $\|X_N(s)\|$ by our convention on ordering], and proceed to bound from below $\|X_N(s)\|$ stochastically. Suppose s is an integer multiple of q. Using the coupling used in the proof of Lemma 2.7, we have that $\|X_N(s)\|$ dominates the minimum at time s of a standard one-dimensional Brunet–Derrida system started from N particles all at the origin. By the monotone coupling of Lemma 2.3, this dominates $S(s) = Z(1) + \cdots + Z(s/q)$, where $Z(i)$ are i.i.d. and distributed as the position of the minimum at time q of a one-dimensional Brunet–Derrida system where all N particles are initially at the origin.

Let $m = m_N = \mathbb{E}[Z(1)] = \mathbb{E}_{(0,\dots,0)}(X_N(q))$. Note that for $N \geq 2$, as $q \to \infty$, $m_N \to \infty$ by Corollary 2.6. Hence we may fix $q > 0$ large enough than $m_N > 0$, which we now assume. Furthermore, $\mathbb{E}[e^{-\lambda Z(1)}] < \infty$ for all $\lambda \geq 0$. Indeed, by coupling, it suffices to observe that the minimum $\overline{M}(q)$ at time q of a free branching Brownian motion started from the origin has exponential moments of all negative orders. This in turn follows from the following argument. Let $\overline{N}(q)$ denote the number of descendants of the N root particles at time q. Then conditionally on $\overline{N}(q)$,

$$\mathbb{P}(\overline{M}(q) \leq y|\overline{N}(q)) \leq \overline{N}(q)\frac{e^{-y^2/2q^2}}{\sqrt{2\pi q^2}}$$

by Markov's inequality, the many-to-one-lemma (Lemma 2.1) and the bound (19). Taking the expectation again, we see that $\mathbb{P}(\overline{M}(q) \leq y) \leq Ce^{-y^2/2q^2}$, where C depends on N and q, but not on y, from which exponential moments of all negative orders follow immediately.

Hence $\psi(\lambda) = \log \mathbb{E}[e^{-\lambda Z(1)}]$ is well defined for $\lambda \geq 0$. Then for any $\lambda \geq 0$,

$$\mathbb{P}\left(S(s) \leq \frac{1}{2q}ms\right) \leq \mathbb{P}\left(e^{-\lambda S(s)} \geq e^{-\frac{1}{2q}\lambda ms}\right) \leq e^{\frac{1}{2q}\lambda ms}\mathbb{E}\left[e^{-\lambda S(s)}\right] \leq \exp\left(\frac{s}{q}f(\lambda)\right),$$

where $f(\lambda) = \frac{1}{2}\lambda m + \psi(\lambda)$. We note that $f(0) = 0$ and $f'(0) = \psi'(0) + \frac{1}{2}m = -\frac{1}{2}m < 0$ and that for some fixed sufficiently small λ, $f(\lambda) \leq \frac{1}{2}\lambda f'(0)$. Therefore,

$$\mathbb{P}\left(S(s) \leq \frac{1}{2q}ms\right) \leq \exp\left(-\frac{1}{4q}\lambda ms\right), \tag{33}$$

where λ is as above. By Jensen's inequality and since $x \mapsto x \wedge 1$ is concave, and using (33) with $s = kq$,

$$\begin{aligned}
&\mathbb{E}\left[1 \wedge \frac{\rho_*((k+1)q)^2}{\|X_N(kq)\|^2}\right] \\
&\leq \mathbb{E}\left[\left(1 \wedge \frac{4\rho_*((k+1)q)^2}{m^2k^2}\right)\mathbf{1}_{\{\|X_N(kq)\| \geq \frac{mk}{2}\}}\right] + \mathbb{P}\left(\|X_N(kq)\| \leq \frac{mk}{2}\right) \qquad (34)\\
&\leq 1 \wedge \frac{4\mathbb{E}[\rho_*((k+1)q)^2]}{m^2k^2} + e^{-ck}.
\end{aligned}$$

It is not hard to see that there exists $C_1 > 0$ depending on N but not s such that $\mathbb{P}(\rho^*(s) > y) \le C_1 e^{-y^2/2}$, and hence $\mathbb{E}[\rho_*(s)^2] \le C_1'$. To see this, note that since C_1 is allowed to depend on N, by summing over all N particles present at time $s - q$, it suffices to establish the bound

$$\mathbb{P}(\rho \ge y) \le C_1 e^{-y^2/2q^2},$$

where ρ is defined as follows: we consider a free branching Brownian motion started at time 0 with N root particles at the origin, and ρ is the maximal distance travelled by time q by any descendants of any of the root N particles during that time. This follows as before by conditioning on $\overline{N}(q)$, the number of descendants of the root particles at time q.

Therefore, we deduce from (34) that for some constant C_1',

$$\mathbb{E}\left[1 \wedge \frac{\rho_*((k+1)q)^2}{\|X_N(kq)\|^2}\right] \le \frac{C_1'}{k^2} + e^{-ck}.$$

Plugging this into (32), we see that for some $C_2 > 0$ and for all integer t of the form $t = kq$, $k \ge 1$

$$\mathbb{E}\left[\theta_*^T(t)^2\right] \le C_2.$$

Since the left-hand side is a monotone function of t by Jensen's inequality, this also holds for all $t > 0$ and so $(\theta_*^T(t), t \ge 0)$ is a martingale bounded in L^2, and so converges almost surely.

Obviously, this implies convergence of θ_* almost surely on the event $\{T = \infty\}$. We now check that almost surely $\rho_*(s) \ge R_*(s)$ eventually never happens for some integer multiple of q. But note that since $\mathbb{E}[\rho_*(s)^2] \le C_1 < \infty$ it follows from Markov's inequality and the Borel–Cantelli lemma that $\rho_*(s) < v_N s/2$ for all s sufficiently large, and hence $\rho_*(s) \le R_*(s)$ for all integer multiples $s = kq$ sufficiently large by Lemma 2.7. Thus if we apply the result that θ_*^T converges almost surely, starting from the initial condition of the system at times $q, 2q, \dots$ we deduce that almost surely there is a k large enough so that the system started from the initial condition $(X_1(kq), \dots, X_N(kq))$ verifies $T = \infty$, and hence θ_* converges. The result follows. $\square$

REMARK 2.14. In this paper, we have concerned ourselves for simplicity with branching Brownian motion with selection. However, there are a variety of possible alternatives: for instance, initially, Brunet and Derrida considered a system where branching occurs at discrete time steps $t = 0, 1, \dots,$ and at each t, each particle branches into two (or possibly even more) individuals, and the displacement follows a random walk with a given distribution. Yet another alternative, taken up by Durrett and Remenik [13], is to have particles branch at rate 1 in continuous time.

As is plain from the above proof, Theorem 1.1 remains true in each of these cases, under the assumption that if $p(x, \cdot)$ is the transition kernel of the underlying random walk then $p(x, y) = p(\|y - x\|)$ is a translation invariant, rotationally symmetric function and $\int_0^\infty e^{\lambda r} p(r)\, dr < \infty$ for some $\lambda > 0$.

3. Displacement of minimum: Proof of Proposition 1.12. We will prove a slightly more general bound than Proposition 1.12. Consider a standard one-dimensional Brunet–Derrida particle system with N particles, started from an initial configuration satisfying, after projection onto $\mathrm{Span}(\lambda)$,

$$\sum_{n=1}^{N} e^{\mu(X_n(0)-x_0)} \le N^{\delta} \tag{35}$$

for some $\delta < 1$ and some $\mu \le \sqrt{2}$ and write $\mu = \sqrt{2-\varepsilon}$ for some $\varepsilon \ge 0$. Our aim will be to prove that for some constants c_δ, c, κ, and for all $\delta < \delta' < 1$,

$$\mathbb{P}\big(X_N(t) \le x_0 + \mu t; \forall t \le T\big) \ge 1 - \frac{e^{\varepsilon T} T \log N}{N^{1-\delta'}}. \tag{36}$$

Then Proposition 1.12 follows by taking $\mu = \sqrt{2 - (2\pi)^2/(\log N)^2}$ and $T = c_\delta (\log N)^3$. Indeed in this case,

$$e^{\varepsilon T} \le \exp\big(\varepsilon c_\delta (\log N)^3\big) = \exp\big(2\pi^2 c_\delta \log N\big) = N^{2\pi^2 c_\delta}.$$

Therefore,

$$\frac{e^{\varepsilon T} T \log N}{N^{1-\delta}} \le N^{-\kappa} (\log N)^4,$$

where $\kappa = 1 - 2\pi^2 c_\delta - \delta > 0$ for a small enough choice of c_δ, so the proposition follows with a slightly different value of κ.

Note that, for any fixed $\zeta > 0$, it suffices to show

$$\mathbb{P}\big(X_N(t) \le \mu t + x_0 + \zeta; \forall t \le T\big) \ge 1 - \frac{e^{\varepsilon T} T \log N}{N^{1-\delta'}}, \tag{37}$$

since if we replace x_0 by $x_0 - \zeta$

$$\sum_{n=1}^{N} e^{\sqrt{2}(X_n(0)-(x_0-\zeta))} = e^{\zeta} \sum_{i=1}^{N} e^{\sqrt{2}(X_i(0)-x_0)} \le e^{\sqrt{2}\zeta} N^{\delta},$$

which also satisfies (35) for N sufficiently large and for a slightly different value of δ. Thus (37) implies (36). (Note that in this argument ζ itself is allowed to depend on δ.)

For ease of notation, we will assume without loss of generality (since the system is translation invariant) that $x_0 = 0$. The key idea of the proof is to compare the Brunet–Derrida system to a free branching Brownian motion where particles are absorbed at a wall, that is, a linear boundary. This is similar to the idea which lies behind papers such as [4], which used a wall of velocity approximately equal to μ in (11) (though with an additional correction term). We shall here consider a wall moving exactly at speed μ.

We use the same natural coupling to a free branching Brownian motion in $\mathbb{R}$ as before, with particles $\overline{X}_i(t)$, $1 \le i \le \overline{N}(t)$, ordered in the usual way right to left. The coupling is obtained by enforcing selection at all times to the free system. Note the key property of this coupling that for each $1 \le i \le N$, $X_i(t) \le \overline{X}_i(t)$, with probability one. Therefore, under this coupling,

$$\mathbb{P}\big(X_N(t) \ge \mu t\big) \le \mathbb{P}\big(\overline{X}_N(t) \ge \mu t\big). \tag{38}$$

Let $T = c_\delta (\log N)^3$ where $c_\delta > 0$ is a small constant depending only on δ which we will fix later on. Let

$$W_t \overset{\text{def}}{=} \sum_{i \ge 1} \mathbf{1}_{\{\overline{X}_i(t) \ge \mu t\}}$$

be the number of particles of the free branching Brownian motion which are greater or equal to μt. Then by (38), we get

$$\mathbb{P}\big(X_N(t) \ge \mu t\big) \le \mathbb{P}(W_t \ge N) \le \frac{\mathbb{E}[W_t]}{N} \tag{39}$$

by Markov's inequality.

In order to estimate W_t we will treat separately the particles that hit the wall and those that do not. More precisely, let $J(t)$ denote the index of particles that never touch the position μs for any $s \le t$. Thus the particles in $J(t)$ are killed when they hit the wall moving at velocity μ, starting from position 0. Let

$$W_1(t) = \sum_{i \in J(t)} \mathbf{1}_{\{\overline{X}_i(t) \ge \mu t\}},$$

$$W_2(t) = \sum_{i \in K(t)} \mathbf{1}_{\{\overline{X}_i(t) \ge \mu t\}},$$

where $K(t)$ is the the set of indices corresponding to particles that do hit the wall at some time $s \le t$, so that

$$W_t = W_1(t) + W_2(t).$$

LEMMA 3.1. *There exists a universal constant $C > 0$ such that for any $t \le T$,*

$$\mathbb{E}\big[W_1(t)\big] \le e^{\frac{1}{2}\varepsilon t} \sum_{i : \overline{X}_i(0) > 0} e^{\mu \overline{X}_i(0)}.$$

PROOF. Suppose initially there is one particle at $x > 0$. Let $N_x(t)$ denote the number of descendants of that particle that do not hit the wall. Then by the many-to-one lemma [Lemma 2.1 with $T = t$ and $g((Y_s)_{s\le t}) = 1_{\{Y_s \ge \mu s \text{ for all } s\le t\}}$] and Girsanov's theorem,

$$\begin{aligned}\mathbb{E}\big(N_x(t)\big) &= e^t \mathbb{P}_x(B_s \ge \mu s \text{ for all } s \le t)\\ &= e^t \mathbb{E}_x\big(e^{-\mu(B_t - x) - \frac{\mu^2}{2}t} 1_{\{B_s \ge 0 \text{ for all } s \le t\}}\big)\\ &\le e^{\frac{1}{2}\varepsilon t} e^{\mu x} \mathbb{P}_x(B_s \ge 0 \text{ for all } s \le t) \le e^{\frac{1}{2}\varepsilon t} e^{\mu x}\end{aligned}$$

from which the result follows by summing over all particles which are initially at some position $x > 0$. □

It remains to treat particles that do hit the wall. Let Δ_t be the number of particles that are killed on the wall up to time t if we kill all particles that hit this wall.

LEMMA 3.2. *Given* Δ_t,

$$\mathbb{E}\big[W_2(t)\big] \le e^{\frac{1}{2}\varepsilon t} \mathbb{E}[\Delta_t].$$

PROOF. Let $\mathcal{H}_t = \sigma(\Delta_s, s \le t)$ and let $t - s_1, \dots, t - s_{\Delta_t}$ be the times at which a particle hits the wall before time t. For each particle killed on the wall at some time $t - s \le t$, the expected number of descendants at time t that are greater or equal to μt is simply, by translation invariance, and the many-to-one lemma [Lemma 2.1 applied with $T = s$, $g((Y_u)_{u\le s}) = 1_{\{Y_s \ge \mu s\}}$], and using (19),

$$e^s \mathbb{P}(B_s \ge \mu s) \le e^s e^{-\mu^2 s/2} = e^{\varepsilon s/2} \le e^{\frac{1}{2}\varepsilon t}. \tag{40}$$

Therefore, the expected contribution to $W_2(t)$ from a particle hitting the wall at time $t - s_i (1 \le i \le \Delta_t)$, conditional on $\mathcal{H}_t$, is at most $e^{(1/2)\varepsilon t}$ and we get, summing over $1 \le i \le \Delta_t$,

$$\mathbb{E}\bigg[\sum_{i \in K(t)} \mathbf{1}_{\{\overline{X}_i(t) \ge \mu t\}} \Big| \mathcal{H}_t\bigg] \le e^{\frac{1}{2}\varepsilon t} \Delta_t.$$

Taking expectations, the lemma follows. □

Hence we have reduced the problem to estimating from above $\mathbb{E}[\Delta_t]$. To this end, we will distinguish between those that started at positive positions and those at negative positions, respectively $K_+(t)$ and $K_-(t)$. Call $\Delta_+(t)$ and $\Delta_-(t)$ the corresponding number of particles killed at the wall.

LEMMA 3.3.

$$\mathbb{E}\big[\Delta_+(t)\big] \le C e^{\frac{1}{2}\varepsilon t} \sum_{i : \overline{X}_i(0) > 0} e^{\mu \overline{X}_i(0)}.$$

PROOF. Any particle that first hits the wall from the right has to descend from an ancestor $\overline{X}_i(0)$ at time 0 with $\overline{X}_i(0) > 0$. We use the following very crude bound: during times t and $t+1$, given $W_1(t)$,

$$\mathbb{E}\big[\Delta_+(t+1) - \Delta_+(t) | W_1(t)\big] \le e W_1(t). \tag{41}$$

This is because the number of particles killed on the wall during $[t, t+1]$ cannot exceed the total number of descendants of particles $i \in J(t)$ such that $X_i(t) > 0$. Since the number of such particles is precisely $W_1(t)$, (41) follows.

Taking expectations in (41), we have by Lemma 3.1,

$$\mathbb{E}\big[\Delta_+(t+1) - \Delta_+(t)\big] \le e\mathbb{E}\big[W_1(t)\big] \le e^{1+\frac{1}{2}\varepsilon t} \sum_{i:\overline{X}_i(0)>0} e^{\mu \overline{X}_i(0)}.$$

Since Δ_s is nondecreasing, the lemma follows by summing over the intervals $[0, 1], \dots, [\lfloor t \rfloor, \lceil t \rceil]$. $\square$

We now address $\Delta_-(t)$.

LEMMA 3.4.

$$\mathbb{E}\big[\Delta_-(t)\big] \le e^{\frac{1}{2}\varepsilon t} \sum_{i:\overline{X}_i(0)<0} e^{\mu \overline{X}_i(0)}.$$

PROOF. Any particle that first hits the wall from the left has to descend from an ancestor $\overline{X}_i(0)$ at time 0 with $\overline{X}_i(0) < 0$. The total number of such particles up to time t, $\Delta_-(t)$, is exactly the number of particles of a branching Brownian motion with drift $-\mu$ that hit level 0 by time t, started from the negative positions in the initial condition.

Fix some constant $A > 0$, and consider a branching Brownian motion with drift $-\mu$ where every particle is *stopped* upon reaching 0 and *killed* upon reaching $-A$. Initially, the starting positions consists precisely of $(\overline{X}_i(0), 1 \le i \le N)$ whenever $\overline{X}_i(0) < 0$. We call $X_i^*(t)$, $i \in N^*(t)$, the corresponding particle locations. Let $\Delta_-^A(t)$ be the number of particles stopped upon reaching 0 by time t. Now consider the process

$$M_s^A = \sum_{i \in N^*(t)} \big(X_i^*(s) + A\big) e^{\mu(X_i^*(t)+A) - \frac{1}{2}(2-\mu^2)s}. \tag{42}$$

Without stopping particles upon reaching 0, it is easy to check that $(M_s^A, s \ge 0)$ defines a nonnegative martingale (see, e.g., Lemma 2 of [17], or Lemma 6 of [4]). However, if we stop particles upon reaching 0, since $2 - \mu^2 = \varepsilon \ge 0$, M_s^A becomes a supermartingale. Therefore,

$$\sum_{i:\overline{X}_i(0)<0} A e^{\mu(\overline{X}_i(0)+A)} \ge \mathbb{E}\big[M_0^A\big] \ge \mathbb{E}\big[M_t^A\big] \ge \mathbb{E}\big[\Delta_-^A(t) A e^{\mu A - \frac{1}{2}\varepsilon t}\big].$$

So, making the cancellations,

$$\mathbb{E}\big[\Delta_-^A(t)\big] \le e^{\frac{1}{2}\varepsilon t} \sum_{i:\overline{X}_i(0)<0} e^{\mu \overline{X}_i(0)}.$$

Letting $A \to \infty$ and using the monotone convergence theorem concludes the proof of Lemma 3.4. $\square$

Putting together Lemmas 3.3 and 3.4, we get

$$\mathbb{E}[\Delta_t] \le \left(e^{\frac{1}{2}\varepsilon t} \sum_{i:\overline{X}_i(0)<0} e^{\mu \overline{X}_i(0)} + C e^{\frac{1}{2}\varepsilon t} \sum_{i:\overline{X}_i(0)>0} e^{\mu \overline{X}_i(0)} \right).$$

Combining with Lemmas 3.1 and 3.2, this yields

$$\begin{aligned}\mathbb{E}[W_t] &\le e^{\varepsilon t} \sum_{i:\overline{X}_i(0)<0} e^{\mu \overline{X}_i(0)} + C e^{\varepsilon t} \sum_{i:\overline{X}_i(0)>0} e^{\mu \overline{X}_i(0)} \\ &\le C e^{\varepsilon t} \sum_i e^{\mu \overline{X}_i(0)}.\end{aligned}$$

Therefore,

$$\mathbb{E}[W_t] \le e^{\varepsilon T} \sum_{i=1}^{N} e^{\mu \overline{X}_i(0)},$$

and because (35) holds,

$$\mathbb{P}(W_t \ge N) \le \frac{e^{\varepsilon T}}{N^{1-\delta}}.$$

By (39), we now have a bound for any fixed time $t \le T$,

$$\mathbb{P}\big(X_N(t) \ge \mu t\big) \le \mathbb{P}\big(\overline{X}_N(t) \ge \mu t\big) \le N^{-\kappa}. \tag{43}$$

This shows that $X_N(t) \ge \mu t$ with probability $1 - N^{-\kappa}$ for a fixed time $t \le T$. We now extend our argument so that we control the position of the minimum throughout the interval $[0, T]$ with high probability. Let $t_k = i(\log N)^{-1}, k = 1, \dots, T(\log N)$, so that t_k forms a regular partition of $[0, T]$ with spacings of size $1/(\log N)$. During each $[t_k, t_{k+1}]$, it is possible to check that $X_N(t)$ has small fluctuations. The key observation here is that $X_N(t)$ is piecewise Brownian and may jump to the right, but *never* to the left. Therefore, during the interval, the minimum cannot travel to the right of $\mu t_{k+1} + \zeta$ since it would then have to move at least ζ to the left in the rest of this interval (as we know it is to the left of μt_{k+1} at the end of the interval). This is unlikely, as displacements to the left are bounded by Brownian displacements, and the interval has a duration of $1/\log N$.

We now fix some large constant $\zeta > 0$ and define the bad events

$$B_k = \Big\{ \sup_{s \in [t_{k-1}, t_k]} X_N(s) \ge \mu t_k + \zeta \Big\}, \tag{44}$$

and the good events $G_k = \{X_N(t_k) \le \mu t_k\}$.

Given the Brunet–Derrida system at time t_{k-1}, consider the coupled free branching Brownian motion started from these N particles and for a particle $\overline{X}_i(t_k)$ at time t_k, let $\overline{Y}_i(s)$ be its ancestor at time $s \le t_k$. We see by the observation above that $X_N(t)$ can only jump to the right. Hence if G_{k+1} holds, in order for B_k to occur, it is necessary that the analogous $\overline{B}_k$ event for the free process also holds, namely

$$\overline{B}_k = \Big\{\exists 1 \le i \le \overline{N}(t_k), \sup_{s\in[t_{k-1},t_k]} \overline{Y}_i(s) \ge \mu t_k + \zeta, \overline{X}_i(t_k) \le \mu t_k\Big\}.$$

On this event, one of the particles being alive during the interval $[t_{k-1}, t_k]$ makes a displacement of at least ζ. Hence by a union bound and the many-to-one lemma [Lemma 2.1, applied with $T = t_k - t_{k-1}$ and $g((Y_s)_{s\le t_k-t_{k-1}}) = 1_{\{\sup_{s\le t_k-t_{k-1}} Y_s - Y_{t_k-t_{k-1}} \ge \zeta\}}$], and the reflection principle,

$$\begin{aligned}\mathbb{P}(\overline{B}_k) &\le N e^{t_k-t_{k-1}} \mathbb{P}\Big(\sup_{s\in[t_{k-1},t_k]} Z(s) - Z(t_k) \ge \zeta\Big)\\ &\le 2N e^{(\log N)^{-1}} \mathbb{P}\big(Z\big((\log N)^{-1}\big) \ge \zeta\big)\\ &\le 4N e^{-\frac{1}{2}\zeta^2 \log N}\end{aligned}$$

using (19), where $(Z(u), u \ge 0)$ is a standard Brownian motion. So for $\zeta = 2 \ge \sqrt{2\delta}$,

$$\mathbb{P}(B_k \cap G_k) \le 4N^{1-\delta}. \tag{45}$$

We can sum the conclusion of (43) and (45) over all $k = 1, \dots, T \log N$ to show that

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\le T}\{X_N(t) - \mu t\} \ge \zeta + \mu(\log N)^{-1}\Big) &\le \sum_{k=1}^{T(\log N)} \big(\mathbb{P}(B_k \cap G_k) + \mathbb{P}\big(G_k^{\complement}\big)\big)\\ &\le \frac{2e^{\varepsilon T} T \log N}{N^{1-\delta}}.\end{aligned}$$

This proves (37) for any $\delta' > \delta$ so the result follows.

4. Genealogical timescale: Proof of Theorem 1.11. Consider a Brunet–Derrida particle system with N particles in $\mathbb{R}^d$, started from an initial configuration satisfying (8) with $\mu = \sqrt{2}$ and some $x_0 \in \mathbb{R}^d$, and driven by the linear score function $s(x) = \langle x, \lambda\rangle = \widehat{x}$. Let $\xi > 0$ be small enough that $\delta' \overset{\text{def}}{=} \delta + \sqrt{2}\xi < 1$. Let $T = c_\delta(\log N)^3$ where c_δ will be chosen later on sufficiently small. Note that (8) implies that if $\widehat{Y}_n(t) = \widehat{X}_n(t) - \widehat{x}_0 + \xi \log N$, then

$$\sum_{n=1}^{N} e^{\sqrt{2}\widehat{Y}_n(0)} \le N^{\delta'}.$$

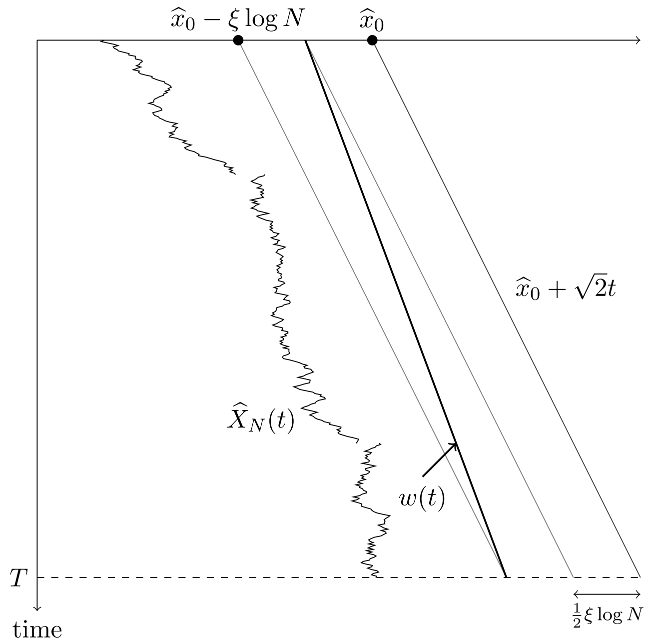

FIG. 5. *Diagram reference for proof of Theorem* 1.11.

Thus by (36), applied with $\mu = \sqrt{2}$, with probability at least $1 - N^{-\kappa}$ for some $\kappa > 0$, for all $t \le T$,

$$\widehat{X}_N(t) \le \widehat{x}_0 - \xi \log N + \sqrt{2}t. \tag{46}$$

On this event,

$$\widehat{X}_N(t) \le w(t) \stackrel{\text{def}}{=} \widehat{x}_0 - \frac{1}{2}\xi \log N + \mu' t, \tag{47}$$

where

$$\mu' = \sqrt{2} - \frac{\xi}{2c_\delta(\log N)^2}. \tag{48}$$

The function $w(t)$ is a linear boundary (see Figure 5) which will act as a *killing wall*. Note that w is chosen so that

$$w(0) = \widehat{x}_0 - \frac{1}{2}\xi \log N, \qquad w(T) = \widehat{x}_0 - \xi \log N + \sqrt{2}T,$$

and thus by (47), if a particle never hits $w(t)$ and starts to its right [i.e., $\widehat{X}_i(0) \ge w(0)$], then it will survive selection in the Brunet–Derrida system.

Now, let $Q_{\mu'}(y)$ be the probability that a branching Brownian motion starting from one particle at $y > 0$ survives killing at a wall $\mu' t$ for all time. By Theorem 1 in [3], we know that if $L := \pi/\sqrt{2 - \mu'^2}$ and $y - L \to \infty$ then $Q'_\mu(y) \to 1$. Note that for N large enough we have that $L \le \pi \log N \sqrt{c_\delta/(2\xi)}$. Let us fix c_δ small enough that $\pi\sqrt{c_\delta/(2\xi)} \le \xi/4$. (Note that ξ, and hence δ' depend only on δ, so c_δ depends only on δ). Then $y = (1/2)\xi \log N$ satisfies $y - L \to \infty$, hence $Q_{\mu'}(y) \to 1$. We deduce that every particle with initial fitness greater than $\widehat{x}_0$ has descendants alive at time T, as desired.

5. Asymptotic shape: Proof of Theorem 1.7. Let $H = \{x \in \mathbb{R}^d : \langle x, \lambda\rangle = 0\}$ be the orthogonal hyperplane to λ. Recall that when a Brunet–Derrida particle system is driven by a linear $s(x) = \langle x, \lambda\rangle$, its projection onto $\mathrm{Span}(\lambda)$ forms a one-dimensional Brunet–Derrida system. For such a system, we have already recalled (in Lemma 2.5) that for any initial condition, if $t > (1+\kappa)\log N$ for some $\kappa > 0$ and $a > 3\sqrt{2}$,

$$\lim_{N\to\infty} \mathbb{P}(\mathrm{diam}_t \le a \log N) = 1.$$

Now suppose that (8) initially holds for $x_0 = X_i(0)$ for some $1 \le i \le N$, and $\mu = \sqrt{2}$. Consider the system at time $u = \log\log N$. This particle branches at rate 1 so with probability at least $1 - 1/(\log N)$ it has branched at least once by time u. Moreover, by Lemma 2.2, all the descendants of this particle will be with high probability located [after projection onto $\mathrm{Span}(\lambda)$] at positions greater or equal to $\widehat{x}_0 - 2\log\log N$. We claim that furthermore at time u the system satisfies (8) again with high probability, with a slightly different value of δ, $\mu = \sqrt{2}$, and x_0 replaced by $x_0' = x_0 - 2(\log\log N)\lambda$. Indeed, letting $\overline{X}_n(t)$ denote a free, one-dimensional branching Brownian motion starting from N particles located in $\widehat{X}_1(0), \dots, \widehat{X}_N(0)$, we have, by linearity of the expectation and the many-to-one lemma [Lemma 2.1 with $T = u$ and $g((Y_s)_{s\le u}) = e^{\sqrt{2}Y_u}$],

$$\begin{aligned}
\mathbb{E}\left(\sum_{n=1}^N e^{\sqrt{2}(\widehat{X}_n(u) - \widehat{x}_0')}\right) &= (\log N)^{2\sqrt{2}}\mathbb{E}\left(\sum_{n=1}^N e^{\sqrt{2}(\overline{X}_n(u) - \widehat{x}_0)}\right) \\
&\le (\log N)^{2\sqrt{2}} \sum_{i=1}^N e^{\sqrt{2}(\widehat{X}_i(0) - \widehat{x}_0)} e^u \mathbb{E}_0\big(e^{\sqrt{2}B_u}\big) \\
&\le (\log N)^4 e^{2u} N^\delta = (\log N)^6 N^\delta,
\end{aligned}$$

where B_t is a standard one-dimensional Brownian motion started at 0. Thus by Markov's inequality $\sum_{n=1}^N e^{\sqrt{2}(\widehat{X}_n(u) - \widehat{x}_0')} \le (\log N)^7 N^\delta$ with probability at least $1 - 1/(\log N)$. Consequently, (8) is satisfied at time u for some $\delta < \delta' < 1$, $\mu = \sqrt{2}$, and x_0 replaced by x_0'. Since we have two particles with fitness greater than $\widehat{x}_0'$ also with probability greater than $1 - 1/(\log N)$, we deduce by Theorem 1.11 that both of these particles have descendants alive with probability tending to 1 as $N \to \infty$ at time $T = c_{\delta'}(\log N)^3$.

Let E be the above event and call $x = X_i(u)$, $y = X_j(u)$ the positions of two particles at time u such that $\widehat{x} \ge \widehat{x}_0'$ and $\widehat{y} \ge \widehat{x}_0'$. Write $X_i(T)$ and $X_j(T)$, with a slight abuse of notation, for any two arbitrarily chosen respective descendants at time T, and $Y_i(t)$, $Y_j(t)$ for the position of their ancestors at time $u \le t \le T$. Then note that if p_H is the orthogonal projection onto H, $(p_H(Y_i(t)) - x, u \le t \le T)$ and $(p_H(Y_j(t)) - y, u \le t \le T)$ are independent $(d-1)$-dimensional Brownian motions on H on the time interval $[u, T]$ (see the end of the proof of Theorem 1.4). Thus by the triangle inequality, on the event E,

$$\begin{aligned}\mathrm{diam}_T^{\perp} &\ge \| p_H(X_i(T)) - p_H(X_j(T)) \| \\ &\ge \| (p_H(X_i(T)) - x) + (y - p_H(X_j(T))) \| - \|x - y\| \\ &\ge (T-u)^{1/2} X - a \log N,\end{aligned}$$

where $a > 3\sqrt{2}$ and X is the norm of the sum of two independent standard $(d-1)$-Gaussian random variables. In particular,

$$\liminf_{\eta \to 0} \liminf_{N \to \infty} \mathbb{P}(\mathrm{diam}_T^{\perp} \ge \eta (\log N)^{3/2}) = 1,$$

as desired.

Acknowledgements. We are grateful to Julien Berestycki for a number of fruitful conversations at several stages of this project. In particular, we learned of the potential relation between multidimensional Brunet–Derrida systems and the role of recombination from him, and we are grateful to have been shown a draft of [5] which raised that issue. We thank the referees for spotting some minor issues with draft versions of this paper and some suggestions which improved the presentation.

REFERENCES

[1] BELL, G. (1982). *The Masterpiece of Nature*. Univ. California Press, Berkeley, CA.

[2] BÉRARD, J. and GOUÉRÉ, J.-B. (2010). Brunet–Derrida behavior of branching-selection particle systems on the line. *Comm. Math. Phys.* **298** 323–342. MR2669438

[3] BERESTYCKI, J., BERESTYCKI, N. and SCHWEINSBERG, J. (2011). Survival of near-critical branching Brownian motion. *J. Stat. Phys.* **143** 833–854. MR2811463

[4] BERESTYCKI, J., BERESTYCKI, N. and SCHWEINSBERG, J. (2013). The genealogy of branching Brownian motion with absorption. *Ann. Probab.* **41** 527–618. MR3077519

[5] BERESTYCKI, J. and YU, F. Unpublished work.

[6] BERESTYCKI, N. (2009). *Recent Progress in Coalescent Theory*. *Ensaios Matemáticos* [*Mathematical Surveys*] **16**. Sociedade Brasileira de Matemática, Rio de Janeiro. MR2574323

[7] BRUNET, E. and DERRIDA, B. (1997). Shift in the velocity of a front due to a cutoff. *Phys. Rev. E* (3) **56** 2597–2604. MR1473413

[8] BRUNET, E. and DERRIDA, B. (1999). Microscopic models of traveling wave equations. *Comput. Phys. Commun.* **121–122** 376–381.

[9] BRUNET, É. and DERRIDA, B. (2001). Effect of microscopic noise on front propagation. *J. Stat. Phys.* **103** 269–282. MR1828730

[10] BRUNET, E., DERRIDA, B., MUELLER, A. H. and MUNIER, S. (2006). Noisy traveling waves: Effect of selection on genealogies. *Europhys. Lett.* **76** 1–7. MR2299937

[11] BRUNET, É., DERRIDA, B., MUELLER, A. H. and MUNIER, S. (2007). Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization. *Phys. Rev. E* (3) **76** 041104, 20. MR2365627

[12] BURT, A. (2000). Perspective: Sex, recombination and the efficacy of selection—Was Weismann right? *Evolution* **54** 337–351.

[13] DURRETT, R. and REMENIK, D. (2011). Brunet–Derrida particle systems, free boundary problems and Wiener–Hopf equations. *Ann. Probab.* **39** 2043–2078. MR2932664

[14] ETHERIDGE, A. M. (2000). *An Introduction to Superprocesses*. *University Lecture Series* **20**. Amer. Math. Soc., Providence, RI. MR1779100

[15] GOLDSCHMIDT, C. and MARTIN, J. B. (2005). Random recursive trees and the Bolthausen–Sznitman coalescent. *Electron. J. Probab.* **10** 718–745. MR2164028

[16] GROISMAN, P. and JONCKHEERE, M. (2013) Front propagation and quasi-stationary distributions: The same selection principle? Available at arXiv:1304.4847.

[17] HARRIS, J. W. and HARRIS, S. C. (2007). Survival probabilities for branching Brownian motion with absorption. *Electron. Commun. Probab.* **12** 81–92. MR2300218

[18] ITÔ, K. and MCKEAN, H. P. JR. (1965). *Diffusion Processes and Their Sample Paths*. *Die Grundlehren der Mathematischen Wissenschaften* **Band 125**. Springer, New York. MR0199891

[19] MAILLARD, P. (2012). Branching Brownian motion with selection. Ph.D. thesis, Univ. Pierre et Marie Curie. Available at arXiv:1210.3500.

[20] MAILLARD, P. (2016). Speed and fluctuations of N-particle branching Brownian motion with spatial selection. *Probab. Theory Related Fields* **166** 1061–1173. MR3568046

[21] ROYNETTE, B., VALLOIS, P. and YOR, M. (2009). Penalisations of multidimensional Brownian motion. VI. *ESAIM Probab. Stat.* **13** 152–180. MR2518544

[22] WEISMANN, A. (1889). The significance of sexual reproduction in the theory of natural selection. In *Essays upon Heredity and Kindred Biological Problems* (E. B. Poulton, S. Schönland and A. E. Shipley, eds.) 251–332. Clarendon Press, Oxford.

[23] WILLIAMS, G. C. (1966). *Adaptation and Natural Selection*. Princeton Univ. Press, Princeton, NJ.

STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE
WILBERFORCE RD.
CAMBRIDGE
CB3 0WB
UNITED KINGDOM
E-MAIL: N.Berestycki@statslab.cam.ac.uk
lzz20@statslab.cam.ac.uk