The Annals of Probability
2017, Vol. 45, No. 6A, 4019–4070
DOI: 10.1214/16-AOP1157


# OIL AND WATER: A TWO-TYPE INTERNAL AGGREGATION MODEL

BY ELISABETTA CANDELLERO*,[1], SHIRSHENDU GANGULY†,[2], CHRISTOPHER HOFFMAN‡,[3] AND LIONEL LEVINE§,[4]

*University of Warwick*, University of California, Berkeley†, University of Washington‡ and Cornell University§*

We introduce a two-type internal DLA model which is an example of a nonunary abelian network. Starting with $n$ "oil" and $n$ "water" particles at the origin, the particles diffuse in $\mathbb{Z}$ according to the following rule: whenever some site $x \in \mathbb{Z}$ has at least 1 oil and at least 1 water particle present, it *fires* by sending 1 oil particle and 1 water particle each to an independent random neighbor $x \pm 1$. Firing continues until every site has at most one type of particles. We establish the correct order for several statistics of this model and identify the scaling limit under assumption of existence.

## CONTENTS



Received August 2014; revised May 2016.

[1]Supported in part by the Austrian Science Fund (FWF): E-1503W01230, and a Marie Curie Career Integration Grant PCIG09-GA2011-293619.

[2]Supported in part by NSF Grant DMS-0847661.

[3]Supported by NSF Grant DMS-1308645 and NSA Grant H98230-13-1-0827.

[4]Supported by NSF Grants DMS-1243606, DMS-1455272 and a Sloan Fellowship.

*MSC2010 subject classifications.* 60J10, 60K35, 82B41, 82C22, 82C24.

*Key words and phrases.* Internal DLA, nonunary abelian network, odometer function, random walk, scaling limit.

## 1. Introduction and main results.

We investigate a new interacting particle system on $\mathbb{Z}$ that can be considered as a model of *mutual diffusion*. Two particle species, called for convenience *oil* and *water*, diffuse on $\mathbb{Z}$ until there is no site that has both an oil and water particle. We start with $n$ oil and $n$ water particles at the origin. At each discrete time step, if at least 1 oil particle and at least 1 water particle are present at $x \in \mathbb{Z}$ then $x$ *fires* by sending 1 oil particle and 1 water particle each to an independent random neighbor $x \pm 1$ with equal probability. The system fixates when no more firing is possible, that is, when every site has particles of at most one type.

How many firings are required to reach fixation? How far is the typical particle from the origin upon fixation? Our main results address these two questions.

DEFINITION 1. For $x \in \mathbb{Z}$, let $u(x)$ be the total number of times $x$ fires before fixation. The random function $u : \mathbb{Z} \to \mathbb{N}$ is called the *odometer* of the process.

In the above informal description, we have assumed that all sites fire in parallel in discrete time, but in fact this system has an *abelian property*: the distribution of the odometer and of the particles upon fixation do not depend on the order of firings (Lemma 2.2).

Our first result concerns the order of magnitude of the odometer.

THEOREM 1.1. *There exist positive numbers* $\epsilon, c, C$ *such that for large enough* $n$:

i.

$$\mathbb{P}\Big(\sup_{x\in\mathbb{Z}} u(x) > Cn^{4/3}\Big) < e^{-n^{\epsilon}},$$

ii.

$$\mathbb{P}\Big(\inf_{x:|x|\le cn^{1/3}} u(x) < cn^{4/3}\Big) < e^{-n^{\epsilon}}.$$

The next result shows that most particles do not travel very far: all but a vanishing fraction of the particles at the end of the process are supported on an interval of length $n^{\alpha}$, for any exponent $\alpha > 1/3$. Formally, for $r > 0$ let $F(r)$ be the number of particles that fixate outside the interval $[-r, r]$.

THEOREM 1.2. *For sufficiently small* $\epsilon > 0$, *there exists* $\delta > 0$ *such that*

$$\mathbb{P}\big(F\big(n^{\frac{1}{3}+\varepsilon}\big) > n^{1-\frac{\epsilon}{2}}\big) < e^{-n^{\delta}}.$$

Theorem 1.1 shows that the odometer function scales like $n^{4/3}$. Theorem 1.2 motivates the conjecture that the proper scaling factor in the spatial direction should be $n^{1/3}$. We conjecture that the scaling limit of the odometer exists, and under the assumption of existence we identify the limiting function. Let $\tilde{u} := \mathbb{E}u$.

CONJECTURE 1.3.

(i) *For any* $\delta > 0$,

$$\mathbb{P}\bigg(\sup_{x\in\mathbb{Z}}\bigg|\frac{u(x)-\tilde{u}(x)}{n^{4/3}}\bigg| > \delta\bigg) \to 0 \qquad \text{as } n \to \infty. \tag{1}$$

(ii) *There is a function* $w : \mathbb{R} \to \mathbb{R}$ *such that*

$$\frac{\tilde{u}(\lfloor n^{1/3}\xi\rfloor)}{n^{4/3}} \to w(\xi), \tag{2}$$

*uniformly in* $\xi$.

Simulations support Conjecture 1.3, as shown in Figure 1. Conditionally on Conjecture 1.3, the following result identifies the limit $w(x)$ exactly.

THEOREM 1.4. *Assuming Conjecture* 1.3, *the function* $w$ *appearing in* (2) *must equal*

$$w(x) = \begin{cases} \dfrac{1}{72\pi}\big((18\pi)^{1/3} - |x|\big)^4, & |x| < (18\pi)^{1/3}, \\ 0, & |x| \ge (18\pi)^{1/3}. \end{cases} \tag{3}$$

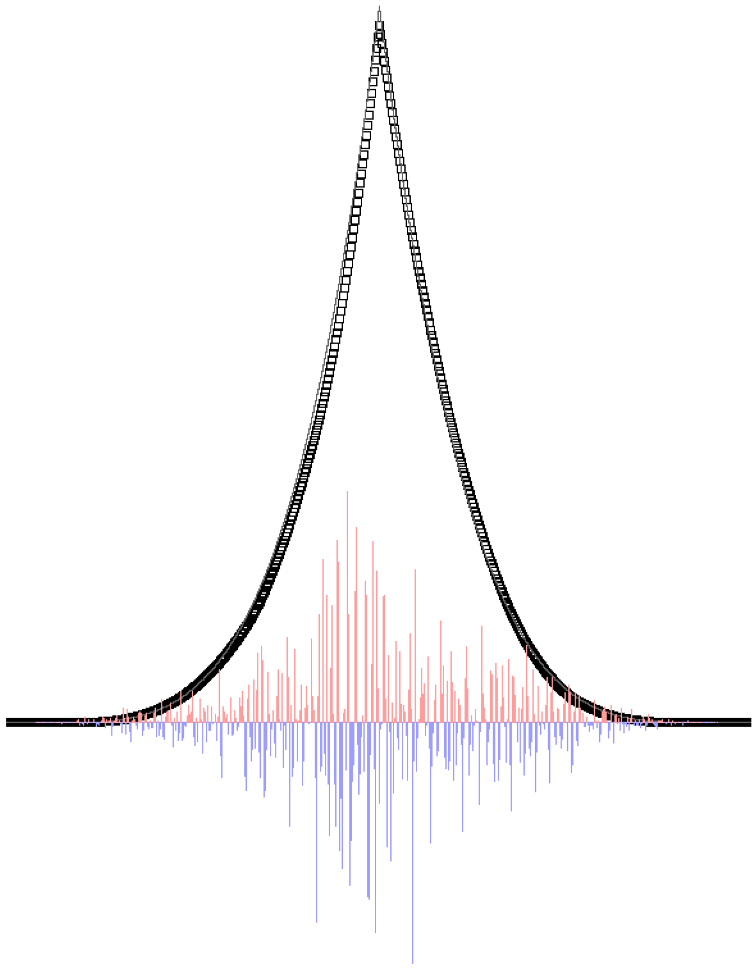

FIG. 1. *Graph of the odometer function of the oil and water model in* $\mathbb{Z}$ *with* $n = 360{,}000$ *particles of each type started at the origin*: *for each* $x \in \mathbb{Z}$ *a box is drawn centered at* $(x, u(x))$ *where* $u(x)$ *is the number of oil-water pairs fired from* $x$. *The curve* $f(x) = \frac{1}{72\pi}((18\pi n)^{1/3} - |x|)^4$ *appears in gray. Red and blue vertical bars represent the final configuration of oil and water particles, respectively*; *the height of the bar is proportional to the number of particles*.

The oil and water model can be defined on any graph and in particular on higher-dimensional lattices. Figure 2 shows an oil and water configuration in $\mathbb{Z}^2$. In Section 7, we conjecture the relevant exponents in $\mathbb{Z}^d$ for $d \geq 2$.

1.1. *Related models*: *Internal DLA and Abelian networks*. In *internal DLA*, each of $n$ particles started at the origin in $\mathbb{Z}^d$ performs a simple random walk until reaching an unoccupied site. The resulting random set of $n$ occupied sites is close to a Euclidean ball [see Lawler, Bramson and Griffeath (1992)]. Internal DLA is

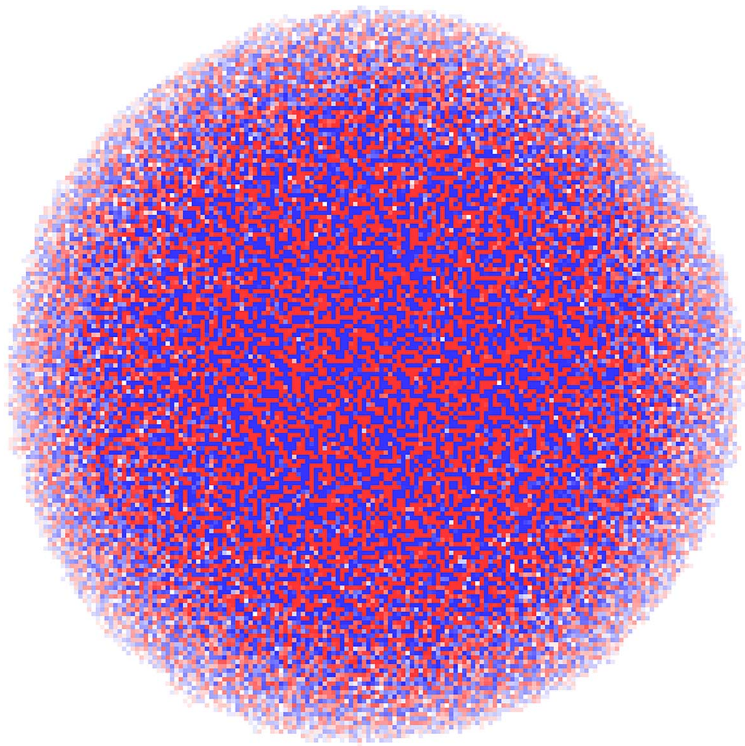

FIG. 2. *Oil and water in $\mathbb{Z}^2$ with $n = 2^{22}$ particles of each type started at the origin. Each site where particles stop is shaded red or blue according to whether oil or water particles stopped there. The intensity of the shade indicates the number of particles. We believe that the limit shape is a disk of radius of order $n^{1/4}$.*

one of several models known to have an *abelian property* according to which the order of moves does not affect the final outcome.

Dhar (1999) proposed certain collections of communicating finite automata as a broader class of models with this property. Until recently the only examples studied in this class have been *unary networks* [or their "block renormalizations" as proposed in Dhar (1999)]. Informally, a unary network is a system of local rules for moving *indistinguishable* particles on a graph, whereas a nonunary network has different types of particles. It is not as easy to construct nonunary examples with an abelian property, but they exist. Alcaraz, Pyatov and Rittenberg (2009) studied a class of nonunary examples which they termed *two-component sandpile models*, and asked whether there is a nontrivial example with two particle species such that the total number of particles of each type is conserved. Oil and water has this conservation property, but differs from the two-component sandpile models in that any number of particles of a single type may accumulate at the same vertex and be unable to fire.

Bond and Levine (2016) developed Dhar's idea into a theory of *abelian networks* and proposed two nonunary examples, *oil and water* and *abelian mobile agents*. Can such models exhibit behavior that is "truly different" from their unary cousins? This is the question that motivates the present paper. Theorem 1.2 shows that oil and water has an entirely different behavior than internal DLA: all but a vanishing fraction of the $2n$ particles started at the origin stop within distance $n^{\frac{1}{3}+\varepsilon}$ (versus $n$ for internal DLA).

In contrast to internal DLA where there is now a detailed picture of the fluctuations in all dimensions [see Asselah and Gaudillière (2013a, 2013b, 2014), Jerison, Levine and Sheffield (2012, 2013, 2014)], there is not even a limiting shape theorem yet for oil and water. Simulations in $\mathbb{Z}^d$ indicate a spherical limit shape (Figure 2) with radius of order $n^{1/(d+2)}$.

1.2. *Main ideas behind the proofs.* In this section, we present informally the key ideas behind the proofs. We start with a definition.

DEFINITION 2. For any $x \in \mathbb{Z}$ and $t \in \{0, 1, \dots, \infty\}$, let $\eta_1(x,t)$ and $\eta_2(x,t)$ represent the number of oil and water particles, respectively, at position $x$ at time $t$.

From now on, let

$$P_t = \sum_{x \in \mathbb{Z}} \min\bigl(\eta_1(x,t), \eta_2(x,t)\bigr)$$

be the number of co-located oil-water pairs at time $t$.

The process is run in phases: we start by firing a pair at the origin and, inductively, at each time $t > 0$ we fire one pair from a vertex $x_t$ determined according to a firing rule.[5] By definition, $P_0 = n$ and the process stops at the first time $\tau$ for which $P_\tau = 0$. Now denote by $l_t = \eta_1(x_t - 1, t) - \eta_2(x_t - 1, t)$ the excess of oil particles over water particles at the left neighbor of $x_t$ and by $r_t = \eta_1(x_t + 1, t) - \eta_2(x_t + 1, t)$ the excess at the right neighbor.

Conditional upon knowing the process up to time $t$, the random variable

$$Z_t := P_{t+1} - P_t \tag{4}$$

can have four possible distributions. To distinguish them, we denote the conditional $Z_t$ by the random variables $\xi_1, \xi_2, \xi_3, \xi_4$ which are described Table 1.

Note that $\xi_1$ has mean zero, whereas $\xi_2$ and $\xi_3$ have negative means, and $\xi_4$ is degenerate at 0. Since all such expected values are less than or equal to zero,

$$P_t \text{ is a supermartingale.}$$

[5] Lemma 2.2 shows that certain statistics of the process remain independent of the choice of the firing rule. The specific rule we use for convenience throughout the article is specified in Lemma 3.5.

TABLE 1
*The four possible conditional distributions for $Z_t$*

| $i$ | $\mathbb{P}(\xi_i = -1)$ | $\mathbb{P}(\xi_i = 0)$ | $\mathbb{P}(\xi_i = 1)$ | **Used when** |
|---|---|---|---|---|
| 1 | $1/4$ | $1/2$ | $1/4$ | $l_t r_t < 0$ |
| 2 | $1/4$ | $3/4$ | 0 | exactly one of $l_t, r_t$ is 0 |
| 3 | $1/2$ | $1/2$ | 0 | $l_t = r_t = 0$ |
| 4 | 0 | 1 | 0 | $l_t r_t > 0$ |

The main idea of the proof for the upper bound in Theorem 1.1 is as follows: Define the auxiliary random variables

$$N_i(t) = \#\{s \le t : Z_s \overset{d}{=} \xi_i\}. \tag{5}$$

It is clear from the definition that

$$\sum_{i=1}^{\tau} Z_i = P_\tau - P_0 = -n,$$

where $\tau$ is the stopping time of the process where all oil and water have been separated ($P_\tau = 0$). Now, informally $\sum_{i=1}^{\tau} Z_i$ is the sum of $N_2(\tau) + N_3(\tau)$ variables with mean at most $-1/4$ and $N_1(\tau)$ variables with mean 0.

Because of the negative and zero drifts, respectively, the sum of the negative mean variables is of order $O(-(N_2(\tau) + N_3(\tau)))$, whereas the sum of the zero mean variables is roughly $O(\sqrt{N_1(\tau)})$ (by square root fluctuations of the symmetric random walk). Thus, roughly

$$\sum_{i=1}^{\tau} Z_i \le -N_2(\tau) - N_3(\tau) + \sqrt{N_1(\tau)}.$$

We then argue by contradiction: conditional upon the event that the odometer is "very large" somewhere (i.e., larger than $Cn^{4/3}$ for a suitable big constant $C$), we show that it is likely that $N_2(\tau) + N_3(\tau)$ is "sufficiently large" compared to $\sqrt{N_1(\tau)}$. This in turn implies that

$$\sum_{i=1}^{\tau} Z_i < -n,$$

which is a contradiction.

To prove the lower bound we first establish a gradient bound on the odometer using the upper bound. In other words, we show that there exists a constant $c$ such that, with high probability, for $x \in [0, cn^{1/3}]$ we have

$$u(x) - u(x+1) \ge n/2.$$

This in turn implies that with high probability,

$$u(0) = \Omega(n^{4/3}),$$

since

$$0 \le u(\lfloor cn^{1/3} \rfloor) \le u(0) - \frac{n}{2}\lfloor cn^{1/3} \rfloor.$$

Theorem 1.2 follows from the proof of the lower bound.

Lastly, we discuss the proof of Theorem 1.4. Clearly, by symmetry of the process about the origin the function $w(\cdot)$ is symmetric as well. We show that Conjecture 1.3 implies that the limiting function is smooth on the positive real axis and, in particular, satisfies

$$w''(x) = \sqrt{\frac{2}{\pi} w(x)},$$

with certain boundary conditions. At this point, Theorem 1.4 follows by identifying a solution of the above boundary value problem, and using uniqueness of the solution.

1.3. *Outline*. This article is structured as follows. In Section 2, we give a rigorous definition of the model. Furthermore, in order to facilitate our proofs, we define two different versions of the process and prove that they both terminate in finite time with probability one. We then construct a coupling between the two versions, which in particular implies that they terminate with the same odometer function and the same distribution of particles.

In Section 3, we show a polynomial bound on the stopping time $\tau$. We start by showing that the number of pairs $P_t$ can be stochastically dominated by a certain lazy random walk with long holding times, started and reflected at $n$. In order to get a rough upper bound on $\tau$, it suffices to bound the hitting time of zero for this walk, which we show is of order $n^4$. Consequently, we get direct polynomial bounds on the support and the maximum of the odometer function. We then improve the last bound, proving the upper bound in Theorem 1.1 in Section 4.

Section 5 is devoted to proving the lower bound in Theorem 1.1. As a byproduct of the proof, Theorem 1.2 follows.

In Section 6, we prove the conditional Theorem 1.4. Section 7 consists of open questions and a conjecture about the process on higher dimensional lattices.

**2. Rigorous definition of the model.** As the underlying randomness for our model, we will take a countable family of independent random variables

$$\omega = (X_k^x, Y_k^x)_{x \in \mathbb{Z}, k \in \mathbb{N}} \tag{6}$$

with $\mathbb{P}(X_k^x = \pm 1) = P(Y_k^x = \pm 1) = \frac{1}{2}$. For each $x \in \mathbb{Z}$, the sequences $(X_k^x)_{k \ge 1}$ and $(Y_k^x)_{k \ge 1}$ are called the *stacks at* $x$. Denote by $\Omega$ the set of all stacks $\omega$.

The stacks will have the following interpretation (described formally below): On the $k$th firing from site $x$, an oil particle steps from $x$ to $x + X_k^x$ and a water particle steps from $x$ to $x + Y_k^x$.

For any value $K \in \mathbb{N}$, a *firing sequence* is a sequence of vertices $s = (x_0, \dots, x_{K-1})$ with each $x_k \in \mathbb{Z}$. Recall from Definition 2 that the random variables $\eta_i(x,t)$ represent the number of particles of type $i$ (type 1 are *oil* and type 2 are *water* particles) at location $x$ at time $t$. Given such a sequence and an initial state $\eta_1(\cdot,0), \eta_2(\cdot,0)$ we define the oil and water process $(\eta_1(\cdot,k), \eta_2(\cdot,k))_{k=0}^K$ inductively by

$$\eta_1(\cdot, k+1) = \eta_1(\cdot, k) - \delta(x_k) + \delta\big(x_k + X_{i_k}^{x_k}\big), \tag{7}$$

$$\eta_2(\cdot, k+1) = \eta_2(\cdot, k) - \delta(x_k) + \delta\big(x_k + Y_{i_k}^{x_k}\big), \tag{8}$$

where $i_k = \#\{j \le k : x_j = x_k\}$. Here, $\delta(x)$ is the function taking value 1 at $x$ and 0 elsewhere.

### 2.1. *Least action principle and Abelian property.*

DEFINITION 3. Let $s = (x_0, \dots, x_{K-1})$ be a firing sequence. We say that $s$ is *legal* for $(\eta_1(\cdot,0), \eta_2(\cdot,0))$ if

$$\min\big(\eta_1(x_k,k), \eta_2(x_k,k)\big) \ge 1,$$

for all $k$, $0 \le k \le K-1$. We say that $s$ is *complete* for $(\eta_1(\cdot,0), \eta_2(\cdot,0))$ if the final configuration $(\eta_1(\cdot,K), \eta_2(\cdot,K))$ satisfies

$$\min\big(\eta_1(x,K), \eta_2(x,K)\big) \le 0,$$

for all $x \in \mathbb{Z}$.

Making Definition 1 precise, we define the *odometer* of a firing sequence $s$ to be the function $u_s : \mathbb{Z} \to \mathbb{N}$ given by

$$u_s(x) = \#\{0 \le k < K : x_k = x\}. \tag{9}$$

How does $u_s$ depend on $s$? The least action principle for abelian networks addresses this question.

LEMMA 2.1 [Least action principle, Bond and Levine (2016)]. *Let $s$ and $s'$ be firing sequences. If $s$ is legal for $(\eta_1(\cdot,0), \eta_2(\cdot,0))$ and $s'$ is complete for $(\eta_1(\cdot,0), \eta_2(\cdot,0))$, then $u_s(x) \le u_{s'}(x)$ for all $x \in \mathbb{Z}$.*

We remark that this statement holds pointwise for any stacks $\omega$, even if $s$ and $s'$ are chosen by an adversary who knows the stacks.

In this paper, we will not use the full strength of Lemma 2.1. Only the following corollaries will be used.

LEMMA 2.2 [Abelian property, Bond and Levine (2016)]. *For fixed stacks* $\omega$ *and fixed initial state* $(\eta_1(\cdot,0),\eta_2(\cdot,0))$:

(i) *If there is a complete firing sequence of length* $K$, *then every legal firing sequence has length* $\leq K$.

(ii) *If firing sequences* $s$ *and* $s'$ *are both legal and complete*, *then* $u_s = u_{s'}$.

(iii) *Any two legal and complete firing sequences result in the same final state* $(\eta_1(\cdot,K),\eta_2(\cdot,K))$.

2.2. *Constant convention.* To avoid cumbersome notation, we will often use the same letter (generally $C$, $C'$, $c$, $c'$, $\epsilon$ and $\delta$) for a constant whose value may change from line to line.

**3. Preliminary bound on the stopping time.** We denote by $\tau \in \mathbb{N}\cup\{\infty\}$, the stopping time of the process:

$$\tau = \min\{k \geq 0 : \min(\eta_1(x,k),\eta_2(x,k)) = 0 \text{ for all } x \in \mathbb{Z}\}. \tag{10}$$

In this section, we prove a preliminary bound on the stopping time $\tau$. The bound is very crude and will be improved in the next section where we prove the upper bound in Theorem 1.1. However, this bound will be used in several places throughout the article. Note that by Lemma 2.2, given the stacks [see (6)], the stopping time $\tau$ and the odometer function $u := u_s$ are well defined and do not depend on the firing sequence $s$.

PROPOSITION 3.1. *Almost surely for any realization of the stacks*, $\tau$ *is finite. Moreover*:

i. *For any given* $\epsilon > 0$, *there exists* $c = c(\epsilon) > 0$ *such that*

$$\mathbb{P}(\tau > n^{4+2\epsilon}) \leq e^{-n^c}.$$

ii. $\mathbb{E}(\tau) \leq 16n^4$.

This result has an immediate consequence, given by the next corollary.

COROLLARY 3.2. *Given* $\epsilon > 0$, *there exists a constant* $c$ *such that*

$$\mathbb{P}(u(x) = 0, \forall x \in \mathbb{Z} : |x| \geq n^{4+\epsilon}) \geq 1 - e^{-n^c}.$$

Before proving Proposition 3.1, we recall from Section 1.2 that

$$P_t = \sum_{x\in\mathbb{Z}} \min(\eta_1(x,t),\eta_2(x,t)) \tag{11}$$

denotes the total number of co-located oil and water pairs at time $t$. By definition, the process stops at $\tau$ when $P_\tau = 0$. In order to prove Proposition 3.1, we follow the four steps described in the following:

- First, we define a lazy reflected random walk $R_i$ on $\{0, 1, \ldots, n\}$ started and reflected at $n$, and stopped upon hitting 0. Define the following stopping time:

$$\tau' := \min_{i \geq 1}\{R_i = 0\}. \tag{12}$$

- Consequently, we use $P_j$ to define a series of stopping times $t_i$, and we show that we can bound the tails of the waiting times defined as

$$\mathrm{Wait}_i := t_{i+1} - t_i. \tag{13}$$

- Next, we show that $P_{t_i}$ is stochastically dominated by $R_i$.
- Finally, we prove Proposition 3.1 by combining information about the distributions of $\tau'$ and the waiting times $\mathrm{Wait}_i$.

The lazy reflected random walk $R_i$ is defined as follows. It starts with $R_0 = n$. Inductively, if $R_i = n$ then $R_{i+1} = n$ or $n - 1$, with probability $3/4$ and $1/4$, respectively. Otherwise, if $R_i \in (0, n)$ then it performs a lazy random walk, that is,

$$R_{i+1} = \begin{cases} R_i + 1, & \text{w.p. } 1/4, \\ R_i - 1, & \text{w.p. } 1/4, \\ R_i, & \text{w.p. } 1/2. \end{cases}$$

The walk terminates at the stopping time $\tau'$, defined in equation (12).

To perform the second and third steps, we analyze $P_t$ and show that it is a supermartingale. This was described heuristically in Section 1.2, and here we make this concept formal, as this fact is the key to many of the subsequent arguments.

If at time $t$ a site $x$ emits a pair, then $P_{t+1}$ is either $P_t - 1$, $P_t$, or $P_t + 1$. Moreover, the distribution of $P_{t+1} - P_t$ depends on the state of the neighbors of $x$ at time $t$ in a fairly simple way.

Define $\mathcal{G}_t$ to be the filtration given by the first $t$ firings. Recall from Definition 2 that $\eta_1(x, t)$ and $\eta_2(x, t)$ are the number of oil and water particles, respectively, at $x$ at time $t$.

DEFINITION 4. Define for any $x \in \mathbb{Z}$ and nonnegative integer time $t$ including infinity:

$$g_t(x) := \eta_1(x, t) - \eta_2(x, t). \tag{14}$$

We say that at time $t$ a site $x \in \mathbb{Z}$ has *type oil*, 0 or *water* depending on the type of the majority of particles at the site, that is, whether $g_t(x)$ is positive, zero or negative, respectively.

We look at how $P_t$ changes conditional on the filtration at time $t$. Formally, we look at

$$Z_t := P_{t+1} - P_t. \tag{15}$$

At this point, we observe that $Z_t$ can have four possible distributions conditioned on $\mathcal{G}_t$. Namely, define four random variables $\xi_1, \xi_2, \xi_3, \xi_4$ as follows:

$$\xi_1 := \begin{cases} 0, & \text{w.p. } 1/2, \\ 1, & \text{w.p. } 1/4, \\ -1, & \text{w.p. } 1/4, \end{cases} \qquad \xi_2 := \begin{cases} 0, & \text{w.p. } 3/4, \\ -1, & \text{w.p. } 1/4, \end{cases}$$

$$\xi_3 := \begin{cases} 0, & \text{w.p. } 1/2, \\ -1, & \text{w.p. } 1/2, \end{cases} \qquad \xi_4 := 0, \quad \text{w.p. } 1.$$

REMARK 1. Every time that a vertex $x$ fires, we divide the possible *types* of the neighbors of $x$ into four groups which determine the law of $Z_t$:

$$Z_t \stackrel{\text{law}}{=} \begin{cases} \xi_1, & \text{if the neighbors have different } \textit{types} \text{ but neither is } 0, \\ \xi_2, & \text{if exactly one of the neighbors has } \textit{type } 0, \\ \xi_3, & \text{if both neighbors have } \textit{type } 0, \\ \xi_4, & \text{if the neighbors have the same nonzero } \textit{type}. \end{cases}$$

DEFINITION 5. A *firing rule* is an inductive way to determine a legal firing sequence. More formally, it is a function that is defined for any $t$ and any atom $A \subset \mathcal{G}_t$ such that $P_t > 0$. The firing rule outputs an integer $z$ such that there is at least one pair at $z$ at time $t$ in $A$.

The stacks combined with a firing rule determine the evolution of the process. Given a firing rule, we inductively define the set of stopping times $t_i$. Set $t_1 = 1$.

Then, inductively, given $t_1, \dots, t_i$ set $t_{i+1}$ to be the first time $j$ after $t_i$ such that $Z_j$ does not have distribution $\xi_4$. Let $t_L$ be the last stopping defined. Thus,

$$L := \#\{i \le \tau : Z_i \stackrel{\text{law}}{\neq} \xi_4\}. \tag{16}$$

Note that with this definition of the stopping times we have

$$P_{t_{i+1}} = P_{t_i+1}.$$

Thus, the distribution of $P_{t_{i+1}} - P_{t_i}$ is either $\xi_1$, $\xi_2$ or $\xi_3$. Finally, define the waiting times:

$$\text{Wait}_i = t_{i+1} - t_i.$$

With these definitions, we are ready to complete the second and third steps of our outline.

LEMMA 3.3. *For any firing rule the sequence $R_i$ stochastically dominates $P_{t_i}$. This in particular implies that $\tau'$ stochastically dominates $L$.*

PROOF. The proof is by induction, and the starting configuration is given by $R_0 = P_0 = n$. If $R_i = n$, then the distribution of $R_{i+1} - R_i$ is $\xi_3$, by definition of the reflected random walk.

Inductively, if $P_{t_i} \geq R_i = n$ then $P_{t_i}$ must also be $n$. This means that all sites have type 0 and the distribution of $P_{t_{i+1}} - P_{t_i}$ is $\xi_3$. Thus we can couple $R_{i+1}$ and $P_{t_{i+1}}$ such that $R_{i+1} \geq P_{t_{i+1}}$.

On the other hand, if $0 < R_i < n$ then the distribution of $R_{i+1} - R_i$ is $\xi_1$ while the distribution of $P_{t_{i+1}} - P_{t_i}$ is either $\xi_1$, $\xi_2$ or $\xi_3$. As all of $\xi_1$, $\xi_2$ and $\xi_3$ are stochastically dominated by $\xi_1$ we can couple $R_{i+1}$ and $P_{t_{i+1}}$ such that $R_{i+1} \geq P_{t_{i+1}}$. $\square$

LEMMA 3.4. *For any $\epsilon > 0$, there exists $\delta$ such that*

$$\mathbb{P}(L > n^{2+\epsilon}) \leq \mathbb{P}(\tau' > n^{2+\epsilon}) < e^{-n^{\delta}},$$

*where $\tau'$ and $L$ are defined in* (12) *and* (16), *respectively.*

PROOF. The first inequality follows from the fact that $\tau'$ stochastically dominates $L$. The second inequality is a standard fact about lazy random walk (stated later in Lemma 5.1). $\square$

LEMMA 3.5. *Let $E_i$ be an i.i.d. sequence of random variables whose distribution is the same as the time taken by simple symmetric random walk started from the origin to hit $\pm 2n$. There exists a firing rule such that $\mathit{Wait}_i$ is stochastically dominated by $E_i$.*

PROOF. At time 1, and at all times $j > 1$ such that $Z_{j-1}$ does not have distribution $\xi_4$, we fire the leftmost oil-water pair. For the remaining times $j$, either both neighbors of the site $x$ fired at time $j-1$ had an excess of water, in which case we fire a pair from the neighbor of $x$ which just received an oil particle; or both neighbors of $x$ had an excess of oil, in which case we fire a pair from the neighbor of $x$ which just received a water particle. This neighbor has the uniform distribution on $\{x+1, x-1\}$ independent of the past. Thus, for each time interval $t_i < j < t_{i+1}$ (where $t_1 = 1$ and $t_{i+1}$ is the first time $j > t_i$ that $Z_j$ does not have distribution $\xi_4$) the location of the site fired at time $j$ performs a simple random walk on $\mathbb{Z}$. We now show that this firing rule allows us to control the waiting times.

Let $z$ be the location of the leftmost pair at time $t_i$ and let $I = (a, b)$ be the interval of integers containing $z$ where every location has at least one particle at time $t_i$. As there are at most $2n$ particles, we have that $I = (a, b) \subset (z - 2n, z + 2n)$.

If there are firings only in the interior of $I$, then there are no particles at $a$ or $b$. If the site being fired reaches the boundary of $I$ at time $k$, then the distribution of $Z_k$ is not $\xi_4$, as one of the neighbors of the pair to be fired (either $a$ or $b$) has no particles and is of *type* 0. Thus, $t_{i+1} \le k$. This tells us that the distribution of the waiting time $\mathrm{Wait}_i = t_{i+1} - t_i$ is bounded by the time taken by simple random walk started at the origin to leave the interval $(-2n, 2n)$. Since this is true for each $i$ independently, the lemma is proven. $\square$

PROOF OF PROPOSITION 3.1. We first notice that by Lemma 2.2, $\tau$ defined in (10) is independent of the firing rule. For the purposes of the proof, we will fix our firing rule to be the one described in Lemma 3.5. Now by Lemmas 3.3 and 3.5,

$$\sum_{i=1}^{\tau'} E_i \text{ stochastically dominates } \tau.$$

Thus, $\mathbb{P}(\tau > n^{4+\epsilon}) \le \mathbb{P}(\sum_{i=1}^{\tau'} E_i \ge n^{4+\epsilon})$. If $\sum_{i=1}^{\tau'} E_i \ge n^{4+\epsilon}$ then either,

$$\tau' > 0.1n^{2+\epsilon/2} \quad \text{or} \quad \sum_{i=1}^{0.1n^{2+\epsilon/2}} E_i > n^{4+\epsilon/2}.$$

Therefore,

$$\mathbb{P}(\tau > n^{4+\epsilon}) \le \mathbb{P}(\tau' > 0.1n^{2+\epsilon/2}) + \mathbb{P}\left(\sum_{i=1}^{0.1n^{2+\epsilon/2}} E_i > n^{4+\epsilon}\right).$$

The first term is bounded by Lemma 3.4. The second is bounded using standard bounds on the distribution of the time for simple random walk started at the origin to leave a fixed interval $(-k, k)$. Both of these probabilities are bounded by $e^{-n^\delta}$ for some $\delta = \delta(\epsilon) > 0$.

Furthermore, by Lemma 3.5 we have that the $E_i$'s are all i.i.d., hence we can apply Wald's identity [see, e.g., Section 3.1 in Durrett (2010)] and get that

$$\mathbb{E}(\tau) \le \mathbb{E}(\tau')\mathbb{E}(E_1) \le 16n^4,$$

which completes the proof. $\square$

For definiteness, throughout the rest of the article we run the process according to the firing rule described in the proof of Lemma 3.5. We remark that any other firing rule would work as well.

## 4. Proof of Theorem 1.1 (upper bound).

4.1. *Notation.* In order to proceed, we need to introduce some further notation. Let

$$\|u\|_\infty = \max\{u(x) : x \in \mathbb{Z}\}.$$

In this section, we will bound from below the probability of the event

$$\mathcal{G} = \{\|u\|_\infty \le Cn^{4/3}\} \tag{17}$$

for some large constant $C$. To this purpose, we define

$$\mathrm{Returns}(x) := \#\{t \le \tau : \eta_1(x,t) = \eta_2(x,t), x_t \in \{x-1, x+1\}\},$$

($x_t$ is the location of firing at time $t$). Thus, $\mathrm{Returns}(x)$ is the number of times such that:

1. $x$ has type 0, (the same number, possibly 0 of oil and water particles), and
2. a pair of particles is emitted from $x-1$ or $x+1$.

Then define

$$\mathrm{Returns} = \sum_{x\in\mathbb{Z}} \mathrm{Returns}(x). \tag{18}$$

Recall the definition of $N_i(t)$ from equation (5) and notice that

$$\mathrm{Returns} = N_2(\tau) + 2N_3(\tau). \tag{19}$$

Our goal is to show that Returns is large. The advantage of the decomposition in (18) is that it will allow us to show the relationship between Returns and the odometer function.

Finally, recall that $\mathcal{G}_t$ is the $\sigma$-algebra generated by the movement of the first $t$ pairs that are emitted and notice that, from the definition of $P_t$ (11) we have that

$$P_t = \sum_{x\in\mathbb{Z}} \min[\eta_1(x,t), \eta_2(x,t)]$$

is the number of pairs at the time of the $t$th emission. We recall that $P_t$ is a supermartingale with respect to $\mathcal{G}_t$.

4.2. *Outline*. In order to show that it is unlikely that $\|u\|_\infty$ is greater than a big constant times $n^{4/3}$, we rely on three main ideas.

(i) *The odometer is fairly regular.* Typically,

$$|u(x) - u(x+1)| \le 2n + u(x)^{1/2}.$$

This regularity implies that under the assumption that $u(x_0)$ is much bigger than $n^{4/3}$ then it is likely that $u(x)$ is much bigger than $n^{4/3}$ for all $x$ such that $|x - x_0| < n^{1/3}$.

(ii) *The variable* Returns *can be expressed approximately in term of the odometer function.* Fix $x \in \mathbb{Z}$ and consider $\eta_1(x,k) - \eta_2(x,k)$ (difference of number of oil particles and water particles) as a function of $k$. This function performs a lazy random walk that takes about

$$\big(u(x-1)+u(x+1)\big) \approx 2u(x)$$

steps, (the approximation is because of the previously mentioned regularity). Thus, we expect Returns$(x)$ to be on the order of $u(x)^{1/2}$. Summing up over all $x$, we expect

$$\text{Returns} = \sum_{x\in\mathbb{Z}} \text{Returns}(x) \approx \sum_{x\in\mathbb{Z}} u(x)^{1/2}. \tag{20}$$

Combined with the previous paragraph this implies that if $\|u\|_\infty$ is much larger than $n^{4/3}$ then it is likely that Returns is much larger than $n$.

(iii) *The process $P_t$ is a supermartingale.* The sum of the negative drifts until the process terminates is, by equation (19),

$$-\frac{1}{4}\big(N_2(\tau)+2N_3(\tau)\big) = -\frac{1}{4}\text{Returns}.$$

Since $P_t$ starts at $n$ and stops when it hits 0, the typical value of Returns should be around $4n$.

We use the Azuma–Hoeffding inequality to show that the probability of the event "Returns is much larger than $4n$" is decaying very rapidly. By the previous paragraphs, we also will get that the probability that $\|u\|_\infty$ is much larger that $n^{4/3}$ is decaying very rapidly.

The main difficulty in implementing this outline comes in the second step as the odometer function and Returns are correlated in a complicated way.

4.3. *The bad events*. To make our outline formal, we now define our set of bad events. The first three of these deal with the odometer function. But first we use the odometer function to define

$$\text{Base} = \big\{x : u(x) > 0\big\} \subset \mathbb{Z}.$$

(0) $\mathcal{B}_0 = \{\text{Base} \not\subset [-n^5, n^5]\}$.

(i) $\mathcal{B}_1 = \{\|u\|_\infty \geq n^{4.05}\}$.

(ii) *Gradient of the Odometer.* It is natural to expect the odometer to decrease as one goes away from the origin (see Figure 1). The next event $\mathcal{B}_2$ is the event where the *gradient* of the odometer function, $u(x)-u(x+1)$, is too large or has the wrong sign.

Let

$$m(x) = \begin{cases} \min\Big\{n, \min\limits_{0\le y\le x} u(y)\Big\}, & \text{for } x > 0 \text{ and} \\ \min\Big\{n, \min\limits_{x\le y\le 0} u(y)\Big\}, & \text{for } x \le 0. \end{cases}$$

We define:

$$\begin{aligned}\mathcal{B}_2 := \{&\exists x \in [-n^5, n^5] : u(x) \geq n^{0.5} \text{ and} \\ &|u(x+1) - u(x)| > 2m(x) + \max(u(x), u(x+1))^{0.51}\} \\ &\cup \{\exists x \in [0, n^5] : u(x) \geq n^{0.5} \text{ and} \\ &(u(x+1) - u(x)) > \max(u(x), u(x+1))^{0.51}\} \\ &\cup \{\exists x \in [-n^5, 0] : u(x) \geq n^{0.5} \text{ and} \\ &(u(x-1) - u(x)) > \max(u(x), u(x+1))^{0.51}\}.\end{aligned}$$

In all the three events above, the $\max(u(x), u(x+1))^{0.51}$ corrections are an upper bound on the fluctuations. The definition of the events are then based on the fact that ignoring corrections due to fluctuation, for any $x > 0$, one has $0 < u(x) - u(x+1) \leq 2m$ as $u(x) - u(x+1)$ represents the number of particles eventually to the right of $x$. More details are provided in the proof of Lemma 4.5.

REMARK 2. While the definition of $\mathcal{B}_2$ is quite technical, we now give one consequence of it that is representative of how we will use it. Let $\mathcal{B}_0^c \cap \mathcal{B}_2^c$ occur and consider the set of $x$ such that $n < u(x) < n^{1.96}$ [or equivalently $1 < \log_n(u(x)) < 1.96$]. Then

$$|u(x) - u(x+1)| < 2m(x) + u(x)^{0.51} \leq 2n + (n^{1.96})^{0.51} \leq 3n. \tag{21}$$

For any $x$ and $y$ in a connected component of this set, then (21) implies

$$\text{if } |\log_n(u(x)) - \log_n(u(y))| \geq 0.05 \qquad \text{then } |x - y| > \frac{1}{6} n^{0.05}.$$

This is because our conditions on $x$ and $y$ imply $|u(x) - u(y)| \geq n^{1.05} - n > 0.5n^{1.05}$. So between $x$ and $y$ the odometer changes by at least $0.5n^{1.05}$ in increments of at most $3n$.

Our proof makes heavy use of this and similar estimates that follow from $\mathcal{B}_0^c \cap \mathcal{B}_2^c$. Much of the complication of our proof comes from the fact that we need to use different estimates depending on whether $m(x) = n$ or $m(x) < n$ and whether $\log_n(u(x))$ is greater than 1.96, between 1 and 1.96 or less than 1. These estimates are used in Lemmas 4.12 and 4.13 which are used to prove Lemma 4.14. The conclusion of Lemma 4.14 is useful in showing that Returns is large because of our next bad event.

(iii) *Regularity of returns*

$$\mathcal{B}_3 = \Bigg\{ \exists i, j, k : |i|, |j| \le n^5, k \in [n^{0.5}, n^5], j - i > 0.1 n^{0.01}, \text{ and } u(x) \ge k \ \forall x \in [i, j], \text{ and } \sum_{x=i+1}^{j-1} \text{Returns}(x) < 0.01 (j-i)\sqrt{k} \Bigg\}.$$

Finally, define the event

$$\mathcal{B}_4 = \{\text{Returns} \ge 20(n + \tau^{0.51})\}. \tag{22}$$

To complete our outline, we show that:

- $\mathcal{G}^c \subset \bigcup_{i=0}^4 \mathcal{B}_i$ and
- $\mathbb{P}(\mathcal{B}_i)$ are small for $i = 1, \dots, 4$.

4.4. *$P_t$ is a supermartingale.* Recall from (4) that

$$Z_k := P_{k+1} - P_k.$$

Recall from Remark 1 that conditional on the past, each $Z_k$ has one of four possible distributions $\xi_1, \dots, \xi_4$.

LEMMA 4.1. *$P_t$ is a supermartingale and*

$$\sum_{t=1}^{\infty} \mathbb{E}(Z_{t+1} | \mathcal{F}_t) = -\frac{1}{4} \text{Returns}.$$

PROOF. This follows from the discussion in Section 1.2. The summands on the left-hand side are all 0, $-\frac{1}{4}$ or $-\frac{1}{2}$. By (19) and (5), Returns is the sum of the number of $-\frac{1}{4}$ terms in the sum plus twice the number of $-\frac{1}{2}$ terms. □

LEMMA 4.2. *There exist $C, C'$ such that $\mathbb{P}(\mathcal{B}_4) < Ce^{-n^{C'}}$.*

PROOF. As Returns $\le \tau$, we have that

$$\mathcal{B}_4 = \bigcup_{r \ge 20n} \mathcal{B}_4 \cap \{\tau = r\}.$$

Note that by Lemma 4.1 and because $P_\tau = 0$ (i.e., $\sum_{i=1}^r Z_i = -n$), for every $r \ge 20n$,

$$\mathcal{B}_4 \cap \{\tau = r\} \subset \Bigg\{ \sum_{i=1}^{r} Z_t - \mathbb{E}(Z_t | \mathcal{F}_{t-1}) \ge 4n + 5r^{0.51} \Bigg\}.$$

Also note that $Z_t - \mathbb{E}(Z_t|\mathcal{F}_{t-1})$ are the increments of a martingale and are bounded by 1. Thus, we can apply the Azuma–Hoeffding inequality to get the bound $\mathbb{P}(\mathcal{B}_4 \cap \{\tau = r\}) \leq e^{-r^{0.02}}$. The lemma follows from the union bound. □

Now define the new event

$$\mathcal{E} = \mathcal{G}^c \cap \mathcal{B}_0^c \cap \mathcal{B}_1^c \cap \mathcal{B}_2^c, \tag{23}$$

where all the events are defined in Section 4.3. Remember the event $\mathcal{G}$ is that the odometer at the origin is less than a large constant times $n^{4/3}$ and the events $\mathcal{B}_0^c$, $\mathcal{B}_1^c$ and $\mathcal{B}_2^c$ are regularity conditions on the odometer. The key lemma needed for the proof of the upper bound is the following.

LEMMA 4.3.

$$\mathcal{E} \cap \mathcal{B}_3^c \subset \mathcal{B}_4, \tag{24}$$

*where* $\mathcal{B}_4$ *was defined in* (22). *Thus*,

$$\mathbb{P}(\mathcal{G}^c) \leq \mathbb{P}(\mathcal{B}_0) + \mathbb{P}(\mathcal{B}_1) + \mathbb{P}(\mathcal{B}_2) + \mathbb{P}(\mathcal{B}_3) + \mathbb{P}(\mathcal{B}_4). \tag{25}$$

Assuming (24), (25) follows by taking the union bound. We postpone the proof of (24) to Section 4.8 and proceed to completing the proof of the upper bound in Theorem 1.1.

LEMMA 4.4. *There exist positive constants* $D$, $C'$ *and* $\gamma$ *such that*

$$\mathbb{P}(\mathcal{B}_0), \mathbb{P}(\mathcal{B}_1), \mathbb{P}(\mathcal{B}_2), \mathbb{P}(\mathcal{B}_3), \mathbb{P}(\mathcal{B}_4) < De^{-C'n^{\gamma}}.$$

PROOF. $\mathbb{P}(\mathcal{B}_0)$ and $\mathbb{P}(\mathcal{B}_1)$ are bounded by Proposition 3.1. $\mathbb{P}(\mathcal{B}_4)$ is bounded by Lemma 4.2. The bounds for $\mathbb{P}(\mathcal{B}_2)$ and $\mathbb{P}(\mathcal{B}_3)$ appear as Lemmas 4.5 and 4.10, which appear in Sections 4.6 and 4.7, respectively. □

4.5. *Proof of the upper bound.* Thus, by (25) and Lemma 4.4, there exists $\epsilon > 0$, such that for large enough $n$,

$$\mathbb{P}(\mathcal{G}) > 1 - e^{-n^{\epsilon}}, \tag{26}$$

where $G$ was defined in (17). The proof of the upper bound is hence complete.

In the following sections, we bound $\mathbb{P}(\mathcal{B}_2)$ and $\mathbb{P}(\mathcal{B}_3)$.

4.6. *The probability of the bad events.*

LEMMA 4.5. *There exist positive constants* $C$, $C'$ *and* $\gamma$ *such that*

$$\mathbb{P}(\mathcal{B}_2) < Ce^{-C'n^{\gamma}}.$$

We first introduce a definition.

DEFINITION 6. Let $\Delta^x(k)$ denote the quantity such that, after $k$ pairs have been emitted from $x$, there are $k + \Delta^x(k)$ particles that moved to the right (i.e., to $x + 1$) and $k - \Delta^x(k)$ particles that have moved to the left (i.e., to $x - 1$). Notice that this is just a function of the stack of variables $X_i^x, Y_i^x$ at the site $x$.

PROOF OF LEMMA 4.5. Without loss of generality, we assume that $x \geq 0$. Recall from Section 4.3(ii) that

$$m(x) = \min\Bigl\{n, \min_{0 \leq y \leq x} u(y)\Bigr\}.$$

Suppose that at some time $t$ exactly $k$ pairs have been emitted from $x$, and exactly $k'$ pairs have been emitted from $x + 1$. Then the number of particles to the right of $x$ is

$$k + \Delta^x(k) - \bigl(k' - \Delta^{x+1}(k')\bigr).$$

This holds for all times $t$, in particular for the time when the process stops [at which time $k = u(x)$ and $k' = u(x + 1)$]. Let $\Delta^x = \Delta^x(u(x))$ and $\Delta^{x+1} = \Delta^{x+1}(u(x + 1))$. Since for any $0 \leq y \leq x$, the number of particles that settle to the right of $x$ is at most the number $2u(y)$ of particles emitted from $y$ (recall $u$ counts emitted *pairs*), we have

$$0 \leq u(x) + \Delta^x - \bigl(u(x + 1) - \Delta^{x+1}\bigr) \leq 2m(x).$$

Rearranging, we get

$$-\Delta^x - \Delta^{x+1} \leq u(x) - u(x + 1) \leq 2m(x) - \Delta^x - \Delta^{x+1}.$$

Then the event $\mathcal{B}_2$ implies that

$$\exists x \in \bigl[-n^5, n^5\bigr] \text{ and } k, \qquad n^{0.5} \leq k \leq n^{4.1} : \bigl|\Delta^x(k)\bigr| > 0.5k^{0.51}.$$

The result follows from standard concentration results of random walks (cf. Appendix A.1) and union bounding over all possible values of $x$ and $k$. Thus, the result holds for some appropriate $C, C'$ and $\gamma$. We omit the details. □

We bound the probability of $\mathcal{B}_3$ next. However, we need a few preliminary results first. Furthermore, in this context we work with a slightly modified but an equivalent setting.

4.7. *Merged stacks*. We describe another process $(\eta_1'(\cdot,\cdot), \eta_2'(\cdot,\cdot))$ which has the same law as the process $(\eta_1(\cdot,\cdot), \eta_2(\cdot,\cdot))$ defined in Definition 2.

In this version, analogous to (6) the source of randomness is a set of independent variables (modified stacks):

$$\omega' = (X_i^x, Y_i^x, \bar{X}_i^x, \bar{Y}_i^x)_{x\in 3\mathbb{Z}, i\in\mathbb{N}} \tag{27}$$

with $\mathbb{P}(X_i^x = \pm 1) = P(Y_i^x = \pm 1) = \mathbb{P}(\bar{X}_i^x = \pm 1) = \mathbb{P}(\bar{Y}_i^x = \pm 1) = \frac{1}{2}$. Denote by $\Omega'$ the set of all stacks $\omega'$. Note that stacks are located only at every $x \in 3\mathbb{Z}$.

Informally, the $i$th firing from $x \in 3\mathbb{Z}$ uses moves $X_i^x$ and $Y_i^x$ as before, but the $i$th firing from the *set* $\{x-1, x+1\}$ uses moves $\pm\bar{X}_i^x$ and $\pm\bar{Y}_i^x$ according to whether the firing was from $x-1$ or $x+1$, respectively. We refer to $(\bar{X}_i^x, \bar{Y}_i^x)_{i\geq 1}$ as the *merged stacks* of $x-1$ and $x+1$.

Formally, given a firing sequence $s = (x_0, \dots, x_{K-1})$, we define the pair $(\eta_1'(\cdot,k), \eta_2'(\cdot,k))$ inductively as follows. If $x_k \in 3\mathbb{Z}$, then

$$\eta_1'(\cdot, k+1) = \eta_1'(\cdot,k) - \delta(x_k) + \delta(x_k + X_{i_k}^{x_k}),$$

$$\eta_2'(\cdot, k+1) = \eta_2'(\cdot,k) - \delta(x_k) + \delta(x_k + Y_{i_k}^{x_k}),$$

where $i_k = \#\{j \leq k : x_j = x_k\}$. If $x_k \in 3\mathbb{Z} \pm 1$, then

$$\eta_1'(\cdot, k+1) = \eta_1'(\cdot,k) - \delta(x_k) + \delta(x_k \mp \bar{X}_{i_k'}^{x_k \mp 1}),$$

$$\eta_2'(\cdot, k+1) = \eta_2'(\cdot,k) - \delta(x_k) + \delta(x_k \mp \bar{Y}_{i_k'}^{x_k \mp 1}),$$

where $i_k' = \#\{j \leq k : x_j \in \{x_k, x_k \mp 2\}\}$.

To compare the modified process to the original, we use the following proposition.

PROPOSITION 4.6. *Let $(Z_i)_{i\in I}$ be independent uniform $\pm 1$-valued random variables indexed by a countable set $I$. Let $i_1, i_2, \dots \in I$ be a sequence of distinct random indices and $\xi_1, \xi_2, \dots$ a sequence of $\pm 1$-valued random variables such that for all $k \geq 1$ both $i_k$ and $\xi_k$ are measurable with respect to $\mathcal{F}_{k-1} := \sigma(Z_{i_\ell})_{1\leq \ell \leq k-1}$. Then $(\xi_k Z_{i_k})_{k\geq 1}$ is an i.i.d. sequence.*

PROOF. We proceed by induction on $\ell$ to show that $(\xi_k Z_{i_k})_{1\leq k\leq \ell}$ is i.i.d. Since $\xi_\ell$ and $i_\ell$ are $\mathcal{F}_{\ell-1}$-measurable, and $i_\ell$ is distinct from $i_1, \dots, i_{\ell-1}$, we have

$$\mathbb{E}(\xi_\ell Z_{i_\ell} | \mathcal{F}_{\ell-1}) = \xi_\ell \mathbb{E}(Z_{i_\ell} | \mathcal{F}_{\ell-1}) = 0.$$

Since $\xi_\ell Z_{i_\ell}$ is $\pm 1$-valued the proof is complete. $\square$

Recall $\Omega$ the set of all stacks $\omega$ defined by the original process, and $\Omega'$, the set of all stacks $\omega'$ defined by the modified process. Let $\tau(\omega)$ denote the stopping time of the sequence $\omega$, and similarly $\tau(\omega')$. Furthermore, denote by $(o_\tau, w_\tau)(\omega)$ the *final configuration* after performing $\omega$, and by $(o_\tau', w_\tau')(\omega')$ the final configuration of the process after performing $\omega'$.

LEMMA 4.7. *There is a measure-preserving map $\phi : \Omega' \to \Omega$ such that with the firing rule described in the proof of Lemma 3.5,*

$$(\eta_1'(\cdot,\cdot), \eta_2'(\cdot,\cdot)) \stackrel{d}{=} (\eta_1(\cdot,\cdot), \eta_2(\cdot,\cdot)).$$

*In particular,* $\tau(\phi(\omega')) = \tau(\omega')$, $(o_\tau', w_\tau')(\omega') = (o_\tau, w_\tau)(\phi(\omega'))$ *and the odometer counts in both the processes are same.*

PROOF. Given $\omega' \in \Omega'$, let $x_1, x_2, \dots$ be the resulting sequence following the firing rule described in the proof of Lemma 3.5, and let $u(x)$ be the number of firings performed at $x$ in the modified process. We set $\phi(\omega') = (X_i^x, Y_i^x)_{x\in\mathbb{Z}, i\in\mathbb{N}}$ where if $i \le u(x)$ then $X_i^x$ (resp., $Y_i^x$) is the direction in which the $i$th oil (resp., $i$th water) exited $x$ in the modified process.

If $i > u(x)$, then we make an arbitrary choice (e.g., split the unused portion of each merged stack into even and odd indices).

It is important to note, that each $x_k$ is measurable with respect to the $\sigma$-algebra $\mathcal{F}'_{k-1}$ generated by the stack variables in $\omega'$ used before time $k$, and that each stack variable in $\omega'$ is used at most once. The stack variables used *at* time $k$ have the form $\xi X, \xi Y$ where $X, Y$ are stack variables not yet used and $\xi = \mathbf{1}_{x_k \in 3\mathbb{Z} \cup (3\mathbb{Z}+1)} - \mathbf{1}_{x_k \in 3\mathbb{Z}-1}$ is a random sign that is measurable with respect to $\mathcal{F}'_{k-1}$. Conditional on $\mathcal{F}'_{k-1}$, $\xi X$ and $\xi Y$ are independent uniform $\pm 1$ random variables, and $\phi$ is measure-preserving by Proposition 4.6. $\square$

Lemma 4.7 allows us to switch between events defined in one version to the other. For $x \in 3\mathbb{Z}$, let

$$W^x(\ell) = \frac{\bar{X}_\ell^x - \bar{Y}_\ell^x}{2},$$

where $\bar{X}_\ell^x, \bar{Y}_\ell^x$ are defined in (27). This represents the change in the difference of oil and water particles at $x$ when the $\ell$th firing takes place from the set $\{x-1, x+1\}$. Clearly, $W^x(\ell)$ has the same distribution as one step of a symmetric lazy random walk. Now define

$$\tilde{R}^x(k) = \#\left\{0 \le j \le k : \sum_{\ell=1}^{j} W^x(\ell) = 0\right\}.$$

Define $\mathcal{B}_3'$ to be the event that there exist three integers $i, j, k$ such that:

(i) $|i|, |j| \le n^5$,
(ii) $j - i \ge 0.1 n^{0.01}$,
(iii) $k \in [n^{0.5}, n^5]$ and
(iv) $\sum_{x \in (i,j), 3|x} \tilde{R}^x(0.9k) < 0.01\sqrt{k}(j-i)$.

We now state a standard fact about number of returns to origin for the simple random walk on $\mathbb{Z}$.

LEMMA 4.8. *Let $\{S_i\}_{i\geq 0}$ be a lazy simple random walk on $\mathbb{Z}$ started at the origin. Let*

$$\mathrm{Zeros}(l) = \#\{i : 0 \leq i \leq l \text{ and } S_i = 0\}.$$

*Then for all $l$,*

$$\mathbb{P}\big(\mathrm{Zeros}(0.9l) > 0.1\sqrt{l}\big) \geq \frac{1}{2}.$$

PROOF. See Chapter III, Section 5 of Feller (1968). □

LEMMA 4.9. *There exist positive constants $C$, $C'$ and $\gamma$ such that*

$$\mathbb{P}\big(\mathcal{B}_3'\big) < Ce^{-C'n^{\gamma}}.$$

PROOF. Fix $i, j, k$ and $x \in (i, j)$. Then, from Lemma 4.8 it follows

$$\mathbb{P}\big(\tilde{R}^x(0.9k) > 0.1\sqrt{k}\big) \geq \frac{1}{2}.$$

Now using the independence of the stacks at the multiples of three, we get

$$\begin{aligned}
&\mathbb{P}\Bigg(\sum_{x\in(i,j),3|x} \tilde{R}^x(0.9k) < 0.01\sqrt{k}(j-i)\Bigg)\\
&\quad < \mathbb{P}\bigg(\#\big\{x \in (i, j), 3|x : \tilde{R}^x(0.9k) > 0.1\sqrt{k}\big\} < \frac{1}{10}(j-i)\bigg)\\
&\quad < ce^{-c'n^{0.1}},
\end{aligned}$$

where $c$, $c'$ are positive constants.

As there are at most $4n^{15}$ choices of $i$, $j$ and $k$ there exists positive constants $c$, $c'$ and $\gamma'$ such that

$$\mathbb{P}\big(\mathcal{B}_3'\big) < 4n^{15}ce^{-c'n^{\gamma'}}$$

so the lemma is true for some choice of $C$, $C'$ and $\gamma$. □

The bound on $\mathbb{P}(\mathcal{B}_3)$ is now a corollary.

LEMMA 4.10. *There exist positive constants $C$, $C'$ and $\gamma$ such that*

$$\mathbb{P}(\mathcal{B}_3) < Ce^{-C'n^{\gamma}}.$$

PROOF. Consider the map $\phi$ defined in Lemma 4.7 and the event $\phi^{-1}(\mathcal{B}_3)$. By the measure preserving property of $\phi$, we have $\mathbb{P}(\mathcal{B}_3) = \mathbb{P}(\phi^{-1}(\mathcal{B}_3))$ where the two probabilities are in the two different probability spaces mentioned in the statement of Lemma 4.7. Note that the event $\phi^{-1}(\mathcal{B}_3)$ implies $\mathcal{B}_3'$. This is because (iv) in the definition of $\mathcal{B}_3'$ says

$$\sum_{x\in(i,j),3|x} \tilde{R}^x(0.9k) < 0.01\sqrt{k}(j-i)$$

whereas in the definition of $\mathcal{B}_3$ we have

$$\sum_{x\in(i,j)} \tilde{R}^x(0.9k) < 0.01\sqrt{k}(j-i).$$

The proof now follows from Lemma 4.9. □

The rest of this section is devoted to the key technical proof of Lemma 4.3, which is all that remains to finish off the proof of the upper bound in Theorem 1.1.

4.8. *Proof of Lemma* 4.3. We split the proof of Lemma 4.3 into several lemmas. In particular, it will suffice to show that

$$\frac{1}{2}\text{Returns} > 20n, \tag{28}$$

$$\frac{1}{2}\text{Returns} > 20\tau^{0.51}. \tag{29}$$

Recall that on the event $\mathcal{E}$ [see (23)], the height is much bigger than $n^{4/3}$ but less than $n^{4.05}$, the odometer is supported inside $[-n^5, n^5]$ and gradient of the odometer is such that $\mathcal{B}_2^c$ occurs.

LEMMA 4.11. *If the event $\mathcal{E}$ occurs, then*

$$\tau \geq n^{5/3} \tag{30}$$

*and if $\mathcal{E}\cap\mathcal{B}_3^c$ occurs, then*

$$\frac{1}{2}\text{Returns} > 20n.$$

PROOF. As usual, we choose to give a direct proof with explicit constants (which might be far from optimal), for sake of exposition.

If $\mathcal{B}_0^c$ occurs, then there exists

$$x_* \in [-n^5, n^5], \qquad \text{such that } u(x_*) = \|u\|_\infty.$$

We now consider two cases.

*Case* 1: If $\|u\|_\infty^{0.51} \leq 2n$, then the events $\mathcal{G}^c$ and $\mathcal{B}_2^c$ imply that

$$u(x) > 700{,}000n^{4/3} > (800n^{2/3})^2, \tag{31}$$

for all $x$ in the interval,

$$X_0 := [x_* - 100n^{1/3}, x_* + 100n^{1/3}] \cap [-n^5, n^5].$$

This follows from the event $\mathcal{B}_0^c \cap \mathcal{B}_2^c$ by the same argument as in Remark 2. Then $|X_0| \geq 100n^{1/3}$, because by $\mathcal{B}_0^c$, we have $x_* \in [-n^5, n^5]$ so at least one of the intervals

$$[x_* - 100n^{1/3}, x_*] \quad \text{or} \quad [x_*, x_* + 100n^{1/3}]$$

is entirely contained in $[-n^5, n^5]$. Thus, looking at the volume of the odometer in $X_0$, we have

$$\tau \geq \sum_{x \in X_0} u(x) \geq 700{,}000 n^{4/3} |X_0| \geq n^{5/3} \tag{32}$$

and

$$\text{Returns} \geq \sum_{x \in X_0} \text{Returns}(x) \geq (100n^{1/3})(0.01)(800n^{2/3}) \geq 40n.$$

The next to last inequality is by (31), the size of $X_0$ and the definition of $\mathcal{B}_3^c$. Thus, we have obtained inequality (28).

*Case* 2: If $\|u\|_\infty^{0.51} > 2n$, then one expects Returns to be even larger. Note that by $\mathcal{B}_2^c$ every gradient is at most $2\|u\|_\infty^{0.51}$. Thus, the odometer is at least $\|u\|_\infty/2$ over an interval of length $\|u\|_\infty^{0.49}/2 \geq n^{0.1}$. We take $X_0$ to be this interval. Thus, by $\mathcal{B}_3^c$,

$$\text{Returns} \geq \sum_{x \in X_0} \text{Returns}(x) \geq (\|u\|_\infty^{0.49}/2)(0.01)(\|u\|_\infty^{0.5}) \geq 40n. \qquad \square$$

4.9. *Partitioning* Base. Inequality (29) is more involved to verify. However, it relies on the observation that at most sites, $\text{Returns}(x) \approx \sqrt{u(x)}$, [since the number of returns to the origin for a random walk in $u(x)$ steps is roughly about $\sqrt{u(x)}$]. Now if the value of $u(x)$ does not change sharply (this is where the gradient bounds in the definition of $\mathcal{B}_2$ are used) this then implies that

$$\text{Returns} = \sum_x \text{Returns}(x) \gg \sqrt{\sum_x u(x)} = \sqrt{\tau}.$$

Thus, the proof proceeds by showing that there is a subset of $Z$ which can be written as the union of not too small intervals, where the odometer is uniformly high and not too rough and contributes a constant fraction of $\sum_x u(x)$. On this set, the above inequality can be made precise.

Formally, we partition Base up into smaller intervals:

$$\text{Base} = \bigcup_{i=-K'}^{K} \text{Base}_i,$$

where, for each $i$, we set each $\mathrm{Base}_i = [a_i, b_i]$ for some $a_i$ and $b_i$ which we describe in the following. We inductively define the $\mathrm{Base}_i$ in a way so that on the event $\mathcal{E}$ and for most $i$ we have

$$b_i - a_i > 0.1n^{0.01}. \tag{33}$$

This will involve a series of estimates like in Remark 2. Define $h_j = n^{4.05-0.05j}$ for $j = 0$ to $j = 68$ and $h_{69} = 0$. Let

$$\mathrm{H} = \{h_j : j = 1, \dots, 68\}.$$

Let $j'$ be such that $h_{j'}$ is the closest value in H to $u(0)$. Let

$$b_0 = \inf\{x : x \geq 0 \text{ and } u(x+1) \notin (h_{j'+1}, h_{j'-1})\}.$$

We say that $\mathrm{Base}_0$ *starts* at height $h_{j'}$ and *ends* at height $h_{j''} \in \mathrm{H}$ where $j''$ is defined so that $h_{j''}$ is the closest element of H to $u(b_0 + 1)$.

Now we inductively define $\mathrm{Base}_{i+1}$. Suppose we have defined $\mathrm{Base}_i = [a_i, b_i]$ which ends at height $h_k$. We will inductively define $\mathrm{Base}_{i+1} = [a_{i+1}, b_{i+1}]$. We let $a_{i+1} = 1 + b_i$ and say $\mathrm{Base}_{i+1}$ *starts* at height $h_k \in \mathrm{H}$. Then we define

$$b_{i+1} = \inf\{x : x \geq a_i \text{ and } u(x+1) \notin (h_{k+1}, h_{k-1})\}.$$

We say the block $\mathrm{Base}_{i+1}$ *ends* at height $h_{k'} \in \mathrm{H}$ where $h_{k'}$ is the closest element of H to $u(b_i + 1)$ and $h_{k'}$ is the closest element of H to $u(b_i + 1)$.

For the case of negative indices, the procedure is totally analogous. Finally, for all $j \in \{0, \dots, 68\}$ define

$$B_j := \text{the union of all } \mathrm{Base}_i \text{ that start at } h_j \text{ and have } |\mathrm{Base}_i| > 0.1n^{0.01}. \tag{34}$$

LEMMA 4.12. *If a block* $\mathrm{Base}_i$ *with* $i \geq 0$ *starts at height* $h_j$ *and ends at height* $h_{j'}$ *then* $j' \neq j$. *On the event* $\mathcal{E}$ *and* $j' < j < 68$, *then*

$$b_i - a_i > 0.1n^{0.01}.$$

PROOF. The first statement is true because

$$u(1 + b_i) \notin (h_{k+1}, h_{k-1})$$

so the closest element of H to $u(1 + b_i)$ is not $h_k$. Consider an interval $\mathrm{Base}_i$ with $i \geq 0$ where $\mathrm{Base}_i$ starts at height $h_j$ and ends at height $h_{j'}$ with $j' < j$. Over the course of such an interval the odometer increased by at least $0.49h_{j-1}$, going from less than $0.5(h_j + h_{j-1})$ to at least $h_{j-1}$. Since $j < 68$, we have

$$u(x) > h_{j+1} \geq h_{68} = n^{0.65} > n^{0.5}$$

for all $x \in \mathrm{Base}_i$. Thus, the event $\mathcal{E}$ implies

$$u(x+1) - u(x) < u(x)^{0.51} < (h_{j-1})^{0.51}.$$

Thus, by the choice of the $h_i$ these intervals must have width at least

$$b_i - a_i \geq \frac{0.49h_{j-1}}{(h_{j-1})^{0.51}} \geq 0.49(h_{j-1})^{0.49} \geq 0.49n^{0.3} > 0.1n^{0.01}.$$

The next to last inequality follows because $j < 68$ so $h_{j-1} \geq n^{0.7}$. $\square$

LEMMA 4.13. *If $\mathcal{E}$ occurs and a block* $\mathrm{Base}_i$ *with $i > 0$ starts at height $h_j$ and ends at height $h_{j'}$ with $j' > j$ then*:

1. $b_i - a_i > 0.1n^{0.01}$ *or*
2. $68 \geq j \geq 60$ *and for no $k$, $0 \leq k < i$ the block* $\mathrm{Base}_k$ *starts at $h_j$ and ends at a height $h_{k'}$ with $k' > j$.*

PROOF. Consider an interval $\mathrm{Base}_i$ with $i \geq 0$ where $\mathrm{Base}_i$ starts at height $h_j$ and ends at height $h_{j'}$ with $j' > j$. Over the course of such an interval, the odometer decreased by at least $0.49h_j$, going from at least $0.5(h_j + h_{j+1})$ to at most $h_{j+1}$.

First, we consider the case that $j < 60$. We have

$$u(x) > h_{j+1} \geq h_{60} = n^{1.05}$$

for all $x \in \mathrm{Base}_i$. Thus, the event $\mathcal{E}$ implies

$$u(x) - u(x+1) < 2m(x) + u(x)^{0.51} = 2n + u(x)^{0.51} < 3(h_j)^{0.96}.$$

Thus, by the choice of the $h_i$ these intervals must have width at least

$$b_i - a_i \geq \frac{0.49h_j}{3(h_j)^{0.96}} \geq 0.15(h_j)^{0.04} > 0.1n^{0.01}.$$

Next, we consider the case that $\mathrm{Base}_i$ ends at $h_j$ with $60 \leq j \leq 68$. Suppose there exists $k < i$ such that $\mathrm{Base}_k$ starts at $h_j$ and ends at a height $h_{k'}$ with $k' > j$. This implies that $j < 68$ and

$$m(x) \leq h_{j+1} \leq (h_j)n^{-0.05}$$

for all $x \in \mathrm{Base}_i$. Thus, the event $\mathcal{E}$ and $j \leq 67$ implies $u(x) > n^{0.65}$ and

$$u(x) - u(x+1) < 2m(x) + u(x)^{0.51} < 3h_j n^{-0.05}.$$

Thus, these intervals must have width at least

$$b_i - a_i \geq \frac{0.49h_j}{3(h_j)n^{-0.05}} \geq 0.15n^{0.05} > 0.1n^{0.01}.$$

$\square$

REMARK 3. We need to only worry about the case $j' > j$ since otherwise the odometer does not decrease and only helps us in our arguments. The reason for separating the study of the case $j' > j$ into $j < 60$ and $68 \geq j \geq 60$ is that a priori we have not ruled out that most of the contribution to $\tau = \sum_x u(x)$ comes from small values of $u(x)$. In the gradient bound of $m(x) = \min\{n, \min_{0\le y\le x} u(y)\}$ for $x > 0$ (see definition of $\mathcal{B}_2$), we use $n$ as an upper bound for $m(x)$ when $u(x)$ is large ($j < 60$) and we use $\min_{0\le y\le x} u(y)$ as an upper bound for $m(x)$ when $u(x)$ is small ($68 \ge j \ge 60$).

4.10. *Consequences of a regular gradient*. Recall the definition of $B_i$ from (34).

LEMMA 4.14. *On the event $\mathcal{E}$, there exists $i \in \{1,\dots,68\}$ such that*

$$\sum_{x\in B_i} u(x) \ge 0.01\tau.$$

PROOF. Define

$$B^* = \bigcup_{i:|\mathrm{Base}_i|\le 0.1n^{0.01}} \mathrm{Base}_i.$$

We first show that

$$\sum_{x\in B^*} u(x) \le 20n^{1.2}. \tag{35}$$

The lemma will follow easily from (30), (35) and the pigeonhole principle.

Let $I = \{i : |\mathrm{Base}_i| \le 0.1n^{0.01}\}$. First, we show that $|I| \le 20$. By Lemmas 4.12 and 4.13 for every $i \in I$, there exists $j$, $60 \le j \le 68$ such that $\mathrm{Base}_i$ starts at height $h_j$. From this, we draw two conclusions. First, by the definitions of $B^*$ and the $\mathrm{Base}_i$ we have $u(x) < n^{1.15}$ on $B^*$. Also for each such $j$, Lemma 4.13 implies there exist at most two $i \in I$ with $\mathrm{Base}_i$ starting at height $h_j$, at most one with $i \ge 0$ and at most one with $i \le 0$. These two facts combine to establish (35) which completes the proof. □

REMARK 4. If we perform the analysis in the previous lemma to nonempty intervals $\mathrm{Base}_i$ that start at $h_j \ge n^{1.4}$, we get that $|\mathrm{Base}_i| \ge 0.1n^{0.4}$. This implies

$$\tau \ge 0.1n^{0.4}h_{j+1} = 0.1n^{0.3}h_{j-1}.$$

LEMMA 4.15. *If $B_j \ne \varnothing$ and $\mathcal{E}$ occurs, then*

$$\frac{\tau^{0.49}}{(h_{j-1})^{0.5}} \ge n^{0.1}.$$

PROOF. If $h_j \le n^{1.35}$, then the result follows from the first part of Lemma 4.11. If $h_j \ge n^{1.4}$, then it follows from the previous remark. □

4.11. *Consequences of regular returns.* Recall the sequence $h_i$ defined in Section 4.9.

LEMMA 4.16. *For every $j \in \{1, \dots, 68\}$, the set $B_j$ satisfies*

$$\sum_{x \in B_j} u(x) \le |B_j| h_{j-1}.$$

*Conditional on $\mathcal{E} \cap \mathcal{B}_3^c$ for every $j \in \{0, \dots, 67\}$ the set $B_j$ satisfies*

$$\sum_{x \in B_j} \mathrm{Returns}(x) \ge 0.01 |B_j| (h_{j+1})^{1/2}.$$

PROOF. From the choice of the intervals $\mathrm{Base}_j$ and $B_i$, we have that

$$h_{i+1} \le u(x) \le h_{i-1},$$

for all $x \in B_i$. Also, by definition, $B_i$ consists of a union of intervals of width at least $0.1 n^{0.01}$ [cf. equation (34)]. Therefore, the second statement follows from the definition of $\mathcal{B}_3^c$. □

LEMMA 4.17. *For all sufficiently large $n$ conditional on $\mathcal{E} \cap \mathcal{B}_3^c$, we have*

$$\frac{1}{2} \mathrm{Returns} > 20 \tau^{0.51}.$$

PROOF. As $\mathcal{E}$ occurs, by Lemma 4.14 we obtain that there exists a $j$ such that

$$\sum_{x \in B_j} u(x) \ge 0.01 \tau.$$

First, consider the case that $j \in \{1, \dots, 67\}$. In this case, from Lemma 4.16 it follows that

$$|B_j| h_{j-1} \ge \sum_{x \in B_j} u(x) > 0.01 \tau. \tag{36}$$

Conditional on the event $\mathcal{B}_3^c$, the set $B_j$ satisfies $\sum_{x \in B_j} \mathrm{Returns}(x) \ge 0.01 |B_j| (h_{j+1})^{1/2}$, which implies

$$\begin{aligned}
\sum_{x \in B_j} \mathrm{Returns}(x) &\ge 0.01 |B_j| (h_{j+1})^{1/2} \\
&\overset{(36)}{\ge} 0.01 \left( \frac{0.01 \tau}{h_{j-1}} \right) (h_{j+1})^{1/2} \\
&= 0.0001 \left( \frac{h_{j+1}}{h_{j-1}} \right)^{1/2} \frac{\tau^{0.49}}{(h_{j-1})^{0.5}} \tau^{0.51} \\
&\ge 0.0001 n^{-0.05} n^{0.1} \tau^{0.51} \\
&\ge 40 \tau^{0.51}.
\end{aligned}$$

The next to last line follows from Lemma 4.15, together with the definition of $h_{j+1}$ and $h_{j-1}$, whereas the last inequality holds whenever $n$ is sufficiently large.

Now consider $j = 68$. In this case, by Lemma 4.11 we have $\tau > n^{5/3}$ so $\tau^{0.49} > n^{0.8}$ and

$$\sum_{x \in B_{68}} \text{Returns}(x) \geq |B_{68}| \geq \frac{0.01\tau}{n^{0.7}} = \frac{0.01\tau^{0.51}\tau^{0.49}}{n^{0.7}} \geq 40\tau^{0.51},$$

which completes the proof. $\square$

PROOF OF LEMMA 4.3. The proof is an easy consequence of Lemmas 4.11 and 4.17. $\square$

**5. Proofs of Theorem 1.1 (lower bound) and Theorem 1.2.** The goal of this section is to prove part ii of Theorem 1.1. Theorem 1.2 will follow. The proof begins by using the upper bound (part i) of Theorem 1.1 to deduce a lower bound on the gradient of the odometer: Namely, there exists a constant $\epsilon$ such that, with high probability, for $x \in [0, \epsilon n^{1/3}]$ we have

$$u(x) - u(x+1) \geq n/4.$$

This in turn implies the lower bound (part ii) of Theorem 1.1: since $u(\epsilon n^{1/3}) \geq 0$, we have $u(x) \geq \frac{1}{8}\epsilon n^{4/3}$ for all $x \in [0, \frac{1}{2}\epsilon n^{4/3}]$.

5.1. *Maximum of lazy random walk.* We collect here a few standard results about the running maximum of $R_i$, the lazy simple symmetric random walk on $\mathbb{Z}$ whose increments $R_{i+1} - R_i$ are $\pm 1$ with probability $\frac{1}{4}$ each and 0 with probability $\frac{1}{2}$. Let

$$M(t) = \sup_{i<t} |R_i|. \tag{37}$$

LEMMA 5.1.

(i) *Given* $\epsilon > 0$, *for all large* $t > 0$,

$$\mathbb{P}\big(M(t) > t^{1/2+\epsilon}\big) < e^{-t^{\epsilon/2}}.$$

(ii) *Given* $\epsilon > 0$, *for all large* $t > 0$,

$$\mathbb{P}\big(M\big(t^{2+\epsilon}\big) < 2t\big) < e^{-t^{\epsilon/2}}. \tag{38}$$

(iii) $\mathbb{E}(M(t)) = \Theta(\sqrt{t})$.

(iv) $\mathbb{E}(M(t)^2) = \Theta(t)$.

(v) $\lim_{t\to\infty} t^{-1/2}\mathbb{E}|R_t| = \sqrt{\frac{1}{\pi}}$.

PROOF. The proofs of parts (i)–(iv) are standard. We refer the reader to Sections 21 and 23 in Spitzer (1976). By the central limit theorem, $t^{-1/2}R_t \xrightarrow{d} Z$ where $Z \sim N(0, \frac{1}{2})$, hence $\mathbb{E}|Z| = \sqrt{1/\pi}$. Part (v) follows since the random variables $t^{-1/2}R_t$ are uniformly integrable; see, for instance, Theorem 3.5 of Billingsley (1999). □

LEMMA 5.2. *Let*

$$\{M^i(n^{\frac{4}{3}})\}_{i=1}^{n^{1/3}}$$

*be a sequence of i.i.d. random variables with the same law as* $M(n^{\frac{4}{3}})$. *Then there exist positive constants* $D, \gamma$ *such that*

$$\mathbb{P}\left(\sum_{i=1}^{n^{1/3}} M^i(n^{\frac{4}{3}}) < Dn\right) > 1 - e^{-n^{\gamma}}.$$

Proof of Lemma 5.2 is deferred to the Appendix.

5.2. *Lower bound on the gradient of the odometer.* For $x \in \mathbb{Z}$ and a positive integer $i$, we define the variable

$$D^x(i) = \mathbf{1}_{(X_i^x=1)} - \mathbf{1}_{(Y_i^x=1)}, \tag{39}$$

where the variables $X_i^x$, $Y_i^x$ appear in (6) in Section 2. Clearly,

$$D^x(i) = \begin{cases} -1, & \text{w.p. } 1/4, \\ 0, & \text{w.p. } 1/2, \\ 1, & \text{w.p. } 1/4. \end{cases}$$

Let $C$ be the constant appearing in the upper bound (part i) of Theorem 1.1. For $x \in \mathbb{Z}$, let

$$S_x = \sup_{i \le Cn^{4/3}} \left|\sum_{j=0}^{i} D^x(j)\right|, \tag{40}$$

$$T_x = \sup_{i \le Cn^{4/3}} \left|\sum_{j=0}^{i} \frac{X_j^x}{2}\right| + \sup_{i \le Cn^{4/3}} \left|\sum_{j=0}^{i} \frac{Y_j^x}{2}\right|. \tag{41}$$

Recall from (14) that $g_\tau(x)$ is the signed count of particles at $x$ at the end of the oil and water process. Since one side of the origin has at least $n/2$ particles at the end of the process, without loss of generality we assume that

$$\sum_{x=0}^{\infty} |g_\tau(x)| \ge n/2. \tag{42}$$

Next, we claim that for any $x \in \mathbb{Z}$,

$$g_\tau(x) = \sum_{i=1}^{u(x-1)} D^{x-1}(i) - \sum_{i=1}^{u(x+1)} D^{x+1}(i), \tag{43}$$

where $D^x(\cdot)$ is defined in (39). To see this, note that the right-hand side is the difference between the number of oil particles sent to $x$ and the number of water particles sent to $x$, whereas the emission of an oil-water pair from $x$ does not change the signed count of particles at $x$.

Recall $\mathcal{G}$ defined in (17). Also given $\epsilon > 0$ define the following events:

$$\mathcal{G}_1 = \left\{ \sum_{x=-\epsilon n^{1/3}}^{\epsilon n^{1/3}} S_x \le \frac{n}{12} \right\}, \tag{44}$$

$$\mathcal{G}_2 = \left\{ \sum_{x=-\epsilon n^{1/3}}^{\epsilon n^{1/3}} T_x \le \frac{n}{12} \right\}, \tag{45}$$

where $S_x$ and $T_x$ are defined in (40) and (41). We suppress the dependence on $\epsilon, n$ in the notation for brevity.

LEMMA 5.3. *For small enough* $\epsilon$, *there exists* $c > 0$ *such that*

$$\mathbb{P}(\mathcal{G} \cap \mathcal{G}_1 \cap \mathcal{G}_2) \ge 1 - e^{-n^c}. \tag{46}$$

PROOF. The proof follows from (26) and Lemma 5.2. $\square$

We now state the following lemma establishing a lower bound on the gradient of the odometer function.

LEMMA 5.4. *Assume* (42). *There exists a constant* $\epsilon > 0$ *such that*

$$u(j) - u(j+1) \ge \frac{n}{4}$$

*for all* $0 \le j \le \epsilon n^{1/3}$, *with failure probability at most* $e^{-n^c}$ *for some positive constant* $c$.

PROOF. Recalling Definition 6 notice that

$$u(j) + \Delta^j\big(u(j)\big) - u(j+1) + \Delta^{j+1}\big(u(j+1)\big) = \sum_{y=j+1}^{\infty} \big|g_\tau(y)\big|.$$

We claim that for any $x \in \mathbb{Z}$,

$$\big|\Delta^x\big(u(x)\big)\big| \mathbf{1}(\mathcal{G}) \le T_x,$$

where $T_x$ is defined in (41). This follows from definitions and the observation that

$$\begin{aligned}\Delta^x\big(u(x)\big) &= \sum_{i=1}^{u(x)}\big[\mathbf{1}\big(X_i^x = 1\big) - 1/2\big] + \sum_{i=1}^{u(x)}\big[\mathbf{1}\big(Y_i^x = 1\big) - 1/2\big] \\ &= \sum_{i=1}^{u(x)} \frac{X_i^x}{2} + \sum_{i=1}^{u(x)} \frac{Y_i^x}{2}.\end{aligned}$$

Thus, we have

$$[u(j) - u(j+1)]\mathbf{1}(\mathcal{G}) \geq \sum_{y=j+1}^{\infty} |g_\tau(y)|\mathbf{1}(\mathcal{G}) - (T_j + T_{j+1})\mathbf{1}(\mathcal{G}). \tag{47}$$

Now on the event $\mathcal{G} \cap \mathcal{G}_1 \cap \mathcal{G}_2$, $\forall j < \epsilon n^{1/3}$,

$$\sum_{i=j}^{\infty} \big|g_\tau(i)\big| \geq n/3. \tag{48}$$

To show this, we upper bound $\sum_{i=1}^{j} |g_\tau(i)|$. By (43), on the event $\mathcal{G}$ we have for all $0 < i \leq \epsilon n^{1/3}$,

$$\big|g_\tau(i)\big| \leq S_{i-1} + S_{i+1}, \tag{49}$$

where $S_i$ is defined in (40). Hence, on the event $\mathcal{G} \cap \mathcal{G}_1 \cap \mathcal{G}_2$,

$$\sum_{i=0}^{j} \big|g_\tau(i)\big| \leq 2 \sum_{i=-1}^{\epsilon n^{1/3}} S_i \leq \frac{n}{6}.$$

Thus, by (42) we have for all $0 < j < \epsilon n^{1/3}$,

$$\sum_{i=j}^{\infty} \big|g_\tau(i)\big| \geq \frac{n}{2} - \sum_{i=0}^{j} \big|g_\tau(i)\big| \geq \frac{n}{3}.$$

Therefore, by (47) and (48), on the event $\mathcal{G} \cap \mathcal{G}_1 \cap \mathcal{G}_2$ we have for all $0 \leq j < \epsilon n^{1/3}$

$$u(j) - u(j+1) \geq \frac{n}{3} - \sum_{x=1}^{k} T_x \geq \frac{n}{4}.$$

□

5.3. *Lower bound on the odometer.* The proof of the lower bound in Theorem 1.1 follows readily.

PROOF OF THEOREM 1.1(II). Since $u \geq 0$, Lemma 5.4 implies that for all large enough $n$ and $j \leq \frac{1}{2}\epsilon n^{1/3}$, we have

$$u(j) \geq \frac{1}{8}\epsilon n^{4/3}.$$

Hence, by (46) it follows

$$\mathbb{P}\Big(\inf_{0\le j\le \frac{1}{2}\epsilon n^{1/3}} u(j) \le \frac{1}{8}\epsilon n^{4/3}\Big) \le e^{-n^c}. \tag{50}$$

To complete the proof, we use the symmetric version of (47) to get the following bound: For $j \le 0$,

$$u(j) - u(j-1) - \Delta^j\big(u(j)\big) - \Delta^{j-1}\big(u(j-1)\big) = \sum_{y=-\infty}^{j-1} \big|g_\tau(y)\big|.$$

The following bound

$$\sum_{y=-\infty}^{i} \big|g_\tau(y)\big| \le 2n$$

is trivial since the total number of particles is $2n$. Using the above and the definition of $\mathcal{G}_2$, we get for all $-\epsilon n^{1/3} < j < 0$

$$u(0) - u(j) \le j\Big(2n + \frac{n}{6}\Big).$$

Thus, $u(0) \ge \frac{1}{4}\epsilon n^{4/3}$ implies that for all $j \le \frac{1}{16}\epsilon n^{1/3}$

$$u(-j) \ge \frac{1}{12}\epsilon n^{4/3},$$

which together with (50) completes the proof. $\square$

5.4. *Upper bound on the number of accumulated particles.* The next lemma will be used in the proof of Theorem 1.2.

LEMMA 5.5. *There exists a constant $C > 0$ such that*

$$\sup_{x\in\mathbb{Z}} \mathbb{E}\big(\big|g_\tau(x)\big|\big) < Cn^{\frac{2}{3}}. \tag{51}$$

*Moreover, for any $\epsilon > 0$ there exists $c > 0$*

$$\mathbb{P}\Big[\sup_{x\in\mathbb{Z}} \big|g_\tau(x)\big| \ge n^{\frac{2}{3}+\epsilon}\Big] \le e^{-n^c}. \tag{52}$$

PROOF. For any $x \in \mathbb{Z}$, we have

$$\big|g_\tau(x)\big| \le [S_{x-1} + S_{x+1}]\mathbf{1}(\mathcal{G}) + 2n\mathbf{1}\big(\mathcal{G}^c\big), \tag{53}$$

where the event $\mathcal{G}$ is defined in (17). The first term follows from (49) and the second term is obvious since the total number of particles is $2n$. The proof of (51) now follows from (iii) of Lemma 5.1 and (26). Additionally, using (i) of Lemma 5.1 we get that for any $x \in \mathbb{Z}$ there exists $c > 0$ such that

$$\mathbb{P}\big(\big|g_\tau(x)\big| \geq n^{\frac{2}{3}+\epsilon}\big) \leq e^{-n^c}. \tag{54}$$

Corollary 3.2 says that with probability at least $1 - e^{-n^c}$ for all $|x| \geq n^5$

$$\big|g_\tau(x)\big| = 0,$$

and (52) now follows from (54) by union bound over all $x \in [-n^5, n^5]$. □

REMARK 5. For small enough $\epsilon$ and $k = \epsilon n^{1/3}$, by (46) and (53)

$$\mathbb{P}\left[\sum_{i=-k}^{k} \big|g_\tau(i)\big| \leq n\right] \geq 1 - e^{-n^c}$$

for some positive constant $c$. Thus, at least $n$ particles are supported outside the interval $[-\epsilon n^{1/3}, \epsilon n^{1/3}]$ with probability at least $1 - e^{-n^c}$.

5.5. *Proof of Theorem* 1.2. Let $x = n^{1/3+\epsilon}$. Under the assumption that there are at least $n^{1-\epsilon/2}$ many particles to the right of $x$, for all $\ell \leq x$,

$$\sum_{i=\ell}^{\infty} \big|g_\tau(i)\big| \geq n^{1-\epsilon/2}.$$

Recalling (47), we have

$$\big(u(0) - u(x)\big)\mathbf{1}(\mathcal{G}) \geq \sum_{i=1}^{x} \sum_{y=i}^{\infty} \big|g_\tau(y)\big| \mathbf{1}(\mathcal{G}) - 2 \sum_{i=0}^{x} T_i \mathbf{1}(\mathcal{G}).$$

Now by part (i) of Lemma 5.1 and union bound over $1 \leq \ell \leq x$ there exists a $c > 0$ such that with probability at least $1 - e^{-n^c}$,

$$\sum_{\ell=1}^{x} T_\ell = O\big(n^{1+2\epsilon}\big).$$

Thus, on the event that there are at least $n^{1-\epsilon/2}$ many particles to the right of $x$ we have

$$u(0) - u(x) \geq x n^{1-\epsilon/2} - O\big(n^{1+\epsilon}\big),$$

except on a set of measure at most $e^{-n^c}$. However, this implies that

$$u(0) \geq n^{4/3+\epsilon/2} - O\big(n^{1+\epsilon}\big).$$

Hence, by the upper bound in Theorem 1.1 we conclude that the probability of the event that there are at least $n^{1-\epsilon/2}$ many particles to the right of $x = n^{1/3+\epsilon}$ is less than $e^{-n^c}$ for some positive $c > 0$. The argument for $x = -n^{1/3+\epsilon}$ is symmetric.

**6. Scaling limit for the odometer.** The goal of this section is to prove Theorem 1.4. The first step will be to show that Conjecture 1.3 implies some regularity of the limiting function: we will argue that $w(\cdot)$ is decreasing and three times differentiable on the positive real axis. Moreover, it is the solution of the boundary value problem:

$$w'' = \sqrt{\frac{2}{\pi} w},$$

$$\lim_{h \to 0^+} \frac{w(h) - w(0)}{h} = -1,$$

$$\lim_{h \to \infty} w(h) = 0.$$

At this point, Theorem 1.4 follows by identifying an explicit solution to the above problem and arguing that it is the unique solution.

6.1. *Properties of the expected odometer.* We first make some easy observations about the expected odometer function, denoted by

$$\tilde{u}(x) := \mathbb{E}\big(u(x)\big).$$

Existence of $\tilde{u}(x)$ follows from (ii) of Proposition 3.1 which says that the stopping time of the process $\tau$ has finite expectation, and clearly for all $x \in \mathbb{Z}$

$$u(x) \leq \tau.$$

We first make the following easy observation.

LEMMA 6.1. *For $x \in \mathbb{Z}$, we have*

$$\tilde{u}(x) - \tilde{u}(x+1) = \mathbb{E}\big[\#\{\textit{particles moving from } x \textit{ to } x+1\}\big] \leq n \qquad \textit{for } x \geq 0,$$
$$\tilde{u}(x) - \tilde{u}(x-1) = \mathbb{E}\big[\#\{\textit{particles moving from } x \textit{ to } x+1\}\big] \leq n \qquad \textit{for } x \leq 0.$$

Clearly, every time there is an emission at a site $x \in \mathbb{Z}$, on average one particle moves to $x-1$ and another to $x+1$. Also by symmetry, the number of particles to the right of any positive $x$ is on an average at most $n$. It is straightforward to make this formal and we omit the details. Similar arguments yield the following as well, whose proof we omit, too.

LEMMA 6.2. *$\tilde{u}(x)$ satisfies the following properties*:

(i) *$\tilde{u}(x)$ is an even function*,
(ii) *restricted to $\mathbb{Z}_+$, $\tilde{u}(x)$ is strictly decreasing*,

(iii) *for every* $x \neq 0$,

$$\Delta\tilde{u}(x) = \mathbb{E}\left(\sum_{y=x}^{\infty}|g_\tau(y)| - \sum_{y=x+1}^{\infty}|g_\tau(y)|\right) = \mathbb{E}(|g_\tau(x)|) > 0,$$

*where* $\Delta$ *is the discrete Laplacian*, *that is*,

$$\Delta\tilde{u}(x) = \tilde{u}(x-1) + \tilde{u}(x+1) - 2\tilde{u}(x). \tag{55}$$

6.2. *The differential equation* $w'' = \sqrt{2w/\pi}$. In this section, we work toward the proof of Theorem 1.4. We recall Conjecture 1.3 stated in the Introduction. Note that we have not assumed a priori that $w$ is continuous. Proving this is our first order of business.

LEMMA 6.3. $w$ *is continuous*, (*in fact*, 1-*Lipschitz*) *on* $\mathbb{R}$. *Moreover*, *it is non-increasing on the positive real axis*.

PROOF. (i) of Lemma 6.2 implies that $w$ is an even function. Hence, it suffices to prove that $w$ is 1-Lipschitz on $[0, \infty)$. By Lemma 6.1, for any $x, k \in \mathbb{Z}_{\geq 0}$ we have

$$0 \leq \tilde{u}(x) - \tilde{u}(x+k) = \sum_{j=0}^{k-1}[\tilde{u}(x+j) - \tilde{u}(x+j+1)] \leq kn.$$

Now let $x = \lfloor n^{1/3}\xi \rfloor$ and $k = \lfloor n^{1/3}(\xi + h)\rfloor - \lfloor n^{1/3}\xi \rfloor$. Dividing by $n^{4/3}$ and taking $n \to \infty$, we obtain

$$0 \leq w(\xi) - w(\xi + h) \leq h$$

thus completing the proof of the lemma. $\square$

Recall the set of random variables

$$X_i^x, Y_i^x$$

defined in (6). To go further, we define the following quantities: For $y = x \pm 1$,

$$O_{x,y}^k := \sum_{i=0}^{k} \mathbf{1}(X_i^x = \pm 1),$$

$$W_{x,y}^k := \sum_{i=0}^{k} \mathbf{1}(Y_i^x = \pm 1).$$

In other words,

$O_{x,y}^k :=$ #{oil particles sent from $x$ to $y$ within the first $k$ firings at $x$},

$W_{x,y}^k :=$ #{water particles sent from $x$ to $y$ within the first $k$ firings at $x$}.

To relate these sums to our earlier notation, recalling Definition 6 we have $O^k_{x,x+1} + W^k_{x,x+1} = k + \Delta^x(k)$, and the signed count of particles remaining at $x$ is given by

$$g_\tau(x) = O^{u(x-1)}_{x-1,x} + O^{u(x+1)}_{x+1,x} - W^{u(x-1)}_{x-1,x} - W^{u(x+1)}_{x+1,x}, \tag{56}$$

where $g_t(\cdot)$ was defined in (14).

Consider now the analogous expression using a deterministic portion of each stack [recall $\tilde{u}(x) = \mathbb{E}u(x)$]:

$$\tilde{g}_\tau(x) := O^{\lfloor \tilde{u}(x-1)\rfloor}_{x-1,x} + O^{\lfloor \tilde{u}(x+1)\rfloor}_{x+1,x} - W^{\lfloor \tilde{u}(x-1)\rfloor}_{x-1,x} - W^{\lfloor \tilde{u}(x+1)\rfloor}_{x+1,x}. \tag{57}$$

Because $\tilde{u}$ is deterministic the four terms on the right-hand side are independent. Moreover, each term $O^k_{x,y}$ and $W^k_{x,y}$ for $|x-y| = 1$ is a sum of $k$ independent Bernoulli$(1/2)$ random variables. So the right-hand side is a sum of $\lfloor \tilde{u}(x-1)\rfloor + \lfloor \tilde{u}(x+1)\rfloor$ independent random variables with the same law as a single step of a lazy symmetric random walk as defined in (39). Setting $x = \lfloor n^{1/3}\xi\rfloor$ for a real number $\xi > 0$, by (ii) in Conjecture 1.3 we have

$$\lim_{n\to\infty} \frac{\lfloor \tilde{u}(x-1)\rfloor + \lfloor \tilde{u}(x+1)\rfloor}{n^{4/3}} = \lim_{n\to\infty} \frac{2\tilde{u}(x)}{n^{4/3}} = 2w(\xi).$$

This is because by Lemma 6.1, $|\tilde{u}(x) - \tilde{u}(x+1)|$ and $|\tilde{u}(x-1) - \tilde{u}(x)|$ are both less than $n$. As $n \to \infty$ by the central limit theorem, since each variable in (39) has variance $1/2$, we obtain

$$n^{-2/3}\tilde{g}_\tau(x) \xrightarrow{d} N\bigl(0, w(\xi)\bigr). \tag{58}$$

By Lemma 5.1(v), we also have

$$n^{-2/3}\mathbb{E}\bigl|\tilde{g}_\tau(x)\bigr| \longrightarrow \sqrt{\frac{2}{\pi} w(\xi)}. \tag{59}$$

Next, we observe that under (i) in Conjecture 1.3, the same kind of central limit theorem holds for $g_\tau$ itself.

LEMMA 6.4. *Let $\xi \geq 0$. For $x = \lfloor n^{1/3}\xi\rfloor$, we have as $n \to \infty$*

(i)

$$n^{-2/3} g_\tau(x) \xrightarrow{d} N\bigl(0, w(\xi)\bigr)$$

(ii)

$$n^{-2/3}\mathbb{E}\bigl|g_\tau(x)\bigr| \longrightarrow \sqrt{\frac{2}{\pi} w(\xi)}.$$

REMARK 6. (ii) along with (55) implies

$$\lim_{n\to\infty} \frac{\Delta \tilde{u}(x)}{n^{2/3}} \to \sqrt{\frac{2}{\pi}} w(\xi). \tag{60}$$

To prove Lemma 6.4, we need the next two results.

LEMMA 6.5. $\mathbb{E}[\tau \mathbf{1}(\tau > n^5)] = O(1)$ *where* $\tau$ *is defined in* (10).

COROLLARY 6.6. *There exists a constant* $C_1 > 0$ *such that*

$$\sup_{x\in\mathbb{Z}} \tilde{u}(x) \le C_1 n^{4/3},$$

*where* $\tilde{u} = \mathbb{E}(u)$.

The proofs of the above two results are deferred to the Appendix.

6.2.1. *Proof of Lemma* 6.4. By (58) and (59), it suffices to show

$$\lim_{n\to 0} n^{-2/3}\mathbb{E}\big|g_\tau(x) - \tilde{g}_\tau(x)\big| = 0. \tag{61}$$

Referring to the definitions of $g$ and $\tilde{g}$ in (56) and (57), respectively, this will be accomplished if we show that as $n \to \infty$ for $y = x \pm 1$:

$$n^{-2/3}\mathbb{E}\big|\big(O_{y,x}^{\lfloor \tilde{u}(y)\rfloor} - W_{y,x}^{\lfloor \tilde{u}(y)\rfloor}\big) - \big(O_{y,x}^{u(y)} - W_{y,x}^{u(y)}\big)\big|$$

tend to 0. For $y = x \pm 1$, the above differences have identical distributions. Hence, it suffices to look at any one. The quantity $(O_{y,x}^{\tilde{u}(y)} - W_{y,x}^{\tilde{u}(y)}) - (O_{y,x}^{u(y)} - W_{y,x}^{u(y)})$ is a sum of

$$N = \big|\lfloor \tilde{u}(y)\rfloor - u(y)\big|$$

independent random variables $X_1, \dots, X_N$ with the same law as in (39). By Conjecture 1.3(i), $N/n^{4/3} \to 0$ in distribution. Fix $\epsilon \ge 0$. Let

- $A_1 = \mathbf{1}(N \le \epsilon n^{4/3}) \sup_{1\le m\le \epsilon n^{4/3}} |\sum_{i=1}^m X_i|$,
- $A_2 = \mathbf{1}(N \ge \epsilon n^{4/3})\mathbf{1}(\mathcal{G}) \sup_{1\le m\le (C+C_1)n^{4/3}} |\sum_{i=1}^m X_i|$,
- $A_3 = 2n^5\mathbf{1}(\mathcal{G}^c)\mathbf{1}(\tau \le n^5)$,
- $A_4 = 2\tau\mathbf{1}(\mathcal{G}^c)\mathbf{1}(\tau > n^5)$,

where $C$ and $C_1$ are the constants appearing in the statement of Theorem 1.1 and Corollary 6.6, respectively, $\tau$ is defined in (10) and $\mathcal{G}$ is defined in (17). We now claim that

$$\big|\big(O_{y,x}^{\lfloor \tilde{u}(y)\rfloor} - W_{y,x}^{\lfloor \tilde{u}(y)\rfloor}\big) - \big(O_{y,x}^{u(y)} - W_{y,x}^{u(y)}\big)\big| \le A_1 + A_2 + A_3 + A_4. \tag{62}$$

The first two terms correspond to the cases:

- $N \le \epsilon n^{4/3}$,
- $\{N \ge \epsilon n^{4/3}\} \cap \mathcal{G}$.

For the last two terms, we use the naive bound that

$$\left|\sum_{i=1}^{N} X_i\right| \le N \le \sup_{x\in\mathbb{Z}} \tilde{u}(x) + \sup_{x\in\mathbb{Z}} u(x) \le C_1 n^{4/3} + \tau, \tag{63}$$

where the last inequality uses Corollary 6.6. Using the above bound and looking at the events $\{\tau \le n^5\} \cap \mathcal{G}^c$ and $\{\tau > n^5\} \cap \mathcal{G}^c$ gives us (62).

- By Lemma 5.1(iii), $\mathbb{E}(A_1) = O(\sqrt{\epsilon} n^{2/3})$.
- By the Cauchy–Schwarz inequality and Lemma 5.1(iv)
$$\mathbb{E}(A_2) = O(n^{2/3})\sqrt{\mathbb{P}(N \ge \epsilon n^{4/3})}.$$
- $\mathbb{E}(A_3 + A_4) = O(1)$ by Theorem 1.1 and Lemma 6.5, respectively.

Thus, for any $\epsilon > 0$

$$\mathbb{E}\left|g_\tau(x) - \tilde{g}_\tau(x)\right| \le \sum_{i=1}^{4} \mathbb{E}(A_i) = n^{2/3}\left(O\left(\sqrt{\epsilon} + \sqrt{\mathbb{P}(N \ge \epsilon n^{4/3})}\right)\right).$$

Hence, (61) follows using the above and Conjecture 1.3(i) ($N/n^{4/3}$ goes to 0 in distribution) and we are done.

REMARK 7. Note that we actually prove (61) uniformly over $x$, that is,

$$\lim_{n\to 0} \sup_{x\in\mathbb{Z}} n^{-2/3}\mathbb{E}\left|g_\tau(x) - \tilde{g}_\tau(x)\right| = 0.$$

We now prove an uniform version of (60).

LEMMA 6.7. *Given $\epsilon > 0$ and $x < y$ such that $w(x), w(y) > 0$, for large enough $n$,*

$$\sup_{\lfloor xn^{1/3}\rfloor \le j \le \lfloor yn^{1/3}\rfloor} \left|\frac{\Delta\tilde{u}(j)}{n^{2/3}} - \sqrt{\frac{2}{\pi} w\left(\frac{j}{n^{1/3}}\right)}\right| \le \epsilon.$$

PROOF. Since $w$ is continuous by Lemma 6.3, and hence uniformly continuous on $[x, y]$, for $\epsilon > 0$ there exists real numbers:

$$x = x_0 < x_1 < \cdots < x_k = y$$

such that

$$\sup_{1\le i\le k} (x_i - x_{i-1}) \le \epsilon,$$

$$\sup_{1\le i\le k} \left(w(x_{i-1}) - w(x_i)\right) \le \epsilon.$$

By Conjecture 1.3(ii) and (60), we have for large enough $n$:

$$\sup_{1\le i\le k}\left|\frac{\tilde{u}(\lfloor n^{1/3}x_i\rfloor)}{n^{4/3}}-w(x_i)\right|\le\epsilon, \tag{64}$$

$$\sup_{1\le i\le k}\left|\frac{\Delta\tilde{u}(\lfloor n^{1/3}x_i\rfloor)}{n^{2/3}}-\sqrt{\frac{2}{\pi}w(x_i)}\right|\le\epsilon. \tag{65}$$

Now for any $\lfloor xn^{1/3}\rfloor\le j\le\lfloor yn^{1/3}\rfloor$ find $0\le i<k$ such that

$$\lfloor x_i n^{1/3}\rfloor\le j\le\lfloor x_{i+1}n^{1/3}\rfloor.$$

Clearly, it suffices to show

$$\left|\Delta\tilde{u}(j)-\Delta\tilde{u}(\lfloor x_i n^{1/3}\rfloor)\right|=O(\sqrt{\epsilon}n^{2/3})$$

or by (55)

$$\left|\mathbb{E}|g_\tau(j)|-\mathbb{E}|g_\tau(\lfloor x_i n^{1/3})|\right|=O(\sqrt{\epsilon}n^{2/3}).$$

Notice that by (61) and Remark 7 we have

$$\left|\mathbb{E}|g_\tau(j)|-\mathbb{E}|\tilde{g}_\tau(j)|\right|=o(n^{2/3}), \tag{66}$$

$$\left|\mathbb{E}|g_\tau(\lfloor n^{1/3}x_i\rfloor)|-\mathbb{E}|\tilde{g}_\tau(\lfloor n^{1/3}x_i\rfloor)|\right|=o(n^{2/3}). \tag{67}$$

Hence, it suffices to show

$$\left|\mathbb{E}|\tilde{g}_\tau(j)|-\mathbb{E}|\tilde{g}_\tau(\lfloor n^{1/3}x_i\rfloor)|\right|=O(\sqrt{\epsilon}n^{2/3}), \tag{68}$$

$$\left|\mathbb{E}|\tilde{g}_\tau(\lfloor n^{1/3}x_{i+1}\rfloor)|-\mathbb{E}|\tilde{g}_\tau(j)|\right|=O(\sqrt{\epsilon}n^{2/3}). \tag{69}$$

By (64)

$$\tilde{u}(j)-\tilde{u}(\lfloor n^{1/3}x_i\rfloor)\le\epsilon n^{4/3},$$
$$\tilde{u}(\lfloor n^{1/3}x_{i+1}\rfloor)-\tilde{u}(j)\le\epsilon n^{4/3}.$$

Now by (57) the quantities on the left-hand side of (68) and (69) are absolute values of a lazy symmetric random walk run for time $\tilde{u}(j)-\tilde{u}(\lfloor n^{1/3}x_i\rfloor)$ and $\tilde{u}(\lfloor n^{1/3}x_{i+1}\rfloor)-\tilde{u}(j)$, respectively. The result now follows from Lemma 5.1(v). □

COROLLARY 6.8.

$$\int_0^\infty\sqrt{\frac{2}{\pi}w(\zeta)}\,d\zeta\le 1,$$

*which in particular implies*

$$\lim_{x\to\infty}w(x)=0.$$

PROOF. By Lemma 6.1,

$$\mathbb{E}\left(\sum_{y=1}^{\infty}|g_\tau(y)|\right)=\sum_{y=1}^{\infty}\Delta\tilde{u}(y)=\tilde{u}(0)-\tilde{u}(1)\le n. \tag{70}$$

Thus, for any positive number $A$

$$\sum_{y=1}^{\lfloor An^{1/3}\rfloor}\Delta\tilde{u}(y)\le n.$$

By Lemma 6.7 and the approximation of an integral by Riemann sum, we have

$$1\ge\lim_{n\to\infty}\sum_{y=1}^{An^{1/3}}\frac{1}{n^{1/3}}\frac{\Delta\tilde{u}(y)}{n^{2/3}}=\int_0^A\sqrt{\frac{2}{\pi}w(\zeta)}\,d\zeta.$$

Since $w$ is nonnegative, it follows that

$$\int_0^\infty\sqrt{\frac{2}{\pi}w(\zeta)}\,d\zeta\le 1.$$

By Lemma 6.3 $w$ is nonincreasing, hence this implies that

$$\lim_{x\to\infty}w(x)=0. \qquad\square$$

REMARK 8. Note that we assumed only convergence of $\tilde{u}$ in Conjecture 1.3(ii) but were able to use a special feature of the oil and water model [namely, the identity $\Delta\tilde{u}(x)=\mathbb{E}|g_\tau(x)|$] to obtain something stronger, convergence of the discrete Laplacian $\Delta\tilde{u}$.

Next, we use this to argue that the scaling limit $w(\xi)$ is actually a twice differentiable function of $\xi>0$.

For any $\epsilon>0$ by Conjecture 1.3(i) and the above corollary, we can choose $L$ large enough so that for large enough $n$,

$$\sup_{|\xi|\ge L}\tilde{u}\left(\lfloor n^{1/3}\xi\rfloor\right)<\varepsilon^2 n^{4/3}$$

for all $|\xi|>L$.

LEMMA 6.9. *Given $\epsilon>0$, let $L$ be as chosen above. Then*

$$\sum_{|x|>n^{1/3}L}\Delta\tilde{u}(x)\le\varepsilon^{1/2}n.$$

PROOF. Since the $n^{1/3}\varepsilon$ differences $\tilde{u}(x) - \tilde{u}(x+1)$ for $x = \lfloor n^{1/3}L \rfloor, \dots, \lfloor n^{1/3}(L+\varepsilon) \rfloor - 1$ are nonnegative and sum to at most $\tilde{u}(\lfloor n^{1/3}L \rfloor) \le \varepsilon^2 n^{4/3}$, the smallest of them (which is the last one) must be at most $\varepsilon n$. Therefore,

$$\sum_{|x| > n^{1/3}(L+\varepsilon)} \Delta\tilde{u}(x) \le \varepsilon n.$$

Now the fact that

$$\sum_{x=\lfloor n^{1/3}L \rfloor}^{\lfloor n^{1/3}(L+\varepsilon) \rfloor} \Delta\tilde{u}(x) \le O(\varepsilon)n$$

follows from Lemma 5.5 and the fact that $\Delta\tilde{u}(x) = \mathbb{E}|g_\tau(x)|$ [see (55)]. Combining the above two results the proof follows. $\square$

LEMMA 6.10. *$w$ is differentiable on the positive real line, and for any $\xi > 0$,*

$$w'(\xi) = -\int_\xi^\infty \sqrt{\frac{2}{\pi} w(\zeta)}\, d\zeta. \tag{71}$$

PROOF. By summation by parts, for positive integers $x, k$,

$$\frac{1}{kn}\big[\tilde{u}(x) - \tilde{u}(x+k)\big] = \frac{1}{kn} \sum_{j=1}^{\infty} \min(j,k) \Delta\tilde{u}(x+j). \tag{72}$$

For positive real numbers $\xi, L, h$ let $x = \lfloor \xi n^{1/3} \rfloor$, $Z = \lfloor L n^{1/3} \rfloor$, $k = \lfloor h n^{1/3} \rfloor$ and consider the first part of the sum in (72):

$$\frac{1}{h} \sum_{j=1}^{Z} \frac{1}{n^{\frac{1}{3}}} \frac{\min(j,k)}{n^{\frac{1}{3}}} \frac{\Delta\tilde{u}(x+j)}{n^{\frac{2}{3}}}.$$

Now given $\delta > 0$, by Lemma 6.7 for large enough $n$:

$$\left| \frac{1}{h} \sum_{j=1}^{Z} \frac{1}{n^{\frac{1}{3}}} \frac{\min(j,k)}{n^{\frac{1}{3}}} \frac{\Delta\tilde{u}(x+j)}{n^{\frac{2}{3}}} - \frac{1}{h} \sum_{j=1}^{Z} \frac{1}{n^{\frac{1}{3}}} \min\left(\frac{j}{n^{\frac{1}{3}}}, h\right) \sqrt{\frac{2}{\pi} w\left(\xi + \frac{j}{n^{\frac{1}{3}}}\right)} \right| \le \delta L.$$

Notice that

$$\frac{1}{h} \sum_{j=1}^{Z} \frac{1}{n^{\frac{1}{3}}} \min\left(\frac{j}{n^{\frac{1}{3}}}, h\right) \sqrt{\frac{2}{\pi} w\left(\xi + \frac{j}{n^{\frac{1}{3}}}\right)}$$

is a Riemann sum approximation of the integral

$$\frac{1}{h} \int_\xi^{L+\xi} \min(\zeta - \xi, h) \sqrt{\frac{2}{\pi} w(\zeta)}\, d\zeta.$$

Thus, as $n$ goes to $\infty$ we see that

$$\frac{1}{h}\sum_{j=1}^{Z}\frac{1}{n^{\frac{1}{3}}}\frac{\min(j,k)}{n^{\frac{1}{3}}}\frac{\Delta\tilde{u}(x+j)}{n^{\frac{2}{3}}} \to \frac{1}{h}\int_{\xi}^{L+\xi}\min(\zeta-\xi,h)\sqrt{\frac{2}{\pi}w(\zeta)}\,d\zeta.$$

Fixing $\epsilon > 0$ and choosing the same $L$ as in the statement of Lemma 6.9, we get that the sum of the remaining terms in (72) $\frac{1}{kn}\sum_{j=Z+1}^{\infty}$ is at most $\varepsilon^{1/2}$ by Lemma 6.9. Hence, as $n \to \infty$ we get from (72):

$$\frac{w(\xi)-w(\xi+h)}{h} = \frac{1}{h}\int_{\xi}^{L}\min(\zeta-\xi,h)\sqrt{\frac{2}{\pi}w(\zeta)}\,d\zeta + O(\epsilon^{1/2}).$$

Sending $h$ to 0 followed by $\epsilon$ to 0 ($L$ to $\infty$), we are done. $\square$

The right-hand side of (71) is manifestly a differentiable function of $\xi$, so we obtain the following.

COROLLARY 6.11. *Under Conjecture 1.3, the function $w$ restricted to the positive real axis is twice continuously differentiable and obeys the differential equation*:

$$w'' = \sqrt{\frac{2}{\pi}w}. \tag{73}$$

LEMMA 6.12. *$w$ is compactly supported. Moreover, on the positive region of support*,

$$w(x) = \left(-\frac{1}{4}\left(\frac{32}{9\pi}\right)^{1/4}x + b\right)^4$$

*for some $b > 0$.*

Before proving the above, we quote the well known Picard existence and uniqueness result for ODEs.

THEOREM 6.13 [Kelley and Peterson (2010), Theorem 8.13]. *Consider an initial value problem (IVP)*:

$$y'(x) = f\big(y(x), x\big), \tag{74}$$

$$y(x_0) = y_0 \tag{75}$$

*with the point $(x_0, y_0)$ belonging to some rectangle $(a, b) \times (A, B)$, that is, $a < a_0 < b$ and $A < y_0 < B$. Also assume that $f$ is $M$-Lipschitz for some $M \geq 0$, that is*,

$$\big|f(z,x) - f(w,x)\big| \leq M|z-w|$$

*for all $x \in (a, b)$ $z, w \in (A, B)$. Then there exists a $h = h(x_0, y_0, M) > 0$ such that*:

- Existence: *There exists a solution to the IVP on the interval* $(x_0 - h, x_0 + h)$.
- Uniqueness: *Any two solutions of the IVP agree on the interval* $(x_0 - h, x_0 + h)$.

PROOF OF LEMMA 6.12. Multiplying (73) by $w'$ on both sides, we get

$$w'w'' = \sqrt{\frac{2}{\pi}} w w'.$$

Integrating both sides from $\xi$ to $\infty$ and using the fact that

$$\lim_{x\to\infty} w(x) = \lim_{x\to\infty} w'(x) = 0$$

(from Corollary 6.8 and Lemma 6.10) and that $w'$ is nonpositive we see that $w(x)$ satisfies the first-order ODE:

$$f' = -\left(\frac{32}{9\pi}\right)^{1/4} f^{3/4}. \tag{76}$$

Now suppose $w$ is positive on the entire real axis. Given any $z \in \mathbb{R}_{+,}$ then $w(z)$ and $w'(z)$ are both nonzero. Thus, we can find $a, b$ such that

$$(az + b)^4 = w(z),$$
$$4a(az + b)^3 = w'(z).$$

By (76),

$$4a = -\left(\frac{32}{9\pi}\right)^{1/4}.$$

Because of the particular choice of $a$ and $b$ the function $(ax + b)^4$ also satisfies (76). Now since $w(z)$ and $w'(z)$ are both nonzero the function $w^{3/4}(z)$ is Lipschitz in a neighborhood of $z$. Hence, by Theorem 6.13 ODE (76) has an unique solution in some neighborhood of $z$. Thus, the functions $w(x)$ and $(ax + b)^4$ are equal in a neighborhood of $z$. Now looking at the biggest interval $I$ containing $z$ such that $w(x) = (ax + b)^4$ on $I$ we conclude that $w(x) = (ax + b)^4$ on $\mathbb{R}_+ \cap \operatorname{supp}(w)$. In particular, since $(ax + b)^4$ is positive only on a compact set this implies that $w(x)$ has compact support. $\square$

Now we find the value for $b$ which completely determines $w$.

LEMMA 6.14.

$$-4ab^3 = \lim_{h\to 0^+} \frac{w(0) - w(h)}{h} = 1.$$

*In particular,*

$$b = \left(\frac{9\pi}{32}\right)^{1/12}.$$

PROOF. That $-4ab^3 = \lim_{h\to 0^+} \frac{w(0)-w(h)}{h}$ follows from Lemma 6.12. To see that, the quantity equals 1 fix $h > 0$. Consider the telescopic sum:

$$\tilde{u}(0) - \tilde{u}(hn^{1/3}) = \sum_{i=0}^{hn^{1/3}} \tilde{u}(i) - \tilde{u}(i+1).$$

Now $\tilde{u}(i) - \tilde{u}(i+1)$ is the expected number of particles on the right of $i$ by Lemma 6.1. By symmetry of the process about the origin and Lemma 5.5,

$$\sum_{x>0} |g_\tau(x)| = n - O(n^{2/3}).$$

Moreover, for any $i > 0$

$$\tilde{u}(i) - \tilde{u}(i+1) = n - O(n^{2/3})i.$$

Summing over $i$, we get

$$\tilde{u}(0) - \tilde{u}(\lfloor hn^{1/3} \rfloor) = hn^{4/3} - h^2 O(n^{4/3}).$$

Dividing throughout by $n^{4/3}$ and taking limit as $n$ goes to infinity, we get

$$w(0) - w(h) = h + O(h^2).$$

Thus, dividing by $h$ and sending $h$ to 0 we are done. □

6.3. *Proof of Theorem* 1.4. From Lemmas 6.12 and 6.14 and using the symmetry of $w$ about the origin, we get

$$w(x) = \left(\left(\frac{9\pi}{32}\right)^{1/12} - \left(\frac{32}{9\pi}\right)^{1/4} \frac{|x|}{4}\right)^4$$

on the region of support. Rearranging we get

$$w(x) = \begin{cases} \dfrac{1}{72\pi}\big((18\pi)^{1/3} - |x|\big)^4, & |x| < (18\pi)^{1/3}, \\ 0, & |x| \geq (18\pi)^{1/3}. \end{cases} \tag{77}$$

**7. Open questions.** Conjecture 1.3 is an obvious target. In this concluding section, we collect some additional open questions.

7.1. *Location of the rightmost particle.* For the oil and water process with $n$ particles of each type started at the origin $\mathbb{Z}$, let $R_n$ be the location of the rightmost particle upon fixation. Is the sequence of random variables $R_n/n^{1/3}$ tight? Does it converge in distribution to a constant? If it does, then Theorem 1.4 suggests that the limit should be at least $(18\pi)^{1/3}$ (and perhaps equal to this value).

7.2. *Order of the variance.* We believe that the standard deviation of the odometer $u$ is of order $n^{7/6}$ in the bulk. Note that Conjecture 1.3 asserts something weaker, namely $o(n^{4/3})$.

Here is a heuristic argument for the exponent $7/6$. The total number of particle exits from $x$ is $2u(x)$; let $N_x$ be the total number of particle *entries* to $x$. Equating entries minus exits with the number of particles left behind, we find that

$$\Delta u(x) := u(x-1) + u(x+1) - 2u(x) = Z(x) + |g_\tau(x)| - 2n\delta_0(x), \tag{78}$$

where $Z(x) = u(x-1) + u(x+1) - N_x$, and $g_\tau(x)$ is the signed count of particles remaining at $x$ in the final state (counting oil as positive, water as negative). Both $Z(x)$ and $g_\tau(x)$ (without the absolute value!) are expressible as sums of independent indicators involving the stack elements at $x \pm 1$. The limits of summation are $u(x \pm 1)$. Assuming Conjecture 1.3 and arguing as in Lemma 6.4, we can replace the limits of summation by their expected values $\tilde{u}(x \pm 1)$, incurring only a small error. The resulting sums $\tilde{Z}$ and $\tilde{g}_\tau$ are asymptotically normal with mean zero and variance of order $n^{4/3}$ [assuming $x$ is in the bulk, $|x| < ((18\pi)^{1/3} - \varepsilon)n^{1/3}$]. Moreover, the function $\tilde{Z} + |\tilde{g}_\tau|$ is 2-*dependent*: its values at $x$ and $y$ are independent if $|x - y| > 2$. By summation by parts,

$$u(x) = \sum_{j=1}^{\infty} j\,\Delta u(x + j).$$

Since most of the support of $u$ is on an interval of length $O(n^{1/3})$, truncating this sum at $Cn^{1/3}$ for a large constant $C$ should not change its variance by much. Replacing $\Delta u$ by its approximation $\tilde{Z} + |\tilde{g}_\tau|$ and using the 2-dependence, we arrive at

$$\operatorname{Var} u(x) = \sum_{j=1}^{Cn^{1/3}} j^2 O(n^{4/3}) = O(n^{7/3}).$$

7.3. *Conjectured exponents in higher dimensions.* For the oil and water model in $\mathbb{Z}^d$ starting with $n$ oil and $n$ water particles at the origin, we believe that the typical order of the odometer (away from 0 and the boundary) is $n^{4/(d+2)}$ and the radius of the occupied cluster is of order $n^{1/(d+2)}$. The reason is by analogy with Section 6.2: if $w : \mathbb{R}^d \to \mathbb{R}$ solves the PDE:

$$\Delta w = -\delta_0 + \sqrt{\frac{2}{\pi} w} \tag{79}$$

then its rescaling

$$v(x) = t^4 w(x/t)$$

satisfies

$$\Delta v = -t^{d+2}\delta_0 + \sqrt{\frac{2}{\pi} v}.$$

If the odometer for $n$ particles has a scaling limit $w$ that satisfies (79), then $v$ is the scaling limit of the odometer for $t^{d+2}n$ particles. So increasing the number of particles a factor of $t^{d+2}$ increases the radius by by a factor of $t$ and the odometer by a factor of $t^4$. This motivates the following conjecture.

CONJECTURE 7.1. *Let $u$ be the odometer for the oil and water model started from $n$ particles of each type at the origin in $\mathbb{Z}^d$. There exists a deterministic function $w : \mathbb{R}^d \to \mathbb{R}$ such that for all $\xi \in \mathbb{R} - \{0\}$ we have almost surely*,

$$\frac{u(\lfloor n^{1/(d+2)}\xi \rfloor)}{n^{4/(d+2)}} \to w(\xi).$$

*Moreover, $w$ is rotationally symmetric, twice differentiable on $\mathbb{R}^d - \{0\}$ and satisfies*

$$\Delta w = \sqrt{\frac{2}{\pi} w},$$

*on $\mathbb{R}^d - \{0\}$ and $\lim_{\xi \to 0} \frac{w(\xi)}{g(\xi)} = 1$ where $g$ is the Green function for the Laplacian on $\mathbb{R}^d$.*

The fourth power scaling is reflected in the even spacing between contour lines of the odometer function in Figure 3.

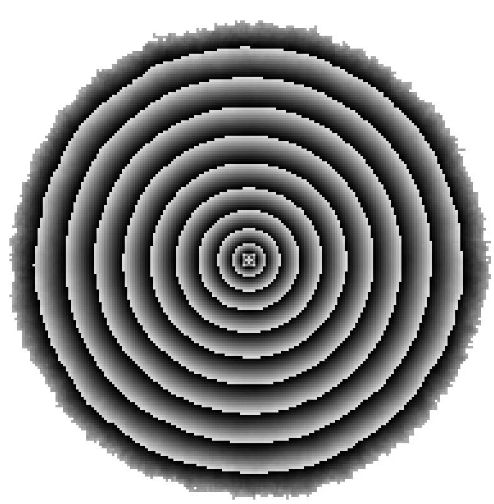

FIG. 3. *Contour lines of the odometer function $u$ of the oil and water model in $\mathbb{Z}^2$ with $n = 2^{22}$ particles of each type started at the origin. Each site is shaded according to the fractional part of $\frac{1}{5}u^{1/4}$.*

## APPENDIX

### A.1. Concentration estimates.

LEMMA A.1. *Suppose for all $x \in \mathbb{Z}$ we have*

$$X_1^x, Y_1^x$$
$$X_2^x, Y_2^x$$
$$\vdots$$

*a sequence of independent uniform $\pm 1$ valued random variables. The sequences across $x$ are also independent of each other. Then there exists constants $C, C', \gamma > 0$ such that for $n$ large enough, with probability at least $1 - C\exp(-C'n^{\gamma})$ for all $k > \sqrt{n}$ and $-n^5 < j < n^5$ we have:*

(i) $|\sum_{i=1}^{k} \mathbf{1}_{(X_i^j=1)} - k/2| < k^{1/2+\epsilon}$;
(ii) $|\sum_{i=1}^{k} \mathbf{1}_{(Y_i^j=1)} - k/2| < k^{1/2+\epsilon}$;
(iii) $|\sum_{i=1}^{k} \mathbf{1}_{(X_i^j=1)}\mathbf{1}_{(Y_i^j=1)} - k/4| < k^{1/2+\epsilon}$;
(iv) $|\sum_{i=1}^{k} \mathbf{1}_{(X_i^j=-1)}\mathbf{1}_{(Y_i^j=-1)} - k/4| < k^{1/2+\epsilon}$.

PROOF. Proof follows by standard bounds from Azuma–Hoeffding's inequality for Bernoulli random variables and union bound over $k \geq \sqrt{n}$ followed by $j \in [-n^5, n^5]$. $\square$

**A.2. Proof of Lemma 5.2.** Let us define the truncated variable:

$$Y = M\big(n^{\frac{4}{3}}\big)\mathbf{1}\big(M\big(n^{\frac{4}{3}}\big) < n^{\frac{2}{3}+\epsilon}\big)$$

for some small but a priori fixed $\epsilon$. Let $Y_i$ be i.i.d. copies of $Y$. Now by using Azuma's inequality,

$$\mathbb{P}\left(\sum_{i=1}^{n^{1/3}}\big(Y_i - \mathbb{E}(Y_i)\big) > t\right) \leq e^{-\frac{t^2}{n^{1/3}n^{4/3+2\epsilon}}}.$$

Taking $t^2 = n^{5/3+3\epsilon}$, we get that

$$\mathbb{P}\left(\sum_{i=1}^{n^{1/3}}\big(Y_i - \mathbb{E}(Y_i)\big) > t\right) < e^{-n^{\epsilon}}.$$

Now by (iv) Lemma 5.1 $\mathbb{E}(Y) = O(n^{2/3})$. Thus,

$$n^{1/3}\mathbb{E}(Y) + t < Dn$$

for some large $D$ as $t = n^{5/6+2\epsilon}$. Hence,

$$\mathbb{P}\left(\sum_{i=1}^{n^{1/3}} Y_i > Dn\right) \le e^{-n^\epsilon}.$$

This implies that

$$\mathbb{P}\left(\sum_{i=1}^{n^{1/3}} M^i(n^{\frac{4}{3}}) > Dn\right) \le Ce^{-n^\gamma}$$

since by (i) Lemma 5.1 and union bound, there exists a positive constant $c > 0$ such that

$$\mathbb{P}(\exists 1 \le i \le n^{1/3} \text{ such that } Y_i \ne M^i) \le e^{-n^c}.$$

**A.3. Proof of Lemma 6.5.** We use the variables $E_i$ defined in the statement of Lemma 3.5. Let

$$Y = \sum_{i=1}^{\tau'} E_i,$$

where $\tau'$ was defined in (12). As mentioned in proof of Proposition 3.1 by Lemmas 3.3 and 3.5 $\tau$ is stochastically dominated by $Y$. Thus,

$$\mathbb{E}[\tau\mathbf{1}(\tau > n^5)] \le \mathbb{E}[Y\mathbf{1}(Y > n^5)].$$

Hence, to prove the lemma it suffices to show the right-hand side is $O(1)$. Now

$$\mathbb{E}[Y\mathbf{1}(Y > n^5)] \le \mathbb{E}[Y\mathbf{1}(\tau' > n^{\frac{5}{2}})] + \mathbb{E}[Y\mathbf{1}(\tau' \le n^{\frac{5}{2}})\mathbf{1}(Y > n^5)] \tag{80}$$

$$\le \mathbb{E}[Y\mathbf{1}(\tau' > n^{\frac{5}{2}})] + \mathbb{E}\left[\sum_{i=1}^{n^{\frac{5}{2}}} E_i\mathbf{1}\left(\sum_{i=1}^{n^{\frac{5}{2}}} E_i \ge n^5\right)\right] \tag{81}$$

$$\le \mathbb{E}[Y\mathbf{1}(\tau' > n^{\frac{5}{2}})] + \mathbb{E}\left[\sum_{i=1}^{n^{\frac{5}{2}}} E_i\right]\left[\sum_{i=1}^{n^{\frac{5}{2}}} \mathbf{1}(E_i \ge n^{\frac{5}{2}})\right], \tag{82}$$

where the last inequality follows from the easy fact

$$\mathbf{1}\left(\sum_{i=1}^{n^{\frac{5}{2}}} E_i \ge n^5\right) \le \sum_{i=1}^{n^{\frac{5}{2}}} \mathbf{1}(E_i \ge n^{\frac{5}{2}}).$$

We use the following tail estimate for $E_1$ and $\tau'$: there exists a constant $c < 1$ such that for $k \ge n^2$:

$$\max(\mathbb{P}(\tau' \ge k), \mathbb{P}(E_1 \ge k)) \le (1-c)^{\lfloor \frac{k}{n^2} \rfloor}, \tag{83}$$

which easy follows from the fact that starting from any point in $[-2n, 2n]$ there exists a constant chance $c$ for the random walk to exit the interval in the next $n^2$ steps. Using (83), independence of $\tau'$, $E_i's$, the theorem now follows from (82). The details are omitted.

**A.4. Proof of Corollary 6.6.** The proof follows from the following observation:

$$u(x) \le Cn^{4/3}\mathbf{1}(\mathcal{G}) + n^5\mathbf{1}(\mathcal{G}^c)\mathbf{1}(\tau \le n^5) + \tau\mathbf{1}(\tau \ge n^5)\mathbf{1}(\mathcal{G}^c), \tag{84}$$

where $C$ is the constant appearing in the statement of Theorem 1.1, $\tau$ is defined in (10) and $\mathcal{G}$ is defined in (17). The first term follows from the definition of $\mathcal{G}$. For the second and third term, we use the trivial bound that

$$u(x) \le \tau.$$

Taking expectation, we get

$$\tilde{u}(x) \le Cn^{4/3} + n^5\mathbb{P}(\mathcal{G}^c) + \mathbb{E}(\tau\mathbf{1}(\tau \ge n^5)).$$

The last two terms are $O(1)$ by Theorem 1.1 and Lemma 6.5, respectively. Hence, we are done.

**Acknowledgments.** We thank the anonymous referee for useful comments. We are grateful to Alexander Holroyd and Yuval Peres for valuable discussions, and to Deepak Dhar for bringing the reference Alcaraz, Pyatov and Rittenberg (2009) to our attention. We thank Wilfried Huss and Ecaterina Sava-Huss for a careful reading of an earlier draft, and for detailed comments which improved the paper.

## REFERENCES

ALCARAZ, F. C., PYATOV, P. and RITTENBERG, V. (2009). Two-component Abelian sandpile models. *Phys. Rev. E* (3) **79** 042102.

ASSELAH, A. and GAUDILLIÈRE, A. (2013a). From logarithmic to subdiffusive polynomial fluctuations for internal DLA and related growth models. *Ann. Probab.* **41** 1115–1159. MR3098673

ASSELAH, A. and GAUDILLIÈRE, A. (2013b). Sublogarithmic fluctuations for internal DLA. *Ann. Probab.* **41** 1160–1179. MR3098674

ASSELAH, A. and GAUDILLIÈRE, A. (2014). Lower bounds on fluctuations for internal DLA. *Probab. Theory Related Fields* **158** 39–53. MR3152779

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

BOND, B. and LEVINE, L. (2016). Abelian networks I. Foundations and examples. *SIAM J. Discrete Math.* **30** 856–874. MR3493110

DHAR, D. (1999). The Abelian sandpile and related models. *Phys. A* **263** 4–25.

DURRETT, R. (2010). *Probability: Theory and Examples*, 4th ed. Cambridge Univ. Press, Cambridge.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications. Vol. I*, 3rd ed. Wiley, New York. MR0228020

JERISON, D., LEVINE, L. and SHEFFIELD, S. (2012). Logarithmic fluctuations for internal DLA. *J. Amer. Math. Soc.* **25** 271–301. MR2833484

JERISON, D., LEVINE, L. and SHEFFIELD, S. (2013). Internal DLA in higher dimensions. *Electron. J. Probab.* **18** 1–14.

JERISON, D., LEVINE, L. and SHEFFIELD, S. (2014). Internal DLA and the Gaussian free field. *Duke Math. J.* **163** 267–308. MR3161315

KELLEY, W. G. and PETERSON, A. C. (2010). *The Theory of Differential Equations*: *Classical and Qualitative*. Springer, Berlin.

LAWLER, G. F., BRAMSON, M. and GRIFFEATH, D. (1992). Internal diffusion limited aggregation. *Ann. Probab.* **20** 2117–2140. MR1188055

SPITZER, F. (1976). *Principles of Random Walk*, 2nd ed. *Graduate Texts in Mathematics* **34**. Springer, New York. MR0388547

E. CANDELLERO
DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
CV4 7AL, COVENTRY
UNITED KINGDOM
E-MAIL: E.Candellero@warwick.ac.uk

S. GANGULY
DEPARTMENT OF STATISTICS,
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: sganguly@berkeley.edu

C. HOFFMAN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
USA
E-MAIL: hoffman@math.washington.edu

L. LEVINE
DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853
USA
E-MAIL: levine@math.cornell.edu