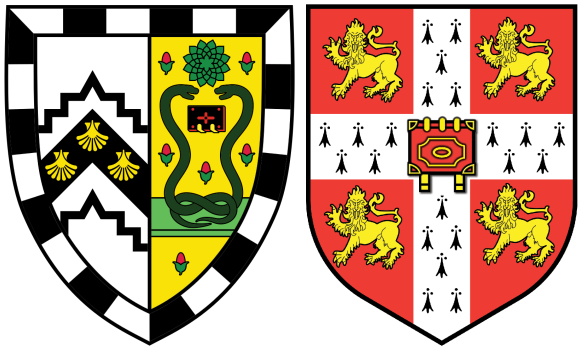
Simulation of Diffusion Limited Aggregation Models and Related Results

James Henry Bell

October 2019

Gonville and Caius College,
University of Cambridge



A thesis submitted for the degree of
Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution.

Acknowledgements

I would like to thank my supervisor Prof James Norris for providing a research topic and being available to answer any questions I had over the last few years, Karen Habermann and Vittoria Silvestri for providing support and encouragement on how to go about a PhD and Kweku Abraham, Tom Berrett, Jo Evans and Benjamin Stokell for advice on PDEs and statistics.

I would also like to thank my parents Ian and Sue and my girlfriend Elinor for supporting me materially, emotionally and with proof-reading over the last stretch of the production of this thesis.

James Bell
Cambridge
12 September 2019

Summary

This thesis is in three parts. All parts are motivated by a desire to gain a better understanding of models of the phenomenon of two-dimensional diffusion limited aggregation, henceforth DLA. The first part proves some generalisations of results relating to Hastings-Levitov with $\alpha = 0$, henceforth HL(0), another two-dimensional growth process. The second part is a study of numerical algorithms for simulating off-grid DLA. The third part describes and reports on some numerical experiments on multiple models of DLA.

Part I provides a generalization of the concept of disturbance flows and of the coalescing Brownian flow, also known as the Brownian web, proving facts about the convergence of the former to the latter and about their time-reversals. This work was motivated as an attempt to generalize known results about the harmonic flow of HL(0) to the case of HL(2), which is supposed to be a model for DLA.

Part II provides the first rigorous analysis of the asymptotic runtimes of four different previously published algorithms for simulating off-grid DLA. A variation on one of these algorithms, incorporating an improvement from another source and a trick new to this work, is implemented in code, with the runtimes comparing favourably to previous work. The runtime of this algorithm, like that of the algorithm it is based on, is $\tilde{O}(n)$, which is optimal.

Part III is a report on experiments testing whether or not off-grid DLA, HL(2) and noise-reduced DLA all have the same limiting shape in the many particle limit. It also contains a heuristic discussion of whether regularized HL can provide a good model for DLA. The results and heuristics indicate that regularizing HL with slit particles is not a promising way to simulate DLA. However, HL with circular particles, off-grid DLA and noise-reduced DLA are found to be in agreement.

Introduction

The unifying topic of the three parts of this thesis is Diffusion Limited Aggregation (DLA). Here we describe the physical motivation for the concept and informally introduce commonly studied models of DLA to refer to later.

DLA is a model for the growth of clusters, where their growth is primarily governed by diffusion. Most clearly from the definition, the relevant situations can involve particles diffusing until they adhere to the cluster, such as in electro deposition. However, similar behaviour occurs in the case of exothermic crystal growth, where the diffusion is of dissipating heat. It is also related to models of dielectric breakdown and can be used to model dielectric breakdown in the high resistance limit [3].

DLA is an irreversible growth process in which a cluster of particles is grown by adding one particle at a time. The particles are added at a point on the boundary chosen according to harmonic measure from infinity, i.e. the point is chosen from the limiting hitting distribution of a random walk or Brownian motion started from far away. The original definition was introduced by Witten and Sander [36] on the square lattice. As lattice based models have been shown to be non-isotropic [3] we shall be primarily interested in lattice free models.

The first lattice free model was that of off-grid DLA introduced by Meakin [18]. In this model, each particle has a fixed radius, say one, and can have any position in $\mathbb{R}^2$. The particles each in turn start from infinity performing a Brownian motion until they make contact with the cluster, at which point they stop and are added to the cluster at their current location. In Part II, we will analyze various algorithms for simulating this model.

The only notable results we know of for the lattice based model are bounds on the fractal dimension (more accurately growth rate of the radius of the cluster). These are due to Kesten [13] and Benjamini and Yadin [5]. No results are known for the Meakin model. This motivates the need for another

model which would allow us to more easily prove results.

Another lattice free model was proposed by Hastings and Levitov [12]. In this model, the growth of the cluster by a particle is simulated by applying a conformal map to the exterior of the cluster. The details of this process are explained more clearly in Part III. The main purpose of Part III is to test experimentally whether this model, and the Meakin model given above, converge to the same distribution of cluster shape.

In fact, Hastings and Levitov proposed a parametrised family of models (also introduced in Part III). Special cases of this model family have yielded a decent level of understanding to the work of Norris and Turner [23] and Silvestri [31]; and a regularized version is considered by Johansson Viklund, Sola and Turner [35]. For this regularized model, they can prove theorems for a larger variety of parameter choices, though the parameters corresponding to DLA remain hard to analyze. This regularized model is discussed in Part III.

Part I

Time-Reversal of Coalescing Diffusive Flows and Weak Convergence of Localized Disturbance Flows

This part is a contribution to the theory of stochastic flows in one dimension, specifically the study of inhomogeneous flows and their time-reversals. We provide two proofs of our main result, Theorem 4.1, which says that the time-reversal of a coalescing diffusive flow with drift b and diffusivity a is (provided the spatial derivative a' of a is Lipschitz) given by a coalescing diffusive flow of drift $-b+\frac{a'}{2}$ and diffusivity a. Theorem 5.4, which establishes convergence of certain families of inhomogeneous disturbance flows to coalescing diffusive flows, may also be of independent interest.

1 Introduction

A disturbance flow, introduced in [24], is a composition of independent random maps of the circle to itself. Unlike [24], we do not require that our maps are identically distributed or that their distributions are invariant under conjugation by a rotation of the circle. For a pair of suitably smooth a, b, we consider limits where the maps F are close to the identity, well localized and have mean of $F(x)-x$ close to $hb(x)$ and variance of $F(x)-x$ close to $ha(x)$ as $h \to 0$. We prove convergence of individual paths to diffusion processes and of the flow as a whole to the coalescing diffusive flow with diffusivity and

drift given by a and b. We also describe the time-reversal of the disturbance flows and use this to describe the time-reversal of a coalescing diffusive flow.

The coalescing diffusive flow consists of a diffusion process starting from each point in space-time, each with drift and diffusivity given by the same functions of space and time. They evolve independently until they collide, at which point they coalesce. The idea of such an object with standard Brownian motions instead of diffusions has been studied widely, starting with Arratia in 1979 [2]. One approach to this is to define a family of random measurable functions $(\phi_{ts} : s \leq t \in \mathbb{R})$ satisfying the flow property

$$\phi_{ts} \circ \phi_{sr} = \phi_{tr}, \qquad r \leq s \leq t$$

and such that every finite collection of trajectories $(\phi_{ts}(x) : t \geq s)$ performs coalescing Brownian motion. This is the approach taken in Arratia [2], Le Jan and Raimond [15] and Tsirelson [34]. A problem, however, with this approach is that the ϕ_{ts} cannot be chosen to be right-continuous, as the composition of two right-continuous functions is not necessarily right-continuous. An alternative approach that avoids this problem is given by Fontes et al. [10] based on completing the set of trajectories to form a compact set of continuous paths, this completion can be done in multiple ways leading to multiple objects known as Brownian webs. Another way around the problem was introduced by Norris and Turner in [24], based on the idea of considering pairs $\{\phi^-, \phi^+\}$ of left and right continuous modifications of the Arratia flow. This setup does not store the information of the value of ϕ_{ts} at a jump, and as a result the flow property must be relaxed to a weak flow property (definition in Section 3 of this part). The space of weak flows with the metric appearing in [24] provides a useful space for studying weak convergence, as it contains flows without continuous trajectories such as disturbance flows. This is the approach that this work builds on. A later paper by Berestycki et al. [6] provides another state-space and topology for the Brownian web, which was based on the quad crossings of Schramm and Smirnov.

Konarovskyi [14] also studies a generalization of coalescing Brownian motions with varying diffusivity. In that work, the diffusions only start from time $t = 0$ and the diffusivity of each is taken to be inversely proportional to the fraction of the diffusions that have coalesced to form it, rather than as as function of position and time, as in this work.

This part is structured as follows. Section 2 proves existence and uniqueness of a simplified version of the coalescing diffusive flows, which consists of only countably many paths. Section 3 defines the metric spaces that our flows take values in, and proves existence and uniqueness of the coalescing diffusive flows (Theorem 3.1). Section 4 defines the time-reversal of a flow and provides the statement of our main result (Theorem 4.1), which identifies the time-reversal of a coalescing diffusive flow. At this point the reader has the option of skipping straight to Section 7 which will not require Sections 5 or 6. Section 5 defines the notion of a disturbance flow, and shows convergence of paths from the flow to diffusions and of countable collections of paths to the simplified flow from Section 2. Section 6 shows convergence of the disturbance flows to coalescing diffusive flows, identifies their time-reversals and uses this to provide a proof of Theorem 4.1. Section 7 provides an alternative proof of Theorem 4.1 that does not require the use of disturbance flows. It also contains as an intermediate weaker version (requiring more smoothness of a and b) Theorem 7.1.

The disturbance flow based approach to our main result is based on [24]; much of the notation is taken from there and some of the proofs are very similar. However, there are multiple places where new ideas are required to handle the generalization. While [24] allows the distribution of disturbances to be random only in that the location of the disturbance is chosen uniformly at random from around the circle, we allow the disturbances to vary in size and shape both randomly and with location in space and time, the shape and size is also allowed to vary a lot more as we take the limit to small disturbances than is allowed in [24]. The new ideas in the proofs are first

evident in the proof of Theorem 5.1, showing that individual trajectories of suitable disturbance flows converge weakly, where the proof of tightness requires bounds that hold despite the possibly varying drift and diffusivity. The time reversal results in Section 6 are generalizations of those in [24]. However, the statement of our main result Theorem 4.1 is not something that you would obviously expect, and the proof had to be modified substantially to deal with the more general disturbance flows.

The proof in Section 7 is original in idea as well as in detail. While it is about the same length as the disturbance flow based proof, the weaker version of our main result Theorem 7.1 (which is identical except it assumes that a and b are Lipschitz in time as well as space) is proved with a substantially smaller amount of work (about 5 pages after the statement has been made rigorous rather than eighteen) and might suffice for future applications. In particular it provides a short proof, without the use of disturbance flows, of the Brownian case which is Corollary 7.2 of [24].

2 Countable Collections of Coalescing Diffusions

In this section, we recall uniqueness in law for weak solutions of SDEs, then define a metric space, D_E, whose elements consist of countable collections of cadlag paths. Finally, using a martingale problem in the style of [32], specifically those corresponding to a countable family of coalescing diffusion processes that are independent until collision; we identify certain elements of D_E.

Given functions $a : \mathbb{R}^2 \to \mathbb{R}_{>0}$ and $b : \mathbb{R}^2 \to \mathbb{R}$ measurable, bounded uniformly on compacts in the first variable and L-Lipschitz in the second, let

$\sigma(t,x) = \sqrt{a(t,x)}$, then the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{1}$$

has uniqueness in law for weak solutions [28], i.e. given $e = (s, x)$ and a triple $(X, W)_{t\geq s}$, $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{t\geq s}$, such that

a) $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space with $(\mathcal{F}_t)_{t\geq s}$ as a filtration satisfying the usual conditions

b) X is adapted to $(\mathcal{F}_t)$, X is continuous and W is an $(\mathcal{F}_t)$-Brownian motion

c) $X_s = x$

d) Almost surely, both X and the quadratic variation of X are bounded on each compact time interval and

e) Almost surely

$$X_t = X_s + \int_s^t b(r, X_r)dr + \int_s^t \sigma(r, X_r)dW_r, \forall t \geq s \tag{2}$$

then the law of X is determined by a,b and e.

We will write this law as $\mu_e^{a,b}$, and say that X is a diffusion process with drift b and diffusivity a. Throughout we will assume that a and b have period 1 in the second variable (as well as the properties above), and X will be considered as a diffusion process on the circle $\mathbb{R}/\mathbb{Z}$.

We will in several proofs use the notation

$$b^* := \sup_{x\in[0,1], r\in I} |b(r, x)|$$

$$a^* := \sup_{x\in[0,1], r\in I} a(r, x)$$

$$a_* := \inf_{x\in[0,1], r\in I} a(r, x)$$

where I is an compact interval of time that contains all the times relevant to the given context. It will only be important that in any given context these

numbers are finite and $a_* > 0$.

Let $D_e = D_x([s,\infty),\mathbb{R})$ be the space of cadlag paths starting from x at time s. Write d_e for the Skorokhod metric on D_e.

Given a sequence $E = (e_k : k \in \mathbb{N})$ in $\mathbb{R}^2$, set

$$D_E = \prod_{k=1}^{\infty} D_{e_k}$$

and define a metric d_E on D_E by

$$d_E(z,z') = \sum_{k=1}^{\infty} 2^{-k}(d_{e_k}(z^k, z'^k) \wedge 1).$$

Then (D_E, d_E) is a complete separable metric space.

Write $e_k = (s_k, x_k)$ and denote by $(Z_t^k)_{t \geq s_k}$ the kth coordinate process on D_E, given by $Z_t^k(z) = z_t^k$. Consider the filtration $(\mathcal{Z}_t)_{t\in\mathbb{R}}$ on D_E, where $\mathcal{Z}_t$ is the σ-algebra generated by $(Z_s^k : s_k < s \leq t \vee s_k, k \in \mathbb{N})$. Write C_E for the (measurable) subset of D_E, where each coordinate path is continuous. Define on C_E

$$T^{jk} = \inf\{t \geq s_j \vee s_k : Z_t^j - Z_t^k \in \mathbb{Z}\}.$$

The T^{jk} are the collision times of the paths considered in $\mathbb{R}/\mathbb{Z}$. The following is a generalization of a reformulation in [24] of a result of Arratia in [2].

Proposition 2.1. *Given a, b measurable and bounded uniformly on compacts in time and L-Lipschitz in space as in* (1)*, there exists a unique Borel probability measure $\mu_E^{a,b}$ on D_E under which, for all $j,k \in \mathbb{N}$, the processes*

$$\left(Z_t^k - \int_{s_k}^t b(s, Z_s^k)ds\right)_{t \geq s_k}$$

and

$$\left(Z_t^k Z_t^j - \int_{s_j \vee s_k}^t Z_s^k b(s, Z_s^j) + Z_s^j b(s, Z_s^k) ds - \int_{T^{jk} \wedge t}^t a(s, Z_s^j) ds \right)_{t \geq s_j \vee s_k}$$

are both continuous local martingales.

We give the following proof sketch. For existence, one can take independent diffusion processes, with coefficients a and b, from each of the given time-space starting points and then impose a rule of coalescence on collision, deleting the path of larger index. The law of the resulting process has the desired properties. On the other hand, given a probability measure such as described in the proposition, on some larger probability space, one can use a supply of independent Brownian motions to build diffusions continuing each of the paths deleted at each collision. Then, the martingale problem characterization of diffusion processes given in [32], can be used to see that one has recovered the set-up used for existence. This gives uniqueness.

3 Existence and Uniqueness of Coalescing Diffusive Flows

We now introduce the space of continuous weak flows $C^\circ(\mathbb{R}, \mathcal{D})$ and the space of cadlag weak flows $D^\circ(\mathbb{R}, \mathcal{D})$, both introduced in [24]. We will then identify certain elements of $C^\circ(\mathbb{R}, \mathcal{D})$ as coalescing diffusive flows, again using a martingale problem. The space $C^\circ(\mathbb{R}, \mathcal{D})$ is sufficient for stating our main result and understanding the proof that doesn't use disturbance flows. However, we will need $D^\circ(\mathbb{R}, \mathcal{D})$ to deal with the fact that the disturbance flows are not continuous in time. The following explanation of notation follows [24] very closely, and all the claims made in italics are proved in [24].

We consider non-decreasing, right-continuous functions $f^+ : \mathbb{R} \to \mathbb{R}$ with

the degree 1 property

$$f^+(x+n) = f^+(x) + n, \qquad x \in \mathbb{R}, \qquad n \in \mathbb{Z}.$$

Let us denote the set of such functions by $\mathcal{R}$ and the set of analogous left-continuous functions by $\mathcal{L}$. Each $f^+ \in \mathcal{R}$ has a left-continuous modification given by $f^-(x) = \lim_{y \uparrow x} f^+(y)$. Let $\mathcal{D}$ denote the set of corresponding pairs $f = \{f^-, f^+\}$. We will write f in place of $f^\pm$ when the choice is irrelevant for the purpose at hand, especially in the case when $f^+ = f^-$, i.e. f^+ is continuous.

Firstly, we define a metric on $\mathcal{D}$. Associate to each function f a function $f^\times$ given by $f^\times(t) = t - x$, where $x \in \mathbb{R}$ is the unique value such that

$$\frac{x + f^-(x)}{2} \leq t \leq \frac{x + f^+(x)}{2}$$

as shown in Figure 1. *We can define a complete locally compact metric* $(\mathcal{D}, d_\mathcal{D})$ *by*

$$d_\mathcal{D}(f, g) = \sup_{t \in [0,1)} |f^\times(t) - g^\times(t)|.$$

Consider $\phi = (\phi_I : I \subseteq \mathbb{R})$, with $\phi_I \in \mathcal{D}$ and I ranging over all non-empty bounded intervals. We say that ϕ is a *weak flow* if given I a disjoint union of intervals I_1 and I_2, with $\sup I_1 =\in I_2$,

$$\phi^-_{I_2} \circ \phi^-_{I_1} \leq \phi^-_I \leq \phi^+_I \leq \phi^+_{I_2} \circ \phi^+_{I_1}.$$

ϕ is said to be *cadlag* if for all $t \in \mathbb{R}$,

$$\phi_{(s,t)} \to \mathrm{id} \quad \text{as} \quad s \uparrow t, \qquad \phi_{(t,u)} \to \mathrm{id} \quad \text{as} \quad u \downarrow t.$$

Here, the convergence of functions is in the uniform norm (also note that this definition is left-right symmetric, we call it cadlag to match previous work).

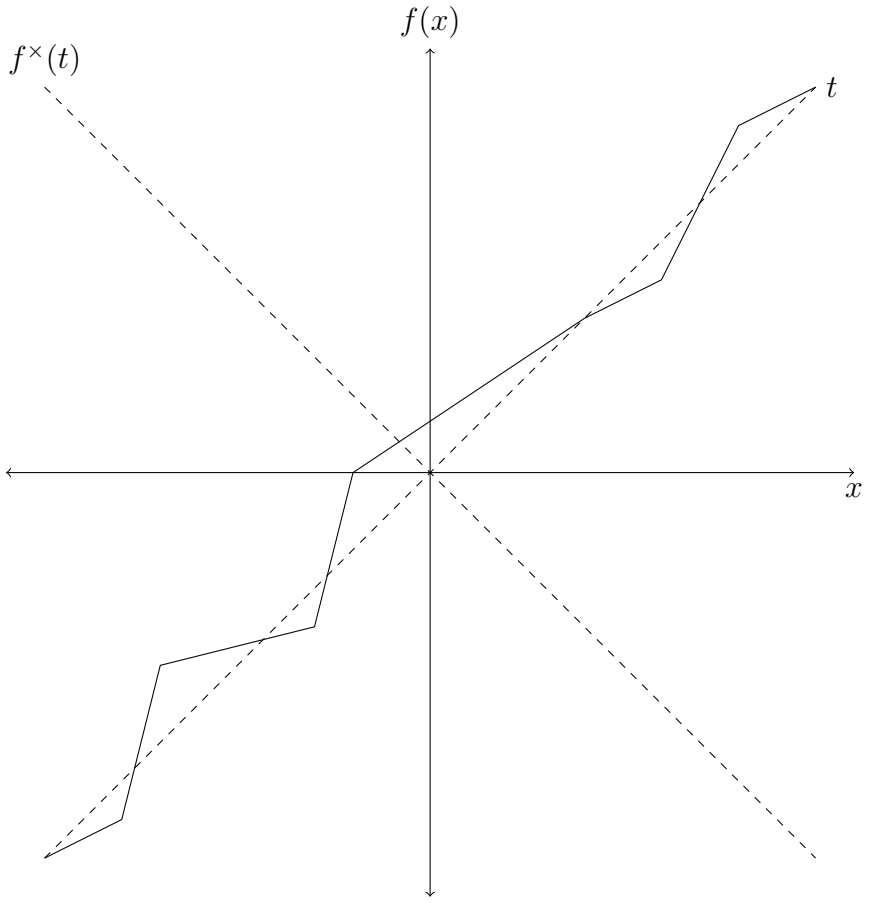


Figure 1: The graph of $f^\times$ can be formed from the graph of f by rotating the axes by $\frac{\pi}{4}$ and scaling both axes up by $\sqrt{2}$

$D^\circ(\mathbb{R},\mathcal{D})$ is the set of cadlag weak flows. We set $\phi_\emptyset = \mathrm{id}$. Given $\{I_n : n \in \mathbb{N}\}$ and I bounded intervals, write $I_n \to I$ if

$$I = \bigcup_n \bigcap_{m\geq n} I_m = \bigcap_n \bigcup_{m\geq n} I_m.$$

For every $\phi \in D^\circ(\mathbb{R},\mathcal{D})$, we have

$$\phi_{I_n} \to \phi_I \text{ whenever } I_n \to I$$

If $\phi \in D^\circ(\mathbb{R},\mathcal{D})$ satisfies $\phi_{\{t\}} = \mathrm{id}$ for all $t \in \mathbb{R}$ then we have that $\phi_{(s,t)} = \phi_{(s,t]} = \phi_{[s,t)} = \phi_{[s,t]}$ for all $s < t$. Denoting these all by ϕ_{ts} we define $C^\circ(\mathbb{R},\mathcal{D})$ to be the set of all such $(\phi_{ts} : s,t \in \mathbb{R}, s < t)$. For $\phi,\psi \in C^\circ(\mathbb{R},\mathcal{D})$ and $n \geq 1$, define

$$d_C^{(n)}(\phi,\psi) = \sup_{s,t\in(-n,n),s<t} d_{\mathcal{D}}(\phi_{ts},\psi_{ts})$$

and then let

$$d_C(\phi,\psi) = \sum_{n=1}^{\infty} 2^{-n}\left(d_C^{(n)}(\phi,\psi)\wedge 1\right).$$

Under this metric $C^\circ(\mathbb{R},\mathcal{D})$ *is complete and separable.*

In the interests of defining a metric on $D^\circ(\mathbb{R},\mathcal{D})$, for λ an increasing homeomorphism of $\mathbb{R}$ we define

$$\gamma(\lambda) = \sup_{t\in\mathbb{R}}|\lambda(t)-t| \vee \sup_{s,t\in\mathbb{R},s<t}\left|\log\left(\frac{\lambda(t)-\lambda(s)}{t-s}\right)\right|,$$

and let χ_n be the cut-off function given by

$$\chi_n(I) = 0 \vee (n+1-R)\wedge 1, \quad R = \sup I \vee (-\inf I).$$

We can now define for $\phi, \psi \in D^\circ(\mathbb{R}, \mathcal{D})$ and $n \geq 1$,

$$d_D^{(n)}(\phi, \psi) = \inf_\lambda \left\{ \gamma(\lambda) \vee \sup_{I \subseteq \mathbb{R}} \|\chi_n(I)\phi_I^\times - \chi_n(\lambda(I))\psi_{\lambda(I)}^\times\|_\infty \right\}$$

where the infimum is taken over the set of increasing homeomorphisms λ of $\mathbb{R}$. Then define

$$d_D(\phi, \psi) = \sum_{n=1}^{\infty} 2^{-n} \left(d_D^{(n)}(\phi, \psi) \wedge 1 \right).$$

Then $(D^\circ(\mathbb{R}, \mathcal{D}), d_D)$ is a *complete and separable* metric space. Moreover d_C *and* d_D *generate the same topology on* $C^\circ(\mathbb{R}, \mathcal{D})$. *For the metric* d_D, *all bounded intervals* I *and all* $x \in \mathbb{R}$, *the* evaluation map

$$\phi \mapsto \phi_I^+(x) : D^\circ(\mathbb{R}, \mathcal{D}) \to \mathbb{R}$$

is *Borel measurable. Moreover the Borel* $\sigma-$*algebra on* $D^\circ(\mathbb{R}, \mathcal{D})$ *is generated by the set of all such evaluation maps with* $I = (s, t]$ *and* s, t *and* x *rational.*

For $e = (s, x) \in \mathbb{R}$ *and* $\phi \in D^\circ(\mathbb{R}, \mathcal{D})$, *the maps*

$$t \mapsto \phi_{(s,t]}^\pm(x) : [s, \infty) \to \mathbb{R}$$

are cadlag. Hence we can define $Z^e = Z^{e,+}$ and $Z^{e,-}$, as maps from $D^\circ(\mathbb{R}, \mathcal{D})$ to D_e, by setting

$$Z^{e,\pm}(\phi) = (\phi_{(s,t]}^\pm(x) : t \geq s).$$

The maps, $t \to Z_t^{e,\pm}(\phi)$ *are continuous when* $\phi \in C^\circ(\mathbb{R}, \mathcal{D})$.

Finally, define a σ-algebra $\mathcal{F}$ and a filtration $(\mathcal{F}_t)_{t \in \mathbb{R}}$ on $C^\circ(\mathbb{R}, \mathcal{D})$ by

$$\mathcal{F} = \sigma(Z_t^{(s,x)} : (s, x) \in \mathbb{R}^2, t \geq s)$$

and

$$\mathcal{F}_t = \sigma(Z_r^{(s,x)} : (s, x) \in \mathbb{R}^2, r \in (-\infty, t] \cap [s, \infty)).$$

Then $\mathcal{F}_t$ *is generated by the random variables* $Z_r^{(s,x)}$ with $(s,x) \in \mathbb{Q}^2$ and $r \in (-\infty, t] \cap [s, \infty)$, and $\mathcal{F}$ *is the Borel* σ*-algebra of the metric* d_C.

The following theorem states the existence of coalescing diffusive flows. The proof given follows the line of argument for the less general result Theorem 3.1 in [24]. The italicized assertions in the proof below are proved in [24]. Generalizing the argument requires generalized versions of results, from [24] which are Proposition 2.1 and Proposition 8.1.

Analogously to T^{jk}, if $e = (s,x)$ and $e' = (s',x')$ we define

$$T^{ee'} = \inf\{t \geq s \vee s' : Z_t^e - Z_t^{e'} \in \mathbb{Z}\}.$$

Theorem 3.1. *Given* a, b *as before, there exists a unique Borel probability measure* $\mu_A^{a,b}$ *on* $C^\circ(\mathbb{R}, \mathcal{D})$ *under which, for all* $e = (s,x), e' = (s',x') \in \mathbb{R}^2$, *the processes*

$$\left(Z_t^e - \int_s^t b(r, Z_r^e) dr \right)_{t \geq s} \tag{3}$$

and

$$\left(Z_t^e Z_t^{e'} - \int_{s \vee s'}^t Z_r^e b(r, Z_r^{e'}) + Z_r^{e'} b(r, Z_r^e) dr - \int_{T^{ee'} \wedge t}^t a(r, Z_r^{e'}) dr \right)_{t \geq s \vee s'} \tag{4}$$

are continuous local martingales with respect to $(\mathcal{F}_t)_{t \in \mathbb{R}}$. *Moreover, for all* $e \in \mathbb{R}^2$ *we have* $\mu_A^{a,b}$*-almost surely* $Z^{e,+} = Z^{e,-}$.

Proof. Fix an enumeration $E = (e_k : k \in \mathbb{N})$ of $\mathbb{Q}^2$. Define $Z^{E,\pm} : C^\circ(\mathbb{R}, \mathcal{D}) \to C_E$ by $Z^{E,\pm}(\phi) = (Z^{e_k,\pm}(\phi) : k \in \mathbb{N})$. Then, *we have* $\mathcal{F}_t = \{(Z^{E,+})^{-1}(B) : B \in \mathcal{Z}_t\}$. Set

$$C_E^{\circ,\pm} = \{Z^{E,\pm}(\phi) : \phi \in C^\circ(\mathbb{R}, \mathcal{D})\}.$$

Then *the sets* $C_E^{\circ,\pm}$ *are measurable subsets of* C_E and, by Proposition 8.1, $\mu_E^{a,b}(C_E^{\circ,\pm}) = 1$. Moreover $Z^{E,\pm}$ maps $C^\circ(\mathbb{R}, \mathcal{D})$ *bijectively to* $C_E^{\circ,\pm}$ *and the inverse bijections* $C_E^{\circ,\pm} \to C^\circ(\mathbb{R}, \mathcal{D})$, *which we denote by* $\Phi^{E,\pm}$, *are measurable.*

Write Z^E for $Z^{E,+}$ and Φ^E for $\Phi^{E,+}$. Then, on $C_E^{\circ,+}$, for all $j,k \in \mathbb{N}$, we have

$$Z^{e_k} \circ \Phi^E = Z^k, \qquad T^{e_j e_k} \circ \Phi^E = T^{jk}$$

and for all $t \in \mathbb{R}$ and $B \in \mathcal{F}_t$ we have $1_B \circ \Phi^E = 1_{B'}$ for some $B' \in \mathcal{Z}_t$. Define $\mu_A^{a,b} = \mu_E^{a,b} \circ (\Phi^E)^{-1}$. By Proposition 2.1, under $\mu_A^{a,b}$, for all $j,k \in \mathbb{N}$, taking $e = e_j$ and $e' = e_k$ makes the processes (3) and (4) into continuous local martingales for $(\mathcal{F}_t)_{t\in\mathbb{R}}$.

On the other hand, for every probability measure μ on $C^\circ(\mathbb{R},\mathcal{D})$ having this property, under the image measure $\mu \circ (Z^E)^{-1}$ on C_E, for all $j,k \in \mathbb{N}$, taking $e = e_j$ and $e' = e_k$ makes the processes (3) and (4) into continuous local martingales for $(\mathcal{Z}_t)_{t\in\mathbb{R}}$, so $\mu \circ (Z^E)^{-1} = \mu_E^{a,b}$ by Proposition 2.1, and so $\mu = \mu_A^{a,b}$.

Given $e_{-1}, e_0 \in \mathbb{R}^2$, all the assertions above hold when E is replaced by the sequence $E' = (e_{-1}, e_0, e_1, e_2, ...)$. We repeat the steps taken to obtain a probability measure $\mu_A'^{a,b} = \mu_{E'}^{a,b} \circ (\Phi^{E'})^{-1}$ on $C^\circ(\mathbb{R},\mathcal{D})$. Then, under $\mu_A'^{a,b}$, taking $e = e_{-1}$ and $e' = e_0$ makes the processes (3) and (4) into continuous local martingales for $(\mathcal{F}_t)_{t\in\mathbb{R}}$. But also, under $\mu_A'^{a,b}$, for all $j,k \in \mathbb{N}$, taking $e = e_j$ and $e' = e_k$ makes the processes (3) and (4) into continuous local martingales for $(\mathcal{F}_t)_{t\in\mathbb{R}}$, so $\mu_A^{a,b} = \mu_A'^{a,b}$.

Finally, *we have* $\Phi^{E',+} = \Phi^{E',-}$ *on* $C_{E'}^{\circ,-} \cap C_{E'}^{\circ,+}$, so

$$Z^{e,-}(\Phi^{E'}) = Z^{e,-}(\Phi^{E',-}) = Z^{e,+}(\Phi^{E'})$$

$\mu_{E'}^{a,b}$-almost surely, and so $Z^{e,-} = Z^{e,+}$, $\mu_A^{a,b}$-almost surely, as claimed. □

We will often write μ_A instead of $\mu_A^{a,b}$ in order to simplify notation.

4 Time Reversal

In this section we quote some definitions and observations from [24] and then state our main theorem. For $f^+ \in \mathcal{R}$ and $f^- \in \mathcal{R}$, define the *left-continuous inverse* from $\mathcal{R}$ to $\mathcal{L}$ and the inverse operation *right-continuous inverse* respectively as follows

$$(f^+)^{-1}(y) = \sup\{x \in \mathbb{R} : f^+(x) < y\},$$

$$(f^-)^{-1}(y) = \inf\{x \in \mathbb{R} : f^-(x) > y\}.$$

Note that these operations are distributive over concatenation. The *inverse* of $f \in \mathcal{D}$ is given by

$$f^{-1} = \{(f^+)^{-1}, (f^-)^{-1}\} \in \mathcal{D}.$$

The *time-reversal* $\hat{\phi}$ of a flow ϕ is given by

$$\hat{\phi}_I = \phi_{-I}^{-1}.$$

The time-reversal map is a well defined isometry of both $D^\circ(\mathbb{R}, \mathcal{D})$ and $C^\circ(\mathbb{R}, \mathcal{D})$.

As before, let a and b be the diffusivity and drift of a diffusive flow with law μ_A. We require that a and b satisfy the smoothness requirements of Section 2 and further require that a is differentiable with respect to x with derivative $a'(t,x)$, which is L-Lipschitz in x and measurable and bounded uniformly on compacts in t. Let $a^\nu(t,x) = a(-t,x)$ and $b^\nu(t,x) = -b(t,-x) + a'(-t,x)/2$, let $\nu_A = \mu_A^{a^\nu, b^\nu}$, i.e. let it be the law of a disturbance flow with drift and diffusivity given by b^ν and a^ν. Finally, write $\hat{\mu}_A$ for the image measure of μ_A under time-reversal.

Theorem 4.1. *The time-reversal of the diffusive flow μ_A is a diffusive flow*

with the new parameters given in the previous paragraph, i.e.

$$\hat{\mu}_A = \nu_A$$

We will provide two proofs of this theorem: one in Section 6 which depends on Section, 5 and one in Section 7 which does not depend on Sections 5 or 6.

5 Disturbance Flows from Countably Many Points on a Circle

This section lays the ground work for Section 6. The reader may skip to Section 7 at this point if they only wish to read the direct proof.

We start this section by defining the notion of a disturbance flow on the circle. This is based on a notion of disturbance flow which was given in [24], but is more general, so as to allow for our disturbance flows to have drift and varying diffusivity. We will then proceed to state and prove two propositions and deduce a theorem. The propositions are as follows: firstly, under appropriate conditions a sequence of single paths from disturbance flows can converge to a diffusion process; and secondly, a sequence of countable families of paths from disturbance flows can converge to a countable family of coalescing diffusions. Combining these propositions with a result from [24], we conclude that disturbance flows can converge to coalescing diffusive flows.

We specify a disturbance flow by a family of probability distributions on $\mathcal{D}$ written

$$\eta = \{\eta_{h,t} : h > 0, t \in \mathbb{R}\}.$$

The parameters of the family are $h > 0$, which corresponds to the size of the disturbance (the limit for our convergence later will be taking h to 0 while

making disturbances more frequent) and time t, which allows our flow to be inhomogeneous in time. We require that η be measurable as a function of t.

Given $f_1, f_2 \in \mathcal{D}$, define $f_2 \circ f_1 := \{f_2^- \circ f_1^-, f_2^+ \circ f_1^+\}$. This is not in general an element of $\mathcal{D}$, however, so long as f_1 sends no interval of positive length to a point of discontinuity of f_2, we will have $f_2 \circ f_1 \in \mathcal{D}$. To avoid this issue, we will only consider families of probability distributions on $\mathcal{D}$ such that, if $F_{h,t} \sim \eta_{h,t}$, then

$$F_{h,t}^+(x) = F_{h,t}^-(x) \text{ a.s. } \forall x, t \in \mathbb{R} \text{ and } h \in \mathbb{R}^+.$$

We denote the set of such families by $\mathcal{D}^*$, and assume from here on that $\eta = \{\eta_{h,t} : h > 0, t \in \mathbb{R}\} \in \mathcal{D}^*$. Let N be a Poisson random measure on $\mathbb{R}$ of intensity h^{-1} and set

$$N_t = \begin{cases} N(0,t], & t \geq 0 \\ -N(t,0], & t < 0. \end{cases}$$

Let

$$t_n = \inf\{t : N_t \geq n\}$$

and $\{F_{h,t_n} : h > 0, n \in \mathbb{N}\}$ be independent random variables with $F_{h,t_n} \sim \eta_{h,t_n}$. We will sometimes write F_n for F_{h,t_n}.

We extend the inverse functions of Section 4 to families of probability distributions $F \in \mathcal{D}^*$ by setting

$$(F^{-1})_{h,t}(y) = (F_{h,t})^{-1}(y)$$

where the inverse on the right hand side is being taken with respect to the x argument (as opposed to the implicit ω argument). Also let $\tilde{F}_{h,t}(x) = F_{h,t}(x) - x$.

Then, for any interval I, define

$$\Phi_I(x) = x + \int_I \tilde{F}_{h,r}(\phi_{I \cap (-\infty, r)}) N(dr).$$

Write Φ for the family of maps Φ_I where I ranges over all bounded intervals in $\mathbb{R}$. We call Φ the *Poisson disturbance flow* or just the *disturbance flow* and write μ_A^η for the distribution of Φ in $D^\circ(\mathbb{R}, \mathcal{D})$.

Fixing $e = (s, x) \in \mathbb{R}^2$ we define 2 processes $X_t^{e,\pm}$ by setting $X_t^{e,\pm} = \Phi^\pm_{(s,t]}(x)$ for $t \geq s$. Because $\Phi \in \mathcal{D}$ a.s. we have a.s. that for all $t \in \mathbb{Q}_{\geq s}$

$$X_t^{e,-} = X_t^{e,+}$$

and thus by right continuity of $X^{e,\pm}$ we have a.s. that for all $t \geq s$,

$$X_t^{e,-} = X_t^{e,+}.$$

Thus, we drop the $\pm$ and write simply X^e. Write μ_e^η for the distribution of X^e on the Skorokhod space D_e. Similarly, for $E = (e_k \in \mathbb{R}^2 : k \in \mathbb{N})$, $(X^{e_k} : k \in \mathbb{N})$ is a random variable in D_E, and we write μ_E^η for its distribution on D_E.

Given a family $\eta \in \mathcal{D}^*$, and coefficients a and b as in Section 2, we define the functions

$$b_h(t, x) = \frac{1}{h}\mathbb{E}(\tilde{F}_{h,t}(x))$$

$$a_h(t, x) = \frac{1}{h}\mathbb{E}(\tilde{F}_{h,t}(x)^2)$$

$$M_h = \sup_{x\in[0,1],t\leq T,\omega\in\Omega} |\tilde{F}_{h,t}(x)|$$

$$B_h = \sup_{x\in[0,1],t\leq T} |b_h - b|$$

$$A_h = \sup_{x\in[0,1],t\leq T} |a_h - a|.$$

The following three conditions will be important for the next proposition and

consequently for the rest of the results:

$$\lim_{h \searrow 0} B_h = 0 \qquad \forall T \in \mathbb{R}^+ \tag{5}$$

$$\lim_{h \searrow 0} A_h = 0 \qquad \forall T \in \mathbb{R}^+ \tag{6}$$

$$\lim_{h \searrow 0} M_h = 0 \qquad \forall T \in \mathbb{R}^+. \tag{7}$$

Proposition 5.1. *Suppose a and b are coefficients as in Section 2, and that η is such that conditions* (5), (6) *&* (7) *hold. Then we have $\mu_e^\eta \to \mu_e^{a,b}$ weakly on D_e, as $h \to 0$.*

Proof. Let $(X^n)_{n\in\mathbb{N}}$ be a sequence of processes distributed according to μ_e^η with $h \to 0$ as $n \to \infty$. By the definition of the Skorokhod metric, it suffices to show that for any $T > s$, the restrictions of X^n to $[s, T]$ converge weakly to a solution of the SDE on $[s, T]$. For the remainder of this proof we consider X^n to be restricted to $[s, T]$. We then take $e = (s, x) = (0, 0)$ and $T = 1$, without loss of generality.

Firstly, we shall calculate (up to an error that is small for small $|t - s|$) two expected values (defined in terms of $s, t \in \mathbb{R}$). We shall then prove a characterization of tightness of the sequence. This will require us to use these calculations to show that the process can't vary too much on a given interval, then deduce the existence of a subsequential limit of each subsequence by Prokhorov's theorem. Finally we will identify the distribution of every subsequential limit as a weak solution of equation (1), using again the 2 expectation calculations. Then we will conclude the proof using the uniqueness in law for such solutions .

Let $\mathcal{F}_t^n$ be the completion of the filtration generated by X^n. For $1 \geq t \geq$

$s \geq 0$ we have

$$\begin{aligned}\mathbb{E}(X_t^n - X_s^n|\mathcal{F}_s^n) &= e^{-\frac{t-s}{h}}\frac{t-s}{h}\mathbb{E}(\tilde{F}_{N_s+1}(X_s^n)|t_{N_s+1} \leq t < t_{N_s+2}) + E_1 \\ &= \int_s^t b_h(r, X_s^n)dr + E_1 + E_2 \\ &= \mathbb{E}\left(\int_s^t b_h(r, X_r^n)dr\right) + E_1 + E_2 + E_3.\end{aligned}$$

Where the approximation errors E_i can be bounded as follows. Let $G_{s,k} = \tilde{F}_{N_s+k} \circ ... \circ \tilde{F}_{N_s+1}$

$$\begin{aligned}|E_1| =& \left|\sum_{k\geq 2} \exp\left(-\frac{t-s}{h}\right)\frac{(t-s)^k}{k!h^k}\mathbb{E}(G_{s,k}(X_s^n)|t_{N_s+k} \leq t < t_{N_s+k+1})\right| \\ \leq& \exp\left(-\frac{t-s}{h}\right)\sum_{k\geq 2}\frac{(t-s)^k}{k!h^k}kM_h \\ =& \exp\left(-\frac{t-s}{h}\right)\frac{t-s}{h}M_h\sum_{k\geq 1}\frac{(t-s)^k}{k!h^k} \\ =& M_h\frac{t-s}{h}\exp\left(-\frac{t-s}{h}\right)\left(\exp\left(\frac{t-s}{h}\right) - 1\right) \\ =& O\left((t-s)^2\right)\end{aligned}$$

Note that $\int_s^t b_h(r, X_s^n)dr = \frac{t-s}{h}\mathbb{E}(\tilde{F}_{N_s+1}(X_s^n)|t_{N_s+1} \leq t < t_{N_s+2})$ to under-

stand E_2.

$$\begin{aligned}
|E_2| &= \left(1 - \exp\left(-\frac{t-s}{h}\right)\right) \frac{t-s}{h} \mathbb{E}(\tilde{F}_{N_s+1}(X_s^n) | t_{N_s+1} \leq t < t_{N_s+2}) \\
&\leq \left(1 - \exp\left(-\frac{t-s}{h}\right)\right) \frac{t-s}{h} M_h \\
&= O\left((t-s)^2\right)
\end{aligned}$$

$$\begin{aligned}
|E_3| &= \mathbb{E}\left(\left|\int_s^t b_h(r, X_s^n) - b_h(r, X_r^n) dr\right|\right) \\
&\leq (t-s)(2b^* + 2B_h)\mathbb{P}(t_{N_s+1} \leq t) \\
&= O\left((t-s)^2\right).
\end{aligned}$$

Breaking the interval $(s, t]$ into a large number of small intervals and taking the limit as the interval sizes go to 0, we have that:

$$\mathbb{E}(X_t^n - X_s^n \mid \mathcal{F}_s^n) = \mathbb{E}\left(\int_s^t b_h(r, X_r^n) dr \middle| \mathcal{F}_s^n\right). \tag{8}$$

Similarly

$$\mathbb{E}((X_t^n - X_s^n)^2 \mid \mathcal{F}_s^n) = \mathbb{E}\left(\int_s^t a_h(r, X_r^n) dr \middle| \mathcal{F}_s^n\right). \tag{9}$$

The characterization of tightness that we shall use is given in Billingsley 1968 [7] Theorem 15.3, it says that tightness is equivalent to the following 2 conditions holding:

1. For all $\epsilon > 0$ there exists a K such that

$$\mathbb{P}\left(\sup_t |X_t^n| \geq K\right) \leq \epsilon, \qquad \forall n \geq 1.$$

2. Taking

$$w''_{X^n}(\delta) = \sup_{\substack{t_1 \le t \le t_2 \\ t_2 - t_1 \le \delta}} \min\{|X^n(t) - X^n(t_1)|, |X^n(t) - X^n(t_2)|\}$$

and

$$w_{X^n}(I) = \sup_{s,t \in I} |X^n_s - X^n_t|$$

for all $\epsilon > 0$ there exists $\delta \in (0,1)$ and $N \in \mathbb{N}$ such that

$$\mathbb{P}(w''_{X^n}(\delta) \ge \epsilon) \le \epsilon, \qquad \forall n \ge N$$

and

$$\mathbb{P}(w_{X^n}[0,\delta) \ge \epsilon) \le \epsilon, \qquad \forall n \ge N$$

and

$$\mathbb{P}(w_{X^n}(1-\delta, 1] \ge \epsilon) \le \epsilon, \qquad \forall n \ge N.$$

Note that B_h and A_h going to 0 as $n \to \infty$ means that $|b_h|$ and $|a_h|$ are bounded uniformly in n, x and $t \in [0,1]$. We call the bounds B and A respectively.

The first condition can be shown as follows, where T_K is the first time t such that $X^n_t \ge K$.

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t \le 1} X^n_t \ge K\right) \\
=&\mathbb{P}\left(T_K \le 1\right) \\
\le&\mathbb{P}\left(X^n_1 \ge \frac{K}{2}\right) + \mathbb{P}\left(T_K \le 1, X^n_1 \le \frac{K}{2}\right) \\
\le&\mathbb{P}\left(X^n_1 \ge \frac{K}{2}\right) + \mathbb{E}\left(\mathbb{P}\left(X^n_1 - X^n_{T_k \wedge 1} \le -\frac{K}{2} \middle| \mathcal{F}_{T_K \wedge 1}\right)\right) \\
\le&\frac{2A}{(\frac{K}{2} - B)^2}
\end{aligned}$$

where in the final inequality we have used Chebyshev's inequality. This bound goes to 0 as $K \to \infty$ uniformly in h. Combining with a corresponding bound for $\inf X_t^n$ gives the first condition.

Note that for the second condition, it suffices to show the following stronger statement, where $\mathbb{I}_\delta$ is the set of subintervals of $[0,1]$ of length δ.

For all $\epsilon > 0$ there exists $\delta \in (0,1)$ and $N \in \mathbb{N}$ such that

$$\mathbb{P}(\exists I \in \mathbb{I}_\delta \text{ such that } w_{X^n}(I) \geq \epsilon) \leq \epsilon \qquad \forall n > N \tag{10}$$

which in turn is implied by the following, where $\mathbb{I}'_\delta$ is the set of intervals of length δ with endpoints that are multiples of $\delta/2$.

For all $\epsilon > 0$ there exists δ with $2 \leq \frac{1}{\delta} \in \mathbb{N}$ and $N \in \mathbb{N}$ such that

$$\mathbb{P}(\exists I \in \mathbb{I}'_\delta \text{ such that } w_{X^n}(I) \geq \epsilon) \leq \epsilon \qquad \forall n > N.$$

There are only $\frac{2}{\delta}$ elements in $\mathbb{I}'_\delta$, so using a union bound it suffices to show that for sufficiently small h and some δ we have

$$\sup_{I \in \mathbb{I}'_\delta} \mathbb{P}(w_{X^n}(I) \geq 4\epsilon) \leq \frac{\delta\epsilon}{2}$$

where a factor of 4 has been included purely for convenience later.

We present the proof for $I = [0, \delta]$ but the same argument and bound will hold for all $I \in \mathbb{I}'_\delta$. We have that

$$\mathbb{P}(w_{X^n}(I) \geq 4\epsilon) \leq \mathbb{P}\left(\sup_{t \leq \delta} X_t^n \geq 2\epsilon\right) + \mathbb{P}\left(\inf_{t \leq \delta} X_t^n \leq -2\epsilon\right).$$

We will bound the first term on the right with a bound that will also apply to the second term by symmetry.

Unfortunately, Chebyshev is not strong enough to bound the first term sufficiently tightly. We will apply the Azuma-Hoeffding inequality which requires the following set-up. Let $X_t'^n = X_t^n - tB$ and note that this is a

super-martingale. Fix $0 < \alpha < \frac{1}{2}$, let $R_0 = 0$ and for $i \geq 1$ let R_i be the first time $t > R_{i-1}$ such that $|X'^n_t - X'^n_{R_{i-1}}| \geq M_h^\alpha$.

Firstly, we show that only about $\delta M_h^{-2\alpha}$ of the R_i are less than δ. Consider the distribution of $R_i - R_{i-1}$ conditional on $\mathcal{F}^n_{R_{i-1}}$, by the same argument used in the first condition we have the following for $l < \frac{M_h^\alpha}{4B}$.

$$\begin{aligned}\mathbb{P}(R_i - R_{i-1} \leq l) &\leq \frac{2lA}{(M_h^\alpha - 2lB)^2} \\ &\leq l\frac{8A}{M_h^{2\alpha}}.\end{aligned}$$

From which we deduce that $R_i - R_{i-1}$ stochastically dominates the uniform distribution on $[0, \frac{M_h^{2\alpha}}{8A}]$ for sufficiently small h. An application of the Azuma-Hoeffding Inequality to uniform random variables gives the following.

$$\mathbb{P}(R_{\lceil \frac{32A\delta}{M_h^{2\alpha}} \rceil} \leq \delta) \leq \exp\left(-\frac{A\delta}{M_h^{2\alpha}}\right)$$

Thus letting $J = \lceil \frac{32A\delta}{M_h^{2\alpha}} \rceil$ and R be the minimum of R_J and the first time R_i such that $X'_{R_i} > \epsilon - \frac{M_h}{2}$,

$$\mathbb{P}\left(\sup_{t\leq\delta} X'^n_t \geq 2\epsilon\right) \leq \exp\left(-\frac{A\delta}{M_h^{2\alpha}}\right) + \mathbb{P}\left(\sup_{i\leq J} X'^n_{R_i} \geq 2\epsilon - M_h\right)$$

$$\begin{aligned}&\mathbb{P}\left(\sup_{i\leq J} X'^n_{R_i} \geq 2\epsilon - M_h\right) \leq \\ &\qquad \mathbb{P}\left(X'^n_{R_J} \geq \epsilon - \frac{M_h}{2}\right) + \mathbb{E}\left(\mathbb{P}\left(X'^n_{R_J} - X'^n_R \leq -\epsilon + \frac{M_h}{2}\middle|\mathcal{F}^n_R\right)\right)\end{aligned}$$

We will bound the first term on the right of the last inequality, and note the second term can be bounded similarly. Let $X''^n_i = X'^n_{R_i} - iM_h$. Note that this

is a discrete super-martingale with step size bounded by $M_h^\alpha + M_h$.

$$\begin{aligned}\mathbb{P}\left(X'^n_{R_J} \geq \epsilon - \frac{M_h}{2}\right) &\leq \mathbb{P}\left(X''^n_J \geq \epsilon - (J+1)M_h\right) \\ &\leq \exp\left(-\frac{(\epsilon-(J+1)M_h)^2}{2J(M_h^\alpha + M_h)^2}\right) \\ &\leq \exp\left(-\frac{\epsilon^2}{65A\delta}\right) \text{ for sufficiently small } h \qquad (11)\end{aligned}$$

where we have used the Azuma-Hoeffding inequality again. Bringing these bounds together gives that for a given δ we have for sufficiently small h that

$$\sup_{I\in\mathbb{I}'_\delta} \mathbb{P}(w_{X^n}(I) \geq 4\epsilon) \leq 2\exp\left(-\frac{A\delta}{M_h^{2\alpha}}\right) + 4\exp\left(-\frac{\epsilon^2}{65A\delta}\right).$$

Thus, by choosing δ so that the second term is less than $\frac{\delta\epsilon}{4}$, and then choosing N such that, for all $n \geq N$ we have, h is sufficiently small that the bound (11) holds and the first term is less than $\frac{\delta\epsilon}{4}$, we can conclude that the second condition holds and the sequence μ_e^η is tight.

By Prokhorov's theorem, we now know that every subsequence has a weakly convergent subsequence, and by standard arguments it suffices to show that the limit of every such sequence is $\mu_e^{a,b}$ (restricted to $[0,1]$). Let μ be the limit of such a subsequence and X be distributed according to μ.

We now show that X is a solution of the SDE (1). Let $(\mathcal{F}_t)_{t\geq s}$ be the completion of the filtration generated by X, and let W be defined as follows.

$$W_t = \int_0^t \frac{1}{\sigma(s,X_s)} dX_s - \int_0^t \frac{b(s,X_s)}{\sigma(s,X_s)} ds$$

Note continuity of X follows from the bound (10), and so $\mathcal{F}$ is right-continuous and thus satisfies the usual conditions. It is immediate that $X_s = x$ and Equation (2) holds by the definition of W.

The identities (8) and (9) show in the limit $n \to \infty$ that both X and the quadratic variation of X are a.s. bounded on each compact interval. The

same argument used to get these identities can also be used to find that

$$\mathbb{E}(W_t - W_s|\mathcal{F}_s) = 0$$

and

$$\mathbb{E}((W_t - W_s)^2|\mathcal{F}_s) = t - s.$$

From the definition of W and the continuity of X, we can deduce W is continuous a.s., putting this together with the above expectations we can conclude by Lévy-Characterization that W is a $(\mathcal{F}_t)$-Brownian motion.

Thus X solves (1) and has the required law. □

Define $\lambda_h(f)$ to be the infimum of λ such that,

$$\lambda \leq |x - y| \leq 1 - \lambda \implies \frac{1}{h}\mathbb{E}(|\tilde{F}_{h,t}(x)\tilde{F}_{h,t}(y)|) < \lambda \qquad \forall t.$$

Proposition 5.2. *Under the conditions of Proposition 5.1 and that $\lambda_h \to 0$, we have $\mu_E^\eta \to \mu_E^{a,b}$ weakly on D_E.*

Proof. We write X^k for X^{e_k}. The family of laws on D_E is tight as each family of marginal laws on D_{e_k} is tight. Let μ be a weak limit law for μ_E^η, then for all j, k and all $t > s \geq s_j \vee s_k$, letting $\mathbb{E}^*(\cdot) = \mathbb{E}(\cdot \mid t_{N_s+1} \leq t < t_{N_s+2}, \mathcal{F}_s)$ we

have:

$$\begin{aligned}
&\mathbb{E}(X_t^j X_t^k - X_s^j X_s^k \mid \mathcal{F}_s) \\
=&\frac{t-s}{h}\mathbb{E}^*(F_{N_s+1}(X_s^j)F_{N_s+1}(X_s^k) - X_s^j X_s^k) + O\left(\frac{(t-s)^2}{h^2}\right) \\
=&\frac{t-s}{h}\mathbb{E}^*(\tilde{F}_{N_s+1}(X_s^j)X_s^k + \tilde{F}_{N_s+1}(X_s^k)X_s^j + \tilde{F}_{N_s+1}(X_s^k)\tilde{F}_{N_s+1}(X_s^j)) \\
&+ O\left(\frac{(t-s)^2}{h^2}\right) \\
=&\int_s^t b_h(r, X_s^j)dr X_s^k + \int_s^t b_h(r, X_s^k)dr X_s^j \\
&+ \frac{(t-s)}{h}\mathbb{E}^*(\tilde{F}_{N_s+1}(X_s^k)\tilde{F}_{N_s+1}(X_s^j)) + O\left(\frac{(t-s)^2}{h^2}\right) \\
=&\mathbb{E}\left(\int_s^t b(r, X_r^j)X_r^k + b(r, X_r^k)X_r^j dr \mid \mathcal{F}\right) + E_1 \\
&+ \frac{(t-s)}{h}\mathbb{E}^*(\tilde{F}_{N_s+1}(X_s^k)\tilde{F}_{N_s+1}(X_s^j)) + O\left(\frac{(t-s)^2}{h^2}\right).
\end{aligned}$$

Where we have (by the same method used to bound E_3 in Proposition 5.1)

$$|E_1| = O\left((t-s)^2\right)$$

and provided $|X_s^j - X_s^k| \geq \lambda_h$ (distance considered modulo one) we have

$$\left|\frac{(t-s)}{h}\mathbb{E}^*(\tilde{F}_{N_s+1}(X_s^k)\tilde{F}_{N_s+1}(X_s^j))\right| \leq (t-s)\lambda_h.$$

So for $(t-s)^{\frac{1}{2}} \ll h \ll 1$ we have

$$\begin{aligned}
&\mathbb{E}\left(X_t^j X_t^k - X_s^j X_s^k \middle| \mathcal{F}_s, |X_s^j - X_s^k| \geq \lambda_h\right) \\
=&\int_s^t b(r, X_r^j)X_r^k + b(r, X_r^k)X_r^j dr + o(t-s).
\end{aligned}$$

Hence, breaking $[s_j \vee s_k, \infty)$ into intervals of length $t-s$ and taking the

limit as $t-s$ and h go to 0, gives that the following process, stopped at time T^{jk} is a martingale.

$$Z_t^j Z_t^k - \int_{s_j \vee s_k}^t Z_s^k b(s, Z_s^j) + Z_s^j b(s, Z_s^k) ds.$$

Further, this process must be continuous because Proposition 5.1 tells us that Z_t^j and Z_t^k are continuous. We know from Proposition 5.1 that, under μ, both $(Z_t^k - \int_{s_k}^t b(Z_s^k)ds)_{t \geq s_k}$ and

$$\left((Z_t^k)^2 - 2\int_{s_k}^t Z_s^k b(s, Z_s^k) ds - \int_{s_k}^t a(s, Z_s^k) ds \right)_{t \geq s_k}$$

are continuous local martingales.

It remains to show that $Z_t^j - Z_t^k$ is constant for $t \geq T^{jk}$ after which the result follows from Proposition 2.1. Let $Y_t = Z_t^j - Z_t^k$ and assume w.l.o.g that $Y_0 > 0$ and $Y_{T^{jk}} = 0$. The process Y inherits the property of not changing sign as our disturbances are order preserving. Given $R \in \mathbb{R}$ and $\epsilon > 0$ localize Y using the stopping time $S = \inf\{t : Y_t > 1 \text{ or } t > R\}$ and note that:

$$\mathbb{E}|Y^S_{T^{jk}+t}| \leq \int_{T^{jk}}^{T^{jk}+t} \mathbb{E}L|Y_s^S| ds = L \int_{T^{jk}}^{T^{jk}+t} \mathbb{E}|Y_s^S| ds$$

where L is the Lipschitz constant of b. So, by Gronwall's inequality, $\mathbb{E}|Y^S_{T^{jk}+t}|$ is identically 0, up to time $t = R$. So $Y_t = 0$ for all $t > T^{jk}$ a.s. and we are done. □

Let $E = (e_k : k \in \mathbb{N})$ be an enumeration of $\mathbb{Q}^2$. Write $Z^{E,\pm}$ for the maps $D^\circ(\mathbb{R}, \mathcal{D}) \to D_E$ given by $Z^{E,\pm} = (Z^{e_k,\pm} : k \in \mathbb{N})$. Write $Z^E = Z^{E,+}$. The following result is a criterion for weak convergence on $D^\circ(\mathbb{R}, \mathcal{D})$, and is Theorem 5.1 of [24].

Theorem 5.3. *Let $(\mu_n : n \in \mathbb{N})$ be a sequence of Borel probability measures*

on $D^\circ(\mathbb{R}, \mathcal{D})$, and let μ be a Borel probability measure on $C^\circ(\mathbb{R}, \mathcal{D})$. Assume that $Z^{E,-} = Z^{E,+}$ holds μ_n-almost surely for all n and μ-almost surely. Assume further that $\mu_n \circ (Z^E)^{-1} \to \mu \circ (Z^E)^{-1}$ weakly on D_E. Then $\mu_n \to \mu$ weakly on $D^\circ(\mathbb{R}, \mathcal{D})$.

The following result is immediate from Proposition 5.2 and Theorem 5.3.

Theorem 5.4. *Given a family of distributions F along with a, b, Lipschitz in space measurable in time, obeying equations* (5)-(7) *and with $\lambda_h \to 0$ then the following convergence holds.*

$$\mu_A^F \to \mu_A^{a,b} \text{ weakly on } D^\circ(\mathbb{R}, \mathcal{D}) \text{ as } h \to 0.$$

In English, this theorem says: if the disturbances defining a sequence of disturbance flows converge nicely, then those disturbance flows converge to a specified continuous stochastic flow.

6 Proof of Theorem 4.1 using Disturbance Flows

In this section, we identify the time-reversal of a generic disturbance flow. We then apply this identification to an explicit sequence of flows and, as the limit of the reversals must be the reversal of the limit, we can deduce Theorem 4.1.

The following proposition is a generalization of the first half of Proposition 7.1 of [24], which can be recovered by assuming that $b_h \equiv 0$ and $a_h \equiv 1$.

Proposition 6.1. *Set $G_{h,t} = F_{h,-t}^{-1}$. The time-reversal of a disturbance flow with disturbance F_h is a disturbance flow with disturbance G_h, for all h. Thus $\hat{\mu}_A^{F_h} = \mu_A^{G_h}$, for all h.*

Proof. The proof is very close to the second half of the proof of proposition 7.1 of [24].

Set m and n to be the minimal and maximal values taken by N_t at jumps in I. Also, take $-\hat{n}$ and $-\hat{m}$ to be the minimal and maximal values taken by N_t at jumps in $-I$. Then, we can define a disturbance flow Φ with disturbance F_h, by

$$\Phi_I^{\pm} = F_{h,t_n}^{\pm} \circ \cdots \circ F_{h,t_m}^{\pm}.$$

Then

$$\hat{\Phi}_I^{\pm} = G_{h,-t_{-\hat{n}}}^{\pm} \circ \cdots \circ G_{h,-t_{-\hat{m}}}^{\pm}.$$

By the properties of the Poisson process $(-t_{-\hat{m}}, \ldots, -t_{-\hat{n}})$ is equal in distribution to $(t_m, \ldots, t_n)$, so $\hat{\Phi}$ is a disturbance flow with disturbance G_h. □

In [24], it is then shown for $a \equiv 1$ and $b \equiv 0$ that μ_A is invariant under time-reversal. We generalize this result to Theorem 4.1.

Theorem 4.1. *If a has spatial derivative a' and a, b and a' are uniformly bounded on compacts in time and Lipschitz in space then*

$$\hat{\mu}_A = \nu_A := \mu_A^{a^\nu, b^\nu}$$

where $a^\nu(t,x) = a(-t,x)$, $b^\nu(t,x) = -b(-t,x) + a'(-t,x)/2$ and $\hat{\mu}_A$ is the time reversal of μ_A.

Proof. The proof is based on the fact that given a family $(F_h)_{h>0}$ (satisfying the conditions of Proposition 5.1) we have that: $\mu_A^{F_{h,-t}^{-1}} = \hat{\mu}_A^{F_h} \to \hat{\mu}_A$. It thus suffices to show for some specific family $(F_h)_{h>0}$ that $\mu_A^{F_{h,-t}^{-1}} \to \nu_A$. This is true by Theorem 5.4 if $(F_{h,-t}^{-1})_{h>0}$ satisfies the conditions that we put on F, but with $a(t,x)$ and $b(t,x)$ replaced by $a^\nu(-t,x)$ and $b^\nu(-t,x)$. Let $\hat{a}_h$, $\hat{a}$, $\hat{b}_h$ and $\hat{b}$ be defined from F^{-1} as a_h and b_h are defined from F.

For every fixed h, consider the sequence t_n, and let θ_{h,t_n} be i.i.d. uniform random variables on $[0,1]$. We will write θ as shorthand for θ_{h,t_n}, and for the

remainder of this proof t will refer to an element of $\{t_n : n \in \mathbb{N}\}$. Let

$$r_{\theta,t} = \frac{h^{\frac{2}{3}}}{2}\left(b\left(t, \theta - \frac{1}{2}\right) - a'\left(t, \theta - \frac{1}{2}\right)\right)$$

and

$$w = \left(\frac{3a(t,\theta)h}{2}\right)^{\frac{1}{3}}.$$

Then, for sufficiently small h, we consider the family of disturbances given by setting,

$$F_{h,t}(x) = \begin{cases} x + r_{\theta,t} & (x-\theta) \in (\frac{1}{2} - h^{\frac{1}{3}}, \frac{1}{2} + h^{\frac{1}{3}}) \\ \frac{1}{2} + h^{\frac{1}{3}} + r_{\theta,t} + \theta & (x-\theta) \in (\frac{1}{2} + h^{\frac{1}{3}}, \frac{1}{2} + h^{\frac{1}{3}} + r_{\theta,t}) \\ \frac{1}{2} - h^{\frac{1}{3}} + r_{\theta,t} + \theta & (x-\theta) \in (\frac{1}{2} - h^{\frac{1}{3}} + r_{\theta,t}, \frac{1}{2} - h^{\frac{1}{3}}) \\ \theta & (x-\theta) \in (-w, w) \\ x & \text{otherwise.} \end{cases}$$

An example from this family is graphed in Figure 2.

Note that $\lambda \to 0$ for both F and F^{-1} (The disturbance of size $r_{\theta,t}$ is negligible in computing λ as it is $O(h^{\frac{2}{3}})$ in magnitude $O(h^{\frac{1}{3}})$ in width and always multiplied by something of size $O(h^{\frac{1}{3}})$ in the definition of λ). The first three cases in the above definition also contribute nothing to either $\lim_{h\to 0} a_h$ or $\lim_{h\to 0} \hat{a}_h$, and their contribution to $\lim_{h\to 0} b_h$ is exactly the negative of their contribution to $\lim_{h\to 0} \hat{b}_h$. So it suffices to prove that the proposition holds for the case $b = a'$, i.e. the case where $r_{\theta,t} \equiv 0$.

We write $w_\pm$ for the largest offsets from x a disturbance can have whilst not mapping x to itself. For sufficiently small h they are given by the following implicit equation:

$$w_\pm = \left(\frac{3a(t, x \pm w_\pm)h}{2}\right)^{\frac{1}{3}}.$$

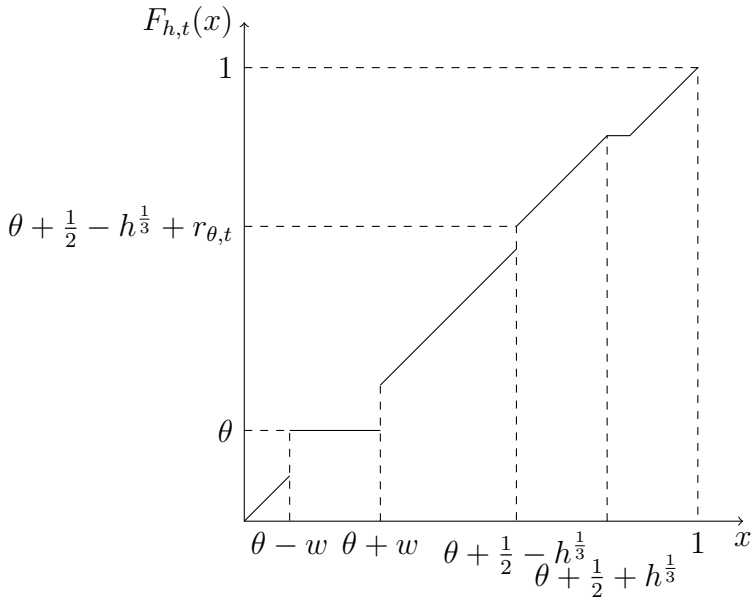


Figure 2: An example from the specific family of disturbances used in this proof.

Expanding this by substitution and Taylor's theorem, and letting

$$c = c_h(x) = \left(\frac{3a(t,x)h}{2}\right)^{\frac{1}{3}}.$$

gives

$$w_\pm = c \pm \frac{a'c^2}{3a} + o(h^{\frac{2}{3}}).$$

where unless otherwise specified a and a' are evaluated at (t, x). We can now

calculate

$$\begin{aligned} a_h &= \frac{1}{h}\mathbb{E}\left(\tilde{F}_{h,t}(x)^2\right) \\ &= \frac{1}{h}\int_{-w_-}^{w_+} \alpha^2 d\alpha \\ &= \frac{1}{3h}\left(w_+^3 + w_-^3\right) \\ &= \frac{2c^3}{3h} + o(1) \\ &\to a \end{aligned}$$

and

$$\begin{aligned} b_h &= \frac{1}{h}\mathbb{E}\left(\tilde{F}_{h,t}(x)\right) \\ &= \frac{1}{h}\int_{-w_-}^{w_+} \alpha d\alpha \\ &= \frac{1}{2h}\left(w_+^2 - w_-^2\right) \\ &= \frac{1}{2h}\left(\left(c^2 + \frac{2a'c^3}{3a} + o(h)\right) - \left(c^2 - \frac{2a'c^3}{3a} + o(h)\right)\right) \\ &= \frac{2a'c^3}{3ha} + o(1) \\ &\to a'. \end{aligned}$$

By Taylor and binomial expansion we also get

$$c_h(x+\alpha) = c_h(x)\left(1 + \frac{a'\alpha}{3a}\right) + o(h^{\frac{1}{3}}\alpha).$$

Which allows us to calculate,

$$\begin{aligned}
\hat{a}_h(-t,x) &= \frac{1}{h}\int_{-w_-}^{0}(\alpha + c_h(x+\alpha))^2 d\alpha + \frac{1}{h}\int_0^{w_+}(\alpha - c_h(x+\alpha))^2 d\alpha \\
&= \frac{1}{h}\int_{-c}^{0}\alpha^2 + 2\alpha c + c^2 d\alpha + \frac{1}{h}\int_0^{c}\alpha^2 - 2\alpha c + c^2 d\alpha + o(1) \\
&= \frac{2}{h}\int_0^{c}\alpha^2 - 2\alpha c + c^2 d\alpha + o(1) \\
&= \frac{2}{h}(\frac{c^3}{3} - c^3 + c^3) + o(1) \\
&\to a
\end{aligned}$$

and

$$\begin{aligned}
\hat{b}_h(-t,x) &= \frac{1}{h}\int_{-w_-}^{0}\alpha + c_h(x+\alpha) d\alpha + \frac{1}{h}\int_0^{w_+}\alpha - c_h(x+\alpha) d\alpha \\
&= b_h + \frac{c}{h}\int_{-w_-}^{0} 1 + \frac{a'\alpha}{3a} + o(\alpha) d\alpha - \frac{c}{h}\int_0^{w_+} 1 + \frac{a'\alpha}{3a} + o(\alpha) d\alpha \\
&= b_h + \frac{c}{h}\left(w_- - w_+ - \frac{a'}{6a}(w_-^2 + w_+^2) + o(h^{\frac{2}{3}})\right) \\
&= b_h + \frac{c}{h}\left(-\frac{2a'c^2}{3a} - \frac{a'c^2}{3a}\right) + o(1) \\
&= b_h - a' - \frac{a'}{2} + o(1) \\
&\to -\frac{a'}{2}.
\end{aligned}$$

So the result holds. □

The following corollary is similar to Corollary 7.3 of [24] (and with an almost identical proof) in that it gives weak convergence for paths running both forward and backward from a given sequence of points. First we define the notation for this result.

Given $e = (s,x) \in \mathbb{R}^2$, define $\bar{D}_e = \{\xi \in D(\mathbb{R},\mathbb{R}) : \xi_s = x\}$ and for

$E = (e_k : k \in \mathbb{N})$ set $\bar{D}_E = \prod_{k=1}^{\infty} \bar{D}_{e_k}$. For $\phi \in D^\circ(\mathbb{R}, \mathcal{D})$, define

$$\bar{Z}_t^{e,\pm}(\phi) = \begin{cases} \phi^{\pm}_{(s,t]}(x), & t \geq s, \\ (\phi^{-1})^{\pm}_{(t,s]}(x), & t < s. \end{cases}$$

Then $\bar{Z}^{e,\pm}(\phi) \in \bar{D}_e$ and extends $Z^{e,\pm}(\phi)$, from $[s, \infty)$ to the whole of $\mathbb{R}$. For all $e \in \mathbb{R}^2$, we have $\bar{Z}^{e,-}$ almost everywhere on $D^\circ(\mathbb{R}, \mathcal{D})$ for both μ_A and μ_A^f, for every disturbance function f. So we drop the $\pm$. Denote by $\bar{\mu}_E^f$ the law of $(\bar{Z}^{e_k} : k \in \mathbb{N})$ on $\bar{D}_E$ under μ_A^f and by $\bar{\mu}_E^{a,b}$ the corresponding law under $\mu_A^{a,b}$

Corollary 6.2. *$\bar{\mu}_E^{F_h} \to \bar{\mu}_E^{a,b}$ weakly on $\bar{D}_E$,*

Proof. Given ϕ with law $\mu_A^{a,b}$, we have that almost surely

$$\bar{Z}^{(s,x\pm\delta),+}(\phi) \to \bar{Z}^{(s,x),+}(\phi)$$

uniformly on $\mathbb{R}$ as $\delta \to 0$. We also have $\phi \in C^\circ(\mathbb{R}, \mathcal{D})$ almost surely and it follows that $\bar{Z}^{(s,x),+}$ is continuous at ϕ almost surely. Thus, the result holds as we already know the convergence holds component wise. □

7 Proof of Theorem 4.1 without Disturbance Flows

In this section we first prove a version of Theorem 4.1 with the extra hypothesis that a and b are Lipschitz in time. Then we use an approximation argument to show Theorem 4.1 in the general case.

Theorem 7.1. *If a has spatial derivative a' and a, b and a' are Lipschitz in both time and space then*

$$\hat{\mu}_A = \nu_A := \mu_A^{a^\nu, b^\nu}$$

where $a^\nu(t,x) = a(-t,x)$, $b^\nu(t,x) = -b(-t,x) + a'(-t,x)/2$ and $\hat{\mu}_A$ is the time reversal of μ_A.

Proof. Let $\phi \sim \mu_A$. It suffices to show that the restriction of $\hat{\phi}$ to E given by $Z^{E,+}(\hat{\phi})$, which we shall call $\hat{\phi}_E$, has distribution ν_E, for each countable set $E \subset \mathbb{R} \times \mathbb{R}$. The distribution ν_E is characterised by its restriction to two point motions by Theorem 3.1.

Coalescence of two motions follows immediately from the definition of time-reversal. As does the continuity of a single motion.

As ϕ_{ts} and ϕ_{su} are independent for $s \in (u,t)$, we have the Markov property. Thus, by Donsker's Invariance Principle, we can identify the two point motion from just the mean and covariance matrix of small increments.

First, we consider each one point motion separately. We will proceed by relating the backward and forward flows, then, noting that increments of the forward process are small, we approximate a and b on an interval that the forward process almost surely won't leave in such a way as to make exact calculations possible. Then we check that the incurred error is small using that a and b are Lipschitz in time, and that the exact calculations give the required answer. Finally, we will show that the increments of each process are independent, conditional on an event of large probability, and so the covariances are small.

We have the relation,

$$\mathbb{P}\left(\hat{\phi}_{t+h,t}(y) < x\right) = \mathbb{P}(\phi_{-t,-t-h}(x) > y)$$

which we can use to determine the distribution of $\hat{\phi}_{t+h,t}(y)$ if we first understand the distributions of the variables $\phi_{-t,-t-h}(x)$.

To study these variables, we first show that the forward paths are localised.

$$\mathbb{P}\left(\sup_{0<\delta t<h}|\phi_{t+\delta t,t}(x)-x|>h^{\frac{1}{2}-\epsilon}\right)\leq\mathbb{P}\left(\sup_{0<\delta t<h}\phi_{t+\delta t,t}(x)-x>h^{\frac{1}{2}-\epsilon}\right)+\mathbb{P}\left(\inf_{0<\delta t<h}\phi_{t+\delta t,t}(x)-x<-h^{\frac{1}{2}-\epsilon}\right)$$

Each of these terms can be bounded in the same way. To bound the first term, consider the process $\phi_{t+\delta t,t}(x)-b^*\delta t$ parametrised by δt. This is a supermartingale with diffusivity bounded by a^*, and thus by the reflection principle

$$\mathbb{P}\left(\sup_{0<\delta t<h}\phi_{t+\delta t,t}(x)-x>h^{\frac{1}{2}-\epsilon}\right)\leq 2\Phi\left(-\frac{h^{\frac{1}{2}-\epsilon}-b^*h}{2(a^*h)^{\frac{1}{2}}}\right).$$

Thus we can derive that

$$\begin{aligned}&\mathbb{P}\left(\sup_{0<\delta t<h}|\phi_{t+\delta t,t}(x)-x|>h^{\frac{1}{2}-\epsilon}\right)\\ \leq&4\Phi\left(-\frac{h^{\frac{1}{2}-\epsilon}-b^*h}{2(a^*h)^{\frac{1}{2}}}\right)\\ \leq&\exp\left(-C(A,B)h^{-2\epsilon}\right)\qquad\text{for sufficiently small } h\end{aligned}$$

where C is positive and independent of h and t.

Now we approximate a and b by $\tilde{a}$ and $\tilde{b}$ which, on the interval $[y-2h^{\frac{1}{2}-\epsilon},y+2h^{\frac{1}{2}-\epsilon}]$, are given by

$$\tilde{a}(s,x)=\frac{a'^2}{4a}\left(x-y+\frac{2a}{a'}\right)^2$$

and

$$\tilde{b}(s,x)=\frac{ba'}{2a}\left(x-y+\frac{2a}{a'}\right).$$

Where we have written a for $a(t,y)$, a' for $a'(t,y)$ and b for $b'(t,y)$. We then extend $\tilde{a}$ and $\tilde{b}$ to $\tilde{L}$-Lipschitz and $\tilde{L}$-Lipschitz differentiable functions on the circle for some $\tilde{L}$. For sufficiently small values of h, this extension can and will be chosen so that $a_* \leq \tilde{a} \leq a^*$. Note that for all s, $a(t,y) = \tilde{a}(s,y)$, $a'(t,y) = \tilde{a}'(s,y)$ and $b(t,y) = \tilde{b}(s,y)$, this will turn out to make them sufficiently good approximations.

We now approximate the diffusion process $\phi_{t+\delta t,t}(x)$ for each $x \in [y - h^{\frac{1}{2}-\epsilon}, y + h^{\frac{1}{2}-\epsilon}]$ by a diffusion process $X_{\delta t}$ started from x, with drift $\tilde{b}$ and diffusivity $\tilde{a}$, but driven by the same Brownian motion $B_{\delta t}$ as $\phi_{t+\delta t,t}(x)$. Note that $\tilde{a}$ and $\tilde{b}$ are constant with respect to time. Let G be the event

$$\left\{ \sup_{0<\delta t<h} |\phi_{t+\delta t,t}(x) - x| < h^{\frac{1}{2}-\epsilon} \right\} \cap \left\{ \sup_{0<\delta t<h} |X_{\delta t} - x| < h^{\frac{1}{2}-\epsilon} \right\}$$

and note the second event in this union has probability bounded like the first, so $\mathbb{P}(G) = 1 - O\left(e^{-Ch^{-2\epsilon}}\right)$. Note also that on this event, X and $\phi_{t+\delta t,t}(x)$ stay within the interval we explicitly defined $\tilde{a}$ and $\tilde{b}$ on.

On this event the error in the approximation is given by

$$\begin{aligned}
\Delta_{\delta t} &:= X_{\delta t} - \phi_{t+\delta t,t}(x) \\
&= \int_0^{\delta t} \tilde{b}(t, X_u) - b(t+u, \phi_{t+u,t}(x)) du \\
&\quad + \int_0^{\delta t} \sqrt{\tilde{a}(t, X_u)} - \sqrt{a(t+u, \phi_{t+u,t}(x))} dB_u.
\end{aligned}$$

We have that if $E_h := \sup_{\delta t<h} |\Delta_{\delta t}|$ then

$$\begin{aligned} E_h \leq \sup_{\delta t<h} \int_0^{\delta t} & |\tilde{b}(t, X_u) - \tilde{b}(t, \phi_{t+u,t}(x))| + |\tilde{b}(t, \phi_{t+u,t}(x)) - b(t, y)| \\ & + |b(t, y) - b(t, \phi_{t+u,t}(x))| \\ & + |b(t, \phi_{t+u,t}(x)) - b(t+u, \phi_{t+u,t}(x))| du \\ + \left| \int_0^{\delta t} \right. & \left(\sqrt{\tilde{a}(t, X_u)} - \sqrt{\tilde{a}(t, \phi_{t+u,t}(x))} + \sqrt{\tilde{a}(t, \phi_{t+u,t}(x))} - \sqrt{a(t, y)} \right. \\ & + \sqrt{a(t, y)} - \sqrt{a(t, \phi_{t+u,t}(x))} \\ & \left. \left. + \sqrt{a(t, \phi_{t+u,t}(x))} - \sqrt{a(t+u, \phi_{t+u,t}(x))} \right) dB_u \right|. \end{aligned}$$

The first integrand is bounded, on G, by $2\tilde{L}h^{\frac{1}{2}-\epsilon} + 2Lh^{\frac{1}{2}-\epsilon} + 2\tilde{L}h^{\frac{1}{2}-\epsilon} + Lh \leq 4(L+\tilde{L})h^{\frac{1}{2}-\epsilon}$. To bound the second integrand we first observe that the square root funciton is Lipschitz with some constant L_s on the interval $[a_*, a^*]$. Secondly, note that we can achieve a stronger bound than for the first integrand as a' is Lipschitz in x. The second integrand is thus bounded, on G, by $d_1 := 2\tilde{L}h^{\frac{1}{2}-\epsilon} + 2Lh^{1-2\epsilon} + 2\tilde{L}h^{1-2\epsilon} + Lh$. The first integral is therefore bounded by $4\delta t(L + \tilde{L})h^{\frac{1}{2}-\epsilon}$. The second integral is a martingale with diffusivity bounded by d_1. Therefore it has absolute value stochastically dominated by that of a martingale with diffusivity d_1. Equivalently, there exists a Brownian motion B' such that, on the event G, $E_h \leq \sup_{\delta t<h} |\Delta'_{\delta t}|$ where

$$\Delta'_{\delta t} = 4\delta t(L + \tilde{L})h^{\frac{1}{2}-\epsilon} + \left| \int_0^{\delta t} d_1 dB'_u \right|.$$

Consider the event $G' = \{\sup_{\delta t<h} |\Delta'_{\delta t}| < h^{1-2\epsilon}\}$. The probability of this event is $1 - O(e^{-Ch^{-2\epsilon}})$. This isn't quite a strong enough bound due to the $h^{\frac{1}{2}-\epsilon}$ term in d_1. However, as that term is proportional to the bound we have on $X_u - \phi_{t+u,t}(x)$ and this result provides a stronger bound, we can

bootstrap this argument. To that effect note that, on $G \cap G'$, the second integrand is bounded by $2\tilde{L}h^{1-2\epsilon} + 2(L+\tilde{L})h^{1-2\epsilon} + Lh \leq 4(L+\tilde{L})h^{1-2\epsilon}$. Thus, as before, there exists a Brownian motion B'' such that, on the event $G \cap G'$, $E_{\delta t} \leq \sup_{v<\delta t} |\Delta''_v|$ where

$$\Delta''_{\delta t} = 4\delta t(L+\tilde{L})h^{\frac{1}{2}-\epsilon} + \left| \int_0^{\delta t} 4(L+\tilde{L})h^{1-2\epsilon} dB''_u \right|.$$

Finally consider $G'' = \{\sup_{v<\delta t} |\Delta''_v| < h^{\frac{3}{2}-3\epsilon}\}$ and note that the probability of this event, conditioned on $G \cap G'$ is $1 - O(e^{-Ch^{-2\epsilon}})$. Thus we can conclude that

$$\mathbb{P}(|\Delta_h| > h^{\frac{3}{2}-3\epsilon}) \leq 1 - \mathbb{P}(G \cap G' \cap G'') = O\left(e^{-Ch^{-2\epsilon}}\right).$$

This result suffices to control the error of the approximation.

Next, we calculate the distribution of X_h. Note that on the event G, we have, for some Brownian motion W, that

$$dX_t = \frac{ba'}{2a}\left(X_t - y + \frac{2a}{a'}\right)dt + \frac{a'}{2\sqrt{a}}\left(X_t - y + \frac{2a}{a'}\right)dW_t.$$

Where we are again writing a for $a(t,y)$, a' for $a'(t,y)$ and b for $b'(t,y)$. Define $f(x) = \frac{2\sqrt{a}}{a'}\log(x - y + \frac{2a}{a'})$. An application of Itō's lemma gives that

$$df(X_t) = \frac{1}{\sqrt{a}}\left(b - \frac{a'}{4}\right)dt + dW_t.$$

The choices for $\tilde{a}$, $\tilde{b}$ and f were made so that this equation has constant coefficients. Thus $f(X_h)$ is normally distributed with mean $f(x) + \frac{h}{\sqrt{a}}\left(b - \frac{a'}{4}\right)$ and variance h. So we can calculate that $\mathbb{P}(X_h > y) = \mathbb{P}(f(X_h) > f(y)) =$

$F_y(x) + O(e^{-Ch^{-2\epsilon}})$, where

$$\begin{aligned} F_y(x) &:= \Phi\left(\frac{f(x)-f(y)}{\sqrt{h}} + \sqrt{\frac{h}{a}}\left(b - \frac{a'}{4}\right)\right) \\ &= \Phi\left(\frac{2\sqrt{a}}{a'\sqrt{h}}\log\left(1 + \frac{a'}{2a}(x-y)\right) + \sqrt{\frac{h}{a}}\left(b - \frac{a'}{4}\right)\right). \end{aligned}$$

This implies that

$$\mathbb{P}(|X_h - x| > h^{\frac{1}{2}-\epsilon}) = O(e^{-Ch^{-2\epsilon}}).$$

We can then calculate for $y = 0$ and $|x| < h^{\frac{1}{2}-\epsilon}$ that

$$\begin{aligned} F_0'(x) &= \left(\frac{1}{\sqrt{2ah\pi}}\frac{1}{1+\frac{a'x}{2a}}\right) \\ &\quad \times \exp\left(-\frac{1}{2}\left(\frac{2\sqrt{a}}{a'\sqrt{h}}\log\left(1+\frac{a'x}{2a}\right) + \sqrt{\frac{h}{a}}\left(b-\frac{a'}{4}\right)\right)^2\right) \\ &= \frac{1}{\sqrt{2ah\pi}}\left(1 - \frac{a'x}{2a} + O\left(h^{1-2\epsilon}\right)\right)e^{-\frac{x^2}{2ah}} \\ &\quad \times \exp\left(-\frac{x}{\sqrt{ah}}\left(-\frac{a'x^2}{4\sqrt{a^3h}} + \sqrt{\frac{h}{a}}\left(b-\frac{a'}{4}\right)\right) + O\left(h^{1-4\epsilon}\right)\right) \\ &= \frac{1}{\sqrt{2ah\pi}}\left(1 - \frac{a'x}{2a} + O\left(h^{1-2\epsilon}\right)\right)e^{-\frac{x^2}{2ah}} \\ &\quad \times \left(1 - \frac{x}{\sqrt{ah}}\left(-\frac{a'x^2}{4\sqrt{a^3h}} + \sqrt{\frac{h}{a}}\left(b-\frac{a'}{4}\right)\right) + O\left(h^{1-6\epsilon}\right)\right) \\ &= \frac{e^{-\frac{x^2}{2ah}}}{\sqrt{2ah\pi}}\left(1 - \frac{x}{a}\left(b - \frac{a'}{4} + \frac{a'}{2}\right) + \frac{a'x^3}{4a^2h} + O(h^{1-6\epsilon})\right). \end{aligned}$$

This is related to $\hat{\phi}$ by

$$\begin{aligned}\mathbb{P}\left(\hat{\phi}_{t+h,t}(y) < x\right) &= \mathbb{P}(\phi_{-t,-t-h}(x) > y) \\ &= F_{y+O\left(h^{\frac{3}{2}-3\epsilon}\right)}(x) + O\left(e^{-Ch^{-2\epsilon}}\right) \\ &= F_y(x) + O\left(h^{\frac{3}{2}-3\epsilon} \sup_{y\in\mathbb{R}} \frac{dF_y(x)}{dy}\right) + O\left(e^{-Ch^{-2\epsilon}}\right)\end{aligned}$$

and on $\left[y - h^{\frac{1}{2}-\epsilon}, y + h^{\frac{1}{2}-\epsilon}\right]$ this is equal to

$$F_y(x) + O(h^{1-3\epsilon}).$$

Finally, we use this to compute,

$$
\begin{aligned}
&\mathbb{E}(\hat{\phi}_{t+h,t}(y)) \\
=&y + \int_y^\infty 1 - \mathbb{P}\left(\hat{\phi}_{t+h,t}(y) < x\right) dx - \int_{-\infty}^y \mathbb{P}\left(\hat{\phi}_{t+h,t}(y) < x\right) dx \\
=&y + \int_y^{y+h^{\frac{1}{2}-\epsilon}} 1 - \mathbb{P}\left(\hat{\phi}_{t+h,t}(y) < x\right) dx \\
&- \int_{y-h^{\frac{1}{2}-\epsilon}}^y \mathbb{P}\left(\hat{\phi}_{t+h,t}(y) < x\right) dx + O\left(e^{-Ch^{-2\epsilon}}\right) \\
=&y + \int_y^{y+h^{\frac{1}{2}-\epsilon}} 1 - F_y(x)dx - \int_{y-h^{\frac{1}{2}-\epsilon}}^y F_y(x)dx + O\left(h^{\frac{3}{2}-4\epsilon}\right) \\
=&y + h^{\frac{1}{2}-\epsilon} - \int_{y-h^{\frac{1}{2}-\epsilon}}^{y+h^{\frac{1}{2}-\epsilon}} F_y(x)dx + O\left(h^{\frac{3}{2}-4\epsilon}\right) \\
=&y + h^{\frac{1}{2}-\epsilon} - \int_{-h^{\frac{1}{2}-\epsilon}}^{h^{\frac{1}{2}-\epsilon}} F_0(x)dx + O\left(h^{\frac{3}{2}-4\epsilon}\right) \\
=&y + h^{\frac{1}{2}-\epsilon} - h^{\frac{1}{2}-\epsilon}(F_0(h^{\frac{1}{2}-\epsilon}) - F_0(-h^{\frac{1}{2}-\epsilon})) + \int_{-h^{\frac{1}{2}-\epsilon}}^{h^{\frac{1}{2}-\epsilon}} xF_0'(x)dx + O\left(h^{\frac{3}{2}-4\epsilon}\right)
\end{aligned}
$$

$$
\begin{aligned}
=&y + \int_{-h^{\frac{1}{2}-\epsilon}}^{h^{\frac{1}{2}-\epsilon}} xF_0'(x)dx + O\left(h^{\frac{3}{2}-4\epsilon}\right) \\
=&y + \int_{-h^{\frac{1}{2}-\epsilon}}^{h^{\frac{1}{2}-\epsilon}} \left(x - \frac{x^2}{a}\left(b + \frac{a'}{4}\right) + \frac{a'x^4}{4a^2h}\right) \frac{e^{-\frac{x^2}{2ah}}}{\sqrt{2ah\pi}} dx + O\left(h^{\frac{3}{2}-7\epsilon}\right) \\
=&y + \int_{-\infty}^{\infty} \left(x - \frac{x^2}{a}\left(b + \frac{a'}{4}\right) + \frac{a'x^4}{4a^2h}\right) \frac{e^{-\frac{x^2}{2ah}}}{\sqrt{2ah\pi}} dx + O\left(h^{\frac{3}{2}-7\epsilon}\right) \\
=&y - h\left(b + \frac{a'}{4}\right) + \frac{3ha'}{4} + O\left(h^{\frac{3}{2}-7\epsilon}\right) \\
=&y + h\left(-b + \frac{a'}{2}\right) + O(h^{\frac{3}{2}-7\epsilon}).
\end{aligned}
$$

Similarly, we find that

$$\mathrm{Var}\left(\hat{\phi}_{t+h,t}(y)\right) = ah + O(h^{2-8\epsilon}).$$

Thus, the single point motions are diffusion processes with the required drift and diffusivity.

Next, we will show that the motions started from y_1 and y_2 have zero covariation until they coalesce, and thus, are independent until they coalesce. This follows immediately from the fact that for $y_1 \neq y_2$

$$\mathrm{Cov}\left(\hat{\phi}_{t+h,t}(y_1), \hat{\phi}_{t+h,t}(y_2)\right) = o(h).$$

To establish this fact consider the events

$$\mathcal{A}_i = \left\{ \sup_{0<\delta t<h} \left|\hat{\phi}_{t+\delta t,t}(y_i) - y_i\right| < \frac{|y_2 - y_1|}{2} \right\} \qquad \text{for } i = 1, 2.$$

On the intersection of these events, we know that the $\hat{\phi}_{t+h,t}(y_i)$ are independent as the forward flows on $[-t-h, t] \times \left[y_i - \frac{|y_2-y_1|}{2}, y_i + \frac{|y_2-y_1|}{2}\right]$ are independent, and each determines the corresponding $\mathcal{A}_i$ and $\hat{\phi}_{t+h,t}(y_i)$. Thus, writing B for the complement of $\mathcal{A}_1 \cap \mathcal{A}_2$,

$$\begin{aligned}
\left|\mathrm{Cov}\left(\hat{\phi}_{t+h,t}(y_1), \hat{\phi}_{t+h,t}(y_2)\right)\right| &= \left|\mathrm{Cov}\left(\mathbb{1}_B\hat{\phi}_{t+h,t}(y_1), \mathbb{1}_B\hat{\phi}_{t+h,t}(y_2)\right)\right| \\
&\leq \sqrt{\mathrm{Var}\left(\mathbb{1}_B\hat{\phi}_{t+h,t}(y_2)\right)\mathrm{Var}\left(\mathbb{1}_B\hat{\phi}_{t+h,t}(y_1)\right)} \\
&\leq \max_{i=1,2}\left\{\mathrm{Var}\left(\mathbb{1}_B\hat{\phi}_{t+h,t}(y_i)\right)\right\} \\
&\leq \max_{i=1,2}\left\{\mathbb{E}\left(\mathbb{1}_B\left(\hat{\phi}_{t+h,t}(y_i) - y_i\right)^2\right)\right\}
\end{aligned}$$

but, as we know that

$$\mathbb{P}\left(\left|\hat{\phi}_{t+h,t}(y_i) - y_i\right| > x\right) \leq 2\left(1 - \Phi\left(\frac{x - b^*h}{\sqrt{a^*h}}\right)\right)$$

we can deduce that

$$\begin{aligned}\left|\mathrm{Cov}\left(\hat{\phi}_{t+h,t}(y_1),\hat{\phi}_{t+h,t}(y_2)\right)\right| &\leq \int_{\frac{|y_2-y_1|}{2}}^{\infty} x^2 \frac{2}{\sqrt{a^*h}}\Phi'\left(\frac{x-b^*h}{\sqrt{a^*h}}\right)dx \\ &= \int_{\frac{|y_2-y_1|-2b^*h}{2\sqrt{a^*h}}}^{\infty} x^2 2\Phi'(u)du \\ &= \sqrt{\frac{2}{a^*h\pi}}\int_{\frac{|y_2-y_1|-2b^*h}{2\sqrt{a^*h}}}^{\infty} a^*h(u+b^*h)^2 e^{-\frac{u^2}{2}}du \\ &= \sqrt{\frac{2a^*h}{\pi}}(1+O(h))\int_{\frac{|y_2-y_1|-2b^*h}{2\sqrt{a^*h}}}^{\infty} u^2 e^{-\frac{u^2}{2}}du.\end{aligned}$$

For sufficiently small h this is bounded by

$$\begin{aligned}&\sqrt{\frac{3a^*h}{\pi}}\int_{\frac{|y_2-y_1|}{3\sqrt{a^*h}}}^{\infty} u^2 e^{-\frac{u^2}{2}}du \\ =&\sqrt{\frac{3a^*h}{\pi}}\left(\frac{|y_2-y_1|}{3\sqrt{a^*h}}e^{-\frac{|y_2-y_1|^2}{18a^*h}}+\int_{\frac{|y_2-y_1|}{3\sqrt{a^*h}}}^{\infty} e^{-\frac{u^2}{2}}du\right) \\ =&O\left(e^{-\frac{|y_2-y_1|^2}{18a^*h}}\right)\end{aligned}$$

This establishes the result. □

Finally, we relax the restriction that a and b are Lipschitz in time.

Theorem 4.1. *If a has spatial derivative a' and a, b and a' are uniformly bounded on compacts in time and Lipschitz in space then*

$$\hat{\mu}_A = \nu_A := \mu_A^{a^\nu, b^\nu}$$

where $a^\nu(t,x) = a(-t,x)$, $b^\nu(t,x) = -b(-t,x) + a'(-t,x)/2$ and $\hat{\mu}_A$ is the time reversal of μ_A.

Proof. Define approximations a_n and b_n by

$$a_n = a * K_n \text{ and } b_n = b * K_n$$

where $*$ denotes convolution in time,

$$K_n(t) = nK(tn)$$

and K is a smooth, non-negative function supported on $[-1, 1]$ with supremum and integral equal to one. The resulting a_n and b_n are smooth. Thus, we will be able to apply Theorem 7.1 to a flow with these parameters.

Let $\phi^n \in C^\circ(\mathbb{R}, \mathcal{D})$ be the coalescing diffusive flow driven by a_n and b_n and let

$$b_k^* = \sup_{[-k-1,k+1]\times[0,1]} |b(t,x)|$$

and

$$a_k^* = \sup_{[-k-1,k+1]\times[0,1]} a(t,x).$$

We define A_N to be the subset of $\phi \in C^\circ(\mathbb{R}, \mathcal{D})$ such that for all k both

$$|\phi_{ts}(x) - x| \leq 4b_k^* k + kN\sqrt{8a_k^*} + 1 \quad \forall x \in [0,1] \quad \forall s,t \in [-k,k] \text{ with } s < t$$

and

$$|\phi_{ts}(x) - x| \leq \frac{1}{k} \quad \forall x \in [0,1] \quad \forall s,t \in [-k,k] \text{ with } t - s \in [0, \delta_{k,N}]$$

where

$$\delta_{k,N} = \min\left\{ \frac{1}{18k^3 N a_k^*(1 + a_k^* + b_k^*)}, \frac{a_k^*}{2{b_k^*}^2} \right\}.$$

In Proposition 8.2, we prove that A_N is compact; and in Proposition 8.3, we prove that $\phi^n \in A_N$ with high probability in N uniformly in n. Thus, we can deduce that the ϕ^n are tight. Let ϕ be a weak sub-sequential limit of ϕ^n.

We will show that $\phi \sim \mu_A$ and that $\hat{\phi} \sim \nu_A$, which establishes the theorem.

We present here only the proof that $\phi \sim \mu_A$. The proof that $\lim_{n\to\infty} \hat{\phi}^n \sim \nu_A$ is identical, but considering $\hat{\phi}^n$ and $-b + \frac{a'}{2}$ instead of ϕ^n and b, it then follows that $\hat{\phi} \sim \nu_A$ as time reversal is an isometry. By Theorem 3.1 it suffices to show that

$$\mathbb{E}\left(\phi_{ts}(x) - \int_s^t b(r, \phi_{rs}(x))dr\right) = x \quad \forall x \in [0,1] \quad \forall s < t$$

and

$$\mathbb{E}\left(M_{st}(x_1, x_2, b, a, \phi)\right) = x_1 x_2 \quad \forall x_1, x_2 \in [0,1] \quad \forall s < t$$

where

$$\begin{aligned} M_{st}(x_1, x_2, b, a, \phi) =& \phi_{ts}(x_1)\phi_{ts}(x_2) - \int_s^t \phi_{rs}(x_1) b(r, \phi_{rs}(x_2)) \\ &+ \phi_{rs}(x_2) b(r, \phi_{rs}(x_1))dr - \int_{T^{(s,x_1)(s,x_2)} \wedge t}^t a(r, \phi_{rs}(x_1))dr \end{aligned}$$

The proof of these two statements are very similar, so we will only provide the more complicated second one here.

Proposition 8.4 says that $\mathbb{E}\left(M_{st}(x_1, x_2, b, a, \phi)\right)$ is a continuous function of x_1 and x_2. Thus, it suffices to show that

$$\mathbb{E}_x \mathbb{E}_\phi \left(M_{st}(x_1, x_2, b, a, \phi)\right) = \mathbb{E}_x(x_1 x_2)$$

where $\mathbb{E}_x$ averages over values of x_1 and x_2 in a pair of intervals I_1 and I_2 respectively, and $\mathbb{E}_\phi$ is the same as $\mathbb{E}$ on previous lines. Proposition 8.6 says

$$\mathbb{E}_\phi \mathbb{E}_x (M_{st}(x_1, x_2, b, a, \phi)) = \lim_n \mathbb{E}_{\phi^n} \mathbb{E}_x \left(M_{st}(x_1, x_2, b, a, \phi^n)\right)$$

which is used in the calculation below. Writing D_n for $M_{st}(x_1, x_2, b, a, \phi^n) - M_{st}(x_1, x_2, b_n, a_n, \phi^n)$ we can calculate, using Proposition 7.1 in the fourth equality, that

$$
\begin{aligned}
&\mathbb{E}_x\mathbb{E}_\phi(M_{st}(x_1,x_2,b,a,\phi))\\
=&\mathbb{E}_\phi\mathbb{E}_x(M_{st}(x_1,x_2,b,a,\phi))\\
=&\lim_n \mathbb{E}_{\phi^n}\mathbb{E}_x\left(M_{st}(x_1,x_2,b,a,\phi^n)\right)\\
=&\lim_n \mathbb{E}_{\phi^n}\mathbb{E}_x\left(M_{st}(x_1,x_2,b_n,a_n,\phi^n)\right)+\lim_n \mathbb{E}_{\phi^n}\mathbb{E}_x\left(D_n\right)\\
=&\mathbb{E}_x\left(x_1x_2\right)+\lim_n \mathbb{E}_x\mathbb{E}_{\phi^n}\left(D_n\right).
\end{aligned}
$$

It remains only to show that $\mathbb{E}_{\phi^n}(D_n)$ goes to 0 uniformly in x as $n\to\infty$.

$$
\begin{aligned}
D_n=&\int_s^t \phi^n_{rs}(x_1)(b_n(r,\phi^n_{rs}(x_2))-b(r,\phi^n_{rs}(x_2)))dr\\
&+\int_s^t \phi^n_{rs}(x_2)(b_n(r,\phi^n_{rs}(x_1))-b(r,\phi^n_{rs}(x_1)))dr\\
&+\int_{T^{(s,x_1)(s,x_2)}}^t a_n(r,\phi^n_{rs}(x_1))-a(r,\phi^n_{rs}(x_1))dr
\end{aligned}
$$

Each of these terms has expectation tending to 0. We will prove this for the first term (the second term is very similar and the third term is even simpler, so the same argument works). We firstly rearrange each half of the first term separately.

$$
\begin{aligned}
&\int_s^t \phi^n_{rs}(x_1) b_n(r, \phi^n_{rs}(x_2)) dr \\
&= \int_s^t \int_{-\frac{1}{n}}^{\frac{1}{n}} \phi^n_{rs}(x_1) b(r+u, \phi^n_{rs}(x_2)) K_n(u) du dr \\
&= \int_{s-\frac{1}{n}}^{t+\frac{1}{n}} \int_{(v-t)\vee -\frac{1}{n}}^{(v-s)\wedge\frac{1}{n}} \phi^n_{v-u,s}(x_1) b(v, \phi^n_{v-u,s}(x_2)) K_n(u) du dv \\
&= \int_{s-\frac{1}{n}}^{s+\frac{1}{n}} \int_{-\frac{1}{n}}^{(v-s)} I_1 du dv + \int_{t-\frac{1}{n}}^{t+\frac{1}{n}} \int_{v-t}^{\frac{1}{n}} I_1 du dv + \int_{s+\frac{1}{n}}^{t-\frac{1}{n}} \int_{-\frac{1}{n}}^{\frac{1}{n}} I_1 du dv
\end{aligned}
$$

where $v = r + u$ and $I_1 = \phi^n_{v-u,s}(x_1) b(v, \phi^n_{v-u,s}(x_2)) K_n(u)$. The first two of these integrals are over an area that is $O(n^{-2})$ and the integrand $I_1 = O(n)$, so only the final integral will contribute to the limit.

$$
\begin{aligned}
&\int_s^t \phi^n_{rs}(x_1) b(r, \phi^n_{rs}(x_2)) dr \\
&= \int_s^t \int_{-\frac{1}{n}}^{\frac{1}{n}} \phi^n_{rs}(x_1) b(r, \phi^n_{rs}(x_2)) K_n(u) du dr \\
&= \int_{s-\frac{1}{n}}^{s+\frac{1}{n}} \int_{-\frac{1}{n}}^{\frac{1}{n}} I_2 du dr + \int_{t-\frac{1}{n}}^{t+\frac{1}{n}} \int_{-\frac{1}{n}}^{\frac{1}{n}} I_2 du dr + \int_{s+\frac{1}{n}}^{t-\frac{1}{n}} \int_{-\frac{1}{n}}^{\frac{1}{n}} I_2 du dr
\end{aligned}
$$

where $I_2 = \phi^n_{rs}(x_1) b(r, \phi^n_{rs}(x_2)) K_n(u)$. Again, the first two terms are $O(n^{-1})$, so only the last term will contribute to the limit. Combining these 2 rear-

rangements together and discarding small terms we find that

$$\begin{aligned}
&\lim_n \mathbb{E}_{\phi^n}(D_n) \\
=&\lim_n \mathbb{E}_{\phi^n}\left(\int_{s+\frac{1}{n}}^{t-\frac{1}{n}}\int_{-\frac{1}{n}}^{\frac{1}{n}} I_1 dudv - \int_{s+\frac{1}{n}}^{t-\frac{1}{n}}\int_{-\frac{1}{n}}^{\frac{1}{n}} I_2 dudr\right) \\
=&\lim_n \mathbb{E}_{\phi^n}\int_{s+\frac{1}{n}}^{t-\frac{1}{n}}\int_{-\frac{1}{n}}^{\frac{1}{n}} I_3 K_n(u) dudr \\
\leq&\lim_n \int_{s+\frac{1}{n}}^{t-\frac{1}{n}}\int_{-\frac{1}{n}}^{\frac{1}{n}} K_n(u) dudr \sup_{\substack{u\in[-\frac{1}{n},\frac{1}{n}]\\ r\in\left[s+\frac{1}{n},t-\frac{1}{n}\right]}} \mathbb{E}_{\phi^n} I_3 \\
\leq&(t-s)\lim_n \sup_{\substack{u\in[-\frac{1}{n},\frac{1}{n}]\\ r\in\left[s+\frac{1}{n},t-\frac{1}{n}\right]}} \mathbb{E}_{\phi^n} I_3
\end{aligned}$$

where

$$I_3 = \left(\phi^n_{r-u,s}(x_1)b(r,\phi^n_{r-u,s}(x_2)) - \phi^n_{rs}(x_1)b(r,\phi^n_{rs}(x_2))\right).$$

As b is Lipschitz in space and the ϕ^n have bounded diffusivity, this final supremum convereges to zero. □

8 Appendix

The following result is required to prove the existence of the coalescing diffusive flows, as stated in Theorem 3.1. It is a generalization of Proposition A.10 of [24] and has a similar proof.

Proposition 8.1. *Let E be a countable subset of $\mathbb{R}^2$ containing $\mathbb{Q}^2$, and let a, b be measurable and uniformly bounded on compacts in time and Lipschitz in space. Then, taking $C_E^\circ = C_E^{\circ,+} \cap C_E^{\circ,-}$, we have $\mu_E^{a,b}(C_E^\circ) = 1$.*

Proof. Following the proof of Proposition 8.10 in [24] we will verify that each

of five conditions hold a.s., and as they characterize C_E° inside C_E [24], the result follows. The first condition is that:

$$z_t^{(s,x+n)} = z_t^{(s,x)} + n, \qquad s,t,x \in \mathbb{Q}, \qquad s < t, \qquad n \in \mathbb{Z}.$$

Taking $e = (s,x)$ and $e' = (s, x+n)$, we have that $T^{ee'} = s$. So by proof of Proposition 2.1, this condition is satisfied.

Next we consider the 3 conditions

$$z_t^{(s,x)} = \inf_{y \in \mathbb{Q}, y > x} z_t^{(s,y)}, \qquad (s,x) \in E, \qquad t \in \mathbb{Q}, \qquad t > s,$$

$$z_t^{(s,x)} = \sup_{y \in \mathbb{Q}, y < x} z_t^{(s,y)}, \qquad (s,x) \in E, \qquad t \in \mathbb{Q}, \qquad t > s$$

and

$$\Phi^-_{(t,u]} \circ \Phi^-_{(s,t]} \leq \Phi^-_{(s,u]} \leq \Phi^+_{(s,u]} \leq \Phi^+_{(t,u]} \circ \Phi^+_{(s,t]}, \qquad s,t,u \in \mathbb{Q}, \qquad s < t < u.$$

Where we define

$$\Phi^-_{(s,t]}(x) = \sup_{y \in \mathbb{Q}, y < x} z_t^{(s,y)}, \qquad \Phi^+_{(s,t]}(x) = \inf_{y \in \mathbb{Q}, y > x} z_t^{(s,y)}.$$

Let $(s,x) \in E$ and $t,u \in \mathbb{Q}$, with $s \leq t < u$. Consider the event

$$A = \left\{ \sup_{y \in \mathbb{Q}, y < Z_t^{(s,x)}} Z_u^{(t,y)} = Z_u^{(s,x)} = \inf_{y' \in \mathbb{Q}, y' > Z_t^{(s,x)}} Z_u^{(t,y')} \right\}.$$

Note that on the countable intersection, over s, x, t, u, of the events A, the above 3 conditions hold. So to show they hold a.s., it suffices to show $\mathbb{P}(A) = 1$. Fix $n \in \mathbb{N}$ and set $Y = n^{-1} \lfloor n Z_t^{(s,x)} \rfloor$ and $Y' = Y + 1/n$. Then Y and Y' are $\mathbb{Q}$ valued, $\mathcal{F}_t$-measurable random variables. Now note that $\mathbb{P}(Y <$

$Z_t^{(s,x)} < Y') = 1$ and

$$\{Y < Z_t^{(s,x)} < Y'\} \cap \{T^{(t,Y)(t,Y')} \leq u\} \subseteq A.$$

Consider the process

$$Z_r^{(t,Y')} - Z_r^{(t,Y)} - 2(r-t)b^*$$

as a function of τ where

$$\tau = \int_t^r a(\rho, Z_\rho^{(t,Y')}) + a(\rho, Z_\rho^{(t,Y)})d\rho$$

is defined to make the diffusivity of this process 1.

This can be bounded above by a Brownian motion B_τ started at $1/n$. For n sufficiently large that $u - t > 1/n$ and

$$\begin{aligned}\mathbb{P}(T^{(t,Y)(t,Y')} \leq u) &\geq \mathbb{P}\left(\inf_{\tau \leq \frac{1}{n}} B_\tau < -\frac{b^*}{na_*}\right) \\ &= 2\Phi\left(\frac{1 + b^*/a_*}{\sqrt{n}}\right) \to 1.\end{aligned}$$

So $\mathbb{P}(A) = 1$ and the conditions hold.

The final condition is that for all $\epsilon > 0$ and all $n \in \mathbb{N}$, there exists $\delta > 0$ such that

$$\|\Phi_{(s,t]} - \mathrm{id}\|_\infty < \epsilon$$

for all $s, t \in \mathbb{Q} \cap (-n, n)$ with $0 < t - s < \delta$.

Define for $\delta > 0$ and $e = (s, x) \in E$,

$$V^e(\delta) = \sup_{s \leq t \leq s+\delta^2} |Z_t^e - x|.$$

Then, letting B be a standard Brownian motion, for sufficiently small δ and

large n

$$\begin{aligned}\mathbb{P}(V^e(\delta) > n\delta) &\leq 2\mathbb{P}\left(\sup_{s\leq t\leq s+\delta^2} B_t - B_s > \frac{n\delta - b^*\delta^2}{a^*}\right)\\ &\leq e^{-\frac{(n-1)^2}{2a^{*2}}}.\end{aligned}$$

Consider, for each $n \in \mathbb{N}$ the set

$$E_n = \left\{(j2^{-2n}, k2^{-n}) : j \in \frac{1}{2}\mathbb{Z} \cap [-n^{\frac{1}{3}}2^{2n}, n^{\frac{1}{3}}2^{2n}), k = 0, 1, \ldots, 2^n - 1\right\}$$

and the event

$$A_n = \bigcup_{e\in E_n} \{V^e(2^{-n}) > n2^{-n}\}.$$

Then, $\mathbb{P}(A_n) \leq |E_n| \sup_{e\in E_n} \mathbb{P}\left(V^e(2^{-n}) > n2^{-n}\right)$.

$$\begin{aligned}\mathbb{P}\left(V^e(2^{-n}) > n2^{-n}\right) \leq &\mathbb{P}\left(\sup_{s\leq t\leq s+2^{-2n}} Z_t^e - x > n2^{-n}\right)\\ &+ \mathbb{P}\left(\inf_{s\leq t\leq s+2^{-2n}} Z_t^e - x < n2^{-n}\right)\end{aligned}$$

For n sufficiently large that $2^n > b^*$, both of these terms are $O\left(e^{-\frac{(n-1)^2}{2}}\right)$ by the reflection principle. As $|E_n| = e^{O(n)}$, we can conclude that $\sum_n \mathbb{P}(A_n) < \infty$, so by Borel-Cantelli, almost surely there exists some $N < \infty$ such that $V^e(2^{-n}) \leq n2^{-n}$ for all $e \in E_n$, for all $n \geq N$.

Given $\epsilon > 0$, choose $n \geq N$ such that $(4n+2)2^{-n} \leq \epsilon$ and set $\delta = 2^{-2n-1}$. Then, for all rationals $s, t \in (-n, n)$ with $0 < t - s < \delta$ and all rationals

$x \in [0,1]$, there exist $e^{\pm} = (r, y^{\pm}) \in E_n$ such that

$$\begin{aligned} r \leq s < t \leq r + 2^{-2n}, \\ x + n2^{-n} < y^{+} \leq x + (n+1)2^{-n}, \\ x - (n+1)2^{-n} \leq y^{-} < x - n2^{-n}, \end{aligned}$$

then, $Z_s^{e^-} < x < Z_s^{e^+}$, so

$$x - \epsilon \leq Z_t^{e^-} \leq Z_t^{(s,x)} \leq Z_t^{e^+} \leq x + \epsilon.$$

Hence, the final condition holds almost surely and thus the proposition holds. □

The rest of the propositions in this appendix are used in the direct proof of Theorem 4.1 in Section 7. The definition of A_N can be found in that proof.

Proposition 8.2. *A_N is compact*

Proof. A_N is a closed subset of $C^{\circ}(\mathbb{R}, \mathcal{D})$, and so is complete. Therefore, by a diagonal argument, it suffices to show that for all $\epsilon > 0$ and for all sequences $\mathcal{S}$ in A_N, there exists a subsequence $\mathcal{S}'$ that is contained in a ball of radius ϵ.

To this end take M such that

$$\sum_{m=M+1}^{\infty} 2^{-m} < \frac{\epsilon}{2}$$

then we have that

$$d_C(\phi, \psi) < \sum_{m=1}^{M} 2^{-m} d_C^{(m)}(\phi, \psi) + \frac{\epsilon}{2} \quad \forall \phi, \psi \in C^{\circ}(\mathbb{R}, \mathcal{D}).$$

Thus, it suffices to find a subsequence $\mathcal{S}'$ where, for $m = 1$ to M, we have

$$d_C^{(m)}(\phi, \psi) < \frac{\epsilon}{2} \quad \forall \phi, \psi \in \mathcal{S}'. \tag{12}$$

As $d_C^{(m)}$ is increasing in m, it suffices for this to hold for $m = M$. Note that $d_C^{(M)}$ only depends on the flows between times in $[-M, M]$. By the definition of A_N, the set of paths from a given point, for each of the flows in $\mathcal{S}$, is uniformly bounded and equicontinuous when restricted to the interval $[-M, M]$. This interval is also compact, so by the Arzelà-Ascoli Theorem, the set of such restricted paths is compact in the uniform norm. Using this compactness we can, for a finite set $E_{\epsilon,M,N} \subset [-M, M] \times [0, 1]$, find a subsequence $\mathcal{S}'$ of $\mathcal{S}$ such that

$$\|\phi_{\cdot s}(x) - \psi_{\cdot s}(x)\|_{L^\infty([s,M])} < \frac{\epsilon}{2} \quad \forall \phi, \psi \in \mathcal{S}' \quad \forall (s, x) \in E_{\epsilon,M,N}. \tag{13}$$

Let $[i] = \{1, \ldots, i\}$. We will take the $\mathcal{S}'$ corresponding to

$$E_{\epsilon,M,N} = \left\{ \left(-M + m\delta_{K,N}, \frac{l\epsilon}{6} \right) : m \in \left[\left\lceil \frac{2M}{\delta_{K,N}} \right\rceil \right], l \in \left[\left\lceil \frac{6}{\epsilon} \right\rceil \right] \right\}$$

where $K = \max\left\{ \lceil \frac{6}{\epsilon} \rceil, M \right\}$. It remains to show from (13) that (12) holds for $m = M$, i.e.

$$d_{\mathcal{D}}(\phi_{ts}, \psi_{ts}) < \frac{\epsilon}{2} \quad \forall s, t \in [-M, M], s < t \quad \forall \phi, \psi \in \mathcal{S}'. \tag{14}$$

By the definition of $d_{\mathcal{D}}$ this is the same as saying there exists s, t, ϕ, ψ as above such that for all x

$$\psi_{ts}\left(x - \frac{\epsilon}{2}\right) < \phi_{ts}(x) + \frac{\epsilon}{2} \tag{15}$$

and

$$\phi_{ts}\left(x - \frac{\epsilon}{2}\right) < \psi_{ts}(x) + \frac{\epsilon}{2}.$$

We will show the first of these the other follows by symmetry.

Given s, t, ϕ, ψ as in (14), there exists

$$(u, y) \in [s, s + \delta_{K,N}] \times \left(x - \frac{\epsilon}{3}, x - \frac{\epsilon}{6} \right) \cap E_{\epsilon,M,N}$$

and by the equicontinuity condition in the definition of A_N

$$\psi_{us}\left(x - \frac{\epsilon}{2}\right) < y$$

$$\phi_{us}(x) > y.$$

Putting these together with (13) we get

$$\psi_{ts}(x - \frac{\epsilon}{2}) \leq \psi_{tu}(y) < \phi_{tu}(y) + \frac{\epsilon}{2} \leq \phi_{ts}(x) + \frac{\epsilon}{2}.$$

This is equation (15) and so we are done. □

Proposition 8.3. *As $N \to \infty$*

$$\mathbb{P}(\phi^n \in A_N) \to 1$$

uniformly in n.

Proof. Throughout W_t is a standard Brownian motion. We start by showing that w.h.p. the condition that gives uniform boundedness on compact intervals holds.

$$\begin{aligned}
&\mathbb{P}\left(|\phi^n_{ts}(x) - x| < 4b^*_k k + kN\sqrt{8a^*_k} + 1 \quad \forall x \in [0,1] \quad \forall s < t \in [-k,k]\right) \\
\geq &\mathbb{P}\left(\sup_{t\in[-k,k]} |\phi^n_{t,-k}(0)| < 2b^*_k k + kN\sqrt{2a^*_k}\right) \\
\geq &1 - 4\mathbb{P}\left(\sqrt{a^*_k} W_{2k} > kN\sqrt{2a^*_k}\right) \\
= &1 - 4\Phi\left(-N\sqrt{k}\right)
\end{aligned}$$

and thus

$$\mathbb{P}(|\phi^n_{ts}(x) - x| < 4b^*_k k + kN\sqrt{8a^*_k} + 1 \quad \forall s < t \in [-k,k] \quad \forall x \in [0,1] \quad \forall k)$$

$$\geq 1 - 4\sum_{k=1}^{\infty} \Phi(-N\sqrt{k}) \to 1$$

Now we will show that w.h.p. the equicontinuity requirement on compact intervals holds. Let

$$E_{k,N} = \left\{ \left(k - m\delta_{k,N}, \frac{l}{3k} \right) : m \in \left\{ 1, \ldots, \left\lceil \frac{2k}{\delta_{k,N}} \right\rceil \right\}, l \in \{1, \ldots, 3k\} \right\}$$

The below calculation says that with high probability for all k paths from each of these points will not move more than $\frac{1}{3k}$ from their stating point within time $2\delta_{k,N}$ and the non-crossing property then implies the required equicontinuity.

$$\begin{aligned}
&\mathbb{P}\left(|\phi_{ts}(x) - x| \leq \frac{1}{k} \quad \forall x \in [0,1] \quad \forall s,t \in [-k,k] \text{ with } t - s \in [0, \delta_{k,N}] \right) \\
\geq &\mathbb{P}\left(\sup_{t \in [s, s+2\delta_{k,N}]} |\phi_{ts}(x) - x| < \frac{1}{3k} \quad \forall (s,x) \in E_{k,N} \right) \\
\geq &1 - 4|E_{k,N}|\mathbb{P}\left(\sqrt{a^*_k} W_{2\delta_{k,N}} + 2\delta_{k,N} b^*_k > \frac{1}{3k} \right) \\
= &1 - 12k \left\lceil \frac{2k}{\delta_{k,N}} \right\rceil \Phi\left(-\frac{1}{\sqrt{2\delta_{k,N} a^*_k}} \left(\frac{1}{3k} - 2\delta_{k,N} b^*_k \right) \right) \\
\geq &1 - \frac{36k^2}{\delta_{k,N}} \Phi\left(-\frac{1}{\sqrt{18k^2 a^*_k \delta_{k,N}}} + \sqrt{\frac{2{b^*_K}^2 \delta_{k,N}}{a^*_k}} \right) \\
\geq &1 - \max\left\{ \frac{72k^2 {b^*_k}^2}{a^*_k}, 648k^5 N a^*_k (1 + a^*_k + b^*_k) \right\} \Phi\left(-\sqrt{kN(1 + a^*_k + b^*_k)} + 1 \right)
\end{aligned}$$

As the maximum can be bounded by a polynomial in k, N, a_k^* and b_k^*, and $\Phi(\dots)$ is decreasing exponentially in all of those variables, we can conclude by use of a union bound that

$$\mathbb{P}\left(|\phi_{ts}(x) - x| \leq \frac{1}{k} \quad \forall x \in [0,1] \quad \forall s,t \in [-k,k] \text{ with } t-s \in [0, \delta_{k,N}] \quad \forall k\right)$$

$\to 1$ as $N \to \infty$. □

Proposition 8.4. *$\mathbb{E}\left(M_{st}(x_1, x_2, b, a, \phi)\right)$ is a continuous function of x_1 and x_2.*

Proof. We will show that

$$|\mathbb{E}_\phi\left(M_{st}(x_1, x_2, b, a, \phi)\right) - \mathbb{E}_\phi\left(M_{st}(x_1', x_2', b, a, \phi)\right)| \to 0$$

uniformly for $d_{eucl}((x_1, x_2), (x_1', x_2')) < \delta$ as $\delta \to 0$. We start by decomposing $M_{st}(x_1, x_2, b, a, \phi)$ into the integrals up to time $s + \delta$ and the rest. The integrals up until time $s + \delta$ are

$$-\int_s^{s+\delta} \phi_{rs}(x_1) b(r, \phi_{rs}(x_2)) + \phi_{rs}(x_2) b(r, \phi_{rs}(x_1)) dr$$

and

$$-\int_{T^{(s,x_1)(s,x_2)} \wedge t}^{(T^{(s,x_1)(s,x_2)} \wedge t) \vee (s+\delta)} a(r, \phi_{rs}(x_1)) dr.$$

Taking expected value w.r.t. ϕ and exchanging order of integration leaves two integrals with length at most δ and integrands bounded by

$$b^* \sup_{r \in [s, s+\delta]} \mathbb{E}_\phi(|\phi_{rs}(x_1)| + |\phi_{rs}(x_2)|) \text{ and } a^*$$

respectively. As $\phi_{rs}(x_i)$ is uniformly integrable for $r \leq t$ these integrals contribute only $O(\delta)$ to M. Thus they can be neglected.

We will use M_{st}^δ to mean M_{st} minus the integrals we have just shown are

$O(\delta)$. Note that

$$\mathbb{E}_\phi\left(M^\delta\right) = \mathbb{E}_\phi\left(\mathbb{E}_\phi\left(M^\delta|\mathcal{F}_{s+\delta}\right)\right)$$

and by the strong Markov property

$$\mathbb{E}_\phi\left(M^\delta_{st}(x_1, x_2, b, a, \phi)|\mathcal{F}_{s+\delta}\right)$$

is a function of $\phi_{s+\delta,s}(x_1)$ and $\phi_{s+\delta,s}(x_2)$. Proposition 8.5 says that

$$d_{TV}((\phi_{s+\delta,s}(x_1), \phi_{s+\delta,s}(x_2)), (\phi_{s+\delta,s}(x_1'), \phi_{s+\delta,s}(x_2'))) \to 0$$

so we can deduce that

$$d_{TV}\left(\mathbb{E}_\phi\left(M^\delta_{st}(x_1, x_2, b, a, \phi)|\mathcal{F}_{s+\delta}\right), \mathbb{E}_\phi\left(M^\delta_{st}(x_1', x_2', b, a, \phi)|\mathcal{F}_{s+\delta}\right)\right) \to 0.$$

Combining this with the fact that $\mathbb{E}_\phi(M^\delta_{st}(x_1, x_2, b, a, \phi)|\mathcal{F}_{s+\delta})$ is uniformly integrable for (x_1, x_2) in each compact set, we are done. □

Proposition 8.5.

$$d_{TV}((\phi_{s+\delta,s}(x_1), \phi_{s+\delta,s}(x_2)), (\phi_{s+\delta,s}(x_1'), \phi_{s+\delta,s}(x_2'))) \to 0$$

uniformly for $d_{eucl}((x_1, x_2), (x_1', x_2')) < \delta$ *as* $\delta \to 0$.

Proof. Let $\tilde{\phi}$ have the same distribution as ϕ but be coupled with ϕ such that, for each $i = 1, 2$, we have $\phi_{t,s}(x_i)$ and $\tilde{\phi}_{t,s}(x_i')$ evolve independently until they take the same value at which point they coalesce. This is possible, as having fixed ϕ we can construct $\tilde{\phi}$ by first constructing $\tilde{\phi}(x_1')$ independently until it hits $\phi(x_1)$, then constructing $\tilde{\phi}(x_2')$ independently until it hits $\tilde{\phi}(x_1')$ or $\phi(x_2)$.

$$
\begin{aligned}
& d_{TV}((\phi_{s+\delta,s}(x_1), \phi_{s+\delta,s}(x_2)), (\phi_{s+\delta,s}(x_1'), \phi_{s+\delta,s}(x_2'))) \\
\leq & 1 - \mathbb{P}\left(\phi_{s+\delta,s}(x_i) = \tilde{\phi}_{s+\delta,s}(x_i) \text{ for both } i = 1, 2\right) \\
\leq & \mathbb{P}\left(\phi_{s+\delta,s}(x_1) \neq \tilde{\phi}_{s+\delta,s}(x_1)\right) + \mathbb{P}\left(\phi_{s+\delta,s}(x_2) \neq \tilde{\phi}_{s+\delta,s}(x_2)\right) \\
\leq & 2(1 - 2\mathbb{P}(a_* W_\delta < -\delta - \delta b^*)) \\
= & 2\left(1 - 2\Phi\left(-\sqrt{\delta}\frac{1+b^*}{a_*}\right)\right) \\
\leq & 2\left(1 - 2\left(0.5 - \sqrt{\frac{\delta}{2\pi}}\frac{1+b^*}{a_*}\right)\right) \\
= & \sqrt{\frac{8\delta}{\pi}}\frac{1+b^*}{a_*} \to 0.
\end{aligned}
$$

□

Proposition 8.6.

$$\mathbb{E}_\phi \mathbb{E}_x(M_{st}(x_1, x_2, b, a, \phi)) = \lim_n \mathbb{E}_{\phi^n} \mathbb{E}_x\left(M_{st}(x_1, x_2, b, a, \phi^n)\right)$$

Proof. We would like to be able to say that $\mathbb{E}_x(M_{st})$ is a continuous function of ϕ, and apply weak convergence. Unfortunately, even after averaging over x, this still isn't true, as $T^{(s,x_1)(s,x_2)}$ is not a continuous function of ϕ, so we now proceed to smooth M_{st} even more. Define

$$T_\eta = \inf\{r \geq s : d(\,\phi_{rs}(x_1), \mathbb{Z} + \phi_{rs}(x_2)\,) < \eta\}$$

and then define

$$\begin{aligned}\tilde{M}^{\epsilon}_{st}(x_1,x_2,b,a,\phi) =&\phi_{ts}(x_1)\phi_{ts}(x_2) - \int_s^t \phi_{rs}(x_1)b(r,\phi_{rs}(x_2)) \\ &+ \phi_{rs}(x_2)b(r,\phi_{rs}(x_1))dr - \frac{1}{\epsilon}\int_0^{\epsilon}\int_{T_\eta\wedge t}^t a(r,\phi_{rs}(x_1))drd\eta.\end{aligned}$$

By applying the triangle inequality the following three claims will now suffice to complete the proof. Firstly

$$\mathbb{E}_{\phi^n}\mathbb{E}_x\tilde{M}^{\epsilon}(\phi^n) \to \mathbb{E}_{\phi}\mathbb{E}_x\tilde{M}^{\epsilon}(\phi) \text{ as } n\to\infty$$

secondly

$$\mathbb{E}_{\phi}\mathbb{E}_x\tilde{M}^{\epsilon}(\phi) \to \mathbb{E}_{\phi}\mathbb{E}_x M(\phi) \text{ as } \epsilon\to 0$$

and thirdly

$$\mathbb{E}_{\phi^n}\mathbb{E}_x\tilde{M}^{\epsilon}(\phi^n) \to \mathbb{E}_{\phi^n}\mathbb{E}_x M(\phi^n) \text{ as } \epsilon\to 0 \text{ uniformly in } n.$$

We first prove the second claim. Note that T_η monotonically increases to T_0 as $\eta\to 0$ and thus $\tilde{M}^{\epsilon}$ is monotonically increasing to M as $\epsilon\to 0$. Thus the second claim holds by the Monotone Convergence Theorem.

We next prove the third claim. We have that

$$\begin{aligned}|\tilde{M}^{\epsilon}(\phi^n) - M(\phi^n)| =&\frac{1}{\epsilon}\int_0^{\epsilon}\int_{T_\eta\wedge t}^{T_0\wedge t} a(r,\phi^n_{rs}(x_1))drd\eta \\ \leq& a^*\left((T_0-T_\epsilon)\wedge(t-s)\right)\end{aligned}$$

and thus

$$\left|\mathbb{E}_{\phi^n}\left(\tilde{M}^{\epsilon}(\phi^n) - M(\phi^n)\right)\right| \leq a^*\left(\epsilon + (t-s)\mathbb{P}(T_0-T_\epsilon>\epsilon)\right).$$

Using the strong Markov property at time T_ϵ we can see that

$$\begin{aligned}\mathbb{P}(T_0 - T_\epsilon > \epsilon) \leq & 1 - 2\mathbb{P}(2a_* W_\epsilon < -\epsilon(1+2b^*)) \\ = & 1 - 2\Phi\left(-\frac{\sqrt{\epsilon}(1+2b^*)}{2a_*}\right) \\ = & O(\sqrt{\epsilon}).\end{aligned}$$

Putting this together and averaging over x, we get the third claim with the order of the expectations swapped. Note that each term of $\tilde{M}^\epsilon(\phi^n)$ and $M(\phi^n)$ have sub-exponential tails. Thus, we can apply Fubini's theorem to deduce the third claim.

Finally, we will show that $\mathbb{E}_x \tilde{M}^\epsilon(\phi)$ is a continuous function of ϕ, from which our first claim immediately follows due to weak convergence. Then we will be done.

Fix $\phi' \in C^\circ(\mathbb{R}, \mathcal{D})$. Let ϕ'' be distance at most δ from ϕ'. We have, for $\delta < 1$, that

$$\phi'(x_1) - 1 - \delta \leq \phi'(x_1 - \delta) - \delta \leq \phi''(x_1) \leq \phi'(x_1 + \delta) + \delta \leq \phi'(x_1) + 1 + \delta.$$

Let $[l_1, u_1]$ be the interval that x_1 is being averaged over, then

$$\begin{aligned}
\mathbb{E}_{x_1}\left(\phi''(x_1)\right) &\leq \frac{1}{u_1 - l_1}\int_{l_1}^{u_1} \phi'(x_1+\delta) + \delta dx_1 \\
&\leq \frac{1}{u_1 - l_1}\int_{l_1}^{u_1-\delta} \phi'(x_1+\delta)dx_1 + \frac{1}{u_1 - l_1}\int_{u_1-\delta}^{u_1} \phi'(x_1) + 1 dx_1 + \delta \\
&\leq \mathbb{E}_{x_1}\left(\phi'(x_1)\right) - \frac{1}{u_1 - l_1}\int_{l_1}^{l_1+\delta} \phi'(x_1)dx_1 \\
&\qquad + \frac{1}{u_1 - l_1}\int_{u_1-\delta}^{u_1} \phi'(x_1)dx_1 + \delta\left(\frac{1}{u_1 - l_1} + 1\right) \\
&\leq \mathbb{E}_{x_1}\left(\phi'(x_1)\right) + \delta\left(1 + \frac{1 + 2\sup_{l_1 \leq x_1 \leq u_1}|\phi'(x_1)|}{u_1 - l_1}\right) \\
&\to \mathbb{E}_{x_1}\left(\phi'(x_1)\right) \text{ as } \delta \to 0.
\end{aligned}$$

The lower bound is similar. We can deduce that $\mathbb{E}_{x_1}(\phi(x_1))$ is continuous in ϕ, and so the first term of $\mathbb{E}_x\tilde{M}^\epsilon$, i.e.

$$\mathbb{E}_{x_1}(\phi(x_1))\mathbb{E}_{x_2}(\phi(x_2))$$

is also continuous in ϕ. To show that the second term of $\mathbb{E}_x\tilde{M}^\epsilon$ is continuous in ϕ, as $\phi''_{rs}(x_1)$ is bounded uniformly over $r \in [s,t]$ and ϕ'' for fixed δ and ϕ', it suffices to show that

$$\mathbb{E}_{x_1} b(t, \phi(x_1))$$

is continuous in ϕ (for all $t > s$).

$$\begin{aligned}
b\left(t, \phi''(x_1)\right) &\leq \sup_{|\delta x| \leq \delta} b\left(t, \phi'(x_1 + \delta x)\right) + L\delta \\
&\leq b\left(t, \phi'(x_1)\right) + L\left(\delta + \sup_{|\delta x| \leq \delta}|\phi'(x_1 + \delta x) - \phi'(x_1)|\right).
\end{aligned}$$

Cut the interval $[l_1 - \delta, u_1 + \delta]$ into pieces of length δ. Let C be the amount that ϕ increases by over that interval. Call a piece bad if ϕ increases by more than $\frac{\sqrt{\delta}}{2}$ on that piece or either of the neighbouring pieces. As ϕ is non-decreasing there can be at most $\frac{6C}{\sqrt{\delta}}$ bad pieces. If x_1 is not in a bad piece then

$$\sup_{|\delta x| \leq \delta} |\phi'(x_1 + \delta x) - \phi'(x_1)| \leq \sqrt{\delta}.$$

So our bound on $b\left(t, \phi''(x_1)\right)$ gives

$$b\left(t, \phi''(x_1)\right) \leq b\left(t, \phi'(x_1)\right) + L\left(\delta + \sqrt{\delta} + C\mathbb{1}_{\{x_1 \in \text{a bad piece}\}}\right).$$

Combining this with the corresponding lower bound whose derivation is similar we find

$$\begin{aligned}
&\left|\mathbb{E}_{x_1}\left(b\left(t, \phi''(x_1)\right)\right) - \mathbb{E}_{x_1}\left(b\left(t, \phi'(x_1)\right)\right)\right| \\
\leq &L\left(\delta + \sqrt{\delta} + C\mathbb{P}_{x_1}(x_1 \in \text{a bad piece})\right) \\
\leq &L\left(\delta + \sqrt{\delta} + \frac{6C^2\sqrt{\delta}}{u_1 - l_1}\right) = O\left(\sqrt{\delta}\right).
\end{aligned}$$

Thus, $\mathbb{E}_{x_1} b(t, \phi(x_1))$ and the second term of $\mathbb{E}_x \tilde{M}^\epsilon$ are continuous in ϕ.

Similarly, we can conclude that $\mathbb{E}_{x_1} a(t, \phi(x_1))$ is continuous with respect to ϕ, and further, as the products of intervals generate the Borel σ-algebra on $\mathbb{R}^2$, that

$$\int a(t, \phi(x_1)) d\mu(x) \tag{16}$$

is a continuous function of ϕ for each measure μ that is bounded, compactly supported and absolutely continuous with respect to Lebesgue measure on

$\mathbb{R}^2$. This will be useful after we rewrite the third term of $\mathbb{E}_x \tilde{M}^\epsilon$ as

$$\begin{aligned}
&\mathbb{E}_x \frac{1}{\epsilon} \int_0^\epsilon \int_{T_\eta \wedge t}^t a(r, \phi_{rs}(x_1)) dr d\eta \\
=&\mathbb{E}_x \int_0^\epsilon \int_s^t a(r, \phi_{rs}(x_1)) \frac{\mathbb{1}_{\{r > T_\eta\}}}{\epsilon} dr d\eta \\
=&\int_s^t \mathbb{E}_x \left(a(r, \phi_{rs}(x_1)) \int_0^\epsilon \frac{\mathbb{1}_{\{r > T_\eta\}}}{\epsilon} d\eta \right) dr.
\end{aligned}$$

To show this is continuous it suffices to show that

$$\mathbb{E}_x \left(a(t, \phi(x_1)) \int_0^\epsilon \frac{\mathbb{1}_{\{t > T_\eta\}}}{\epsilon} d\eta \right)$$

is continuous and uniformly bounded for all $t > s$. The boundedness is immediate. The continuity is not immediate from (16) being continuous, because T_η depends on ϕ. However, it can be shown as follows. Let T_η^0 be the T_η corresponding to ϕ' and define T_η^δ similarly. Let μ' be the measure on $\mathbb{R}^2$ with Radon-Nikodym derivative

$$\int_0^\epsilon \frac{\mathbb{1}_{\{t > T_\eta\}}}{\epsilon} d\eta$$

with respect to the uniform probability measure on $I_1 \times I_2$ and define μ'' similarly. Then

$$\begin{aligned}
&\left| \mathbb{E}_x \left(a\left(t, \phi''(x_1)\right) \int_0^\epsilon \frac{\mathbb{1}_{\{t > T_\eta^\delta\}}}{\epsilon} d\eta \right) - \mathbb{E}_x \left(a\left(t, \phi'(x_1)\right) \int_0^\epsilon \frac{\mathbb{1}_{\{t > T_\eta^0\}}}{\epsilon} d\eta \right) \right| \\
=&\left| \int a\left(t, \phi''(x_1)\right) d\mu''(x_1) - \int a\left(t, \phi'(x_1)\right) d\mu'(x_1) \right| \\
\leq&\left| \int a\left(t, \phi''(x_1)\right) d\mu''(x_1) - \int a\left(t, \phi''(x_1)\right) d\mu'(x_1) \right| \\
&+ \left| \int a\left(t, \phi''(x_1)\right) d\mu'(x_1) - \int a\left(t, \phi'(x_1)\right) d\mu'(x_1) \right|.
\end{aligned}$$

The second of these terms is small due to the continuity of (16), the first term is bounded by

$$\frac{a^*}{\epsilon}\int_{l_1}^{u_1}\int_{l_2}^{u_2}\int_0^{\epsilon}\left|\mathbb{1}_{\{t>T_\eta^\delta\}}-\mathbb{1}_{\{t>T_\eta^0\}}\right|d\eta dx_2 dx_1.$$

The contribution to this integral when $|x_1 - x_2| < 2\delta$ is clearly small. We will show that the contribution when $x_1 \geq x_2 + 2\delta$ is small and, as the case for $x_1 \leq x_2 - 2\delta$ is similar, we will then be done. Conditional on $x_1 \geq x_2 + 2\delta$ we have

$$T^0_{\eta+2\delta}(x_1-\delta, x_2+\delta) \leq T^\delta_\eta(x_1,x_2) \leq T^0_{\eta-2\delta}(x_1+\delta, x_2-\delta)$$

and thus our integrand is zero unless

$$t \in \left[T^0_{\eta+2\delta}(x_1-\delta, x_2+\delta), T^0_{\eta-2\delta}(x_1+\delta, x_2-\delta)\right]. \tag{17}$$

To see that the integral is small we will change the variables of the integral. We will use an orthonormal substitution to change the variables to, $\alpha = (2\eta - x_1 + x_2)/\sqrt{6}$, $\beta = (-\eta - x_1 + x_2)/\sqrt{3}$, and $\gamma = (x_1 + x_2)/\sqrt{2}$. For some values of the endpoints, the result is as follows

$$\frac{a^*}{\epsilon}\int_{\gamma_1}^{\gamma_2}\int_{\beta_1}^{\beta_2}\int_{\alpha_1}^{\alpha_2}\left|\mathbb{1}_{\{t>T_\eta^\delta\}}-\mathbb{1}_{\{t>T_\eta^0\}}\right|d\alpha d\beta d\gamma.$$

Note that the end points of the integral are independent of δ. Thus the whole expression has at most the same order as the inner integral for small δ. For any fixed value of β and γ, the interval of α for which Equation (17) can hold has length at most $2\sqrt{6}\delta$ (as for values of α differeing by more than that, the corresponding intervals on the right hand side are disjoint). As the integrand is bounded by one, the inner integral is bounded by $2\sqrt{6}\delta$. Therefore, the total expression is $O(\delta)$. □

Part II

Off-Grid Diffusion Limited Aggregation can be Simulated in $\tilde{O}(n)$ Time and Space

In this part we look at previous algorithms proposed for simulating off-grid DLA. We provide the first analytic analysis of the asymptotic runtimes of four previously proposed algorithms. An algorithm from the early 1990s is found to have better asymptotic runtime and memory usage than more recent proposals. We also wrote code for a variation of that algorithm combining it with tricks from a couple of the more recent papers and a new trick presented here. Our code can be downloaded at `www.github.com/mathsjames/fastDLA` and can also simulate noise reduced DLA.

1 Introduction

DLA is a model of irreversible growth proposed by Witten and Sander in 1981 [36]. The definition works in any dimension but work has focussed primarily on the two and three dimensional cases. It has attracted a great degree of interest in the following four decades. The original model was defined on an underlying lattice and was studied through the 1980s with simulations in e.g. [18] [17] [3] and with an analytical result in [13]. The off-grid version introduced in [18] was studied numerically in e.g. [17] [25] [26]. A good survey of work on DLA prior to 1995 can be found in [19].

Further models supposed to have the same asymptotic behaviour were introduced by Hastings and Levitov in [12] and Ball et al. in [4]. The Hastings-Levitov method is specific to two dimensions but unifies analogues

of DLA, dielectric breakdown and the Eden model in a family of models with one real parameter. Some parts of this family, not including those corresponding to DLA, have had analytic results proved about them in [23] [31], however extending these results to the DLA case seems to be a hard problem. Further generalisation of the Hastings-Levitov model is suggested and analysed in [35]. The noise reduced DLA of Ball et al. has been demonstrated to converge faster into its asymptotic growth rate than the original model does [4] [8], but shows no signs of accessibility by analytic methods. The algorithms presented here for off-grid DLA generalise easily to the noise reduced case and so even if one is interested in exploring the asymptotics of this faster converging process this work is very relevant and the code available at `www.github.com/mathsjames/fastDLA` will also compute this model.

More recent work on off-grid DLA has included a study by Menshutin in 2012 [21] that demonstrated that finite size effects are quite substantial for clusters with up to 10^7 particles. Generating even one cluster of this size is not trivial and being able to do so efficiently is important for being able to understand the asymptotics of DLA. In the rest of the introduction we look at the history of simulating DLA.

Originally the DLA process took place on a lattice, however in 1985 Ball and Brady [3] demonstrated that the asymptotic shape of this original DLA model was anisotropic and dependent on the underlying lattice. As a result the *non-lattice* or *off-grid* model introduced by Meakin in 1983 [18] is a better candidate for a universal model. The algorithm in [3] was only very briefly described and they claimed that, for a cluster of n particles, it only requires $O\left(\log^2(n)\right)$ steps per particle. This algorithm was generalised to the off-grid case by Ossadnik in 1991 [25]. Here, the algorithm is well described and claimed to require $O\left(\log^2(n)\right)$ steps per particle. However, the focus in that paper is also the results of the simulations, so no rigorous analysis and limited runtime data are provided. Ossadnik was able to produce 100 clusters of 10^6 particles using RAM only, and managed to grow a cluster of 6×10^6

particles using swap space. We will show in Section 5 that this algorithm requires $O\left(\log^2(n)\right)$ steps per particle and in Section 7 that each step can be implemented in $O\left(\log(n)\right)$ time, thus this algorithm runs in $O\left(n\log^3(n)\right)$ time. We also show in Section 7 that the algorithm requires $O\left(n\right)$ space.

As in this part, and in Part III, we are primarily concerned with Meakin's off-grid DLA, the acronym DLA will refer to Meakin's DLA unless otherwise specified.

More recent work in simulating square-lattice DLA has been conducted by Loh [16]. This was recently used, by Grebenkov and Beliaev[11], to demonstrate that such clusters appear to converge to a one dimensional limit. However, this is mainly concerned with reducing biases introduced by approximating the lattice. Therefore it is not relevant to the case we are concerned with.

In 2006 Menshutin and Shchur [20] generated 100 clusters of 5×10^7 particles. Their algorithm is discussed in Section 4.4.1 where it is shown to have a runtime of $\Omega\left(n^{\frac{8}{4+d}+o(1)}\right)\approx\Omega\left(n^{1.40}\right)$ given certain physical assumptions, including that the dimension of the cluster is given by $d\approx1.71$. Whether this algorithm is good for the cluster sizes considered is unclear as they provide no data on runtimes, and the increase in cluster size generated compared to [25] could be explained by better hardware. In 2008 Alves, Ferreira and Martins [1] proposed an algorithm for off-grid DLA. They cite Ball and Brady but present an algorithm that runs, as we will show in Section 4.4.2, in $n^{1+\frac{2}{d}+o(1)}\approx\Theta\left(n^{2.17}\right)$ time, whilst claiming that runtimes grow as $n^{1.4}$. The largest cluster that they grow has 10^6 particles and they do so only twenty times faster than [25], despite a factor of 100 improvement in CPU speed and a lot more available memory which Ossadnik had needed to optimize for. In 2014, Kuijpers, de Martín and Van Ommen [27] proposed another algorithm which they claimed runs in $n^{1.08}$ time. We will show in Section 4.4.3 that this algorithm runs in at best $n^{1+\frac{1}{d}+o(1)}\approx n^{1.58}$ time, depending on how parameters are chosen. The runtimes they give are significantly faster than those

of [1] but are still slow compared to the algorithm of [25]. Furthermore their algorithm requires 100 times as much space as that of [25]. Details of these runtimes are given and considered in Section 6.

The algorithm proposed in this work, in Section 5, differs only slightly from that of [25]. In Section 4.2 we describe a technique for resetting particles from far away to the starting circle, that first appeared in [30]; and in Section 4.3 a new technique for adhering particles to the cluster without a step size dependent approximation is presented. Both of these are then incorporated into our recommended algorithm. In Section 7 we describe Ossadnik's idea for keeping memory consumption down; however we found this to be an unnecessary and inefficient trade off with runtime for our application and so have not implemented it.

In Section 8 we discuss how well this algorithm generalizes to higher dimensions and what would be required to do better in that context. In Section 9 we discuss how the logarithmic factors might be improved and possible directions for future work.

2 Definition of Off-Grid Diffusion Limited Aggregation

DLA models the growth of a cluster of particles. As is standard, we take each particle to be a circle of radius one. However, the idea of all the algorithms discussed here could be extended to the setting where the radius of each particle is generated independently before the particle is added to the cluster (although fat-tailed distributions of radii would lead to a slow down). Let $C(r)$ denote the circle of radius r centred on the origin. The process of DLA is described in Algorithm 1.

Initially, the cluster consists of a single particle at the origin. Each of n particles in turn follows a Brownian motion from infinity until it touches the cluster. At this point the particle is fixed at its current location and added

Algorithm 1: Ideal DLA

Input: Number of particles to add n
Output: $P \in (\mathbb{R}^2)^{n+1}$

1: $P = [(0,0)]$
2: for i in $\{1..n\}$
3: Let B_t be 2D Brownian motion starting from ∞
4: Let $T = \inf\{t | d(B_t, P) \leq 2\}$
5: Append B_T onto P
6: endfor
7: **Return** P ▷ `The sequence of particle centres`

to the cluster. Once all particles have been added, the cluster is complete. We are here using a Brownian motion started from infinity to mean that the limit of this distribution (of cluster or attachment point) should be taken as the starting point of the Brownian motions goes to infinity. We describe how to approximate Algorithm 1 numerically in the following sections.

3 The Basic Algorithm

To make Algorithm 1 implementable on a computer we make the following changes.

- Instead of starting from infinity each particle will start from a uniformly randomly selected point on a circle containing the cluster up to this aggregation, referred to as the starting circle.
- The Brownian motion is replaced by a discrete time random walk with a fixed step size in a uniformly chosen direction.
- If during a random walk the particle gets very far from the cluster (say ten times the radius of the starting circle) then it is reset to a uniformly random point on the starting circle.

- Upon collision with the cluster a particle is placed on the line between the position it was in immediately before collision and the position it had stepped to, so that it just touches the cluster.

- All positions and steps are computed in floating point arithmetic

The first two changes above are not approximations at all but are exact, so long as the particle has distance from the cluster at least given by the step size. This is because harmonic measure from the centre of a disc to its boundary is precisely the uniform distribution on the boundary.

The third and fourth changes incur an error. The error from resetting the particle can be made small by only resetting once the particle gets very far away; however this incurs extra computational cost which isn't negligible, even with optimisations to the step sizes. Crucially this error will not shrink with the number of particles in the cluster. A means to remove this approximation at little computational cost which first appeared in [30], is described in Section 4.2.

The error due to the fourth change seems likely to become negligible as the number of particles becomes large. More precisely, it seems it will be of the same order as the error between the model and the real world physical processes it is meant to model. However, in Section 4.3 we present an approach to removing this approximation in a way similar to addressing the approximation in the third change, because we want to have a single simple model without an extraneous parameter, step size, affecting the distribution of the resulting cluster.

The use of floating point numbers is a necessary approximation as computers can't do infinite precision of arbitrary real numbers and the continuous randomness entails generic real numbers. The resulting algorithm is specified in Algorithm 2

In the following sections we shall optimize this basic algorithm, but before that we shall assess its time and space requirements.

Algorithm 2: Basic DLA

Parameter: Step size l, reset ratio σ.
Input: Number of particles to add n
Output: $P \in (\mathbb{R}^2)^{n+1}$

$P = [(0,0)]$
$r = 2$
for i in $\{1..n\}$
    Sample $p \leftarrow C(r)$
    while $d(p, P) > 2$
        Sample $s \leftarrow C(l)$
        $p = p + s$
        if $d(0, p) > \sigma r$ then sample $p \leftarrow C(r)$
    endwhile
    $p = p - as$ for smallest positive a such that $d(p - as, P) = 2$
    $r = \max\{r, |p| + 2\}$
    Append p onto P
endfor
Return P ▷ `The sequence of particle centres`

3.1 Complexity

The spatial complexity of this basic algorithm is $O(n)$ as it is only required to store the points in the cluster so far and a constant number of working variables.

The time complexity can be broken down as,

$$n \times \text{cost per step} \times \text{steps per particle.} \tag{18}$$

The most complicated part of understanding the time complexity is ascertaining how many steps each particle will take i.e. how many times the while loop is executed. To get a lower bound that matches the order of the upper bound we will require the following assumption.

Assumption 3.1. *If P is a DLA cluster, $r = \max_{p \in P} |p| + 2$, and x is*

sampled from $C(r)$ then

$$\mathbb{E}\left(\frac{d(x,P)}{r}\right) = \Omega(1)$$

This assumption roughly says that DLA clusters are not asymptotically circular discs. Hopefully the reader will agree this is clear from simulations although there is no known proof.

Lemma 3.2. *Given Assumption 3.1, the while loop in Algorithm 2 is on average executed $\Theta(r^2)$ times for each execution of the for loop.*

Proof. The number of while loop executions within a for loop execution is the number of steps a single particle takes before colliding with the cluster. For fixed sigma the expected number of times each particle is reset is $O(1)$ as the probability of being reset before adhering is bounded away from 1. Thus the order of the expected number of steps a particle will take is equal to the order of the expected number of steps it will take before it is first reset.

For each iteration of the for loop. Let P be the cluster of particles so far, $r = \max_{p\in P}|p| + 2$ be the radius of the cluster (plus 1) and p_s be the starting point of the next particle.

The number of steps before the particle either adheres or is reset for the first time is bounded above by the number of steps it would take to get distance $(\sigma + 2)r$ from p_s and below by how long it takes to get distance $d(p_s, P) - 2$ which by assumption is $\Omega(r)$ with probability bounded away from zero. A random walk takes order k^2 expected time to get distance k from its starting point so the result follows. □

Assumption 3.3. *There exists $d \in \mathbb{R}$, called the dimension of DLA such that if P is a DLA cluster of n points and $r = \max_{p\in P}|p| + 2$ then*

$$r = n^{\frac{1}{d}+o(1)}$$

This assumption is uncontroversial in the physics literature and the value of d has been found to be about 1.71 ± 0.01 [4].

Proposition 3.4. *Under Assumptions 3.1 and 3.3, the time complexity of the basic algorithm is* $n^{2+\frac{2}{d}+o(1)}$

Proof. By Lemma 3.2 each particle requires $\Theta\left(r^2\right) = n^{\frac{2}{d}+o(1)}$ steps to place and each time we wish to check for a collision (the condition of the while loop) we must check each of the $\Theta(n)$ particles already in the cluster. Therefore, by Expression (18), the runtime is of order $n \times n^{\frac{2}{d}+o(1)} \times n$. □

This runtime is approximately $\Theta\left(n^{3.17}\right)$. We now move onto the methods used in previous work to speed this process up. These assumptions are needed in the analysis of previous attempts to speed up the algorithm given in [20], [1] and [27], and analysing the algorithm of [20] will require further assumptions. However, no unproven assumptions are required to prove the bounds on the algorithm of [25] and the variation on it that we recommend.

4 Improvements to Runtime and Accuracy

We break down the improvements that we recommend into four changes. Most of the runtime improvement comes from the first and last, whilst the second and third improve the accuracy of the simulation. The first improvement, in Section 4.1, is employed by all previous proposed algorithms [25] [20] [1] [27] and provides a way to check for collisions in $O(1)$ time. The second, in Section 4.2, is from [30] and allows us to reset a particle that moves away from the cluster cheaply and exactly. The third improvement, in Section 4.3, allows us to not incur any approximation in a step that could result in the particle colliding with the cluster. The fourth improvement, in Section 4.4, is to take larger steps. All previous algorithms do this in different ways and we look at how they each approached reducing the number of steps per

particle, analysing how much of a reduction they each achieve and at what computational cost.

After implementing the third improvement there is no checking for collisions step. However, the idea behind the first improvement is used in implementing the third improvement, so is still important.

4.1 Fast Collision Checking

It is convenient to store the particles according to where they are in the cluster so as to speed up checking whether or not there is a collision after each step.

To achieve this goal we can lay a grid over the space the cluster will grow into (this idea of storing information in a grid laid over the cluster will also be used in all approaches to increasing step size in Section 4.4). The grid must cover a large enough area that we can be confident that the cluster we wish to grow will not exceed its bounds, i.e. the area must be $\Omega\left(r^2\right) \approx \Omega\left(n^{1.17}\right)$. Further we will take the side length of each square of the grid to be a constant, in [20] the side length of twenty is suggested. For each square of the grid we store a vector containing all of the points within that square. Now when checking for collisions we need only check the distance from each of the points in grid squares close to our particle.

This alteration reduces the cost of each step considerably to $O(1)$, reducing the time complexity of Algorithm 2 to $n^{1+\frac{2}{d}+o(1)} \approx \Theta\left(n^{2.17}\right)$. It comes at a cost of increasing the space complexity to $n^{\frac{2}{d}+o(1)} \approx \Theta\left(n^{1.17}\right)$ however, as will be discussed in more detail in Section 6. For practical cluster sizes this is a small price to pay as the constant of the $n^{\frac{2}{d}+o(1)}$ term can be made to be quite small and still achieve a very substantial speed up. In Section 7 we will explain how the space requirement can be reduced to $O(n)$.

4.2 Exact Resetting

If the particle is not reset when it gets far away from the cluster then, even if steps are always taken to be of size equal to the distance to the cluster, the expected number of steps to get within distance ϵ of the cluster is infinite. This is because the logarithm of the distance from the origin is a martingale (see Equation (19) in the proof of Lemma 4.1) with quadratic variation growing at a bounded rate. This was dealt with in the basic algorithm above by resetting the particle to a uniformly random point on the starting circle whenever it gets too far from the cluster. This however incurs errors. A better approach is to use an analytic method to sample from the distribution of the point where the particle would next hit the starting circle if it were doing a true Brownian motion, and advance the particle to there. This method is described in [30] and in appendix A of [20], but we present another derivation here that we find more intuitive and better matches the method of sampling used in our code.

In this section (as in our code) we view $\mathbb{R}^2$ as the complex plane. The path of a Brownian motion is invariant under conformal maps so we apply conformal maps to simplify the situation. First apply a conformal map that sends the starting circle to the real line and the current position of the particle to the upper half plane, then scale and shift by real values to map the image of the current position to i. Now we can sample from the standard Cauchy distribution on $\mathbb{R}$, which is the hitting distribution of the real line by a Brownian motion from i, and map that via the inverses of our conformal maps to the starting circle.

If r is the radius of the starting circle and p is the current position of the particle which is outside of the starting circle, then this process works as follows. Define

$$f(z) = i\frac{z-r}{z+r}$$

and note that f is conformal as it is a Möbius map, that it maps the starting

circle to $\mathbb{R}$ and p to the upper half plane and finally that f^{-1} is given by

$$f^{-1}(z) = -r\frac{z+i}{z-i}.$$

The map that takes $f(p)$ to i is

$$g(z) = \frac{z - \Re\left(f(p)\right)}{\Im\left(f(p)\right)}$$

with inverse

$$g^{-1}(z) = \Re\left(f(p)\right) + \Im\left(f(p)\right) z.$$

Thus, if we now generate a standard Cauchy random variable X then $f^{-1}\left(g^{-1}\left(X\right)\right)$ has the correct distribution to be the point we advance our particle to.

This process uses only as much entropy as a single ordinary step and not much more CPU time. It therefore speeds up the algorithm, marginally, as well as making it more accurate.

4.3 An Exact Approach to Adhering

In all previous algorithms once the active particle gets within some distance of the cluster it takes a step that has a positive probability of overlapping with the cluster. After this step, if the particle has collided it is moved out by some rule so that it only touches the cluster. This process does not result in the same distribution of collision point as a true Brownian motion would. To sample from that true distribution we can use a process similar to that used in Section 4.2 to reset the active particle. This process works as follows.

Firstly we find the nearest particle in the cluster, whose centre we call q, and the centre to centre distance to the second nearest particle, which we call a. We wish to sample from the distribution of the hitting point, of a Brownian motion starting from p, on the union of the circle of radius $a-2$ centred at p and the circle of radius 2 centred at q. Call the intersection

points of the two circles in our union w_+ and w_-, and let θ be the acute angle formed at q by the line through p and the line through w_+. According to the cosine rule θ is given by

$$\theta = \cos^{-1}\left(\frac{(a-2)^2 - 4 - |q-p|^2}{4|q-p|}\right).$$

Let f_1 translate and rotate p to 0 and q to $(0, \infty)$ i.e.

$$f_1(z) = |q-p|\frac{z-p}{q-p}.$$

We can now see by trigonometry that

$$f_1(w_\pm) = |q-p| - 2\cos(\theta) \pm 2i\sin(\theta)$$

and so we can map f_1 of our boundary to a pair of rays from 0 using

$$f_2(z) = \frac{z - f_1(w_+)}{z - f_1(w_-)}.$$

Then the angle between these rays is given by $\phi = \arg\left(\frac{f_2(|q-p|-2)}{f_2(2-a)}\right)$ and we can now conformally map this pair of rays to the real line by applying

$$f_3(z) = \frac{z^\phi}{f_2(2-a)^\phi}.$$

Let $\alpha + i\beta = f_3(f_2(0))$, now letting X be a standard Cauchy random variable we can sample a new position for the particle by computing

$$f_1^{-1} \circ f_2^{-1} \circ f_3^{-1}(\alpha + \beta X).$$

This process is a little complicated but requires no more entropy and not much more computational time than an ordinary step, and removes any error that could be caused by taking finishing steps to be too large or from too far

away.

4.4 Taking Larger Steps

Algorithm 3: Generic Large Step DLA

Not Specified: How and when to reset, how to and size of step.
Input: Number of particles to add n
Output: $P \in (\mathbb{R}^2)^{n+1}$

```
1: P = [(0, 0)]
2: r = 2
3: for i in {1..n}
4:         Sample p ← C(r)
5:         Set fixed = False
6:         while not fixed
7:                 if |p| too large
8:                         Reset p to C(r)
9:                 else
10:                        Take a step    ▷ This may include adhering
11:                        If touching the cluster set fixed = True
12:                endif
13:        endwhile
14:        r = max{r, |p| + 2}
15:        Append p onto P
16: endfor
17: Return P                    ▷ The sequence of particle centres
```

The basic idea behind all previous attempts to reduce the number of steps per particle is the observation that we incur no error by increasing the size of each step so long as the step size is smaller than the current distance to the cluster. Algorithm 3 is a generic algorithm that all the following algorithms fit into and it allows us to state the following lemma. This lemma will be used for analysing the algorithm from [20] as well as the algorithm of [25] and our variation of it in Section 5. We call a step ordinary if it consists of moving the particle a fixed amount in a uniformly random direction and

can't collide with the cluster. We call a step a *δ-finishing step* if, conditional on the previous history, the probability of the partice adhereing with that step is at least δ.

Lemma 4.1. *If in Algorithm 3 the following conditions hold, then the number of steps and resets required for each particle is $O\left(\log^2(n)\right)$. There exists $C, \epsilon, \delta > 0$ independent of n satisfying the following conditions.*

- *The particle is reset, using the trick in Section 4.2, whenever it goes beyond the starting circle but is never reset otherwise.*
- *Whenever $d(p, P) < \epsilon + 2$ the next step is a δ-finishing step.*
- *Every step is either a δ-finishing step or an ordinary step of size between $(d(p, P) - 2)\, C$ and $d(p, P) - 2$.*

Proof. Note that the expected number of δ-finishing steps is $O(1)$, so it suffices to show that the expected number of steps before the next δ-finishing step is always $O\left(\log^2(n)\right)$. Further, as every reset is followed by a step, it suffices to show that the expected number of steps is $O\left(\log^2(n)\right)$.

Let S be the number of ordinary steps required before the next δ-finishing step and r be the radius of the starting circle. We will show that X_m, defined to be $\log\left(d(p, P) - 2\right)$ right before the $m+1$th step, and $Y_m := \left(\log(3r) - X_m\right)^2$ are such that for some $v > 0$,

- $\mathbb{E}\left(Y_{m+1} - Y_m | Y_m\right) \geq v$
- and $\mathbb{E}\left(\min\left(0, X_S - \log(\epsilon)\right)^2\right) = O(1 + \log(\epsilon)^2)$.

It follows from the final point that

$$\begin{aligned}
\mathbb{E}\left(Y_S\right) &= \mathbb{E}\left(\left(\log(3r) - \log(\epsilon) + \log(\epsilon) - \log\left(X_S\right)\right)^2\right) \\
&\leq 2\left(\log(3r) - \log(\epsilon)\right)^2 + 2\mathbb{E}\left(\min\left(0, X_S - \log(\epsilon)\right)^2\right) \\
&\leq 2\left(\log(3r) - \log(\epsilon)\right)^2 + O(1 + \log(\epsilon))
\end{aligned}$$

but as $\mathbb{E}(Y_m)$ grows by v with each step we must have $\mathbb{E}(Y_S) \geq v\mathbb{E}(S)$. Therefore $\mathbb{E}(S) \leq \left(2(\log(3r) - \log(\epsilon))^2 + O(1 + \log(\epsilon)^2)\right)/v$ and the result is attained.

X_m is mapped to X_{m+1} by the effects of one ordinary step followed possibly by a reset.

The possible reset can be simulated to arbitrary accuracy as follows. Whilst $d(p, 0) < 2r$, take an ordinary step with size equal to the distance to return to the starting circle. If p gets close enough to the starting circle then stop; otherwise, once $d(p, 0) \geq 2r$, reset the particle in the usual way. Note that resetting the particle in the usual way from a position with $d(p, 0) \geq 2r$ always decreases $\log(d(p, P) - 2)$, and as we only do that from positions with $d(p, 0) < 3r$, this must result in an increase in Y. Therefore, to show that the reset does not decrease Y it suffices to show that ordinary steps of any size less than $d(p, P)$ do not increase $\log(d(p, P) - 2)$ in expectation.

Let l be the step size, s be the step and q be the point in P closest to p before the step. Then as $d(p + s, P) \leq d(p + s, q)$ it suffices to show that

$$\mathbb{E}(\log(d(p + s, q) - 2)) = \log(d(p, q) - 2). \tag{19}$$

The expression on the left is given by the following integral.

$$\frac{1}{2\pi} \int_0^{2\pi} \log\left(\sqrt{(d(p, q) - 2 + l\sin(\theta))^2 + l^2\cos(\theta)^2}\right) d\theta$$

This expression is equal to $\log(d(p, q) - 2)$ and so the reset does not decrease Y.

Next we establish that Y_m grows by some $v > 0$ in expectation with each ordinary step of size at least $C(d(p, P) - 2)$, that cannot reach beyond distance $3r$ from the origin. Let p be the position of the incoming particle before the step, $p + s$ the position afterwards and q the point in P closest to p. We know that $d(p + s, P) \leq d(p + s, q) \leq 3r$ and that Equation (19) holds

so

$$\begin{aligned}\mathbb{E}\left(Y_{m+1}|Y_m\right) &= \mathbb{E}\left(\left(\log(3r) - \log\left(d(p+s,P)-2\right)\right)^2\right) \\ &\geq \mathbb{E}\left(\left(\log(3r) - \log\left(d(p+s,q)-2\right)\right)^2\right) \\ &= \left(\log(3r) - \log\left(d(p,q)-2\right)\right)^2 + \mathrm{Var}\left(\log\left(d(p+s,q)-2\right)\right) \\ &= Y_m + \mathrm{Var}\left(\log\left(d(p+s,q)-2\right)\right).\end{aligned}$$

This variance is guaranteed to be greater than some $v > 0$ as the ratio of $l = |s|$ to $d(p,q) - 2$ is bounded away from zero.

Finally, we must show that $\mathbb{E}\left(\min\left(0, X_S - \log(\epsilon)\right)^2\right) \leq 1$. Let $f_\epsilon(x) = \min\left(0, x - \log(\epsilon)\right)^2$. To complete the proof it suffices to show that $\mathbb{E}\left(f_\epsilon(X_{m+1})|S = m+1, p\right) \leq 1$ for all m and p. Note that as X_{m+1} is increasing in $d(p',q)$ and $\{d(p',q)-2 < \epsilon\} \subset \{S = m+1\}$ we have that

$$\mathbb{E}\left(f_\epsilon(X_{m+1})|S = m+1, p\right) \leq \mathbb{E}\left(f_\epsilon(X_{m+1})|d(p',q) - 2 < \epsilon, p\right).$$

If $l < d(p,q) - 2 - epsilon/2$ then $d(p',q) > 2 + \epsilon/2$ and $\log(d(p',q)-2) - \log(\epsilon))^2 < \log(2)^2$, thus we are done. Therefore we can assume that

$$l \geq d(p,q) - 2 - \frac{\epsilon}{2}. \tag{20}$$

Note that the step generating p' from p is an ordinary step with some step size l. Suppose w.l.o.g. that $q = (0,0)$ and $p = (d(p,q), 0)$. Let A be the arc of radius l about p that is within $2 + \epsilon$ of q and let a be sampled uniformly from A.

$$\mathbb{E}\left(f_\epsilon(X_{m+1})|d(p',q) - 2 < \epsilon, p\right) \leq \mathbb{E}\left(f_\epsilon(\log(d(a,q)-2))\right)$$

Let B be the result of projecting A onto the line $x = d(p,q) - l$. Let b be a uniformly random point on B, note that the distribution of $d(q,b)$ is

stochastically dominated by that of $d(q,a)$, therefore

$$\mathbb{E}\left(f_\epsilon(\log(d(a,q)-2))\right) \leq \mathbb{E}\left(f_\epsilon(\log(d(b,q)-2))\right).$$

This expectation can be written, in terms of the length y of B, as

$$\frac{2}{y}\int_0^{y/2} f_\epsilon(\log(\sqrt{x^2+(d(p,q)-l)^2}-2))dx.$$

As the arc A must be centred outside of the circle of radius $2+\epsilon$ about q, y must satisfy $y/2 \geq \frac{\sqrt{3}}{2}(2+\epsilon-(d(p,q)-l))$. Defining $z = d(p,q)-l$ and assuming the worst case (minimal) value of y gives

$$\mathbb{E}\left(f_\epsilon(\log(d(b,q)-2))\right) \leq \frac{1}{\sqrt{3}(2+\epsilon-z)}\int_0^{\sqrt{3}(2+\epsilon-z)} (\log(\sqrt{x^2+z^2}-2)-\log(\epsilon))^2 dx.$$

At least point we can invoke Equation (20) to note that $2 \leq z \leq 2+\epsilon/2$ and thus

$$\mathbb{E}\left(f_\epsilon(\log(d(b,q)-2))\right) \leq \frac{2}{\sqrt{3}\epsilon}\int_0^{\sqrt{3}\epsilon} (\log(\sqrt{x^2+2^2}-2)-\log(\epsilon))^2 dx.$$

Given that $\epsilon < 2$ we can bound the square root.

$$\begin{aligned}\mathbb{E}\left(f_\epsilon(\log(d(b,q)-2))\right) &\leq \frac{2}{\sqrt{3}\epsilon}\int_0^{\sqrt{3}\epsilon} (\log(x^2/6)-\log(\epsilon))^2 dx \\ &= \frac{2}{\sqrt{3}\epsilon}\int_0^{\sqrt{3}\epsilon} (\log(x^2/6\epsilon))^2 dx\end{aligned}$$

Into this integral we can substitute $u = x/\sqrt{6\epsilon}$ to get

$$\frac{2\sqrt{2}}{\sqrt{\epsilon}}\int_0^{\sqrt{\frac{\epsilon}{2}}} 4(\log(u))^2 du = 2((\log(\epsilon/2)-1)^2+4) = O(\log(\epsilon)^2+1).$$

This completes the proof. □

Remark 4.2. *In Section 9 we present empirical evidence that this is a tight upper bound.*

The straightforward way to take large steps is to always take steps of size $d(p, P) - 2$ until very close to the cluster and then switch to taking steps with a positive probability of finishing, which could be either ordinary steps slightly bigger than $d(p, P) - 2$, followed by a check for collision, or could be the exact finishing steps described in Section 4.3. The number of steps per particle for this algorithm is $O\left(\log^2(n)\right)$ and the cost per step with a naive implementation is $\Theta(n)$, so this gives a $O\left(n^2 \log^2(n)\right)$ algorithm.

We now look at the three more recent papers and their approaches to choosing step size. For all three of these analyses we assume the truth of Assumptions 3.1 and 3.3, conditional on which we can show the following. All three have $n^{1+\frac{2}{d}+o(1)} \approx \Theta\left(n^{2.17}\right)$ runtimes as described in the relevant papers, though by choosing parameters depending on n the first and third can be made to run in order $n^{\frac{8}{4+d}+o(1)} \approx n^{1.40}$ and order $n^{1+\frac{1}{d}+o(1)} \approx n^{1.58}$ time respectively (though further assumptions are required to prove this in the case of the first algorithm). The second may also have better asymptotic runtimes with parameter choices dependent on n, but they don't seem to have optimised their parameters for the values of n they look at. The algorithm proposed in the older paper [25] is analysed in Sections 5.1 and 7 as we recommend only a slight variation of it (this analysis is independent of all non-proven assumptions).

4.4.1 Menshutin and Shchur

The first algorithm we consider is given in [20] from 2006 by Menshutin and Shchur. They used a grid of squares to store their points. Their squares have side length twenty in their paper but we will call their side length L so we can discuss the effects of changing it more easily.

Each of their L by L grid squares has a 32-bit integer associated to it, indicating whether each of twenty five sub-squares within that square has a

particle in it. When they wish to know what size of step to take next, they explore outwards from the square containing the current value of p, checking every square in increasing order of distance from p until they come across a particle in the cluster. They then search the contents of each square that might contain the closest particle, allowing them to compute $d(p, P)$ and take steps with size $d(p, P) - 2$. They also optimize their algorithm by storing with each square the size of the previous step from that square so that all but the first step taken from that square can work from that distance inwards. We will compute an upper and lower bound on the runtime of this algorithm in terms of L and n.

We will bound the runtime of this algorithm conditional on the following two extra assumptions, the first of which roughly says that the radius of the cluster grows continuously as new particles are added.

Assumption 4.3. *Let a DLA cluster of n particles be grown and P_{start} be the set of points where particles first come within their corresponding r of the origin (i.e. the starting points of particles in this algorithm). If a grid of side length L is superimposed over the cluster, then the number of squares containing an element of P_{start} is*

$$\Theta\left(\min\left\{n, \frac{r_{final}^2}{L^2}\right\}\right)$$

Furthermore, for $\Theta\left(\min\left\{n, \frac{r_{final}^2}{L^2}\right\}\right)$ of those squares the first element of P_{start} inside that square is distance $\Omega(r)$ from the cluster as it was when that point was added.

This assumption makes sense as we expect the n starting points to be roughly evenly spread out. Thus, if the number of squares in the grid (within r_{final} of the origin) is much bigger than n, then we would expect the n starting points to be mostly in a square on their own, whilst if the number of squares is much smaller than n, then we would expect most of the squares to have a

starting point in them. Furthermore, due to Assumption 3.1 we expect most of these squares to be $\Omega(r)$ from the cluster.

Our second assumption says that the density of the cluster around any given particle is expected to be roughly the same as the global density.

Assumption 4.4. *For $1 \ll L \ll r_{final}$, most particles of the cluster are in a grid square that contains $L^{d+o(1)}$ particles.*

This assumption is reasonable as DLA clusters seem to be self-similar and have already been assumed to have dimension d on a global scale.

Firstly note that by Lemma 4.1 the total number of steps is $O\left(n \log^2(n)\right)$ and thus $\tilde{\Theta}(n)$.

Assumption 4.3 provides a lower bound on how many steps will be the first from the square they start from, and the total number of steps and the total number of squares provide an almost matching upper bound. Thus, we can conclude the number of such steps is $\tilde{\Theta}\left(\min\left\{n, \frac{r_{\text{final}}^2}{L^2}\right\}\right)$. Furthermore, $\tilde{\Theta}\left(\min\left\{n, \frac{r_{\text{final}}^2}{L^2}\right\}\right)$ of these will be distance $\Omega\left(r_{\text{final}}\right)$ from the cluster at the time, therefore the total number of squares that the algorithm has to read the bit for is $\tilde{\Theta}\left(\min\left\{n, \frac{r_{\text{final}}^2}{L^2}\right\} \frac{r_{\text{final}}^2}{L^2}\right)$.

Assumption 4.4 implies most of the particles (i.e. $\tilde{\Theta}(n)$ of them) are added to a square that already contains $L^{d+o(1)}$ particles. Over the course of growing the whole cluster this will require checking $\Omega\left(nL^{d+o(1)}\right)$ points for a collision. Thus the total runtime can be lower bounded by

$$\Omega\left(\min\left\{n, \frac{r_{\text{final}}^2}{L^2}\right\} \frac{r_{\text{final}}^2}{L^2} + nL^{d+o(1)}\right).$$

For constant L this clearly gives a lower bound of $\Omega\left(n^{1+\frac{2}{d}+o(1)}\right)$ on the runtime. This lower bound is minimised when L is chosen to be $n^{\frac{1}{d}\frac{4-d}{4+d}}$ when it becomes $\Omega\left(n^{\frac{8}{4+d}+o(1)}\right) \approx n^{1.40}$.

To get an upper bound on the runtime, we need an upper bound on the number of squares that we have to iterate over the contents of whilst finding

a large step size. The obvious bound is $\frac{r_{\text{final}}}{L}$ which gives a runtime of

$$O\left(\min\left\{n, \frac{r_{\text{final}}^2}{L^2}\right\}\frac{r_{\text{final}}^2}{L^2} + nr_{\text{final}}L^{d-1+o(1)}\right).$$

With an optimal choice of L this becomes $O\left(n^{\frac{8}{3+d}+o(1)}\right) \approx O\left(n^{1.70}\right)$. We consider it far more likely that the number of squares whose contents are iterated over in an average step is $O(1)$, though this would have to be taken as another assumption. Thus, we believe that this algorithm has a runtime of $n^{\frac{8}{4+d}+o(1)} \approx n^{1.40}$, if L is set appropriately.

Menshutin and Shchur do not provide any asymptotic analysis or runtimes for their code in their paper, presumably as they were only interested in finding an algorithm good enough for the simulations that they wanted to do at the time. However, it is not clear whether their choice of L was optimal or close to optimal, and it is impossible to compare the runtimes that they got in practice for small clusters with those achieved by the other algorithms discussed in this paper.

Menshutin and Shchur cite Ball and Brady [3] but say of the idea of a hierarchical model that "This approach seems rather memory consumptive: in growing a large cluster, it would become a bottleneck of an algorithm." This is a subjective statement, but we hope to convince the reader later that it is not a practical concern. The absence of the data from grids of larger mesh size is the reason their algorithm is not $O\left(n\text{polylog}(n)\right)$.

4.4.2 Alves, Ferreira and Martins

We now look at the algorithm given in [1] in 2008 by Alves, Ferreira and Martins. They cite Ball and Brady [3] and use three grids in a hierarchical structure, with squares of side length 4, 8 and 16 respectively.

Each square in each grid is labelled with a 1 if either it or any of the neighbouring eight cells contains the centre of a particle in the cluster and a 0 otherwise. To find the size for a step they check the grids in decreasing

order of mesh size. For each grid, if the square containing p is labelled with a 1 then they move onto the next grid, if it is labelled with a 0 then the step size is taken to be the side length of the square minus two.

If all three grids return the label 1, then the particle is close to the cluster and takes a step of size 1. Then a check to see if there has been a collision is performed. They present multiple methods for checking for collisions, finding the best method to be having a grid of squares of side length 4 and storing all the particles in each square in a place pointed to by that square (as described in Section 4.1). This allows the collision checks to be made in $O(1)$ time.

The labels need to be updated each time a point is added to the cluster but this requires updating only nine values on each grid for each particle added, and so only takes $O(1)$ time per particle.

The problem with this algorithm is that it never takes steps of size > 14. As a result, the argument of Lemma 3.2 (given Assumption 3.1) implies that this algorithm will run in $n^{1+\frac{2}{d}+o(1)} \approx \Theta\left(n^{2.17}\right)$ time.

This algorithm is very similar to the one in Section 5, but with a fixed number of grids of fixed mesh size. This is arguably not a problem so long as their choice for the number of grids is optimised for the cluster sizes they consider. However, the choice of sixteen as the largest mesh size is far too small for the sizes of cluster that are considered in the paper, a problem that can be seen quite clearly in Section 6. They present simulation runtimes suggesting a growth of $O\left(n^{1.4}\right)$, but this averages the increase in runtimes between $n = 10^3$ and $n = 10^6$. Looking at their increase in runtime between $n = 5 \times 10^5$ and $n = 10^6$ suggests runtimes are growing like $n^{1.98}$, and as n gets bigger this will approach the $n^{2.17}$ shown above. This algorithm fails to be competitive because it can never take large steps.

4.4.3 Kuijpers, Martín and van Ommen

We now look at the algorithm given in [27] in 2014 by Kuijpers, Martín and van Ommen.

They use only a single grid of squares with a side length of 1. Each square contains two integers: a first describing if there is a particle there and, if so, where in memory it is stored; and a second which is the distance from there to the nearest particle, if that distance is less than a parameter $D_{\max}$. Otherwise it is $D_{\max}$.

When a step is to be taken, it is chosen according to the distance to the nearest particle given in the relevant square. This process is very fast and gives a close to optimal step size, so long as the particle is within $D_{\max}$ of the cluster. There are two downsides to this approach. Firstly, most particles start $\Theta(r_{\text{final}})$ from the cluster. If this is much bigger than $D_{\max}$, then the step size will be too small and the walk will take a long time, $r_{\text{final}}^2 D_{\max}^{-2}$. Secondly, most insertions require updating $\Theta(D_{\max}^2)$ entries in the grid, so together the insertions take $\Theta(nD_{\max}^2)$. This leads to a trade off in the choice of $D_{\max}$.

For a fixed value of $D_{\max} = 140$, the analysis of numerical runtimes given in the paper claims the algorithm runs in $O(n^{1.08})$ time. For all the values of n they consider in the paper, $D_{\max} = 90$ is, by their own published runtimes, faster than taking $D_{\max} = 140$. Because $D_{\max} = 90$ is much faster for the smaller values of n, it gives worse apparent asymptotics.

The asymptotics for this algorithm are better if $D_{\max}$ is chosen depending on n. Then the trade off above gives a runtime of $\Theta\left(\frac{nr_{\text{final}}^2}{D_{\max}^2} + nD_{\max}^2\right)$ which, for an optimal choice of $D_{\max} = \sqrt{r_{\text{final}}}$, gives a runtime of $\Theta(nr_{\text{final}}) = n^{1+\frac{1}{d}+o(1)} \approx \Theta(n^{1.58})$.

They note that the space requirements for their algorithm are quite high and that this is due to the grid with squares of side length as small as 1. They suggest future work could use a larger mesh size to save memory. Whilst they don't draw attention to the fact, this would also improve runtimes as well due to having fewer updates to make after each insertion. These optimizations still wouldn't allow runtimes of $O(n^{1.08})$ though.

The runtimes for the largest value of n they both look at, 10^6, are better

by a factor of 5 than those in Alves et al. They achieve this as taking steps of size up to 140, or even 90, is a significant improvement over taking steps of size at most 14 when a particle is distance 1500 from the cluster. However, this is really an improvement of a constant factor and not an improvement to the exponent of the runtime.

5 The Recommended Algorithm

We now present the algorithm that we would recommend to anyone looking to do large scale simulations of DLA. It is available implemented in C++ at `github.com/mathsjames/fastDLA`. It is a variation on the algorithm in [25] and thus a generalisation of the lattice DLA algorithm proposed in [3] to the off-grid case. We use the method of generalizing the grids of [3] given in [1]; keep the particle from wandering far from the cluster with the technique described in Section 4.2 and taken from [30]; and deal with adhering particles to the cluster using the new technique described in Section 4.3. An overview of the algorithm is given as Algorithm 4 and the following few paragraphs describe it in words.

Firstly we choose $r_{\max}$ large enough that the cluster radius won't exceed it and then find $L = 2^{-a} r_{\max} = O(1)$ for some $a \in \mathbb{Z}$. We chose $L \in [12, 24]$ for all values of n in our implementation by trial and error to minimise runtime. The points of the cluster are stored in an array P with length $n + 1$. We have two types of grid, a *points grid* with mesh size $2L$ and a hierarchy of grids with mesh sizes $L2^k$ for each k from 0 to a. Each grid is realised as an array with each entry containing the data for one square. Each square in the points grid contains a vector of integers which are the indices in P of the points contained in that square. Each square in the other grids contains a single boolean variable indicating if there is a point either within that square or any of the neighbouring eight squares (This is how the boolean variables in [1] are set for each grid and unlike [25] where they indicated whether or

Algorithm 4: Fast Grid DLA

Input: Number of particles to add n
Output: $P \in (\mathbb{R}^2)^{n+1}$

```
P = [(0, 0)]
Initialise all grids
Reference (0, 0) in points grid
Mark grids for (0, 0)
r = 2
for i in {1..n}
    Sample p ← C(r)
    while True
        Binary search grids for step size
        if step size found
            Take step
        else
            Find nearest and second nearest points
            Take finishing step
            if touched cluster
                breakwhile
            endif
        endif
        if d(0, p) > r then reset to C(r)
    endwhile
    r = max {r, |p| + 2}
    Append p onto P
    Reference p in points grid
    Mark grids for p
endfor
Return P                ▷ The sequence of particle centres
```

not that square alone contained a particle, this only affects runtime by a constant factor).

The cluster is grown one particle at a time. Each particle starts at a uniformly selected point on a circle, with radius r, containing the cluster so far. The particle then proceeds to step as follows: if the particle is outside

the starting circle, return it to the starting circle as described in Section 4.2. Otherwise, if the hierarchy of grids provides a valid step size as described in the next paragraph, take a step in a uniform direction of that size. If not, the particle is close to the cluster and we can search the few surrounding points grid squares to find the nearest two particles in the cluster and perform a finishing step as described in Section 4.3.

To find a step size from the hierarchy of grids, we wish to find the grid of largest mesh size which contains a 0 in the square corresponding to the particle. The returned step size will then be the mesh size minus two to account for the size of the particles. We can find this grid by binary searching because if any grid contains a 1 in the relevant square so will every grid of larger mesh size. If no such grid exists, then the particle is close to the cluster and this process fails to provide a valid step size.

Once a finishing step results in the particle touching the cluster, we add the particle to the cluster. This requires adding it to the array P, adding a reference to it in the points grid and updating the other grids.

Once the final particle is added, P is an array containing all of the particle positions in the order they were added to the cluster.

5.1 Complexity

The following is an analysis of the time and space complexity of the recommended algorithm, only Assumption 3.3 is needed here and only for the analysis of the initialisation phase and the space requirements. The alterations detailed in Section 7, while we don't recommend them in practice, remove the need for this assumption whilst providing better asymptotics.

Due to the fact that $L = O(1)$, the space requirements of the algorithm are $O\left(r_{\max}^2\right) \approx O\left(n^{1.17}\right)$ due to having to store the finest grids. Note that each grid in the hierarchy uses a quarter as much memory as the next finest grid, so the entire hierarchy only requires four thirds as much memory as the finest grid on its own. This is why we think [20] was wrong to dismiss the

hierarchical structure due to memory concerns. In Section 7, we will show how to reduce this to $O(n)$. However, for values of n up to 10^8, the array P uses about as much memory as the grids do and the changes would incur increases to the constants both for space and runtime.

The time complexity of initialising the grids is of course $O\left(r_{\max}^2\right) \approx O\left(n^{1.17}\right)$. This is reduced to $O\left(1\right)$ by the changes proposed in Section 7. However, initialisation accounts for less than 1% of the runtime for a cluster of size 10^8 so in practice this is negligible and only the growing of the cluster matters.

Each of n particles takes $O\left(\log^2(n)\right)$ steps by Lemma 4.1, for each step all computations run in $O\left(1\right)$ time except for the binary search of the $O\left(\log(n)\right)$ grids which takes $O\left(\log\log(n)\right)$ time. When each particle is fixed, there are $O\left(\log\left(n\right)\right)$ grids to update, and each one only takes $O(1)$ time to update. Thus the runtime after initialisation is $O\left(n\log^2(n)\log\log(n)\right)$.

6 Numerical Assessment

The table below shows the runtimes, in seconds, for the algorithm given here and the two most recent works considered in this paper. Alves et al. used a 3.0GHz processor, Kuijpers et al. a 3.1GHz processor and our simulations were run on a 3.4GHz processor. While these clock speeds are similar, due to the ten years of development in processor design between the Alves et al. work and this work, caution should be taken when comparing these runtimes.

CPU runtimes in seconds

n	This work	Kuijpers et al. [27]	Alves et al. [1]
10^3	0.018	0.093	0.10
10^4	0.18	1.0	2.4
10^5	2.0	12	30
10^6	26	313	1560
5×10^6	158	3398	-
10^7	336	-	-
10^8	4230	-	-

These results show clearly that this approach beats the other recent suggestions (with published runtimes) for small values of n, as well as asymptotically. It is unclear how well the algorithm in [20] performs as they do not publish runtimes. The next table shows memory usage for the algorithm given here and the minimum amount required to hold the particles in the cluster i.e. the output.

Space requirements

n	This work (MB)	Minimum (MB)	Others
10^3	0.093	0.015	-
10^4	0.31	0.15	-
10^5	2.6	1.5	-
10^6	28	15	14MB Ossadnik [25]
5×10^6	134	76	-
10^7	325	153	$>$ 16GB Kuijpers et al. [27]
10^8	3956	1535	-

None of the three recent papers say how much memory their programs used. However, Kuijpers et al. [27] say that they do not include runtimes for growing clusters of size 10^7, as their 16GB machine didn't have enough memory. Ossadnik [25] says that his entire program uses only 14MB of

memory which is lower than what we claim to be the minimum. We suspect the reason for this discrepancy is that he was working with single precision floating point rather than double. If rewritten to use single precision, our algorithm would require about 21MB of memory rather than 28MB.

It is difficult to compare our runtimes to those of Ossadnik [25], who reports using about 9 hours of CPU time to generate a 10^6 particle cluster which we can produce in 26 seconds. He reports using a Sun SPARCstation 1, which had a 20MHz processor, suggesting that he required a little over 7 times the number of clock cycles that we do. There were almost three decades of processor development between these two measurements and he had to optimize for memory usage, so this number should not be taken too seriously. However combined with the fact that he was working in single precision we take this as a good indicator that our code is competitive.

7 Reducing Asymptotic Memory Usage

We now present the idea explained in [25] and generalised from [3] for reducing the memory usage of the algorithm. As is noted in Section 5.1 only $O(n\text{polylog}(n))$ time is needed after initialisation, and therefore at least asymptotically almost all memory will never be written to during the growing of the cluster. The idea is to avoid ever assigning most of that memory at all.

To that end we will structure the memory as a tree, with the points grid identified with one of the other grids, typically the finest one, so that it fits into the structure. Each cell in a grid that isn't the finest contains, along with its boolean value, four pointers to the locations in memory that the four cells beneath it are stored. If a cell and all of its descendants have not had data written to them since the initialisation phase then that cell does not need to be stored. Its parent will have a null pointer instead of a pointer to its location and this will tell the executing program that whatever value it

was looking for in that branch of the tree is zero.

At initialisation only the root node needs to be created which is a $O(1)$ operation. The use of binary search when trying to find the next step size is no longer possible as in order to read a value from one of the grids one has to follow a length $\log(n)$ path down the tree. Writing to an entry also requires navigating a path through the tree however, as the tree is only ever written to when adding a particle and every level of the tree in the direction of the particle is written to when doing so. This doesn't worsen the bound we presented above. Overall this increases the runtime of the main section of the algorithm to $O\left(n\log^3(n)\right)$, which is also the order of the runtime of the whole algorithm.

We now show that the space used is $O(n)$. We refer to the 5×5 sub-grid of a grid centred at a certain cell as the surroundings of that cell. There are $O(1)$ cells in each grid whose surroundings contain the origin. Now consider the grid with cells of side length l. A cell only has to be allocated memory if a particle is inserted into it or one of the neighbouring eight cells. For cells that don't have the origin within their surroundings, that requires there to be a touching sequence of particles stretching from outside of the surroundings to a neighbouring cell. As these must cross a cell, the surroundings must contain at least $\frac{l-2}{2}$ particles. Thus, as there are n particles, the grid can have at most $\frac{50n}{l-2}+O(1)$ occupied cells. As we go through the grids from finest to least fine, the value of l grows exponentially and thus summing this bound over all $O\left(\log(n)\right)$ grids gives a $O(n)$ bound on the total number of cells that need allocating. Clearly the number of particle positions that need storing is n and everything else is $O(1)$. So we are done.

The order of the runtime of the algorithm has been increased, though in Section 9 we will explain how this could probably be removed. The reason that we are wary of this approach is that the apparent need to store multiple pointers in each cell applies a massive constant to the space requirement that more than negates the reduction in order for realistic values. In Section 9,

we will also discuss how this could be optimised and we can tell from the results presented in [25] that Ossadnik must have managed to implement this in a way to save space. However one would have to be careful with the implementation and it might necessarily entail a significant cost to runtime.

8 Higher Dimensions

Most of the ideas in these algorithms generalise straight-forwardly to higher dimensions.

The technique for exact resetting generalises to higher dimensions fine with the use of a higher dimensional Möbius map. However the exact finishing step does not generalize easily as it requires the use of a non-Möbius conformal map, of which there are none in higher dimensions.

For fixed dimension, the process of finding a step size still runs in time $O(\log\log(n))$ by the same argument and the step size is still bounded below by some positive factor of the maximum step size. However the number of steps that are required per particle can't be shown to be $O\left(\log^2(n)\right)$ in the same way. Indeed we would not expect this to be true in sufficiently high dimension. The problem is that higher dimensional random walks are transient and so we suspect that when the co-dimension of the cluster gets too large the Brownian motion could move around amongst the cluster, going around fingers for a very long time before it collides.

9 Future Improvements

As the output of two-dimensional DLA is $O(n)$, the space and runtime requirements must be at least $O(n)$. Thus there is no room for improving the order of the space requirements. Any attempt to improve the constants significantly is likely to depend on the details of the implementation including the language and hardware used to implement the algorithm, and so we do

not consider this here.

The runtime of both the algorithm recommended here and the tree based version can probably be reduced to $O\left(n\log^2(n)\right)$ using the following observation. The binary search, for finding a step size, makes no use of our knowledge of the previous step size. We could instead start our search from the grid corresponding to the previous step size and then search linearly from that grid through finer or rougher grids. The number of grids searched is of the order of the number of grids between the one corresponding to the previous step size and the one corresponding to the current step size. Heuristically, if we consider a particle at a uniform point inside a grid cell of side length $2^k l$ taking a step of size $l-2$, the probability of this particle leaving the cell is $< 2^{2-k}$. Thus, we suspect that the expected number of grids to search in order to find the next step size and thus the runtime for finding the step size will be $O(1)$. However, we do not have a rigorous proof of this fact.

Reducing the remaining $log^2(n)$ seems harder. Within the paradigm of making steps in a uniformly selected direction, with step size independent of direction, there is a limit to the amount of improvement that would be possible. We suspect though, that this has not yet been achieved. Figure 3 is a log-log plot of the average number of steps required to add the kth particle to a cluster against $\log(k)$ for both the algorithm recommended here and an algorithm that takes maximal step sizes every time. For the algorithm described here, the curve seems to be settling towards a gradient of 2, suggesting that the upper bound of $O\left(\log^2(n)\right)$ on the number of required steps is tight. For the algorithm taking maximally sized steps though, the gradient seems to be 1, thus indicating that only $O\left(\log(n)\right)$ steps are required. This indicates that there is something to be gained from taking larger steps and the question is then how to go about quickly finding the maximal step size or a closer approximation. One possible idea would be to have even more grids with mesh sizes between those already used, though we have not explored this idea here. Between the finest grids this would require a lot of memory,

but between all but the finest grids it could be quite cheap. As the grids no longer fit together nicely, the monotonicity property that the binary search (and the improvement suggested above) relies on will not work. Therefore, how to search these grids efficiently and the trade off between space, step size and time to find each step size becomes complicated. We doubt that trying to exploit this idea is worth the effort, but it is the best approach we have thought of.

Another interesting and related question which has a significant impact on what is possible in the next part of this thesis, is how to efficiently simulate Hastings-Levitov clusters. The straight-forward approach for this task takes $O(n^2)$ time and we are unaware of any way to achieve a lower order runtime. We are of the opinion that a $O(n\text{polylog}(n))$ runtime algorithm for simulating HL(2) would be the biggest contribution to the numerical simulation of 2-dimensional growth processes since the algorithm of [3] in 1985.

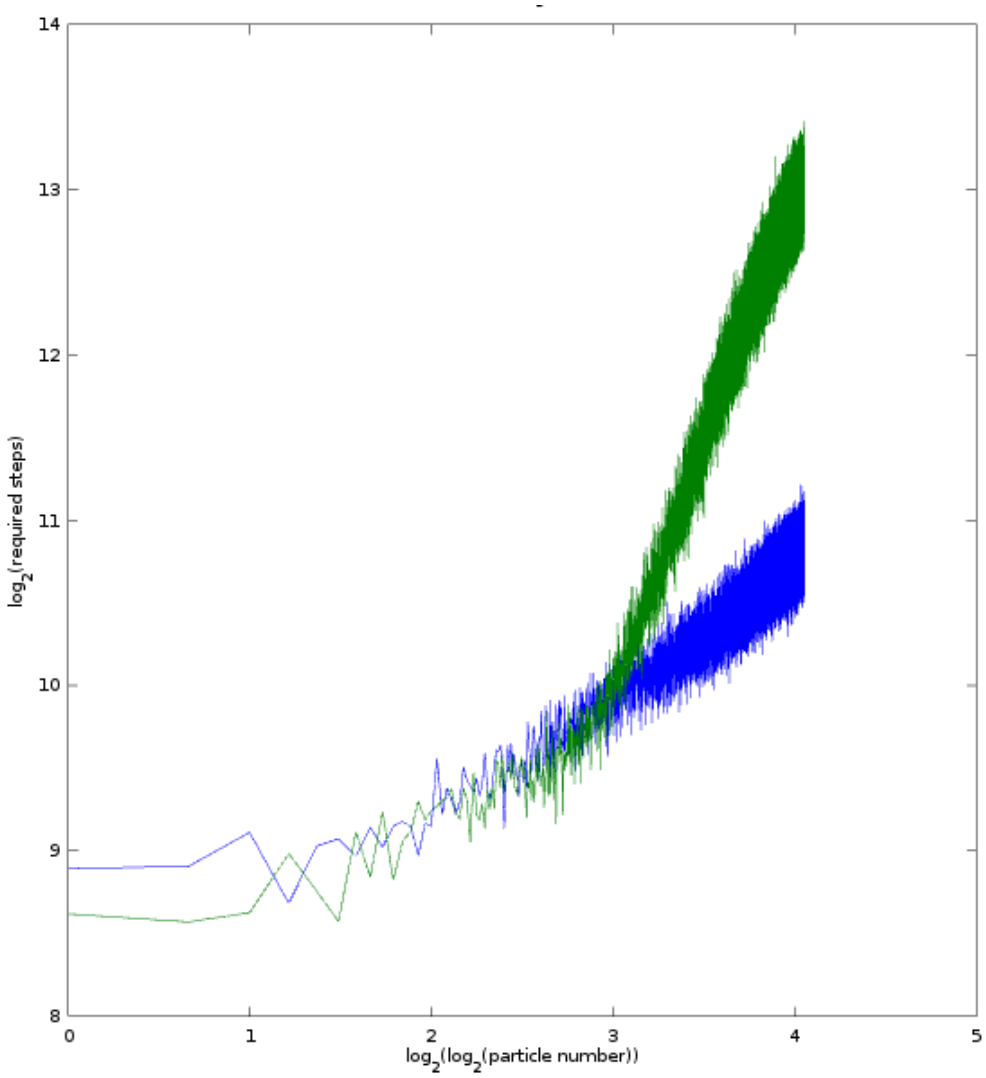


Figure 3: A plot of the number of steps required to add each particle in a cluster of 10^5 particles, summed over 100 clusters. Green is for the algorithm recommended here, blue is for an algorithm that always takes the maximum allowable step.

Part III

Is Hastings Levitov a Good Model of Diffsion Limited Aggregation?

1 Introduction

In this part we heuristically and numerically investigate whether three different models of diffusion limited aggregation converge to the same limiting shape. We look at Meakin's version of the off-grid diffusion limited aggregation model [18] (refered to as DLA or off-grid DLA), the noise reduced version introduced by Ball et al. in [4] and the Hastings-Levitov model with $\alpha = 2$ introduced in [12].

[4] considered the growth rate of the radius of each of these clusters and found that their noise reduced clusters settle into their asymptotic growth faster, thus making a better model for exploring asymptotic properties numerically. However, they do not verify that these models match asymptotically except in the growth rates of their radii.

Hastings-Levitov clusters have received more attention, largely due to hopes that it might be possible to prove theorems about them. Results have been proven about HL clusters with $\alpha = 0$ [23] [31] and for regularized versions of HL with an extra parameter σ [35]. Simulations of HL have been conducted in [4] and [22], however they have been held back compared to simulation of other models due to the lack of a sub quadratic time algorithm, meaning that the simulated clusters have previously had at most 10^5 particles. A variation of the Hastings-Levitov model was proposed by Rohde and Zinsmeister [29] and will be briefly discussed later.

In Section 2 we will define the models under consideration. In Section 3 we provide a heuristic discussion together with pictures of clusters to illustrate our conclusions where they are clearly visible. In Section 4 we will describe the methodology for statistically testing some of the questions which remain unclear after Section 3, and in Section 5 we present the results of these tests. A discussion of what might be interpreted from these results beyond that which was explicitly tested for is then provided in Section 6.

2 Definition of the Models

Off-lattice or off-grid diffusion limited aggregation (DLA) is a model in which a cluster of circular particles of radius one is grown. To begin with the cluster consists of a single particle at the origin. Particles are added to the cluster one at a time. Each particle does a Brownian motion started from infinity until it touches the cluster, at which point it is added to the cluster and will never move from that position. Let $C(r)$ denote the circle of radius r. The process of DLA and how to simulate it was described Part II.

This process can straight forwardly be generalised to particles with different sizes. Given any distribution $\mathcal{D}$ on $[0, \infty)$, one can draw the size of each particle independently from $\mathcal{D}$ before setting it off from infinity. We denote this process as $\mathrm{DLA}_{\mathcal{D}}$, see [9] for an analysis of simulations of this model.

Hastings and Levitov proposed a one-parameter family of two-dimensional growth processes based on conformal maps. This family is denoted $HL(\alpha)$ for $\alpha \in [0, \infty)$. Let $\mathbb{D}$ be the closed unit disc in $\mathbb{C}$ and $\Delta = \mathbb{C} \backslash \mathbb{D}$ its complement. To construct a Hastings-Levitov cluster we must first choose a family of particle maps $f_c : \Delta \to \Delta$, where c can take any value in $\mathbb{R}_{\geq 0}$. This family must have the following properties.

- f_c is conformal.
- ∞ is a fixed point of f_c.

- The derivative of f_c at ∞ is equal to e^c, that is the log-capacity of f_c should be equal to c.
- $\Delta \backslash f_c(\Delta)$ is a contiguous set contained in a $o(1)$ radius ball around 1 as $c \to 0$, this should be thought of as a single particle attached to the unit disc at 1.

For an angle $\theta \in [0, 2\pi]$, we now define $f_c^\theta : z \to e^{i\theta} f_c\left(ze^{-i\theta}\right)$. This map attaches a particle to the unit disc at $e^{i\theta}$.

Given a sequence $(c_k)_{k=1}^n$ of capacities and a sequence $(\theta_k)_{k=1}^n$ of attachment angles, which we will explain how to choose momentarily, we define

$$\Phi_k = f_{c_1}^{\theta_1} \circ f_{c_2}^{\theta_2} \circ \ldots \circ f_{c_k}^{\theta_k}$$

and the corresponding cluster of n particles as

$$\mathbb{C} \backslash \Phi_n(\mathbb{D}).$$

Where the attachment point of the kth particle is therefore given by

$$\Phi_{k-1}\left(e^{i\theta_k}\right).$$

The θ_k in this model are chosen independently and uniformly at random from $[0, 2\pi)$. Thus, due to the conformal invariance of harmonic measure, the attachment point of the kth particle is distributed according to harmonic measure from infinity on the boundary of the cluster of the first $k-1$ particles i.e. the distribution of the hitting point of Brownian motion from infinity. Therefore, it is natural to expect this to be a good model of DLA so long as the particles are all the same size in some appropriate sense. However, as the kth particle is distorted by Φ_{k-1}, we need to choose c_k in a way that is dependent on Φ_{k-1}. Hastings and Levitov reasoned that the particle will be enlarged by a factor of approximately $\left|\Phi_{k-1}'\left(e^{i\theta_k}\right)\right|$ by Φ_{k-1}, and so they

chose to set

$$c_k = \frac{c}{\left|\Phi'_{k-1}\left(e^{i\theta_k}\right)\right|^\alpha}$$

for some fixed value of c, and when $\alpha = 2$ they suggested this should give a model for DLA. The extent to which this holds and attempts to vary this model are discussed in Section 3.

Another generalisation of DLA was proposed in [4] called noise reduced diffusion limited aggregation. Here, particles are circles of radius one and are added to the cluster one at a time exactly as in DLA except that, after they touch the cluster, but before being fixed in place, they first "sink into" the particle they made contact with. That is to say, if p is the position of the particle at the moment it makes contact, and q is the position of the particle it makes contact with, then the final position for the centre of the added particle is $Ap + (1 - A)q$ for some parameter $A \in (0, 1)$. We denote this process NRDLA(A). In [4] it is shown by simulation that the growth of the radius of NRDLA(A) clusters for $A < 1$ settles into the believed asymptotic rate much faster than for regular DLA.

3 HL Particle Maps and Regularization

Multiple choices for the particle maps f_c have been used, but we will focus on two families here. The first is that of slit particles which Hastings and Levitov used for their simulations in the case $\alpha = 0$ and which Viklund, Sola and Turner used in their work on regularizing HL(α) [35]. The second is of circular particles which are similar to the bump particles used by Hastings and Levitov in their original work for the $\alpha = 1.5$ and $\alpha = 2$ case.

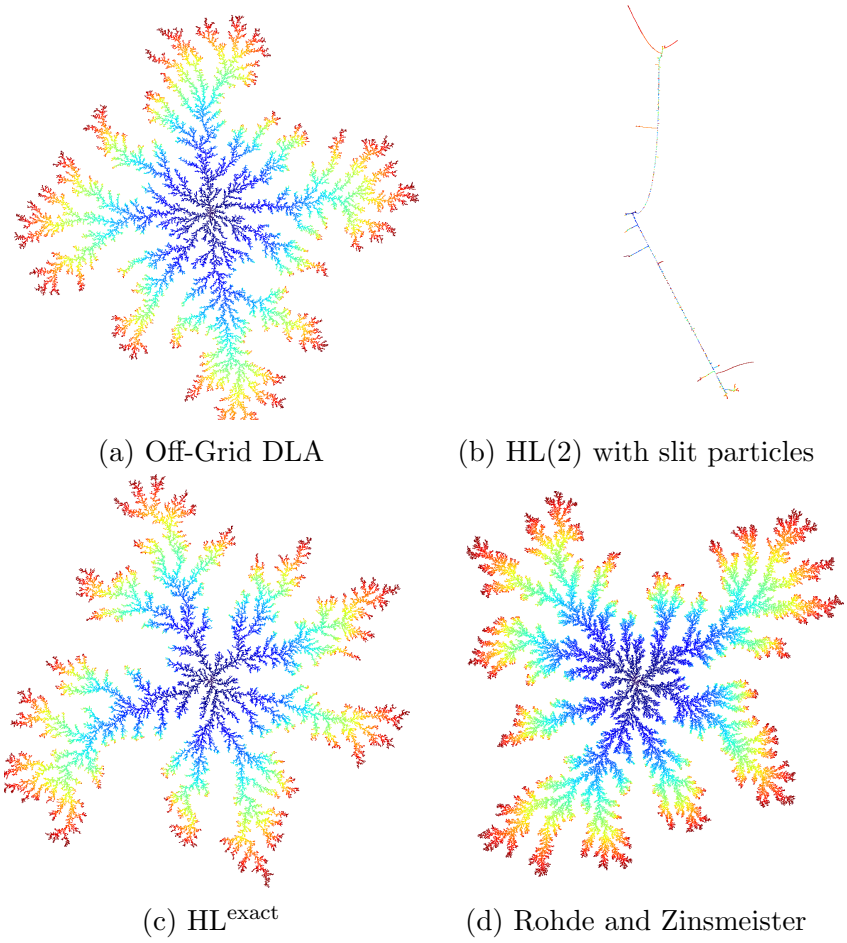
3.1 Slit Particles

A particle map for a slit particle is fixed by the fact that $f_c(\Delta) = \Delta \backslash (1, 1+d]$ for some $d(c)$, thus the particle that is added is a one dimensional line (or

curve if it is distorted by the maps of previous particles). We will write d_k for $d(c_k)$ copying [35]. We observe, as in [35], that the actual length of the kth particle is given by

$$\int_1^{1+d_k} \left|\Phi'_{k-1}\left(re^{i\theta_k}\right)\right| dr = d_k \left|\Phi'_{k-1}\left(r_0 e^{i\theta_k}\right)\right| \tag{21}$$

for some $r_0 \in [1, 1+d_k]$.



(a) Off-Grid DLA

(b) HL(2) with slit particles

(c) $\mathrm{HL}^{\mathrm{exact}}$

(d) Rohde and Zinsmeister

Figure 4: Diffusion Limited Aggregation

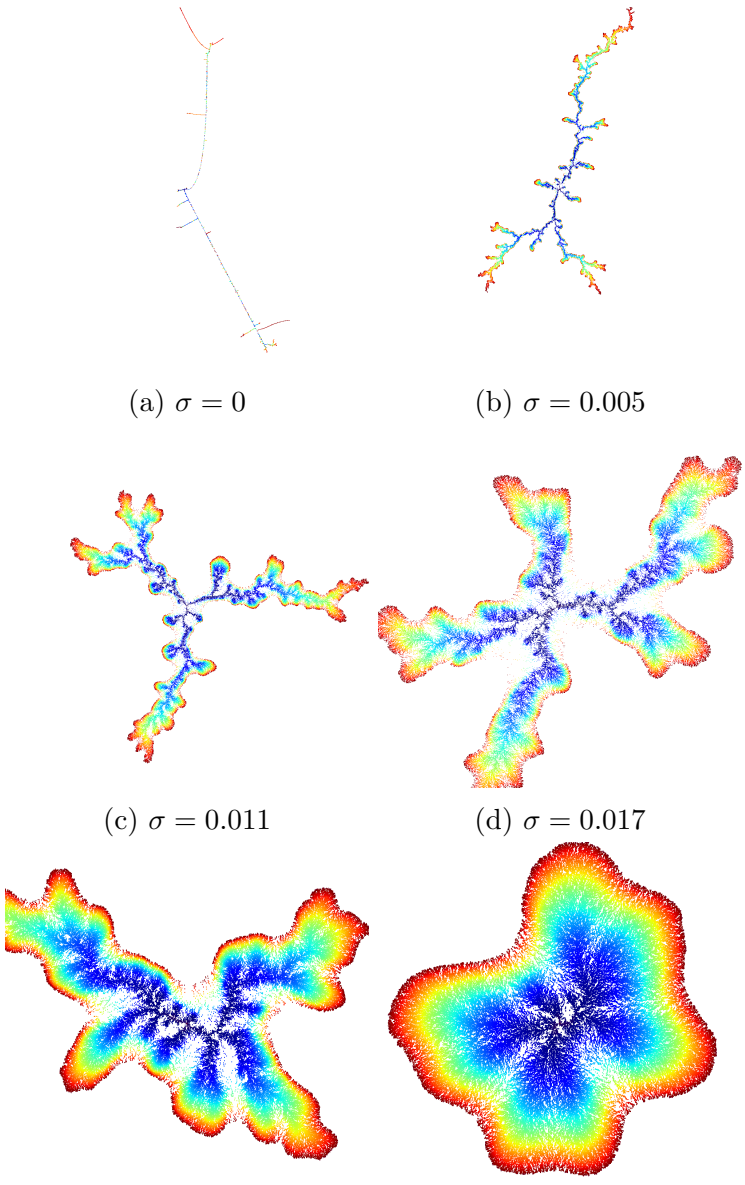
Figure 4 shows three clusters of 10^5 particles. One was generated by

DLA and another by HL(2) with slit particles. We hope it is clear that these clusters are very different. In the HL(2) case some of the particles have macroscopic size i.e. on the order of the size of the cluster. The idea of the HL(2) model is that $\left|\Phi'_{k-1}\left(e^{i\theta_k}\right)\right|$ will be a good enough approximation to $\left|\Phi'_{k-1}\left(r_0 e^{i\theta_k}\right)\right|$. This has failed because the derivative of Φ_{k-1} can change very quickly near the boundary of Δ and in particular can approach 0. In order to explore whether this is the only issue, we simulate a model where each d_k is chosen by a numerical method so that the distance from the base of each particle to the tip is always the same. The third cluster in Figure 4 is generated from this model which we call $\mathrm{HL}^{\mathrm{exact}}$. This cluster looks a lot more like DLA and so there is hope that an appropriate method for choosing the d_k (and thus the c_k) whilst avoiding the complexity of the implicit scheme in $\mathrm{HL}^{\mathrm{exact}}$, could be a good model for DLA. We call the value of d_k that $\mathrm{HL}^{\mathrm{exact}}$chooses for a given cluster of $k-1$ particles, $\hat{d}_k$. The fourth cluster in Figure 4 is generated by the method of Rohde and Zinsmeister. In this model the rescaling factor for the size of the particle is chosen to be the inverse of the derivative at the tip of the particle, as opposed to the base as in Hastings Levitov. This is simpler than $\mathrm{HL}^{\mathrm{exact}}$however it is still an implicit scheme and so does not fulfil the above goal. It is not obvious and is unaddressed here whether or not this model converges to the same limiting shape as DLA.

One place to look for such a model is [35] which defines a regularized version of HL(α), called HL(α,σ). σ is called the regularization parameter. The model is the same as Hastings-Levitov except the c_k are chosen according to the rule

$$c_k = \frac{c}{\left|\Phi'_{k-1}\left(e^{i\theta_k+\sigma}\right)\right|^{\alpha}}.$$

The idea of this σ is to avoid measuring the derivative on the boundary where it can take extreme values and thus hopefully avoid the problem of macroscopic particles. It is stated in [35] that $\sigma \asymp d_1$ is a natural choice of σ due to Equation (21). This makes sense in the small particle limit which they consider, i.e. where $c \asymp \frac{1}{n}$, because they take $cn \asymp 1$ and thus could



(a) $\sigma = 0$ (b) $\sigma = 0.005$

(c) $\sigma = 0.011$ (d) $\sigma = 0.017$

(e) $\sigma = 0.023$ (f) $\sigma = 0.060$

Figure 5: Hastings-Levitov with $\alpha = 2$ and various σ, note that only the base point of each particle is marked

reasonably expect some average of d_k to be $\asymp d_1$. More generally, one could choose $\sigma \asymp \mathbb{E}\left(d_n|\Phi_{n-1}\right)$ when generating a cluster of n particles (note this is an implicit definition). Figure 5 shows a collection of clusters for various values of σ. Whilst they do show non-trivial geometric behaviour they are clearly not the same shape as DLA.

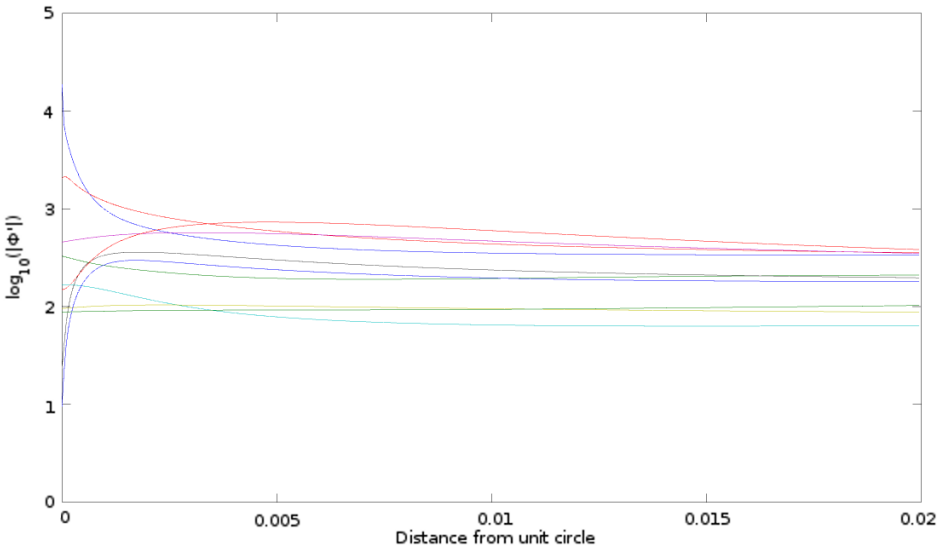


Figure 6: $\log_{10}$ of absolute value of derivative along radial lines

To understand why this doesn't work as a model for DLA take a look at the plot in Figure 6. It shows how the derivative of an instance of Φ_{10^4} (from $\mathrm{HL}^{\mathrm{exact}}$) changes along ten rays from the origin as the distance from $\mathbb{D}$ increases (note the logarithmic scale on the y-axis). If $e^{i\theta_{10001}}$ was on one of these rays, our task in $\mathrm{HL}^{\mathrm{exact}}$would be to find the value of d_{10001} such that the length of the particle, i.e. the integral in Equation (21) of the functions plotted, is d. In the $HL(2)$ case, these functions are approximated by their values on $\mathbb{D}$. Amongst the ten rays shown here, one of them has derivative ≈ 10 on $\mathbb{D}$ rising very quickly to > 200. This particle would therefore end up roughly twenty times larger than it should be. More extreme values will occur as more samples beyond these ten are taken.

In the $HL(2, \sigma)$ case, the behaviour of the cluster depends on σ. If it is

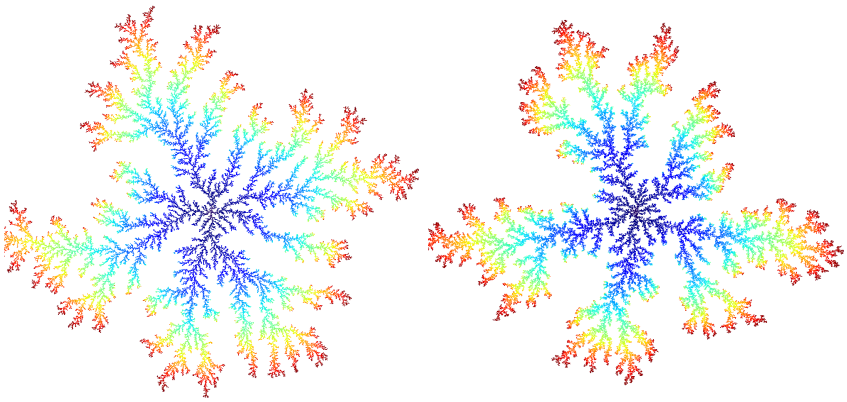
small compared to the possible values of $\hat{d}_k$, then the growth will proceed like $HL(2)$, with many oversized particles leading to long thin fingers. If it is large compared to the possible values of $\hat{d}_k$, then the differences in the derivatives near $\mathbb{D}$ will not be seen and so they will not be corrected for. The particle capacities are then close to deterministic, leading to $HL(0)$ like behaviour where the gaps are filled in and the cluster ends up roughly circular. This observation is made precise and proved in [35]. The most interesting case is when σ is amongst the possible values of $\hat{d}_k$, with a non-trivial shape emerging. However, this growth is not very DLA like. The issue is that the spread of possible values for $\hat{d}_k$ extends over orders of magnitude. Therefore in some places (e.g. deep inside gaps), σ is relatively large and we get an $HL(0)$ style smoothing effect, whilst in other areas (e.g. the ends of fingers), σ is relatively small, meaning some particles end up very significantly oversized which leads to smoothing on a global scale.

It seems that any method that tries to imitate DLA by choosing the d_k as a function of Φ'_{k-1}, must measure this derivative at a point that is distance of order $\hat{d}_k$ from $\mathbb{D}$. As the cluster gets larger and more intricate, the spread of values of $\hat{d}_k$ will extend over more orders of magnitude, therefore, at least naively, the number of measurements taken must be unbounded to find the right level to look at. This unbounded number of measurements is the approach used in our implementation of $\mathrm{HL}^{\mathrm{exact}}$. However, whilst we can not rule out the existence of a method for approximating $\hat{d}_k$ using a bounded number of measurements, there is a much simpler approach. The problem with $HL(2)$ was that $|\Phi'_{k-1}(z)|$ could be very close to zero for z on the boundary of $\mathbb{D}$, but increase away from zero very quickly as z moves away from $\mathbb{D}$. In the next subsection, we argue that all that is required to avoid this is to take a particle map whose derivative is bounded away from zero. Then, $HL(2)$ will be a good model for DLA. From this we believe it is clear that attempting to regularize $\mathrm{HL}(\alpha)$ with slit particles, whilst providing interesting behaviour to study for its own sake, is not a good approach to

trying to model DLA.

3.2 Circular Particles

A particle map for a circular particle is fixed by the fact that $f_{c_k}(\Delta) = \Delta \backslash C_k$, where C_k is a disc whose boundary meets the unit circle at right angles and whose centre is in $\mathbb{R}_{>0}$. The correspondence between the radius r_k of C_k and the capacity c_k is non-trivial. However, to get the correct behaviour for small particle sizes, it suffices to assume either that $c_k \propto r_k^2$ or, asymptotically equivalently, that the distance a_k from 1 to the furthest part of the particle satisfies $c_k \propto a_k^2$. Thus, we relax the definition of HL(2) and take $a_k = \frac{a_0}{\left|\Phi'_{k-1}\left(e^{i\theta_k}\right)\right|}$ in our simulations. Figure 7 shows two clusters of 10^5 particles, one from this model and one from off-grid DLA. This model will be tested statistically later on, however, we believe there are good heuristic reasons to think that this is a good model for DLA which we outline in the rest of this section.



(a) Off-Grid DLA (b) HL(2) with circular particles

Figure 7: Diffusion Limited Aggregation

The important point is that for the asymptotic shape to match that of DLA, there is no need to require every particle to have the same size, even

asymptotically. We expect it to be sufficient for the expected size of the particle to be very close to independent of the global shape of the cluster and this particle's global location, and for the size distribution to be concentrated enough that all particles are small compared to the cluster. With slit particles the concentration failed, as when a new particle attaches near the tip of a previous particle, $|\Phi'|$ is measured on the boundary as being very close to zero compared to its value on the rest of the resulting particle. This causes macroscopic particles to arise. In the circular particle case, this doesn't happen as the derivative of the particle map is bounded away from zero. Of course, in this case, the size of the particle is determined by the cluster so far and the particle's location. However, we claim that it is reasonable to suppose that, asymptotically, the distribution of the size of a particle is determined by the shape of the cluster within any $\omega(1)$ ball around the attachment point. This is because the asymptotic effect of macroscopic features of the cluster outside of this ball on the derivatives near the particle, is just going to be to change them all by the same factor. Thus, the global location and shape are not important. The histogram in Figure 8 shows the empirical distribution of the sizes of particles from HL(2) with circular particles. It is clearly highly concentrated, with not one of the 10^5 particles exceeding twice the median size.

We will call the empirical distribution shown in Figure 8 $\mathcal{D}$. We will later simulate both $\mathrm{DLA}_{\mathcal{D}}$with this distribution and also simulate a version of $\mathrm{HL}^{\mathrm{exact}}$with circular particles, where a numerical scheme is used to set the resulting length of the kth particle, defined as $\left|\Phi_{k-1}\left(e^{i\theta_k}\right) - \Phi_{k-1}\left(f_{c_k}^{\theta_k}\left(e^{i\theta_k}\right)\right)\right|$, to an independent sample from $\mathcal{D}$, which we denote $\mathrm{HL}_{\mathcal{D}}^{\mathrm{exact}}$.

4 Methods

We wish to test a variety of different models to see if there is a difference in the distribution of the shape of the resulting cluster. We do this by

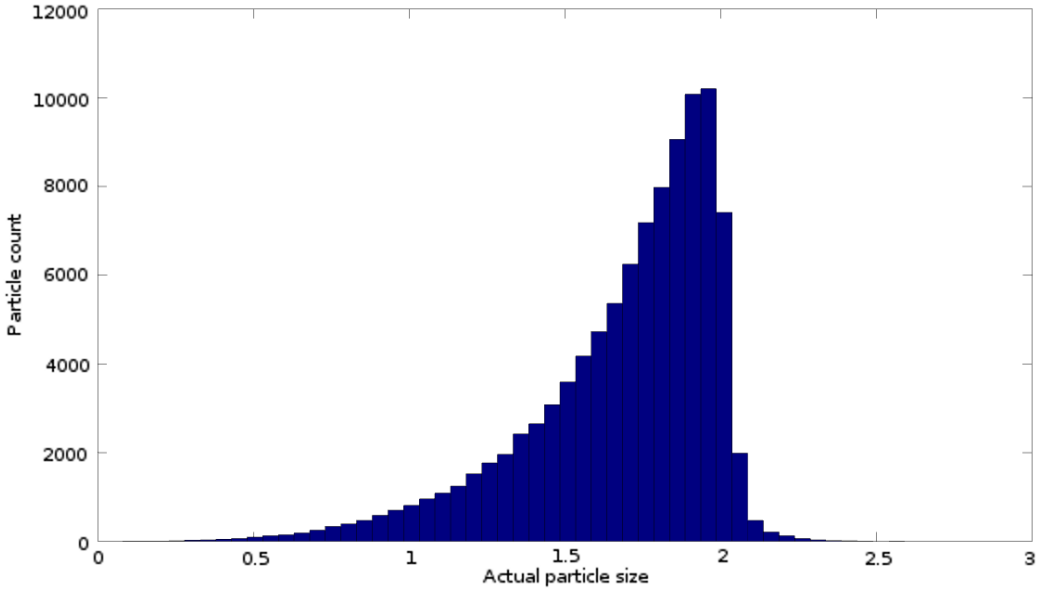


Figure 8: A histogram the actual size of every tenth particle from ten clusters of 10^5 particles generated from HL(2), a similar picture for clusters of 10^4 particles was indistinguishable by eye.

simulating many clusters from each model, projecting each cluster into a Euclidean vector space and then using a test for equality of distribution in Euclidean space proposed in [33].

4.1 The Test Used

The models produce different sizes of clusters, but we wish to disregard this information. Therefore we normalise all clusters to have the same radius. The different kinds of cluster are very easy to tell apart on the microscopic level. We are not interested in these differences so it is important to choose a projection that disregards the microscopic information. To achieve this goal we first rescale every cluster so that the outermost centre, or in the case of Hastings Levitov type models, the base point, of any particle is distance 1 from the origin. Then we replace each cluster with the set of points within

distance 0.04 of a centre or base point. This results in a normalised and fattened cluster. Finally, the projection itself consists of taking the empirical distribution function of the distance from a random point in the disc of radius 1.1 to this new cluster using 10000 sampled points. This gives us a CDF for each cluster, which we can consider as an element of a function space with the L^2 norm.

The numbers in the previous paragraph are to a fair extent arbitrary. The important property of 0.04 is that it is much smaller than 1. Therefore, whilst it is much bigger than the ratio of the particle radius to the cluster radius, so does smother the microscopic details, it isn't so large that it smothers the entire cluster. The important property of 1.1 is merely being bigger than 1.04, without being so much bigger that a significant fraction of the noise from the sampling is a result of the distance of the sampled points from the circle containing the cluster. Finally, the power of the test increases to a limit with the number of points sampled to construct each CDF. We chose 10000 as this gets a good approximation of the actual CDF for each cluster, without taking too long to compute.

We now have collections of functions drawn from different distributions of a Euclidean function space, and so can apply the method of [33], which we now describe. If $f_1, ..., f_{n_1}$ and $g_1, ..., g_{n_2}$ are elements of a Euclidean vector space drawn independently from two distributions $\mathcal{D}_f$ and $\mathcal{D}_g$ which we wish to test for equality, the energy $\mathcal{E}_{n_1,n_2}$ is defined to be

$$\frac{n_1 n_2}{n_1 + n_2} \left(\frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{m=1}^{n_2} \|f_i - g_m\| - \frac{1}{n_1^2} \sum_{i,j=1}^{n_1} \|f_i - f_j\| - \frac{1}{n_2^2} \sum_{l,m=1}^{n_2} \|g_l - g_m\| \right).$$

This statistic is guaranteed to be non-negative and, under the null hypothesis that the distributions are equal, has a limiting distribution as n_1 and n_2 converge to infinity with n_1/n_2 converging away from 0 [33]. Furthermore, the test which rejects the null hypothesis given $\mathcal{E}_{n_1,n_2} > c_\alpha$, is consistent in

that limit [33].

Unfortunately, the limiting distribution is not straightforward to calculate, and it is more complicated to understand the speed of convergence to it. As a result, the authors of [33] recommend using a permutation test, which we do for each pair of models that we consider.

In order to give a clearer picture of what is going on here, we also compute the energy distance between the two samples given by

$$\sqrt{\frac{n_1+n_2}{n_1 n_2}\mathcal{E}_{n_1,n_2}}.$$

For large n_1, n_2, this is approximately $\sqrt{\mathbb{E}\left(2\|f_1-g_1\|-\|f_1-f_2\|-\|g_1-g_2\|\right)}$.

4.2 Models Tested and Sample Sizes

At first the models we tested were DLA, $\text{DLA}_{\mathcal{D}}$, HL(2) (with circular particles), HL^{exact}(with slit particles) and $\text{HL}^{\text{exact}}_{\mathcal{D}}$(with circular particles). The distribution $\mathcal{D}$ was the empirical distribution of the actual sizes of particles from HL(2) clusters (which were not otherwise used in any analysis). In all of these cases we used 10^5 particles per cluster.

As a basic rule, we chose the sample sizes to be proportional to the reciprocal of the square root of the computation time of a single instance of the model, as this maximises the power attained for a fixed level of computational resources. However, we deviated from that basic rule by doubling the sample sizes for HL(2) and DLA as we thought these cases were the most interesting. The resulting sample sizes are given in the following table.

Model	DLA	$\text{DLA}_{\mathcal{D}}$	HL(2)	HL^{exact}	$\text{HL}^{\text{exact}}_{\mathcal{D}}$
Number of Clusters	16000	5000	3000	500	500

After seeing the results of these tests and reading [21] and [4] we decided to add two more categories to our experiments. Firstly, as finite size effects had been found to be significant in DLA for millions of particles in [21],

and our results so far had shown DLA to be far from the Hastings-Levitov type models, we added DLA clusters with 10^8 particles to see if these would be substantially closer to the Hastings-Levitov clusters. We also simulate NRDLA(0.03) clusters with 10^5 and 10^7 particles, as the radius of these clusters settles into its asymptotic behaviour much faster than in the case of regular DLA, in order to test if they converge faster toward some universal limiting shape. The sample sizes for these three cases are given in the next table.

Model	DLA $n = 10^8$	NRDLA(0.03) $n = 10^5$	NRDLA(0.03) $n = 10^7$
Number of Clusters	3000	5000	3000

Finally, having seen the results of all of those experiments, we also simulated HL clusters with 10^4 and 10^6 particles to get a clearer picture of whether HL was converging towards the same limit as the others, sampling 3000 and 200 cases respectively.

5 Results

The following table contains the value of the energy between different cluster types.

Energies	$\text{DLA}_{\mathcal{D}}$	HL	HL^{exact}	$\text{HL}_{\mathcal{D}}^{\text{exact}}$	DLA 10^8	NRDLA	NRDLA 10^7	HL 10^4	HL 10^6
DLA	2.56	5.42	1.12	1.27	44.1	28.9	48.3	19.5	1.42
$\text{DLA}_{\mathcal{D}}$	-	2.00	0.488	0.519	21.8	11.2	24.5	22.0	0.735
HL	-	-	0.0157	0.0361	14.4	4.49	16.3	17.4	0.450
HL^{exact}	-	-	-	0.0175	3.79	1.02	4.30	5.26	0.294
$\text{HL}_{\mathcal{D}}^{\text{exact}}$	-	-	-	-	3.54	0.885	4.05	5.72	0.257
DLA 10^8	-	-	-	-	-	10.1	0.0990	59.0	0.495
NRDLA	-	-	-	-	-	-	11.7	40.2	0.389
NRDLA 10^7	-	-	-	-	-	-	-	62.1	0.629
HL 10^4	-	-	-	-	-	-	-	-	4.23

This table gives the corresponding p-values.

p-values	$\text{DLA}_{\mathcal{D}}$	HL	HL^{exact}	$\text{HL}_{\mathcal{D}}^{\text{exact}}$	DLA 10^8	NRDLA	NRDLA 10^7	HL 10^4	HL 10^6
DLA	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}
$\text{DLA}_{\mathcal{D}}$	-	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}
HL	-	-	0.750	0.268	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}
HL^{exact}	-	-	-	0.670	10^{-5}	10^{-5}	10^{-5}	10^{-5}	5×10^{-5}
$\text{HL}_{\mathcal{D}}^{\text{exact}}$	-	-	-	-	10^{-5}	10^{-5}	10^{-5}	10^{-5}	7×10^{-5}
DLA 10^8	-	-	-	-	-	10^{-5}	0.00794	10^{-5}	10^{-5}
NRDLA	-	-	-	-	-	-	10^{-5}	10^{-5}	10^{-5}
NRDLA 10^7	-	-	-	-	-	-	-	10^{-5}	10^{-5}
HL 10^4	-	-	-	-	-	-	-	-	10^{-5}

Finally, this table gives the energy distances between the samples.

Distances	$\text{DLA}_{\mathcal{D}}$	HL	HL^{exact}	$\text{HL}_{\mathcal{D}}^{\text{exact}}$	DLA 10^8	NRDLA	NRDLA 10^7	HL 10^4	HL 10^6
DLA	0.0259	0.0463	0.0481	0.0511	0.132	0.0871	0.138	0.0878	0.0846
$\text{DLA}_{\mathcal{D}}$	-	0.0326	0.0328	0.0338	0.108	0.0670	0.114	0.108	0.0618
HL	-	-	0.00606	0.00918	0.0979	0.0489	0.104	0.108	0.0490
HL^{exact}	-	-	-	0.00837	0.0940	0.0473	0.100	0.111	0.0454
$\text{HL}_{\mathcal{D}}^{\text{exact}}$	-	-	-	-	0.0909	0.0441	0.0973	0.116	0.0424
DLA 10^8	-	-	-	-	-	0.0734	0.00812	0.198	0.0514
NRDLA	-	-	-	-	-	-	0.0789	0.147	0.0450
NRDLA 10^7	-	-	-	-	-	-	-	0.204	0.0579
HL 10^4	-	-	-	-	-	-	-	-	0.150

6 Conclusion

The table of p-values shows that all of the different kinds of cluster studied could be distinguished with p-value < 0.01, except for the different kinds of HL clusters with the same number of particles. This indicates that we have the necessary resolution, and that the different types of cluster are still distinguishable. To understand the data, we turn to looking at the distances rather than the p-values, as they provide more information.

Looking at the distances for the first five types of cluster we can see that, for HL, randomizing the particle sizes or changing between different particle shapes (so long as all particles are microscopic) makes much less difference

to the shape of the resulting cluster than randomizing the particle sizes for ordinary DLA.

The smallest distance (excluding those between different types of HL with 10^5 particles) was between DLA, with 10^8, and NRDLA, with 10^7, particles. These are the two kinds that were expected to be closest to the limit, so this is a good indication that DLA and NRDLA are converging to the same limit. The distances from NRDLA 10^7 to HL with 10^4, 10^5 and 10^6 particles were respectively 0.204, 0.104 and 0.0579, which strongly suggests that HL is also converging to the same place.

As all of these models are converging to the same limiting shape, another interesting question is whether they are approaching from the same direction. To answer this, for each pair of models (excluding those with NRDLA 10^7 or DLA 10^8) we have embedded the three-point metric space, consisting of those three models with distances given by the distance table, into Euclidean space. We then computed the angle formed by joining each of the pair with a line to NRDLA 10^7. The largest resulting angle was $0.619 < \frac{\pi}{5}$ between NRDLA 10^5 and DLA 10^5. We think this is quite a small angle given that the space is reasonably high dimensional and the expected angle formed by random directions would be $\frac{\pi}{2}$.

In conclusion, we believe that HL(2) with circular particles, NRDLA and DLA all converge to the same limiting shape, and that they do so from the same direction to first order.

References

[1] Sidiney Alves, Silvio Ferreira, and M L. Martins. Strategies for Optimize Off-Lattice Aggregate Simulations. *Brazilian Journal of Physics*, 38, 04 2008.

[2] Richard Alejandro Arratia. *Coalescing Brownian Motions on the Line.* ProQuest LLC, Ann Arbor, MI, 1979. Thesis (Ph.D.)–The University

of Wisconsin - Madison.

[3] R C Ball and R M Brady. Large Scale Lattice Effect in Diffusion-Limited Aggregation. *Journal of Physics A: Mathematical and General*, 18(13):L809, 1985.

[4] Robin Ball, Neill E. Bowler, Leonard M. (Leonard Michael) Sander, and Ellák Somfai. Off-Lattice Noise Reduction and the Ultimate Scaling of Diffusion-Limited Aggregation in Two Dimensions. *Physical Review E*, Vol.66(No.2), August 2002.

[5] Itai Benjamini and Ariel Yadin. Upper Bounds on the Growth Rate of Diffusion Limited Aggregation. 05 2017.

[6] Nathanaël Berestycki, Christophe Garban, and Arnab Sen. Coalescing Brownian Flows: A New Approach. *arXiv e-prints*, page arXiv:1307.4313, Jul 2013.

[7] Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[8] Neill Bowler and Robin C Ball. Off-Lattice Noise Reduced Diffusion-Limited Aggregation in Three Dimensions. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 71:011403, 02 2005.

[9] F. L. Braga, O. A. Mattos, V. S. Amorin, and A. B. Souza. Diffusion Limited Aggregation of Particles with Different Sizes: Fractal Dimension Change by Anisotropic Growth. *Physica A Statistical Mechanics and its Applications*, 429:28–34, July 2015.

[10] L. R. G. Fontes, M. Isopi, C. M. Newman, and K. Ravishankar. The Brownian Web: Characterization and Convergence. *Ann. Probab.*, 32(4):2857–2883, 2004.

[11] Denis Grebenkov and Dmitry Beliaev. How anisotropy beats fractality in two-dimensional on-lattice diffusion-limited-aggregation growth. *Physical Review E*, 96, 10 2017.

[12] Matthew B Hastings and Leonid S Levitov. Laplacian Growth as One-Dimensional Turbulence. *Physica D: Nonlinear Phenomena*, 116(1-2):244–252, 1998.

[13] Harry Kesten. Upper Bounds for the Growth Rate of DLA. *Physica A: Statistical Mechanics and its Applications*, 168(1):529 – 535, 1990.

[14] Vitalii Konarovskyi. A System of Coalescing Heavy Diffusion Particles on the Real Line. *Ann. Probab.*, 45(5):3293–3335, 09 2017.

[15] Yves Le Jan and Olivier Raimond. Flows, Coalescence and Noise. *Ann. Probab.*, 32(2):1247–1315, 2004.

[16] Yen Lee Loh. Bias-free simulation of diffusion-limited aggregation on a square lattice. 07 2014.

[17] P Meakin. The Structure of Two-Dimensional Witten-Sander Aggregates. *Journal of Physics A: Mathematical and General*, 18(11):L661, 1985.

[18] Paul Meakin. Diffusion-Controlled Cluster Formation in 2—6-Dimensional Space. *Phys. Rev. A*, 27:1495–1507, Mar 1983.

[19] Paul Meakin. Progress in DLA Research. *Physica D: Nonlinear Phenomena*, 86(1):104 – 112, 1995. Chaos, Order and Patterns: Aspects of Nonlinearity - @'The Gran Finale@'.

[20] A. Y. Menshutin and L. N. Shchur. Test of Multiscaling in a Diffusion-Limited-Aggregation Model Using an Off-Lattice Killing-Free Algorithm. *Physical Review E.*, 73(1):011407, January 2006.

[21] Anton Menshutin. Scaling in the Diffusion Limited Aggregation Model. *Physical review letters*, 108:015501, 01 2012.

[22] F Mohammadi, A A Saberi, and S Rouhani. Scaling and Multiscaling Behavior of the Perimeter of a Diffusion-Limited Aggregation Generated by the Hastings–Levitov Method. *Journal of Physics: Condensed Matter*, 21(37):375110, 2009.

[23] James Norris and Amanda Turner. Hastings–Levitov Aggregation in the Small-Particle Limit. *Communications in Mathematical Physics*, 316(3):809–841, 2012.

[24] James Norris and Amanda Turner. Weak Convergence of the Localized Disturbance Flow to the Coalescing Brownian Flow. *Ann. Probab.*, 43(3):935–970, 2015.

[25] Peter Ossadnik. Multiscaling Analysis of Large-Scale Off-Lattice DLA. *Physica A: Statistical Mechanics and its Applications*, 176(3):454 – 462, 1991.

[26] Peter Ossadnik. Multiscaling Analysis and Width of the Active Zone of Large Off-Lattice DLA. *Physica A: Statistical Mechanics and its Applications*, 195(3):319 – 323, 1993.

[27] Kasper R. Kuijpers, Lilian de Martín, and J. Ruud Van Ommen. Optimizing Off-Lattice Diffusion-Limited Aggregation. *Computer Physics Communications*, 185:841–846, 03 2014.

[28] L. C. G. Rogers and David Williams. *Diffusions, Markov Processes, and Martingales. Vol. 2.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1987. Itô calculus.

[29] Steffen Rohde and Michel Zinsmeister. Some Remarks on Laplacian Growth. *Topology and its Applications*, 152:26–43, 07 2005.

[30] Evelyn Sander, Leonard Sander, and Robert Ziff. Fractals and Fractal Correlations. *Computers in Physics*, 8:420–425, 07 1994.

[31] Vittoria Silvestri. Fluctuation Results for Hastings–Levitov Planar Growth. *Probability Theory and Related Fields*, 167(1-2):417–460, 2017.

[32] Daniel W. Strook and S.R.S. Varadhan. *Multidimensional Diffusion Processes*. Classics in Mathematics. Springer-Verlag Berlin Heidelberg, 2006.

[33] Gabor Szekely and Maria Rizzo. Testing for Equal Distributions in High Dimension. *InterStat*, 5, 11 2004.

[34] Boris Tsirelson. Nonclassical Stochastic Flows and Continuous Products. *Probab. Surv.*, 1:173–298, 2004.

[35] Fredrik Johansson Viklund, Alan Sola, and Amanda Turner. Small-Particle Limits in a Regularized Laplacian Random Growth Model. *Communications in Mathematical Physics*, 334(1):331–366, 2015.

[36] T. A. Witten and L. M. Sander. Diffusion-Limited Aggregation, a Kinetic Critical Phenomenon. *Phys. Rev. Lett.*, 47:1400–1403, Nov 1981.