Adv. Appl. Probab. **56,** 354–388 (2024)
*doi:*10.1017/apr.2023.24

INTERLACEMENT LIMIT OF A STOPPED RANDOM WALK TRACE ON A TORUS

ANTAL A. JÁRAI[*] [**] AND
MINWEI SUN,[*] [***] *University of Bath*

Abstract

We consider a simple random walk on $\mathbb{Z}^d$ started at the origin and stopped on its first exit time from $(-L, L)^d \cap \mathbb{Z}^d$. Write L in the form $L = mN$ with $m = m(N)$ and N an integer going to infinity in such a way that $L^2 \sim AN^d$ for some real constant $A > 0$. Our main result is that for $d \geq 3$, the projection of the stopped trajectory to the N-torus locally converges, away from the origin, to an interlacement process at level $Ad\sigma_1$, where σ_1 is the exit time of a Brownian motion from the unit cube $(-1, 1)^d$ that is independent of the interlacement process. The above problem is a variation on results of Windisch (2008) and Sznitman (2009).

Keywords: Interlacement; hitting probability; hashing; loop-erased random walk

2020 Mathematics Subject Classification: Primary 60K35

1. Introduction

A special case of a result of Windisch [15]—extended further in [1]—states that the trace of a simple random walk on the discrete d-dimensional torus $(\mathbb{Z}/N\mathbb{Z})^d$, for $d \geq 3$, started from stationarity and run for time uN^d converges, in a local sense, to an interlacement process at level u, as $N \to \infty$. In this paper we will be concerned with a variation on this result, for which our motivation was a heuristic analysis of an algorithm we used to simulate high-dimensional loop-erased random walks and the sandpile height distribution [7]. Let us first describe our main result and then discuss the motivating problem.

Consider a discrete-time lazy simple random walk $(Y_t)_{t \geq 0}$ starting at the origin o on $\mathbb{Z}^d$. We write $\mathbf{P}_o$ for the probability measure governing this walk. We stop the walk at the first time T_L when it exits the large box $(-L, L)^d$, where L is an integer. We will take $L = L(N)$ of the form $L = mN$, where $m = m(N)$ and N is an integer, such that $L^2 \sim AN^d$ for some $A \in (0, \infty)$, as $N \to \infty$. We consider the projection of the trajectory $\{Y_t : 0 \leq t < T_L\}$ to the N-torus $\mathbb{T}_N = [-N/2, N/2)^d \cap \mathbb{Z}^d$. The projection is given by the map $\varphi_N : \mathbb{Z}^d \to \mathbb{T}_N$, where for any $x \in \mathbb{Z}^d$, $\varphi_N(x)$ is the unique point of $\mathbb{T}_N$ such that $\varphi_N(x) \equiv x \pmod{N}$, where congruence (mod N) is understood coordinate-wise.

Let σ_1 denote the exit time from $(-1, 1)^d$ of a standard Brownian motion started at o. We write $\mathbb{E}$ for the expectation associated to this Brownian motion. For any finite set $K \subset \mathbb{Z}^d$,

Received 19 August 2021; accepted 26 April 2023.
[*] Postal address: Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, UK.
[**] Email address: A.Jarai@bath.ac.uk
[***] Email address: M.Sun@bath.ac.uk

let Cap(K) denote the capacity of K [9]. For any $0 < R < \infty$ and $x \in \mathbb{Z}^d$, we define $B_R(x) = \{y \in \mathbb{Z}^d : |y - x| < R\}$, where $|\cdot|$ is the Euclidean norm. Let $\mathcal{K}_R$ denote the collection of all subsets of $B_R(o)$. Given $\mathbf{x} \in \mathbb{T}_N$, let $\tau_{\mathbf{x},N} : \mathbb{T}_N \to \mathbb{T}_N$ denote the translation of the torus by $\mathbf{x}$. Let $g : \mathbb{N} \to (0, \infty)$ be any function satisfying $g(N) \to \infty$.

Theorem 1.1. *Let $d \geq 3$. For any $0 < R < \infty$, any $K \in \mathcal{K}_R$, and any $\mathbf{x}$ satisfying $\tau_{\mathbf{x},N}\varphi_N(B_R(o)) \cap \varphi_N(B_{g(N)}(o)) = \emptyset$ we have*

$$\mathbf{P}_o\left[\varphi_N(Y_t) \notin \tau_{\mathbf{x},N}\varphi_N(K),\ 0 \leq t < T_L\right] = \mathbb{E}\left[e^{-dA\sigma_1 \mathrm{Cap}(K)}\right] + o(1) \quad \text{as } N \to \infty. \tag{1}$$

The error term depends on R and g, but is uniform in K and $\mathbf{x}$.

Note that the trace of the lazy simple random walk stopped at time T is the same as the trace of the simple random walk stopped at the analogous exit time. We use the lazy walk for convenience of the proof.

Our result is close in spirit—although the details differ—to a result of Sznitman [12] that is concerned with a simple random walk on a discrete cylinder. The *interlacement process* was introduced by Sznitman in [13]. It consists of a one-parameter family $(\mathcal{I}^u)_{u>0}$ of random subsets of $\mathbb{Z}^d$ ($d \geq 3$), where the distribution of $\mathcal{I}^u$ can be characterized by the relation

$$\mathbf{P}[\mathcal{I}^u \cap K = \emptyset] = \exp\left(-u\mathrm{Cap}(K)\right) \quad \text{for any finite } \emptyset \neq K \subset \mathbb{Z}^d. \tag{2}$$

The precise construction of a process satisfying (2) represents $\mathcal{I}^u$ as the trace of a Poisson cloud of bi-infinite random walk trajectories (up to time-shifts), where u is an intensity parameter. We refer to [13] and the books [3, 14] for further details. Comparing (1) and (2), we now formulate precisely what we mean by saying that the stopped trajectory, locally, is described by an interlacement process at the random level $u = Ad\sigma_1$.

Let $g' : \mathbb{N} \to (0, \infty)$ be any function satisfying $g'(N) \to \infty$. Note this does not have to be the same function as $g(N)$. Let $\mathbf{x}_N$ be an arbitrary sequence satisfying $\tau_{\mathbf{x}_N,N}\varphi_N(B_{g'(N)}(o)) \cap \varphi_N(B_{g(N)}(o)) = \emptyset$. Define the sequence of random configurations $\omega_N \subset \mathbb{Z}^d$ by

$$\omega_N = \left\{x \in \mathbb{Z}^d : \tau_{\mathbf{x}_N,N}\varphi_N(x) \in \left\{\varphi_N(Y_t) : 0 \leq t < T_L\right\}\right\}.$$

Define the process $\tilde{\mathcal{I}}$ by requiring that for all finite $K \subset \mathbb{Z}^d$ we have

$$\mathbf{P}\big[\tilde{\mathcal{I}} \cap K = \emptyset\big] = \mathbb{E}\big[e^{-dA\sigma_1 \mathrm{Cap}(K)}\big].$$

To see that this formula indeed defines a process that is also unique, write the right-hand side as

$$\int_0^\infty e^{-u\mathrm{Cap}(K)} f_{\sigma_1}(u)\, du = \int_0^\infty \mathbf{P}\big[\mathcal{I}^u \cap K = \emptyset\big] f_{\sigma_1}(u)\, du,$$

where f_{σ_1} is the density of $Ad\sigma_1$. Then via the inclusion–exclusion formula, we see that we necessarily have for all finite sets $B \subset K$ the equality

$$\mathbf{P}[\tilde{\mathcal{I}} \cap K = B] = \int_0^\infty \mathbf{P}\big[\mathcal{I}^u \cap K = B\big] f_{\sigma_1}(u)\, du,$$

and the right-hand side can be used as the definition of the finite-dimensional marginals of $\tilde{\mathcal{I}}$. Note that $\tilde{\mathcal{I}}$ lives in a compact space (the space can be identified with $\{0, 1\}^{\mathbb{Z}^d}$ with the product topology). Hence the finite-dimensional marginals uniquely determine the distribution of $\tilde{\mathcal{I}}$, by Kolmogorov's extension theorem.

Theorem 1.2. *Let $d \geq 3$. Under $\mathbf{P}_o$, the law of the configuration ω_N converges weakly to the law of $\tilde{\mathcal{I}}$, as $N \to \infty$.*

Proof of Theorem 1.2 assuming Theorem 1.1. For events of the form $\{\omega_N \cap K = \emptyset\}$, Theorem 1 immediately implies that

$$\mathbf{P}_o[\omega_N \cap K = \emptyset] \stackrel{N\to\infty}{\longrightarrow} \mathbf{P}\big[\tilde{\mathcal{I}} \cap K = \emptyset\big].$$

For events of the form $\{\omega_N \cap K = B\}$, the inclusion–exclusion formula represents $\mathbf{P}_o[\omega_N \cap K = B]$ as a linear combination of probabilities of the former kind, and hence convergence follows. □

Our motivation for studying the question in Theorem 1.1 was a simulation problem that arose in our numerical study of high-dimensional sandpiles [7]. We refer the interested reader to [2, 6, 11] for background on sandpiles. In our simulations we needed to generate loop-erased random walks (LERWs) from the origin o to the boundary of $[-L, L]^d$, where $d \geq 5$. The LERW is defined by running a simple random walk from o until it hits the boundary, and erasing all loops from its trajectory chronologically, as they are created. We refer to the book [9] for further background on LERWs (which is not needed to understand the results in this paper). It is known from results of Lawler [8] that in dimensions $d \geq 5$ the LERW visits on the order of L^2 vertices, the same as the simple random walk generating it. As the number of vertices visited is much smaller than the volume cL^d of the box, an efficient way to store the path generating the LERW is provided by the well-known method of *hashing*. We refer to [7] for a discussion of this approach, and only provide a brief summary here. Assign to any $x \in [-L, L]^d \cap \mathbb{Z}^d$ an integer value $f(x) \in \{0, 1, \ldots, CL^2\}$ that is used to label the information relevant to position x, where C can be a large constant or slowly growing to infinity. Thus f is necessarily highly non-injective. However, we may be able to arrange that with high probability the restriction of f to the simple random walk trajectory is not far from injective, and then memory use can be reduced from order L^d to roughly $O(L^2)$.

A simple possible choice of the hash function f can be to compose the map $\varphi_N : [-L, L]^d \cap \mathbb{Z}^d \to \mathbb{T}_N$ with a linear enumeration of the vertices of $\mathbb{T}_N$, whose range has the required size. (This is slightly different from what was used in [7].) The method can be expected to be effective, if the projection $\varphi_N(Y[0, T))$ spreads roughly evenly over the torus $\mathbb{T}_N$ with high probability. Our main theorem establishes a version of such a statement, as the right-hand-side expression in (1) is independent of $\mathbf{x}$.

We now make some comments on the proof of Theorem 1.1. We refer to [3, Theorem 3.1] for the strategy of the proof in the case when the walk is run for a fixed time uN^d. The argument presented there proceeds by decomposing the walk into stretches of length $\lfloor N^\delta \rfloor$ for some $2 < \delta < d$, and then estimating the (small) probability in each stretch that $\tau_{\mathbf{x},N}\varphi_N(K)$ is hit by the projection. We follow the same outline for the stopped lazy random walk. However, the elegant time-reversal argument given in [3] is not convenient in our setting, and we need to prove a delicate estimate on the probability that $\tau_{\mathbf{x},N}\varphi_N(K)$ is hit, *conditional* on the starting point and endpoint of the stretch. For this, we only want to consider stretches with 'well-behaved' starting points and endpoints. We also classify a stretch as a 'good stretch' if the total displacement is not too large, and as a 'bad stretch' otherwise. We do this in such a way that the expected number of 'bad stretches' is small, and summing over the 'good stretches' gives us the required behaviour.

Possible generalizations.

(1) It is not essential that we restrict to the simple random walk: any random walk for which the results in Section 2 hold (such as finite-range symmetric walks) would work equally well.

(2) The paper [15] considers several distant sets $K^1, \ldots, K^r$, and we believe this would also be possible here, but would lead to further technicalities in the presentation.

(3) It is also not essential that the rescaled domain be $(-1, 1)^d$, and we believe it could be replaced by any other domain with sufficient regularity of the boundary.

A note on constants. All constants will be positive and finite. Constants denoted by C or c will depend only on the dimension d and may change from line to line. If we need to refer to a constant later, it will be given an index, such as C_1.

We now describe the organization of this paper. In Section 2, we first introduce some basic notation, then recall several useful known results on random walks and state the key propositions required for the proof of the main theorem, Theorem 1.1. Section 3 contains the proof of the main theorem, assuming the key propositions. Finally, in Section 4 we provide the proofs of the propositions stated in Section 2.

2. Preliminaries

2.1. Some notation

We first introduce some notation used in this paper. In Section 1 we defined the discrete torus $\mathbb{T}_N = [-N/2, N/2)^d \cap \mathbb{Z}^d$, $d \geq 3$, and the canonical projection map $\varphi_N : \mathbb{Z}^d \to \mathbb{T}_N$. From here on, we will omit the N-dependence and write φ and $\tau_{\mathbf{x}}$ instead.

We write vertices and subsets of the torus in bold, i.e. $\mathbf{x} \in \mathbb{T}_N$ and $\mathbf{K} \subset \mathbb{T}_N$. In order to simplify notation, in the rest of the paper we abbreviate $\mathbf{K} = \tau_{\mathbf{x}}\varphi(K)$.

Let $(Y_t)_{t\geq 0}$ be a discrete-time lazy simple random walk on $\mathbb{Z}^d$; that is,

$$\mathbf{P}\big[Y_{t+1} = y' \mid Y_t = x'\big] = \begin{cases} \frac{1}{2} & \text{when } y' = x'; \\ \frac{1}{4d} & \text{when } |y' - x'| = 1. \end{cases}$$

We denote the corresponding lazy random walk on $\mathbb{T}_N$ by $(\mathbf{Y}_t)_{t\geq 0} = (\varphi(Y_t))_{t\geq 0}$. Let $\mathbf{P}_{x'}$ denote the distribution of the lazy random walk on $\mathbb{Z}^d$ started from $x' \in \mathbb{Z}^d$, and write $\mathbf{P}_{\mathbf{x}}$ for the distribution of the lazy random walk on $\mathbb{T}_N$ started from $\mathbf{x} = \varphi(x') \in \mathbb{T}_N$. We write $p_t(x', y') = \mathbf{P}_{x'}[Y_t = y']$ for the t-step transition probability. Further notation we will use includes the following:

- $L = mN$, where $L^2 \sim AN^d$ as $N \to \infty$ for some constant $A \in (0, \infty)$;
- $D = (-m, m)^d$, the rescaled box, indicates which copy of the torus the walk is in;
- $n = \lfloor N^\delta \rfloor$ for some $2 < \delta < d$, long enough for the mixing property on the torus, but short compared to L^2;
- $\mathbf{x_0} \in \mathbf{K}$ is a fixed point of $\mathbf{K}$;
- we write points in the original lattice $\mathbb{Z}^d$ with a prime, such as y', and decompose a point y' as $yN + \mathbf{y}$ with y in another lattice isomorphic to $\mathbb{Z}^d$ and $\mathbf{y} = \varphi(y') \in \mathbb{T}_N$;

- $T = \inf\left\{t \geq 0 : Y_t \notin (-L, L)^d\right\}$, the first exit time from $(-L, L)^d$;
- $S = \inf\left\{\ell \geq 0 : Y_{n\ell} \notin (-L, L)^d\right\}$, so that the first multiple of n when the rescaled point $Y_{n\ell}/N$ is not in $(-m, m)^d$ equals $S \cdot n$.

For simplicity, we omit the dependence on d and N from some of the notation above.

2.2. Some auxiliary results on random walks

In this section, we collect some known results required for the proof of Theorem 1.1. We will rely heavily on the local central limit theorem (LCLT) [9, Chapter 2], with error term, and the martingale maximal inequality [9, Equation (12.12) of Corollary 12.2.7]. We will also use [9, Equation (6.31)], which relates Cap(K) to the probability that a random walk started from the boundary of a large ball with radius n hits the set K before exiting the ball. In estimating some error terms in our arguments, sometimes we will use the Gaussian upper and lower bounds [5]. We also need to derive a lemma related to the mixing property on the torus [10, Theorem 5.6] to show that the starting positions of different stretches are not far from uniform on the torus; see Lemma 2.1.

We recall the LCLT from [9, Chapter 2]. The following is a specialization of [9, Theorem 2.3.11] to lazy simple random walks. The covariance matrix Γ and the square root $J^*(x)$ of the associated quadratic form are given by

$$\Gamma = (2d)^{-1} I, \qquad J^*(x) = (2d)^{\frac{1}{2}} |x|,$$

where I is the $(d \times d)$-unit matrix.

Let $\bar{p}_t(x')$ denote the estimate of $p_t(x')$ that one obtains by the LCLT, for a lazy simple random walk. We have

$$\begin{aligned}
\bar{p}_t(x') &= \frac{1}{(2\pi t)^{d/2}\sqrt{\det \Gamma}} \exp\left(-\frac{J^*(x')^2}{2t}\right) \\
&= \frac{1}{(2\pi t)^{d/2}(2d)^{-d/2}} \exp\left(-\frac{2d\,|x'|^2}{2t}\right) \\
&= \frac{\bar{C}}{t^{d/2}} \exp\left(-\frac{d\,|x'|^2}{t}\right).
\end{aligned}$$

The lazy simple random walk $(Y_t)_{t\geq 0}$ in $\mathbb{Z}^d$ is aperiodic and irreducible with mean zero, finite second moment, and finite exponential moments. All joint third moments of the components of Y_1 vanish.

Theorem 2.1. ([9, Theorem 2.3.11].) *For a lazy simple random walk $(Y_t)_{t\geq 0}$ in $\mathbb{Z}^d$, there exists $\rho > 0$ such that for all $t \geq 1$ and all $x' \in \mathbb{Z}^d$ with $|x'| < \rho t$,*

$$p_t(x') = \bar{p}_t(x') \exp\left\{O\left(\frac{1}{t} + \frac{|x'|^4}{t^3}\right)\right\}.$$

The martingale maximal inequality in [9, Equation (12.12) of Corollary 12.2.7] is stated as follows. Let $\left(Y_t^{(i)}\right)_{t\geq 0}$ denote the ith coordinate of $(Y_t)_{t\geq 0}$ $(1 \leq i \leq d)$. The standard deviation σ of $Y_1^{(i)}$ is given by $\sigma^2 = (2d)^{-1}$. For all $t \geq 1$ and all $r > 0$ we have

$$\mathbf{P}_o\left[\max_{0\leq j\leq t} Y_j^{(i)} \geq r\sigma\sqrt{t}\right] \leq e^{-r^2/2} \exp\left\{O\left(\frac{r^3}{\sqrt{t}}\right)\right\}. \tag{3}$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

Now we state the result of [9, Equation (6.31)]. Recall that $B_r(o)$ is the discrete ball centred at o with radius r. For any subset $B \subset \mathbb{Z}^d$, let

$$\xi_B = \inf\left\{t \geq 1 : Y_t \notin B)\right\}.$$

Let $\partial B_r(o) = \left\{y' \in \mathbb{Z}^d \setminus B_r(o) : \exists x' \in B_r(o) \text{ such that } |x' - y'| = 1\right\}$. For a given finite set $K \subseteq \mathbb{Z}^d$, let H_K denote the hitting time

$$H_K = \inf\left\{t \geq 1 : Y_t \in K\right\}.$$

Then we have

$$\frac{1}{2}\mathrm{Cap}(K) = \lim_{r\to\infty} \sum_{y' \in \partial B_r(o)} \mathbf{P}_{y'}\left[H_K < \xi_{B_r(o)}\right]. \tag{4}$$

Here Cap(K) is the capacity of K; see [9, Section 6.5], which states the analogous statement for the simple random walk. Since we consider the lazy random walk, this introduces a factor of $1/2$.

In estimating some error terms in our arguments, sometimes we will use the Gaussian upper and lower bounds [5]: there exist constants $C = C(d)$ and $c = c(d)$ such that

$$\begin{aligned} p_t\left(x', y'\right) &\leq \frac{C}{t^{d/2}} \exp\left(-c\frac{|y' - x'|^2}{t}\right), \quad \text{for } x', y' \in \mathbb{Z}^d \text{ and } t \geq 1; \\ p_t\left(x', y'\right) &\geq \frac{c}{t^{d/2}} \exp\left(-C\frac{|y' - x'|^2}{t}\right), \quad \text{for } |y' - x'| \leq ct. \end{aligned} \tag{5}$$

Recall that the norm $|\cdot|$ refers to the Euclidean norm.

Regarding mixing times, recall that the lazy simple random walk on the N-torus mixes in time N^2 [10, Theorem 5.6]. With this in mind we derive the following simple lemma.

Recall that $2 < \delta < d$ and $n = \lfloor N^\delta \rfloor$.

Lemma 2.1. *There exists $C = C(d)$ such that for any $N \geq 1$ and any $t \geq n$ we have*

$$\mathbf{P_o}[\mathbf{Y}_t = \mathbf{x}] \leq \frac{C}{N^d}, \quad \mathbf{x} \in \mathbb{T}_N.$$

Proof. Using the Gaussian upper bound, the left-hand side can be bounded by

$$\begin{aligned} \sum_{x\in\mathbb{Z}^d} p_t(o, xN + \mathbf{x}) &\leq \frac{C}{t^{d/2}} \sum_{x\in\mathbb{Z}^d} \exp\left(-c\frac{|xN + \mathbf{x}|^2}{t}\right) \leq \frac{C}{t^{d/2}} \sum_{x\in\mathbb{Z}^d} \exp\left(-c\frac{|xN|^2}{t}\right) \\ &\leq \frac{C}{t^{d/2}} \sum_{k=0}^{\infty} (k+1)^{d-1} \frac{t^{d/2}}{N^d} \exp\left(-ck^2\right) \\ &\leq \frac{C}{N^d} \sum_{k=0}^{\infty} (k+1)^{d-1} \exp\left(-ck^2\right) \leq \frac{C}{N^d}. \end{aligned}$$

Here we bounded the number of x in $\mathbb{Z}^d$ satisfying $k\sqrt{t}/N \leq |x| < (k+1)\sqrt{t}/N$ by $C(k+1)^{d-1}t^{d/2}/N^d$, where $k = 0, 1, 2, \ldots$. □

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

2.3. Key propositions

In this section we state some propositions to be used in Section 3 to prove Theorem 1.1. The propositions will be proved in Section 4.

The strategy of the proof is to consider stretches of length n of the walk, and estimate the small probability in each stretch that $\mathbf{K}$ is hit by the projection. What makes this strategy work is that we can estimate, conditionally on the starting point and endpoint of a stretch, the probability that $\mathbf{K}$ is hit, and this event asymptotically decouples from the number of stretches. The number of stretches will be the random variable S. Since $nS \approx T$, and T is $\Theta\big(N^d\big)$ in probability, we have that S is $\Theta\big(N^d/n\big)$. In Lemma 2.2 below we show a somewhat weaker estimate for S (which suffices for our needs).

The main part of the proof will be to show that during a fixed stretch, $\mathbf{K}$ is not hit with probability

$$1 - \frac{1}{2}\mathrm{Cap}(\mathbf{K})\frac{n}{N^d}(1 + o(1)). \tag{6}$$

Heuristically, conditionally on S this results in the probability

$$\left(1 - \frac{1}{2}\mathrm{Cap}(\mathbf{K})\frac{n}{N^d}(1 + o(1))\right)^S \approx \exp\left(-\frac{1}{2}\mathrm{Cap}(\mathbf{K})\frac{n}{N^d}S\right),$$

and we will conclude by showing that $\big(n/N^d\big)\, S$ converges in distribution to a constant multiple of the Brownian exit time σ_1.

The factor n/N^d in (6) arises as the expected time spent by the projected walk at a fixed point of the torus during a given stretch. The capacity term arises as we pass from expected time to hitting probability.

For the above approach to work, we need a small-probability event on which the number of stretches or endpoints of stretches are not sufficiently well-behaved. First, we will need to restrict to realizations where $\big(\sqrt{\log\log n}\big)^{-1}\big(N^d/n\big) \le S \le \log N\, \big(N^d/n\big)$, which occurs with high probability as $N \to \infty$ (see Lemma 2.2 below). Second, suppose that the ℓth stretch starts at the point $y'_{\ell-1}$ and ends at the point y'_ℓ; that is, $y'_{\ell-1}$ and y'_ℓ are realizations of $Y_{(\ell-1)n}$ and $Y_{\ell n}$. In order to have a good estimate of the probability that $\mathbf{K}$ is hit during this stretch, we will need to impose a few conditions on $y'_{\ell-1}$ and y'_ℓ. One of these is that the displacement $|y'_\ell - y'_{\ell-1}|$ is not too large: we will require that for all stretches, it is at most $f(n)\sqrt{n}$, for a function to be chosen later that increases to infinity with N. We will be able to choose $f(n)$ of the form $f(n) = C_1\sqrt{\log N}$ in such a way that this restriction holds for all stretches with high probability. A third condition we need to impose, that will also hold with high probability, is that $y'_{\ell-1}$ is at least a certain distance N^ζ from $\varphi^{-1}(\mathbf{K})$ for a parameter $0 < \zeta < 1$ (this will only be required for $\ell \ge 1$, and is not needed for the first stretch starting with $y'_0 = o$). The reason we need this is to be able to appeal to (4) to extract the $\mathrm{Cap}(\mathbf{K})$ contribution, when we know that $\mathbf{K}$ is hit from a long distance (we will take $r = N^\zeta$ in (4)). The larger the value of ζ, the better error bound we get on the approach to $\mathrm{Cap}(\mathbf{K})$. On the other hand, ζ should not be too close to 1, because we want the separation of $y'_{\ell-1}$ from $\mathbf{K}$ to occur with high enough probability.

The set $\mathcal{G}_{\zeta,C_1}$ defined below represents realizations of S and the sequence $Y_n, Y_{2n}, \ldots, Y_{Sn}$ satisfying the above restrictions. Proposition 2.1 below implies that these restrictions hold with high probability. First, we will need $2 < \delta < d$ to satisfy the inequality

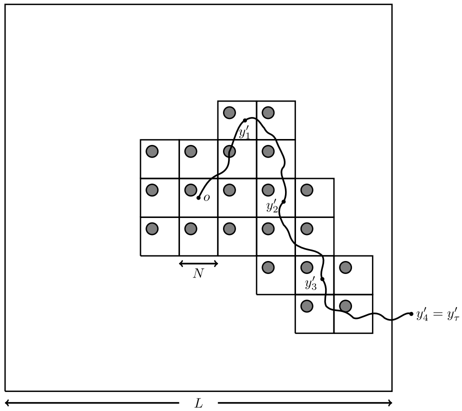

FIGURE 1. This figure explains the properties of the set $\mathcal{G}_{\zeta,C_1}$ (not to scale). The shaded regions represent the balls of radius N^ζ in each copy of the torus. None of the y'_ℓ, for $\ell \geq 1$, is in a shaded region.

$$\delta - \frac{2\delta}{d} > d - \delta \qquad \Leftrightarrow \qquad 2\delta > \frac{d^2}{d-1}. \tag{7}$$

This can be satisfied if $d \geq 3$ and δ is sufficiently close to d, say $\delta = \frac{7}{8}d$. Since the left-hand side of the left-hand inequality in (7) equals $(\delta/d)(d-2)$, we can subsequently choose ζ such that we also have

$$0 < \zeta < \frac{\delta}{d}, \qquad \zeta(d-2) > d - \delta. \tag{8}$$

With the parameter ζ fixed satisfying the above, we now define

$$\mathcal{G}_{\zeta,C_1} = \left\{ \left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^{\tau}\right) : \begin{array}{l} \left(\sqrt{\log\log n}\right)^{-1} \frac{N^d}{n} \leq \tau \leq \log N \frac{N^d}{n}; \\ y_\ell \in D,\, \mathbf{y}_\ell \in \mathbb{T}_N \setminus B\left(\mathbf{x_0}, N^\zeta\right) \text{ for } 1 \leq \ell < \tau; \\ y_\tau \in D^c \text{ and } \mathbf{y}_\tau \in \mathbb{T}_N \setminus B\left(\mathbf{x_0}, N^\zeta\right); \\ |y'_\ell - y'_{\ell-1}| \leq f(n) n^{\frac{1}{2}} \text{ for } 1 \leq \ell \leq \tau \end{array} \right\}, \tag{9}$$

where $f(n) = C_1\sqrt{\log N}$ and recall that we write $y'_\ell = y_\ell N + \mathbf{y}_\ell$ and $y'_{\ell-1} = y_{\ell-1}N + \mathbf{y}_{\ell-1}$, and we define $y'_0 = o$. The time τ is corresponding to a particular value of the exit time S, so $y_\ell \in D$ for $1 \leq \ell < \tau$ and $y_\tau \notin D$. The parameter C_1 will be chosen in the course of the proof. See Figure 1 for a visual illustration of the sequence $y'_0, y'_1, \ldots$ in the definition of $\mathcal{G}_{\zeta,C_1}$.

The next lemma shows that the restriction made on the time-parameter τ in the definition of $\mathcal{G}_{\zeta,C_1}$ holds with high probability for S.

Lemma 2.2. *We have*

$$\mathbf{P}_o\left[\left\{\frac{Sn}{N^d} < \left(\sqrt{\log\log n}\right)^{-1}\right\} \cup \left\{\frac{Sn}{N^d} > \log N\right\}\right] \to 0$$

as $N \to \infty$.

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

Proof. By the definitions of S and T, we first notice that

$$\mathbf{P}_o\left[S<\left(\sqrt{\log\log n}\right)^{-1}\frac{N^d}{n}\right]\le\mathbf{P}_o\left[T<\left(\sqrt{\log\log n}\right)^{-1}N^d\right]$$
$$\le\sum_{1\le i\le d}\left(\mathbf{P}_o\left[\max_{0\le j\le(\sqrt{\log\log n})^{-1}N^d}Y_j^{(i)}\ge L\right]\right.$$
$$\left.+\mathbf{P}_o\left[\max_{0\le j\le(\sqrt{\log\log n})^{-1}N^d}-Y_j^{(i)}\ge L\right]\right),$$

where $Y^{(i)}$ denotes the ith coordinate of the d-dimensional lazy random walk.

We are going to use (3). Setting $t=\left(\sqrt{\log\log n}\right)^{-1}N^d$ and $r\sigma\sqrt{t}=L$, we can evaluate each term $\left(\text{similarly for the event with } -Y_j^{(i)}\right)$ in the sum

$$\mathbf{P}_o\left[\max_{0\le j\le(\sqrt{\log\log n})^{-1}N^d}Y_j^{(i)}\ge L\right]$$
$$\le\exp\left\{-\frac{1}{2}\frac{L^2}{\sigma^2\left(\sqrt{\log\log n}\right)^{-1}N^d}+O\left(\frac{L^3}{\sigma^3\left(\sqrt{\log\log n}\right)^{-2}N^{2d}}\right)\right\}. \tag{10}$$

Recall that $L^2\sim AN^d$ and $\sigma^2=1/2d$. For the main term in the exponential in (10), we have the upper bound

$$\exp\left(-(1+o(1))\frac{1}{2}\frac{2d\cdot AN^d}{\left(\sqrt{\log\log n}\right)^{-1}N^d}\right)=\exp\left(-(1+o(1))Ad\sqrt{\log\log n}\right)$$
$$\to 0,\qquad \text{as } N\to\infty.$$

The big-O term in the exponential in (10) produces an error term because

$$\exp\left\{O\left(\frac{(AN^d)^{3/2}}{\sigma^3\left(\sqrt{\log\log n}\right)^{-2}N^{2d}}\right)\right\}=\exp\left\{O\left(N^{-d/2}(\log\log n)\right)\right\}$$
$$=1+o(1),\qquad \text{as } N\to\infty.$$

Coming to the second event $\left\{\frac{S_n}{N^d}>\log N\right\}$, observe that the central limit theorem applied to $\left(Y_{kn\lfloor N^d/n\rfloor)+t}-Y_{kn\lfloor N^d/n\rfloor}\right)_{t\ge 0}$ implies that

$$\mathbf{P}_o\left[S>(k+1)\left\lfloor\frac{N^d}{n}\right\rfloor\,\middle|\,S>k\left\lfloor\frac{N^d}{n}\right\rfloor\right]\le\max_{z\in(-L,L)^d}\left(1-\mathbf{P}_z\left[Y_{n\lfloor N^d/n\rfloor}\notin(-2L,2L)^d\right]\right)\le 1-c$$

for some $c>0$. Hence we have $\mathbf{P}_o\left[S>k\frac{N^d}{n}\right]\le e^{-ck}$ for all $k\ge 0$.

Applying this with $k=\log N$, we obtain

$$\mathbf{P}_o\left[S>\log N\frac{N^d}{n}\right]\le e^{-c\log N}\to 0$$

as required. □

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

The starting point for the proof of Theorem 1.1 is the following proposition, which decomposes the probability we are interested in into terms involving single stretches of duration n.

Proposition 2.1. *For a sufficiently large value of C_1, we have that*

$$\mathbf{P}_o\left[\left(S,\, (Y_{\ell n},\, \varphi(Y_{\ell n})_{\ell=1}^{S}\right) \notin \mathcal{G}_{\zeta,C_1}\right] = o(1) \quad \text{as } N \to \infty. \tag{11}$$

Furthermore,

$$\begin{aligned}&\mathbf{P}_o\left[Y_t \notin \varphi^{-1}(\mathbf{K}),\ 0 \le t < T\right]\\ &\quad = \sum_{(\tau,(y_\ell,\mathbf{y}_\ell)_{\ell=1}^{\tau})\in\mathcal{G}_{\zeta,C_1}} \prod_{\ell=1}^{\tau} \mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell,\ Y_t \notin \varphi^{-1}(\mathbf{K})\, \textit{for}\ 0 \le t < n\right] + o(1),\end{aligned} \tag{12}$$

where $o(1) \to 0$ as $N \to \infty$.

Central to the proof of Theorem 1.1 is the following proposition, which estimates the probability of hitting a copy of $\mathbf{K}$ during a 'good stretch' where the displacement $|y'_\ell - y'_{\ell-1}|$ is almost of order $\sqrt{n}$. This will not hold for all stretches with high probability, but the fraction of stretches for which it fails will vanish asymptotically.

Proposition 2.2. *There exists a sufficiently large value of C_1 so that the following holds. Let $\left(\tau,\, (y_\ell,\, \mathbf{y}_\ell)_{\ell=1}^{\tau}\right) \in \mathcal{G}_{\zeta,C_1}$. Then for all $2 \le \ell \le \tau$ such that $|y'_\ell - y'_{\ell-1}| \le 10\sqrt{n}\log\log n$ we have*

$$\begin{aligned}&\mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell,\ Y_t \notin \varphi^{-1}(\mathbf{K})\, \textit{for}\ 0 \le t < n\right]\\ &\qquad = \mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell\right]\left(1 - \frac{1}{2}\frac{\mathrm{Cap}(\mathbf{K})\, n}{N^d}\,(1 + o(1))\right).\end{aligned} \tag{13}$$

In addition to the above proposition (that we prove in Section 4.2), we will need a weaker version for the remaining 'bad stretches' that have less restriction on the distance $|y'_\ell - y'_{\ell-1}|$. This will be needed to estimate error terms arising from the 'bad stretches', and it will also be useful in demonstrating some of our proof ideas for other error terms arising later in the paper. It will be proved in Section 4.1.

Proposition 2.3. *Let $\left(\tau,\, (y_\ell,\, \mathbf{y}_\ell)_{\ell=1}^{\tau}\right) \in \mathcal{G}_{\zeta,C_1}$. For all $2 \le \ell \le \tau$ we have*

$$\mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell,\ Y_t \notin \varphi^{-1}(\mathbf{K})\, \textit{for all}\ 0 \le t < n\right] = \mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell\right]\left(1 - O\left(\frac{n}{N^d}\right)\right), \tag{14}$$

and for the first stretch we have

$$\mathbf{P}_o\left[Y_n = y'_1,\ Y_t \notin \varphi^{-1}(\mathbf{K})\, \textit{for all}\ 0 \le t < n\right] = \mathbf{P}_o\left[Y_n = y'_1\right]\,(1 - o(1)). \tag{15}$$

Here the $-O\big(n/N^d\big)$ and $-o(1)$ error terms are negative.

Our final proposition is needed to estimate the number of stretches that are 'bad'.

Proposition 2.4. *We have*

$$\begin{aligned}&\mathbf{P}_o\left[\#\left\{1 \le \ell \le \frac{N^d}{n}C_1\log N : |Y_{n\ell} - Y_{n(\ell-1)}| > 10\sqrt{n}\log\log n\right\} \ge \frac{N^d}{n}\frac{1}{\log\log n}\right]\\ &\quad \to 0,\end{aligned} \tag{16}$$

as $N \to \infty$.

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

3. Proof of the main theorem assuming the key propositions

This section gives the proof of Theorem 1.1.

Proof of Theorem 1.1 assuming Propositions 2.1–2.4.

First, given any $\left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^{\tau}\right) \in \mathcal{G}_{\zeta,C_1}$, we define

$$\begin{aligned} \mathcal{L} &= \left\{2 \le \ell \le \tau : \left|y'_\ell - y'_{\ell-1}\right| \le 10\sqrt{n}\log\log n\right\}, \\ \mathcal{L}' &= \left\{2 \le \ell \le \tau : \left|y'_\ell - y'_{\ell-1}\right| > 10\sqrt{n}\log\log n\right\}. \end{aligned} \tag{17}$$

Thus we have

$$\{1, \dots, \tau\} = \{1\} \cup \mathcal{L} \cup \mathcal{L}'.$$

We further define

$$\mathcal{G}'_{\zeta,C_1} = \left\{\left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^{\tau}\right) \in \mathcal{G}_{\zeta,C_1} : |\mathcal{L}'| \le \frac{N^d}{n}\frac{1}{\log\log n}\right\}.$$

We have by Proposition 2.1 that

$$\begin{aligned} &\mathbf{P}_o\left[Y_t \notin \varphi^{-1}(\mathbf{K}),\ 0 \le t < T\right] \\ &= o(1) + \sum_{\left(\tau,(y_\ell,\mathbf{y}_\ell)_{\ell=1}^{\tau}\right)\in\mathcal{G}_{\zeta,C_1}} \prod_{\ell=1}^{\tau} \mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell,\ Y_t \notin \varphi^{-1}(\mathbf{K}) \text{ for } 0 \le t < n\right]. \end{aligned} \tag{18}$$

By Proposition 2.4, we can replace the summation over elements of $\mathcal{G}_{\zeta,C_1}$ by summation over just elements of $\mathcal{G}'_{\zeta,C_1}$, at the cost of an $o(1)$ term. That is,

$$\begin{aligned} &\mathbf{P}_o\left[Y_t \notin \varphi^{-1}(\mathbf{K}),\ 0 \le t < T\right] \\ &= o(1) + \sum_{\left(\tau,(y_\ell,\mathbf{y}_\ell)_{\ell=1}^{\tau}\right)\in\mathcal{G}'_{\zeta,C_1}} \prod_{\ell=1}^{\tau} \mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell,\ Y_t \notin \varphi^{-1}(\mathbf{K}) \text{ for } 0 \le t < n\right]. \end{aligned} \tag{19}$$

Applying Proposition 2.2 for the factors $\ell \in \mathcal{L}$ and Proposition 2.3 for the factors $\ell \in \{1\} \cup \mathcal{L}'$, we get

$$\begin{aligned} &\mathbf{P}_o\left[Y_t \notin \varphi^{-1}(\mathbf{K}),\ 0 \le t < T\right] \\ &= o(1) + \sum_{\left(\tau,(y_\ell,\mathbf{y}_\ell)_{\ell=1}^{\tau}\right)\in\mathcal{G}'_{\zeta,C_1}} \prod_{\ell=1}^{\tau} \mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell\right] \\ &\quad \times (1 - o(1)) \prod_{\ell\in\mathcal{L}} \left(1 - \frac{1}{2}\mathrm{Cap}(\mathbf{K})\frac{n}{N^d}(1 + o(1))\right) \prod_{\ell\in\mathcal{L}'} \left(1 - O\left(\frac{n}{N^d}\right)\right). \end{aligned} \tag{20}$$

Note that since the summation is over elements of $\mathcal{G}'_{\zeta,C_1}$ only, we have

$$|\mathcal{L}'| \le \frac{N^d}{n}\frac{1}{\log\log n}. \tag{21}$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

By (21), we can lower-bound the last product in (20) by

$$\exp\left(-O\left(\frac{n}{N^d}\right)\frac{N^d}{n}\frac{1}{\log\log n}\right)=e^{o(1)}=(1+o(1)).$$

Since the product is also at most 1, it equals $1+o(1)$.

Also, by (21), we have

$$\tau-1-\frac{N^d}{n}\frac{1}{\log\log n}\leq|\mathcal{L}|\leq\tau.$$

Since $\tau\geq\frac{N^d}{n}\big(\sqrt{\log\log n}\big)^{-1}$, we have $|\mathcal{L}|=(1+o(1))\tau$. This implies that the penultimate product in (20) equals

$$\left(1-\frac{1}{2}\mathrm{Cap}(\mathbf{K})\frac{n}{N^d}(1+o(1))\right)^{(1+o(1))\tau}=\exp\left(-\frac{1}{2}\mathrm{Cap}(\mathbf{K})\frac{n}{N^d}\tau(1+o(1))\right).\qquad(22)$$

Recall that $S=\inf\left\{\ell\geq 0{:}Y_{n\ell}\notin(-L,L)^d\right\}$. By summing over $(y_\ell,\mathbf{y}_\ell)_{\ell=1}^{\tau}$ and appealing to (11), we get that (20) equals

$$o(1)+\sum_{\tau}{}'\,\mathbf{E}\left[\mathbf{1}_{S=\tau}\exp\left(-\frac{1}{2}\mathrm{Cap}(\mathbf{K})\frac{n}{N^d}\tau(1+o(1))\right)\right],\qquad(23)$$

where the primed summation denotes restriction to

$$\frac{N^d}{n}\big(\sqrt{\log\log n}\big)^{-1}\leq\tau\leq(\log N)\frac{N^d}{n}.$$

Since, by Lemma 2.2, S satisfies the bounds on τ with probability going to 1, the latter expression equals

$$o(1)+\mathbf{E}\left[e^{-\frac{1}{2}\mathrm{Cap}(\mathbf{K})\frac{n}{N^d}S}\right].\qquad(24)$$

Let Γ_n denote the covariance matrix for Y_n, so that $\Gamma_n=\frac{n}{2d}I$. Let $Z_1=\sqrt{\frac{2d}{n}}\,Y_n$, with the covariance matrix $\Gamma_Z=I$. Let $Z_\ell=\sqrt{\frac{2d}{n}}\,Y_{n\ell}$ for $\ell\geq 0$.

Since $L^2\sim A\,N^d$, the event $\left\{Y_{n\ell}\notin(-L,L)^d\right\}$ is the same as

$$\left\{Y_{n\ell}\notin\left(-(1+o(1))\sqrt{A}N^{d/2},\,(1+o(1))\sqrt{A}N^{d/2}\right)^d\right\}.$$

Converting to events in terms of Z we have

$$Z_\ell\notin\left(-\sqrt{2dA(1+o(1))}\big(N^d/n\big)^{1/2},\,\sqrt{2dA(1+o(1))}\big(N^d/n\big)^{1/2}\right)^d.$$

Now we can write S as

$$S=\inf\left\{\ell\geq 0:Z_\ell\notin\left(-\sqrt{2dA(1+o(1))}\big(N^d/n\big)^{1/2},\,\sqrt{2dA(1+o(1))}\big(N^d/n\big)^{1/2}\right)^d\right\}.$$

Let $\sigma_1=\inf\left\{t>0:B_t\notin(-1,1)^d\right\}$ be the exit time of Brownian motion from $(-1,1)^d$. By Donsker's theorem [4, Theorem 8.1.5] we have

$$\mathbf{P}\left[S\leq 2dA(1+o(1))\frac{N^d}{n}t\right]\to\mathbf{P}[\sigma_1\leq t].$$

Then we have that $\frac{n}{N^d}S$ converges in distribution to $c\sigma_1$, with $c=2dA$. This completes the proof. □

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

4. Proofs of the key propositions

4.1. Proof of Proposition 2.3

In the proof of the proposition we will need the following lemma, which bounds the probability of hitting some copy of **K** in terms of the Green's function of the random walk. Recall that the Green's function is defined by

$$G(x', y') = \sum_{t=0}^{\infty} p_t(x', y'),$$

and in all dimensions $d \geq 3$ it satisfies the bound [9]

$$G(x', y') \leq \frac{C_G}{|y' - x'|^{d-2}}$$

for a constant $C_G = C_G(d)$. For Part (ii) of the lemma recall that $\mathbf{K} \cap \varphi(B_{g(N)}(o)) = \emptyset$. We also define $\mathrm{diam}(\mathbf{K})$ as the maximum Euclidean distance between two points in **K**.

Lemma 4.1. *Let $d \geq 3$. Assume that N is large enough so that $N^{\zeta} \geq \mathrm{diam}(\mathbf{K})$.*

(i) *If $y' \in \mathbb{Z}^d$ satisfies $\varphi(y') \notin B(\mathbf{x}_0, N^{\zeta})$, then for all sufficiently small $\varepsilon > 0$ we have*

$$\sum_{t=0}^{N^{2+6\varepsilon}} \sum_{x \in \mathbb{Z}^d} \sum_{x' \in \mathbf{K}+xN} p_t(y', x') \leq \frac{C}{N^{\zeta(d-2)}}. \tag{25}$$

(ii) *If $g(N) \leq N^{\zeta}$, then for all sufficiently small $\varepsilon > 0$ we have*

$$\sum_{t=0}^{N^{2+6\varepsilon}} \sum_{x \in \mathbb{Z}^d} \sum_{x' \in \mathbf{K}+xN} p_t(o, x') \leq \frac{C}{g(N)^{(d-2)}}. \tag{26}$$

(iii) *If $y' \in \mathbb{Z}^d$ satisfies $\varphi(y') \notin B(\mathbf{x}_0, N^{\zeta})$, then for all sufficiently small $\varepsilon > 0$ we have*

$$\sum_{x \in \mathbb{Z}^d} \sum_{\substack{x' \in \mathbf{K}+xN \\ |x'-y'| \leq n^{\frac{1}{2}+\varepsilon}}} G(y', x') \leq \frac{C}{N^{d-\delta-2\delta\varepsilon}}. \tag{27}$$

Proof. (i) We split the sum according to whether $|y' - x'| > N^{1+4\varepsilon}$ or $\leq N^{1+4\varepsilon}$. In the first case we use (5) and write $r = \lfloor |x' - y'| \rfloor$ to get

$$\begin{aligned} \sum_{t=0}^{N^{2+6\varepsilon}} \sum_{\substack{x' \in \varphi^{-1}(\mathbf{K}) \\ |x'-y'| > N^{1+4\varepsilon}}} p_t(y', x') &\leq \sum_{t=0}^{N^{2+6\varepsilon}} \sum_{r=\lfloor N^{1+4\varepsilon} \rfloor}^{\infty} C\, r^{d-1} \exp\left(-\frac{c\, r^2}{N^{2+6\varepsilon}}\right) \\ &\leq N^{2+6\varepsilon} \sum_{r=\lfloor N^{1+4\varepsilon} \rfloor}^{\infty} C\, r^{d-1} \exp\left(-\frac{c\, r^2}{N^{2+6\varepsilon}}\right) \\ &\leq N^{O(1)} \exp(-cN^{2\varepsilon}). \end{aligned}$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

For the remaining terms, we have the upper bound

$$\sum_{t=0}^{N^{2+6\varepsilon}} \sum_{\substack{x' \in \varphi^{-1}(\mathbf{K}) \\ |x'-y'| \leq N^{1+4\varepsilon}}} p_t(y', x') \leq \sum_{\substack{x' \in \varphi^{-1}(\mathbf{K}) \\ |x'-y'| \leq N^{1+4\varepsilon}}} G(y', x').$$

Let $Q(k\,N)$ be the cube with radius kN centred at o. Then $y' + (Q(kN) \backslash Q((k-1)N))$ are disjoint annuli for $k = 1, 2, \ldots$. We decompose the sum over x' according to which annulus x' falls into. For $k \geq 2$ we have

$$\sum_{\substack{x' \in \varphi^{-1}(\mathbf{K}) \\ x'-y' \in Q(kN) \backslash Q((k-1)N)}} \frac{C_G}{|y'-x'|^{d-2}} \leq |\mathbf{K}| C k^{d-1} C_G (Nk)^{2-d} \leq |\mathbf{K}| C k N^{2-d},$$

where C_G is the Green's function constant. The contribution from any copy of $\mathbf{K}$ in $y' + Q(N)$ will be of order N^{2-d} if its distance from y' is at least $N/3$, say. Note that there is at most one copy of $\mathbf{K}$ within distance $N/3$ of y', which may have a distance as small as N^{ζ}.

We have to sum over the following values of k:

$$k = 1, \ldots, \frac{N^{1+4\varepsilon}}{N} = N^{4\varepsilon}.$$

Since $x' \in \varphi^{-1}(\mathbf{K})$ and $y' \notin \varphi^{-1}\left(B(\mathbf{x}_0, N^{\zeta})\right)$ for $\mathbf{x}_0 \in \mathbf{K}$, the distance between x' and y' is at least N^{ζ}. Therefore, we get the upper bound as follows:

$$\begin{aligned} \sum_{\substack{x' \in \varphi^{-1}(\mathbf{K}) \\ |x'-y'| \leq N^{1+4\varepsilon}}} G(y', x') &\leq |\mathbf{K}| N^{\zeta(2-d)} + \sum_{k=1}^{N^{4\varepsilon}} |\mathbf{K}| C k N^{2-d} \\ &\leq |\mathbf{K}| N^{\zeta(2-d)} + C |\mathbf{K}| N^{2-d} \times N^{8\varepsilon} \leq C |\mathbf{K}| N^{\zeta(2-d)}. \end{aligned}$$

Here the last inequality follows from the choice of ζ, (8), for sufficiently small $\varepsilon > 0$.

(ii) The proof is essentially the same, except for the contribution of the 'nearest' copy of $\mathbf{K}$, which is now $C|\mathbf{K}| g(N)^{2-d}$.

(iii) The proof is very similar to that in Part (i). Recall that $n = \lfloor N^{\delta} \rfloor$. This time we need to sum over $k = 1, \ldots, n^{\frac{1}{2}+\varepsilon}/N$, which results in the bound

$$C|\mathbf{K}| N^{-\zeta(d-2)} + C|\mathbf{K}| N^{2-d} \times N^{\delta + 2\delta\varepsilon - 2} = C|\mathbf{K}| \left[N^{-\zeta(d-2)} + N^{\delta - d + 2\delta\varepsilon} \right].$$

Here, for $\varepsilon > 0$ small enough, the second term dominates thanks to the choice of ζ; see (8). □

Proof of Proposition 2.3. Since

$$\begin{aligned} &\mathbf{P}_{y'_{\ell-1}} \left[Y_n = y'_\ell,\ Y_t \notin \varphi^{-1}(\mathbf{K}) \text{ for } 0 \leq t < n \right] \\ &\quad = \mathbf{P}_{y'_{\ell-1}} \left[Y_n = y'_\ell \right] - \mathbf{P}_{y'_{\ell-1}} \left[Y_n = y'_\ell,\ Y_t \in \varphi^{-1}(\mathbf{K}) \text{ for some } 0 \leq t < n \right], \end{aligned}$$

we need to show that

$$\mathbf{P}_{y'_{\ell-1}} \left[Y_n = y'_\ell,\ Y_t \in \varphi^{-1}(\mathbf{K}) \text{ for some } 0 \leq t < n \right] = O\left(\frac{n}{N^d} \right) \mathbf{P}_{y'_{\ell-1}} \left[Y_n = y'_\ell \right].$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

Define $A(x) = \{Y_n = y'_\ell,\ Y_t \in xN + \mathbf{K} \text{ for some } 0 \le t < n\}$, so that

$$\mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell,\ Y_t \in \varphi^{-1}(\mathbf{K}) \text{ for some } 0 \le t < n\right] \le \sum_{x \in \mathbb{Z}^d} \mathbf{P}_{y'_{\ell-1}}[A(x)]. \tag{28}$$

We have

$$\mathbf{P}_{y'_{\ell-1}}[A(x)] \le \sum_{n_1+n_2=n} \sum_{x' \in \mathbf{K}+xN} p_{n_1}\left(y'_{\ell-1}, x'\right) p_{n_2}\left(x', y'_\ell\right). \tag{29}$$

We bound this by splitting up the sum into different contributions. Let $\varepsilon > 0$; this will be chosen sufficiently small in the course of the proof.

Case 1: $n_1, n_2 \ge N^{2+6\varepsilon}$ and $|y'_{\ell-1} - x'| \le n_1^{\frac{1}{2}+\varepsilon}$, $|x' - y'_\ell| \le n_2^{\frac{1}{2}+\varepsilon}$. By the LCLT we have that

$$\begin{aligned} p_{n_1}\left(y'_{\ell-1}, x'\right) &\le C p_{n_1}\left(y'_{\ell-1}, u'\right) \quad \text{for any } u' \in \mathbb{T}_N + xN, \\ p_{n_2}\left(x', y'_\ell\right) &\le C p_{n_2}\left(u', y'_\ell\right) \quad \text{for any } u' \in \mathbb{T}_N + xN. \end{aligned} \tag{30}$$

For this note that we have

$$\begin{aligned} \left| \frac{d|y'_{\ell-1} - x'|^2}{n_1} - \frac{d|y'_{\ell-1} - u'|^2}{n_1} \right| &\le \frac{d|x' - u'|^2}{n_1} + \frac{2d|\langle x' - u', y'_{\ell-1} - x'\rangle|}{n_1} \\ &\le C\frac{N^2}{n_1} + \frac{CN \cdot n_1^{\frac{1}{2}+\varepsilon}}{n_1}, \end{aligned}$$

where the first term tends to 0 and the rest equals

$$CNn_1^{-\frac{1}{2}+\varepsilon} \le CN \cdot N^{(2+6\varepsilon)\left(-\frac{1}{2}+\varepsilon\right)} = CN^{-\varepsilon+6\varepsilon^2} \to 0, \quad \text{as } N \to \infty.$$

Here we choose ε so that $-\varepsilon + 6\varepsilon^2 < 0$. A similar observation shows the estimate for $p_{n_2}\left(x', y'_\ell\right)$.

The way we are going to use (30) is to replace the summation over x' by a summation over all $u' \in \mathbb{T}_N + xN$, at the same time inserting a factor $|\mathbf{K}|/N^d$. Hence the contribution of the values of n_1, n_2 and x in Case 1 to the right-hand side of (28) is at most

$$\begin{aligned} \frac{C|\mathbf{K}|}{N^d} \sum_{n_1+n_2=n} \sum_{u' \in \mathbb{Z}^d} p_{n_1}\left(y'_{\ell-1}, u'\right) p_{n_2}(u', y'_\ell) &= \frac{C|\mathbf{K}|}{N^d} \sum_{n_1+n_2=n} p_n\left(y'_{\ell-1}, y'_\ell\right) \\ &\le \frac{C|\mathbf{K}|n}{N^d} p_n\left(y'_{\ell-1}, y'_\ell\right). \end{aligned}$$

This completes the bound in Case 1. For future use, note that if $\varepsilon_n \to 0$ is any sequence, and we add the restriction $n_1 \le \varepsilon_n n$ to the conditions in Case 1, we obtain the upper bound

$$C|\mathbf{K}|\frac{\varepsilon_n n}{N^d} p_n\left(y'_{\ell-1}, y'_\ell\right) = o(1)\frac{n}{N^d} p_n\left(y'_{\ell-1}, y'_\ell\right). \tag{31}$$

Case 2a: $n_1, n_2 \ge N^{2+6\varepsilon}$ but $\left|x' - y'_{\ell-1}\right| > n_1^{\frac{1}{2}+\varepsilon}$. In this case we bound $p_{n_2}\left(x', y'_\ell\right) \le 1$ and have that the contribution of this case to the right-hand side of (28) is at most

$$\begin{aligned} \sum_{\substack{n_1+n_2=n \\ n_1,n_2 \ge N^{2+6\varepsilon}}} \mathbf{P}_{y'_{\ell-1}}\left[\left|Y_{n_1} - y'_{\ell-1}\right| > n_1^{1/2+\varepsilon}\right] &\le \sum_{\substack{n_1+n_2=n \\ n_1,n_2 \ge N^{2+6\varepsilon}}} C\exp\left(-cn_1^{2\varepsilon}\right) \\ &\le Cn\exp\left(-cN^{4\varepsilon}\right) = o\left(\frac{n}{N^d}\right) p_n\left(y'_{\ell-1}, y'_\ell\right), \end{aligned}$$

where in the first step we used (3) and in the last step we used the Gaussian lower bound (5) for p_n. Indeed, the requirement for the Gaussian lower bound is satisfied for sufficiently large N because $\left|y'_\ell - y'_{\ell+1}\right| \le C_1\sqrt{\log n}\sqrt{n} \le c\, n$. Therefore, we have

$$p_n\big(y'_{\ell-1}, y'_\ell\big) \ge \frac{c}{n^{d/2}} \exp\left(-\frac{C\left|y'_\ell - y'_{\ell-1}\right|^2}{n}\right) \ge \frac{c}{n^{d/2}} \exp\left(-C\ \log n\right). \tag{32}$$

Then we have

$$\frac{Cn\exp\left(-cN^{4\varepsilon}\right)}{cn^{-d/2}\exp\left(-C\ \log n\right)} \le Cn^{1+d/2}\exp\left(-cN^{4\varepsilon} + C\ \log n\right) = o\left(\frac{n}{N^d}\right), \text{ as } N \to \infty.$$

Case 2b: $n_1, n_2 \ge N^{2+6\varepsilon}$ but $|y'_\ell - x'| > n_2^{1/2+\varepsilon}$. This case can be handled very similarly to Case 2a.

Case 3a: $n_1 < N^{2+6\varepsilon}$ and $\left|x' - y'_{\ell-1}\right| \le N^{\frac{\delta}{2}-\varepsilon}$. By the LCLT we have

$$p_{n_2}\big(x', y'_\ell\big) = \frac{C}{n_2^{d/2}} \exp\left(-\frac{d|y'_\ell - x'|^2}{n_2}\right)(1+o(1)),$$
$$p_n\big(y'_{\ell-1}, y'_\ell\big) = \frac{C}{n^{d/2}} \exp\left(-\frac{d|y'_\ell - y'_{\ell-1}|^2}{n}\right)(1+o(1)).$$

We claim that

$$p_{n_2}\big(x', y'_\ell\big) \le C\, p_n\big(y'_{\ell-1}, y'_\ell\big). \tag{33}$$

We first note that $n_2 = n - n_1 = n(1+o(1))$, then deduce that $n_2^{-d/2} = O\big(n^{-d/2}\big)$ and

$$\exp\left(-\frac{d|y'_\ell - y'_{\ell-1}|^2}{n}\right) \ge \exp\left(-\frac{d|y'_\ell - y'_{\ell-1}|^2}{n_2}\right).$$

Since we have $\left|x' - y'_{\ell-1}\right| \le N^{\frac{\delta}{2}-\varepsilon}$ in the exponent, as $N \to \infty$ we have

$$\frac{|x' - y'_{\ell-1}|^2}{n_2} \le \frac{N^{\delta-2\varepsilon}}{n_2} \to 0$$

and

$$\frac{|y'_\ell - y'_{\ell-1}||x' - y'_{\ell-1}|}{n_2} \le \frac{n^{\frac{1}{2}}C_1\sqrt{\log n}N^{\frac{\delta}{2}-\varepsilon}}{n_2} \to 0.$$

These imply that

$$\left|\frac{|y'_\ell - y'_{\ell-1}|^2 - |y'_\ell - x'|^2}{n_2}\right| \le \left|\frac{|y'_\ell - y'_{\ell-1}|^2 - |(y'_\ell - y'_{\ell-1}) + (y'_{\ell-1} - x')|^2}{n_2}\right|$$
$$\le \frac{|x' - y'_{\ell-1}|^2}{n_2} + \frac{2|y'_\ell - y'_{\ell-1}||x' - y'_{\ell-1}|}{n_2} \to 0.$$

Thus (33) follows from comparing the LCLT approximations of the two sides.

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

We now have that the contribution of this case to the right-hand side of (28) is at most

$$
\begin{aligned}
Cp_n(y'_{\ell-1}, y'_\ell) \sum_{n_1 < N^{2+6\varepsilon}} \sum_{x \in \mathbb{Z}^d} \sum_{x' \in \mathbf{K}+xN} p_{n_1}\big(y'_{\ell-1}, x'\big) &\le \frac{C}{N^{\zeta(d-2)}} p_n\big(y'_{\ell-1}, y'_\ell\big) \\
&\le o(1) \frac{n}{N^d} p_n\big(y'_{\ell-1}, y'_\ell\big),
\end{aligned}
$$

where in the first step we used Lemma 4.1(i) and the last step holds for the value of ζ we chose; cf. (8).

Case 3b: $n_1 < N^{2+6\varepsilon}$ but $\left|x' - y'_{\ell-1}\right| > N^{\frac{\delta}{2}-\varepsilon}$. Use the Gaussian upper bound (5) to bound p_{n_1}, and bound the sum over all $x' \in \mathbb{Z}^d$ of p_{n_2} by 1 using symmetry of p_{n_2}, to get

$$
\begin{aligned}
&\sum_{\substack{n_1+n_2=n \\ n_1 < N^{2+6\varepsilon}}} \sum_{x \in \mathbb{Z}^d} \sum_{x' \in \mathbf{K}+xN} p_{n_1}\big(y'_{\ell-1}, x'\big) p_{n_2}\big(x', y'_\ell\big) \\
&\quad \le \sum_{n_1 < N^{2+6\varepsilon}} \frac{C}{n_1^{d/2}} \exp\left(-\frac{N^{\delta-2\varepsilon}}{N^{2+6\varepsilon}}\right) \sum_{x' \in \mathbb{Z}^d} p_{n-n_1}\big(x', y'_\ell\big) \\
&\quad \le \sum_{n_1 < N^{2+6\varepsilon}} \frac{C}{n_1^{d/2}} \exp\left(-\frac{N^{\delta-2\varepsilon}}{N^{2+6\varepsilon}}\right) \\
&\quad \le CN^{O(1)} \exp\left(-N^{\delta-2-8\varepsilon}\right) = o\left(\frac{n}{N^d}\right) p_n\big(y'_{\ell-1}, y'_\ell\big), \qquad \text{as } N \to \infty.
\end{aligned}
$$

In the last step we used a Gaussian lower bound for p_n; cf. (32).

Case 4a: $n_2 < N^{2+6\varepsilon}$ and $|y'_\ell - x'| \le N^{\frac{\delta}{2}-\varepsilon}$. This case can be handled very similarly to Case 3a.

Case 4b: $n_2 < N^{2+6\varepsilon}$ and $|y'_\ell - x'| > N^{\frac{\delta}{2}-\varepsilon}$. This case can be handled very similarly to Case 3b.

Therefore, we have discussed all possible cases and proved statement (14) of the proposition as required.

The proof of (15) is similar to that of the first part, with only a few modifications. In this part we have to show that

$$
\mathbf{P}_o\left[Y_n = y'_1,\ Y_t \in \varphi^{-1}(\mathbf{K}) \text{ for some } 0 \le t < n\right] = o(1)\mathbf{P}_o\left[Y_n = y'_1\right].
$$

Define $A_0(x) = \left\{Y_n = y'_1, Y_t \in xN + \mathbf{K} \text{ for some } 0 \le t < n\right\}$, so that

$$
\mathbf{P}_o\left[Y_n = y'_1,\ Y_t \in \varphi^{-1}(\mathbf{K}) \text{ for some } 0 \le t < n\right] \le \sum_{x \in \mathbb{Z}^d} \mathbf{P}_o[A_0(x)]. \tag{34}
$$

We have

$$
\mathbf{P}_o[A_0(x)] \le \sum_{n_1+n_2=n} \sum_{x' \in \mathbf{K}+xN} p_{n_1}(o, x') p_{n_2}\big(x', y'_1\big). \tag{35}
$$

We bound the term above by splitting up the sum into the same cases as in the proof of (14). The different cases can be handled very similarly to the first part. The difference is only in Case 3a while applying the Green's function bound Lemma 4.1.

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

In Case 3a, by the LCLT, we can deduce that

$$p_{n_2}\big(x', y'_1\big) \le C\, p_n\big(o, y'_1\big).$$

If $g(N) > N^\zeta$, the bound of Lemma 4.1(i) can be used as before. If $g(N) \le N^\zeta$, by Lemma 4.1(ii), we have that the contribution of this case to the right-hand side of (34) is at most

$$Cp_n\big(o, y'_1\big) \sum_{n_1 < N^{2+6\varepsilon}} \sum_{x\in\mathbb{Z}^d} \sum_{x'\in\mathbf{K}+xN} p_{n_1}(o, x') \le \frac{C}{g(N)^{d-2}} p_n\big(o, y'_1\big) = o(1) p_n\big(o, y'_1\big).$$

Here we used that $g(N) \to \infty$.

Note that Case 4a can be handled in the same way as in the proof of (14), since the distance between y'_1 and any copy of $\mathbf{K}$ is at least N^ζ.

Therefore, we have discussed all possible cases and proved (15) as required. □

For future use, we extract a few corollaries of the proof of Proposition 2.3.

Corollary 4.1. *Assume that $y', y'' \in \mathbb{Z}^d$ are points such that $|y'' - y'| \le 2C_1\sqrt{\log n}\sqrt{n}$. Then for all $n/2 \le m \le n$ we have*

$$\sum_{n_1+n_2=m} \sum_{x\in\mathbb{Z}^d} \sum_{\substack{x'\in\mathbf{K}+xN \\ |y'-x'|,|y''-x'|>N^\zeta}} p_{n_1}(y', x') p_{n_2}(x', y'') = O\left(\frac{n}{N^d}\right) p_m(y', y''). \tag{36}$$

Proof. In the course of the proof of Proposition 2.3, we established the above with $m = n$, where $y' = y'_{\ell-1}$ and $y'' = y'_\ell$, and with C_1 in place of $2C_1$ in the upper bound on the displacement $|y'' - y'|$. Note that in this case the restriction $|y' - x'|, |y'' - x'| > N^\zeta$ in the summation holds for all x, because of conditions imposed on $y'_{\ell-1}$ and y'_ℓ in the definition of $\mathcal{G}_{\zeta,C_1}$.

The arguments when $n/2 \le m < n$ and with the upper bound increased by a factor of 2 are essentially the same. The information that $y'_{\ell-1}$ and y'_ℓ are at least distance N^ζ from $\varphi^{-1}(\mathbf{K})$ was only used in Cases 3a and 4a to handle terms x' close to these points. Since in (36) we exclude such x' from the summation, the statement follows without restricting the location of y', y''. □

The following is merely a restatement of what was observed in (31) (with Part (ii) below holding by symmetry).

Corollary 4.2.

(i) *For $\ell \ge 2$ and any sequence $\varepsilon_n \to 0$, we have*

$$\sum_{\substack{n_1+n_2=n \\ n_1\le\varepsilon_n n \\ n_1,n_2\ge N^{2+6\varepsilon}}} \sum_{x\in\mathbb{Z}^d} \sum_{\substack{x'\in\mathbf{K}+xN: \\ |y'_{\ell-1}-x'|\le n_1^{\frac12+\varepsilon} \\ |x'-y'_\ell|\le n_2^{\frac12+\varepsilon}}} p_{n_1}\big(y'_{\ell-1}, x'\big) p_{n_2}\big(x', y'_\ell\big) = o(1)\frac{n}{N^d} p_n\big(y'_{\ell-1}, y'_\ell\big). \tag{37}$$

(ii) *The same right-hand-side expression is valid if we replace the restriction $n_1 \le \varepsilon_n n$ by $n_2 \le \varepsilon_n n$.*

The following is a restatement of the bounds of Cases 2a and 2b.

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

Corollary 4.3. *For $\ell \geq 2$ we have*

(i)

$$\sum_{\substack{n_1+n_2=n \\ n_1,n_2 \geq N^{2+6\varepsilon}}} \sum_{x \in \mathbb{Z}^d} \sum_{\substack{x' \in \mathbf{K}+xN: \\ |y'_{\ell-1}-x'|>n_1^{\frac{1}{2}+\varepsilon}}} p_{n_1}\big(y'_{\ell-1}, x'\big) p_{n_2}\big(x', y'_\ell\big) = o(1) \frac{n}{N^d} p_n\big(y'_{\ell-1}, y'_\ell\big), \tag{38}$$

(ii)

$$\sum_{\substack{n_1+n_2=n \\ n_1,n_2 \geq N^{2+6\varepsilon}}} \sum_{x \in \mathbb{Z}^d} \sum_{\substack{x' \in \mathbf{K}+xN: \\ |x'-y'_\ell|>n_2^{\frac{1}{2}+\varepsilon}}} p_{n_1}\big(y'_{\ell-1}, x'\big) p_{n_2}\big(x', y'_\ell\big) = o(1) \frac{n}{N^d} p_n\big(y'_{\ell-1}, y'_\ell\big). \tag{39}$$

The following is a restatement of the bounds of Cases 3a and 3b combined.

Corollary 4.4. *For $\ell \geq 2$ we have*

$$\sum_{\substack{n_1+n_2=n \\ n_1 < N^{2+6\varepsilon}}} \sum_{x \in \mathbb{Z}^d} \sum_{x' \in \mathbf{K}+xN} p_{n_1}\big(y'_{\ell-1}, x'\big) p_{n_2}\big(x', y'_\ell\big) = o(1) \frac{n}{N^d} p_n\big(y'_{\ell-1}, y'_\ell\big). \tag{40}$$

4.2. Proof of Proposition 2.2

In this section we need C_1 large enough so that we have

$$e^{-df(n)^2} N^d n^{1+3d/2} \to 0. \tag{41}$$

We have

$$\mathbf{P}_{y'_{\ell-1}}\big[Y_n = y'_\ell,\ Y_t \in \varphi^{-1}(\mathbf{K}) \text{ for some } 0 \leq t < n\big] = \mathbf{P}_{y'_{\ell-1}}\big[\cup_{x \in \mathbb{Z}^d} A(x)\big],$$

where

$$A(x) = \big\{Y_n = y'_\ell,\, Y_t \in xN + \mathbf{K} \text{ for some } 0 \leq t < n\big\}.$$

The strategy is to estimate the probability via the Bonferroni inequalities:

$$\begin{aligned} \sum_x \mathbf{P}_{y'_{\ell-1}}[A(x)] - \sum_{x_1 \neq x_2} \mathbf{P}_{y'_{\ell-1}}[A(x_1) \cap A(x_2)] &\leq \mathbf{P}_{y'_{\ell-1}}\big[\cup_{x \in \mathbb{Z}^d} A(x)\big] \\ &\leq \sum_x \mathbf{P}_{y'_{\ell-1}}[A(x)]. \end{aligned} \tag{42}$$

We are going to use a parameter A_n that we choose as $A_n = 10 \log \log n$, so that in particular $A_n \to \infty$.

4.2.1. *The main contribution.* In this section, we consider only stretches with $|y'_\ell - y'_{\ell-1}| \leq A_n\sqrt{n}$. We will show that the main contribution in (42) comes from x in the set

$$G = \Big\{x \in \mathbb{Z}^d : |y'_{\ell-1} - xN| \leq A_n^2\sqrt{n},\ |xN - y'_\ell| \leq A_n^2\sqrt{n}\Big\}.$$

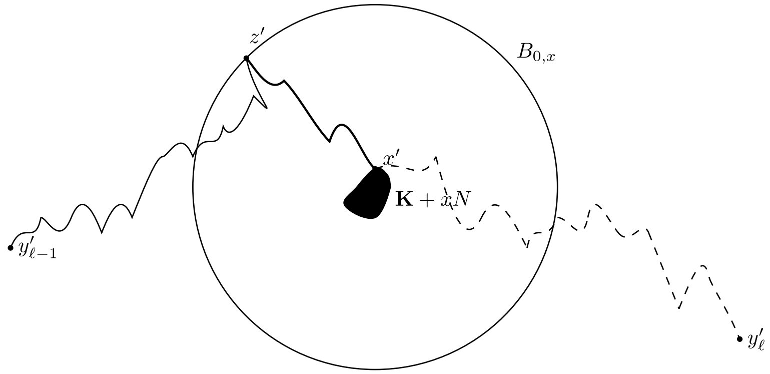

FIGURE 2. The decomposition of a path hitting a copy of **K** into three subpaths (not to scale).

We first examine $\mathbf{P}_{y'_{\ell-1}}[A(x)]$ for $x \in G$. Putting $B_{0,x} = B(\mathbf{x}_0 + xN, N^\zeta)$, let n_1 be the time of the last visit to $\partial B_{0,x}$ before hitting $\mathbf{K} + xN$, let $n_1 + n_2$ be the time of the first hit of $\mathbf{K} + xN$, and let $n_3 = n - n_1 - n_2$. See Figure 2 for an illustration of this decomposition.

Then we can write

$$\begin{aligned}\mathbf{P}_{y'_{\ell-1}}[A(x)] = \sum_{n_1+n_2+n_3=n} \sum_{z' \in \partial B_{0,x}} \sum_{x' \in \mathbf{K}+xN} \widetilde{p}^{(x)}_{n_1}\big(y'_{\ell-1}, z'\big) \\ \times \mathbf{P}_{z'}\big[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}},\ Y_{H_{\mathbf{K}+xN}} = x'\big]\, p_{n_3}\big(x', y'_\ell\big),\end{aligned} \tag{43}$$

where

$$\widetilde{p}^{(x)}_{n_1}\big(y'_{\ell-1}, z'\big) = \mathbf{P}_{y'_{\ell-1}}\big[Y_{n_1} = z',\ Y_t \notin \mathbf{K} + xN \text{ for } 0 \le t \le n_1\big].$$

We are going to use another parameter ε_n that will need to go to 0 slowly. We choose it as $\varepsilon_n = (10 \log\log n)^{-1} \to 0$. The main contribution to (43) will be when $n_1 \ge \varepsilon_n n$, $n_3 \ge \varepsilon_n n$, and $n_2 \le N^{2\delta/d} \sim n^{2/d}$. Therefore, we split the sum over n_1, n_2, n_3 in (43) into a main contribution $I(x)$ and an error term $II(x)$. In order to define these, let

$$\begin{aligned}&F\big(n_1, n_2, n_3, x, y'_{\ell-1}, y'_\ell\big) \\ &\quad = \sum_{z' \in \partial B_{0,x}} \sum_{x' \in \mathbf{K}+xN} \widetilde{p}^{(x)}_{n_1}\big(y'_{\ell-1}, z'\big) \mathbf{P}_{z'}\big[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}},\ Y_{H_{\mathbf{K}+xN}} = x'\big] p_{n_3}\big(x', y'_\ell\big).\end{aligned} \tag{44}$$

Then with

$$\begin{aligned}I(x) &:= \sum_{\substack{n_1+n_2+n_3=n \\ n_1, n_3 \ge \varepsilon_n n,\ n_2 \le N^{2\delta/d}}} F\big(n_1, n_2, n_3, x, y'_{\ell-1}, y'_\ell\big), \\ II(x) &:= \sum_{\substack{n_1+n_2+n_3=n \\ n_1 < \varepsilon_n n \text{ or } n_3 < \varepsilon_n n \\ \text{or } n_2 > N^{2\delta/d}}} F\big(n_1, n_2, n_3, x, y'_{\ell-1}, y'_\ell\big),\end{aligned} \tag{45}$$

we have

$$\mathbf{P}_{y'_{\ell-1}}[A(x)] = I(x) + II(x).$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

Lemma 4.2. *When $x \in G$ and $n_3 \geq \varepsilon_n n$, we have*

$$p_{n_3}\big(x', y'_\ell\big) = (1+o(1))p_{n_3}\big(u', y'_\ell\big) \quad \textit{for all } x' \in \boldsymbol{K} + xN \textit{ and all } u' \in \mathbb{T}_N + xN,$$

with the o(1) term uniform in x' and u'.

Proof. By the LCLT, we have

$$\begin{aligned} p_{n_3}\big(x', y'_\ell\big) &= \frac{C}{n_3^{d/2}} \exp\left(-\frac{d|y'_\ell - x'|^2}{n_3}\right)(1+o(1)), \\ p_{n_3}\big(u', y'_\ell\big) &= \frac{C}{n_3^{d/2}} \exp\left(-\frac{d|y'_\ell - u'|^2}{n_3}\right)(1+o(1)). \end{aligned}$$

We compare the exponents

$$\begin{aligned} \left|\frac{d\big|y'_\ell - x'\big|^2}{n_3} - \frac{d\big|y'_\ell - u'\big|^2}{n_3}\right| &\leq \frac{d|x'-u'|^2}{n_3} + \frac{2d\big|\langle x'-u', y'_\ell - x'\rangle\big|}{n_3} \\ &\leq C\frac{N^2}{n_3} + \frac{CN \cdot A_n^2\sqrt{n}}{n_3} \to 0, \end{aligned}$$

as $N \to \infty$. □

Lemma 4.3. *When $x \in G$ and $n_1 \geq \varepsilon_n n$, we have*

$$\widetilde{p}_{n_1}^{(x)}\big(y'_{\ell-1}, z'\big) = (1+o(1))p_{n_1}\big(y'_{\ell-1}, u'\big) \quad \textit{for all } z' \in \partial B_{0,x} \textit{ and all } u' \in \mathbb{T}_N + xN,$$

with the o(1) term uniform in z' and u'.

Proof. The statement will follow if we show the following claim:

$$\mathbf{P}_{y'_{\ell-1}}\big[Y_{n_1} = z',\ Y_t \in \mathbf{K} + xN \text{ for some } 0 \leq t \leq n_1\big] = o(1)p_{n_1}\big(y'_{\ell-1}, z'\big).$$

For this, observe that by (5) we have

$$\begin{aligned} p_{n_1}\big(y'_{\ell-1}, z'\big) &\geq \frac{c}{n_1^{d/2}} \exp\left(-C\frac{|z'-y'_{\ell-1}|^2}{n_1}\right) \\ &\geq \frac{c}{n^{d/2}} \exp\left(-c\frac{A_n^2 n + N^\zeta}{\varepsilon_n n}\right) \\ &\geq \frac{c}{n^{d/2}} \exp\left(-C(\log\log n)^{O(1)}\right) \\ &= n^{-d/2+o(1)}. \end{aligned} \tag{46}$$

On the other hand, using the Markov property, (5), and the fact that for $x' \in \mathbf{K} + xN$ we have $\big|y'_{\ell-1} - x'\big| \geq cN^\zeta$ and $|x'-z'| \geq cN^\zeta$, we get

$$
\begin{aligned}
\mathbf{P}_{y'_{\ell-1}}&\left[Y_{n_1}=z',\ Y_t\in\mathbf{K}+xN \text{ for some } 0\le t\le n_1\right]\\
&\le \sum_{1\le m\le n_1-1}\ \sum_{x'\in\mathbf{K}+xN} p_m\big(y'_{\ell-1},x'\big)\,p_{n_1-m}(x',z')\\
&\le C\sum_{1\le m\le n_1-1}\frac{1}{m^{d/2}}\frac{1}{(n_1-m)^{d/2}}\exp\left(-c\frac{N^{2\zeta}}{m}\right)\exp\left(-c\frac{N^{2\zeta}}{n_1-m}\right)\\
&\le C\sum_{1\le m\le n_1/2}\frac{1}{m^{d/2}}\frac{1}{(n_1-m)^{d/2}}\exp\left(-c\frac{N^{2\zeta}}{m}\right)\exp\left(-c\frac{N^{2\zeta}}{n_1-m}\right).
\end{aligned}
\tag{47}
$$

We note here that the sum over $1\le m\le n_1/2$ and the sum over $n_1/2\le m\le n_1-1$ are symmetric. Bounding the sum over $1\le m\le n_1/2$ gives

$$
\begin{aligned}
&\frac{C}{n_1^{d/2}}\sum_{1\le m\le n_1/2}\frac{1}{m^{d/2}}\exp\left(-c\frac{N^{2\zeta}}{m}\right)\\
&=\frac{C}{n_1^{d/2}}\left[\sum_{1\le m\le N^{2\zeta}}\frac{1}{m^{d/2}}\exp\left(-c\frac{N^{2\zeta}}{m}\right)+\sum_{N^{2\zeta}<m\le n_1/2}\frac{1}{m^{d/2}}\exp\left(-c\frac{N^{2\zeta}}{m}\right)\right].
\end{aligned}
\tag{48}
$$

In the second sum we can bound the exponential by 1 and get the upper bound

$$
\frac{C}{n_1^{d/2}}N^{\zeta(2-d)}=o\big(n^{-d/2+o(1)}\big).
$$

In the first sum, we group terms on dyadic scales k so that $2^k\le N^{2\zeta}/m\le 2^{k+1}$, $k=0,\dots,\lfloor\log_2 N^{2\zeta}\rfloor+1$. This gives the bound

$$
\frac{C}{n_1^{d/2}}\sum_{k=0}^{\lfloor\log_2 N^{2\zeta}\rfloor+1}\frac{(2^{k+1})^{d/2}}{(N^{2\zeta})^{d/2}}\exp\left(-c2^k\right)\le\frac{C}{n_1^{d/2}}\frac{1}{N^{\zeta d}},
$$

which is also $o\big(n^{-d/2+o(1)}\big)$. □

In order to apply the previous two lemmas to analyse $I(x)$ in (45), we first define a modification of F in (44), where z' and x' are both replaced by a vertex $u'\in\mathbb{T}_N+xN$. That is, we define

$$
\begin{aligned}
&\widetilde{F}\big(n_1,n_2,n_3,u',x,y'_{\ell-1},y'_\ell\big)\\
&\quad=\sum_{z'\in\partial B_{0,x}}\ \sum_{x'\in\mathbf{K}+xN}p_{n_1}\big(y'_{\ell-1},u'\big)\mathbf{P}_{z'}\left[H_{\mathbf{K}+xN}=n_2<\xi_{B_{0,x}},\ Y_{H_{\mathbf{K}+xN}}=x'\right]p_{n_3}\big(u',y'_\ell\big).
\end{aligned}
$$

Then Lemmas 4.2 and 4.3 allow us to write, for $x\in G$, the main term $I(x)$ in (45) as

$$
\begin{aligned}
I(x)&=\sum_{\substack{n_1+n_2+n_3=n\\ n_1,n_3\ge\varepsilon_n n,\ n_2\le N^{2\delta/d}}}F\big(n_1,n_2,n_3,x,y'_{\ell-1},y'_\ell\big)\\
&=\frac{1+o(1)}{N^d}\sum_{u'\in\mathbb{T}_N+xN}\ \sum_{\substack{n_1+n_2+n_3=n\\ n_1,n_3\ge\varepsilon_n n,\ n_2\le N^{2\delta/d}}}\widetilde{F}\big(n_1,n_2,n_3,u',x,y'_{\ell-1},y'_\ell\big)
\end{aligned}
\tag{49}
$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

$$= \frac{1+o(1)}{N^d} \sum_{u' \in \mathbb{T}_N + xN} \sum_{\substack{n_1+n_2+n_3=n \\ n_1,n_3 \geq \varepsilon_n n \\ n_2 \leq N^{2\delta/d}}} p_{n_1}\left(y'_{\ell-1}, u'\right) p_{n_3}\left(u', y'_\ell\right)$$

$$\times \sum_{z' \in \partial B_{0,x}} \mathbf{P}_{z'}\left[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}}\right],$$

where the sum over x' is removed since

$$\sum_{x' \in \mathbf{K}+xN} \mathbf{P}_{z'}\left[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}},\ Y_{H_{\mathbf{K}+xN}} = x'\right] = \mathbf{P}_{z'}\left[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}}\right].$$

Lemma 4.4. *Assume that $n_1, n_3 \geq \varepsilon_n n$ and $n_2 \leq N^{2\delta/d}$.*

(i) *We have*

$$p_{n_1+n_3}\left(y'_{\ell-1}, y'_\ell\right) = (1+o(1)) p_n\left(y'_{\ell-1}, y'_\ell\right).$$

(ii) *We have*

$$\sum_{x \in G} \sum_{u' \in \mathbb{T}_N + xN} p_{n_1}\left(y'_{\ell-1}, u'\right) p_{n_3}\left(u', y'_\ell\right) = (1+o(1)) p_{n_1+n_3}\left(y'_{\ell-1}, y'_\ell\right). \tag{50}$$

Proof.

(i) When $n_2 \leq N^{2\delta/d} \sim n^{2/d}$, we have

$$n_1 + n_3 = n\left(1 - O\left(n^{-1+2/d}\right)\right).$$

Hence the exponential term in the LCLT for $p_{n_1+n_3}\left(y'_{\ell-1}, y'_\ell\right)$ is

$$\exp\left(-\frac{|y'_\ell - y'_{\ell-1}|^2}{n}\left(1 + O\left(n^{-1+2/d}\right)\right)\right) = (1+o(1)) \exp\left(-\frac{\left|y'_\ell - y'_{\ell-1}\right|^2}{n}\right),$$

where we used that $A_n = 10 \log\log n$, and hence $\left|y'_\ell - y'_{\ell-1}\right|^2 \leq A_n^2 n = n\, o\left(n^{1-2/d}\right)$.

(ii) If we summed over all $x \in \mathbb{Z}^d$, we would get exactly $p_{n_1+n_3}\left(y'_{\ell-1}, y'_\ell\right)$. Thus the claim amounts to showing that

$$\sum_{x \in \mathbb{Z}^d \setminus G} \sum_{u' \in \mathbb{T}_N + xN} p_{n_1}\left(y'_{\ell-1}, u'\right) p_{n_3}\left(u', y'_\ell\right) = o(1) p_{n_1+n_3}\left(y'_{\ell-1}, y'_\ell\right). \tag{51}$$

First, note that from the LCLT we have

$$p_{n_1+n_3}\left(y'_{\ell-1}, y'_\ell\right) = (1+o(1)) \overline{p}_{n_1+n_3}\left(y'_{\ell-1}, y'_\ell\right).$$

In order to estimate the left-hand side of (51), using (5), we can estimate the contribution of $\left\{x \in \mathbb{Z}^d \setminus G :\ \max\left\{\left|y'_{\ell-1} - xN\right|, \left|xN - y'_{\ell-1}\right|\right\} > A_n^2 \sqrt{n}\right\}$ as follows. First, we have

$$p_{n_1+n_3}\left(y'_{\ell-1}, y'_\ell\right) \geq \frac{c}{n^{d/2}} \exp\left(-C A_n^2 (1+o(1))\right) \geq \frac{c}{n^{d/2}} \exp\left(-C(\log\log n)^2\right).$$

Here we used $\left|y'_\ell - y'_{\ell-1}\right|^2 \leq A_n^2 n$ and $n_1 + n_3 = n(1 - o(1))$.

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

On the other hand, note that either $n_1 \geq n/3$ or $n_3 \geq n/3$. Without loss of generality, assume that $n_3 \geq n/3$. Then the contribution to the left-hand side of (51), using (5), and by summing in dyadic shells with radii $2^k A_n^2 \sqrt{n}$, $k = 0, 1, 2, \ldots$, we get the bound

$$\begin{aligned}
&\sum_{k=0}^{\infty} C\left(A_n^2\sqrt{n}\right)^d 2^{dk} \frac{C}{n_1^{d/2}} \exp\left(-c2^{2k}A_n^4 n/n_1\right) \frac{1}{n_3^{d/2}} \exp\left(-c2^{2k}A_n^4 n/n_3\right) \\
&\quad \leq \sum_{k=0}^{\infty} C\, A_n^{2d}\, 2^{dk} \frac{1}{\varepsilon_n^{d/2}} \exp\left(-c2^{2k}A_n^4\right) \frac{1}{n^{d/2}} \exp\left(-c2^{2k}A_n^4\right) \\
&\quad \leq \frac{C}{n^{d/2}} \frac{A_n^{2d}}{\varepsilon_n^{d/2}} \sum_{k=0}^{\infty} \exp\left(-c2^{2k}(\log\log n)^4 + dk\log 2\right) \\
&\quad = \frac{C}{n^{d/2}}\, o\left(\exp\left(-100(\log\log n)^2\right)\right).
\end{aligned} \tag{52}$$

□

The above lemma allows us to write

$$\begin{aligned}
\sum_{x\in G} I(x) &= \frac{1+o(1)}{N^d}\, p_n\left(y'_{\ell-1}, y'_\ell\right) \sum_{\substack{n_1+n_2+n_3=n \\ n_1,n_3\geq \varepsilon_n n \\ n_2 \leq N^{2\delta/d}}} \sum_{z'\in\partial B_{0,x}} \mathbf{P}_{z'}\left[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}}\right] \\
&= \frac{(1+o(1))\, n}{N^d}\, p_n\left(y'_{\ell-1}, y'_\ell\right) \sum_{n_2 \leq N^{2\delta/d}} \sum_{z'\in\partial B_{0,x}} \mathbf{P}_{z'}\left[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}}\right].
\end{aligned} \tag{53}$$

The next lemma will help us extract the Cap(K) contribution from the right-hand side of (43).

Lemma 4.5. *We have*

$$\sum_{n_2=0}^{N^{2\delta/d}} \sum_{z'\in\partial B_{0,x}} \mathbf{P}_{z'}\left[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}}\right] = \frac{1}{2}\mathrm{Cap}(K)\,(1+o(1)). \tag{54}$$

Proof. Performing the sum over n_2 allows us to rewrite the expression in the left-hand side of (54) as

$$\begin{aligned}
&\left(\sum_{z'\in\partial B_{0,x}} \mathbf{P}_{z'}\left[H_{\mathbf{K}+xN} < \xi_{B_{0,x}}\right]\right) - \left(\sum_{z'\in\partial B_{0,x}} \mathbf{P}_{z'}\left[N^{2\delta/d} < H_{\mathbf{K}+xN} < \xi_{B_{0,x}}\right]\right) \\
&\quad = \frac{1}{2}\mathrm{Cap}(K) + o(1) - \sum_{z'\in\partial B_{0,x}} \sum_{\mathbf{x}\in\mathbf{K}} \mathbf{P}_{z'}\left[N^{2\delta/d} < H_{\mathbf{K}+xN} < \xi_{B_{0,x}},\ Y_{H_{\mathbf{K}+xN}} = \mathbf{x} + xN\right].
\end{aligned}$$

Here the $1/2$ before Cap(K) comes from the random walk being lazy; see (4). Using time-reversal for the summand in the last term, we get the expression

$$\begin{aligned}
&= \frac{1}{2}\mathrm{Cap}(K) + o(1) - \sum_{\mathbf{x}\in\mathbf{K}} \sum_{z'\in\partial B_{0,x}} \mathbf{P}_{\mathbf{x}+xN}\left[N^{2\delta/d} < \xi_{B_{0,x}} < H_{\mathbf{K}+xN},\ Y_{\xi_{B_{0,x}}} = z'\right] \\
&\quad = \frac{1}{2}\mathrm{Cap}(K) + o(1) - \sum_{\mathbf{x}\in\mathbf{K}} \mathbf{P}_{\mathbf{x}+xN}\left[N^{2\delta/d} < \xi_{B_{0,x}} < H_{\mathbf{K}+xN}\right].
\end{aligned} \tag{55}$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

The subtracted term in the right-hand side of (55) is at most

$$|\mathbf{K}| \max_{\mathbf{x}\in\mathbf{K}} \mathbf{P}_{\mathbf{x}+xN}\Big[\xi_{B_{0,x}} > N^{2\delta/d}\Big].$$

Since $\zeta < \delta/d$, this expression is $o(1)$. □

From the above lemma we get that the main contribution equals

$$\sum_{x\in G} I(x) = (1+o(1))\frac{n}{N^d}\,\frac{1}{2}\,\mathrm{Cap}(K)\,p_n\big(y'_{\ell-1}, y'_\ell\big). \tag{56}$$

It remains to estimate all the error terms.

4.2.2. *The error terms.*

Lemma 4.6. *We have*

$$\sum_{x\in G} II(x) = o(1)\frac{n}{N^d} p_n\big(y'_{\ell-1}, y'_\ell\big).$$

Proof. We split the estimates according to which condition is violated in the sum. Recall that in the proof of Proposition 2.3 we chose $\varepsilon > 0$ such that $-\varepsilon + 6\varepsilon^2 < 0$. Here we make the further restriction that $\varepsilon < 2\delta/d - 2\zeta$.

Case 1: $n_2 > N^{2\delta/d}$. We claim that

$$\mathbf{P}_{z'}\big[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}},\, Y_{H_{\mathbf{K}+xN}} = x'\big] \le C\exp\big(-N^{\varepsilon/2}\big)p_{n_2}\big(z', x'\big). \tag{57}$$

Since, in every time interval of duration $N^{2\zeta}$, the walk has a positive chance of exiting the ball $B_{0,x}$, we have

$$\begin{aligned}\mathbf{P}_{z'}\big[H_{\mathbf{K}+xN} = n_2 < \xi_{B_{0,x}},\, Y_{H_{\mathbf{K}+xN}} = x'\big] &\le \mathbf{P}_{z'}\Big[\xi_{B_{0,x}} > N^{2\delta/d}\Big] \le C\exp\left(-c\frac{N^{2\delta/d}}{N^{2\zeta}}\right)\\ &\le C\exp\big(-N^{\varepsilon}\big).\end{aligned}$$

By (5) on p_{n_2}, and since $\zeta < \delta/d$ and $N^{2\delta/d} < n_2 < n$, we have

$$p_{n_2}(z', x') \ge \frac{c}{n_2^{d/2}}\exp\left(-C\frac{N^{2\zeta}}{n_2}\right) \ge c\exp\big(-N^{\varepsilon/2}\big).$$

Here we lower-bounded $\exp\left(-C\frac{N^{2\zeta}}{n_2}\right)$ by c. The claim (57) is proved.

We also have the bound

$$\widetilde{p}^{(x)}_{n_1}\big(y'_{\ell-1}, z'\big) \le p_{n_1}\big(y'_{\ell-1}, z'\big).$$

We then get (summing over z' and x') that the contribution to $\sum_{x\in\mathbb{Z}^d} II(x)$ from Case 1 is at most

$$\begin{aligned}&\sum_{n_1+n_2+n_3=n}\;\sum_{z'\in\mathbb{Z}^d}\;\sum_{x'\in\mathbb{Z}^d} p_{n_1}\big(y'_{\ell-1}, z'\big)\,C\exp\big(-N^{\varepsilon/2}\big)p_{n_2}(z', x')\,p_{n_3}\big(x', y'_\ell\big)\\ &\quad\le C\exp\big(-N^{\varepsilon/2}\big)\sum_{n_1+n_2+n_3=n} p_n\big(y'_{\ell-1}, y'_\ell\big)\end{aligned}$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

$$\leq Cn^2 \exp\left(-N^{\varepsilon/2}\right) p_n\left(y'_{\ell-1}, y'_\ell\right)$$
$$= o(1) \frac{n}{N^d} p_n\left(y'_{\ell-1}, y'_\ell\right).$$

Case 2: $n_2 \leq N^{2\delta/d}$ and $n_1 < \varepsilon_n n$. Note that since $n_2 \leq N^2 \leq \varepsilon_n n$ for large enough N, if we put $n'_1 = n_1 + n_2$ and $n'_2 = n_3$, we can upper-bound the contribution of this case by

$$\sum_{\substack{n'_1+n'_2=n \\ n'_1 \leq 2\varepsilon_n n}} \sum_{x\in\mathbb{Z}^d} \sum_{x'\in\mathbf{K}+xN} p_{n'_1}\left(y'_{\ell-1}, x'\right) p_{n'_2}\left(x', y'_\ell\right).$$

Now we can make use of the corollaries stated after the proof of Proposition 1 as follows.

Case 2-(i). $N^{2+6\varepsilon} \leq n'_1 \leq 2\varepsilon_n n$, $\left|y'_{\ell-1} - x'\right| \leq \left(n'_1\right)^{\frac{1}{2}+\varepsilon}$ and $\left|x' - y'_\ell\right| \leq \left(n'_2\right)^{\frac{1}{2}+\varepsilon}$. Note that for large enough N we have $n'_2 \geq (n - 2\varepsilon_n n) \geq N^{2+6\varepsilon}$. Hence, by Corollary 4.2(i) (with ε_n there replaced by $2\varepsilon_n$), the contribution of this case is

$$o(1) \frac{n}{N^d} p_n\left(y'_{\ell-1}, y'_\ell\right).$$

Case 2-(ii). $N^{2+6\varepsilon} \leq n'_1 \leq 2\varepsilon_n n$ but either $\left|y'_{\ell-1} - x'\right| > \left(n'_1\right)^{\frac{1}{2}+\varepsilon}$ or $\left|x' - y'_\ell\right| > \left(n'_2\right)^{\frac{1}{2}+\varepsilon}$. Again, we have $n'_2 \geq N^{2+6\varepsilon}$. Hence, neglecting the requirement $n'_1 \leq 2\varepsilon_n n$, Corollary 4.3 immediately implies that the contribution of this case is

$$o(1) \frac{n}{N^d} p_n\left(y'_{\ell-1}, y'_\ell\right).$$

Case 2-(iii). $n'_1 < N^{2+6\varepsilon}$. It follows immediately from Corollary 4.4 that the contribution of this case is

$$o(1) \frac{n}{N^d} p_n\left(y'_{\ell-1}, y'_\ell\right).$$

Case 3: $n_2 \leq N^{2\delta/d}$ and $n_3 < \varepsilon_n n$. By symmetry, this case can be handled very similarly to Case 2. □

Lemma 4.7. *We have*

$$\sum_{x\in\mathbb{Z}^d\setminus G} \mathbf{P}[A(x)] = o(1) \frac{n}{N^d} p_n\left(y'_{\ell-1}, y'_\ell\right).$$

Proof. By the same arguments as in Lemma 4.4(ii), we have

$$p_n\left(y'_{\ell-1}, y'_\ell\right) \geq \frac{C}{n^{d/2}} \exp\left(-100(\log\log n)^2\right).$$

For $x \in \mathbb{Z}^d \setminus G$, let k be the dyadic scale that satisfies

$$2^k A_n^2 \sqrt{n} \leq \left|x' - y'_{\ell-1}\right| < 2^{k+1} A_n^2 \sqrt{n}.$$

The same bounds hold up to constants for $\left|x' - y'_\ell\right|$.

Then we have

$$\begin{aligned}\mathbf{P}[A(x)] &\leq \sum_{1\leq m\leq n-1}\sum_{x'\in\mathbf{K}+xN} p_m\big(y'_{\ell-1}, x'\big)p_{n-m}\big(x', y'_\ell\big)\\ &\leq C|\mathbf{K}| \sum_{1\leq m\leq n-1} \frac{1}{m^{d/2}}\frac{1}{(n-m)^{d/2}} \exp\left(-c\frac{2^{2k}A_n^4 n}{m}\right)\exp\left(-c\frac{2^{2k}A_n^4 n}{n-m}\right).\end{aligned}$$

By symmetry of the right-hand side, it is enough to consider the contribution of $1\leq m\leq n/2$, which is bounded by

$$\begin{aligned}&\frac{C}{n^{d/2}}\exp\left(-c2^{2k}A_n^4\right)\sum_{1\leq m\leq n/2}\frac{1}{m^{d/2}}\exp\left(-c\frac{2^{2k}A_n^4 n}{m}\right)\\ &\leq \frac{C}{n^{d/2}}\exp\left(-c2^{2k}A_n^4\right)\sum_{k'=1}^{\lfloor \log_2 n\rfloor}\sum_{m\,:\,2^{k'}\leq n/m<2^{k'+1}}\frac{2^{k'd/2}}{n^{d/2}}\exp\left(-c2^{2k}A_n^4 2^{k'}\right)\\ &\leq \frac{C}{n^d}\exp\left(-c2^{2k}A_n^4\right)\sum_{k'=1}^{\infty}\frac{n}{2^{k'}}\exp\left(-c2^{2k}A_n^4 2^{k'}+k'd/2\log 2\right)\\ &\leq \frac{Cn}{n^d}\exp\left(-c2^{2k}A_n^4\right).\end{aligned}$$

Now, summing over $x\in\mathbb{Z}^d\setminus G$, we have that the number of the copies of the torus at dyadic scale $2^kA_n^2\sqrt{n}$ is at most $C\frac{1}{N^d}\left(2^kA_n^2\sqrt{n}\right)^d$. Hence

$$\begin{aligned}\sum_{x\in\mathbb{Z}^d\setminus G}\mathbf{P}[A(x)] &\leq \frac{Cn}{n^d}\sum_{k=0}^{\infty}\frac{1}{N^d}\left(2^kA_n^2\sqrt{n}\right)^d\exp\left(-c2^{2k}A_n^4\right)\\ &\leq \frac{C}{n^{d/2}}\frac{n}{N^d}\sum_{k=0}^{\infty}\exp\left(-c2^{2k}A_n^4+kd\log 2+2d\log A_n\right)\\ &= o(1)\frac{1}{n^{d/2}}\frac{n}{N^d}\exp\left(-100(\log\log n)^2\right).\end{aligned}$$

□

Lemma 4.8. *We have*

$$\sum_{x_1\neq x_2\in\mathbb{Z}^d}\mathbf{P}[A(x_1)\cap A(x_2)] = o(1)\frac{n}{N^d}p_n\big(y'_{\ell-1}, y'_\ell\big).$$

Proof. The summand on the left-hand side is bounded above by

$$\begin{aligned}\mathbf{P}[A(x_1)\cap A(x_2)] \leq \sum_{m_1+m_2+m_3=n}\sum_{\substack{x'_1\in\mathbf{K}+x_1N\\ x'_2\in\mathbf{K}+x_2N}} &\big[p_{m_1}\big(y'_{\ell-1}, x'_1\big)p_{m_2}\big(x'_1, x'_2\big)p_{m_3}\big(x'_2, y'_\ell\big)\\ &+p_{m_1}\big(y'_{\ell-1}, x'_2\big)p_{m_2}\big(x'_2, x'_1\big)p_{m_3}\big(x'_1, y'_\ell\big)\big].\end{aligned}$$

By symmetry it is enough to consider the first term inside the summation. The estimates are again modelled on the proof of Proposition 2.3.

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

Case 1: $m_1 + m_2 \geq n/2$ and $|x_2' - y_{\ell-1}'| \leq 2C_1\sqrt{n}\sqrt{\log n}$. In this case we can use Corollary 4.1 with $y' = y_{\ell-1}'$ and $y'' = x_2'$ to perform the summation over x_1' and x_1 and get the upper bound

$$C\frac{n}{N^d} \sum_{m_1'+m_2'=n} \sum_{x_2 \in \mathbb{Z}^d} \sum_{x_2' \in \mathbf{K}+x_2N} p_{m_1'}(y_{\ell-1}', x_2') p_{m_2'}(x_2', y_\ell'), \tag{58}$$

where we have written $m_1' = m_1 + m_2$ and $m_2' = m_3$. Using Corollary 4.1 again, this time with $y' = y_{\ell-1}'$ and $y'' = y_\ell'$, yields the upper bound

$$C\left(\frac{n}{N^d}\right)^2 p_n(y_{\ell-1}', y_\ell') = o(1)\frac{n}{N^d} p_n(y_{\ell-1}', y_\ell'). \tag{59}$$

Case 2: $m_1 + m_2 \geq n/2$ and $2C_1\sqrt{n}\sqrt{\log n} < |x_2' - y_{\ell-1}'| \leq n^{\frac{1}{2}+\varepsilon}$. We are going to use that $\varepsilon \leq 1$, which we can clearly assume. First sum over all $x_1' \in \mathbb{Z}^d$ to get the upper bound

$$Cn \sum_{m_1'+m_2'=n} \sum_{x_2 \in \mathbb{Z}^d} \sum_{x_2' \in \mathbf{K}+x_2N}{}' p_{m_1'}(y_{\ell-1}', x_2') p_{m_2'}(x_2', y_\ell'), \tag{60}$$

where the primed summation denotes the restriction $2C_1\sqrt{n}\sqrt{\log n} < |x_2' - y_{\ell-1}'| \leq n^{\frac{1}{2}+\varepsilon}$. The choice of C_1 (recall (41)) implies that $p_{m_1'}$ is $o(1/n^{1+3d/2}N^d)$. By the triangle inequality we also have $|y_\ell' - x_2'| > C_1\sqrt{n}\sqrt{\log n}$. Using the LCLT for $p_{m_2'}$ we get that

$$p_{m_2'}(x_2', y_\ell') \leq \frac{C}{(m_2')^{d/2}} \exp\left(-dC_1^2 n \log n/m_2'\right) \leq \frac{C}{n^{d/2}} \exp\left(-dC_1^2 \log n\right) \leq Cp_n(y_{\ell-1}', y_\ell'). \tag{61}$$

Substituting this bound and $p_{m_1'} = o(1/n^{1+3d/2}N^d)$ into (60), we get

$$\begin{aligned} &Cn\, o(1)\left(\frac{1}{n \cdot n^{3d/2} \cdot N^d}\right) \sum_{m_1'+m_2'=n} \sum_{x_2'}{}' p_n(y_{\ell-1}', y_\ell') \\ &\leq o(1)\left(\frac{n}{N^d}\right) p_n(y_{\ell-1}', y_\ell') \sum_{x_2'}{}' \frac{1}{(n^{1/2+\varepsilon})^d} \\ &= o\left(\frac{n}{N^d}\right) p_n(y_{\ell-1}', y_\ell'). \end{aligned}$$

Case 3: $m_1 + m_2 \geq n/2$ and $|x_2' - y_{\ell-1}'| > n^{\frac{1}{2}+\varepsilon}$. Summing over all $x_1' \in \mathbb{Z}^d$, we get the transition probability $p_{m_1+m_2}(y_{\ell-1}', x_2')$. This is stretched-exponentially small, and hence this case satisfies the required bound.

Case 4: $m_2 + m_3 \geq n/2$. By symmetry, this case can be handled analogously to Cases 1–3. ☐

4.3. Proof of Proposition 2.1

Proof of Proposition 2.1. We start with the proof of the second claim. We denote the error term in (12) by E, which we claim to satisfy $|E| \leq E_1 + E_2 + E_3 + E_4$, with

$$E_1 = \mathbf{P}_o\left[\left\{\frac{Sn}{N^d} < \left(\sqrt{\log\log n}\right)^{-1}\right\} \cup \left\{\frac{Sn}{N^d} > \log N\right\}\right],$$

$$E_2 = \mathbf{P}_o\left[\exists \ell\colon 1 \leq \ell \leq \log N \tfrac{N^d}{n} \text{ such that } Y_{\ell n} \in \varphi^{-1}(B(\mathbf{x_0}, N^\zeta))\right],$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

$$E_3 = \mathbf{P}_o\left[\exists\, t : T \le t < Sn \text{ such that } Y_t \in \varphi^{-1}(\mathbf{K})\right],$$
$$E_4 = \mathbf{P}_o\left[\exists \ell\colon 1 \le \ell \le \log N \frac{N^d}{n} \text{ such that } |Y_{\ell n} - Y_{(\ell-1)n}| > f(n) n^{\frac{1}{2}}\right].$$

Since $T \le Sn$, we have

$$\left|\mathbf{P}_o\left[Y_t \notin \varphi^{-1}(\mathbf{K}),\ 0 \le t < T\right] - \mathbf{P}_o\left[Y_t \notin \varphi^{-1}(\mathbf{K}),\ 0 \le t < Sn\right]\right| \le E_3.$$

By the Markov property, for $\left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^\tau\right) \in \mathcal{G}_{\zeta, C_1}$,

$$\prod_{\ell=1}^{\tau} \mathbf{P}_{y'_{\ell-1}}\left[Y_n = y'_\ell,\ Y_t \notin \varphi^{-1}(\mathbf{K}) \text{ for } 0 \le t < n\right]$$
$$= \mathbf{P}_o\left[\begin{matrix} Y_{n\ell} = y'_\ell \text{ for } 0 \le \ell \le \tau; Y_t \notin \varphi^{-1}(\mathbf{K}) \text{ for } 0 \le t < \tau n; Y_{n\ell} \notin \\ \varphi^{-1}(B(\mathbf{x_0}, N^\zeta)) \text{ for } 0 < \ell \le \tau \end{matrix}\right].$$

We denote the probability on the right-hand side by $p\left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^\tau\right)$. On the event of the right-hand side, since $y_\tau \in D^c$, we have $S = \tau$. We claim that

$$\left|\sum_{\left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^\tau\right) \in \mathcal{G}_{\zeta, C_1}} p\left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^\tau\right) - \mathbf{P}_o\left[Y_t \notin \varphi^{-1}(\mathbf{K}) \text{ for } 0 \le t < Sn\right]\right| \le E_1 + E_2 + E_4. \tag{62}$$

Let $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_4$ be the events in the definitions of E_1, E_2, E_4 respectively. Let

$$A\left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^\tau\right) := \left\{\begin{matrix} Y_{n\ell} = y'_\ell \text{ for } 0 \le \ell \le \tau\, ; Y_t \notin \varphi^{-1}(\mathbf{K}) \text{ for } 0 \le t < \tau n\, ; Y_{n\ell} \notin \\ \varphi^{-1}(B(\mathbf{x_0}, N^\zeta)) \text{ for } 0 < \ell \le \tau \end{matrix}\right\}.$$

Then we have

$$\begin{aligned} &\mathbf{P}_o\left[\left\{Y_t \notin \varphi^{-1}(\mathbf{K}) \text{ for } 0 \le t < Sn\right\} \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_4^c\right] \\ &= \sum_{\frac{N^d}{n\sqrt{\log\log n}} \le \tau \le \frac{N^d \log N}{n}} \mathbf{P}_o\left[\{S = \tau\} \cap \left\{Y_t \notin \varphi^{-1}(\mathbf{K}) \text{ for } 0 \le t < \tau n\right\} \cap \mathcal{E}_2^c \cap \mathcal{E}_4^c\right] \\ &= \sum_{\left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^\tau\right) \in \mathcal{G}_{\zeta, C_1}} \mathbf{P}_o\left[A\left(\tau, (y_\ell, \mathbf{y}_\ell)_{\ell=1}^\tau\right) \cap \mathcal{E}_2^c \cap \mathcal{E}_4^c\right]. \end{aligned}$$

From the above, the claim (62) follows.

The proof of the second claim of Proposition 2.1 follows subject to Lemmas 4.9, 4.11, and 4.12 below, which show that $E_j \to 0$ for $j = 2, 3, 4$. We have already shown $E_1 \to 0$ in Lemma 2.2.

Similarly, the first claim of the proposition follows from Lemmas 2.2, 4.9, and 4.12. □

We bound E_2, E_3, and E_4 in the following lemmas.

Lemma 4.9. *We have $E_2 \le C \log N \frac{N^{d\zeta}}{n}$ for some C. Consequently, $E_2 \to 0$.*

Proof. Since the number of points in $B(\mathbf{x_0}, N^\zeta)$ is $O\left(N^{d\zeta}\right)$, and since we are considering times after the first stretch, the random walk is well mixed, so by Lemma 2.1 the probability

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

of visiting any point in the torus is $O\big(1/N^d\big)$. Using a union bound we have

$$E_2 \le C \log N \frac{N^d}{n} N^{d\zeta} \frac{1}{N^d} = C \log N \frac{N^{d\zeta}}{n}.$$

Since $\zeta < \delta/d$, we have

$$E_2 \to 0, \quad \text{as } N \to \infty. \qquad \square$$

Before we bound the error term E_3, we first introduce the following lemma. Let $T_0^{(i)} = \inf\left\{t \ge 0 : Y_t^{(i)} = 0\right\}$, where we recall that $Y^{(i)}$ denotes the ith coordinate of the d-dimensional lazy random walk. We will denote by t_0 an instance of $T_0^{(i)}$.

Lemma 4.10. *For all $1 \le i \le d$, for any integer y such that $-t_0 \le y < 0$ and $0 < t_0 \le n$ we have*

$$\mathbf{E}_y\left[Y_n^{(i)} \mid T_0^{(i)} = t_0,\ Y_n^{(i)} > 0\right] \le C n^{\frac{1}{2}},$$
$$\mathbf{E}_y\left[\left(Y_n^{(i)}\right)^2 \mid T_0^{(i)} = t_0,\ Y_n^{(i)} > 0\right] \le C n.$$

Proof. Using the Markov property at time t_0, we get

$$\mathbf{E}_y\left[Y_n^{(i)} \mid T_0^{(i)} = t_0, Y_n^{(i)} > 0\right] = \mathbf{E}_0\left[Y_{n-t_0}^{(i)} \mid Y_{n-t_0}^{(i)} > 0\right] = \frac{\mathbf{E}_0\left[Y_{n-t_0}^{(i)} \mathbf{1}_{\left\{Y_{n-t_0}^{(i)} > 0\right\}}\right]}{\mathbf{P}_0\left[Y_{n-t_0}^{(i)} > 0\right]}$$
$$\le C_0 \left(\mathbf{E}_0\left[\left(Y_{n-t_0}^{(i)}\right)^2\right]\right)^{\frac{1}{2}} \le C(n-t_0)^{\frac{1}{2}} \le C n^{\frac{1}{2}},$$

where the third step is due to the Cauchy–Schwarz inequality and $\mathbf{P}_0\left[Y_{n-t_0}^{(i)} > 0\right] \ge c_0$ for some $c_0 > 0$, and the second-to-last step is due to $\mathbf{E}_0\left[\left(Y_{n-t_0}^{(i)}\right)^2\right] = (n-t_0)/2d$.

We can similarly bound the conditional expectation of $\left(Y_n^{(i)}\right)^2$ as follows:

$$\mathbf{E}_y\left[\left(Y_n^{(i)}\right)^2 | T_0^{(i)} = t_0, Y_n^{(i)} > 0\right] = \mathbf{E}_0\left[\left(Y_{n-t_0}^{(i)}\right)^2 | Y_{n-t_0}^{(i)} > 0\right]$$
$$= \frac{\mathbf{E}_0\left[\left(Y_{n-t_0}^{(i)}\right)^2 \mathbf{1}_{\left\{Y_{n-t_0}^{(i)} > 0\right\}}\right]}{\mathbf{P}_0\left[Y_{n-t_0}^{(i)} > 0\right]} \le C_0\left(\mathbf{E}_0\left[\left(Y_{n-t_0}^{(i)}\right)^2\right]\right) \le C(n-t_0) \le Cn.$$

$\square$

Lemma 4.11. *We have $E_3 \to 0$ as $N \to \infty$.*

Proof. First we are going to bound the time difference between T and Sn. We are going to consider separately the cases when Y_T is in each face of the cube $(-L, L)^d$. Assume that we have $Y_T^{(i)} = L$ for some $1 \le i \le d$. (The arguments needed are very similar when $Y_T^{(i)} = -L$ for some $1 \le i \le d$, and these will not be given.)

Let us consider the lazy random walk $(Y_t)_{t\geq 0}$ in multiples of n steps. Let

$$s_1 = \min\{\ell n : \ell n \geq T\} - T,$$

and similarly, let

$$s_{r+1} = rn + \min\{\ell n : \ell n \geq T\} - T, \quad r \geq 1.$$

We let $M_0 = L - Y^{(i)}_{T+s_1}$ and $M_r = L - Y^{(i)}_{T+s_{r+1}}$ for $r \geq 1$. We have that $(M_r)_{r\geq 0}$ is a martingale. Let $\tilde{S} = \inf\{r \geq 0 : M_r \leq 0\}$, and we are going to bound $\mathbf{P}\big[\tilde{S} > N^{\varepsilon_1}\big]$ for some small ε_1 that we will choose in the course of the proof. We are going to adapt an argument in [10, Proposition 17.19] to this purpose.

Define

$$T_h = \inf\big\{r \geq 0 : M_r \leq 0 \text{ or } M_r \geq h\big\},$$

where we set $h = \sqrt{n}\sqrt{N^{\varepsilon_1}}$. Let $(\mathcal{F}_r)_{r\geq 0}$ denote the filtration generated by $(M_r)_{r\geq 0}$. We have

$$\mathrm{Var}\big(M_{r+1} \mid \mathcal{F}_r\big) = n\sigma^2 \quad \text{for all } r \geq 0; \tag{63}$$

here, recall that σ^2 is the variance of $Y^{(i)}_1$.

We first estimate $\mathbf{E}\big[M_0 \mid \tilde{S} > 0\big]$. Since $0 \leq s_1 < n$, by the same argument as in Lemma 4.10 we have that

$$\mathbf{E}\Big[M_0 \mid Y^{(i)}_{T+s_1} < L\Big] = \mathbf{E}\Big[L - Y^{(i)}_{T+s_1} \mid L - Y^{(i)}_{T+s_1} > 0\Big] \leq Cn^{\frac{1}{2}}.$$

We first bound $\mathbf{P}\big[M_{T_h} \geq h \mid M_0\big]$. Since $\big(M_{r\wedge T_h}\big)$ is bounded, by the optional stopping theorem, we have

$$\begin{aligned}
M_0 = \mathbf{E}\big[M_{T_h} | M_0\big] &= \mathbf{E}\left[M_{T_h}\mathbf{1}_{\{M_{T_h}\leq 0\}} | M_0\right] + \mathbf{E}\left[M_{T_h}\mathbf{1}_{\{M_{T_h}\geq h\}} | M_0\right] \\
&= -m_-(h) + \mathbf{E}\left[M_{T_h}\mathbf{1}_{\{M_{T_h}\geq h\}} | M_0\right] \\
&\geq -m_-(h) + h\mathbf{P}\left[M_{T_h} \geq h | M_0\right],
\end{aligned}$$

where we denote $\mathbf{E}\Big[M_{T_h}\mathbf{1}_{\{M_{T_h}\leq 0\}} | M_0\Big]$ by $-m_-(h) \leq 0$, and the last step is due to $M_{T_h}\mathbf{1}_{\{M_{T_h}\geq h\}} \geq h\mathbf{1}_{\{M_{T_h}\geq h\}}$. Hence, we have

$$M_0 + m_-(h) \geq h\mathbf{P}\left[M_{T_h} \geq h | M_0\right].$$

We bound $m_-(h)$ using Lemma 4.10:

$$m_-(h) \leq \max_{y\leq L} \mathbf{E}_y\Big[Y^{(i)}_n - L | Y^{(i)}_n > L\Big] \leq Cn^{\frac{1}{2}}.$$

Hence, we have

$$\mathbf{P}[M_{T_h} \geq h \mid M_0] \leq \frac{M_0}{h} + \frac{Cn^{\frac{1}{2}}}{h}.$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

We now estimate $\mathbf{P}[T_h \geq r \mid M_0]$. Let $G_r = M_r^2 - hM_r - \sigma^2 nr$. The sequence (G_r) is a martingale by (63). We can bound both the 'overshoot' above h and the 'undershoot' below 0 by Lemma 4.10. For the 'undershoot' below 0 we have

$$\begin{aligned}\mathbf{E}\Big[\big(M_{T_h} - h\big)M_{T_h} \mid M_{T_h} \leq 0, M_0\Big] &= \mathbf{E}\Big[M_{T_h}^2 \mid M_{T_h} \leq 0, M_0\Big] + \mathbf{E}\Big[-hM_{T_h} \mid M_{T_h} \leq 0, M_0\Big] \\ &\leq Cn + Chn^{1/2}.\end{aligned}$$

For the 'overshoot' above h, write $M_{T_h} =: N_{T_h} + h$; then we have

$$\big(M_{T_h} - h\big)M_{T_h} = N_{T_h}\big(h + N_{T_h}\big).$$

Hence

$$\begin{aligned}\mathbf{E}\Big[\big(M_{T_h} - h\big)M_{T_h} \mid M_{T_h} \geq h, M_0\Big] &= \mathbf{E}\Big[hN_{T_h} \mid N_{T_h} \geq 0, M_0\Big] + \mathbf{E}\Big[N_{T_h}^2 \mid N_{T_h} \geq 0, M_0\Big] \\ &\leq Chn^{1/2} + Cn.\end{aligned}$$

For $r < T_h$, we have $(M_r - h)M_r < 0$. Therefore, we have

$$\mathbf{E}\left[M_{r\wedge T_h}^2 - hM_{r\wedge T_h} | M_0\right] \leq Chn^{1/2} + Cn.$$

Since $(G_{r\wedge T_h})$ is a martingale,

$$\begin{aligned}-hM_0 \leq G_0 \leq \mathbf{E}\big[G_{r\wedge T_h} | M_0\big] &= \mathbf{E}\Big[M_{r\wedge T_h}^2 - hM_{r\wedge T_h} | M_0\Big] - \sigma^2 n\mathbf{E}[r \wedge T_h | M_0] \\ &\leq Cn^{\frac{1}{2}}h + Cn - \sigma^2 n\mathbf{E}[r \wedge T_h | M_0].\end{aligned}$$

We conclude that

$$\mathbf{E}[r \wedge T_h \mid M_0] \leq \frac{h\Big(M_0 + Cn^{\frac{1}{2}}\Big) + Cn}{\sigma^2 n}.$$

Letting $r \to \infty$, by the monotone convergence theorem,

$$\mathbf{E}[T_h \mid M_0] \leq \frac{h\Big(M_0 + Cn^{\frac{1}{2}}\Big) + Cn}{\sigma^2 n},$$

where $h = \sqrt{n}\sqrt{N^{\varepsilon_1}}$. This gives

$$\mathbf{P}[T_h > N^{\varepsilon_1} | M_0] \leq \frac{1}{N^{\varepsilon_1}}\left[\frac{\sqrt{n}\sqrt{N^{\varepsilon_1}}M_0 + Cn\sqrt{N^{\varepsilon_1}} + Cn}{\sigma^2 n}\right].$$

Taking expectations of both sides, we have

$$\begin{aligned}\mathbf{P}[T_h > N^{\varepsilon_1}] &\leq \frac{1}{N^{\varepsilon_1}}\left[\frac{\sqrt{n}\sqrt{N^{\varepsilon_1}}\mathbf{E}M_0 + Cn\sqrt{N^{\varepsilon_1}} + Cn}{\sigma^2 n}\right] \\ &= \frac{\mathbf{E}M_0}{\sigma^2\sqrt{n}\sqrt{N^{\varepsilon_1}}} + \frac{C}{\sigma^2\sqrt{N^{\varepsilon_1}}} + \frac{C}{\sigma^2 N^{\varepsilon_1}} \leq \frac{C}{\sqrt{N^{\varepsilon_1}}}.\end{aligned}$$

Combining the above bounds, we get

$$\mathbf{P}\big[\tilde{S} > N^{\varepsilon_1}\big] \leq \mathbf{P}[M_{T_h} \geq h] + \mathbf{P}[T_h > N^{\varepsilon_1}] \leq \frac{\mathbf{E}[M_0]}{h} + \frac{Cn^{\frac{1}{2}}}{h} + \frac{C}{\sqrt{N^{\varepsilon_1}}} \leq \frac{C}{\sqrt{N^{\varepsilon_1}}}.$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

We now bound the probability that a copy of $\mathbf{K}$ is hit between times T and $s_{N^{\varepsilon_1}}$.

We first show that the probability that the lazy random walk on the torus is in the ball $B(\mathbf{x}_0, N^\zeta)$ at time T goes to 0. Indeed, we have

$$\mathbf{P}_o\left[Y_T \in \varphi^{-1}\left(B(\mathbf{x}_0, N^\zeta)\right)\right] = \sum_{y' \in \partial(-L,L)^d \cap \varphi^{-1}(B(\mathbf{x}_0, N^\zeta))} \mathbf{P}_o\left[Y_T = y'\right]$$
$$\leq CN^{\zeta(d-1)} \frac{L^{d-1}}{N^{d-1}} \frac{C}{L^{d-1}} = CN^{(\zeta-1)(d-1)},$$

where we have $\zeta < \delta/d < 1$, so the last expression goes to 0. Here we used that $\mathbf{P}_o[Y_T = y'] \leq C/L^{d-1}$, for example using a half-space Poisson kernel estimate [9, Theorem 8.1.2]. As for the number of terms in the sum, we have that there are CL^{d-1}/N^{d-1} copies of the torus within ℓ_∞ distance $\leq N$ of the boundary $\partial(-L, L)^d$. Considering the intersection of the union of balls $\varphi^{-1}(B(\mathbf{x}_0, N^\zeta))$ and the boundary $\partial(-L, L)^d$, the worst case is that within a single copy of the torus the intersection has size at most $CN^{\zeta(d-1)}$.

Condition on the location y' of the walk at the exit time T. For $y' \notin \varphi^{-1}\left(B(\mathbf{x}_0, N^\zeta)\right)$, we first bound the probability of hitting $\mathbf{K}$ between the times between 0 and s_2. After time s_2, the random walk is well mixed, and we can apply a simpler union bound.

We thus have the upper bound

$$\sum_{t=0}^{s_2} \sum_{x' \in \varphi^{-1}(\mathbf{K})} p_t(y', x') \leq \mathbf{P}\left[\max_{0 \leq t \leq s_2} |Y_t^{(i)} - y'| \geq n^{\frac{1}{2}+\varepsilon}\right] + \sum_{\substack{x' \in \varphi^{-1}(\mathbf{K}) \\ |x'-y'| \leq n^{\frac{1}{2}+\varepsilon}}} G(y', x').$$

The first term is stretched-exponentially small by the martingale maximal inequality (3). The Green's function term is bounded by Lemma 4.1(iii).

After time s_2, by Lemma 2.1, we have that

$$\sum_{t=s_2}^{s_{N^{\varepsilon_1}}} \mathbf{P}_y\left[Y_t \in \varphi^{-1}(K)\right] \leq n \cdot N^{\varepsilon_1} |\mathbf{K}| \frac{C}{N^d} = C\, N^{\delta+\varepsilon_1-d}.$$

Therefore, combining the above upper bounds, we have the required result:

$$E_3 \leq C \cdot N^{-\frac{\varepsilon_1}{2}} + C \cdot N^{\delta-d+2\delta\varepsilon} + C \cdot N^{\delta-d+\varepsilon_1} \to 0, \quad \text{as } N \to \infty,$$

if ε and ε_1 are sufficiently small. □

Lemma 4.12. *We have $E_4 \leq Ce^{-cf(n)^2} \frac{N^d \log N}{n}$ for some C. Consequently, there exists C_1 such that if $f(n) \geq C_1\sqrt{\log N}$, then $E_4 \to 0$.*

Proof. By the martingale maximal inequality (3), we have that

$$E_4 \leq Ce^{-cf(n)^2} \frac{N^d \log N}{n}.$$

Taking, say, $C_1 > \sqrt{d/c}$ implies that if $f(n) \geq C_1\sqrt{\log N}$, we have

$$E_4 \to 0, \quad \text{as } N \to \infty. \qquad \square$$

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

4.4. Proof of Proposition 2.4

Proof. By the martingale maximal inequality (3) used in the second step, we have

$$\mathbf{P}_{y'_{\ell-1}}\Big[\big|Y_n - y'_{\ell-1}\big| > \sqrt{n}\big(10\log\log n\big)\Big] = \mathbf{P}_0\Big[|Y_n| > \sqrt{n}\big(10\log\log n\big)\Big] \le \exp\Big(-c100\big(\log\log n\big)^2\Big).$$

Hence we have

$$\mathbf{E}\left[\#\left\{1 \le \ell \le \frac{N^d}{n}C_1\log N : |Y_{n\ell} - Y_{n(\ell-1)}| > 10\sqrt{n}\log\log n\right\}\right] \le \frac{N^d}{n}C_1\log N\exp\Big(-c\big(\log\log n\big)^2\Big) \le \frac{N^d}{n}C\exp\Big(-(c/2)\big(\log\log n\big)^2\Big).$$

By Markov's inequality, it follows that

$$\mathbf{P}\left[\#\left\{1 \le \ell \le \frac{N^d}{n}C_1\log N : |Y_{n\ell} - Y_{n(\ell-1)}| > 10\sqrt{n}\log\log n\right\} \ge \frac{N^d}{n}\big(\log\log n\big)^{-1}\right] \le \frac{\frac{N^d}{n}C\exp\Big(-(c/2)\big(\log\log n\big)^2\Big)}{\frac{N^d}{n}\big(\log\log n\big)^{-1}} \le C\frac{\exp\Big(-(c/2)\big(\log\log n\big)^2\Big)}{\big(\log\log n\big)^{-1}} \to 0,$$

as $N \to \infty$. □

Acknowledgements

We thank two anonymous referees for their constructive criticism.

Funding information

The research of M. Sun was supported by an EPSRC doctoral training grant to the University of Bath with project reference EP/N509589/1/2377430.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ČERNÝ, J. AND TEIXEIRA, A. (2016). Random walks on torus and random interlacements: Macroscopic coupling and phase transition. *Ann. Appl. Prob.* **26**, 2883–2914.
[2] DHAR, D. (2006). Theoretical studies of self-organized criticality. *Physica A* **369**, 29–70.
[3] DREWITZ, A., RÁTH, B. AND SAPOZHNIKOV, A. (2014). *An Introduction to Random Interlacements*, 1st edn. Springer, Berlin.
[4] DURRETT, R. (2019). *Probability: Theory and Examples*, 5th edn. Cambridge University Press.
[5] HEBISCH, W. AND SALOFF-COSTE, L. (1993). Gaussian estimates for Markov chains and random walks on groups. *Ann. Prob.* **21**, 673–709.
[6] JÁRAI, A. A. (2018). Sandpile models. *Prob. Surveys* **15**, 243–306.
[7] JÁRAI, A. A. AND SUN, M. (2019). Toppling and height probabilities in sandpiles. *J. Statist. Mech.* **2019**, article no. 113204.
[8] LAWLER, G. F. (2013). *Intersections of Random Walks*. Birkhäuser, New York.
[9] LAWLER, G. F. AND LIMIC, V. (2010). *Random Walk: A Modern Introduction*. Cambridge University Press.

https://doi.org/10.1017/apr.2023.24 Published online by Cambridge University Press

[10] LEVIN, D. A., PERES, Y. AND WILMER, E. L. (2017). *Markov Chains and Mixing Times*, 2nd edn. American Mathematical Society, Providence, RI.

[11] REDIG, F. (2006). Mathematical aspects of the abelian sandpile model. In *Mathematical Statistical Physics: Lecture Notes of the Les Houches Summer School 2005*, eds A. BOVIER, F. DUNLOP, A. VAN ENTER, F. den Hollander and J. Dalibard, Elsevier, Amsterdam, pp. 657–729.

[12] SZNITMAN, A.-S. (2009). Random walks on discrete cylinders and random interlacements. *Prob. Theory Relat. Fields* **145**, 143–175.

[13] SZNITMAN, A.-S. (2010). Vacant set of random interlacements and percolation. *Ann. Math.* **171**, 2039–2087.

[14] SZNITMAN, A.-S. (2012). *Topics in Occupation Times and Gaussian Free Fields*. European Mathematical Society, Zürich.

[15] WINDISCH, D. (2008). Random walk on a discrete torus and random interlacements. *Electron. Commun. Prob.* **13**, 140–150.