

A FLEMING-VIOT PARTICLE REPRESENTATION OF THE DIRICHLET LAPLACIAN

Krzysztof Burdzy Robert Hołyst Peter March

Abstract: We consider a model with a large number N of particles which move according to independent Brownian motions. A particle which leaves a domain D is killed; at the same time, a different particle splits into two particles. For large N , the particle distribution density converges to the normalized heat equation solution in D with Dirichlet boundary conditions. The stationary distributions converge as $N \rightarrow \infty$ to the first eigenfunction of the Laplacian in D with the same boundary conditions.

1. Introduction. Our article is closely related to a model studied by Burdzy, Hołyst, Ingberman and March (1996) using heuristic and numerical methods. Although we are far from proving conjectures stated in that article, the present paper seems to lay solid theoretical foundations for further research in this direction. The model is related to many known ideas in probability and physics—we review them in the Appendix (Section 3). We present the model and state our main results in this section. Section 2 is entirely devoted to proofs.

We will be concerned with a multiparticle process. The motion of an individual particle will be represented by Brownian motion in an open subset of \mathbf{R}^d . Probably all our results can be generalized to other processes. However, the present paper is motivated by the article of Burdzy, Hołyst, Ingberman and March (1996) whose results are very specific to Brownian motion and so we will limit ourselves to this special case. We note that Theorem 1.1 below uses only the strong Markov property of the process representing a particle and the continuity of the density of the hitting time of a set. Theorem 1.2 is similarly easy to generalize. At the other extreme, the proof of Theorem 1.4 uses Brownian properties in an essential way and would be hard to generalize. It might be therefore of some interest to see if Theorem 1.4 holds for a large class of processes.

Research partially supported by NSF grant DMS-9700721, KBN grant 2P03B12516 and Maria Curie-Skłodowska Joint Fund II.

Consider an open set $D \subset \mathbf{R}^d$ and an integer $N \geq 2$. Let $\mathbf{X}_t = (X_t^1, X_t^2, \dots, X_t^N)$ be a process with values in D^N whose evolution can be described as follows. Suppose $\mathbf{X}_0 = (x^1, x^2, \dots, x^N) \in D^N$. The processes $X_t^1, X_t^2, \dots, X_t^N$ evolve as independent Brownian motions until the first time τ_1 when one of them, say, X^j hits the boundary of D . At this time one of the remaining particles is chosen in a uniform way, say, X^k , and the process X^j jumps at time τ_1 to $X_{\tau_1}^k$. The processes $X_t^1, X_t^2, \dots, X_t^N$ continue as independent Brownian motions after time τ_1 until the first time $\tau_2 > \tau_1$ when one of them hits ∂D . At the time τ_2 , the particle which approaches the boundary jumps to the current location of a randomly (uniformly) chosen particle among the ones strictly inside D . The subsequent evolution of the process \mathbf{X}_t proceeds along the same lines.

Before we start to study properties of \mathbf{X}_t , we have to check if the process is well defined. Since the distribution of the hitting time of ∂D has a continuous density, only one particle can hit ∂D at time τ_k , for every k , a.s. However, the process \mathbf{X}_t can be defined for all $t \geq 0$ using the informal recipe given above only if $\tau_k \rightarrow \infty$ as $k \rightarrow \infty$. This is because there is no obvious way to continue the process \mathbf{X}_t after the time $\tau_\infty = \lim_{k \rightarrow \infty} \tau_k$ if $\tau_\infty < \infty$. Hence, the question of the finiteness of τ_∞ has a fundamental importance for our model.

Theorem 1.1. *We have $\lim_{k \rightarrow \infty} \tau_k = \infty$ a.s.*

Consider an open set D which has more than one connected component. If at some time t all processes X_t^k belong to a single connected component of D then they will obviously stay in the same component from then on. Will there be such a time t ? The answer is yes, according to the theorem below, and so we could assume without loss of generality that D is a connected set, especially in Theorem 1.4.

Theorem 1.2. *With probability 1, there exists $t < \tau_\infty$ such that all processes X_t^k belong to a single connected component of D at time t .*

Before we continue the presentation of our results, we will provide a slightly more formal description of the process \mathbf{X}_t than that at the beginning of the introduction. The fully rigorous definition would be a routine but tedious task and so it is left to the reader. One can show that given $(x^1, x^2, \dots, x^N) \in D^N$, there exists a strong Markov process \mathbf{X}_t , unique in the sense of distribution, with the following properties. The process starts from $\mathbf{X}_0 = (x^1, x^2, \dots, x^N)$, a.s. Let

$$\tau_1 = \inf_{1 \leq m \leq N} \inf\{t > 0 : \lim_{s \rightarrow t^-} X_s^m \in D^c\},$$

and for $n \geq 1$,

$$\tau_{n+1} = \inf_{1 \leq m \leq N} \inf\{t > \tau_n : \lim_{s \rightarrow t-} X_s^m \in D^c\}.$$

Then $\tau_{n+1} > \tau_n$ for every $n \geq 1$, a.s. For every $n \geq 1$, there exists a unique k_n such that $\lim_{s \rightarrow \tau_n-} X_s^{k_n} \in D^c$, a.s. We have $X_{\tau_n}^m = X_{\tau_n-}^m$, for every $m \neq k_n$. For some random $j = j(n, k_n) \neq k_n$ we have $X_{\tau_n}^{k_n} = X_{\tau_n}^j$. The distribution of $j(n, k_n)$ is uniform on the set $\{1, 2, \dots, N\} \setminus \{k_n\}$ and independent of $\{\mathbf{X}_t, 0 \leq t < \tau_n\}$. For every n , the process $\{\mathbf{X}_{(t \wedge \tau_{n+1})-}, t \geq \tau_n\}$ is a Brownian motion on D^N stopped at the hitting time of ∂D^N .

Let $P_t^D(x, dy)$ be the transition probability for the Brownian motion killed at the time of hitting of D^c . Given a probability measure $\mu_0(dx)$ on D , we define measures μ_t for $t > 0$ by

$$\mu_t(A) = \frac{\int_D P_t^D(x, A) \mu_0(dx)}{\int_D P_t^D(x, D) \mu_0(dx)}, \quad (1.1)$$

for open sets $A \subset D$. Note that μ_t is a probability measure, for every $t \geq 0$. Let $\delta_x(dy)$ be the probability measure with the unit atom at x . We will write $\mathcal{X}_t^N(dy) = (1/N) \sum_{k=1}^N \delta_{X_t^k}(dy)$ to denote the empirical (probability) distribution representing the particle process \mathbf{X}_t .

We will say that D has a regular boundary if for every $x \in \partial D$, Brownian motion starting from x hits D^c for arbitrarily small $t > 0$, a.s.

Theorem 1.3. *Suppose that D is bounded and has a regular boundary. Fix a probability distribution μ_0 on D and recall the definition (1.1). Suppose that for every N , the initial distribution \mathcal{X}_0^N is a non-random measure μ_0^N . If the measures μ_0^N converge as $N \rightarrow \infty$ to μ_0 then for every fixed $t > 0$ the empirical distributions \mathcal{X}_t^N converge to μ_t in the sense that for every set $A \subset D$, the sequence $\mathcal{X}_t^N(A)$ converges to $\mu_t(A)$ in probability.*

The regularity of ∂D seems to be only a technical assumption, i.e., Theorem 1.3 is likely to hold without this assumption.

We conjecture that for any $S > 0$, the measure-valued processes $\{\mathcal{X}_t^N(\cdot), 0 \leq t \leq S\}$ converge to $\{\mu_t(\cdot), 0 \leq t \leq S\}$ in the Skorohod topology, as $N \rightarrow \infty$. The arguments presented in this paper do not seem to be sufficient to justify this claim.

One may wonder whether $E\mathcal{X}_t^N(A) = \mu_t(A)$ for all sets A and $t > 0$ if we assume that $\mathcal{X}_0^N = \mu_0$. One can find intuitive arguments both for and against this claim but none of them seemed to be quite clear to us. We will have to resort to brute calculation to show that the statement is false. The word “brute” refers only to the lack of a clear intuitive

explanation and not to the difficulty of the example which is in fact quite elementary (see Example 2.1 in Section 2). The example is concerned with a process on a finite state space. We presume that a similar example can be based on the Brownian motion process.

We will say that an open set $D \subset \mathbf{R}^d$ satisfies the interior ball condition if for some $r > 0$ and every $x \in D$ there exists an open ball $B(y, r) \subset D$ such that $x \in B(y, r)$.

Theorem 1.4. *Suppose that $D \subset \mathbf{R}^d$ is a bounded domain, has a regular boundary and satisfies the interior ball condition.*

(i) *For every N , there exists a unique stationary probability measure \mathbf{M}^N for \mathbf{X}_t . The process \mathbf{X}_t converges to its stationary distribution exponentially fast, i.e., there exists $\lambda > 0$ such that for every $A \subset D^N$, and every $\mathbf{x} \in D^N$,*

$$\lim_{t \rightarrow \infty} e^{\lambda t} |P^{\mathbf{x}}(\mathbf{X}_t \in A) - \mathbf{M}^N(A)| = 0.$$

(ii) *Let $\mathcal{X}_{\mathbf{M}}^N$ be the stationary empirical measure, i.e., let $\mathcal{X}_{\mathbf{M}}^N$ have the same distribution as $(1/N) \sum_{k=1}^N \delta_{X_t^k}(dy)$, assuming that \mathbf{X}_t has the distribution \mathbf{M}^N . Let $\varphi(x)$ be the first eigenfunction for the Laplacian in D with the Dirichlet boundary conditions, normalized so that $\int_D \varphi = 1$. Then the random measures $\mathcal{X}_{\mathbf{M}}^N$ converge as $N \rightarrow \infty$ to the (non-random) measure with density $\varphi(x)$, in the sense of weak convergence of random measures.*

We are grateful to David Aldous, Wilfrid Kendall, Tom Kurtz, Jeff Rosenthal, Dan Stroock, Kathy Temple and Richard Tweedie for very useful advice. We would like to thank the anonymous referee for many suggestions for improvement.

2. Proofs. This section is devoted to proofs of the main results. It also contains an example related to Theorem 1.3.

Proof of Theorem 1.1. Fix an arbitrary $S < \infty$. Let B_t be a Brownian motion, and

$$h(x, t) = P(\inf\{s > t : B_s \notin D\} > S \mid B_t = x).$$

We will first prove the theorem for $N = 2$ as this special case presents the main idea of the proof in a clear way. Let $M_t = h(X_{t-}^1, t) + h(X_{t-}^2, t)$. Consider an arbitrary $y \in D$ and assume that $X_0^1 = X_0^2 = y$. Let $a = h(y, 0)$ and $\tau_* = \tau_1 \wedge S$. An application of the

optional stopping theorem to the martingale $M_{t \wedge \tau_*}$ gives

$$\begin{aligned} 2h(y, 0) &= EM_0 = EM_{\tau_*} \\ &= E(M_{\tau_*} \mid \tau_* = S)P(\tau_* = S) + E(M_{\tau_*} \mid \tau_* < S)P(\tau_* < S) \\ &= 2 \cdot P(\tau_* = S) + E(M_{\tau_*} \mid \tau_1 < S)P(\tau_1 < S) \\ &= 2 \cdot h(y, 0)^2 + E(M_{\tau_*} \mid \tau_1 < S)(1 - h(y, 0)^2). \end{aligned}$$

From this we obtain

$$E(M_{\tau_*} \mid \tau_1 < S) = \frac{2h(y, 0) - 2h(y, 0)^2}{1 - h(y, 0)^2} = \frac{2h(y, 0)}{1 + h(y, 0)} \geq h(y, 0).$$

The process X_t^k which hits ∂D at time τ_1 jumps to the location of $X_{\tau_1}^{3-k}$, so we have

$$\begin{aligned} E(h(X_{\tau_1}^1, \tau_1) + h(X_{\tau_1}^2, \tau_1) \mid \tau_1 < S) &\geq 2h(y, 0) \\ &= E(h(X_0^1, 0) + h(X_0^2, 0)). \end{aligned}$$

By applying the strong Markov property at the stopping time τ_1 we obtain

$$E(h(X_{\tau_2}^1, \tau_2) + h(X_{\tau_2}^2, \tau_2) \mid \tau_2 < S) \geq E(h(X_0^1, 0) + h(X_0^2, 0)).$$

By induction, for all $k \geq 1$,

$$E(h(X_{\tau_k}^1, \tau_k) + h(X_{\tau_k}^2, \tau_k) \mid \tau_k < S) \geq E(h(X_0^1, 0) + h(X_0^2, 0)) = 2a. \quad (2.1)$$

Let $J_k = h(X_{\tau_k}^1, \tau_k) + h(X_{\tau_k}^2, \tau_k)$. Since $h(x, t) \leq 1$, we have $J_k \leq 2$. Hence,

$$\begin{aligned} E(J_k \mid \tau_k < S) &\leq 2P(J_k \geq a \mid \tau_k < S) + aP(J_k < a \mid \tau_k < S) \\ &= P(J_k \geq a \mid \tau_k < S)(2 - a) + a, \end{aligned}$$

and so, using (2.1),

$$\begin{aligned} P(h(X_{\tau_k}^1, \tau_k) = h(X_{\tau_k}^2, \tau_k) \geq a/2 \mid \tau_k < S) \\ = P(J_k \geq a \mid \tau_k < S) \geq \frac{E(J_k \mid \tau_k < S) - a}{2 - a} \geq \frac{2a - a}{2 - a} = \frac{a}{2 - a}. \end{aligned}$$

It follows that

$$\begin{aligned} &P(\tau_{k+1} \geq S \mid \tau_k < S) \\ &= P(\inf\{s > \tau_k : X_{s-}^1 \notin D\} > S, \inf\{s > \tau_k : X_{s-}^2 \notin D\} > S \mid \tau_k < S) \\ &= \int [P(\inf\{s > \tau_k : X_s^1 \notin D\} > S \mid X_{\tau_k}^1 = x)]^2 P(X_{\tau_k}^1 \in dx \mid \tau_k < S) \\ &= \int h(x, \tau_k)^2 P(X_{\tau_k}^1 \in dx \mid \tau_k < S) \\ &\geq (a/2)^2 \cdot \frac{a}{2 - a} = \frac{a^3}{8 - 4a}. \end{aligned}$$

This implies that

$$P(\tau_{k+1} < S) = \prod_{j=1}^k P(\tau_{j+1} < S \mid \tau_j < S) \leq \left(1 - \frac{a^3}{8 - 4a}\right)^k,$$

and so

$$P(\tau_\infty < S) = 0.$$

Recall that we have assumed that $X_0^1 = X_0^2$. If X_0^1 is not equal to X_0^2 , we can apply the argument to the post- τ_1 process to see that $P(\tau_\infty < S) = 0$ for every starting position of \mathbf{X}_t . Since $S < \infty$ is arbitrarily large, the proof of the theorem is complete in the special case $N = 2$.

Now we generalize the argument to arbitrary $N \geq 2$. Recall S, τ_* and $h(x, t)$ from the first part of the proof. Let

$$M_t = \sum_{k=1}^N h(X_{t-}^k, t),$$

and $a_k = h(X_0^k, 0)$. Then

$$\begin{aligned} \sum_{k=1}^N a_k &= EM_0 = EM_{\tau_*} \\ &= E(M_{\tau_*} \mid \tau_* = S)P(\tau_* = S) + E(M_{\tau_*} \mid \tau_* < S)P(\tau_* < S) \\ &= N \cdot P(\tau_* = S) + E(M_{\tau_*} \mid \tau_1 < S)P(\tau_1 < S) \\ &= N \prod_{k=1}^N a_k + E(M_{\tau_*} \mid \tau_1 < S) \left(1 - \prod_{k=1}^N a_k\right). \end{aligned}$$

From this we obtain

$$E(M_{\tau_*} \mid \tau_1 < S) = \frac{\sum_{k=1}^N a_k - N \prod_{k=1}^N a_k}{1 - \prod_{k=1}^N a_k}. \quad (2.2)$$

Our immediate goal is to prove that the right hand side of (2.2) is bounded below by $((N-1)/N) \sum_{k=1}^N a_k$.

The derivative of the function $x \rightarrow (\sum_{k=1}^N a_k - Nx)/(1-x)$ is equal to

$$\frac{\sum_{k=1}^N a_k - N}{(1-x)^2}.$$

The derivative is non-positive since $\sum_{k=1}^N a_k \leq N$. If we let $b = \sum_{k=1}^N a_k / N$ then for a fixed value of b , the value of the product $\prod_{k=1}^N a_k$ is maximized if we take $a_k = b$ for all k . These facts imply that

$$\frac{\sum_{k=1}^N a_k - N \prod_{k=1}^N a_k}{1 - \prod_{k=1}^N a_k} \geq \frac{\sum_{k=1}^N a_k - N \cdot b^N}{1 - b^N} = \frac{Nb - N \cdot b^N}{1 - b^N} = Nb \cdot \frac{1 - b^{N-1}}{1 - b^N}. \quad (2.3)$$

We will show that

$$Nb \cdot \frac{1 - b^{N-1}}{1 - b^N} \geq (N-1)b, \quad (2.4)$$

for $b \in [0, 1)$. The last inequality is equivalent to

$$N(1 - b^{N-1}) \geq (N-1)(1 - b^N).$$

After multiplying out and regrouping the terms we obtain

$$1 + Nb^N - Nb^{N-1} - b^N \geq 0. \quad (2.5)$$

The function $f(b) = 1 + Nb^N - Nb^{N-1} - b^N$ has the derivative $f'(b) = N(N-1)b^{N-2}(b-1)$ which is negative for $b < 1$. Since $f(1) = 0$, we have $f(b) \geq 0$ for $b \in [0, 1)$, i.e., (2.5) holds. Consequently, (2.4) is true as well.

Combining (2.2), (2.3) and (2.4) yields

$$E(M_{\tau_*} \mid \tau_1 < S) = \frac{\sum_{k=1}^N a_k - N \prod_{k=1}^N a_k}{1 - \prod_{k=1}^N a_k} \geq (N-1)b = \frac{N-1}{N} \sum_{k=1}^N a_k.$$

The process X^k which hits the boundary at time τ_1 jumps to the location of a process X^j , uniformly chosen from other processes. Hence,

$$\begin{aligned} E\left(\sum_{k=1}^N h(X_{\tau_*}^k, t) \mid \tau_1 < S\right) &= \left(1 + \frac{1}{N-1}\right) E\left(\sum_{k=1}^N h(X_{\tau_*-}^k, t) \mid \tau_1 < S\right) \\ &= \frac{N}{N-1} E(M_{\tau_*} \mid \tau_1 < S) \geq \frac{N}{N-1} \cdot \frac{N-1}{N} \sum_{k=1}^N a_k = \sum_{k=1}^N a_k. \end{aligned}$$

By induction and the strong Markov property applied at τ_k 's, we have for every $k \geq 1$,

$$E\left(\sum_{k=1}^N h(X_{\tau_k}^k, t) \mid \tau_k < S\right) \geq \sum_{k=1}^N a_k = Nb. \quad (2.6)$$

Let $J_k = \sum_{j=1}^N h(X_{\tau_k}^j, \tau_k)$. Since $h(x, t) \leq 1$, we have $J_k \leq N$. Recall that $b = (1/N) \sum_{k=1}^N a_k$. Hence,

$$\begin{aligned} E(J_k \mid \tau_k < S) &\leq NP(J_k \geq b \mid \tau_k < S) + bP(J_k < b \mid \tau_k < S) \\ &= P(J_k \geq b \mid \tau_k < S)(N - b) + b. \end{aligned}$$

This and (2.6) imply that

$$P(J_k \geq b \mid \tau_k < S) \geq \frac{E(J_k \mid \tau_k < S) - b}{N - b} \geq \frac{Nb - b}{N - b}.$$

It follows that

$$P(\exists j : h(X_{\tau_k}^j, \tau_k) \geq b/N \mid \tau_k < S) \geq \frac{Nb - b}{N - b}. \quad (2.7)$$

Fix some $t \in (0, S)$. Suppose that $h(X_t^j, t) \geq b/N$ for some j and assume that j is the smallest number with this property. Let $T = \inf\{s > t : h(X_s^j, s) \notin (b/(2N), 1)\}$. Note that $h(X_T^j, T) = 1$ if and only if $T = S$. The process $h(X_s^j, s)$ is a martingale on the interval (t, T) . By the martingale property and the optional stopping theorem, the probability of not hitting $b/(2N)$ before time S is greater than or equal to

$$\frac{b/N - b/(2N)}{1 - b/(2N)} = \frac{b}{2N - b}. \quad (2.8)$$

Consider the event A that $h(X_t^j, t) \geq b/N$ at some time t and that the process $h(X_s^j, s)$ does not hit $b/(2N)$ between t and S . Given this event, for any $k \neq j$, the process X^k may jump at most once before time S with probability greater than $(1/(N-1)) \cdot (b/(2N))$, independent of other processes X^m , $m \neq k, j$. To see this, observe that X^k might not hit ∂D before time S at all; or it may hit ∂D , then jump to the location of X^j with probability $1/(N-1)$. If the jump takes place, the process X^k lands at some time u at a place where we have $h(X_u^k, u) = h(X_u^j, u) \geq b/(2N)$, because we are assuming that A holds. The definition of the function h now implies that after u , the process X^k will not hit ∂D before time S , with probability greater than $b/(2N)$.

Multiplying the probabilities for all $k \neq j$ and using (2.8), we conclude that if we have $h(X_t^j, t) \geq b/N$ then the probability that there will be at most $N-1$ jumps (counting all particles) before time S is greater than

$$p_0 = \frac{b}{2N - b} \cdot \left(\frac{b}{2N(N-1)} \right)^{N-1}.$$

Hence, in view of (2.7),

$$\begin{aligned} P(\tau_{k+N} \geq S \mid \tau_k < S) &\geq P(\exists j : h(X_{\tau_k}^j, \tau_k) \geq b/N \mid \tau_k < S)P(\tau_{k+N} > S \mid \exists j : h(X_{\tau_k}^j, \tau_k) \geq b/N) \\ &\geq \frac{Nb - b}{N - b} \cdot p_0. \end{aligned}$$

Thus

$$P(\tau_{(m+1)N} < S) = \prod_{j=1}^m P(\tau_{(j+1)N} < S \mid \tau_{jN} < S) \leq \left(1 - \frac{Nb - b}{N - b} \cdot p_0\right)^m,$$

and so

$$P(\tau_\infty < S) = 0.$$

Since S is arbitrarily large, the proof is complete. \square

Proof of Theorem 1.2. Fix arbitrary points $x^j \in D$ and suppose that $X_0^j = x^j$ for all j . Let τ^j be the first jump time for the process X^j . Since there are $N!$ permutations of $\{1, 2, \dots, N\}$, there exists a permutation (j_1, j_2, \dots, j_N) , such that

$$P(\tau^{j_1} < \tau^{j_2} < \dots < \tau^{j_N}) \geq 1/N!.$$

In order to simplify the notation we will assume that $(j_1, j_2, \dots, j_N) = (1, 2, \dots, N)$. Thus we have

$$P(\tau^1 < \tau^2 < \dots < \tau^N) \geq 1/N!$$

and

$$P(\tau^1 < \tau^2 < \dots < \tau^N, X_{\tau^1}^1 = X_{\tau^1}^N) \geq \frac{1}{N \cdot N!}.$$

Let τ_2^j denote the time of the second jump of process X_t^j . By independence,

$$\begin{aligned} &P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N \mid \tau^1, \tau^2, \dots, \tau^{N-1}, X_{\tau^1}^1 = X_{\tau^1}^N) \\ &= P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \tau^N, X_{\tau^1}^1 = X_{\tau^1}^N \mid \tau^1, \tau^2, \dots, \tau^{N-1}, X_{\tau^1}^1 = X_{\tau^1}^N) \\ &\quad \times P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \tau_2^1, X_{\tau^1}^1 = X_{\tau^1}^N \mid \tau^1, \tau^2, \dots, \tau^{N-1}, X_{\tau^1}^1 = X_{\tau^1}^N) \\ &= P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \tau^N, X_{\tau^1}^1 = X_{\tau^1}^N \mid \tau^1, \tau^2, \dots, \tau^{N-1}, X_{\tau^1}^1 = X_{\tau^1}^N)^2. \end{aligned}$$

It follows that,

$$\begin{aligned}
& P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N) \\
&= EP(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N \mid \tau^1, \tau^2, \dots, \tau^{N-1}, X_{\tau^1}^1 = X_{\tau^1}^N) \\
&= E[P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \tau^N, X_{\tau^1}^1 = X_{\tau^1}^N \mid \tau^1, \tau^2, \dots, \tau^{N-1}, X_{\tau^1}^1 = X_{\tau^1}^N)^2] \\
&\geq [EP(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \tau^N, X_{\tau^1}^1 = X_{\tau^1}^N \mid \tau^1, \tau^2, \dots, \tau^{N-1}, X_{\tau^1}^1 = X_{\tau^1}^N)]^2 \\
&= P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \tau^N, X_{\tau^1}^1 = X_{\tau^1}^N)^2 \\
&\geq \frac{1}{(N \cdot N!)^2}.
\end{aligned}$$

We proceed by induction. Let us display one induction step. We start with

$$P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N) \geq \frac{1}{N} \cdot \frac{1}{(N \cdot N!)^2}.$$

Then we observe that

$$\begin{aligned}
& P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1, \tau_2^2), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N \\
&\quad \mid \tau^1, \tau^2, \dots, \tau^{N-1}, \tau_2^1, X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N,) \\
&= P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N \\
&\quad \mid \tau^1, \tau^2, \dots, \tau^{N-1}, \tau_2^1, X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N,) \\
&\quad \times P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau_2^2, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N \\
&\quad \mid \tau^1, \tau^2, \dots, \tau^{N-1}, \tau_2^1, X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N,) \\
&= P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N \\
&\quad \mid \tau^1, \tau^2, \dots, \tau^{N-1}, \tau_2^1, X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N,)^2.
\end{aligned}$$

From this we deduce that

$$\begin{aligned}
& P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1, \tau_2^2), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N) \\
&= EP(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1, \tau_2^2), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N \\
&\quad \mid \tau^1, \tau^2, \dots, \tau^{N-1}, \tau_2^1, X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N,) \\
&= E[P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N \\
&\quad \mid \tau^1, \tau^2, \dots, \tau^{N-1}, \tau_2^1, X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N,)^2] \\
&\geq [EP(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N \\
&\quad \mid \tau^1, \tau^2, \dots, \tau^{N-1}, \tau_2^1, X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N,)]^2 \\
&= P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1), X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N)^2 \\
&\geq \left(\frac{1}{N} \cdot \frac{1}{(N \cdot N!)^2} \right)^2.
\end{aligned}$$

Proceeding in this way, we can prove that

$$\begin{aligned} P(\tau^1 < \tau^2 < \dots < \tau^{N-1} < \min(\tau^N, \tau_2^1, \tau_2^2, \dots, \tau_2^{N-1}), \\ X_{\tau^1}^1 = X_{\tau^1}^N, X_{\tau^2}^2 = X_{\tau^2}^N, \dots, X_{\tau^{N-1}}^{N-1} = X_{\tau^{N-1}}^N) \geq c_1, \end{aligned} \quad (2.9)$$

where $c_1 > 0$ is a constant which depends on N but not on the starting position of X_t^k 's. If the event in (2.9) occurs then at time τ_{N-1} all particles are present in the same connected component of D as X^N . They will stay in this connected component of D forever. If the event in (2.9) does not occur then we wait until the time $\max(\tau^N, \tau_2^1, \tau_2^2, \dots, \tau_2^{N-1})$ and restart our argument, using the strong Markov property. We can construct in this way a sequence of events whose conditional probabilities (given the outcomes of previous "trials") are bounded below by c_1 . With probability 1, at least one of these events will occur and so all particles will end up in a single connected component of D . \square

Proof of Theorem 1.3. Fix some $S \in (0, \infty)$. We will prove that \mathcal{X}_S^N converges to μ_S . Our proof will consist of three parts.

Part 1. In this part of the proof, we will define the tree of *descendants* of a particle and estimate its size.

Fix an arbitrarily small $\varepsilon_1 > 0$. Let B_t denote a Brownian motion and $T_A = \inf\{t > 0 : B_t \in A\}$. Find open subsets A_1, A_2 and A_3 of D such that $A_1 \subset A_2 \subset A_3$, $\mu_0(A_1) \geq \varepsilon_1 > 0$, and for some $p_1, p_2 > 0$,

$$\inf_{x \in A_1} P^x(T_{A_2^c} > S) > p_1 \quad \text{and} \quad \inf_{x \in A_2} P^x(T_{A_3^c} > S) > p_2.$$

We would like to set aside a small family of particles starting from A_1 and containing about $\varepsilon_1 N$ particles. Since the measure μ_0 may be purely atomic with all atoms greater than ε_1 , we cannot designate the particles in that family just by their starting position. We have assumed that the measures \mathcal{X}_0^N converge to μ_0 , so we must have $\mathcal{X}_0^N(A_1) > \varepsilon_1/2$ for large N . Let $[b]$ denote the integer part of a number b . For each sufficiently large N , we arbitrarily choose $[N\varepsilon_1/2]$ particles with the property that their starting positions lie inside A_1 . The choice is deterministic (non-random). The family of all $N - [N\varepsilon_1/2]$ remaining particles will be called \mathcal{H} . By the law of large numbers, for any $p_3 < 1$ and sufficiently large N , more than $N\varepsilon_1 p_1/4$ particles will stay inside A_2 until time S , with probability greater than p_3 .

We will say that a particle has label k if its motion is represented by X_t^k . We will identify the families \mathcal{H}^c and \mathcal{H} with the sets of labels so that we can write, for example, $k \in \mathcal{H}^c$.

Let F be the event that at least $N_1 = N\varepsilon_1 p_1/4$ particles from the family \mathcal{H}^c stay inside A_2 until time S . Consider the motion of a particle X^k belonging to the \mathcal{H} family, conditional on F . Given F , the probability that the particle lands on a particle from the family \mathcal{H}^c at the time of a jump is not less than $(N\varepsilon_1 p_1/4)/(N-1)$. If this event occurs, then the k -th particle can stay within the set A_3 until time S with probability p_2 or higher. Hence, each jump of particle k has at least probability $p_2(N\varepsilon_1 p_1/4)/(N-1) \geq p_2 \varepsilon_1 p_1/4 \equiv p_4$ of being the last jump for this particle before time S . We see that the total number of jumps of X^k before time S is stochastically bounded by the geometric distribution with mean $1/p_4$.

In the rest of Part 1, we will assume that F occurred, i.e., all the probabilities will be conditional probabilities given F , even if the conditioning is not reflected in the notation.

We will now define a tree \mathcal{T}^m of particle trajectories representing *descendants* of particle m (see Fig. 1.). Informally, the family of all descendants of a particle X^m can be described as the smallest family of points (t, n) with the following properties. The particle X_t^m is its own descendant for all t , i.e., $(t, m) \in \mathcal{T}^m$ for all t . If a particle X^k jumps on a descendant of X^m at time s then X_t^k becomes a descendant of X^m for all $t \geq s$, i.e., $(t, k) \in \mathcal{T}^m$ for all $t \geq s$. We now present a more formal definition. We will say that $(t, n) \in \mathcal{T}^m$ if there exists a set of pairs (s_j, y_j) , $0 \leq j \leq j_0$, with $(s_0, y_0) = (0, x^m)$, $(s_{j_0}, y_{j_0}) = (t, X_t^n)$, such that there exist distinct $k(j) \in \mathcal{H}$, $j \geq 1$, with the following properties: (i) $(s_{j-1}, X_{s_{j-1}}^k) = (s_{j-1}, y_{j-1})$ and $(s_j, X_{s_j}^k) = (s_j, y_j)$ for all j , (ii) X^k does not jump at time s_j for $0 \leq j \leq j_0 - 1$, and (iii) $k(0) = m$ and $k(j_0) = n$. Note that the definition assumes that we use only pieces of trajectories of particles from the family \mathcal{H} .

Figure 1. Descendants of the m -th particle are represented by thick lines.

The domain D is the interval $(0, a)$.

Let \mathcal{K}_t^m be the set of all descendants of particle m until t , i.e., the set of all k such that $(s, k) \in \mathcal{T}^m$ for some $s \leq t$. The function $t \rightarrow \mathcal{K}_t^m$ is monotone.

Fix some $m \in \mathcal{H}$. Let α_1^k be the number of jumps made by particle k from the family \mathcal{H} during the time interval $[0, S]$ but before the first time T_m^k when it becomes a descendant of m (if there is no such time, we count all jumps before S). Let α_2^k be the number of jumps made after T_m^k but before S ($\alpha_2^k = 0$ if k does not become a descendant of m before S). It is easy to see that every random variable α_1^k and α_2^k is stochastically bounded by the geometric distribution with mean $1/p_4$ — we can use the same argument as earlier in the proof. We will need a substantially stronger bound, though. It is not very hard to see that one can define our processes on a probability space which will also carry random variables $\hat{\alpha}_1^k$ and $\hat{\alpha}_2^k$ for all $k \in \mathcal{H}$, such that $\alpha_1^k \leq \hat{\alpha}_1^k$ and $\alpha_2^k \leq \hat{\alpha}_2^k$ for all $k \in \mathcal{H}$, and every random variable $\hat{\alpha}_1^k$ and $\hat{\alpha}_2^k$ is geometric with mean $1/p_4$. Moreover, random variables $\hat{\alpha}_1^k$ and $\hat{\alpha}_2^k$ can be constructed so that they are jointly independent and independent of the process

$t \rightarrow \mathcal{K}_t^m$. The construction of such a family of random variables is standard so we will only sketch it. We start with constructing $\widehat{\alpha}_1^k$'s and $\widehat{\alpha}_2^k$'s and then we use them to construct α_1^k 's and α_2^k 's. We consider a probability space which carries independent sequences of Bernoulli coin tosses with success probability p_4 . We then identify $\widehat{\alpha}_1^k$'s and $\widehat{\alpha}_2^k$'s with the number of tosses until and including the first success in different sequences—we need one sequence of Bernoulli trials for each $\widehat{\alpha}_1^k$ and $\widehat{\alpha}_2^k$. The results of coin tosses corresponding to $\widehat{\alpha}_1^k$ are used to determine whether particle k jumps onto a particle from the family \mathcal{H}^c and then stays inside A_3 until S (this would be considered a “success”), for all jumps before the time T_m^k . The analogous events after time T_m^k are determined by the sequence of coin tosses corresponding to $\widehat{\alpha}_2^k$. All other aspects of the motion of particle k are determined by some other random mechanism. Such mechanisms need not be independent for different particles.

Let $|\mathcal{K}_t^m|$ denote the cardinality of \mathcal{K}_t^m . Note that if a particle from the family $\mathcal{H} \setminus \mathcal{K}_t^m$ jumps then with probability $1/(N-1)$ it lands on any other given particle. The value of $|\mathcal{K}_t^m|$ increases by 1 at the time t of a jump of a particle from $\mathcal{H} \setminus \mathcal{K}_t^m$ with probability equal to $|\mathcal{K}_{t-}^m|/(N-1)$. Hence,

$$P(|\mathcal{K}_t^m| = a + 1 \mid |\mathcal{K}_{t-}^m| = a, \exists k \in \mathcal{H} \setminus \mathcal{K}_{t-}^m : X_t^k \neq X_{t-}^k) = \frac{a}{N-1},$$

so for integer $a, r \geq 1$, and $c_1 = c_1(r) < \infty$,

$$\begin{aligned} E(|\mathcal{K}_t^m|^r \mid |\mathcal{K}_{t-}^m| = a, \exists k \in \mathcal{H} \setminus \mathcal{K}_{t-}^m : X_t^k \neq X_{t-}^k) &= a^r \left(1 - \frac{a}{N-1}\right) + (a+1)^r \frac{a}{N-1} \\ &= a^r \left(1 - \frac{a}{N-1} + \left(\frac{a+1}{a}\right)^r \frac{a}{N-1}\right) \\ &\leq a^r \left(1 - \frac{a}{N-1} + \left(1 + \frac{c_1 r}{a}\right) \frac{a}{N-1}\right) \\ &= a^r \left(1 + \frac{c_1 r}{N-1}\right). \end{aligned}$$

Hence, the expectation of $|\mathcal{K}_t^m|^r$ jumps by at most the factor of $1 + c_1 r/(N-1)$ at the time of a jump of a particle from $\mathcal{H} \setminus \mathcal{K}_t^m$. Let $\alpha = \sum_{k \in \mathcal{H}} \alpha_1^k$. By conditioning on the times of jumps, we obtain for $m \in \mathcal{H}$,

$$E|\mathcal{K}_t^m|^r \leq E \left(1 + \frac{c_1 r}{N-1}\right)^\alpha.$$

We estimate this quantity as follows, using the fact that the family of random variables α_1^k may be simultaneously stochastically bounded by a sequence of independent geometric

random variables $\widehat{\alpha}_1^k$ with mean $1/p_4$. The number of particles in \mathcal{H} is obviously bounded by N . In the following calculation we will pretend that the number of $\widehat{\alpha}_1^k$'s is N ; this is harmless because if the number of particles in \mathcal{H} is smaller than N , we can always add a few independent $\widehat{\alpha}_1^k$'s to the family. If

$$\log \left(1 + \frac{c_1 r}{N-1} \right) = c_2,$$

then c_2 is small for large N , and the following holds,

$$\begin{aligned} E \left(1 + \frac{c_1 r}{N-1} \right)^\alpha &= E \exp(c_2 \alpha) = E \exp \left(c_2 \sum_{k \in \mathcal{H}} \alpha_1^k \right) \\ &\leq E \exp \left(c_2 \sum_{k \in \mathcal{H}} \widehat{\alpha}_1^k \right) = E \prod_{k \in \mathcal{H}} \exp(c_2 \widehat{\alpha}_1^k) \\ &\leq E \prod_{k=1}^N \exp(c_2 \widehat{\alpha}_1^k) = (E \exp(c_2 \widehat{\alpha}_1^k))^N \\ &= \left(\sum_{j=0}^{\infty} e^{c_2(j+1)} (1-p_4)^j p_4 \right)^N = \left(\frac{e^{c_2} p_4}{1 - e^{c_2} (1-p_4)} \right)^N \\ &= \left(1 + \frac{1}{p_4} (e^{-c_2} - 1) \right)^{-N} = \left(1 + \frac{1}{p_4} \left(\frac{1}{1 + \frac{c_1 r}{N-1}} - 1 \right) \right)^{-N} \\ &= \left(1 - \frac{c_1 r}{p_4} \frac{1}{N-1 + c_1 r} \right)^{-N} \leq c_3 = c_3(r, p_4) < \infty. \end{aligned}$$

Thus, for some c_3 which depends only on r and p_4 ,

$$E |\mathcal{K}_t^m|^r \leq c_3. \quad (2.10)$$

Next we will estimate the total number of jumps β^m on the tree of descendants of particle m . For each descendant, this will include not only the first jump, at the time of which a particle becomes a descendant of m , but also all subsequent jumps by the descendant. Recall that given the whole genealogical tree $\{\mathcal{K}_t^m, 0 \leq t \leq S\}$, the numbers of jumps of descendants of m can be simultaneously bounded by $\widehat{\alpha}_2^k$'s, i.e., independent geometric random variables with mean $1/p_4$. We have, using (2.10), for some $c_4 = c_4(r) < \infty$,

$$E(\beta^m)^r \leq E \left(|\mathcal{K}_t^m| + \sum_{k=1}^{|\mathcal{K}_t^m|} \widehat{\alpha}_2^k \right)^r$$

$$\begin{aligned}
&\leq c_4 E |\mathcal{K}_t^m|^r + c_4 E \left(\sum_{k=1}^{|\mathcal{K}_t^m|} \hat{\alpha}_2^k \right)^r \\
&\leq c_4 E |\mathcal{K}_t^m|^r + c_4 E E \left(\left(\sum_{k=1}^a \hat{\alpha}_2^k \right)^r \mid |\mathcal{K}_t^m| = a \right) \\
&\leq c_4 E |\mathcal{K}_t^m|^r + c_4 E E \left(a^r \sum_{k=1}^a (\hat{\alpha}_2^k)^r \mid |\mathcal{K}_t^m| = a \right) \\
&\leq c_4 E |\mathcal{K}_t^m|^r + c_4 E E \left(a^{r+1} (\hat{\alpha}_2^k)^r \mid |\mathcal{K}_t^m| = a \right) \\
&\leq c_4 E |\mathcal{K}_t^m|^r + c_4 E E \left(a^{r+1} c_5 \mid |\mathcal{K}_t^m| = a \right) \\
&= c_4 E |\mathcal{K}_t^m|^r + c_4 c_5 E |\mathcal{K}_t^m|^{r+1} \leq c_6 = c_6(r) < \infty. \tag{2.11}
\end{aligned}$$

Part 2. This part of the proof is devoted to some qualitative estimates of the transition probabilities of the killed Brownian motion. Suppose $A \subset D$ is an open set.

We recall our assumption of regularity of ∂D . It implies that the function $(x, t) \rightarrow P_t^D(x, A)$ vanishes continuously as $x \rightarrow D^c$ and so it has a continuous extension to $\bar{D} \times [0, \infty)$.

The notation of the following remarks partly anticipates the notation in Part 3 of the proof. Fix some $t > 0$ and arbitrarily small $\delta_1 > 0$. The set $\bar{D} \times [0, t]$ is compact, so the continuous function $(x, t) \rightarrow P_t^D(x, A)$ is uniformly continuous on this set. It follows that we can find an integer $n < \infty$ and $\delta_2 > 0$ such that $|P_{t-s_j}^D(x, A) - P_{t-s}^D(y, A)| < \delta_1$ when $s \in [s_j, s_{j+1}]$, $|s_j - s_{j+1}| \leq t/n$, and $|x - y| \leq \delta_2$.

Fix arbitrarily small $t_1, \delta_1 > 0$. Let D_{δ_2} denote the set of points whose distance to D^c is greater than δ_2 . The transition density $p_t(x, y)$ of the free Brownian motion is bounded by $r_1 < \infty$ for $x, y \in \mathbf{R}^d$ and $t \geq t_1$. The same bound holds for the transition densities $p_t^D(x, y)$ for the killed Brownian motion because $p_t^D(x, y) \leq p_t(x, y)$. Choose $\delta_2 > 0$ so small that the volume of $D \setminus D_{\delta_2}$ is less than δ_1/r_1 . Then for every $s_j \geq t_1$ and $x \in D$,

$$P_{s_j}^D(x, D_{\delta_2}^c) = \int_{D \setminus D_{\delta_2}} p_{s_j}^D(x, y) dy \leq (\delta_1/r_1) \cdot r_1 = \delta_1.$$

Part 3. We start with the definition of *marks* which we will use to label particles. We will prove, in a sense, that the theorem holds separately for each family of particles bearing same marks. Typically, a particle X^j will bear different marks at different times.

The family of marks Θ is defined as the smallest set which contains 0, and which has the property that if $\theta_1, \theta_2 \in \Theta$ then the ordered pair (θ_1, θ_2) also belongs to Θ . Note that we do not assume that $\theta_1 \neq \theta_2$. We will write $(\theta_1 \mapsto \theta_2)$ rather than (θ_1, θ_2) . Here are some examples of marks:

$$0, \quad (0 \mapsto 0), \quad (0 \mapsto (0 \mapsto 0)), \quad ((0 \mapsto 0) \mapsto 0).$$

Our marks can be identified with vertices of a binary tree and are introduced only because our notation seems more intuitive in our context. Every mark will have an associated “height.” The height of 0 is defined to be 1. The height of $(\theta_1 \mapsto \theta_2)$ is one plus the maximum of heights of θ_1 and θ_2 .

We assign marks as follows. If a particle X^j has not jumped before time t then its mark is equal to 0 on the interval $(0, t)$. The mark of every particle is going to change every time it jumps, and only at such times. If a particle X^j jumps at time t onto a particle X^k , the mark of X^j has been θ_1 just before t , and the mark of X^k has been θ_2 just before t then the mark of X^j will be $(\theta_1 \mapsto \theta_2)$ on the interval between (and including) t and the first jump of X^j after t . To see that the above definition uniquely assigns marks to all particles at all times, note that we assign mark 0 to all particles until the first jump by any particle. Recall that $\tau_1 < \tau_2 < \tau_3 < \dots$ denote the jump times of all particles. If we know the marks on the interval $[\tau_j, \tau_{j+1})$ then it is easy to assign them in a unique way on the interval $[\tau_{j+1}, \tau_{j+2})$. An easy inductive procedure allows us to assign marks to all particles at all times.

The mark of X_t^k will be denoted $\theta(X_t^k)$. For any $\theta_1 \in \Theta$, let

$$\mathcal{X}_t^{N, \theta_1}(dy) = \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{\{\theta_1\}}(\theta(X_t^k)) \delta_{X_t^k}(dy).$$

Note that $\mathcal{X}_t^{N, \theta_1}(dy)$ is a (sub-probability) empirical measure supported by the particles marked with θ_1 at time t .

The law of large numbers and the continuity of $x \rightarrow P_t^D(x, dy)$ (see Part 2 of the proof) imply that for every fixed $t \leq S$, the measures $\mathcal{X}_t^{N, 0}(dy)$ converge in probability as $N \rightarrow \infty$ to the measure $\int_D P_t^D(x, dy) \mu_0(dx)$, in the sense of weak convergence of measures. In particular, $\mathcal{X}_S^{N, 0}(dy)$ converge weakly to a multiple of $\mu_S(dy)$. The main goal of this part of the proof is to show that $\mathcal{X}_S^{N, \theta}(dy)$ converge weakly to a multiple of $\mu_S(dy)$, for any fixed mark θ . This will be achieved by an inductive argument. We will elaborate the details of one inductive step, showing how the convergence of $\mathcal{X}_t^{N, 0}(dy)$ to $\int_D P_t^D(x, dy) \mu_0(dx)$ for

every $t \leq S$ implies the convergence of $\mathcal{X}_t^{N,(0 \rightarrow 0)}(dy)$ to $c_1 \int_D P_t^D(x, dy) \mu_0(dx)$ for every $t \leq S$.

Consider some $t \in (0, S]$. We will show that for any $\delta > 0$, $p < 1$, open $A \subset D$ and $0 < t_1 < t_2 < t$, we have

$$\begin{aligned} (1 - \delta)\mu_t(A) &\leq \inf_{t_1 \leq s \leq t_2} \frac{\int_D P_{t-s}^D(x, A) \mathcal{X}_s^{N,0}(dx)}{\int_D P_t^D(x, D) \mu_0(dx)} \\ &\leq \sup_{t_1 \leq s \leq t_2} \frac{\int_D P_{t-s}^D(x, A) \mathcal{X}_s^{N,0}(dx)}{\int_D P_t^D(x, D) \mu_0(dx)} \leq (1 + \delta)\mu_t(A), \end{aligned} \quad (2.12)$$

with probability greater than p , when N is sufficiently large. If we fix any integer $n \geq 1$, then for every $s_j = jt_2/n$, $j = 0, 1, \dots, n$, and every open set A_1 we have $\mathcal{X}_{s_j}^{N,0}(A_1) \rightarrow \int_D P_{s_j}^D(x, A_1) \mu_0(dx)$ in probability as $N \rightarrow \infty$, by the law of large numbers and the continuity of $x \rightarrow P_{s_j}^D(x, A_1)$. This and the continuity of $x \rightarrow P_{t-s_j}^D(x, A)$ imply that

$$\frac{\int_D P_{t-s_j}^D(x, A) \mathcal{X}_{s_j}^{N,0}(dx)}{\int_D P_t^D(x, D) \mu_0(dx)} \rightarrow \mu_t(A),$$

in probability. Since there is only a finite number of s_j 's, we immediately obtain a weak version of (2.12), namely,

$$\begin{aligned} (1 - \delta)\mu_t(A) &\leq \inf_{0 \leq j \leq n} \frac{\int_D P_{t-s_j}^D(x, A) \mathcal{X}_{s_j}^{N,0}(dx)}{\int_D P_t^D(x, D) \mu_0(dx)} \\ &\leq \sup_{0 \leq j \leq n} \frac{\int_D P_{t-s_j}^D(x, A) \mathcal{X}_{s_j}^{N,0}(dx)}{\int_D P_t^D(x, D) \mu_0(dx)} \leq (1 + \delta)\mu_t(A), \end{aligned} \quad (2.13)$$

with probability greater than p , when N is sufficiently large.

Fix arbitrarily small $\delta_1, p_1 > 0$ and let $\delta_2 > 0$ be so small and n so large that the following conditions are satisfied, according to Part 2 of the proof. First, $|P_{t-s}^D(x, A) - P_{t-s}^D(y, A)| < \delta_1$ when $s \in [s_j, s_{j+1}]$, $0 \leq j \leq n-1$, and $|x-y| \leq \delta_2$. Second, if D_{δ_2} denotes the set of points whose distance to D^c is greater than δ_2 , we want to have $P_{s_j}^D(x, D_{\delta_2}^c) < \delta_1$, for every x and j with $s_j \geq t_1$. Finally, increase n if necessary so that the probability that a Brownian path has an oscillation larger than δ_2 within a subinterval of $[0, S]$ of length t_2/n or less, is less than p_1 . With this choice of various constants, we see that for large N , with probability greater than p the following will be true for all j with $t_1 \leq s_j \leq S$. First, the proportion of X^k 's which will be within distance δ_2 of the boundary at time s_j will be less than $2\delta_1$ and the proportion of X^k 's which will jump during the interval $[s_j, s_{j+1}]$ will be less than $3\delta_1$. Because of this and the other parameter choices, we will

have $|P_{t-s_j}^D(X_{s_j}^k, A) - P_{t-s}^D(X_s^k, A)| < \delta_1$ for all $j, s \in [s_j, s_{j+1}]$ and all labels k in a subset of $\{1, 2, \dots, N\}$ whose cardinality would be bounded below by $(1 - 2p_1 - 3\delta_1)/N$. This implies that simultaneously for all j and $s \in [s_j, s_{j+1}]$, for large N ,

$$\left| \int_D P_{t-s_j}^D(x, A) \mathcal{X}_{s_j}^{N,0}(dx) - \int_D P_{t-s}^D(x, A) \mathcal{X}_s^{N,0}(dx) \right| \leq \delta_1 + 2p_1 + 3\delta_1,$$

with probability greater than p . This, the fact that p can be arbitrarily large and δ_1 and p_1 arbitrarily small, and (2.13) prove (2.12).

We will now prove that a suitable version of (2.12) holds when we replace $\mathcal{X}_s^{N,0}$ with $\mathcal{X}_s^{N,(0 \mapsto 0)}$. Consider an arbitrary $t \leq S$, and $0 < t_1 < t_2 < t$. Suppose that a particle X^k with mark 0 hits the boundary of D at a time $s \in (t_1, t_2)$. Then it will jump onto a randomly chosen particle. If X^k jumps onto a particle marked 0, its label will change to $(0 \mapsto 0)$. Given this event, conditional on the value of $\mathcal{X}_s^{N,0}$, the distribution of X^k at time t will be $\int_D P_{t-s}^D(x, \cdot) \mathcal{X}_s^{N,0}(dx)$, by the strong Markov property. The same holds true for all other particles with mark 0 which hit D^c between times t_1 and t_2 . Since these particles evolve independently after the jump, we see from (2.12) that the empirical distribution at time t of all particles marked $(0 \mapsto 0)$ which received this mark at a time between t_1 and t_2 converges in probability to a constant multiple of μ_t , as $N \rightarrow \infty$. If we fix $t > 0$, it is easy to see that for sufficiently small $t_1 > 0$ and large $t_2 < t$, the probability that a particle with mark 0 will hit the boundary of D in one of the intervals $[0, t_1]$ or $[t_2, t]$ will be arbitrarily small. Hence, $\mathcal{X}_t^{N,(0 \mapsto 0)}$ converges to a constant multiple of μ_t , in probability.

Given the last fact, the same argument which proves (2.13) yields for some $\eta(\theta) \in (0, 1]$,

$$\begin{aligned} \eta((0 \mapsto 0))(1 - \delta)\mu_t(A) &\leq \inf_{0 \leq j \leq n} \int_D P_{t-s_j}^D(x, A) \mathcal{X}_{s_j}^{N,(0 \mapsto 0)}(dx) \\ &\leq \sup_{0 \leq j \leq n} \int_D P_{t-s_j}^D(x, A) \mathcal{X}_{s_j}^{N,(0 \mapsto 0)}(dx) \leq \eta((0 \mapsto 0))(1 + \delta)\mu_t(A), \end{aligned} \quad (2.14)$$

with large probability, when N is large. The argument following (2.13) is not specific to the case when the particles have the mark 0 and so it can be applied to the present case of particles marked $(0 \mapsto 0)$. Hence, we obtain the following formula, which differs from (2.12) only in that the normalizing constant is $\eta((0 \mapsto 0))$ and not $\int_D P_t^D(x, D)\mu_0(dx)$,

$$\begin{aligned} \eta((0 \mapsto 0))(1 - \delta)\mu_t(A) &\leq \inf_{t_1 \leq s \leq t_2} \int_D P_{t-s}^D(x, A) \mathcal{X}_s^{N,(0 \mapsto 0)}(dx) \\ &\leq \sup_{t_1 \leq s \leq t_2} \int_D P_{t-s}^D(x, A) \mathcal{X}_s^{N,(0 \mapsto 0)}(dx) \leq \eta((0 \mapsto 0))(1 + \delta)\mu_t(A). \end{aligned} \quad (2.15)$$

The last formula holds with probability greater than p if N is sufficiently large, for any fixed $t \leq S$, any $0 < t_1 < t_2 < t$, and any $p < 1$.

Proceeding by induction, one can show that (2.15) applies not only to the mark $\theta = (0 \mapsto 0)$ but also to $(0 \mapsto (0 \mapsto 0))$, $((0 \mapsto 0) \mapsto 0)$, $((0 \mapsto 0) \mapsto (0 \mapsto 0))$, and to every other mark θ . Imbedded in an induction step for a mark θ is the proof that the measures $\mathcal{X}_S^{N,\theta}(dy)$ converge to a constant (deterministic) multiple of $\mu_S(dy)$. It follows that for every finite deterministic subset Θ_1 of Θ , the measures $\sum_{\theta \in \Theta_1} \mathcal{X}_S^{N,\theta}(dy)$ converge to a multiple of $\mu_S(dy)$.

It will now suffice to show that for any $p_2, p_3 > 0$, there exists a finite set Θ_1 such that $\sum_{\theta \notin \Theta_1} \mathcal{X}_S^{N,\theta}(D) < p_2$ with probability greater than $1 - p_3$. In other words, we want to show that for some finite Θ_1 , the number of particles with a mark from Θ_1^c at time S is less than $p_2 N$ with probability greater than $1 - p_3$. In order to prove this, we will use the result proved in Part 1 of the proof.

Recall the notion of the “height” of a mark from the beginning of the second part of the proof. Suppose that a particle X^k has a mark with height j at time S . Let t_j be the infimum of t with the property that X_t^k has a mark with height j . Then t_j must be the time of a jump of X^k . Let X^{n_j} be the particle on which X^k jumps at time t_j . By definition, the height of the mark of X^k or the height of the mark of X^{n_j} must be equal to $j - 1$ just before time t_j . We define k_j to be k or n_j , so that the height of the mark of X^{k_j} is equal to $j - 1$ prior to t_j . We proceed by induction. Suppose we have identified a particle X^{k_m} which has a mark with height $m - 1$ prior to a time t_m , where $m \leq j$. Then we let t_{m-1} be the infimum of $t < t_m$ with the property that the height of the mark of $X_t^{k_m}$ is $m - 1$. We see that X^{k_m} must jump at time t_{m-1} on a particle $X^{n_{m-1}}$. We choose k_{m-1} to be either k_m or n_{m-1} , so that the height of the mark of $X^{k_{m-1}}$ is $m - 2$ just before the time t_{m-1} . Proceeding in this way, we will end up with a particle X^{k_2} which has a mark with height 1. The mark of this particle is 0, as it is the only mark with height 1. This implies that $t_1 = 0$. We claim that for all $m \leq j$ and $t \in [t_{m-1}, t_m)$, we have $(t, k_m) \in \mathcal{T}^{k_2}$. In other words, every particle X^{k_m} is a descendant of X^{k_2} at times $t \geq t_{m-1}$. To see this, note that the claim is obviously true for $m = 2$. If $k_3 = k_2$ then the claim is true for $m = 3$, because X^{k_2} always remains its own descendant. If $k_3 \neq k_2$, it is clear that the particle X^{k_3} has jumped at time t_2 on X^{k_2} , a descendant of k_2 , and so became a descendant of k_2 at this time. Proceeding by induction, we can show that all particles in our chain are descendants of k_2 on the intervals specified above. Note that at every time t_m , either a descendant of k_2 jumps or a descendant of k_2 is born. Hence, if

X^k has a mark with height j at time S , it must belong to the family of descendants of a particle for which the sum of descendants and their jumps is not less than j .

Recall the event F from Part 1 of the proof and choose the parameters in Part 1 so that the probability of F^c is less than $p_3/2$ and the cardinality of \mathcal{H}^c is less than $Np_2/2$. Conditional on F , we have the following estimate. If the sum of the number of descendants of a particle k and the number of all their jumps is equal to j then a crude estimate says that at most j descendants of k end up at time S with marks of height j or lower; by the argument in the previous paragraph, the marks cannot be higher than j . Hence, the expected number of particles with marks higher than n at time S among descendants of a particle $k \in \mathcal{H}$ can be bounded using (2.11) by

$$\sum_{j \geq n} j P(\beta^k \geq j) \leq \sum_{j \geq n} j \frac{E(\beta^k)^3}{j^3} \leq c_1 \sum_{j \geq n} j^{-2} \leq c_2/n.$$

The expected number of all particles in the family \mathcal{H} with marks higher than n is bounded by Nc_2/n .

Choose Θ_1 to be the set of all marks of height less than n , where n is so large that $Nc_2/n < (p_3/2)(p_2N/2)$. Then, conditional on F , the probability that the total number of particles in \mathcal{H} with marks higher than n is bounded by $Np_2/2$ with probability $1 - p_3/2$ or higher. We add to that estimate all particles from \mathcal{H}^c —their number is bounded by $Np_2/2$, so the total number of particles with marks higher than n is bounded by Np_2 , with probability greater than or equal to $1 - p_3/2$. This estimate was obtained under the assumption that F holds. Since the probability of F is less than $p_3/2$, we are done. \square

Example 2.1 We will show that for some process X_t , some t, A, μ_0 and N we have $E\mathcal{X}_t^N(A) \neq \mu_t(A)$. Our process has a finite state space; we presume that a similar example can be constructed for Brownian motion.

We will consider a continuous time Markov process on the state space $\{0, 1, 2\}$. The set $\{1, 2\}$ will play the role of D . The possible jumps of the process are $0 \rightarrow 1$, $1 \rightarrow 2$, $2 \rightarrow 1$ and $1 \rightarrow 0$. The jump rates are equal to 1 for each one of these possible transitions. The measure μ_0 will be uniform on D , i.e., $\mu_0(1) = \mu_0(2) = 1/2$.

First we will argue that $\mu_t = \mu_0$ for all $t > 0$. To see this, note that if we condition the process not to jump to 0, it will jump from the state 1 to the state 2 and from 2 to 1 at the rates 1, i.e., at the original rates. This is because all four types of jumps $0 \rightarrow 1$, $1 \rightarrow 2$, $2 \rightarrow 1$ and $1 \rightarrow 0$ may be thought of as coming from four independent Poisson

processes. Conditioning on the lack of jumps of one of these processes does not influence the other three jump processes.

Since the conditioned process makes jumps from 1 to 2 and vice versa with equal rates, the symmetry of μ_0 on the set $\{1, 2\}$ is preserved forever, i.e., $\mu_t = \mu_0$ for all $t > 0$.

Now let $N = 2$ and consider the distribution of the process \mathbf{X}_t . Let A_t denote the number of particles X^1 and X^2 at the state 1 at time t . The process A_t is a continuous time Markov process with possible values 0, 1 and 2. Its possible transitions are $0 \rightarrow 1$, $1 \rightarrow 2$, $2 \rightarrow 1$ and $1 \rightarrow 0$, just like for the original process X_t . If $A_t = 0$, i.e., if both particles are at the state 2, the waiting time for a jump of A_t has expectation $1/2$ because each of the particles jumps independently of the other one with the jump rate 1. When both particles are in the state 1, and one of them jumps to 0, it immediately returns to 1 (the location of the other particle) so the jumps of X^k 's from 1 to 0 have no effect on A_t , if $A_t = 2$. It follows that the rate for the jumps of A_t from 2 to 1 is 2, i.e., it is the same as for the jumps of A_t from 0 to 1. Finally, let us analyze the case $A_t = 1$. When only one of the particles is at 2, it jumps to 1 at the rate 1, so the rate of the transitions $1 \rightarrow 2$ for the A_t process is 1. However, its rate of transitions $1 \rightarrow 0$ is equal to 2 because any jump of a particle from the state 1 will result in its landing at 2, either directly or through the instantaneous visit to 0. Given these transition rates, it is elementary to check that the stationary distribution of A_t assigns probabilities $1/3$, $1/2$ and $1/6$ to the states 0, 1 and 2. This implies that $E\mathcal{X}_t^2(\{1\}) \approx 5/12$ for large t (no matter what \mathcal{X}_0^2 is) and so $E\mathcal{X}_t^2(\{1\}) \neq \mu_t(\{1\})$ for some t when we choose $\mu_0(\{1\})$ to be equal to $1/2$. \square

Proof of Theorem 1.4. For a point $x \in D$ let ρ_x be the supremum of $\text{dist}(x, \partial B(y, r))$ over all open balls $B(y, r)$ such that $x \in B(y, r) \subset D$. For each $x \in D$ we will choose a ball B_x with radius r , such that $x \in B_x \subset D$ and $\text{dist}(x, \partial B_x) > \rho_x/2$. The center of B_x will be denoted v_x . We would like the mapping $x \rightarrow v_x$ to be measurable. One way to achieve this goal is to construct a countable family of balls with radius r and make the mapping $x \rightarrow v_x$ constant on every element of a countable family of squares, closed on two sides, disjoint, and summing up to D . Such a construction is known as ‘‘Whitney squares,’’ it is quite elementary and so it is left to the reader.

We will construct X_t^k 's in a special way. Two 1-dimensional processes U_t^k and R_t^k will be associated with each X_t^k . The processes U_t^k and R_t^k will take their values in $[0, r]$. The processes R_t^k will be independent d -dimensional Bessel processes reflected at r . In other words, every process R_t^k will have the same distribution as the radial part

of the d -dimensional Brownian motion reflected inside the ball $B(0, r)$. We will define U_t^k so that $U_t^k \leq R_t^k$ for all k and t . The processes R_t^k will give us a bound on the distance of X_t^k from D^c ; more precisely, we will have, according to our construction, $\text{dist}(X_t^k, D^c) \geq r - U_t^k \geq r - R_t^k$.

No matter what distribution for \mathbf{X}_0 is desirable, it is easy to see that we can choose the starting values for R_t^k 's so that $R_0^k = \text{dist}(X_0^k, v_{X_0^k})$, a.s.

In our construction, we assume that R_t^k 's are given and we proceed to describe how to define U_t^k 's and X_t^k 's given R_t^k 's. We will first fix a k . Let T_1^k be the first time when the process R_t^k hits r . On the interval $[0, T_1^k)$, we can define X_t^k as Brownian motion in \mathbf{R}^d such that $R_t^k = \text{dist}(X_t^k, v_{X_0^k})$. This requires only generating an angular part for X_t^k , relative to the initial positions X_0^k . A classical ‘‘skew-product’’ decomposition (see Itô and McKean (1974)) achieves the goal by generating a Brownian motion on a sphere (independent of R_t^k) and then time-changing it according to a clock defined by R_t^k . Note that according to the definition of v_x , this constructed process X_t^k will remain inside D for $t \in [0, T_1^k)$. We let $U_t^k = R_t^k$ for $t \in [0, T_1^k)$. At time T_1^k , the process U_t^k jumps to the value $\text{dist}(X_{T_1^k-}^k, v(X_{T_1^k-}^k))$, i.e., we let $U_{T_1^k}^k = \text{dist}(X_{T_1^k-}^k, v(X_{T_1^k-}^k))$. We let the process U_t^k evolve after time T_1^k as a d -dimensional Bessel process independent of R_t^k , until time $T_2^k = \inf\{t \geq T_1^k : U_t^k = R_t^k\}$. Let $T_3^k = \inf\{t \geq T_2^k : R_t^k = r\}$. We couple the processes U_t^k and R_t^k on the interval $[T_2^k, T_3^k)$, i.e., we let $U_t^k = R_t^k$ for $t \in [T_2^k, T_3^k)$. For $t \in [T_1^k, T_3^k)$, we construct X_t^k so that $\text{dist}(X_t^k, v(X_{T_1^k}^k)) = U_t^k$. The spherical part is constructed in an ‘‘independent’’ way, in the sense of the skew-product decomposition.

We proceed by induction. Recall that R_t^k is given, and suppose that processes X_t^k and U_t^k are defined on the interval $[0, T_{2j-1}^k)$. Moreover, suppose that R_t^k approaches r as $t \uparrow T_{2j-1}^k$. We define $U_{T_{2j-1}^k}^k$ to be $\text{dist}(X_{T_{2j-1}^k-}^k, v(X_{T_{2j-1}^k-}^k))$ and we let the process U_t^k evolve after time T_{2j-1}^k as a d -dimensional Bessel process independent of R_t^k , until time $T_{2j}^k = \inf\{t \geq T_{2j-1}^k : U_t^k = R_t^k\}$. Note that $U_t^k < R_t^k \leq r$ for $t \in [T_{2j-1}^k, T_{2j}^k)$. Let $T_{2j+1}^k = \inf\{t \geq T_{2j}^k : R_t^k = r\}$. We couple the processes U_t^k and R_t^k (i.e., we make them equal) on the interval $[T_{2j}^k, T_{2j+1}^k)$. The Brownian motion X_t^k is defined on $[T_{2j-1}^k, T_{2j+1}^k)$ so that $\text{dist}(X_t^k, v(X_{T_{2j-1}^k}^k)) = U_t^k$. Its spherical part is generated in an independent way from other elements of the construction and then time-changed according to the skew-product recipe. We see that $U_t^k < R_t^k \leq r$ for $t \in [T_{2j-1}^k, T_{2j+1}^k)$ and $\lim_{t \rightarrow T_{2j+1}^k} R_t^k = r$. This implies that X_t^k stays inside D on every interval $[T_{2j-1}^k, T_{2j+1}^k)$.

Let $\tau_1^k = \lim_{j \rightarrow \infty} T_j^k$ and note that typically, $\tau_1^k < \infty$. The above procedure allows us to define the processes X_t^k and U_t^k on the interval $[0, \tau_1^k)$. We repeat the construction for

all particles X_t^k in such a way that the processes in the family $\{(X_t^k, U_t^k)\}_{1 \leq k \leq N}$ are jointly independent.

Let $\tau_1 = \min_{1 \leq j \leq N} \tau_1^j$ and suppose that the minimum is attained at k , i.e., $\tau_1^k = \tau_1 < \infty$. Since infinitely many independent Bessel processes $\{U_t^k, t \in [T_{2j-1}^k, T_{2j+1}^k]\}$ traveled from $\text{dist}(X_{T_{2j-1}^k}^k, v(X_{T_{2j-1}^k}^k))$ to r , and their travel times sum up to a finite number, bounded by τ_1 , it follows that $\lim_{j \rightarrow \infty} \text{dist}(X_{T_{2j-1}^k}^k, v(X_{T_{2j-1}^k}^k)) = r$. We will show that X_t^k must approach ∂D at time $\tau_1^k = \tau_1$.

Recall the function ρ_x from the beginning of the proof. If x belongs to an open ball $B(y, r) \subset D$ then the same holds for all points in a small neighborhood of x . The definition of ρ_x now easily implies that the function $x \rightarrow \rho_x$ is Lipschitz inside D . Since, by assumption, ρ_x does not vanish inside D , every sequence x_n satisfying $\text{dist}(x_n, v_{x_n}) \rightarrow r$ also satisfies $\rho_{x_n} \rightarrow 0$ and so must approach ∂D as $n \rightarrow \infty$. This finishes the proof that $\lim_{t \rightarrow \tau_1^k-} \text{dist}(X_t^k, D^c) = 0$. Since two independent Brownian particles cannot hit ∂D at the same time, we see that there is only one process X_t^k with $\tau_1^k = \tau_1 < \infty$.

Still assuming that $\tau_1^k = \tau_1 < \infty$, we uniformly and independently of everything else choose $j \neq k$ and let $X_{\tau_1}^k = X_{\tau_1}^j$ and $U_{\tau_1}^k = U_{\tau_1}^j$. We then proceed with the construction of X_t^k and U_t^k on the interval $[\tau_1^k, \tau_2^k)$, such that $\lim_{t \rightarrow \tau_2^k-} \text{dist}(X_t^k, D^c) = 0$. The construction is completely analogous to that outlined above. Note that we necessarily have $U_{\tau_1}^k < R_{\tau_1}^k$ so we have to start our construction as in the inductive step of the original algorithm. Recall that the construction generates a process U_t^k satisfying $U_t^k \leq R_t^k$ for $t \in [\tau_1^k, \tau_2^k)$.

We let $\tau_2 = \tau_2^k \wedge \min_{1 \leq j \leq N, j \neq k} \tau_1^j$. A particle X^j will have to approach ∂D at time τ_2 . We will make this particle jump and then proceed by induction. Theorem 1.1 shows that there will be no accumulation of jumps of X_t^k 's at any finite time.

Recall that the inner ball radius $r > 0$ is a constant depending only on the domain D . It is well known that the reflected process R_t^k spends zero time on the boundary (i.e., at the point r) so if it starts from $R_0^k = r$ then its distribution at time $t = 1$ is supported on $(0, r)$. It follows that for any $p_1 < 1$ there exists $r_1 \in (0, r)$ such that we have, with $r_2 = r$,

$$P(R_1^k \in [0, r_1] \mid R_0^k = r_2) > p_1.$$

This estimate can be extended to all $r_2 \in [0, r]$, by an easy coupling argument. It follows from this and the independence of processes R_t^k that there exists $p_2 > 0$ such that with probability greater than p_2 , more than $Np_1/2$ processes R_t^k happen to be in $[0, r_1]$ at time 1, no matter what their starting positions are at time 0, for every $N > 0$.

Let D_a be the set of all points in D whose distance from D^c is greater than or equal to a . The processes X_t^k have been constructed in such a way that a.s., for every k and

t we have $\text{dist}(X_t^k, D^c) \geq r - R_t^k$. This and the claim in the previous paragraph show that for any starting position of X_t^k 's, with probability greater than p_2 , more than $Np_1/2$ processes X_t^k happen to be in D_{r_3} at time 1, where $r_3 = r - r_1$. We will now proceed as at the beginning of Part 1 in the proof of Theorem 1.3. Fix an arbitrary $p_1 < 1$, a corresponding $r_1 = r_1(p_1) \in (0, r)$, $r_3 = r - r_1$, and arbitrary $0 < r_5 < r_4 < r_3$. Let \mathcal{H} be the family of all processes X^k such that $X_1^k \in D_{r_3}$. Assume that \mathcal{H} has at least $Np_1/2$ elements. There is $p_3 > 0$ (depending on N, p_1 and r_j 's) such that with probability greater than p_3 , all processes in \mathcal{H} will stay in D_{r_4} for all $t \in [1, 2]$. For some $p_4 > 0$, all processes in \mathcal{H}^c will have a jump in the interval $[1, 2]$, will land on a particle from the \mathcal{H} family, and subsequently stay in D_{r_5} until time $t = 2$. Altogether, there is a strictly positive probability $p_2 p_3 p_4 \equiv p_5$ that all particles will be in D_{r_5} at time $t = 2$, given any initial distribution at time $t = 0$.

Now let us rephrase the last statement in terms of the vector process \mathbf{X}_t whose state space is D^N . We have just shown that with probability higher than p_5 the process \mathbf{X}_t can reach a compact set $D_{r_5}^N$ within 2 units of time. This and the strong Markov property applied at times $2, 4, 6, \dots$ show that the hitting time of $D_{r_5}^N$ is stochastically bounded by an exponential random variable with the expectation independent of the starting point of \mathbf{X}_t . Since the transition densities $p_t^{\mathbf{X}}(\mathbf{x}, \mathbf{y})$ for \mathbf{X}_t are bounded below by the densities for the Brownian motion killed at the exit time from D^N , we see that $p_t^{\mathbf{X}}(\mathbf{x}, \mathbf{y}) > c_1 > 0$ for $\mathbf{x}, \mathbf{y} \in D_{r_5}^N$. Fix arbitrarily small $s > 0$ and consider the ‘‘skeleton’’ $\{\mathbf{X}_{ns}\}_{n \geq 0}$. It is standard to prove that the properties listed in this paragraph imply that the skeleton has a stationary probability distribution and that it converges to that distribution exponentially fast. This can be done, for example, using Theorem 2.1 in Down, Meyn and Tweedie (1995). Extending the convergence claim to the continuous process $t \rightarrow \mathbf{X}_t$ from its skeleton can be done in a very general context, as was kindly shown to us by Richard Tweedie. In our case, a simple argument based on ‘‘continuity’’ can be supplied. More precisely, one can use a lower estimate for $p_t^{\mathbf{X}}(\mathbf{x}, \mathbf{y})$ in terms of the transition densities for Brownian motion killed upon leaving D^N , which are continuous. We leave the details to the reader. This completes the proof of part (i) of the theorem.

Recall that we have proved that for any $p_1 < 1$ there exists $r_1 < r$ such that for any starting position of X_0^k , the particle X^k is in D_{r-r_1} at time $t = 1$, with probability greater than p_1 . It follows that for any N , the mean measure $E\mathcal{X}_{\mathbf{M}}^N$ of the compact set D_{r-r_1} is not less than p_1 . Hence, the mean measures $E\mathcal{X}_{\mathbf{M}}^N$ are tight in D . Lemma 3.2.7 of Dawson (1992, p. 32) implies that the sequence of random measures $\mathcal{X}_{\mathbf{M}}^N$ is tight and so it contains

a convergent subsequence.

Choose a subsequence N_j such that the sequence $\mathcal{X}_{\mathbf{M}}^{N_j}$ is convergent to a probability measure $\Pi(\mu)$, carried by the family of probability measures on D . It will be enough to prove part (ii) of the theorem for this sequence. Consider the sequence of processes $\mathbf{X}_t = \mathbf{X}_t^{N_j}$, each with the stationary distribution $\mathcal{X}_{\mathbf{M}}^{N_j}$ as its starting distribution. Fix an open set $A \subset D$. By an argument totally analogous to the proof of Theorem 1.3, the following holds in the sense of convergence in probability,

$$\lim_{j \rightarrow \infty} \mathcal{X}_t^{N_j}(A) = \int \frac{\int_D P_t^D(x, A) \mu(dx)}{\int_D P_t^D(x, D) \mu(dx)} \Pi(d\mu). \quad (2.16)$$

We will now apply a few results from Bass and Burdzy (1992). Check Section 3 of that paper for the definition of a *John domain*. It is elementary to see that our domain D is a John domain, because it satisfies the interior ball condition. By Proposition 3.2 in Bass and Burdzy (1992), every John domain is a twisted Hölder domain of order 1. Hence, the parabolic boundary Harnack principle (Theorem 1.2 of Bass and Burdzy (1992)) holds for D . That theorem says that if $p_t^D(x, y)$ denotes the transition densities for Brownian motion killed upon exiting D , then for each $u > 0$ there exists $c = c(D, u) \in (0, 1)$ such that

$$\frac{p_t^D(x, y)}{p_t^D(x, z)} \geq c \frac{p_s^D(v, y)}{p_s^D(v, z)} \quad (2.17)$$

for all $s, t \geq u$ and all $v, x, y, z \in D$. We will need a stronger version of this inequality. The proof will be based on a lemma of Burdzy, Toby and Williams (1989). The following version of that lemma is taken from Burdzy and Khoshnevisan (1998).

Suppose that functions $h(x, y), g(x, y)$ and $h_1(x, y)$ are defined on product spaces $W_1 \times W_2, W_2 \times W_3$ and $W_1 \times W_3$, resp. Assume that for some constant $c_1, c_2 \in (0, 1)$ the functions satisfy for all $x, y, x_1, x_2, y_1, y_2, z_1, z_2$,

$$h_1(x, y) = \int_{W_2} h(x, z) g(z, y) dz,$$

$$\frac{h(x_1, z_1)}{h(x_1, z_2)} \geq \frac{h(x_2, z_1)}{h(x_2, z_2)} (1 - c_1),$$

and

$$\frac{g(z_1, y_1)}{g(z_1, y_2)} \geq c_2 \frac{g(z_2, y_1)}{g(z_2, y_2)}.$$

Then

$$\frac{h_1(x_1, y_1)}{h_1(x_1, y_2)} \geq \frac{h_1(x_2, y_1)}{h_1(x_2, y_2)} (1 - c_1 + c_2^2 c_1). \quad (2.18)$$

We will apply the lemma with $p_2^D(x, y)$ in place of $h_1(x, y)$, and $p_1^D(x, y)$ in place of $h(x, z)$ and $g(z, y)$. We see from (2.18) that the constant $c(D, 2)$ in (2.17) may be taken to be $c(D, 1) + c(D, 1)^2(1 - c(D, 1))$. By induction, we see that the constants $c(D, 2^n)$ may be chosen in such a way that $c(D, 2^n) = c(D, 2^{n-1}) + c(D, 2^{n-1})^2(1 - c(D, 2^{n-1}))$. Then $c(D, 2^n) \rightarrow 1$ as $n \rightarrow \infty$. Obviously, we may assume that the function $u \rightarrow c(D, u)$ is non-decreasing. Hence, (2.17) holds for some $c(D, u)$ satisfying $c(D, u) \rightarrow 1$ as $u \rightarrow \infty$.

The inequality (2.17) easily implies that

$$c(D, t) \frac{P_t^D(y, A)}{P_t^D(y, D)} \leq \frac{P_t^D(x, A)}{P_t^D(x, D)} \leq c(D, t)^{-1} \frac{P_t^D(y, A)}{P_t^D(y, D)},$$

for all $x, y \in D$. This in turn shows that

$$c(D, t) \frac{\int_D P_t^D(x, A) \mu_2(dx)}{\int_D P_t^D(x, D) \mu_2(dx)} \leq \frac{\int_D P_t^D(x, A) \mu_1(dx)}{\int_D P_t^D(x, D) \mu_1(dx)} \leq c(D, t)^{-1} \frac{\int_D P_t^D(x, A) \mu_2(dx)}{\int_D P_t^D(x, D) \mu_2(dx)},$$

for any probability measures μ_1 and μ_2 on D . Since $c(D, t) \rightarrow 1$ as $t \rightarrow \infty$, we see that for some fixed probability measure μ_1 on D and any $\Pi(\mu)$,

$$\lim_{t \rightarrow \infty} \left(\frac{\int_D P_t^D(x, A) \mu_1(dx)}{\int_D P_t^D(x, D) \mu_1(dx)} \right)^{-1} \int \frac{\int_D P_t^D(x, A) \mu(dx)}{\int_D P_t^D(x, D) \mu(dx)} \Pi(d\mu) = 1. \quad (2.19)$$

The normalized distribution of the killed Brownian motion in D converges to the normalized first eigenfunction φ_1 of the Dirichlet Laplacian in D , i.e.,

$$\lim_{t \rightarrow \infty} \frac{\int_D P_t^D(x, A) \mu_1(dx)}{\int_D P_t^D(x, D) \mu_1(dx)} = \frac{\int_A \varphi_1(y) dy}{\int_D \varphi_1(y) dy},$$

by the eigenfunction expansion for $p_t^D(x, y)$. In view of (2.19),

$$\int \frac{\int_D P_t^D(x, A) \mu(dx)}{\int_D P_t^D(x, D) \mu(dx)} \Pi(d\mu) \rightarrow \frac{\int_A \varphi_1(y) dy}{\int_D \varphi_1(y) dy}, \quad (2.20)$$

as $t \rightarrow \infty$. By the stationarity of $\mathcal{X}_{\mathbf{M}}^{N_j}$, the right hand side of (2.16) does not depend on t and so (2.20) is in fact an equality. This observation combined with (2.16) completes the proof. \square

3. Appendix. Related probabilistic and physical models. We will discuss a few well known models and problems in probability and mathematical physics to which our paper

is related. Before we do so, let us note that the original impulse for the article came from heuristic and numerical results presented in Burdzy, Hołyst, Ingberman and March (1996). This largely determined the direction of our research. The notes below may include some ideas for future research on our model, perhaps different in their flavor from the present article.

(i) *Superprocesses with interactions.* Superprocesses, also known as measure-valued diffusions or Dawson-Watanabe diffusions, are processes whose states are measures. Super-Brownian motion and the Fleming-Viot process with Brownian spatial motion are two of the most studied models in this class. The model introduced in this paper resembles most the Fleming-Viot process, which can be described in a heuristic way as follows. Consider N particles performing independent Brownian motions in \mathbf{R}^d . Every ε units of time, two particles are chosen uniformly and the first particle jumps to the location of the second one. Between the jumps, the particles are independent Brownian motions. Assume for simplicity that all particles start from a fixed point. If $N \rightarrow \infty$ and $\varepsilon \rightarrow 0$, at a rate related to N , then the empirical distributions of the particles converge for every time $t \geq 0$ to a random measure. It is known that in dimensions $d \geq 2$, the measures are carried by sets of fractal nature.

The original Fleming-Viot model sketched above assumes independence of the branching mechanism from the spatial distribution of the particles. In recent years, a number of papers have been devoted to processes which are similar but whose branching mechanism does depend on the spatial distribution of the particles. Roughly speaking, two closely related models have been considered—in one of them a “catalyst” is present, facilitating the branching of particles; the other model assumes that branching can be influenced by the local density of particles (see, e.g., Adler and Ivanitskaya (1996) Dawson and Fleischmann (1997), Dawson and Greven (1996), Dawson and Perkins (1998) and Klenke (1999)).

Our model goes in a slightly different direction because we consider an “obstacle” (the set D^c) where the particles are killed although the offspring are generated in a uniform way across the whole population as in the original model. Our process might possibly represent a biological population, with a region D^c having fatal effect on individuals. The assumption of the constant number of individuals is an idealization of the constant carrying capacity of an environment. Fleming-Viot models are sometimes applied to “populations” whose individual members are genes.

The main qualitative difference between our model and the classical Fleming-Viot process with Brownian spatial motion is that in the limit, we obtain measures with smooth

densities.

(ii) *Propagation of chaos.* When we consider a large number of interacting particles then under some assumptions, two tagged particles will behave in an almost independent way (see, e.g., Sznitman (1991)). In our case, two particles X_t^k and X_t^j are almost independent when N is large. Even stronger result is true—the propagation of chaos holds for the entire trees of descendants of particles labeled k and j . The two claims are quite clear in view of the theorems and techniques presented in the paper but we will not give a rigorous proof here.

(iii) *Genetic algorithms.* A very active area of applied and theoretical research deals with “genetic algorithms.” We mention a book of Man, Tang and Kwong (1999) as a possible starting entry point to this rapidly growing field. A genetic algorithm is a way to search for an answer to a problem by imitating biological genetic processes. Our model might be thought of as a genetic algorithm generating the first eigenvalue and the corresponding eigenfunction for the Dirichlet Laplacian. We do not make any claims of direct applicability of our model, especially in view of the fact that we do not present any theoretical estimates of the rate of convergence or computer simulations. We note however, that a related problem of finding the second Neumann eigenvalue (the “spectral gap”) is one of the most studied problems from both theoretical and practical points of view, for various Markov processes.

(iv) *Minimization of entropy production.* It is postulated in physics (Wio (1994) III.5) that an irreversible system achieves a stationary state characterized by the minimum entropy production. See Prigogine-de Groot Theorem in Yourgrau, van der Merwe and Raw (1982); consult also a recent article of Ruelle (1997) on this topic. Entropy production has been studied in the context of stochastic processes, for example by Gong and Qian (1997) but we could not find a direct relationship between that paper and our model.

We will explain how our model relates the principle of minimum entropy production to a minimizing property of the first Laplacian eigenfunction. In order to simplify the presentation, we will consider a slightly modified model in which the branching rate is constantly equal to λ_1 , the first eigenvalue of the Laplacian in D with Dirichlet boundary conditions. In general, the branching rate does not have to be a constant. In the limit, when $N \rightarrow \infty$, we obtain the following formula for the evolution of the density of the

particle process,

$$\frac{\partial p(x, t)}{\partial t} = \Delta p(x, t) + \lambda_1 p(x, t), \quad (3.1)$$

where Δ represents the Laplacian (we ignored the probabilistic constant $1/2$). The first term on the right hand side represents the Brownian motion effect and the second one represents branching.

We will use the notion of entropy proposed by Rényi (1961). He introduced a family of entropy measures parametrized by β ,

$$\mathcal{S}_t(\beta) = (1 - \beta)^{-1} \log \int_D p(x, t)^\beta dx.$$

We will consider one of these definitions corresponding to $\beta = 2$,

$$\mathcal{S}_t = - \log \int_D p(x, t)^2 dx.$$

Using this definition of entropy, we obtain from (3.1),

$$\frac{d\mathcal{S}}{dt} = -\lambda_1 + \frac{\int_D |\nabla p(x, t)|^2 dx}{\int_D p(x, t)^2 dx}.$$

The first term represents the decrease of the entropy in the system due to the flux of particles through the boundary. The second term represents the entropy production. The last quantity is always positive and is minimal in the stationary state, i.e., when $d\mathcal{S}/dt = 0$. We see that the entropy production is minimal when

$$\frac{\int_D |\nabla p(x, t)|^2 dx}{\int_D p(x, t)^2 dx} = \lambda_1.$$

However, the same minimization problem defines the first eigenfunction of the Laplacian in D with Dirichlet boundary conditions leading to the equation (3.1) in the stationary regime ($dp/dt = 0$). In this sense, the first eigenfunction minimizes the entropy production. We note that since λ_1 is the mean escape rate from the system, the property of minimum entropy production is equivalent to the property of the minimum mean escape rate from the system.

The Rényi entropy belongs to the class of entropies introduced in the nonextensive thermostatics (Pennini, Plastino and Plastino (1998)). In ordinary physical systems it is usually assumed—in view of the second law of thermodynamics—that entropy is an additive quantity and therefore has a properly defined density. This is the case when the

boundary conditions do not strongly influence the bulk properties of the system. This does not hold for the stochastic process considered in this paper since the process of branching in the middle of the system is induced by the flux of particles through the boundary.

REFERENCES

- [1] R.J. Adler and L. Ivanitskaya (1996) “A superprocess with a disappearing self-interaction,” *J. Theoret. Probab.* **9**, 245–261.
- [2] R. Bass and K. Burdzy (1992) “Lifetimes of conditioned diffusions,” *Probab. Th. Rel. Fields* **91**, 405–443.
- [3] K. Burdzy, R. Hołyst, D. Ingerman and P. March (1996) “Configurational transition in a Fleming-Viot-type model and probabilistic interpretation of Laplacian eigenfunctions,” *J. Phys. A* **29**, 2633–2642.
- [4] K. Burdzy and D. Khoshnevisan (1998) “Brownian motion in a Brownian crack” *Ann. Appl. Probab.* **8**, 708–748
- [5] K. Burdzy, E. Toby and R.J. Williams (1989) “On Brownian excursions in Lipschitz domains. Part II. Local asymptotic distributions,” in *Seminar on Stochastic Processes 1988* (E. Cinlar, K.L. Chung, R. Gettoor, J. Glover, editors), 55–85, Birkhäuser, Boston.
- [6] D.A. Dawson (1992) “Infinitely divisible random measures and superprocesses,” in *Stochastic Analysis and Related Topics*, H. Körezlioglu and A.S. Üstünel, Birkhäuser, Boston.
- [7] D.A. Dawson and K. Fleischmann (1997) “Longtime behavior of a branching process controlled by branching catalysts,” *Stoch. Process. Appl.* **71**, 241–257.
- [8] D.A. Dawson and A. Greven (1996) “Multiple Space-Time Scale Analysis For Interacting Branching Models,” *Electronic J. Probab.* **1**, paper no. 14, pages 1–84.
- [9] D.A. Dawson and E.A. Perkins (1998) “Long-time behavior and coexistence in a mutually catalytic branching model,” *Ann. Probab.* **26**, 1088–1138.
- [10] D. Down, S.P. Meyn and R.L. Tweedie (1995) “Exponential and uniform ergodicity of Markov processes,” *Ann. Probab.* **23**, 1671–1691.
- [11] G. Gong and M. Qian (1997) “Entropy production of stationary diffusions on non-compact Riemannian manifolds,” *Sci. China Ser. A* **40**, 926–931.
- [12] K. Itô and P. McKean (1974) *Diffusion Processes and Their Sample Paths*. Springer-Verlag, New York, 2nd edition.

- [13] A. Klenke (1999) “A Review on Spatial Catalytic Branching,” in *Festschrift in honour of D. Dawson*, to appear.
- [14] K.F. Man, K.S. Tang and S. Kwong (1999) *Genetic algorithms. Concepts and designs*. Springer-Verlag, London.
- [15] F. Pennini, A.R. Plastino and A. Plastino (1998) ”Rényi entropies and Fisher informations as measures of nonextensivity in a Tsallis setting”, *Physica A* **258**, 446-457.
- [16] A. Rényi (1961) “On measures of entropy and information,” *Proc. 4-th Berkeley Symp. Math. Stat. Probab.*, vol. **1**, 547–561.
- [17] D. Ruelle (1997) “Entropy production in nonequilibrium statistical mechanics,” *Comm. Math. Phys.* **189**, 365–371.
- [18] A.-S. Sznitman (1991) “Topics in propagation of chaos,” *École d’Été de Probabilités de Saint-Flour XIX—1989*, 165–251, Lecture Notes in Math., 1464, Springer, Berlin.
- [19] H.S. Wio (1994) *An Introduction to Stochastic Processes and Nonequilibrium Statistical Physics*, World Scientific, Singapore.
- [20] W. Yourgrau, A. van der Merwe, and G. Raw (1982) *Treatise on Irreversible and Statistical Thermophysics*. Dover Publications Inc, pp 48-52 New York, 2nd edition.

Krzysztof Burdzy
 Department of Mathematics
 University of Washington
 Box 354350
 Seattle, WA 98195-4350, USA
 e-mail: burdzy@math.washington.edu

Robert Hołyst
 Institute of Physical Chemistry
 Polish Academy of Sciences, Dept. III
 Kasprzaka 44/52,
 01224 Warsaw, Poland
 e-mail: holyst@saka.ichf.edu.pl

Peter March
 Department of Mathematics
 Ohio State University,
 Columbus, OH 43210, USA
 e-mail: march@math.ohio-state.edu