HOW LONG DOES IT TAKE FOR INTERNAL DLA TO FORGET ITS INITIAL PROFILE?

LIONEL LEVINE AND VITTORIA SILVESTRI

September 11, 2018

ABSTRACT. Internal DLA is a discrete model of a moving interface. On the cylinder graph $\mathbb{Z}_N \times \mathbb{Z}$, a particle starts uniformly on $\mathbb{Z}_N \times \{0\}$ and performs simple random walk on the cylinder until reaching an unoccupied site in $\mathbb{Z}_N \times \mathbb{Z}_{\geq 0}$, which it occupies forever. This operation defines a Markov chain on subsets of the cylinder. We first show that a typical subset is rectangular with at most logarithmic fluctuations. We use this to prove that two Internal DLA chains started from different typical subsets can be coupled with high probability by adding order $N^2 \log N$ particles. For a lower bound, we show that at least order N^2 particles are required to forget which of two independent typical subsets the process started from.

CONTENTS

1. INTRODUCTION

Internal Diffusion-Limited Aggregation (IDLA) models a random subset of $\mathbb{Z}^d$ that grows in proportion to the harmonic measure on its boundary seen from an internal point. More precisely, let $A(0) = \{0\}$ denote the subset ("cluster") at time $t = 0$. For integer times $t \geq 1$ inductively define

$$A(t) = A(t-1) \cup \{Z_t\}, \tag{1}$$

_____

where Z_t denotes the exit location from $A(t-1)$ of a simple random walk on $\mathbb{Z}^d$ starting from 0, independent of the past. The process $(A(t))_{t \geq 0}$ is a Markov chain on the space of connected subsets of $\mathbb{Z}^d$. As $t \to \infty$, the asymptotic shape of $A(t)$ is an origin-centered Euclidean ball [11] and its fluctuations from the ball are at most logarithmic in dimensions $d \geq 2$ [2, 4, 7, 10]; see also [3] for a lower bound on the fluctuations. Space-time averages of the fluctuations converge to a logarithmically correlated Gaussian field [8, 9].

In the setting of the cylinder graph [9] it was asked what happens if the process is initiated with a cluster $A(0)$ other than a singleton: How long does it take for IDLA to forget the shape of $A(0)$? Our main results, Theorems 1.3 and 1.4 below, give upper and lower bounds that match up to a log factor.

Fix an integer $N \geq 3$, and let $\mathbb{Z}_N \times \mathbb{Z}$ denote the cylinder graph with the cycle on N vertices as base graph. We refer to $x \in \mathbb{Z}_N$ as the horizontal coordinate, and to $y \in \mathbb{Z}$ as the vertical coordinate. For $k \in \mathbb{Z}$, we call $\{y = k\} := \mathbb{Z}_N \times \{k\}$ the kth level of the cylinder, and $R_k := \{(x,y) : y \leq k\}$ the infinite rectangle of height k.

It is sometimes convenient to formulate the growth in terms of aggregation of particles. Let $A(0)$ be the union of the lower half-cylinder R_0 with a finite (possibly empty) set of sites in the upper half-cylinder. At time zero, each site in $A(0)$ is occupied by a particle. At each discrete time step, a new particle is released from a uniform point on level 0, and performs a simple random walk until reaches an unoccupied site, which it occupies forever. Motion of particles is instantaneous: we do not take into account how many random walk steps are required for a particle reach an unoccupied site, but rather increment the time by 1 after it does so. Formally, we define $A(t)$ inductively according to (1), where Z_t is the the exit location from $A(t-1)$ of a simple random walk on the cylinder graph starting from a uniform random site on level 0, independent of the past. Note that this is equivalent to adding a new site to the cluster according to the harmonic measure on its exterior boundary seen from level $-\infty$. When $A(0) = R_0$ we say that the process is "starting from flat".

In the cylinder setting there are two parameters, the size N of the base graph, and the time t. Just as large IDLA clusters on $\mathbb{Z}^d$ are logarithmically close to Euclidean balls, it is natural to expect large IDLA clusters on the cylinder to be logarithmically close to filled rectangles. When $t = N^2$ this was stated in [9] (but not proved there, as the proof method is the same as in [7]). Our first result extends this result to large times $t \leq N^m$, which we will later use to control the fluctuations of stationary clusters.

Theorem 1.1. *Let $(A(t))_{t \geq 0}$ be an IDLA process on $\mathbb{Z}_N \times \mathbb{Z}$ starting from the flat configuration $A(0) = R_0$. For any $\gamma > 0$, $m \in \mathbb{N}$ there exists a constant $b_{\gamma,m}$, depending only on γ, m, such that*

$$\mathbb{P}\Big(R_{\frac{t}{N} - b_{\gamma,m} \log N} \subseteq A(t) \subseteq R_{\frac{t}{N} + b_{\gamma,m} \log N} \quad \forall t \leq N^m \Big) \geq 1 - N^{-\gamma} \tag{2}$$

for N large enough.

We prove the above result in two steps. To start with, we argue that (2) holds for $T \leq (N \log N)^2$, which can be shown by adapting the Jerison, Levine and Sheffield arguments [7] to the cylinder setting (cf. Theorem 5.1). We then invoke the Abelian

property of IDLA (cf. Section 2) to build large clusters by piling up nearly rectangular blocks of $\mathcal{O}(N^2 \log N)$ particles each.

Suppose now that the IDLA process on $\mathbb{Z}_N \times \mathbb{Z}$ is not initiated from the flat configuration R_0, but rather from an arbitrary connected cluster $A(0) \supset R_0$. How long does it take for the process $A(t)$ to forget that it did not start from flat? Clearly, the answer to this question very much depends on $A(0)$. For example, it will take an arbitrarily large time to forget an arbitrarily tall initial profile. On the other hand, most profiles are unlikely for IDLA dynamics. This leads us to the following related question:

How long does it take for IDLA to forget a <u>typical</u> initial profile?

To define "typical" let

$$\Omega := \{A \subseteq \mathbb{Z}_N \times \mathbb{Z} : A = R_0 \cup F \text{ for some finite } F\}$$

denote the set of clusters which are completely filled up to level 0. This is the state space of IDLA. On Ω we introduce the following shift procedure: each time the cluster is completely filled up to level $k > 0$, we shift the cluster down by k.

Definition 1.1 (Shifted IDLA)**.** Let $\mathcal{S}$ be the map from the space of IDLA configurations to itself defined as follows. For a given cluster A, let

$$k_A := \max\{k \geq 0 : R_k \subseteq A\}$$

be the height of the maximal filled (infinite) rectangle contained in A. The *downshift* of A is the cluster

$$\mathcal{S}(A) := \{(x, y - k_A) : (x, y) \in A\},$$

Note that $k_{\mathcal{S}(A)} = 0$, so $\mathcal{S}(\mathcal{S}(A)) = \mathcal{S}(A)$. Now from the IDLA chain $(A(t))_{t \geq 0}$ we set

$$A^*(t) := \mathcal{S}(A(t))$$

for all $t \geq 0$. We will refer to $(A^*(t))_{t \geq 0}$ as the shifted process associated to $(A(t))_{t \geq 0}$.

The shifted process defines a new Markov chain on the same configuration space Ω. While the original IDLA chain is transient, we will see that Shifted IDLA is a positive recurrent Markov chain (cf. Remark 6.1) and it thus has a stationary distribution, which we denote by μ_N.
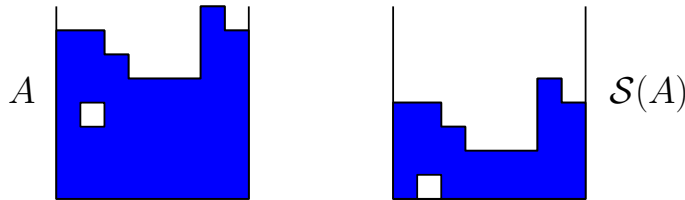

FIGURE 1. An IDLA cluster A and its shifted version $\mathcal{S}(A)$.

Recall from [1] that a probability distribution is called a *warm start* for a given Markov chain if it is upper bounded by a constant factor times the stationary distribution of the chain. We relax this definition below.

Definition 1.2 (Lukewarm start). For $k \in \mathbb{N}$, a probability measure ν_N on Ω is said to be a *k-lukewarm start* (for Shifted IDLA) if

$$\nu_N(A) \leq N^k \mu_N(A) \qquad \forall A \in \Omega.$$

We say that ν_N is a lukewarm start if it is a k-lukewarm start for some $k \in \mathbb{N}$.

Definition 1.3 (Typical clusters). A random cluster $A \in \Omega$ is said to be a *typical cluster* if it is distributed according to some lukewarm start ν_N for Shifted IDLA. If $\nu_N = \mu_N$ then A is said to be a *stationary cluster*.

For $A \in \Omega$, let $|A|$ denote the number of occupied sites above level 0, that is

$$|A| := \sharp\{(x,y) \in A : y > 0\}.$$

Let further $h(A)$ denote the height of A, that is

$$h(A) = \max\{y : (x,y) \in A \text{ for some } x \in \mathbb{Z}_N\},$$

so that $h(A) \geq 0$ for all $A \in \Omega$.

Our next result is a bound for the height of a typical cluster: it is at most logarithmic in N with high probability.

Theorem 1.2 (Height of typical clusters). *For any $\gamma > 0$, $k \in \mathbb{N}$ there exists a constant $c_{\gamma,k}$, depending only on γ and k, such that for all k-lukewarm starts ν_N it holds*

$$\nu_N\Big(\{A : h(A) > c_{\gamma,k} \log N\}\Big) \leq N^{-\gamma}. \tag{3}$$

for N large enough.

In particular, taking $k = 0$ we see that stationary IDLA clusters have logarithmic height with high probability. This gives nontrivial information on the stationary measure of a nonreversible Markov chain on an infinite state space.

Remark 1.1. We believe that this bound is tight, in the sense that for $A_N \sim \mu_N$ the ratio $h(A_N)/\log N$ converges in probability to a positive constant as $N \to \infty$. See [6, Figure 9] for some numerical evidence.

We use Theorem 1.2 to bound from above the time it takes for IDLA to forget a typical initial profile.

Theorem 1.3 (The upper bound). *For any $\gamma > 0$, $k \in \mathbb{N}$ there exist a constant $d_{\gamma,k}$ and a set $\Omega_{\gamma,k} \subseteq \Omega$, depending only on γ and k, such that the following hold for all sufficiently large N. For any k-lukewarm start ν_N on $\mathbb{Z}_N \times \mathbb{Z}$, we have*

$$\nu_N(\Omega_{\gamma,k}) \geq 1 - N^{-\gamma}.$$

Moreover, for any $A_0, A_0' \in \Omega_{\gamma,k}$ with $|A_0| = |A_0'|$, writing $t_{\gamma,k} = d_{\gamma,k} N^2 \log N$ for brevity, we have

$$\|P(t_{\gamma,k}) - P'(t_{\gamma,k})\|_{TV} \leq N^{-\gamma}$$

where $(A(t))_{t \geq 0}$ and $(A'(t))_{t \geq 0}$ are IDLA processes starting from A_0 and A_0' respectively, and $P(t)$ and $P'(t)$ denote the laws of $A(t)$ and $A'(t)$ respectively.

Thus IDLA forgets a typical initial state, and hence in particular a stationary initial state, in $\mathcal{O}(N^2 \log N)$ steps with high probability.

Remark 1.2. The assumption $|A_0| = |A_0'|$ could be relaxed to $|A_0| \equiv |A_0'| \pmod{N}$, by shifting the smaller cluster to include some full rows of N sites each. To remove the cardinality assumption entirely, one could consider "lazy" IDLA processes: at each discrete time step, a particle is added with probability $1/2$, and otherwise nothing happens. Then the difference $|A_t| - |A_t'| \pmod{N}$ is a lazy simple random walk on an N-cycle, so with probability at least $1 - N^{-\gamma}$ there is a time $T \le d_\gamma N^2 \log N$ such that $|A_T| \equiv |A_T'| \pmod{N}$.

When the initial profiles are stationary, we can complement the upper bound in Theorem 1.3 with the following lower bound.

Theorem 1.4 (The lower bound). *For any $\delta, \varepsilon > 0$ there exist disjoint subsets $\Omega_\delta, \Omega_\delta'$ of Ω such that*

$$\mu_N(\Omega_\delta) \ge \frac{1}{2} - \delta, \qquad \mu_N(\Omega_\delta') \ge \frac{1}{2} - \delta \tag{4}$$

and a constant $\alpha = \alpha(\delta, \varepsilon) > 0$ such that the following holds. Let $(A(t))_{t \ge 0}$ and $(A'(t))_{t \ge 0}$ be two IDLA processes on $\mathbb{Z}_N \times \mathbb{Z}$ starting from $A_0 \in \Omega_\delta$, $A_0' \in \Omega_\delta'$, and denote by $P(t), P'(t)$ the laws of $A(t)$, $A'(t)$ respectively. Then

$$\|P(\alpha N^2) - P'(\alpha N^2)\|_{TV} > 1 - \varepsilon \tag{5}$$

for N large enough.

The above theorem tells us that two independently sampled stationary profiles A, A' are, with probability arbitrarily close to $1/2$, different enough for IDLA to need order N^2 steps to forget from which one it started. To prove this, we identify a slow-mixing statistic based on the second eigenvector of simple random walk on $\mathbb{Z}_N$ (see Definition 8.1 in Section 8.3).

1.1. **Outline of the proofs.** We start by showing that IDLA forgets polynomially high profiles in polynomial time, as stated in the following theorem.

Theorem 1.5. *Let A_0, A_0' be any two clusters in Ω with $|A_0| = |A_0'|$ and define*

$$h_0 = \max\{h(A_0), h(A_0')\}.$$

Assume that $h_0 \le N^m$ for some fixed $m \in \mathbb{N}$. Let $(A(t))_{t \ge 0}$ and $(A'(t))_{t \ge 0}$ be two IDLA processes starting from A_0, A_0' and denote the laws of $A(t)$, $A'(t)$ by $P(t), P'(t)$ respectively. Then for any $\gamma > 0$ there exists a constant $d_{\gamma,m}'$, depending only on γ and m, such that for

$$t_{\gamma,m} = h_0 N + d_{\gamma,m}' N^2 \log N$$

and N large enough, it holds

$$\|P(t_{\gamma,m}) - P'(t_{\gamma,m})\|_{TV} \le N^{-\gamma}.$$

Theorem 1.5 is proved via a coupling argument that we spell out below. We then combine it with Theorem 1.1 to show that stationary clusters have at most logarithmic height with high probability (cf. Theorem 1.2). The upper bound (cf. Theorem 1.3), is a simple corollary of these two results. We sketch here the main ideas behind these proofs.

1.1.1. *Theorem 1.5: the water level coupling.* Let A_0 and A'_0 be two clusters in Ω with $n_0 := |A_0| = |A'_0|$ and $h_0 = \max\{h(A_0), h(A'_0)\} \leq N^m$. We are going to build large IDLA clusters starting from A_0, A'_0 in a convenient way. Let $t_{\gamma,m} = h_0 N + d'_{\gamma,m} N^2 \log N$ as in Theorem 1.5. Introduce a new IDLA cluster W_0, independent of everything else, built by adding $t_{\gamma,m}$ particles to the flat configuration R_0. Note that since W_0 contains polynomially many particles, by Theorem 1.1 it will be completely filled up to height $h_0 + b_{\gamma,m} N \log N$ for a suitable constant $b_{\gamma,m}$ with high probability.
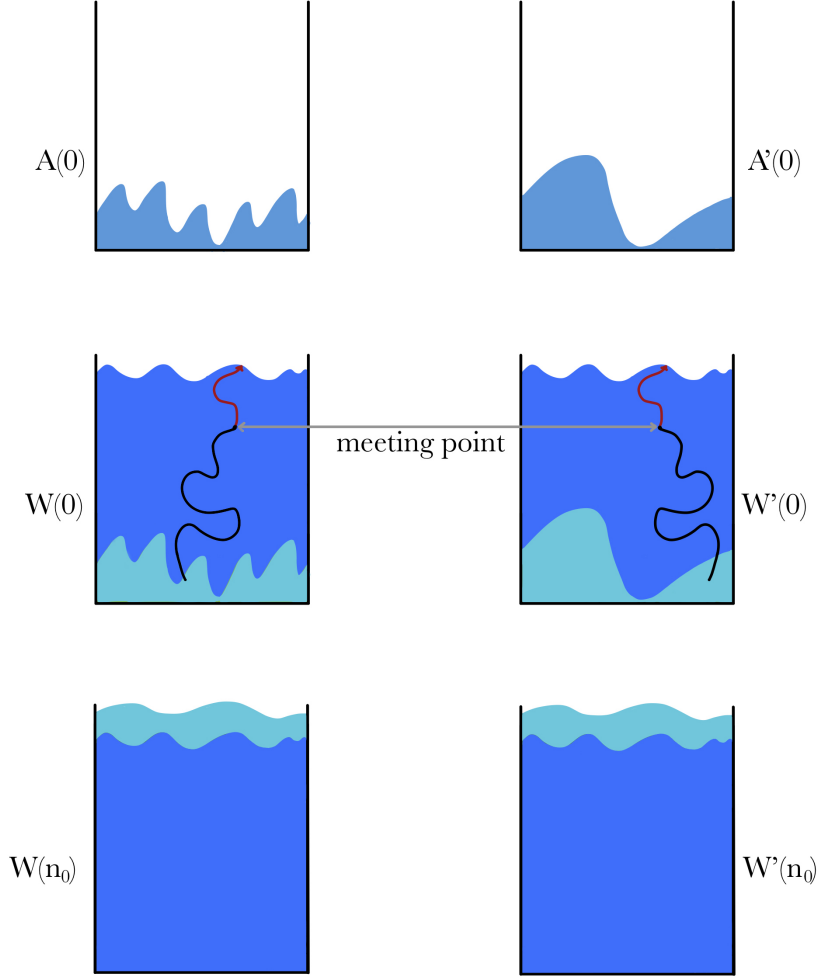

FIGURE 2. The "water level coupling" in the proof of Theorem 1.5. Top row: initial clusters. Middle row: the water cluster $W(0)$, with frozen particles in light blue. Frozen particles are released in pairs, and their trajectories (in black) are coupled so that they meet with high probability before exiting $W(0)$. After meeting, they follow the same trajectory (in red). As a result, they exit in the same exit location. Bottom row: identical clusters $W(n_0) = W'(n_0)$ resulting from the release of all frozen particles.

We take W_0 to be the initial configuration of two auxiliary processes $(W(t))_{t \leq n_0}$ and $(W'(t))_{t \leq n_0}$, that we think of as water flooding the clusters. Water falling in A_0 (resp. A'_0)

freezes, and it is only released at a later time. Frozen water particles are released in pairs, and their trajectories are coupled so to make the particles meet with high probability before exiting the respective clusters. Clearly, by taking the initial water cluster W_0 large enough we can ensure that all pairs of frozen particles meet with high probability before exiting their respective clusters, in which case we have $W(s) = W'(s)$ for all $s \leq n_0$. The theorem then follows by invoking the Abelian property (cf. Section 2) for the equalities in law

$$A(t_{\gamma,m}) \overset{(d)}{=} W(n_0), \qquad A'(t_{\gamma,m}) \overset{(d)}{=} W'(n_0).$$

See Figure 2 for an illustration of this argument, and Section 4 for the details.

1.1.2. *Theorem 1.2: typical clusters are shallow.* It suffices to prove the result for $\nu_N = \mu_N$. Assume Theorem 1.1, according to which an IDLA process starting from flat has logarithmic fluctuations for polynomially many steps, with high probability. It suffices to show that such process reaches stationarity in polynomial time to conclude. This would follow from Theorem 1.5, if we knew that stationary clusters have at most polynomial height in N, with high probability. To see this, we first observe that stationary clusters are dense, since an IDLA process spends only a small amount of time at low density configurations. Here the density of a cluster $A \in \Omega$ is measured via its *excess height* $\mathcal{E}(A) = h(A) - |A|/N$, that is the difference between the actual height and ideal height of A. We show that if the excess height is too large, then it has a negative drift under IDLA dynamics. The advantage of measuring the clusters' density via their excess height lies on the fact that a bound on the latter easily translates on a bound on the number of empty sites below the top level. Once we have such bound, we can try to *fill the holes* below the top level by releasing enough (but at most polynomially many) additional particles. This will leave us with clusters of at most polynomial height. Take any deterministic time of the form $t = N^k$ for large enough k. Then the cluster $A(t)$ is both stationary and it has at most polynomial height with high probability, which implies that typical clusters have at most polynomial height with high probability, as claimed.
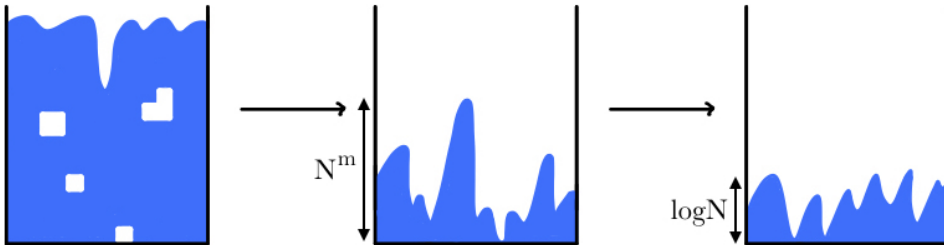

FIGURE 3. Sketch of the proof of Theorem 1.2. Starting from a high density configuration (left), the associated Shifted IDLA process reaches a polynomially high profile (centre), and then a logarithmically high one (right). Both transitions take at most polynomially many steps.

1.2. **Organization of the paper.** We start by recalling the Abelian property of Internal DLA in Section 2. We then collect some useful preliminary results in Section 3. In Section 4 we prove Theorem 1.5, concerning deterministic initial profiles. In Section 5 we bound the fluctuations of IDLA clusters with polynomially many particles (cf. Theorem 1.1), which we then use in Section 6 to prove Theorem 1.2. The upper bound (cf. Theorem 1.3) is a simple corollary of Theorems 1.5 and 1.2, as we briefly explain in Section 7. The corresponding lower bound (cf. Theorem 1.4) is proved in Section 8. We conclude the paper with a short review of the logarithmic fluctuations result by Jerison, Levine and Sheffield (cf. Theorem 5.1), that we include in Appendix C.

Acknowledgement. We are very grateful to Tom Holding, James Norris and Yuval Peres for many valuable comments and suggestions, which substantially improved our results. We are also indebted to the referee for a very careful reading of the paper. V.S. would like to thank Cornell University, where this work was initiated, for the kind hospitality.

2. The Abelian property

The Abelian property of Internal DLA was first observed by Diaconis and Fulton [5]. More recently, it has been used to generate exact samples of Internal DLA clusters in less time than it takes to run the constitutent random walks [6]. For the related model of *activated random walkers* on $\mathbb{Z}$, the Abelian property was used to prove existence of a phase transition [13].

To state the version of the Abelian property that will be used in our arguments, let us start by defining the Diaconis-Fulton *smash sum* in our setting. Given a set $A \in \Omega$ and a vertex $z \in \mathbb{Z}_N \times \mathbb{Z}$, define the set $A \oplus \{z\}$ as follows:

(i) if $z \notin A$, then $A \oplus \{z\} := A \cup \{z\}$,

(ii) if $z \in A$, then $A \oplus \{z\}$ is the random set obtained by adding to A the endpoint of a simple random walk started at z and stopped upon exiting A.

Remark 2.1. Note that if $A(0) = R_0$ and $\{z_1, z_2, \ldots, z_t\}$ are t independent, uniformly distributed vertices at level zero, then we can build an IDLA cluster $A(t)$ by setting

$$A(t) = ((A(0) \oplus \{z_1\}) \oplus \{z_2\}) \oplus \cdots \oplus \{z_t\}.$$

The Abelian property, stated below, gives some freedom on how to build IDLA clusters without changing their laws.

Proposition 2.1 (Abelian property, [5]). *Given any finite set $\{z_1, z_2, \ldots, z_t\}$ of vertices of $\mathbb{Z}_N \times \mathbb{Z}$, and a set $A \in \Omega$, the law of*

$$((A \oplus \{z_1\}) \oplus \{z_2\}) \oplus \cdots \oplus \{z_t\}$$

does not depend on the order of the z_i's. More precisely, if $\sigma : \{1, \ldots, t\} \to \{1, \ldots, t\}$ is an arbitrary permutation of $\{1, \ldots, t\}$, we have the equality in distribution

$$((A \oplus \{z_1\}) \oplus \{z_2\}) \oplus \cdots \oplus \{z_t\} \overset{(d)}{=} ((A \oplus \{z_{\sigma(1)}\}) \oplus \{z_{\sigma(2)}\}) \oplus \cdots \oplus \{z_{\sigma(t)}\}.$$

In light of the above result, given $A \in \Omega$ and a finite subset B of $\mathbb{Z}_N \times \mathbb{Z}$, we write $A \oplus B$ to denote $((A \oplus \{z_1\}) \oplus \{z_2\}) \oplus \cdots \oplus \{z_n\}$, where $z_1, \ldots z_n$ is an arbitrary enumeration of the elements of B.

Remark 2.2. We point out that Proposition 2.1 is stated and proved in [5] for finite sets, while in our setting the set A is infinite. Nevertheless, we can easily reduce to working with finite sets by changing the jump rates of our random walks at level zero to mimic the hitting distribution of a simple random walk after an excursion below level zero. This has the effect of contracting all excursions below level zero to a single step, and it clearly does not change the law of the model.

3. PRELIMINARIES

3.1. A mixing bound.
For $x \in \mathbb{Z}_N$, let P_x^t denote the law of a lazy[1] simple random walk on $\mathbb{Z}_N$ starting from x, and denote its stationary measure, uniform on $\mathbb{Z}_N$, by π_N. For $\varepsilon > 0$ define

$$\tau_N(\varepsilon) := \inf \left\{ t \geq 0 : \max_{x \in \mathbb{Z}_N} \|P_x^t - \pi_N\|_{TV} \leq \varepsilon \right\}. \tag{6}$$

The Total Variation (TV) mixing time of this walk is defined to be $\tau_N(1/4)$, which we simply denote by τ_N. It is well known that

$$\tau_N \leq N^2, \tag{7}$$

and moreover

$$\tau_N(N^{-\gamma}) \leq \lceil \log_2 N^\gamma \rceil \tau_N \leq 3\gamma N^2 \log N \tag{8}$$

for any $\gamma > 0$ (see e.g. [14]). Let $\omega = (x(t), y(t))_{t \geq 0}$ be a simple random walk on $\mathbb{Z}_N \times \mathbb{Z}$ and define

$$\tau_n^y := \inf\{t \geq 0 : y(t) = n\}$$

to be the first time it reaches level n. The next lemma tells us that by the time the walker has travelled for about $N \log N$ levels in the vertical coordinate, it has mixed well in the horizontal one.

Lemma 3.1. *For all $\gamma > 0$ and all $n \geq 10\gamma N \log N$, it holds*

$$\max_{x \in \mathbb{Z}_N} \mathbb{P}_{(x,0)}(\tau_n^y < 3\gamma N^2 \log N) \leq N^{-\gamma}$$

for N large enough.

Proof. Let $(\hat{x}(t))_{t \geq 0}$ be the jump process associated to the motion of the horizontal coordinate, obtained by only looking at $(x(t))_{t \geq 0}$ when the random walk makes a horizontal step. Similarly, let $(\hat{y}(t))_{t \geq 0}$ be the jump process associated to the motion of the vertical coordinate. Let

$$\hat{\tau}_n^y := \inf\{t \geq 0 : \hat{y}(t) = n\}$$

denote the number of vertical steps made by the walk to reach level n. For $i \geq 0$, let $G_i - 1$ denote the number of moves in the x coordinates between the ith and the $(i+1)^{th}$ move in the y coordinate. Then $(G_i)_{i \geq 0}$ is a collection of i.i.d. Geometric random variables of mean 2, and we have

$$\tau_n^y = \sum_{i=1}^{\hat{\tau}_n^y - 1} G_i.$$

---

[1] A lazy walk stays in place with probability $1/2$, and otherwise takes a simple random walk step.

It follows that

$$
\max_{x \in \mathbb{Z}_N} \mathbb{P}_{(x,0)}(\tau_n^y < N^2 \log N^{3\gamma}) = \mathbb{P}_0 \left(\sum_{i=1}^{\hat{\tau}_n^y - 1} G_i < N^2 \log N^{3\gamma} \right)
$$

$$
\leq \mathbb{P}_0 \left(\sum_{i=1}^{8\gamma N^2 \log N} G_i < 3\gamma N^2 \log N \right) + \mathbb{P}_0 \left(\hat{\tau}_n^y \leq 9\gamma N^2 \log N \right). \tag{9}
$$

We estimate the two terms separately: the first one is controlled by large deviations estimates, while the second one by using the explicit expression for the moment generating function of $\hat{\tau}_n^y$. More precisely, we have

$$
\mathbb{P}_0 \left(\sum_{i=1}^{8\gamma N^2 \log N} G_i < 3\gamma N^2 \log N \right) = \mathbb{P}_0 \left(2^{-\sum_{i=1}^{8\gamma N^2 \log N} G_i} > 2^{-3\gamma N^2 \log N} \right)
$$

$$
\leq \left[2\mathbb{E}(2^{-G_1})^{\frac{8}{3}} \right]^{3\gamma N^2 \log N} \leq \left(\frac{2}{9} \right)^{3\gamma N^2 \log N} \leq N^{-3\gamma},
$$

where we have used that $\mathbb{E}(2^{-G_1}) = 1/3$ for the second inequality. For the second term, it is simple to check that, for $z \in (0, 1)$,

$$
\mathbb{E}_0 \left(z^{\hat{\tau}_1^y} \right) = \frac{1 - \sqrt{1 - z^2}}{z}, \qquad \mathbb{E}_0 \left(z^{\hat{\tau}_n^y} \right) = \mathbb{E}_0 \left(z^{\hat{\tau}_1^y} \right)^n.
$$

This gives, for any $z \in (0, 1)$,

$$
\mathbb{P}_0 \left(\hat{\tau}_n^y < 9\gamma N^2 \log N \right) \leq \mathbb{E}_0 \left(z^{\hat{\tau}_n^y} \right) z^{-9\gamma N^2 \log N}
$$

$$
= \left(\frac{1 - \sqrt{1 - z^2}}{z} \right)^n \cdot \left(\frac{1}{z} \right)^{9\gamma N^2 \log N}.
$$

Choose $z \in (0, 1)$ of the form $z^2 = 1 - 1/\alpha^2$ for some $\alpha \gg 1$ as $N \to \infty$, to have that

$$
\left(\frac{1 - \sqrt{1 - z^2}}{z} \right)^n \cdot \left(\frac{1}{z} \right)^{9\gamma N^2 \log N} \leq \left(1 - \frac{1}{\alpha} \right)^{n/2} \left(1 - \frac{1}{\alpha^2} \right)^{-\frac{9}{2}\gamma N^2 \log N}
$$

$$
\leq \exp \left(-\frac{n}{2\alpha} + \frac{5\gamma}{\alpha^2} N^2 \log N \right).
$$

To have the far r.h.s. smaller than, say, $N^{-5\gamma/4}$ it suffices to take

$$
n \geq \frac{10\gamma}{\alpha} N^2 \log N + \frac{5}{2}\gamma \alpha \log N.
$$

Optimizing over α suggests to take $\alpha = 2N$, to get $n \geq 10\gamma N \log N$. $\square$

3.2. **The role of starting locations.** The following result tells us that if the walkers have time to mix in the horizontal coordinate before exiting the cluster, the resulting IDLA configuration does not depend too much on their initial positions.

Proposition 3.1. *Fix any $T > 0$ such that $T \leq N^m$ for some finite $m \in \mathbb{N}$. Let $\{(x_i, y_i)\}_{1 \leq i \leq T}$ and $\{(x_i', y_i')\}_{1 \leq i \leq T}$ denote two fixed collections of vertices of $\mathbb{Z}_N \times \mathbb{Z}$ such that $y_i, y_i' \leq 0$ for all $i \leq T$. Let, moreover, $(A(t))_{t \leq T}$ and $(A'(t))_{t \leq T}$ be two IDLA processes with starting configurations $A(0) = A'(0)$, and such that the ith walkers start from (x_i, y_i) and (x_i', y_i') respectively. Then there exists a coupling of the two processes such that*

the following holds. For any $\gamma > 0$ there exists a finite constant $C_{\gamma,m}$, depending only on γ and m, such that if

$$A(0) = A'(0) \supseteq R_{C_{\gamma,m} N \log N} \tag{10}$$

then

$$\mathbb{P}(A(t) = A'(t) \text{ for all } t \leq T) \geq 1 - N^{-\gamma}. \tag{11}$$

Remark 3.1. In particular, this tells us that, as long as the IDLA cluster is filled up to level $C_{\gamma,m} N \log N$, releasing the next T walkers from fixed initial locations below level 0 or from uniform locations at level 0 results in the same final cluster with high probability.

Proof. This is an easy consequence of Lemma 3.1. For $i \leq T$, let $\omega_i = ((x_i(k), y_i(k)))_{k \geq 0}$ and $\omega_i' = ((x_i'(k), y_i'(k)))_{k \geq 0}$ denote the simple random walk trajectories of the ith walkers starting from (x_i, y_i) and (x_i', y_i') respectively. These are coupled as follows. If, say, $y_i < y_i'$ then ω_i' stays in place until ω_i reaches level y_i' (and vice versa if $y_i > y_i'$). We can therefore assume that $y_i = y_i'$ without loss of generality. If, moreover, N is even and $|x_i - x_i'|$ is odd, then the first time that ω_i moves in the horizontal coordinate we keep ω_i' in place. Since the probability of ω_i reaching level N before making a horizontal step is $o(2^{-N})$ for large N, we can assume that $|x_i - x_i'|$ is even without loss of generality.

The walks move as follows. If ω_i moves in the y coordinate, then so does ω_i', and the two walks take the same step. Thus $y_i(0) = y_i'(0)$ implies that $y_i(k) = y_i'(k)$ for all $k \geq 0$. If, on the other hand, ω_i moves in the x coordinate, then so does ω_i', and the two walks move according to the reflection coupling on the N-cycle (cf. Section 1.1). Finally, the walkers ω_i and ω_i' stick together upon meeting, that is if $\omega_i(k) = \omega_i'(k)$ for some $k \geq 0$ then $\omega_i(j) = \omega_i'(j)$ for all $j \geq k$. Note that if $A(i-1) = A'(i-1)$ and the ith walkers meet before exiting the identical clusters, then $A(i) = A'(i)$.

Let, consistently with the notation introduced in the proof of the previous result, $(\hat{x}_i(k))_{k \geq 0}$ and $(\hat{x}_i'(k))_{k \geq 0}$ be the jump processes associated to $(x_i(k))_{k \geq 0}$ and $(x_i'(k))_{k \geq 0}$. Then, if

$$\hat{\tau}_x := \inf\{k \geq 0 : \hat{x}_i(k) = \hat{x}_i'(k)\}$$

we have, by comparison with a simple random walk,

$$\mathbb{P}(\hat{\tau}_x > 3\gamma N^2 \log N) \leq N^{-\gamma} \tag{12}$$

for all $\gamma > 0$ and N large enough. Moreover, by setting

$$\tau_x := \inf\{k \geq 0 : x_i(k) = x_i'(k)\},$$

we see that $x_i(\tau_x) = x'(\tau_x) = \hat{x}(\hat{\tau}_x) = \hat{x}'(\hat{\tau}_x)$ and

$$\tau_x = \sum_{k=1}^{\hat{\tau}_x - 1} G_i,$$

for $(G_i)_{i \geq 1}$ i.i.d. Geometric random variables of mean 2. Let $\gamma' = 6(\gamma + m)$ and $\gamma'' = \gamma'/3$. Recall that R_n denotes the infinite rectangle of height n, and assume that $A(0) = A'(0) \supseteq R_n$ for $n = 10\gamma' N \log N$. Denote by τ_n the first time both walkers reach level n. Then by

Lemma 3.1 and (12) we have

$$\mathbb{P}(\omega_i \text{ and } \omega_i' \text{ exit } R_n \text{ before meeting}) = \mathbb{P}(\tau_x > \tau_n) = \mathbb{P}\Big(\sum_{k=1}^{\hat{\tau}_x - 1} G_k > \tau_n \Big)$$

$$\leq \mathbb{P}\Big(\sum_{k=1}^{\hat{\tau}_x} G_k > 3\gamma' N^2 \log N \Big) + \mathbb{P}(\tau_n < 3\gamma' N^2 \log N)$$

$$\leq \mathbb{P}\Big(\sum_{k=1}^{3\gamma'' N^2 \log N} G_k > 3\gamma' N^2 \log N) \Big) + \mathbb{P}\big(\hat{\tau}_x > 3\gamma'' N^2 \log N \big) + N^{-\gamma'}$$

$$\leq \mathbb{P}\Big(\sum_{k=1}^{3\gamma'' N^2 \log N} G_k > 9\gamma'' N^2 \log N \Big) + N^{-\gamma''} + N^{-\gamma'}$$

$$\leq \left[\frac{\mathbb{E}(e^{\lambda G_k})}{e^{3\lambda}} \right]^{3\gamma'' N^2 \log N} + 2N^{-\gamma''},$$

where the third inequality follows from (12). Finally, taking $\lambda = \log(3/2) > 0$ makes the term in the square brackets equal to $8/9$, from which we conclude that

$$\mathbb{P}(\omega_i \text{ and } \omega_i' \text{ exit } R_n \text{ before meeting}) \leq N^{-3\gamma'' N^2 \log \frac{9}{8}} + 2N^{-\gamma''} \leq N^{-(\gamma+m)}$$

for N large enough. In all, we have found that

$$\mathbb{P}(\exists t \leq T \text{ such that } A(t) \neq A'(t)) \leq \mathbb{P}(\exists i \leq T \text{ such that } \omega_i \text{ and } \omega_i' \text{ exit } R_n \text{ before meeting})$$

$$\leq T N^{-(\gamma+m)} \leq N^{-\gamma}.$$

This shows that we can take $C_{\gamma,m} = 60(\gamma + m)$ in (10) to have (11), thus concluding the proof. $\qquad\square$

4. THE WATER LEVEL COUPLING

In this section we prove Theorem 1.5.

Proof of Theorem 1.5. Let A_0, A_0' be any two clusters in Ω with $|A_0| = |A_0'| = n_0$ and such that

$$h_0 = \max\{h(A_0), h(A_0')\} \leq N^m.$$

Let $C_{\gamma+1,m+1}$ and $b_{\gamma+1,m+2}$ be defined as in Proposition 3.1 and Theorem 1.1 respectively, and define

$$d'_{\gamma,m} := 2\max\{C_{\gamma+1,m+1}, b_{\gamma+1,m+2}\}.$$

We build an auxiliary water cluster W_0 by adding $t_{\gamma,m} = h_0 N + d'_{\gamma,m} N^2 \log N$ particles to the flat configuration R_0 according to IDLA rules. Then, since $t_{\gamma,m} \leq N^{m+2}$, by Theorem 1.1 we have

$$\mathbb{P}\Big(W_0 \supseteq R_{\frac{t_{\gamma,m}}{N} - b_{\gamma+1,m+2} \log N} \Big) \geq 1 - N^{-(\gamma+1)}$$

for N large enough. In particular, since $d'_{\gamma,m} \geq C_{\gamma+1,m+1} + b_{\gamma+1,m+2}$, this gives

$$\mathbb{P}\Big(W_0 \supseteq R_{h_0 + C_{\gamma+1,m+1} N \log N} \Big) \geq 1 - N^{-(\gamma+1)},$$

i.e. the water cluster is completely filled up to height $h_0 + C_{\gamma+1,m+1}N\log N$ with high probability. Write

$$A_0 = \{z_1, z_2, \ldots, z_{n_0}\}, \qquad A_0' = \{z_1', z_2', \ldots, z_{n_0}'\}$$

for arbitrary enumerations of the sites in A_0, A_0' above level 0. We define two auxiliary processes $(W(t))_{t \le n_0}$, $(W'(t))_{t \le n_0}$ by setting $W(0) = W'(0) = W_0$ and inductively defining for $t \le n_0$

$$W(t) = W(t-1) \cup \{Z_t\}, \qquad W'(t) = W'(t-1) \cup \{Z_t'\},$$

where Z_t, Z_t' denote the exit locations from $W(t-1)$, $W'(t-1)$ of simple random walks on $\mathbb{Z}_N \times \mathbb{Z}$ starting from z_t, z_t'. These walks are coupled as follows. If z_t is at a lower level than z_t', then the walk starting from z_t moves freely (independently of everything else) until it reaches the level of z_t', while the other walk stays in place. Once at the same level, the walks move together in the vertical coordinate, whereas the horizontal coordinates evolve according to the *reflection coupling*: if one steps to the left, the other one steps to the right, and vice versa[2]. Once the walks meet, they move together in both coordinates. Since $n_0 \le h_0 N \le N^{m+1}$, and all the walks start at distance at least $C_{\gamma+1,m+1}N\log N$ from the boundary of the cluster, Proposition 3.1 gives[3]

$$\mathbb{P}\big(W(n_0) \ne W'(n_0)\big) \le \mathbb{P}\Big(W(n_0) \ne W'(n_0)\Big|W_0 \supseteq R_{\frac{t_{\gamma,m}}{N}+C_{\gamma+1,m+1}\log N}\Big) +$$
$$+ \mathbb{P}\Big(W_0 \not\supseteq R_{\frac{t_{\gamma,m}}{N}+C_{\gamma+1,m+1}\log N}\Big) \le 2N^{-(\gamma+1)} \le N^{-\gamma}$$

for N large enough. The result then follows by observing that if $(A(t))_{t \ge 0}$ and $(A'(t))_{t \ge 0}$ denote two IDLA processes starting from A_0 and A_0' respectively, then

$$A(t_{\gamma,m}) \overset{(d)}{=} W(n_0), \qquad A'(t_{\gamma,m}) \overset{(d)}{=} W'(n_0)$$

by the Abelian property. Indeed, if we denote by $w_1, w_2, \ldots, w_{t_{\gamma,m}}$ the starting locations of the $t_{\gamma,m}$ walkers used to grow W_0, then, with the notation introduced in Section 2, we have

$$W(n_0) = (((R_0 \oplus \{w_1\}) \oplus \{w_2\}) \oplus \cdots \oplus \{w_{t_{\gamma,m}}\}) \oplus A_0,$$

while

$$A(t_{\gamma,m}) = ((R_0 \oplus A_0) \oplus \{w_1\}) \oplus \{w_2\}) \oplus \cdots \oplus \{w_{t_{\gamma,m}}\}.$$

The claimed equality in law is then given by Proposition 2.1. $\qquad\square$

Remark 4.1. Note that to prove Theorem 1.5 we have constructed a coupling of the final clusters $A(t_{\gamma,m})$, $A'(t_{\gamma,m})$, not of the whole processes $(A(t))_{t \ge 0}$, $(A'(t))_{t \ge 0}$.

---

[2]Here we again use the first horizontal step of the walks to adjust the parity of the difference of the horizontal coordinates, if needed, as explained in the proof of Proposition 3.1.

[3]Although the particles' starting positions were taken to be below level 0 in Proposition 3.1 for notational convenience, the result applies in this setting by invariance under vertical shifts.

5. Logarithmic fluctuations for large clusters

In this section we bound the fluctuations of an IDLA cluster with polynomially many particles, thus proving Theorem 1.1. To start with, we claim that the following holds.

Theorem 5.1. *Let $(A(t))_{t \geq 0}$ denote an IDLA process on $\mathbb{Z}_N \times \mathbb{Z}$ starting from the flat configuration $A(0) = R_0$. Fix any $T \leq (N \log N)^2$. Then for any $\gamma > 0$ there exists a finite constant a_γ, depending only on γ, such that*

$$\mathbb{P}\Big(R_{\frac{T}{N} - a_\gamma \log N} \subseteq A(T) \subseteq R_{\frac{T}{N} + a_\gamma \log N} \Big) \geq 1 - N^{-\gamma}$$

for N large enough.

The above theorem is stated for $T = N^2$ in [9], where the authors use it to show convergence of space-time averages of IDLA fluctuations to the Gaussian free field. It can be proved by the same arguments used in [7] for planar IDLA, which are in fact rather simplified by the structure of the cylinder graph. Since this proof does not appear anywhere, and since we need to extend it to larger values of T, we give it in Appendix C.

Assuming Theorem 5.1, we can proceed with the proof of Theorem 1.1.

Proof of Theorem 1.1. It suffices to show that (2) holds for any fixed $T \leq N^m$. The result will then follow by replacing γ with $\gamma + m$ and using the union bound over all $T \leq N^m$.

If $T = \mathcal{O}(N^2 \log N)$ then we are done by Theorem 5.1, so assume $T \gg N^2 \log N$. We are going to iteratively reduce the size of the cluster, until it becomes $\mathcal{O}(N^2 \log N)$. To this end, let $a_{\gamma+2+m}$ and $C_{\gamma+2,m}$ denote the constants in Theorem 5.1 and Proposition 3.1 respectively, and set $a := a_{\gamma+2+m}$, $c := C_{\gamma+2,m}$ for brevity. At cost of increasing these constants, we can assume that $2c \geq 1$ and both $a \log N$ and $c \log N$ are integers. Define

$$n := aN \log N + 2cN^2 \log N.$$

We build a large cluster, with the same law of $A(T)$, by using the water processes mentioned in the introduction. To this end, let W_1 denote the cluster obtained by adding n particles to $A(0) = R_0$ according to IDLA rules. Then by Theorem 5.1

$$\mathbb{P}\big(R_{2cN \log N} \subseteq W_1 \subseteq R_{2a \log N + 2cN \log N} \big) \geq 1 - N^{-(\gamma+2+m)} \qquad (13)$$

for N large enough. Let

$$W_1^f := W_1 \cap R_{2cN \log N}$$

denote the region which is filled with high probability after n releases. Write further

$$F_1 := W_1 \setminus W_1^f$$

for the fluctuation region. Water particles in the fluctuation region are declared frozen, and they will be released at a later time. On top of the water-filled region W_1^f we are going to build a second cluster with again n particles, so to fill a rectangle of height $4cN \log N$ with high probability. Let W_2 denote such a cluster, obtained by adding n particles to W_1^f according to IDLA rules (thus, new water particles can, and will, settle inside F_1). Write W_2^f and F_2 for the filled region and fluctuation region of such cluster. Then, as in (13), we have

$$\mathbb{P}\Big(R_{4cN \log N} \subseteq W_1^f \cup W_2 \subseteq R_{2aN \log N + 4cN \log N} \Big) \geq 1 - 2N^{-(\gamma+2+m)}.$$

We again declare particles in F_2 frozen, and treat their locations as empty for subsequent walkers. This procedure is iterated for $k = \lfloor T/n \rfloor - 1$ rounds.

Let Ω_1 denote the event that the fluctuations bound (13) holds for all k rounds, that is

$$\Omega_1 := \left\{ W_1^f \cup W_2^f \cup \ldots \cup W_k^f = R_{2kcN \log N} \right\} \cap \bigcap_{l=1}^{k} \left\{ F_l \subseteq R_{2a \log N + 2lcN \log N} \setminus R_{2lcN \log N} \right\}.$$

Then by (13)

$$\mathbb{P}(\Omega_1) \geq 1 - kN^{-(\gamma+2+m)} \geq 1 - N^{-(\gamma+2)}.$$

Let us restrict to this good event. It remains to release the $akN \log N$ frozen particles
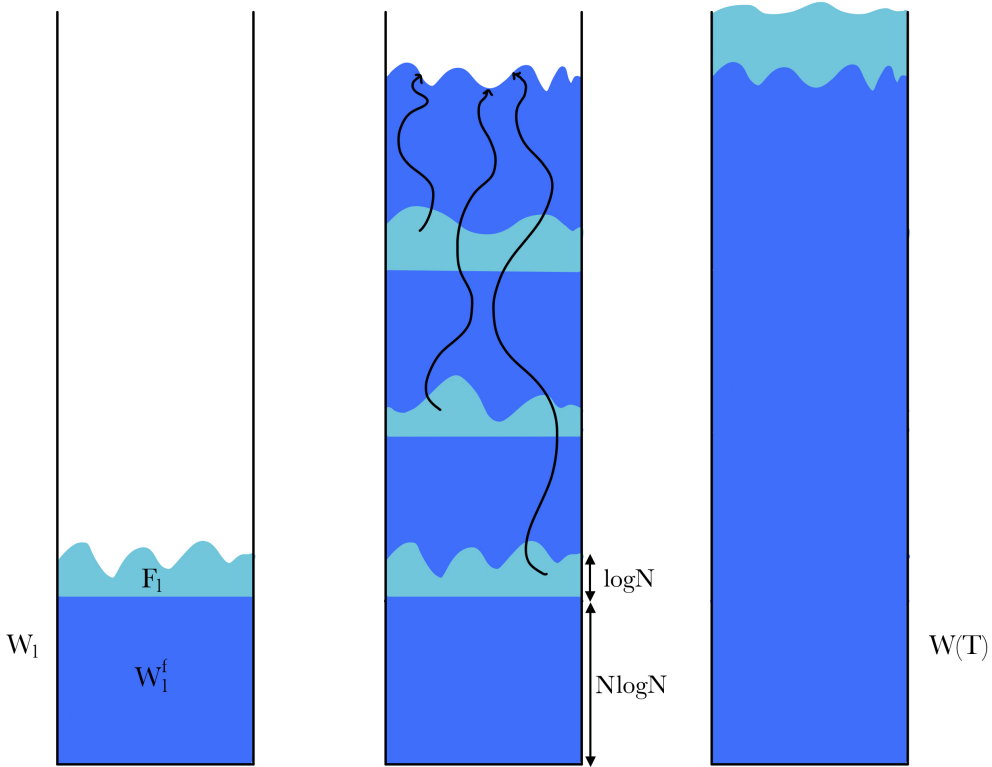

FIGURE 4. Sketch of the proof of Theorem 1.1.

from their locations, plus $T - kn$ new ones uniformly from level 0. Denote by $W(T)$ the cluster obtained after all the released particles have settled, so that $|W(T)| = T$. By the Abelian property, $W(T)$ has the same distribution as $A(T)$. Write $T' = T - kn + akN \log N$ for brevity, and let $W'(T')$ denote the cluster obtained by adding T' particles to the filled rectangle $R_{2kcN \log N}$ according to IDLA rules (more precisely, we start T' random walks uniformly from level zero, independently of everything else, and add their exit locations to the cluster). We argue that we can couple $W(T)$ and $W'(T')$ so that they coincide with high probability. To see this, we proceed as follows. First release $T - kn$ new particles uniformly from level 0, and note that, since $T - kn \geq n$, on the event Ω_1 these will fill a rectangle of height $2cN \log N$ with probability at least $1 - N^{-(\gamma+2+m)}$. If this happens,

then all the frozen particles are at distance at least $cN \log N$ from the boundary of the cluster. Thus by Proposition 3.1 we can couple $W(T)$ and $W'(T')$ so that

$$\mathbb{P}\Big(W(T) = W'(T')\Big) \geq 1 - \mathbb{P}(\Omega_1) - N^{-(\gamma+2+m)} - N^{-(\gamma+2)} \geq 1 - N^{-(\gamma+1)}.$$

In all, we have reduced the problem of bounding the fluctuations of $W(T) \overset{(d)}{=} A(T)$ to the same one for the smaller cluster

$$A(T') \overset{(d)}{=} W'(T') \setminus R_{2kcN \log N},$$

at the price of a small probability of failure. Since $T' \leq 2 \max\{2n, aT/N\}$, this either makes the number of particles $\mathcal{O}((N \log N)^2)$, in which case we stop, or it decreases it by a multiplicative factor $2a/N$. Thus after at most m iterations of the above procedure we are back to clusters with $\mathcal{O}((N \log N)^2)$ particles, which we know to have logarithmic fluctuations. This shows that

$$\mathbb{P}\Big(R_{\frac{T}{N} - a \log N} \subseteq A(T) \subseteq R_{\frac{T}{N} + a \log N}\Big) \geq 1 - mN^{-(\gamma+1)} \geq 1 - N^{-\gamma},$$

so (2) holds with $b_{\gamma,m} = a$, as wanted.

$\square$

6. Typical profiles are shallow

In this section we show that typical IDLA profiles have at most logarithmic height, thus proving Theorem 1.2. This is achieved by combining Theorem 1.5 with a control on the density of stationary clusters.

6.1. **Decay of the excess height.** We distinguish between high and low density clusters by looking at their excess height, that we now define. Recall that Ω denotes the set of clusters completely filled up to level 0, so that $h(A) \geq 0$ for all $A \in \Omega$, while $|A|$ denotes the total number of sites in A strictly above level 0.

Definition 6.1 (Excess height). For $A \in \Omega$, the *excess height* of A, denoted by $\mathcal{E}(A)$, is defined as the difference between the height of A and the minimum possible height, i.e.,

$$\mathcal{E}(A) := h(A) - \frac{|A|}{N}.$$

Note that $\mathcal{E}(A) \geq 0$. We say that a cluster A has high density if $\mathcal{E}(A) \leq \mathcal{E}^*$, where $\mathcal{E}^*$ is a constant to be chosen later depending only on N. If instead $\mathcal{E}(A) > \mathcal{E}^*$, then A is said to have low density.
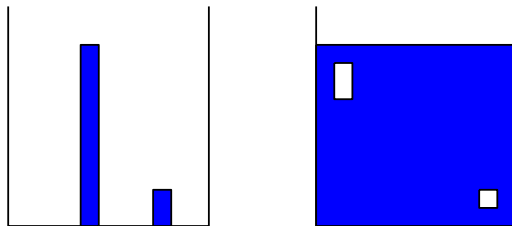

FIGURE 5. Clusters with high excess height (left) and low excess height (right).

To start with, we prove that, for a suitable choice of $\mathcal{E}^*$, the excess height drops below $\mathcal{E}^*$ quickly under the IDLA dynamics.

Lemma 6.1. *Let $(A(t))_{t \geq 0}$ denote an IDLA process on $\mathbb{Z}_N \times \mathbb{Z}$ with $A(0) \supseteq R_0$. For $t \geq 0$ let $\mathcal{E}(t) := \mathcal{E}(A(t))$ denote the excess height of $A(t)$. Then for any $\eta \in (0,1)$ there exists a constant $\mathcal{E}^* = \mathcal{E}^*(N, \eta)$, depending only on N and η, such that if*

$$T_{\mathcal{E}^*} := \inf\{t \geq 0 : \mathcal{E}(t) \leq \mathcal{E}^*\}$$

denotes the first time that the excess height drops below $\mathcal{E}^*$, then

$$\mathbb{P}(T_{\mathcal{E}^*} > t) \leq e^{-\frac{\eta^2}{8N^2}t},$$

for $t > \frac{N}{\eta}(\mathcal{E}(0) - \mathcal{E}^*)$ and N large enough.

Lemma 6.1 is an easy consequence of the following result, which tells us that if a cluster has low density then its excess height has a negative drift under the IDLA dynamics.

Lemma 6.2. *Let $\mathcal{F}_t := \sigma\{A(s) : s \leq t\}$, and write $h(t)$ in place of $h(A(t))$ for brevity. Then for all $\eta \in (0,1)$ there exists a constant $\mathcal{E}^* = \mathcal{E}^*(N, \eta)$, depending only on N and η, such that if $\mathcal{E}(t) > \mathcal{E}^*$ then it holds*

$$\mathbb{E}\big(h(t+1) - h(t)|\mathcal{F}_t\big) < \frac{1-\eta}{N}$$

for N large enough.

Proof. Fix $\eta \in (0,1)$ throughout. For $k \leq h(t)$, we say that level k is *bad* if it contains at least one empty site, that is if $A(t)^c \cap \{y = k\} \neq \emptyset$. We claim that if $\mathcal{E}(t) \geq \mathcal{E}^*$ then there are at least $\mathcal{E}^*$ bad levels between 0 and the top one $h(t)$. Indeed, since $h(t) \geq \frac{|A(t)|}{N} + \mathcal{E}^*$ then there are at least $\frac{|A(t)|}{N} + \mathcal{E}^*$ levels above level 0 in the cluster, and at most $\lfloor \frac{|A(t)|}{N} \rfloor$ of them can be completely filled.

Recall from (6) the definition of $\tau_N(N^{-\gamma})$, and note that by (8) and Lemma 3.1 with $\gamma = 2$ we have

$$\max_{x \in \mathbb{Z}_N} \mathbb{P}_{(x,0)}(\tau_n^y < \tau_N(N^{-2})) \leq \max_{x \in \mathbb{Z}_N} \mathbb{P}_{(x,0)}(\tau_n^y < 6N^2 \log N) \leq N^{-2}$$

as long as $n \geq n^* := 20N \log N$. Then, if $\omega = (x(k), y(k))_{k \geq 0}$ is a simple random walk on $\mathbb{Z}_N \times \mathbb{Z}$, the above implies that

$$\min_{x,x' \in \mathbb{Z}_N} \mathbb{P}_{(x,0)}\big(y(\tau_{n^*}^y) = (x', n^*)\big) \geq \frac{1}{N} - \frac{1}{N^2} \geq \frac{1}{2N}$$

for N large enough. In words, a simple random walk on $\mathbb{Z}_N \times \mathbb{Z}$ starting at level zero has probability at least $1/2N$ to reach level n^* for the first time at any given vertex, uniformly over the starting location. Now, since there are at least $\mathcal{E}^*$ bad levels, we can find at least $\frac{\mathcal{E}^*}{n^*}$ bad levels at distance at least n^* from each other. We treat these as traps. More precisely, for a new particle to increase the height of the cluster $A(t)$, the particle must travel through all the bad levels without exiting the cluster, until it reaches the top. Since we are taking the bad levels sufficiently far apart, the particle has time to mix in between, so whenever it reaches a bad level for the first time it has probability at least $\frac{1}{2N}$ to fall outside the cluster. By taking enough bad levels, then, we can make the probability for the particle to survive all of them arbitrarily small.
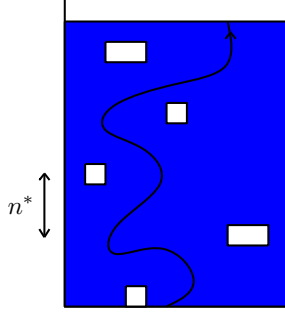
FIGURE 6. An example of the trajectory of a particle surviving all the bad
levels at distance at least n^* one from the other.

Formally, since the walker has probability at least $\frac{1}{2N}$ to exit the cluster upon reaching
a bad level for the first time, we have

$$\mathbb{P}(h(t+1) - h(t) = 1 | \mathcal{F}_t) \leq \left(1 - \frac{1}{2N} \right)^{\mathcal{E}^*/n^*}. \tag{14}$$

To make this smaller than $\frac{1-\eta}{N}$ it suffices to take $\mathcal{E}^*$ large enough, precisely

$$\mathcal{E}^* \geq \left\lceil 2n^* N \log \left(\frac{N}{1-\eta} \right) \right\rceil, \tag{15}$$

which concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Note in particular that

$$\mathcal{E}^* > 40(N \log N)^2, \tag{16}$$

which will be useful later on. We can now prove that low density clusters tend to decrease
their excess height under IDLA dynamics.

Proof of Lemma 6.1. Let $n_0 = |A(0)|$ denote the number of sites in $A(0)$ of positive height,
and assume that $\mathcal{E}(0) = h(0) - n_0/N > \mathcal{E}^*$, with $\mathcal{E}^*$ as in (15). It follows from Lemma 6.2
that

$$M(t) := h(t) - \frac{n_0 + (1-\eta)t}{N}, \quad t \geq 0$$

is a supermartingale up to the stopping time $T_{\mathcal{E}^*}$, with $M(0) = \mathcal{E}(0)$. As a consequence, the
stopped process $M^*(t) = M(t \wedge T_{\mathcal{E}^*})$ is a supermartingale for all $t \geq 0$, with $M^*(t) = M(t)$
for $t \leq T_{\mathcal{E}^*}$, and

$$|M^*(t+1) - M^{T_{\mathcal{E}^*}}(t)| \leq |h(t+1) - h(t)| + \frac{1-\eta}{N} \leq 2$$

for all $t \geq 0$. Now, since $|A(t)| = n_0 + t$, for $t < T_{\mathcal{E}^*}$ we have

$$\{\mathcal{E}(t) > \mathcal{E}^*\} = \left\{ h(t) - \frac{n_0 + t}{N} > \mathcal{E}^* \right\} \subseteq \left\{ M(t) > \mathcal{E}^* + \frac{\eta}{N}t \right\} = \left\{ M^*(t) > \mathcal{E}^* + \frac{\eta}{N}t \right\}.$$

It thus follows from Azuma's inequality that, for $t > \frac{N}{\eta}(\mathcal{E}(0) - \mathcal{E}^*)$,

$$\mathbb{P}(T_{\mathcal{E}^*} > t) \leq \mathbb{P}\big(T_{\mathcal{E}^*} > t, \mathcal{E}(t) > \mathcal{E}^*\big) \leq \mathbb{P}\big(T_{\mathcal{E}^*} > t, M^*(t) > \mathcal{E}^* + \frac{\eta}{N}t\big)$$

$$\leq \mathbb{P}\Big(M^*(t) - M^*(0) > \mathcal{E}^* + \frac{\eta}{N}t - M(0)\Big)$$

$$\leq \exp\Big\{-\frac{(\mathcal{E}^* + \eta t/N - M(0))^2}{8t}\Big\}$$

$$\leq \exp\Big\{-\frac{(\mathcal{E}^* - \mathcal{E}(0))^2 + (\eta t/N)^2}{8t}\Big\}$$

$$\leq \exp\Big\{-\frac{\eta^2}{8N^2}t\Big\},$$

as claimed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 6.1. Lemma 6.1 implies that Shifted IDLA is positive recurrent, thus proving the existence of the stationary distribution μ_N. To see this it suffices to show, for example, that the flat configuration R_0 is positive recurrent. Let $(A(t))_{t\geq 0}$ be an IDLA process with $A(0) = R_0$, and define T_0 to be the first time t such that $A(t) = R_k$ for some $k \geq 1$. Here is a wasteful way to show that $\mathbb{E}_{R_0}(T_0) < \infty$. Fix any $\eta \in (0,1)$ and let $\mathcal{E}^* = \mathcal{E}^*(N, \eta)$ be the integer constant in Lemma 6.1. Starting from R_0, release $\mathcal{E}^*N$ particles: with probability at least $N^{-\mathcal{E}^*N}$ the final configuration will be the filled rectangle $R_{\mathcal{E}^*}$. If not, then the excess height has increased by at most $\mathcal{E}^*N$. We keep releasing particles until the excess height falls again below $\mathcal{E}^*$: by Lemma 6.1 this takes a random time with exponential tails, and hence finite expectation. Once the excess height is at most $\mathcal{E}^*$, there are at most $\mathcal{E}^*N$ empty sites below the top level in the cluster. We release as many particles as the number of the empty sites below the top level: with probability at least $N^{-\mathcal{E}^*N}$ the final configuration will be a filled rectangle R_k for some $k \geq 1$. If not, the excess height is at most $\mathcal{E}^* + \mathcal{E}^*N$: we again wait for it to fall below $\mathcal{E}^*$, and iterate. After at most a Geometric$\big(N^{-\mathcal{E}^*N}\big)$ number of attempts, the final configuration will be a filled rectangle. Each attempt takes $\mathcal{E}^*N$ releases, plus the time it takes for the excess height to fall below $\mathcal{E}^*$ starting from $\mathcal{E}^* + \mathcal{E}^*N$, which has exponential tails. In all, we conclude that the total time it takes to go from R_0 back to a flat configuration R_k for some $k \geq 1$ has finite expectation.

6.2. **Typical clusters are dense.** In this section we show that, for N large enough, μ_N gives high probability to high density clusters. This shows that stationary clusters are dense, and hence that typical clusters are dense, with high probability.

Proposition 6.1. *For $\mathcal{E}^* = \mathcal{E}^*(N, \eta)$ as in* (15) *and N large enough, it holds*

$$\mu_N\Big(\big\{A : \mathcal{E}(A) > 2\mathcal{E}^*\big\}\Big) \leq e^{-N/2}.$$

Proof. Recall that for $(A(t))_{t\geq 0}$ IDLA process on $\mathbb{Z}_N \times \mathbb{Z}$, we denote by $(A^*(t))_{t\geq 0}$ the associated shifted process. Let us denote by $\Omega^* := \{A^* : A \in \Omega\}$ its state space.

Define the sets

$$\mathcal{A} = \{A \in \Omega^* : \mathcal{E}(A) \leq \mathcal{E}^*\}, \qquad \mathcal{B} = \{A \in \Omega^* : \mathcal{E}(A) \leq 2\mathcal{E}^*\}.$$

We seek to bound $\mu_N(\mathcal{B}^c)$. By the Ergodic theorem for positive recurrent Markov chains,

$$\mu_N(\mathcal{B}^c) = \lim_{t \to \infty} \frac{1}{t} \mathbb{E}\bigg(\sum_{i=1}^{t} \mathbb{1}(A^*(i) \in \mathcal{B}^c) \bigg). \tag{17}$$

Define the following stopping times:

$$\tau_{\mathcal{A},\partial\mathcal{B}}^1 := \inf\{t \geq 0 : A^*(t) \in \partial\mathcal{B}\},$$
$$\tau_{\partial\mathcal{B},\mathcal{A}}^1 := \inf\{t \geq \tau_{\mathcal{A},\partial\mathcal{B}}^1 : A^*(t) \in \mathcal{A}\} - \tau_{\mathcal{A},\partial\mathcal{B}}^1,$$
$$\tau_{\mathcal{A},\partial\mathcal{B}}^i := \inf\{t \geq \tau_{\partial\mathcal{B},\mathcal{A}}^{i-1} : A^*(t) \in \partial\mathcal{B}\} - \tau_{\partial\mathcal{B},\mathcal{A}}^{i-1}, \quad i \geq 2,$$
$$\tau_{\partial\mathcal{B},\mathcal{A}}^i := \inf\{t \geq \tau_{\mathcal{A},\partial\mathcal{B}}^i : A^*(t) \in \mathcal{A}\} - \tau_{\mathcal{A},\partial\mathcal{B}}^i, \quad i \geq 2,$$

where $\partial\mathcal{B} = \{A \in \Omega^* : \mathcal{E}(A) \in [2\mathcal{E}^*, 2\mathcal{E}^* + 1)\}$. Note that, since the excess height always changes by either $1 - 1/N$ or $1/N$,

$$\mathcal{E}\bigg(\sum_{i=1}^{k} \big(\tau_{\mathcal{A},\partial\mathcal{B}}^i + \tau_{\partial\mathcal{B},\mathcal{A}}^i \big) \bigg) \leq \mathcal{E}^* + 1, \qquad \forall\, k \geq 1.$$

We divide the interval $[0, t]$ into excursions from $\mathcal{A}$ to $\partial\mathcal{B}$ and then from $\partial\mathcal{B}$ to $\mathcal{A}$. Since during excursions from $\mathcal{A}$ to $\partial\mathcal{B}$ the process is in $\mathcal{B}$, only excursions from $\partial\mathcal{B}$ to $\mathcal{A}$ contribute to the expectation in (17). Let us say that the concatenation of an excursion from $\mathcal{A}$ to $\partial\mathcal{B}$ and from $\partial\mathcal{B}$ to $\mathcal{A}$ is a *complete excursion*. In the next lemma we bound the number of complete excursions by time t.

Lemma 6.3. *Let*

$$K(t) := \sup\bigg\{ k \geq 1 : \sum_{i=1}^{k} \big(\tau_{\mathcal{A},\partial\mathcal{B}}^i + \tau_{\partial\mathcal{B},\mathcal{A}}^i \big) \leq t \bigg\}$$

denote the number of complete excursions by time t. Then for $\gamma = \frac{6}{Ne^N}$ and N large enough it holds

$$\mathbb{P}(K(t) \geq \gamma t) \leq \exp\Big(-\frac{t}{Ne^N} \Big). \tag{18}$$

Proof. Let

$$\tilde{K}(t) := \sup\bigg\{ k \geq 1 : \sum_{i=1}^{k} \tau_{\mathcal{A},\partial\mathcal{B}}^i \leq t \bigg\}.$$

Then clearly $\tilde{K}(t) \geq K(t)$, and so $\mathbb{P}(K(t) \geq \gamma t) \leq \mathbb{P}(\tilde{K}(t) \geq \gamma t)$. Moreover, since the excess height changes by at most 1 at each step, at the start of each excursion from $\mathcal{A}$ to $\partial\mathcal{B}$ the excess height must lie between $\mathcal{E}^* - 1$ and $\mathcal{E}^*$. We know (cf. Lemma 6.2) that when the excess height is greater than $\mathcal{E}^*$ it has a negative drift, but we do not have any information on its drift when the process is in the set $\mathcal{A}$. To overcome this problem, we simply ignore the time spent in $\mathcal{A}$. Indeed, we will see that the time spent in $\mathcal{B} \setminus \mathcal{A}$ is large enough to give us what we want.

We seek to stochastically bound $\tau^i_{\mathcal{A},\partial\mathcal{B}}$ from below. To this end, define the following auxiliary random walk on $\frac{1}{N}\mathbb{N} = \left\{\frac{n}{N} : n \in \mathbb{N}\right\}$, with a reflecting barrier at zero:

$$X_0 = 0, \qquad \mathbb{P}\left(X_{i+1} = 1 - \frac{1}{N} \,\Big|\, X_i = 0\right) = 1,$$

$$X_{i+1} - X_i = \begin{cases} 1 - \frac{1}{N}, & \text{with probability } \frac{1-\eta}{N}, \\ -\frac{1}{N}, & \text{otherwise} \end{cases} \qquad \text{for } X_i > 0. \tag{19}$$

Note that, while the Shifted IDLA process $(A^*(t))_{t\geq0}$ is in $\mathcal{B} \setminus \mathcal{A}$, the walk X away from 0 stochastically dominates the associated excess height process by Lemma 6.2. In particular, let N_0 denote the total number of visits of X to 0 before reaching $[\mathcal{E}^*, \infty)$. Then

$$\tau^i_{\mathcal{A},\partial\mathcal{B}} \succeq N_0 N, \tag{20}$$

since each time the walk reaches 0 it jumps deterministically to $1 - 1/N$, after which it takes at least N steps to reach 0 again. Now, N_0 is a geometric random variable with success probability $\mathbb{P}_{1-1/N}(X \text{ reaches } [\mathcal{E}^*, \infty) \text{ before } 0)$. The next result tells us that this probability is very small, and so N_0 is typically large.

Lemma 6.4. *For $\mathcal{E}^*$ as in (15) and N large enough, it holds*

$$\mathbb{P}_{1-1/N}(X \text{ reaches } [\mathcal{E}^*, \infty) \text{ before } 0) \leq e^{-N}. \tag{21}$$

Let us postpone the proof of the above lemma to Appendix A, and explain how from this one can deduce the bound of Lemma 6.3. Note that (20) and (21) together imply that

$$\mathbb{E}(\tau^i_{\mathcal{A},\partial\mathcal{B}}) \geq Ne^N$$

for N large enough. Let now $T^X_{\mathcal{E}^*} := \inf\{i \geq 0 : X_i \geq \mathcal{E}^*\}$ denote the first hitting time of $[\mathcal{E}^*, \infty)$ for the walk X, and notice that $T^X_{\mathcal{E}^*} \geq N_0 N$. We define $\left(T^{X,(i)}_{\mathcal{E}^*}\right)_{i\geq1}$ to be i.i.d. copies of $T^X_{\mathcal{E}^*}$, independent of everything else. Recall the definition of $\tilde{K}(t)$, and define further

$$\tilde{K}'(t) := \sup\left\{k \geq 1 : \sum_{i=1}^{k} T^{X,(i)}_{\mathcal{E}^*} \leq t\right\}.$$

Then we have

$$\mathbb{P}(\tilde{K}(t) \geq \gamma t) \leq \mathbb{P}(\tilde{K}'(t) \geq \gamma t) = \mathbb{P}\left(\sum_{i=1}^{\gamma t} T^{X,(i)}_{\mathcal{E}^*} \leq t\right). \tag{22}$$

Lemma 6.5. *Let $\gamma = \frac{6}{Ne^N}$. Then for N large enough it holds*

$$\mathbb{P}\left(\sum_{i=1}^{\gamma t} T^{X,(i)}_{\mathcal{E}^*} \leq t\right) \leq \exp\left(-\frac{t}{Ne^N}\right).$$

This lemma is an easy consequence of the fact that $T^X_{\mathcal{E}^*} \geq N_0 N$ and Lemma 6.4, so we leave the proof for Appendix B. This finishes the proof of Lemma 6.3. $\qquad\square$

Using (18) we can conclude the proof of Proposition 6.1. Indeed, writing $\mathcal{E}(t)$ in place of $\mathcal{E}(A^*(t))$ for brevity, and taking γ as above, we find

$$
\mu_N(\mathcal{B}^c) = \lim_{t\to\infty} \frac{1}{t}\mathbb{E}\bigg(\sum_{i=1}^{t} \mathbb{1}(\mathcal{E}(t) \geq 2\mathcal{E}^*)\bigg) \leq \lim_{t\to\infty}\frac{1}{t}\mathbb{E}\bigg(t - \sum_{i=1}^{K(t)} \tau^i_{\mathcal{A},\partial\mathcal{B}}\bigg)
$$

$$
\leq \lim_{t\to\infty}\bigg[\frac{1}{t}\mathbb{E}\bigg(t - \sum_{i=1}^{K(t)} \tau^i_{\mathcal{A},\partial\mathcal{B}}\,;\, K(t) < \gamma t\bigg) + \mathbb{P}(K(t) \geq \gamma t)\bigg]
$$

$$
\leq \lim_{t\to\infty}\bigg[\frac{1}{t}\mathbb{E}\bigg(\sum_{i=1}^{\gamma t} \tau^i_{\partial\mathcal{B},\mathcal{A}}\bigg) + \exp\Big(-\frac{t}{Ne^N}\Big)\bigg] = \gamma\mathbb{E}(\tau^1_{\partial\mathcal{B},\mathcal{A}}) = \frac{2\mathbb{E}(\tau^1_{\partial\mathcal{B},\mathcal{A}})}{Ne^N}.
$$

Lemma 6.1 with $t_0 := \frac{N}{\eta}\mathcal{E}^*$ then yields

$$
\mathbb{E}(\tau^1_{\partial\mathcal{B},\mathcal{A}}) = \int_0^\infty \mathbb{P}(\tau^1_{\partial\mathcal{B},\mathcal{A}} > t)dt \leq t_0 + \int_{t_0}^\infty \exp\Big(-\frac{\eta^2 t}{8N^2}\Big)dt
$$

$$
= \frac{N\mathcal{E}^*}{\eta} + \frac{8N^2}{\eta^2}\exp\Big(-\frac{\eta\mathcal{E}^*}{8N}\Big) \leq \frac{2N\mathcal{E}^*}{\eta}
$$

for N large enough, since $\mathcal{E}^* > (N\log N)^2$ by (16). Thus we conclude that

$$
\mu_N(\mathcal{B}^c) \leq \frac{4\mathcal{E}^*}{\eta}e^{-N} \leq e^{-N/2}
$$

for N large enough, as claimed. $\qquad\square$

6.3. **From high density to low height.** We have shown in the previous section that typical clusters are dense. While this does not give any information on the height of A, it provides an upper bound on the number of empty sites, that we will call *holes*, below the top level $h(A)$. Indeed, if $\mathcal{E}(A) \leq \mathcal{E}^*$ then there can be at most $N\mathcal{E}^*$ holes in the cluster. We now obtain a bound on the time it takes to fill these holes (cf. Proposition 6.2), showing that it is at most polynomial in N, and use this to prove that stationary clusters have at most polynomial height (cf. Proposition 6.3).

Proposition 6.2. *Let $(A^*(t))_{t\geq 0}$ denote a Shifted IDLA process on $\mathbb{Z}_N \times \mathbb{Z}$, and assume that $\mathcal{E}(A^*(0)) \leq \mathcal{E}^*$, with $\mathcal{E}^*$ as in (15). Let*

$$
\Delta := \lceil 2N^2\mathcal{E}^* \rceil + 1. \tag{23}
$$

Write $h^*(t)$ in place of $h(A^*(t))$ for brevity, and assume that $h^*(0) > \Delta$. Define $T_\Delta := \inf\{t \geq 0 : h^*(t) \leq \Delta\}$ to be the first time the height of the Shifted IDLA process drops below Δ. Then there exists a constant $C_\eta > 0$, depending only on η, such that

$$
\mathbb{P}(T_\Delta > t) \leq 4\exp\Big(-\frac{C_\eta t}{N\Delta}\Big) \tag{24}
$$

for N large enough.

Remark 6.2. Since $N\Delta$ is polynomial in N, this tells us that, when starting from a dense configuration, the height of a Shifted IDLA process drops below Δ after at most polynomially many releases.

Proof. We argue as follows. Each time we add a new particle to the cluster, the lowest hole has probability at least $1/N$ to be filled, independently of everything else. Hence it will take at most a geometric number of releases of parameter $1/N$ to fill the lowest hole. In total, then, it will take at most the sum of $N\mathcal{E}^*$ i.i.d. Geometric$(1/N)$ to fill all the holes up to the top level $h(A)$. Some care is needed, though: we cannot let the excess height increase too much while releasing these extra particles.

Let $(G_i)_{i=1}^{N\mathcal{E}^*}$ be a collection of i.i.d. Geometric$(1/N)$ random variables, and note that since $\Delta > 2N^2\mathcal{E}^*$ we have

$$\mathbb{P}\left(\sum_{i=1}^{N\mathcal{E}^*} G_i > \Delta \right) < \frac{1}{2}. \tag{25}$$

It follows that if we release Δ particles then we have probability at least $1/2$ to fill all the holes below the top level. If we fail, then the excess height has increased by at most Δ. If so, we keep on releasing particles until the excess height falls again below $\mathcal{E}^*$, and iterate. After a Geometric$(1/2)$ number of attempts we will have filled all the holes below the top level, which implies that the resulting Shifted IDLA cluster will have height at most Δ.

To formalise the above strategy, write $\mathcal{E}(t)$ in place of $\mathcal{E}(A^*(t))$ and define the following random times:

$$\begin{aligned}
\tau_0 &:= 0, \\
\tau_k &:= \inf\{t \geq \tau_{k-1} + \Delta : \mathcal{E}(t) \leq \mathcal{E}^*\}, \quad k \geq 1, \\
s_k &:= \tau_k - (\tau_{k-1} + \Delta), \quad k \geq 1.
\end{aligned}$$

Then, by definition, at times τ_i there are at most $N\mathcal{E}^*$ holes below the top level. Since the number of releases needed to fill all these holes is stochastically dominated by the sum of $N\mathcal{E}^*$ i.i.d. Geometric$(1/N)$ random variables, (25) implies that at times $\tau_i + \Delta$ we have filled all the holes with probability at least $1/2$. If this happens, then $h^*(\tau_i + \Delta) \leq \Delta$, and we stop. Otherwise $\mathcal{E}(\tau_i + \Delta) \leq \mathcal{E}^* + \Delta$, and we start again.

Let

$$K := \min\{k \geq 0 : h^*(\tau_k + \Delta) \leq \Delta\}$$

denote the number of attempts needed to succeed. Then, denoting by $\preceq$ stochastic domination, we have $K \preceq K'$ for K' Geometric$(1/2)$ random variable. Note that the times s_i's are not in general identically distributed. It is convenient to make them i.i.d. by assuming that the excess height always starts from the maximum value $\mathcal{E}^* + \Delta$. More precisely, let

$$\hat{s} := \inf\{t \geq 0 : \mathcal{E}(t) \leq \mathcal{E}^* \text{ starting from } \mathcal{E}(0) = \mathcal{E}^* + \Delta\}, \tag{26}$$

and let $(\hat{s}_i)_{i \geq 1}$ be a sequence of i.i.d. random variables equal in law to $\hat{s}$. Then $s_i \preceq \hat{s}_i$ and we have that

$$T_\Delta \preceq K\Delta + \sum_{i=1}^{K} \hat{s}_i \preceq K'\Delta + \sum_{i=1}^{K'} \hat{s}_i.$$

We use this to estimate the moment generating function of T_Δ. For $\lambda \in \mathbb{R}$, set $M(\lambda) := \mathbb{E}(e^{\lambda \hat{s}})$. We have

$$
\begin{aligned}
\mathbb{E}(e^{\lambda T_\Delta}) &\leq \mathbb{E}\left[\exp\left\{ \lambda\left(K'\Delta + \sum_{i=1}^{K'} \hat{s}_i \right) \right\} \right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\exp\left\{ \lambda\left(k\Delta + \sum_{i=1}^{k} s_i \right) \right\} \mathbb{1}(K' = k) \right] \\
&\leq \sum_{k=1}^{\infty} \left(\frac{e^{\lambda\Delta} M(\lambda)}{2} \right)^k,
\end{aligned}
\tag{27}
$$

where the first equality above follows from the Monotone Convergence Theorem, and the last inequality from the independence of K' and the $\hat{s}_i$'s.

Lemma 6.6. *For $\lambda < \frac{\eta^2}{32N^2}$ and N large enough, it holds $M(\lambda) \leq \frac{3}{2}\exp\left(\frac{2\Delta N\lambda}{\eta}\right)$.*

The above lemma implies that $\frac{e^{\lambda\Delta} M(\lambda)}{2} \leq \frac{3}{4}\exp\left\{ \lambda\Delta\left(1 + \frac{2N}{\eta}\right) \right\}$, which can be made smaller than, say, $4/5$ by taking

$$
\lambda < \frac{\log(16/15)}{\Delta\left(1 + \frac{2N}{\eta}\right)}.
$$

Let $C_\eta = \frac{\log(16/15)}{4}\eta$, so that $\frac{C_\eta}{N\Delta} < \frac{\log(16/15)}{\Delta\left(1 + \frac{2N}{\eta}\right)}$ for large N. If $\lambda^* = \frac{C_\eta}{N\Delta}$ then we have

$$
\mathbb{E}(e^{\lambda^* T_\Delta}) \leq \sum_{k=1}^{\infty} \left(\frac{4}{5} \right)^k = 4.
$$

Thus we conclude that

$$
\mathbb{P}(T_\Delta > t) \leq \mathbb{E}(e^{\lambda^* T_\Delta}) e^{-\lambda^* t} \leq 4e^{-\lambda^* t} = 4\exp\left(-\frac{C_\eta t}{N\Delta} \right),
$$

as claimed. It remains to prove Lemma 6.6.

Proof of Lemma 6.6. Recall the definition of $\hat{s}$ from (26), so that $M(\lambda) = \mathbb{E}(e^{\lambda\hat{s}})$. Introduce the auxiliary random walk $(X_k)_{k\geq 0}$ defined as in (19), and note that

$$
\hat{s} \preceq \hat{\tau}_\Delta := \inf\{ t \geq 0 : X_t \leq \mathcal{E}^* \text{ starting from } \mathcal{E}^* + \Delta \}.
$$

Then it follows from Lemma 6.1 that, for $t \geq \frac{2\Delta N}{\eta}$ and N large enough, it holds

$$
\mathbb{P}(\hat{\tau}_\Delta > t) \leq \exp\left(-\frac{\eta^2 t}{8N^2} \right).
$$

Therefore, for $\lambda < \frac{\eta^2}{32N^2}$, we find

$$\mathbb{E}(e^{\lambda \hat{s}}) \leq \mathbb{E}(e^{\lambda \hat{\tau}_\Delta}) = \int_0^\infty \mathbb{P}(e^{\lambda \hat{\tau}_\Delta} \geq t)dt = \int_0^\infty \mathbb{P}\Big(\hat{\tau}_\Delta \geq \frac{\log t}{\lambda}\Big)dt$$

$$\leq \exp\Big(\frac{2\Delta N\lambda}{\eta}\Big) + \int_{\exp\big(\frac{2\Delta N\lambda}{\eta}\big)}^\infty \exp\Big(-\frac{\eta^2 \log t}{8N^2\lambda}\Big)dt$$

$$= \exp\Big(\frac{2\Delta N\lambda}{\eta}\Big) + \frac{\exp\big(-\frac{2\Delta N\lambda}{\eta}\big(\frac{\eta^2}{8N^2\lambda}-1\big)\big)}{\frac{\eta^2}{8N^2\lambda}-1}$$

$$\leq \frac{3}{2}\exp\Big(\frac{2\Delta N\lambda}{\eta}\Big),$$

where the last inequality holds for N large enough. □

This concludes the proof of Proposition 6.2. □

We use this to show that stationary clusters have at most polynomial height.

Proposition 6.3. *For N large enough, it holds*

$$\mu_N\Big(\{A : h(A) > N^8\}\Big) \leq 2e^{-N/2}.$$

Proof. Let $(A(t))_{t\geq 0}$ be an IDLA process with $A(0) = A \sim \mu_N$. Then $A(t) \sim \mu_N$ for all deterministic $t \geq 0$. Write $h(t)$ in place of $h(A(t))$ for brevity. Take $\mathcal{E}^* = N^2 \log^3 N$, and note that it satisfies (15). Define Δ as in (23). Then we have

$$\mathbb{P}(h(A) > N^8) = \mathbb{P}(h(N^7) > N^8 | A(0) = A)$$

$$\leq \mathbb{P}(h(N^7) > N^8 | \mathcal{E}(A(0)) \leq 2\mathcal{E}^*) + \mathbb{P}(\mathcal{E}(A(0)) > 2\mathcal{E}^*)$$

$$\leq \mathbb{P}(T_\Delta > N^7 | \mathcal{E}(A(0)) \leq 2\mathcal{E}^*) + e^{-N/2} \leq 2e^{-N/2},$$

where we have used that if $T_\Delta \leq N^7$ then $h(N^7) \leq \Delta + N^7 < N^8$. □

Remark 6.3. It is worth pointing out that the proof of Proposition 6.1, and hence of Proposition 6.3 above, would also work for driving random walks on $\mathbb{Z}_N \times \mathbb{Z}$ with a vertical drift. This would still give a polynomial bound for the height of typical clusters, perhaps with a larger exponent than the one in Proposition 6.3. Reasoning as in Theorem 1.5, such height bound could in turn be translated into a (largely non-optimal) upper bound for the time it takes for IDLA with transient driving walks to forget its initial profile.

6.4. **Typical clusters are shallow.** We now finish the proof of Theorem 1.2. To start with, note that it suffices to prove the result for stationary clusters. Indeed, suppose we showed that for any $\gamma > 0$ there exists a constant c_γ such that

$$\mu_N\left(\{A : h(A) > c_\gamma \log N\}\right) \leq N^{-\gamma}$$

for N large enough. Then, if ν_N is any k-lukewarm start for Shifted IDLA, we have

$$\nu_N\left(\{A : h(A) > c_\gamma \log N\}\right) \leq N^{-(\gamma-k)},$$

so that (3) still holds with $\gamma + k$ in place of γ. It thus suffices to prove Theorem 1.2 for stationary clusters. We argue as follows. Let $(A(t))_{t\geq 0}$ be an IDLA process starting from

$A(0) \sim \mu_N$, so that $A(t) \sim \mu_N$ for all deterministic $t \geq 0$. Introduce an auxiliary IDLA process $(\bar{A}(t))_{t \geq 0}$ starting from the flat profile $\bar{A}(0) = R_0$. Then, if $|A(0)| = n_0$, we have

$$|A(t)| = |\bar{A}(n_0 + t)| = n_0 + t, \qquad \forall t \geq 0.$$

Write $A'(t) = \bar{A}(n_0 + t)$ to shorten the notation. Following the ideas presented in the proof of Theorem 1.5, we will couple the clusters $A(N^{10})$ and $A'(N^{10})$ so that they match with high probability. Since $A(N^{10}) \sim \mu_N$, and $A'(N^{10})$ has logarithmic fluctuations with high probability, this will allow us to conclude.

To start with, note that by Proposition 6.3 we have

$$\mathbb{P}\big(h(A(0)) > N^8\big) \leq 2e^{-N/2}$$

for N large enough. In particular this shows that $\mathbb{P}(n_0 > N^9) \leq 2e^{-N/2}$. We can use this to bound the height of $A'(0)$. Indeed, for γ as in the statement of Theorem 1.2, we have

$$\mathbb{P}(h(A'(0)) > 2N^8) \leq \mathbb{P}(h(\bar{A}(n_0)) > 2N^8, n_0 \leq N^9) + \mathbb{P}(n_0 > N^9)$$
$$\leq \mathbb{P}\big(h(\bar{A}(N^9)) > 2N^8\big) + 2e^{-N/2} \leq N^{-2\gamma} + 2e^{-N/2} \leq 2N^{-2\gamma}$$

for N large enough, where in the last inequality we have used Theorem 1.1 to argue that an IDLA cluster built by adding N^9 particles to R_0 has height $\mathcal{O}(N^8)$ with high probability. Introduce a water cluster W_0 obtained by adding N^{10} particles to the flat configuration R_0 according to IDLA rules, and note that

$$\mathbb{P}\big(W_0 \supseteq R_{3N^8}\big) \geq 1 - N^{-2\gamma} \tag{28}$$

by Theorem 1.1 for N large enough. Define two auxiliary water processes $(W(t))_{t \geq 0}$ and $(W'(t))_{t \geq 0}$ as follows. Set $W(0) = W'(0) = W_0$. Particles in $W(0) \cap A(0)$ and $W'(0) \cap A'(0)$ are declared frozen, and will be released at a later time. Note that

$$\mathbb{P}\Big(\max\{h(A(0)), h(A'(0))\} \leq 2N^8; W_0 \supseteq R_{3N^8} \Big) \geq 1 - 4N^{-2\gamma}, \tag{29}$$

which shows that all frozen particles are at distance at least N^8 from the boundary of W_0 with high probability. Fix arbitrary enumerations of the two sets of frozen particles, and accordingly denote their locations by $\{z_1, z_2, \ldots, z_{n_0}\}$ and $\{z'_1, z'_2, \ldots, z'_{n_0}\}$. For $t \geq 0$, then, let $W(t)$ (respectively $W'(t)$) be the cluster obtained by adding to $W(t-1)$ (respectively $W'(t-1)$) the exit location from it of a simple random walk on $\mathbb{Z}_N \times \mathbb{Z}$ starting from z_t (respectively z'_t). The random walks starting from z_t and z'_t are coupled as explained in the introduction: the higher one stays in place until the other one reaches its level, after which they move together in the vertical coordinate, and according to the reflection coupling in the horizontal one[4]. Then, writing E_0 for the event appearing in (29) for brevity, by Proposition 3.1 we find

$$\mathbb{P}(W(n_0) \neq W'(n_0)) \leq \mathbb{P}(W(n_0) \neq W'(n_0) \,|\, E_0) + \mathbb{P}(E_0^c)$$
$$\leq N^{-2\gamma} + 4N^{-2\gamma} = 5N^{-2\gamma}$$

---

[4]If needed, we use the first horizontal step to adjust the parity of the difference of the horizontal coordinates, as explained in the proof of Proposition 3.1.

for N large enough. Thus $W(n_0) = W'(n_0)$ with high probability. Since $A(N^{10}) \overset{(d)}{=} W(n_0)$ and $A'(N^{10}) \overset{(d)}{=} W'(n_0)$, this shows that we can couple $A(N^{10})$ and $A'(N^{10})$ so that

$$\mathbb{P}\big(A(N^{10}) = A'(N^{10})\big) \geq 1 - 5N^{-2\gamma}.$$

Now, $A(N^{10})$ is stationary since $A(0)$ is. We want to argue that $A'(N^{10}) = \bar{A}(n_0 + N^{10})$ has logarithmic fluctuations. If n_0 were deterministic, this would follow from Theorem 1.1. Instead n_0 is random since such is $A(0)$, and as already observed it satisfies $\mathbb{P}(n_0 > N^9) \leq 2e^{-N/2}$ for N large enough. Moreover, by Theorem 1.1 there exists a constant b_γ, depending only on γ, such that

$$\mathbb{P}\Big(R_{\frac{t}{N} - b_\gamma \log N} \subseteq A(t) \subseteq R_{\frac{t}{N} + b_\gamma \log N} \ \ \forall t \leq 2N^{10}\Big) \geq 1 - N^{-2\gamma}$$

for N large enough. We thus conclude that

$$\mathbb{P}\Big(\Big\{R_{\frac{n_0+N^{10}}{N} - b_\gamma \log N} \subseteq A'(N^{10}) \subseteq R_{\frac{n_0+N^{10}}{N} + b_\gamma \log N}\Big\}^c\Big) =$$

$$= \mathbb{P}\Big(\Big\{R_{\frac{n_0+N^{10}}{N} - b_\gamma \log N} \subseteq \bar{A}(n_0 + N^{10}) \subseteq R_{\frac{n_0+N^{10}}{N} + b_\gamma \log N}\Big\}^c\Big) \leq$$

$$\leq \mathbb{P}\Big(\Big\{R_{\frac{t}{N} - b_\gamma \log N} \subseteq \bar{A}(t) \subseteq R_{\frac{t}{N} + b_\gamma \log N} \ \ \forall t \leq 2N^{10}\Big\}^c\Big) + \mathbb{P}(n_0 > N^9)$$

$$\leq N^{-2\gamma} + 2e^{-N/2} \leq 2N^{-2\gamma}$$

for N large enough. This shows that $A'(N^{10})$ has logarithmic fluctuations with high probability. Recall that for $A \in \Omega$ we denote by A^* its shifted version (cf. Definition 1.1). Then we have found that

$$\mu_N\Big(\big\{A : h(A) > 2b_\gamma \log N\big\}\Big) = \mathbb{P}\big(h(A^*(N^{10})) > 2b_\gamma \log N\big)$$

$$\leq \mathbb{P}\big(h(A'^*(N^{10})) > 2b_\gamma \log N\big) + \mathbb{P}\big(A(N^{10}) \neq A'(N^{10})\big)$$

$$\leq 7N^{-2\gamma} \leq N^{-\gamma}$$

for N large enough, which concludes the proof.

7. The upper bound

We briefly explain how one can deduce Theorem 1.3 from Theorems 1.5 and 1.2. For any $\gamma, k > 0$ let $c_{\gamma,k}$ be defined as in Theorem 1.2, while we take $d'_{\gamma,2}$ as in Theorem 1.5 with $m = 2$. Set

$$d_{\gamma,k} := c_{\gamma,k} + d'_{\gamma,2}, \qquad t_{\gamma,k} := d_{\gamma,k} N^2 \log N.$$

Then, if ν_N is any k-lukewarm distribution, we can define

$$\Omega_{\gamma,k} := \{A : h(A) \leq c_{\gamma,k} \log N\}$$

to have that, by Theorem 1.2,

$$\mathbb{P}\big(\Omega_{\gamma,k}\big) \geq 1 - N^{-\gamma}$$

for N large enough. Take any two clusters A_0, A'_0 in $\Omega_{\gamma,k}$ with $|A_0| = |A'_0|$, and let $(A(t))_{t \geq 0}$ and $(A'(t))_{t \geq 0}$ be two IDLA processes starting from A_0 and A'_0 respectively. Then Theorem 1.5 tells us that we can build $A(t_{\gamma,k})$ and $A'(t_{\gamma,k})$ on the same probability space so that

$$\mathbb{P}\big(A(t_{\gamma,k}) \neq A'(t_{\gamma,k})\big) \leq N^{-\gamma}$$

for N large enough. We thus gather that

$$\|P(t_{\gamma,k}) - P'(t_{\gamma,k})\|_{TV} = \sup_{S \subseteq \Omega} \left| \mathbb{P}(A(t_{\gamma,k}) \in S) - \mathbb{P}(A'(t_{\gamma,k}) \in S) \right|$$

$$\leq 2\mathbb{P}\big(A(t_{\gamma,k}) \neq A'(t_{\gamma,k})\big) \leq 2N^{-\gamma}$$

for N large enough. Since γ is arbitrary, this concludes the proof.

8. The lower bound

In this section we specialise to stationary initial clusters, and prove that if two such clusters are sampled *independently* from μ_N, then the IDLA dynamics will remember from which one it started for at least order N^2 steps, as stated in Theorem 1.4. We proceed as follows. We first recall the GFF fluctuations result by Jerison, Levine and Sheffield [9], saying that the average IDLA fluctuations, appropriately measured, converge to the restriction of the Gaussian Free Field to the unit circle. We then use this to define an observable which we show to be large, with positive probability, for stationary IDLA clusters (cf. Proposition 8.3). Finally, we argue that a necessary condition for IDLA to forget the initial configuration is for this observable to reach 0, and show that this takes time at least αN^2 for some $\alpha > 0$, thus proving the result.

8.1. **Average IDLA fluctuations and the GFF.** Let us start by briefly recalling the average IDLA fluctuations result by Jerison, Levine and Sheffield [9]. Let $(A(t))_{t \geq 0}$ be an IDLA process on $\mathbb{Z}_N \times \mathbb{Z}$ starting from flat, i.e. $A(0) = R_0$. For $n = (n_1, n_2) \in \mathbb{Z}_N \times \mathbb{Z}$, define the rescaled square

$$Q_N(n) := \left\{ (x,y) \in \mathbb{T} \times \mathbb{R} : x \in \left(\frac{n_1 - 1}{N}, \frac{n_1}{N} \right], \ y \in \left(\frac{n_2 - 1}{N}, \frac{n_2}{N} \right] \right\}$$

with side–length $1/N$ and $\left(\frac{n_1}{N}, \frac{n_2}{N} \right)$ as top–right corner, and set

$$A_N(t) := \bigcup_{n \in A(t)} Q_N(n), \tag{30}$$

so that $A_N(t) \subset \mathbb{T} \times \mathbb{R}$. We use the rescaled and filled cluster $A_N(t)$ to define the *discrepancy function*

$$D_{N,t}(x,y) := N \left(\mathbb{1}_{A_N(t)}(x,y) - \mathbb{1}_{\left\{ y \leq \frac{t}{N^2} \right\}}(x,y) \right) \tag{31}$$

for $(x,y) \in \mathbb{T} \times \mathbb{R}$. Note that $D_{N,t}$ is supported on the symmetric difference $A_N(t) \Delta R_{t/N^2}$. Finally, let $\varphi \in C^\infty(\mathbb{T} \times \mathbb{R})$ be of the form

$$\varphi(x,y) = \sum_{|k| \leq K} \alpha_k(y) e^{2\pi i k x}$$

for some finite integer K and with $\alpha_{-k} = \overline{\alpha_k}$, so that φ is real–valued. Jerison, Levine and Sheffield proved the following.

Theorem 8.1 (Theorem 3, [9])**.** *Let $T = y_0 N^2$ and φ be as above. Then, as $N \to \infty$,*

$$D_{N,T}(\varphi) := \int_{\mathbb{T} \times \mathbb{R}} D_{N,T}(x,y) \varphi(x,y) \mathrm{d}x \mathrm{d}y$$

converges in distribution to a Gaussian random variable with mean zero and variance

$$v(\varphi) = \sum_{0 < |k| \leq K} |\alpha_k(y_0)|^2 \Big(\frac{1 - e^{-4\pi|k|y_0}}{4\pi|k|} \Big).$$

The next result tells us that the above theorem can be generalised to larger times.

Theorem 8.2. *Let $T = CN^2 \log N$ for some absolute constant C large enough. Then, as $N \to \infty$,*

$$\int_{\mathbb{T} \times \mathbb{R}} D_{N,T}(x, y) \varphi \Big(x, y - \frac{T}{N^2} \Big) \mathrm{d}x \mathrm{d}y$$

converges in distribution to a Gaussian random variable with mean zero and variance

$$v_\mu(\varphi) = \sum_{0 < |k| \leq K} \frac{|\alpha_k(0)|^2}{4\pi|k|}.$$

Remark 8.1. Note that, as observed in [9], the exponential term in $v(\varphi)$ is due to the fact that the process started from flat. Indeed, this term does not appear in the limiting variance $v_\mu(\varphi)$, since T is large enough for the process to have reached stationarity.

The above result can be proved exactly as in [9], Theorem 3, by replacing the maximal fluctuations result with Theorem 1.1 above. For this reason we choose to skip the proof.

We now use the fact that $A(T)$ is close to a stationary cluster to control the size of the average fluctuations at stationarity. Let A^μ denote a stationary cluster, that is $A^\mu \sim \mu_N$, and denote by A^μ_N its rescaled and filled version as in (30). We start an IDLA process $(A^\mu(t))_{t \geq 0}$ from $A^\mu(0) = A^\mu$. Let $n_0 = |A^\mu(0)|$ be the number of particles in $A^\mu(0)$ above level zero. Then, if $(A(t))_{t \geq 0}$ is an IDLA process starting from flat, we have

$$|A(n_0)| = |A^\mu(0)| = n_0$$

and hence $|A(n_0 + t)| = |A^\mu(t)| = n_0 + t$ for all $t \geq 0$. Moreover, by Theorems 1.1 and 1.2, for any $\gamma > 0$ there exists $a_\gamma < \infty$ such that, with $h_0 = \max\{h(A(n_0)), h(A^\mu(0))\}$, it holds

$$\mathbb{P}\big(h_0 > a_\gamma \log N\big) \leq 2N^{-\gamma}$$

for N large enough. Consider the time–shifted IDLA process $A'(t) = A(n_0 + t)$ starting from logarithmic height with high probability. Then by Theorem 1.5 for any $\gamma > 0$ there exists a finite constant d_γ such that for $t_\gamma = d_\gamma N^2 \log N$ we can couple the clusters $A^\mu(t_\gamma), A'(t_\gamma)$ so that

$$\mathbb{P}\big(A^\mu(t_\gamma) \neq A'(t_\gamma), \ h_0 \leq a_\gamma \log N\big) \leq 3N^{-\gamma}$$

for N large enough. Let $T = t_\gamma + n_0$. Recall from (31) the definition of $D_{N,T}$, and define

$$D^\mu_{N,t_\gamma}(x, y) := N\Big(\mathbb{1}_{A^\mu_N(t_\gamma)}(x, y) - \mathbb{1}_{\{y \leq \frac{T}{N^2}\}}(x, y) \Big). \tag{32}$$

For φ as above, introduce the random variables

$$X^\varphi_N := \int_{\mathbb{T} \times \mathbb{R}} D_{N,T}(x, y) \varphi \Big(x, y - \frac{T}{N^2} \Big) \mathrm{d}x \mathrm{d}y,$$

$$X^{\mu,\varphi}_N := \int_{\mathbb{T} \times \mathbb{R}} D^\mu_{N,t_\gamma}(x, y) \varphi \Big(x, y - \frac{T}{N^2} \Big) \mathrm{d}x \mathrm{d}y.$$

Then for any $\delta > 0$ we have

$$\mathbb{P}(X_N^{\mu,\varphi} > \delta) \geq \mathbb{P}(X_N^{\varphi} > \delta, A^{\mu}(t_{\gamma}) = A(T)) \geq \mathbb{P}(X_N^{\varphi} > \delta) - 5N^{-\gamma}.$$

Moreover, by Theorem 8.2 for any $\varepsilon > 0$ we can take N large enough to ensure that, if $\mathcal{N}$ is a standard Gaussian random variable,

$$\mathbb{P}(X_N^{\varphi} > \delta) \geq \mathbb{P}\Big(\mathcal{N} > \frac{\delta}{\sqrt{v(\varphi)}}\Big) - \varepsilon \geq \frac{1}{2} - \Big(\frac{\delta}{\sqrt{2\pi v_{\mu}(\varphi)}} + \varepsilon\Big),$$

from which

$$\mathbb{P}(X_N^{\mu,\varphi} > \delta) \geq \frac{1}{2} - \Big(\frac{\delta}{\sqrt{2\pi v_{\mu}(\varphi)}} + \varepsilon + 5N^{-\gamma}\Big), \tag{33}$$

for any $\gamma, \delta, \varepsilon > 0$ and N large enough.

8.2. **The initial imbalance.** We now make a choice for the test function φ in the above discussion. Let

$$\phi(x,y) := \sin(2\pi x) = \frac{e^{2\pi i x} - e^{-2\pi i x}}{2},$$

for which $v_{\mu}(\phi) = \frac{1}{8\pi}$. Then by (33) we can take $\varepsilon = \delta$ and N large enough to get

$$\mathbb{P}(X_N^{\mu,\phi} > \delta) \geq \frac{1}{2} - 4\delta. \tag{34}$$

On the other hand,

$$X_N^{\mu,\phi} = \int_{\mathbb{T}\times\mathbb{R}} D_{N,t_{\gamma}}^{\mu}(x,y)\phi(x,y)\mathrm{d}x\mathrm{d}y = N\int_{\mathbb{T}\times\mathbb{R}} \phi(x,y)\mathbb{1}_{A_N^{\mu}(t_{\gamma})}(x,y)\mathrm{d}x\mathrm{d}y$$

$$= N\int_{A_N^{\mu}(t_{\gamma})} \phi(x,y)\mathrm{d}x\mathrm{d}y \overset{(d)}{=} N\int_{A_N^{\mu}} \phi(x,y)\mathrm{d}x\mathrm{d}y,$$

where $A^{\mu} \sim \mu_N$ and the last equality holds in distribution. In order to build a martingale, we now approximate ϕ by its discrete harmonic extension ψ away from the line $\{y = 0\}$ on the high probability event

$$E_{\mu} = \big\{h(A^{\mu}) \leq c_{\gamma} \log N\big\},$$

with c_{γ} as in Theorem 1.2 with $k = 0$. Following [9], to define ψ we introduce q_N solution of

$$\cosh(q_N/N) = 2 - \cos(2\pi/N), \tag{35}$$

so that $q_N = 2\pi + \mathcal{O}(N^{-2})$, and for $n = (n_1, n_2) \in \mathbb{Z}_N \times \mathbb{Z}$ set

$$\psi(n) := e^{q_N n_2/N} \sin\Big(\frac{2\pi n_1}{N}\Big). \tag{36}$$

It is easy to check that ψ is discrete harmonic on $\mathbb{Z}_N \times \mathbb{Z}$. Moreover, for $n = (n_1, n_2) \in A^{\mu}$ and $(x,y) \in Q_N(n)$ we have that on the good event E_{μ}

$$|\phi(x,y) - \psi(n)| = \Big|\sin(2\pi x) - e^{q_N n_2/N}\sin\Big(\frac{2\pi n_1}{N}\Big)\Big| \leq \frac{12\pi c_{\gamma}\log N}{N},$$

from which

$$N\int_{A_N^{\mu}} \phi(x,y)\mathrm{d}x\mathrm{d}y = \frac{1}{N}\sum_{n\in A^{\mu}} \psi(n) + \mathcal{O}\Big(\frac{\log^2 N}{N}\Big)$$

for N large enough. Combining this with (34), we get

$$\mathbb{P}\Big(\frac{1}{N}\sum_{n\in A^\mu}\psi(n)>\frac{\delta}{2}\Big)\geq\mathbb{P}\Big(N\int_{A_N^\mu}\phi(x,y)\mathrm{d}x\mathrm{d}y>\delta,E_\mu\Big)\geq\frac{1}{2}-4\delta-\mathbb{P}(E_\mu)\geq\frac{1}{2}-5\delta$$

for any $\delta>0$ and N large enough. The same arguments can be used to prove the reverse inequality, thus obtaining the following.

Proposition 8.3. *Let $A^\mu\sim\mu_N$ be a stationary IDLA cluster. For any $\delta>0$ and N large enough, it holds*

$$\mathbb{P}\Big(\frac{1}{N}\sum_{n\in A^\mu}\psi(n)>\delta\Big)\geq\frac{1}{2}-10\delta,\qquad\mathbb{P}\Big(\frac{1}{N}\sum_{n\in A^\mu}\psi(n)<-\delta\Big)\geq\frac{1}{2}-10\delta.$$

Let c_2 be defined as in Theorem 1.2 with $k=0$. We define

$$\begin{aligned}\Omega_\delta&:=\Big\{A\in\Omega:h(A)\leq c_2\log N\text{ and }\frac{1}{N}\sum_{n\in A}\psi(n)>\frac{\delta}{20}\Big\},\\\Omega_\delta'&:=\Big\{A\in\Omega:h(A)\leq c_2\log N\text{ and }\frac{1}{N}\sum_{n\in A}\psi(n)<-\frac{\delta}{20}\Big\}\end{aligned}\tag{37}$$

to have that

$$\mu_N(\Omega_\delta)\geq 1-\mu_N(\{A:h(A)>c_2\log N\})-\mathbb{P}\Big(\frac{1}{N}\sum_{n\in A^\mu}\psi(n)>\delta\Big)$$

$$\geq\frac{1}{2}-\frac{1}{N^2}-\frac{\delta}{2}\geq\frac{1}{2}-\delta$$

for N large enough. Similarly, $\mu_N(\Omega_\delta')\geq\frac{1}{2}-\delta$ for N large enough, and thus (4) in Theorem 1.4 is satisfied.

Remark 8.2. It follows from the above result that if A and A' are two independent samples of μ_N, then for any $\delta>0$ we can take N large enough so that

$$\mathbb{P}\Big(\Big|\frac{1}{N}\sum_{n\in A}\psi(n)-\frac{1}{N}\sum_{n\in A'}\psi(n)\Big|>2\delta\Big)\geq 2\Big(\frac{1}{2}-10\delta\Big)^2\geq\frac{1}{2}-20\delta,$$

which can be made arbitrarily close to $1/2$ by taking δ small enough.

8.3. **The observable.** We use the above remark to define a convenient observable which, loosely speaking, measures the difference in the horizontal imbalance of two IDLA processes. Take $A_0\in\Omega_\delta$ and $A_0'\in\Omega_\delta'$, and assume $|A_0|=|A_0'|$ without loss of generality, as if not then two IDLA processes starting from A_0 and A_0' will never meet. We take A_0,A_0' as starting configurations of two IDLA processes $(A(t))_{t\geq0}$ and $(A'(t))_{t\geq0}$.

Definition 8.1 (Imbalance). For $A\in\Omega$, define the *horizontal imbalance* of A by

$$u_A:=\frac{1}{N}\sum_{n\in A}\psi(n)=\frac{1}{N}\sum_{n\in A}e^{q_N n_2/N}\sin\Big(\frac{2\pi n_1}{N}\Big),$$

with q_N as in (35).

We use this to define an observable $u(t)$ which measures the difference in the imbalance of $A(t)$ and $A'(t)$, namely

$$u(t) := u_{A(t)} - u_{A'(t)} = \frac{1}{N} \sum_{n \in A(t)} \psi(n) - \frac{1}{N} \sum_{n \in A'(t)} \psi(n).$$

Remark 8.3. Since ψ is discrete harmonic on $\mathbb{Z}_N \times \mathbb{Z}$, we have that $(u(t))_{t \geq 0}$ is a discrete time martingale.

By Proposition 8.3 and Remark 8.2 we have

$$\mathbb{P}\big(|u(0)| > \delta\big) \geq \frac{1}{2} - 10\delta \tag{38}$$

for any $\delta > 0$ and N large enough. Clearly $A(t) = A'(t)$ implies $u(t) = 0$, so if we define

$$T_0 = \inf\{t \geq 0 : u(t) = 0\}$$

then for any $\alpha > 0$ we have

$$\mathbb{P}(T_0 \leq \alpha N^2, |u(0)| > \delta) \leq \mathbb{P}\Big(\sup_{t \leq \alpha N^2} |u(t) - u(0)| > \delta \Big)$$

$$\leq \frac{1}{\delta^2} \mathbb{E}\Big(\sup_{t \leq \alpha N^2} |u(t) - u(0)|^2 \Big) \leq \frac{c \mathbb{E}(Q)}{\delta^2} \tag{39}$$

for c absolute constant and

$$Q := \sum_{t=1}^{\alpha N^2} \mathbb{E}(|u(t) - u(t-1)|^2 | \mathcal{F}_{t-1}),$$

where the last inequality in (39) follows by the Burkholder-Davis-Gundy inequality. Now, if we let n_t, n_t' denote the settling locations of the tth walkers in the two IDLA processes, we have

$$Q = \frac{1}{N^2} \sum_{t=1}^{\alpha N^2} \mathbb{E}\Big(|\psi(n_t) - \psi(n_t')|^2 \Big| \mathcal{F}_{t-1} \Big),$$

so

$$\mathbb{E}(Q) \leq \mathbb{E}\Big(\frac{2}{N^2} \sum_{t=1}^{\alpha N^2} |\psi(n_t)|^2 \Big) + \mathbb{E}\Big(\frac{2}{N^2} \sum_{t=1}^{\alpha N^2} |\psi(n_t')|^2 \Big).$$

To estimate the above expectations we have to control the height of the clusters $A(\alpha N^2)$ and $A'(\alpha N^2)$, as the function ψ grows exponentially with them. We explain how to control the first expectation, the second one following from the same arguments. Introduce the good event

$$E := \{h(A(\alpha N^2)) \leq 100\alpha N\} \supseteq \{h(A(\alpha N^2)) \leq h(A_0) + 60\alpha N\}.$$

Then the a priori bound in Lemma C.1 with $m = 60$ gives

$$\mathbb{P}(E^c) \leq e^{-100\alpha N}$$

for N large enough. On E we have

$$\frac{1}{N^2} \sum_{t=1}^{\alpha N^2} |\psi(n_t)|^2 \leq \alpha\Big(\max_{n \in A(\alpha N^2)} |\psi(n)|^2 \Big) \leq \alpha \exp\Big\{ \frac{2q_N}{N} h(A(\alpha N^2)) \Big\} \leq C_\alpha$$

for some constant C_α depending only on α, and N large enough. On E^c, on the other hand, we trivially have that $h(A(\alpha N^2)) \leq h(A_0) + \alpha N^2 \leq 2\alpha N^2$, from which

$$\frac{1}{N^2} \sum_{t=1}^{\alpha N^2} |\psi(n_t)|^2 \leq \alpha \Big(\max_{n \in A(\alpha N^2)} |\psi(n)|^2 \Big) \leq \alpha \exp\Big\{ \frac{2q_N}{N} h(A(\alpha N^2)) \Big\} \leq \alpha \exp\{32\alpha N\},$$

since $q_N \leq 8$ for N large enough. In all, we have found

$$\mathbb{E}\left(\frac{1}{N^2} \sum_{t=1}^{\alpha N^2} |\psi(n_t)|^2 \right) \leq \mathbb{E}\left(\frac{1}{N^2} \sum_{t=1}^{\alpha N^2} |\psi(n_t)|^2; E \right) + \mathbb{E}\left(\frac{1}{N^2} \sum_{t=1}^{\alpha N^2} |\psi(n_t)|^2; E^c \right)$$

$$\leq C_\alpha + \alpha e^{32\alpha N} \mathbb{P}(E^c) \leq C_\alpha + \alpha e^{-68\alpha N} \leq 2C_\alpha$$

for N large enough. Similarly one can control the same expectation involving $A'(\alpha N^2)$. Thus

$$\mathbb{P}(T_0 \leq \alpha N^2, |u(0)| > \delta) \leq \frac{8cC_\alpha}{\delta^2},$$

where c is the absolute constant in (39). Let ε be as in Theorem 1.4. Since $C_\alpha \to 0$ as $\alpha \to 0$, we can take α small enough so that $\frac{8cC_\alpha}{\delta^2} \leq \varepsilon\big(\frac{1}{2} - 10\delta\big)$, to find

$$\mathbb{P}(T_0 \leq \alpha N^2, |u(0)| > \delta) \leq \varepsilon\Big(\frac{1}{2} - 10\delta\Big).$$

Finally, putting this together with (38) we gather that

$$\mathbb{P}\big(T_0 \leq \alpha N^2 \big| |u(0)| > \delta\big) = \frac{\mathbb{P}(T_0 \leq \alpha N^2, |u(0)| > \delta)}{\mathbb{P}(|u(0)| > \delta)} \leq \varepsilon,$$

which concludes the proof of Theorem 1.4.

APPENDIX A. PROOF OF LEMMA 6.4

Recall that we consider a random walk X on $\frac{1}{N}\mathbb{N}$, with a reflecting barrier at 0, with transition probabilities given by

$$X_0 = 0, \qquad \mathbb{P}\Big(X_{i+1} = 1 - \frac{1}{N} \big| X_i = 0\Big) = 1,$$

$$X_{i+1} - X_i = \begin{cases} 1 - \frac{1}{N}, & \text{with probability } \frac{1-\eta}{N}, \\ -\frac{1}{N}, & \text{otherwise} \end{cases} \qquad \text{for } X_i > 0. \tag{40}$$

We aim to show that

$$q := \mathbb{P}_{1-1/N}(X \text{ reaches } [\mathcal{E}^*, \infty) \text{ before } 0) \leq e^{-N}$$

for N large enough. Denote $\mathbb{P}_{1-1/N}$ simply by $\mathbb{P}$ to shorten the notation, and let $T_{0,\mathcal{E}^*}^X$ be the first time X reaches either 0 or $[\mathcal{E}^*, \infty)$. Then we have

$$q = \mathbb{P}\Big(X_{T_{0,\mathcal{E}^*}^X} \geq \mathcal{E}^*\Big) \leq \mathbb{P}\Big(X_{T_{0,\mathcal{E}^*}^X} \geq \mathcal{E}^*, T_{0,\mathcal{E}^*}^X \leq N^3 \log N\Big) + \mathbb{P}\Big(T_{0,\mathcal{E}^*}^X > N^3 \log N\Big)$$

$$\leq \mathbb{P}\Big(\sup_{t \leq N^3 \log N} X_t^{T_{0,\mathcal{E}^*}^X} \geq \mathcal{E}^*\Big) + \mathbb{P}\Big(T_0^X > N^3 \log N\Big),$$

where T_0^X is the first time the random walk X reaches 0, and $X^{T_{0,\mathcal{E}^*}^X}$ denotes the process X stopped upon reaching 0 or $[\mathcal{E}^*, \infty)$. We bound the two terms above separately.

For the rightmost term, it is easy to check that $M(t) = X_t + \frac{\eta t}{N}$ is a martingale up to time T_0^X. If we thus define $M_0(t) = M(t \wedge T_0^X)$, then $(M_0(t))_{t \geq 0}$ is a martingale for all times, with increments bounded by 1. It therefore follows from Azuma's inequality that for $t \geq 2N/\eta$,

$$
\begin{aligned}
\mathbb{P}\big(T_0^X > t\big) &= \mathbb{P}(T_0^X > t, X_t \geq 1/N) = \mathbb{P}\Big(T_0^X > t, X_t - X_0 \geq -1 + \frac{2}{N}\Big) \\
&= \mathbb{P}\Big(T_0^X > t, M(t) - M(0) \geq \frac{\eta t + 2}{N} - 1\Big) \leq \mathbb{P}\Big(M_0(t) - M_0(0) > \frac{\eta t}{N} - 1\Big) \\
&\leq \exp\Big(-\frac{(\frac{\eta t}{N} - 1)^2}{8t}\Big) \leq \exp\Big(-\frac{\eta^2 t}{32N^2}\Big),
\end{aligned}
$$

which can be made smaller than e^{-2N} by choosing $t \geq N^3 \log N$.

For the remaining term, again by Azuma we have

$$
\mathbb{P}\Big(\sup_{t \leq N^3 \log N} X_t^{T_{0,\mathcal{E}^*}^X} \geq \mathcal{E}^*\Big) \leq \mathbb{P}\Big(\sup_{t \leq N^3 \log N} M(t \wedge T_{0,\mathcal{E}^*}^X) \geq \mathcal{E}^*\Big) \leq \exp\Big(-\frac{(\mathcal{E}^*)^2}{8N^3 \log N}\Big) \leq e^{-2N}
$$

for N large enough. In all, we have found that

$$
q = \mathbb{P}\Big(X_{T_{0,\mathcal{E}^*}^X} \geq \mathcal{E}^*\Big) \leq e^{-2N} + e^{-2N} \leq e^{-N}
$$

for N large enough. This concludes the proof of Lemma 6.4.

Appendix B. Proof of Lemma 6.5

We have, for any $\lambda > 0$,

$$
\mathbb{P}\Big(\sum_{i=1}^{\gamma t} T_{\mathcal{E}^*}^{X,(i)} \leq t\Big) \leq \mathbb{P}\Big(\exp\Big(-\lambda \sum_{i=1}^{\gamma t} T_{\mathcal{E}^*}^{X,(i)}\Big) \geq e^{-\lambda t}\Big) \leq e^{\lambda t}\Big[\mathbb{E}\Big(e^{-\lambda T_{\mathcal{E}^*}^X}\Big)\Big]^{\gamma t}.
$$

To bound the right hand side above, recall that $T_{\mathcal{E}^*}^X \geq N_0 N$ for $N_0 \sim \text{Geometric}(p)$, with $p \leq e^{-N}$. Thus if we let $\hat{N}_0$ be a Geometric random variable of parameter $\hat{p} := e^{-N}$, then we find

$$
\begin{aligned}
\mathbb{E}\Big(e^{-\lambda T_{\mathcal{E}^*}^X}\Big) &= \int_0^\infty \mathbb{P}\big(e^{-\lambda T_{\mathcal{E}^*}^X} \geq t\big)\mathrm{d}t = \int_0^1 \mathbb{P}\Big(T_{\mathcal{E}^*}^X \leq -\frac{\log t}{\lambda}\Big)\mathrm{d}t \\
&\leq \int_0^1 \mathbb{P}\Big(\hat{N}_0 \leq -\frac{\log t}{\lambda N}\Big)\mathrm{d}t \leq \int_0^1 \Big(1 - (1 - \hat{p})^{-\frac{\log t}{\lambda N}}\Big)\mathrm{d}t \\
&\leq 1 - \int_0^1 \exp\Big(\frac{2 \log t}{\lambda N e^N}\Big)\mathrm{d}t = 1 - \frac{\lambda N e^N}{\lambda N e^N + 2},
\end{aligned}
$$

where for the last inequality we have used that $\log(1 - x) \geq -2x$ for $x \in (0, 1/2)$. Thus

$$
\begin{aligned}
\mathbb{P}\Big(\sum_{i=1}^{\gamma t} T_{\mathcal{E}^*}^{X,(i)} \leq t\Big) &\leq e^{\lambda t}\Big(1 - \frac{\lambda N e^N}{\lambda N e^N + 2}\Big)^{\gamma t} \\
&= \exp\Big\{\lambda t + \gamma t \log\Big(1 - \frac{\lambda N e^N}{\lambda N e^N + 2}\Big)\Big\} \\
&= \exp\Big\{-\lambda t\Big(-1 + \frac{\gamma N e^N}{\lambda N e^N + 2}\Big)\Big\}
\end{aligned}
$$

since $\log(1-x) \leq -x$. Taking $\lambda = \frac{1}{Ne^N}$ and recalling that $\gamma = \frac{6}{Ne^N}$, it is then easy to check that

$$\exp\left\{ -\lambda t \left(-1 + \frac{\gamma Ne^N}{\lambda Ne^N + 2} \right) \right\} \leq \exp(-\lambda t) = \exp\left(-\frac{t}{Ne^N} \right),$$

which concludes the proof.

Appendix C. Logarithmic fluctuations for IDLA at large times

We survey the proof of the logarithmic fluctuations bound by Jerison, Levine and Sheffield for IDLA on the cylinder graph $\mathbb{Z}_N \times \mathbb{Z}$, and extend it to larger times to prove Theorem 5.1.

C.1. **A priori bound.** We obtain an a priori bound on the height of $A(T)$ following the outer bound argument by Lawler, Bramson and Griffeath [11], for arbitrary starting configurations.

Lemma C.1. *Let $(A(t))_{t \geq 0}$ denote an IDLA process on $\mathbb{Z}_N \times \mathbb{Z}$, and let $h_0 = h(A(0))$ denote its initial height. Then for any $m \geq 3$ there exists $\beta = \beta(m) \in (0,1)$ such that, for $T \gg N \log N$ and N large enough, it holds*

$$\mathbb{P}\left(A(T) \not\subseteq R_{h_0 + \frac{mT}{N}} \right) \leq \beta^{T/N}.$$

Thus if, in particular, the process starts from the flat configuration $A(0) = A_0$, then $h(A(T)) \leq mT/N$ with high probability for N large enough.

Proof. Let $Z_k(t) := |A(t) \cap \{y = k\}|$ denote the number of particles in $A(t)$ at level k, and set $\mu_k(t) := \mathbb{E}(Z_k(t))$. Then

$$\mathbb{P}\left(A(T) \not\subseteq R_{h_0 + \frac{mT}{N}} \right) \leq \mathbb{P}\left(Z_{h_0 + \frac{mT}{N} + 1}(T) \geq 1 \right) \leq \mu_{h_0 + \frac{mT}{N} + 1}(T).$$

We claim that

$$\mu_k(t) \leq \left(\frac{1}{N} \right)^{k - h_0 - 1} \frac{t^{k - h_0}}{(k - h_0)!} \tag{41}$$

for all $k > h_0$ and $t \geq 0$. Indeed, clearly $\mu_1(t) \leq N$ and $\mu_k(0) = 0$ for all $k > h_0$. For other values of k, j we have

$$\mu_k(t+1) - \mu_k(t) = \mathbb{E}(Z_k(t+1) - Z_k(t)) = \mathbb{P}(Y_{t+1} \in A(t)^c \cap \{y = k\})$$

$$\leq \frac{1}{N} \mathbb{E}(|A(t) \cap \{y = k - 1\}|) = \frac{1}{N} \mu_{k-1}(t),$$

where in the above inequality we have used that the probability that the $(t+1)^{th}$ walker reaches level $k-1$ inside $A(t)$ is maximised when $A(t)$ is completely filled up to level $k-2$. Thus for $k > h_0$

$$\mu_k(t) = \sum_{s=0}^{t-1} (\mu_k(s+1) - \mu_k(s)) \leq \frac{1}{N} \sum_{s=0}^{t-1} \mu_{k-1}(s),$$

and (41) follows by a simple iteration. Now take $t = T$, $k = h_0 + \frac{mT}{N} + 1$ and recall that $k! \geq k^k e^{-k}$ to get

$$\mu_{h_0 + \frac{mT}{N} + 1}(T) \leq \left(\frac{1}{N} \right)^{\frac{mT}{N}} \frac{T^{\frac{mT}{N} + 1}}{(\frac{mT}{N} + 1)!} \leq N \left[\left(\frac{e}{m} \right)^m \right]^{\frac{T}{N}} \leq \beta^{T/N}$$

for any $\beta \in \left(\left(\frac{e}{m} \right)^m, 1 \right)$ and N large enough. $\hfill\square$

Remark C.1. By the above result with $h_0 = 0$, it suffices to prove Theorem 5.1 for large T. Suppose, indeed, that there exists a finite constant b such that $T \leq bN \log N$. Since $\frac{T}{N} \leq b \log N$, it suffices to take $a > b$ to have that the inner bound is trivially satisfied. For the outer bound we set $c = \frac{a+b}{3}$ and note that, as long as $a \geq 2b$, it holds

$$\mathbb{P}\left(\left\{ R_{\frac{T}{N} - a\log N} \subseteq A(T) \subseteq R_{\frac{T}{N} + a\log N} \right\}^c \right) = \mathbb{P}\left(A(T) \not\subseteq R_{(a+b)\log N} \right)$$
$$\leq \mathbb{P}\left(A(cN\log N) \not\subseteq R_{3c\log N} \right)$$
$$\leq \beta^{c\log N} = N^{-c\log \frac{1}{\beta}},$$

which can be made smaller than $N^{-\gamma}$ by taking $a \geq \max\left\{ 2b; \frac{3\gamma}{\log 1/\beta} - b \right\}$. In light of this observation, from now on we can assume that $T \gg N \log N$ as $N \to \infty$.

The proof of Theorem 5.1 follows an iterative argument, which we now sketch.

Definition C.1. A point (x, y) with $y \geq 0$ is said to be:
- m-early if $(x, y) \in A((y - m)N)$,
- ℓ-late if $(x, y) \notin A((y + \ell)N)$.

For $t \leq T$, let

$$\mathcal{E}_m[t] := \bigcup_{(x,y)\in A(t)} \{(x, y) \text{ is } m\text{-early}\}, \qquad \mathcal{L}_\ell[t] := \bigcup_{(x,y)\in R_{t/N}} \{(x, y) \text{ is } \ell\text{-late}\}.$$

Clearly,

$$\left\{ R_{\frac{T}{N} - a\log N} \subseteq A(T) \subseteq R_{\frac{T}{N} + a\log N} \right\}^c \subseteq \mathcal{E}_{a\log N}[T] \cup \mathcal{L}_{a\log N}[T],$$

so we bound the probability of the right hand side. Note that the a priori bound in the previous section (cf. Lemma C.1) with $m = 3$ tells us that there exists an absolute constant $\beta \in (0, 1)$ such that

$$\mathbb{P}\left(\mathcal{E}_{\frac{2T}{N}}[T] \right) \leq \beta^{T/N} \ll \beta^{\log N}$$

for N large enough. We use this to initialise the iteration, which consists of showing that, in turn,
- no m-early point implies no ℓ-late point ($\ell \ll m$), and
- no ℓ-late point implies no m'-early point ($m' \asymp \ell$).

To perform the above steps, we will use an explicit discrete harmonic function with a pole close to the early/late point to build a martingale, which will be then controlled via its quadratic variation.

C.2. **No early points implies no late points.** The main goal of this section is to prove the following.

Proposition C.1. *For any $\gamma > 0$ there exists a finite constant $C = C(\gamma)$, depending only on γ, such that, if $\ell \geq C \log N$ and $m \leq \ell^2/C \log N$, then*

$$\mathbb{P}(\mathcal{L}_\ell[T] \cap \mathcal{E}_m[T]^c) \leq N^{-(\gamma+1)}$$

for N large enough.

The above result tells us that on the event that there is no m-early point, there is no ℓ-late point with high probability, with $\ell = \lceil \sqrt{Cm \log N} \rceil \ll m$.

To prove this, we argue as follows. For $\zeta \in \mathbb{Z}_N \times \mathbb{Z}$, let $L(\zeta) = \{\zeta \text{ is } \ell\text{-late}\}$. Then, since

$$\mathcal{L}_\ell[T] = \bigcup_{\zeta \in R_{\frac{T}{N} - \ell}} L(\zeta),$$

we have

$$\mathbb{P}(\mathcal{L}_\ell[T] \cap \mathcal{E}_m[T]^c) \leq \sum_{\zeta \in R_{\frac{T}{N} - \ell}} \mathbb{P}(L(\zeta) \cap \mathcal{E}_m[T]^c).$$

It therefore suffices to show that

$$\mathbb{P}(L(\zeta) \cap \mathcal{E}_m[T]^c) \leq N^{-(\gamma+4)}$$

for arbitrary $\zeta \in R_{\frac{T}{N} - \ell}$. To see this, we use a discrete harmonic function to build a martingale with pole at ζ. We then show that on the event $L(\zeta)$ this martingale is large and negative, while on the event $\mathcal{E}_m[T]^c$ its quadratic variation is small. This will then imply that the probability that both $L(\zeta)$ and $\mathcal{E}_m[T]^c$ hold is small.

Recall that $(A(t))_{t \geq 0}$ denotes the IDLA process. For $\zeta = (\zeta_x, \zeta_y)$ and $z \in \mathbb{Z}_N \times \mathbb{Z}_+$ define

$$H_\zeta(z) := \mathbb{P}_z(\text{a SRW reaches level } \zeta_y \text{ for the first time at } \zeta),$$

where SRW stands for simple random walk on $\mathbb{Z}_N \times \mathbb{Z}$. Then $H_\zeta(z) \in [0,1]$ for all z and $H_\zeta(z) \to 1/N$ as $z_y \to -\infty$. Moreover, H_ζ is discrete harmonic up to level ζ_y (in fact, it is the discrete harmonic extension of the function $\mathbf{1}(x = \zeta_x)$ at level ζ_y to the region $\{(x,y) : y \leq \zeta_y\}$).

We embed the driving random walks in continuous time by mean of time-changed Brownian motions on the lattice (see [7] for a precise definition), that we denote by $\{(B_n(t))_{t \in [0,1]}, n \geq 1\}$, so that $B_n(0)$ and $B_n(1)$ are the starting and settling location of the nth random walk respectively. This turns out to be technically convenient, since it makes our discrete time martingales into continuous time ones. Finally, for real $t \in [0, \infty)$ we define

$$M_\zeta(t) := \sum_{n=1}^{\lfloor t \rfloor} \left(H_\zeta(B_n(1)) - \frac{1}{N} \right) + \left(H_\zeta(B_{\lfloor t \rfloor}(t - \lfloor t \rfloor)) - \frac{1}{N} \right).$$

Note that the first sum represents the contribution of the settled walkers, while the last term gives the contribution of the active one at time t. Since the function H_ζ is discrete harmonic up to level ζ_y, M_ζ above is a continuous-time martingale up to the first time the cluster reaches level ζ_y. We split it into a continuous part and a jump part, namely we set $M_\zeta(t) = M_\zeta^1(t) + M_\zeta^2(t)$ with

$$M_\zeta^1(t) = \sum_{n=1}^{\lfloor t \rfloor - 1} \left(H_\zeta(B_n(1)) - H_\zeta(B_n(0)) \right) + H_\zeta(B_{\lfloor t \rfloor}(t - \lfloor t \rfloor)) - B_{\lfloor t \rfloor}(0),$$

and

$$M_\zeta^2(t) = \sum_{n=1}^{\lfloor t \rfloor} \left(H_\zeta(B_n(0)) - \frac{1}{N} \right).$$

Let S_ζ, S_ζ^1 and S_ζ^2 denote the quadratic variation of M_ζ, M_ζ^1 and M_ζ^2 respectively. Since M_ζ^1 is a continuous martingale starting from 0, it is a time-changed Brownian Motion, that is

$$M_\zeta^1(t) = B(S_\zeta^1(t))$$

for B one-dimensional Brownian motion. Moreover, since M_ζ^2 is piecewise constant and by independence of the starting locations, we have

$$S_\zeta^2(t) = \sum_{n=1}^{\lfloor t \rfloor} \mathbb{E}\Big[\Big(H_\zeta(B_n(0)) - \frac{1}{N}\Big)^2\Big].$$

Recall that we want to bound the quadratic variation of M_ζ on the event that no point is m-early, which means that the cluster has a controlled height. The next lemma shows how a control on the shape of the cluster can be translated into a control of the quadratic variation of the martingale.

Lemma C.2. *Assume that $m \gg 1$ and that $\zeta_y \geq \frac{t}{N} + 2m + 1$. Then*

$$\mathbb{E}\Big(e^{S_\zeta(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}\Big) \leq e^2 t^{800}.$$

Proof. We have $S_\zeta(t) \leq 2(S_\zeta^1(t) + S_\zeta^2(t))$, from which

$$\mathbb{E}(e^{S_\zeta(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}) \leq \mathbb{E}(e^{2(S_\zeta^1(t)+S_\zeta^2(t))\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}}) \leq \Big[\mathbb{E}(e^{4S_\zeta^1(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c})\mathbb{E}(e^{4S_\zeta^2(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c})\Big]^{1/2}.$$

Let us estimate the two factors separately. As a technical trick, we will assume that the driving random walks are released from a uniformly chosen location at level $-N^2$, so that

$$\Big|H_\zeta(B_n(0)) - \frac{1}{N}\Big| \leq \frac{1}{N \log N}. \tag{42}$$

This clearly does not change the law of the process, and it has the technical advantage of transferring some amount of the quadratic variation of M_ζ from S_ζ^2 to S_ζ^1. We have:

$$\mathbb{E}(e^{4S_\zeta^2(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}) \leq \mathbb{E}\Big(\exp\Big\{4\sum_{n=1}^{\lfloor t \rfloor}\Big(H_\zeta(B_n(0)) - \frac{1}{N}\Big)^2\Big\}\Big)$$

$$\leq \mathbb{E}\Big(\exp\Big\{\frac{4\lfloor t \rfloor}{(N\log N)^2}\Big\}\Big) \leq e^4.$$

Note that the above bound does not depend on m, but only on the assumption (42). It remains to show that

$$\mathbb{E}(e^{4S_\zeta^1(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}) \leq t^{1600}.$$

To this end, we use that on the event $\mathcal{E}_{m+1}[t]^c$ the height on the cluster is controlled, which in turn implies that the quadratic variation of the continuous martingale part cannot be too large. Let us take t to be integer to simplify the writing. Then we find

$$\mathbb{E}(e^{4S_\zeta^1(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}) \leq \mathbb{E}\Big(\exp\Big\{4\sum_{n=1}^{t}(S_\zeta^1(n) - S_\zeta^1(n-1))\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}\Big\}\Big).$$

We claim that on the event $\mathcal{E}_{m+1}[t]^c$ the martingale increments are bounded, that is

$$M_\zeta^1(s) - M_\zeta^1(\lfloor s \rfloor) \in [-a, b_n]$$

for all $s \in [n, n+1)$ and some (non-random) a, b_n. Indeed,

$$M_\zeta^1(s) - M_\zeta^1(\lfloor s \rfloor) = H_\zeta(B_{\lfloor s \rfloor}(s - \lfloor s \rfloor)) - H_\zeta(B_{\lfloor s \rfloor}(0)) \geq -H_\zeta(B_{\lfloor s \rfloor}(0)) \geq -\frac{2}{N} =: -a.$$

Moreover, on the event $\mathcal{E}_{m+1}[t]^c$ we have that for all $n \leq t$ the cluster $A(n)$ is contained in $R_{\frac{n}{N}+m+1}$, from which

$$M_\zeta^1(s) - M_\zeta^1(\lfloor s \rfloor) = H_\zeta(B_{\lfloor s \rfloor}(s - \lfloor s \rfloor)) - H_\zeta(B_{\lfloor s \rfloor}(0))$$
$$\leq \max_{z: z_y \leq \frac{n}{N}+m+1} (H_\zeta(z) - H_\zeta(B_n(0)))$$
$$\leq \frac{10}{\zeta_y - \frac{n}{N} - m - 1} =: b_n,$$

where the last inequality follows from the explicit computation of $H_\zeta(z)$ for large N, and it holds as long as $m \gg 1$ (see [12] for details). By a standard argument on Brownian motion, this implies that

$$S_\zeta^1(n) - S_\zeta^1(n-1) \in \tau_n(-a, b_n),$$

where $\tau_n(a, b)$ denotes the exit time from the interval $(-a, b)$ for a one-dimensional Brownian motion starting from 0. We thus have that

$$\log \mathbb{E}\left(\exp\left\{ 4 \sum_{n=1}^t (S_\zeta^1(n) - S_\zeta^1(n-1)) \mathbf{1}_{\mathcal{E}_{m+1}[t]^c} \right\} \right) \leq \log \mathbb{E}\left(\exp\left\{ 4 \sum_{n=1}^t \tau_n(-a, b_n) \right\} \right)$$
$$\leq \sum_{n=1}^t \log \mathbb{E}\left(e^{4\tau_n(-a, b_n)} \right). \tag{43}$$

It is easy to check (cf. [7], Lemma 5) that $\mathbb{E}(e^{\lambda \tau(-a,b)}) \leq 1 + 10\lambda ab$ provided $\sqrt{\lambda}(a+b) \leq 3$. For our choice of parameters $\lambda = 4$ and

$$a + b_n = \frac{2}{N} + \frac{10}{\zeta_y - \frac{n}{N} - m - 1} \leq \frac{2}{N} + \frac{10}{\frac{t-n}{N} + m} \leq 1$$

for N large enough, since $m \gg 1$. Thus

$$\sum_{n=1}^t \log \mathbb{E}\left(e^{4\tau_n(-a, b_n)} \right) \leq \sum_{n=1}^t \log(1 + 40ab_n) \leq 40a \sum_{n=1}^t b_n \leq 400a \sum_{n=1}^t \frac{1}{\frac{t-n}{N} + 1}$$
$$\leq \frac{800}{N}\left(1 + \int_1^t \frac{dx}{x/N} \right) = \frac{800}{N}(1 + N \log t) \leq 1600 \log t,$$

which concludes the proof. $\qquad \square$

We can now show that no early points implies no late points.

Proof of Proposition C.1. Take $m = \lfloor \frac{\ell^2}{C \log N} \rfloor$ without loss of generality, with C as in the statement, and note that $m \geq \ell$ since $\ell \geq C \log N$ by assumption. Let $\zeta = (\zeta_x, \zeta_y)$ be such that $\zeta_y \leq \frac{T}{N} - \ell$, and set $T_1 = N(\zeta_y + \ell)$. Note that $T_1 \leq T$. Recall that $L(\zeta)$ denotes the event that ζ is ℓ-late, that is $\zeta \notin A(T_1)$. As already observed, it will suffice to show that

$$\mathbb{P}(L(\zeta) \cap \mathcal{E}_m[T]^c) \leq N^{-(\gamma+4)}$$

for arbitrary $\zeta \in R_{\frac{T}{N}-\ell}$. On the event $L(\zeta)$ we use the martingale M_ζ with pole at ζ. We will show that both $M_\zeta(T_1)$ is large and negative, and $S_\zeta(T_1)$ is small, so that the probability of both happening simultaneously is tiny.

Let $A_\zeta(t)$ denote the IDLA cluster at time t with points stopped upon reaching level ζ_y, counted with multiplicity. We have

$$M_\zeta(T_1) = \sum_{A_\zeta(T_1)} \left(H_\zeta(z) - \frac{1}{N} \right).$$

On the event $L(\zeta)$, no particle reaches ζ by time T_1. Moreover, $H_\zeta(z) = 0$ for all $z \neq \zeta$ at level ζ_y, while $H_\zeta(z) > 0$ if $z_y < \zeta_y$. It follows that the sum defining $M_\zeta(T_1)$ is maximised when as many particles as possible are below level ζ_y, that is when the cluster is completely filled up to level $\zeta_y - 1$. Hence

$$M_\zeta(T_1) \leq \underbrace{\sum_{z \in R_{\zeta_y-1}} \left(H_\zeta(z) - \frac{1}{N} \right)}_{0 \text{ (mean value pr.)}} + \sum_{z \in A_\zeta(T_1),\, z_y=\zeta_y} \left(H_\zeta(z) - \frac{1}{N} \right)$$

$$= -\frac{1}{N} \underbrace{\sharp\{\text{particles stopped at level } \zeta_y\}}_{T_1-N(\zeta_y-1)=N(\ell+1)} \leq -\ell.$$

To show that, on the other hand, the quadratic variation is small, we use the following lemma.

Lemma C.3. *Assume that $m \gg 1$ and $\ell \leq m$. Fix any ζ, and let $t = N(\zeta_y + \ell)$. Then*

$$\mathbb{E}(e^{S_\zeta(t)} \mathbf{1}_{\mathcal{E}_m[t]^c}) \leq e^{200m} t^{800}.$$

Using the above with $t = T_1$ and since $T_1 \leq e^m$ we gather that

$$\mathbb{E}(e^{S_\zeta(T_1)} \mathbf{1}_{\mathcal{E}_m[T_1]^c}) \leq e^{200m} T_1^{800} \leq e^{1000m}.$$

In conclusion, for any $s > 0$ we have

$$\mathbb{P}(L(\zeta) \cap \mathcal{E}_m[T]^c) \leq \mathbb{P}(\mathcal{E}_m[T]^c \cap \{S_\zeta(T_1) > s\}) + \mathbb{P}(\{S_\zeta(T_1) \leq s\} \cap L(\zeta)),$$

with

$$\mathbb{P}(\mathcal{E}_m[T]^c \cap \{S_\zeta(T_1) > s\}) \leq \mathbb{P}(S_\zeta(T_1)\mathbf{1}_{\mathcal{E}_m[T_1]^c} > s) \leq \mathbb{E}(e^{S_\zeta(T_1)}\mathbf{1}_{\mathcal{E}_m[T_1]^c})e^{-s} \leq e^{-1000m}$$

taking $s = 2000m$, and

$$\mathbb{P}(\{S_\zeta(T_1) \leq s\} \cap L(\zeta)) \leq \mathbb{P}(S_\zeta(T_1) \leq s,\, M_\zeta(T_1) \leq -\ell) \leq e^{-\ell^2/2s}$$

by a standard large deviations estimate for Brownian motion. In all,

$$\mathbb{P}(L(\zeta) \cap \mathcal{E}_m[T]^c) \leq e^{-1000m} + e^{-\ell^2/4000m} \leq 2N^{-(\gamma+5)},$$

as long as $m \geq \ell \geq (\gamma+5)\log N$ and $\frac{\ell^2}{4000m} \geq (\gamma+5)\log N$. This shows that Proposition C.1 holds with $C \geq 4000(\gamma+5)$.

In order to conclude, it remains to prove the above lemma.

Proof of Lemma C.3. As in the proof of Lemma C.2, we have

$$\mathbb{E}(e^{S_\zeta(t)}\mathbf{1}_{\mathcal{E}_m[t]^c}) \leq \mathbb{E}(e^{2(S^1_\zeta(t)+S^2_\zeta(t))\mathbf{1}_{\mathcal{E}_m[t]^c}}) \leq \left[\mathbb{E}(e^{4S^1_\zeta(t)}\mathbf{1}_{\mathcal{E}_m[t]^c})\mathbb{E}(e^{4S^2_\zeta(t)}\mathbf{1}_{\mathcal{E}_m[t]^c})\right]^{1/2}.$$

The factor involving S^2_ζ is bounded above by an absolute constant, so it suffices to show that, say,

$$\mathbb{E}(e^{4S^1_\zeta(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}) \leq e^{320m}t^{1600}.$$

Note that $\zeta_y = \frac{t}{N} - \ell$, so we cannot use Lemma C.2, which requires $\zeta_y \geq \frac{t}{N} + 2m + 1$. On the other hand, note that if $t_0 \leq t$ then $\mathcal{E}_m[t]^c \subseteq \mathcal{E}_m[t_0]^c$. Moreover, for t_0 such that $\zeta_y \geq \frac{t_0}{N} + 2m + 1$ we can apply Lemma C.2. Take therefore $t_0 = N(\zeta_y - 2m - 1) \vee 0$. Then

$$\mathbb{E}(e^{4S^1_\zeta(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}) \leq \left[\mathbb{E}(e^{8(S^1_\zeta(t)-S^1_\zeta(t_0))}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c})\underbrace{\mathbb{E}(e^{8S^1_\zeta(t_0)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c})}_{\leq t^{3200}}\right]^{1/2}$$

$$\leq \left[\mathbb{E}\left(\exp\left\{8\sum_{n=t_0+1}^{t}(S^1_\zeta(n)-S^1_\zeta(n-1))\mathbf{1}_{\mathcal{E}_m[t_0]^c}\right\}\right)\right]^{1/2}t^{1600}$$

$$\leq \left[\underbrace{\mathbb{E}(e^{8\tau(-\frac{2}{N},1)})}_{\leq 1+\frac{160}{N}\leq e^{160/N}}\right]^{\frac{t-t_0}{2}}t^{1600} \leq e^{\frac{80}{N}(t-t_0)}t^{1600}.$$

Now, if $t_0 = 0$ then $\zeta_y \leq 2m + 1$, so $\frac{t}{N} = \zeta_y + \ell \leq 2m + 1 + \ell \leq 4m$. Otherwise, $\frac{t-t_0}{N} = \ell + 2m + 2 \leq 4m$. In all,

$$\mathbb{E}(e^{4S^1_\zeta(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c}) \leq e^{320m}t^{1600},$$

as wanted. $\qquad\square$

This concludes the proof of Proposition C.1. $\qquad\square$

C.3. **No late points implies no early points.** We prove the following.

Proposition C.2. *Assume $m, \ell \leq \sqrt{3CN\log N}$, with C as in Proposition C.1. Then for any $\gamma > 0$, there exist a finite constant $C' = C'(\gamma)$, depending only on γ, and an absolute constant $b > 0$ such that, if $m \geq C'\log N$ and $\ell \leq \frac{bm}{48\pi}$, it holds*

$$\mathbb{P}(\mathcal{E}_m[T] \cap \mathcal{L}_\ell[T]^c) \leq N^{-(\gamma+1)}$$

for N large enough.

Proof. We assume $\ell = \frac{bm}{48\pi}$ without loss of generality, with b to be determined. Let

$$Q_{z,t} := \{z \text{ is the first } m\text{-early point, absorbed at time } t\}.$$

Then

$$\mathbb{P}(\mathcal{E}_m[T] \cap \mathcal{L}_\ell[T]^c) = \sum_{t=1}^{T}\sum_{z\in R_T}\mathbb{P}(Q_{z,t}\cap\mathcal{L}_\ell[T]^c).$$

On the event $Q_{z,t}$ take $\zeta = (\zeta_x, \zeta_y)$ such that $\zeta_y = z_x$ and $\zeta_y = z_y + m + 1 = \frac{t}{N} + 2m + 1$. Then for any $s > 0$

$$\mathbb{P}(Q_{z,t} \cap \mathcal{L}_\ell[T]^c) \leq \mathbb{P}(Q_{z,t} \cap \{S_\zeta(t) > s\}) +$$
$$\mathbb{P}\Big(Q_{z,t} \cap \big\{M_\zeta(t) < \frac{b}{1000}m\big\} \cap \mathcal{L}_\ell[T]^c\Big) +$$
$$\mathbb{P}\Big(\{S_\zeta(t) > s\} \cap \big\{M_\zeta(t) \geq \frac{b}{1000}m\big\}\Big).$$

Taking $s = (2\gamma + 1000)\log N$, we find

$$\mathbb{P}\Big(\{S_\zeta(t) > s\} \cap \big\{M_\zeta(t) \geq \frac{b}{1000}m\big\}\Big) \leq e^{-s/2} \leq N^{-(\gamma+500)},$$

and, using that $Q_{z,t} \subseteq \mathcal{E}_{m+1}[t]^c$,

$$\mathbb{P}(Q_{z,t} \cap \{S_\zeta(t) > s\}) \leq \mathbb{P}(\mathcal{E}_{m+1}[t]^c \cap \{S_\zeta(t) > s\}) = \mathbb{P}(S_\zeta(t)\mathbf{1}_{\mathcal{E}_{m+1}[t]^c} > s)$$
$$\leq \mathbb{E}(e^{S_\zeta(t)}\mathbf{1}_{\mathcal{E}_{m+1}[t]^c})e^{-s} \leq e^8 t^{800} e^{-s} \leq N^{-(2\gamma+100)}.$$

To finish, we show that

$$\mathbb{P}\Big(Q_{z,t} \cap \big\{M_\zeta(t) < \frac{b}{1000}m\big\} \cap \mathcal{L}_\ell[T]^c\Big) \leq N^{-(\gamma+5)}.$$

Note that on the event $Q_{z,t}$ we have $A(t) \subseteq R_{\frac{t}{N}+m+1}$. Let $B_r(z)$ denote the Euclidean ball of radius r around z. Partition $A(t)$ as follows:

$$A(t) = \Big(\underbrace{A(t) \cap R_{\frac{t}{N}-\ell}}_{A_1}\Big) \cup \Big(\underbrace{A(t) \cap B_m(z)}_{A_2}\Big) \cup \Big(\underbrace{A(t) \setminus (A_1 \cup A_2)}_{A_3}\Big).$$

Then

$$M_\zeta(t) = \sum_{z \in A_1} \Big(H_\zeta(z) - \frac{1}{N}\Big) + \sum_{z \in A_2} \Big(H_\zeta(z) - \frac{1}{N}\Big) + \sum_{z \in A_3} \Big(H_\zeta(z) - \frac{1}{N}\Big).$$

Since $A_1 = R_{\frac{t}{N}-\ell}$ on the event $\mathcal{L}_\ell[T]^c$,

$$\sum_{z \in A_1} \Big(H_\zeta(z) - \frac{1}{N}\Big) = 0.$$

Moreover,

$$\sum_{z \in A_3} \Big(H_\zeta(z) - \frac{1}{N}\Big) \geq -\frac{|A_3|}{N} = -\frac{t - (t - \ell N) - |A_2|}{N} \geq -\ell.$$

It remains to estimate the contribution to the martingale of points in A_2, which we show to be large. To this end, observe that if $i \leq m$ and $m \leq k \leq 2m + 1$, then (with $m \gg 1$)

$$\mathbb{P}_0(\text{a SRW reaches level } k \text{ for the first time at } (i,k)) \geq \frac{1}{4\pi} \log\Big(1 + \frac{4k}{i^2 + (k-1)^2}\Big) + \mathcal{O}\Big(\frac{1}{N^2}\Big)$$
$$\geq \frac{1}{2\pi} \frac{k}{i^2 + (k-1)^2} \geq \frac{1}{12\pi m}.$$

This implies that $H_\zeta(z) \geq \frac{1}{12\pi m}$ for $z \in A_2$. Moreover, the next lemma shows that A_2 contains a positive proportion of points, and it identifies the absolute constant b in Proposition C.2.

Lemma C.4 (Thin tentacles, cf. [7] Lemma 2)**.** *Assume $m \ll N$. Then there exist absolute constants b, C_0, c_0 such that for all z with $z_y \geq m$ it holds*

$$\mathbb{P}(z \in A(t), \ |A(t) \cap B_m(z)| \leq bm^2) \leq C_0 e^{-c_0 m},$$

for all t.

Proof. This can be proven exactly as in the $\mathbb{Z}^2$ case treated in [7], as the argument is completely local and $m \ll N$. □

This allows us to conclude that, on the event $Q_{z,t} \cap \mathcal{L}_\ell[T]^c \cap \{|A(t) \cap B_m(z)| \leq bm^2\}$, it holds

$$M_\zeta(t) \geq \sum_{z \in A_2} \left(H_\zeta(z) - \frac{1}{N} \right) - \ell \geq \left(\frac{1}{12\pi m} - \frac{1}{N} \right) |A_2| - \ell \geq \frac{bm}{12\pi} - \frac{bm^2}{N} - \ell \geq \frac{bm}{48\pi},$$

where the last inequality follows by recalling that $\ell = bm/48\pi$ and that, by assumption, $m^2 \leq 3CN \log N$ and $m \geq C' \log N$, from which

$$\frac{bm^2}{N} \leq 3bC \log N \leq \frac{bm}{24\pi}$$

as long as $3C \leq C'/24\pi$, which can always be ensured by taking C' large enough, depending only on C.

In all, we conclude that

$$\mathbb{P}\Big(Q_{z,t} \cap \Big\{ M_\zeta(t) < \frac{b}{1000} m \Big\} \cap \mathcal{L}_\ell[T]^c \Big) =$$

$$= \mathbb{P}\Big(Q_{z,t} \cap \Big\{ M_\zeta(t) < \frac{b}{1000} m \Big\} \cap \mathcal{L}_\ell[T]^c \cap \{ |A(t) \cap B_m(z)| \leq bm^2 \} \Big)$$

$$\leq \mathbb{P}(z \in A(t), \{ |A(t) \cap B_m(z)| \leq bm^2 \})$$

$$\leq C_0 e^{-c_0 m} \leq N^{-(\gamma+5)},$$

where the last inequality holds as long as $m \geq \frac{2}{c_0}(\gamma+5) \log N$, which can always be ensured at cost of increasing C', only depending on γ. This concludes the proof of Proposition C.2. □

C.4. **Iterative scheme.** To finally prove Theorem 5.1 we iteratively apply Propositions C.1 and C.2, starting with $m = 2T/N \gg \log N$ and iterating as long as $m, \ell \geq a_\gamma \log N$, with $a_\gamma = \max\{C, C'\}$. Assume that $T \gg N \log N$, otherwise the result holds by Remark C.1. We have from Lemma C.1 with $m = 3$ that

$$\mathbb{P}(A(T) \not\subseteq R_{3T/N}) \leq \mathbb{P}(A(N^2 \log^2 N) \not\subseteq R_{3T/N}) \leq \beta^N \leq N^{\gamma+1},$$

where, for all fixed $\gamma > 0$, the last inequality holds for N large enough. Let $m_0 = 3N \log^2 N$. For C, C' as in Proposition C.1 and C.2 respectively, iteratively define

$$\begin{cases} \ell_k = \sqrt{C m_{k-1} \log N}, \\ m_k = \frac{48\pi}{b} \ell_k, \end{cases} \quad k \geq 1.$$

Then, as long as $m_k, \ell_k \geq a_\gamma \log N$, we have

$$\mathbb{P}(\mathcal{E}_{m_{k-1}}[T]^c \cap \mathcal{L}_{\ell_k}[T]^c) \leq N^{-(\gamma+1)},$$

$$\mathbb{P}(\mathcal{L}_{\ell_k}[T]^c \cap \mathcal{E}_{m_k}[T]) \leq N^{-(\gamma+1)}$$

by Proposition C.1 and C.2 respectively. Thus

$$\mathbb{P}(\mathcal{E}_{m_0}[T]) \leq N^{-(\gamma+1)},$$

$$\mathbb{P}(\mathcal{L}_{\ell_1}[T]) \leq \mathbb{P}(\mathcal{L}_{\ell_1}[T] \cap \mathcal{E}_{m_0}[T]^c) + \mathbb{P}(\mathcal{E}_{m_0}[T]) \leq 2N^{-(\gamma+1)},$$

$$\mathbb{P}(\mathcal{E}_{\mu_1}[T]) \leq \mathbb{P}(\mathcal{E}_{m_1}[T] \cap \mathcal{L}_{\ell_1}[T]^c) + \mathbb{P}(\mathcal{L}_{\ell_1}[T]) \leq 3N^{-(\gamma+1)},$$

$$\vdots$$

$$\mathbb{P}(\mathcal{L}_{\ell_k}[T]) \leq 2kN^{-(\gamma+1)},$$

$$\mathbb{P}(\mathcal{E}_{m_k}[T]) \leq (2k+1)N^{-(\gamma+1)}.$$

We want to iterate as much as possible, that is to take k as large as possible so that $\ell_k, m_k \geq a_\gamma \log N$. It is easy to check that if $k > c_0 \log \log N$, for a suitable absolute constant c_0, then $\ell_k < a_\gamma \log N$, so we can iterate at most $c_0 \log \log N$ times. In all, this shows that by taking $\ell = \max\{\ell_{c_0 \log \log N}; a_\gamma \log N\}$ and $m = \max\{m_{c_0 \log \log N}; a_\gamma \log N\}$ we have

$$\mathbb{P}(\mathcal{E}_m[T] \cup \mathcal{L}_\ell[T]) \leq \mathbb{P}(\mathcal{E}_m[T]) + \mathbb{P}(\mathcal{L}_\ell[T]) \leq 5c_0 N^{-(\gamma+1)} \log \log N \leq N^{-\gamma},$$

for N large enough, as wanted.

REFERENCES

[1] David J Aldous and James Propp. *Microsurveys in discrete probability: DIMACS workshop, June 2-6, 1997*, volume 41. American Mathematical Soc., 1998.

[2] Amine Asselah and Alexandre Gaudillière. From logarithmic to subdiffusive polynomial fluctuations for internal DLA and related growth models. *Ann. Probab.*, 41(3A):1115–1159, 2013.

[3] Amine Asselah and Alexandre Gaudilliere. Lower bounds on fluctuations for internal DLA. *Probability Theory and Related Fields*, 158(1-2):39–53, 2014.

[4] Amine Asselah, Alexandre Gaudillière, et al. Sublogarithmic fluctuations for internal DLA. *The Annals of Probability*, 41(3A):1160–1179, 2013.

[5] Persi Diaconis and William Fulton. A growth model, a game, an algebra, lagrange inversion, and characteristic classes. *Rend. Sem. Mat. Univ. Pol. Torino*, 49(1):95–119, 1991.

[6] Tobias Friedrich and Lionel Levine. Fast simulation of large-scale growth models. *Random Structures & Algorithms*, 42(2):185–213, 2013.

[7] David Jerison, Lionel Levine, and Scott Sheffield. Logarithmic fluctuations for internal DLA. *J. Amer. Math. Soc.*, 25(1):271–301, 2012.

[8] David Jerison, Lionel Levine, and Scott Sheffield. Internal DLA and the Gaussian free field. *Duke Math. J.*, 163(2):267–308, 2014.

[9] David Jerison, Lionel Levine, and Scott Sheffield. Internal DLA for cylinders. *Advances in Analysis: The Legacy of Elias M. Stein*, page 189, 2014.

[10] David Jerison, Lionel Levine, Scott Sheffield, et al. Internal DLA in higher dimensions. *Electronic Journal of Probability*, 18, 2013.

[11] Gregory F. Lawler, Maury Bramson, and David Griffeath. Internal diffusion limited aggregation. *Ann. Probab.*, 20(4):2117–2140, 1992.

[12] Gregory F Lawler and Vlada Limic. *Random walk: a modern introduction*, volume 123. Cambridge University Press, 2010.

[13] Leonardo T Rolla and Vladas Sidoravicius. Absorbing-state phase transition for driven-dissipative stochastic dynamics on $\mathbb{Z}$. *Inventiones mathematicae*, 188(1):127–150, 2012.

[14] EL Wilmer, David A Levin, and Yuval Peres. Markov chains and mixing times. *American Mathematical Soc., Providence*, 2009.

Lionel Levine. Department of Mathematics, Cornell University, Ithaca, NY 14853.
E-mail address: levine@math.cornell.edu

Vittoria Silvestri. Statslab, Wilberforce road, Cambridge, CB3 0WA, United Kingdom.
E-mail address: V.Silvestri@maths.cam.ac.uk