J. Eur. Math. Soc. 10, 133–172

Itai Benjamini · Alain-Sol Sznitman

Giant component and vacant set for random walk on a discrete torus

Received November 22, 2006

Abstract. We consider random walk on a discrete torus E of side-length N, in sufficiently high dimension d. We investigate the percolative properties of the vacant set corresponding to the collection of sites which have not been visited by the walk up to time uN^d. We show that when u is chosen small, as N tends to infinity, there is with overwhelming probability a unique connected component in the vacant set which contains segments of length $\text{const} \log N$. Moreover, this connected component occupies a non-degenerate fraction of the total number of sites N^d of E, and any point of E lies within distance N^β of this component, with β an arbitrary positive number.

0. Introduction

We investigate random walk on a d-dimensional torus of large side-length N, and we are interested in the set of points that have not been visited by the walk up to times of order N^d. This time scale is much shorter than the typical time it takes the walk to cover the discrete torus. Indeed, the cover time of the discrete torus is known to be of order $N^d \log N$ when $d \geq 3$, and $N^2(\log N)^2$ when $d = 2$ (cf. [1], [2], [4], [5], [6], and the references therein, for this and much more). In fact, when $d \geq 3$, and u is an arbitrary positive number, the probability that the walk visits a given point of the discrete torus up to time uN^d remains bounded away from 0 and 1 as N tends to infinity. This makes the time scale N^d an appropriate choice to discuss the percolative properties of the vacant set left by the walk. Incidentally these questions are closely related to the analysis of the disconnection time of a discrete cylinder by a random walk, recently investigated in [7], [15]. The main object of this work is to discuss the typical presence of a well-defined giant connected component in the vacant set left by the walk by time uN^d, for large N, when the dimension d is large enough, and u suitably small. We expect a different behavior when u is large, but this work does not present results in this direction. We believe some of our methods and results are pertinent to improved bounds of the disconnection time of a discrete cylinder by a random walk recently derived in [7] (see also [15]).

Before discussing our results any further, we describe the model in more detail. We consider $d \geq 3$, $N \geq 1$, and denote by E the d-dimensional discrete torus of side-

I. Benjamini: Department of Mathematics, The Weizmann Institute of Science, POB 26, Rehovot 76100, Israel; e-mail: itai.benjamini@weizmann.ac.il

A.-S. Sznitman: Departement Mathematik, ETH Zürich, CH-8092 Zürich, Switzerland; e-mail: sznitman@math.ethz.ch

length N:

$$E = (\mathbb{Z}/N\mathbb{Z})^d. \tag{0.1}$$

We write P, resp. P_x when $x \in E$, for the law on $E^{\mathbb{N}}$ endowed with the product σ-algebra $\mathcal{F}$, of simple random walk on E started with uniform distribution, resp. at x. We let $X_.$ stand for the canonical process on $E^{\mathbb{N}}$, and $X_{[0,t]}$ for the set of sites visited by the walk up to time $[t]$:

$$X_{[0,t]} = \{z \in E : \text{for some } 0 \le n \le t,\ X_n = z\} \quad \text{for } t \ge 0. \tag{0.2}$$

Our main focus is on the percolative properties of the vacant set $E \setminus X_{[0,uN^d]}$ left by the walk up to time uN^d, when N is large and $u > 0$ is some fixed positive number. We show that when $d \ge 4$ and u is suitably small, the vacant set by time uN^d typically contains a profusion of segments of logarithmic size in N, for large N. More precisely, for $K > 0$, $0 < \beta < 1$, $t \ge 0$, we define the event which specifies that for every point of E there is in each coordinate direction, within N^β steps, a segment of length $[K \log N]$ in the vacant set left by the walk at time t:

$$\begin{aligned}\mathcal{V}_{K,\beta,t} = \{&\text{for all } x \in E,\ 1 \le j \le d, \text{ for some } 0 \le m < N^\beta, \\ &X_{[0,t]} \cap \{x + (m + [0, [K \log N]])e_j\} = \emptyset\},\end{aligned} \tag{0.3}$$

where $(e_i)_{1 \le i \le d}$ stands for the canonical basis of $\mathbb{R}^d$. We show in Theorem 1.2 that for d, K, β as above,

$$\lim_N P[\mathcal{V}_{K,\beta,uN^d}] = 1 \quad \text{for small } u > 0. \tag{0.4}$$

We also show in Proposition 1.1 that when $d \ge 3$, for $u > 0$,

$$e^{-cu} \le \liminf_N P[0 \notin X_{[0,uN^d]}] \le \limsup_N P[0 \notin X_{[0,uN^d]}] \le e^{-c'u}, \tag{0.5}$$

with c, c' suitable positive dimension dependent constants (more is known, see [2, Chapter 3, Proposition 20 and Chapter 13, Proposition 8]). This feature motivates the interest of the time scale N^d in the investigation of the vacant set left by the walk. We sharpen this result by showing in Corollary 4.5 that

$$\lim_N P[e^{-cu} \le |E \setminus X_{[0,uN^d]}|/N^d \le e^{-c'u}] = 1 \quad \text{for } u > 0, \tag{0.6}$$

where for $A \subseteq E$, $|A|$ denotes the cardinality of A.

When the dimension d is suitably large, i.e. $d \ge d_0$ (cf. (2.41)), we introduce a dimension dependent constant c_0, and events $\mathcal{G}_{\beta,t} \subseteq \mathcal{V}_{c_0,\beta,t}$, increasing with $\beta \in (0, 1)$, such that for any such β:

$$\begin{aligned}&\text{(i) } \lim_N P[\mathcal{G}_{\beta,uN^d}] = 1 \text{ for small } u > 0, \\ &\text{(ii) on } \mathcal{G}_{\beta,t} \text{ there is a unique connected component } O \text{ in } E \setminus X_{[0,t]} \\ &\quad \text{which contains segments of length } L_0 = [c_0 \log N]\end{aligned} \tag{0.7}$$

(see (2.53) for a more general claim). The connected component O is thus well-defined on the nested events $\mathcal{G}_{\beta,t}$. In view of (0.3) and since $\mathcal{G}_{\beta,t} \subseteq \mathcal{V}_{c_0,\beta,t}$, the connected component

O is ubiquitous on E. We refer to it as the *giant component*. We also show in Corollary 4.6 that when u is small, O typically has a non-degenerate volume in E. More precisely, we prove that for $d \geq d_0$, $\beta, \gamma \in (0, 1)$,

$$\lim_N P[\mathcal{G}_{\beta, uN^d} \cap \{|O| \geq \gamma N^d\}] = 1 \quad \text{when } u > 0 \text{ is small.} \tag{0.8}$$

However, our results do not rule out the possible existence of other components of the vacant set with non-degenerate volume as well (cf. Remark 4.7). In fact, the present work raises many questions. How do percolative properties of the vacant set compare to the picture stemming from Bernoulli bond-percolation? Is there a small u regime with typically one single giant component and all other components of small volume and size, a large u regime with only small connected components, and in between a critical regime (see for instance [10] and references therein)? Simulations performed when $d = 3, 4, 5, 6, 7$ seem to support this picture, with a critical threshold located near $u = 3$. If such a critical regime can be extracted, do components in the vacant set in this regime inherit some of the invariance properties of Brownian motion viewed as a scaling limit of simple random walk? What are the relevant values of the dimension d? It is maybe instructive to consider these problems also on other graphs, such as expanders (where a small u regime with some giant component and a large u regime with only small components can easily be established), random d-regular graphs, hypercubes etc. (see [3] for a study of percolation on such graphs). These are just a few examples of the many questions raised by the present article.

We now try to describe some of the ideas and methods involved in the proof of (0.4), (0.7), (0.8).

Behind (0.4) lies a type of coupon-collector heuristics. We show in Proposition 1.1 that up to time uN^d about const uN^{d-2} excursions take place in and out of two concentric balls centered at the origin with radius some suitable fraction of N. At most const uN^β of these excursions hit a given segment of length N^β starting at the origin. Chopping this segment into $M = N^\beta/[K \log N]$ segments of length $[K \log N]$, and neglecting the possible hits of more than one segment by one such excursion, a coupon-collector heuristics (cf. Durrett [8, Chapter 2, Example 6.6]) makes it plausible that it takes about $M \log M \sim (\beta/K)N^\beta$ such excursions to hit each of these segments. However, when u is chosen small, const $uN^\beta \ll (\beta/K)N^\beta$, and not all segments can be hit by the walk up to time N^d. The above lines describe the intuition behind the proof of (0.4) in Theorem 1.2.

The key to the uniqueness statement in (0.7) is an exponential estimate proved in Theorem 2.1. It shows in particular that when $d \geq 5$, and $\lambda > 0$ is such that

$$e^{2\lambda}\left(\frac{2}{d} + \left(1 - \frac{2}{d}\right) q(d-2)\right) < 1, \tag{0.9}$$

where for any integer $\nu \geq 1$,

$$q(\nu) = \text{the return probability to the origin of simple random walk on } \mathbb{Z}^\nu, \tag{0.10}$$

then for $N \geq N(d, \lambda)$ and $u \leq u(d, \lambda)$,

$$P[X_{[0,uN^d]} \supseteq A] \leq \exp\{-\lambda |A|\} \tag{0.11}$$

for any subset A of E contained in the canonical projection F on E of a two-dimensional affine plane generated by two coordinate directions in $\mathbb{Z}^d$. When $e^\lambda > 7$ can be achieved (this is the requirement which specifies d_0, cf. (2.41)), the exponential bound (0.11) combined with a Peierl-type argument yields in Corollary 2.5 the key uniqueness statement behind (0.7). The claim (0.7) is then proved in Corollary 2.6. We also explain in Remark 2.4 why a restriction on the class of sets A that appear in (0.11) is needed. There is an independent interest to the above exponential bound: a variation of it and of (0.4) should lead to a sharpening of the lower bounds on the disconnection time of discrete cylinders $(\mathbb{Z}/N\mathbb{Z})^d \times \mathbb{Z}$ derived in [7], at least when d is large enough.

To prove (0.8), in essence we control fluctuations of the proportion of sites in E which at time uN^d are connected by a vacant path in some two-dimensional F, as below (0.11), to sites at distance $[c_0 \log N]$. Such sites belong to the giant component O when $\mathcal{G}_{\beta,uN^d}$ occurs (cf. (2.56)). This leads us to develop estimates on the covariance of "local functions" of the vacant sites left by the walk up to time of order uN^d in the neighborhood of two sufficiently distant point on the torus (see Proposition 4.2). Qualitatively similar issues appear for instance in [6]. To this end we develop in Theorem 3.1 a bound on the total variation norm between the law of a suitable "limit model" and the law $Q_{u,w}$ of a recentered excursion of the walk. This excursion runs from the time of the first up to the last visit to $C(x_1) \cup C(x_2)$, where the $C(x_i)$ are boxes of side-length $2L$ centered at x_i, $i = 1, 2$, in E, with mutual distance at least $2r+3$, where $r \geq 10L$, and the walk is conditioned to start at a point u at distance at least r from $\{x_1, x_2\}$ and exit the r-neighborhood of $\{x_1, x_2\}$ at the point w, and stop there. The "limit model" with law Q corresponds to excursions of the simple random walk on $\mathbb{Z}^d$ starting with the normalized harmonic measure viewed from infinity of the box C centered at the origin with side-length $2L$, stopped at its last visit of C. In Theorem 3.1 we show that

$$\|Q_{u,w} - Q\|_{TV} \leq c\frac{L^2}{r}, \tag{0.12}$$

where c is a dimension dependent constant and $\|\cdot\|_{TV}$ the total variation norm. This estimate is of independent interest and can be straightforwardly extended to the case of finitely many points x_i (cf. Remark 3.2). Our main control on fluctuations of spatial averages on E of local functions is then stated in Theorem 4.3, and enables us to show (0.6) in Corollary 4.5, and (0.8) in Corollary 4.6. In Corollary 4.8 we also show that when $d \geq 3$, the largest cube contained in the vacant set at time uN^d typically has size of order $(\log N)^{1/(d-2)}$, for large N. This should be contrasted with the case of Bernoulli bond-percolation on the torus, where for large N the largest cube contained in a cluster typically has much smaller size of order $(\log N)^{1/d}$.

Let us now describe the organization of this article.

In Section 1, we introduce some further notation, and mainly provide the proof of (0.4) in Theorem 1.2. On the way we show (0.5) in Proposition 1.1.

In Section 2 we prove a more general version of (0.11) in Theorem 2.1, and use it in Corollaries 2.5, 2.6 to prove (0.7), and thereby construct the giant component in the vacant set, which is shown to be typical when $d \geq d_0$ and u is small enough.

In Section 3, we obtain the total variation estimate (0.12) in Theorem 3.1. This comes as a preparation for the control of fluctuations of certain spatial averages of local functions in the next section.

In Section 4, we show (0.8) in Corollary 4.6, the simpler (0.6) in Corollary 4.5, and the controls on the largest cube contained in the vacant set in Corollary 4.8. The variance bounds of Proposition 4.2 make strong use of Theorem 3.1. Our general control on fluctuations of averages of local functions appears in Theorem 4.3.

Finally, throughout the text, c or c' denote positive constants which solely depend on d, with values that change from place to place. The numbered constants $c_0, c_1, \dots$ are fixed and refer to the value at their first place of appearance in the text. Dependence of constants on additional parameters appears in the notation. For instance, $c(K, \beta)$ denotes a positive constant depending on d, K, β.

1. Ubiquity of vacant segments of logarithmic size

The main object of this section is to show that when $d \geq 4$, for large N, up to times that are small multiples of N^d the vacant set left by the walk on the discrete torus E contains with overwhelming probability segments of size of order $\log N$ in the vicinity of each point of E (cf. Theorem 1.2 and (0.4)). We also prove the estimate (0.5) on the probability that a point belongs to the vacant set up to time uN^d, with $d \geq 3$ (cf. Proposition 1.1). We first need some additional notation.

We denote by $|\cdot|$ and $|\cdot|_\infty$ the Euclidean and ℓ^∞-distances on $\mathbb{Z}^d$, or the corresponding distances on E. We write $B(x, r)$ for the closed ball relative to $|\cdot|_\infty$, with radius $r \geq 0$ and center x in $\mathbb{Z}^d$ or E. We denote by $S(x, r)$ the corresponding $|\cdot|_\infty$-sphere. We say that x, y in $\mathbb{Z}^d$ or E are *neighbors*, resp. $\star$-*neighbors*, if $|x - y| = 1$, resp. $|x - y|_\infty = 1$. The notions of connected or $\star$-connected subsets of $\mathbb{Z}^d$ or E are then defined accordingly, and so are the notions of nearest neighbor path or $\star$-nearest neighbor path on $\mathbb{Z}^d$ or E. For A, B subsets of $\mathbb{Z}^d$ or E, we denote by $A + B$ the subset of points of the form $x + y$ with $x \in A$, $y \in B$. When U is a subset of $\mathbb{Z}^d$ or E, we let $|U|$ stand for the cardinality of U, and ∂U for the boundary of U:

$$\partial U = \{x \in U^c : \exists y \in U,\ |x - y| = 1\}. \tag{1.1}$$

We denote by π_E the canonical projection from $\mathbb{Z}^d$ onto E. For $1 \leq m \leq d$, we write $\mathcal{L}_m$ for the collection of subsets of E that are projections under π_E of affine lattices in $\mathbb{Z}^d$ generated by m distinct vectors of the canonical basis:

$$\begin{aligned} \mathcal{L}_m = \Big\{ F \subseteq E : &\text{ for some } I \subseteq \{1, \dots, d\} \text{ with } |I| = m, \text{ and some } y \in \mathbb{Z}^d, \\ & F = \pi_E\Big(y + \sum_{i \in I} \mathbb{Z} e_i\Big)\Big\}, \end{aligned} \tag{1.2}$$

where as below (0.3), $(e_i)_{1 \leq i \leq d}$ denotes the canonical basis of $\mathbb{R}^d$.

We let $(\theta_n)_{n\ge 0}$ and $(\mathcal{F}_n)_{n\ge 0}$ stand for the canonical shift on $E^{\mathbb{N}}$ and the filtration of the canonical process. For $U \subseteq E$, H_U and T_U stand for the entrance time and exit time in or from U:

$$H_U = \inf\{n \ge 0 : X_n \in U\}, \quad T_U = \inf\{n \ge 0 : X_n \notin U\}. \tag{1.3}$$

We write $\widetilde{H}_U$ for the hitting time of U:

$$\widetilde{H}_U = \inf\{n \ge 1 : X_n \in U\}. \tag{1.4}$$

When $U = \{x\}$, we write as a subscript x in place of $\{x\}$, for simplicity. Given $A \subseteq \widetilde{A} \subseteq E$, we often consider the successive return times to A and departures from $\widetilde{A}$:

$$\begin{aligned}
&R_1 = H_A, \quad D_1 = T_{\widetilde{A}} \circ \theta_{R_1} + R_1, \quad \text{and for } k \ge 1,\\
&R_{k+1} = H_A \circ \theta_{D_k} + D_k, D_{k+1} = D_1 \circ \theta_{D_k} + D_k, \quad \text{so that}\\
&0 \le R_1 \le D_1 \le \cdots \le R_k \le D_k \le \cdots \le \infty,
\end{aligned} \tag{1.5}$$

and P-a.s. the above inequalities are strict except maybe for the first one. We also set $R_0 = 0 = D_0$ by convention. The transition density of the walk on E is denoted by

$$p_k(x, y) = P_x[X_k = y], \quad k \ge 0,\ x, y \in E. \tag{1.6}$$

We write $P_x^{\mathbb{Z}^\nu}$, or $E_x^{\mathbb{Z}^\nu}$, for $x \in \mathbb{Z}^\nu$, $\nu \ge 1$, to indicate the law or expectation for simple random walk on $\mathbb{Z}^\nu$ starting from x. We otherwise keep the same notation as above. We let $g_\nu(\cdot)$ stand for the Green function of simple random walk on $\mathbb{Z}^\nu$, $\nu \ge 1$, with a pole at the origin:

$$g_\nu(z) = E_z^{\mathbb{Z}^\nu}\Big[\sum_{n\ge 0} 1_{\{X_n=0\}}\Big] \quad \text{for } z \in \mathbb{Z}^\nu \tag{1.7}$$

(which of course is identically infinite unless $\nu \ge 3$). As a direct consequence of the geometric number of returns of the walk to the origin, one classically has

$$g_\nu(0) = (1 - q(\nu))^{-1}, \tag{1.8}$$

where (cf. (0.10)) $q(\nu) = P_0^{\mathbb{Z}^\nu}[\widetilde{H}_0 < \infty]$ denotes the return probability to the origin.

We are now ready to begin and consider, for $N \ge 1$,

$$B = \pi_E([-N/8, N/8]^d \cap \mathbb{Z}^d) \subseteq \widetilde{B} = \pi_E([-N/4, N/4]^d \cap \mathbb{Z}^d), \tag{1.9}$$

as well as (cf. (1.5))

$$\mathcal{R}_k, \mathcal{D}_k, k \ge 1, \text{ the successive returns to } B \text{ and departures from } \widetilde{B}. \tag{1.10}$$

The following estimates will be useful. We also prove the controls (0.5) on the probability that a point belongs to the vacant set (see also (2.26) of [4]).

Proposition 1.1. $(d \geq 3)$

$$P[\mathcal{R}_{k^*} \leq uN^d] \leq c\exp\{-cuN^{d-2}\}, \tag{1.11}$$

$$P[\mathcal{R}_{k_*} \geq uN^d] \leq c\exp\{-cuN^{d-2}\}, \tag{1.12}$$

for $u > 0$*,* $N \geq 1$*, with* $k^* = [c_1 uN^{d-2}]$*,* $k_* = [c_2 uN^{d-2}]$*, and* $c_1 > c_2$*. Moreover, for* $u > 0$*,*

$$e^{-cu} \leq \liminf_N P[0 \notin X_{[0,uN^d]}] \leq \limsup_N P[0 \notin X_{[0,uN^d]}] \leq e^{-c'u}. \tag{1.13}$$

Proof. We begin with the proof of (1.11). As a direct consequence of the invariance principle, we see that for $N \geq 1$, $x \in \widetilde{B}^c$, $y \in B$,

$$E_x\left[\exp\left\{-\frac{H_B}{N^2}\right\}\right] \leq 1 - c, \quad E_y\left[\exp\left\{-\frac{T_{\widetilde{B}}}{N^2}\right\}\right] \leq 1 - c. \tag{1.14}$$

Hence for $k \geq 2$ and $x \in E$ one finds, using the strong Markov property and induction,

$$\begin{aligned} E_x\left[\exp\left\{-\frac{\mathcal{R}_k}{N^2}\right\}\right] &\leq (1-c)E_x\left[\exp\left\{-\frac{\mathcal{D}_{k-1}}{N^2}\right\}\right] \\ &\leq (1-c)^2 E_x\left[\exp\left\{-\frac{\mathcal{R}_{k-1}}{N^2}\right\}\right] \leq (1-c)^{2(k-1)}. \end{aligned}$$

As a result we see that (with the convention below (1.5))

$$E_x\left[\exp\left\{-\frac{\mathcal{R}_k}{N^2}\right\}\right] \leq c\exp\{-ck\} \quad \text{for } k \geq 0, \tag{1.15}$$

and therefore for $u > 0$, $N \geq 1$, $k \geq 0$, we find

$$P[\mathcal{R}_k \leq uN^d] \leq c\exp\{uN^{d-2} - ck\}, \tag{1.16}$$

from which (1.11) readily follows.

We now turn to the proof of (1.12). With similar arguments to the proof of Lemma 1.3 of [7], we see that

$$E_x\left[\exp\left\{\frac{c}{N^2}H_B\right\}\right] \leq 2, \;\; E_x\left[\exp\left\{\frac{c}{N^2}T_{\widetilde{B}}\right\}\right] \leq 2 \;\text{ for } N \geq 1 \text{ and } x \in E. \tag{1.17}$$

Therefore by the strong Markov property and induction we find that for $k \geq 1$,

$$E\left[\exp\left\{\frac{c}{N^2}\mathcal{R}_k\right\}\right] \leq 2E\left[\exp\left\{\frac{c}{N}\mathcal{D}_{k-1}\right\}\right] \leq 4E\left[\exp\left\{\frac{c}{N^2}\mathcal{R}_{k-1}\right\}\right] \leq 4^k. \tag{1.18}$$

It now follows that for $k \geq 0$,

$$P[\mathcal{R}_k \geq uN^d] \leq \exp\{-cuN^{d-2} + 2(\log 2)k\}, \tag{1.19}$$

from which (1.12) easily follows.

We next prove (1.13). Using a comparison between the Green function of the random walk killed when exiting $\widetilde{B}$ and of simple random walk in $\mathbb{Z}^d$ (see for instance (1.10), (1.11) in Lemma 1.2 of [7]), we have

$$c'(|x-y|+1)^{-(d-2)} \le P_x[H_y < T_{\widetilde{B}}] \le c(|x-y|+1)^{-(d-2)} \quad \text{for } N \ge 1,\, x, y \in B. \tag{1.20}$$

Note that for $k \ge 1$, one has

$$\{H_0 > \mathcal{D}_k\} = \{H_0 > \mathcal{R}_1\} \cap \theta^{-1}_{\mathcal{R}_1}\{H_0 > T_{\widetilde{B}}\} \cap \cdots \cap \theta^{-1}_{\mathcal{R}_k}\{H_0 > T_{\widetilde{B}}\}. \tag{1.21}$$

Hence by the strong Markov property and the left-hand inequality of (1.20) we see that

$$P[H_0 > \mathcal{D}_k] \le (1 - cN^{-(d-2)})^k \quad \text{for } k \ge 1. \tag{1.22}$$

Similarly we see that for $k \ge 1$, $0 < \epsilon < 1/8$,

$$\begin{aligned} P[H_0 > \mathcal{D}_k] &\ge P[X_0 \notin B(0, \epsilon N),\ H_0 > \mathcal{D}_k] \\ &\ge \left(1 - \frac{c}{N^{d-2}}\right)^{k-1} P[X_0 \notin B(0,\epsilon N),\ H_0 > T_{\widetilde{B}}] \\ &\ge (1 - cN^{-(d-2)})^{k-1}(1 - c(\epsilon N)^{-(d-2)})_+ \left(1 - \frac{|B(0,\epsilon N)|}{N^d}\right). \end{aligned} \tag{1.23}$$

We can now write, for large N,

$$\begin{aligned} P[H_0 > uN^d] &\le P[\mathcal{R}_{k_*} \ge uN^d] + P[H_0 > \mathcal{D}_{k_*-1}] \\ &\overset{(1.12),(1.22)}{\le} c\exp\{-cuN^{d-2}\} + (1 - cN^{-(d-2)})^{k_*-1}, \end{aligned} \tag{1.24}$$

as well as

$$\begin{aligned} P[H_0 > uN^d] &\ge P[H_0 > \mathcal{D}_{k^*}, \mathcal{R}_{k^*} \ge uN^d] \\ &\overset{(1.11),(1.23)}{\ge} -c\exp\{-cuN^{d-2}\} \\ &\quad + (1 - cN^{-(d-2)})^{k^*-1}(1 - c(\epsilon N)^{-(d-2)})(1 - N^{-d}|B(0,\epsilon N)|). \end{aligned} \tag{1.25}$$

Inserting the value of k_* and k^* (see below (1.12)), we can let N tend to infinity in (1.24), (1.25) and then ϵ to 0 in (1.25), and find (1.13). □

We now come to the main result of this section that shows the ubiquity of segments of logarithmic size in the vacant set left by the walk at times which are small multiples of N^d. As explained in the Introduction, the heuristics underlying this result stems from the coupon-collector problem (cf. below (0.8)). We refer to (0.3) for the definition of the event $\mathcal{V}_{K,\beta,t}$ with $K > 0$, $0 < \beta < 1$, $t \ge 0$. Our main result is:

Theorem 1.2. $(d \ge 4)$ *For any* $K > 0$, $0 < \beta < 1$,

$$\lim_N P[\mathcal{V}_{K,\beta,uN^d}] = 1 \quad \textit{for small } u > 0. \tag{1.26}$$

Proof. We pick β_1, β_2 such that

$$0 < \beta_2 < \beta_1 < \beta < 1 \quad \text{and} \quad 2\beta_1 - \beta_2 < \beta. \tag{1.27}$$

Using translation invariance and isotropy the claim (1.26) follows once we show that

$$\lim_N N^d P\Big[\bigcap_{0 \le m < N^\beta} \{H_{(m+[0,L])e_1} \le uN^d\}\Big] = 0 \quad \text{for small } u > 0, \tag{1.28}$$

with the notation $L = [K \log N]$. We now prove (1.28), and for this purpose consider the segments S_i in E defined by

$$S_i = \pi_E((2i[N^{\beta_1 - \beta_2}] + [0, L])e_1), \quad 1 \le i \le \ell := [N^{\beta_1}], \tag{1.29}$$

and write

$$S = \bigcup_{1 \le i \le \ell} S_i. \tag{1.30}$$

We want to show that when $u > 0$ is chosen small, with overwhelming probability as N tends to infinity, some of the segments S_i, $1 \le i \le \ell$, remain vacant up to time uN^d. With the help of (1.27), (1.28) will then follow. We then introduce (cf. (1.10) for notation),

$$\begin{aligned}
&\mathcal{S}_0 = S, \quad \tau_1 = \inf\{k \ge 1 : H_S \circ \theta_{\mathcal{R}_k} < T_{\widetilde{B}} \circ \theta_{\mathcal{R}_k}\},\\
&\widetilde{R}_1 = H_S \circ \theta_{\mathcal{R}_{\tau_1}} + \mathcal{R}_{\tau_1}, \quad j_1 = \text{the unique } j \in \{1, \dots, \ell\} \text{ such that } X_{\widetilde{\mathcal{R}}_1} \in S_j,\\
&\mathcal{S}_1 = S \setminus S_{j_1}, \quad \tau_2 = \inf\{k > \tau_1 : H_{\mathcal{S}_1} \circ \theta_{\mathcal{R}_k} < T_{\widetilde{B}} \circ \theta_{\mathcal{R}_k}\},\\
&\widetilde{R}_2 = H_{\mathcal{S}_1} \circ \theta_{\mathcal{R}_{\tau_2}} + \mathcal{R}_{\tau_2}, \quad j_2 = \text{the unique } j \in \{1, \dots, \ell\} \setminus \{j_1\} \text{ such that}\\
&X_{\widetilde{R}_2} \in S_j, \text{ and so on until } \mathcal{S}_{\ell-1} = S \setminus \textstyle\bigcup_{j \in \{j_1, \dots, j_{\ell-1}\}} S_j,\ \tau_\ell,\ \widetilde{R}_\ell \text{ and}\\
&j_\ell \text{ with } \{1, \dots, \ell\} = \{j_1, \dots, j_\ell\}.
\end{aligned} \tag{1.31}$$

In this fashion we label the successive excursions in B and out of $\widetilde{B}$ giving rise to hits of new segments, and for the time being disregard the fact that possibly more than one segment may be hit during one such excursion. As a straightforward consequence of the above definition, one has

$$\begin{aligned}
&\widetilde{R}_i,\ 1 \le i \le \ell, \text{ are } (\mathcal{F}_n)\text{-stopping times},\\
&j_m,\ 1 \le m \le i, \text{ are } \mathcal{F}_{\widetilde{R}_i}\text{-measurable},\\
&\mathcal{D}_{\tau_i} = T_{\widetilde{B}} \circ \theta_{\widetilde{R}_i} + \widetilde{R}_i,\ 1 \le i \le \ell, \text{ are } (\mathcal{F}_n)\text{-stopping times as well.}
\end{aligned} \tag{1.32}$$

By (1.20), we see that for large N, when $U \subseteq S$,

$$P_x[H_U < T_{\widetilde{B}}] \le \frac{c_3}{2} \frac{|U|}{N^{d-2}} \quad \text{for } x \in B \cap \partial(B^c), \tag{1.33}$$

and using the reversibility of the walk on E,

$$\begin{aligned} P[H_U < T_{\widetilde{B}}] &\le N^{-d} \sum_{x\in E,\, y\in U,\, k\ge 0} P_x[X_k = y,\ T_{\widetilde{B}} > k] \\ &= N^{-d} \sum_{x\in E,\, y\in U,\, k\ge 0} P_y[X_k = x,\ T_{\widetilde{B}} > k] = N^{-d} \sum_{y\in U} E_y[T_{\widetilde{B}}] \\ &\le \frac{c_3}{2}\, \frac{|U|}{N^{d-2}}, \end{aligned} \tag{1.34}$$

using the fact that $\sup_{z\in E} E_z[T_{\widetilde{B}}] \le cN^2$ (cf. the second inequality of (1.17)), and increasing if necessary the value of c_3 in (1.33). We now introduce, for $U \subseteq S$,

$$\eta(U) = \inf\{k \ge 1 : H_U \circ \theta_{\mathcal{R}_k} < T_{\widetilde{B}} \circ \theta_{\mathcal{R}_k}\}, \tag{1.35}$$

and note that for $\lambda > 0$ and $2 \le i \le \ell$, as a result of the strong Markov property applied at time $\mathcal{D}_{\tau_{i-1}}$ (cf. (1.32)),

$$E[\exp\{-\lambda(\tau_i - \tau_{i-1})\} \,|\, \mathcal{F}_{\mathcal{D}_{\tau_{i-1}}}] = \sum_{k\ge 1} e^{-\lambda k} P_{X_{\mathcal{D}_{\tau_{i-1}}}}[\eta(\mathcal{S}_{i-1}) = k], \tag{1.36}$$

where in the last expression $\mathcal{S}_{i-1}$ is a frozen variable ($\mathcal{F}_{\mathcal{D}_{\tau_{i-1}}}$-measurable).

Note that for $z \in \widetilde{B}^c$, $U \subseteq S$,

$$\sum_{k\ge 1} e^{-\lambda k} P_z[\eta(U) = k] = (1 - e^{-\lambda}) \sum_{k\ge 1} e^{-\lambda k} P_z[\eta(U) \le k]. \tag{1.37}$$

Moreover for large N, with z and U as above,

$$\begin{aligned} P_z[\eta(U) \le k] &= 1 - P_z[\eta(U) > k] \\ &= 1 - P_z[H_U \circ \theta_{\mathcal{R}_m} > T_{\widetilde{B}} \circ \theta_{\mathcal{R}_m} \text{ for } 1 \le m \le k] \\ &\le 1 - \Big(1 - \frac{c_3}{2}\, \frac{|U|}{N^{d-2}}\Big)^k \quad \text{for } k \ge 0, \end{aligned} \tag{1.38}$$

with the help of (1.33) and strong Markov property applied at times $\mathcal{R}_m$, $1 \le m \le k$. We thus see that under P_z, $\eta(U)$ stochastically dominates a geometric variable with success probability

$$p(U) = c_3 \frac{|U|}{N^{d-2}} \tag{1.39}$$

(the factor 2 multiplying the expression subtracted from 1 under the parenthesis in (1.38) is there in order to obtain (1.41) below). Thus coming back to (1.37), we see that for large N, $U \subseteq S$, $z \in \widetilde{B}^c$,

$$\begin{aligned} E_z[\exp\{-\lambda\eta(U)\}] &\le \sum_{k\ge 1} e^{-\lambda k} p(U)(1 - p(U))^{k-1} \\ &= \frac{e^{-\lambda} p(U)}{1 - (1 - p(U))e^{-\lambda}} \quad \text{for } \lambda \ge 0. \end{aligned} \tag{1.40}$$

In the same fashion, using (1.33), (1.34) (note that $P[H_S \circ \theta_{\mathcal{R}_1} < T_{\widetilde{B}} \circ \theta_{\mathcal{R}_1}] \le c_3 |S|/N^{d-2}$, when bounding the term $m = 1$ in the expression corresponding to the second line of (1.38)), we find that for large N,

$$E[\exp\{-\lambda \eta(S)\}] \le \frac{e^{-\lambda} p(S)}{1 - (1 - p(S)) e^{-\lambda}} \quad \text{for } \lambda \ge 0. \tag{1.41}$$

Hence by (1.36), (1.40), (1.41), and the fact that P-a.s., for $1 \le i \le \ell$,

$$p_i := p(\mathcal{S}_{i-1}) \overset{(1.39)}{\underset{(1.31)}{=}} c_3 (L+1)(\ell - i + 1) N^{-(d-2)}, \tag{1.42}$$

we find that for large N and $\ell' \le \ell$, $\lambda \ge 0$,

$$E[\exp\{-\lambda \tau_{\ell'}\}] \le \prod_{1 \le i \le \ell'} \frac{e^{-\lambda} p_i}{1 - e^{-\lambda}(1 - p_i)}, \tag{1.43}$$

and hence with the notation below (1.12),

$$P[\tau_{\ell'} \le k^*] \le \exp\Big\{\lambda k^* - \lambda \ell' + \sum_{1 \le i \le \ell'} \log\Big(\frac{p_i}{p_i + (1 - e^{-\lambda})(1 - p_i)}\Big)\Big\}. \tag{1.44}$$

We now specify ℓ' and λ by setting (cf. (1.27))

$$\ell' = \ell - [N^{\beta_2}], \qquad \rho := d - 2 - \frac{\beta_1 + \beta_2}{2} \; (> 0), \qquad \lambda = N^{-\rho}. \tag{1.45}$$

As a result for large N, by (1.42), (1.27), we have

$$\begin{aligned} &\text{(i)} \quad 10^{-3} p_1 \ge N^{-\rho} \ge \frac{1 - e^{-\lambda}}{2} \ge \frac{N^{-\rho}}{4} \ge 10^3 c_3 \frac{L+1}{N^{d-2}}, \\ &\text{(ii)} \quad \tfrac{1}{2} \ge p_1 \ge p_i \quad \text{for } 1 \le i \le \ell, \\ &\text{(iii)} \quad (\ell - \ell') c_3 \frac{L+1}{N^{d-2}} < \frac{N^{-\rho}}{10}. \end{aligned} \tag{1.46}$$

As a result we see that for large N,

$$\begin{aligned} \sum_{1 \le i \le \ell'} \log\Big(\frac{p_i + (1 - e^{-\lambda})(1 - p_i)}{p_i}\Big) &\ge \sum_{1 \le i \le \ell'} \int_{p_i}^{p_i + N^{-\rho}/4} \frac{dt}{t} \\ &\overset{(1.42)}{=} \int_0^\infty \sum_{\ell - \ell' < j \le \ell} 1\Big\{c_3 \frac{L+1}{N^{d-2}} j < t < c_3 \frac{L+1}{N^{d-2}} j + \frac{N^{-\rho}}{4}\Big\} \frac{dt}{t}. \end{aligned} \tag{1.47}$$

From (1.46) it now follows that for large N, when $t \in (N^{-\rho}/5, c_3(L+1)\ell/N^{d-2})$, the sum under the above integral is greater than $cN^{-\rho}N^{d-2}/L$. As a result we see that for large N, using (1.27), the definition of L below (1.28), and (1.45),

$$\sum_{1\le i\le \ell'} \log\Big(\frac{p_i + (1-e^{-\lambda})(1-p_i)}{p_i}\Big) \ge cN^{-\rho}\,\frac{N^{d-2}}{L}\,\log(5c_3(L+1)\ell N^{\rho-(d-2)})$$
$$\ge \frac{c}{K} N^{d-2-\rho}(\beta_1 + \rho - (d-2)) = \frac{c}{K}(\beta_1-\beta_2)N^{(\beta_1+\beta_2)/2}. \qquad (1.48)$$

Inserting this bound in (1.44), by (1.45) we see that for large N,

$$P[\tau_{\ell'} \le k^*] \le \exp\Big\{c_1 u N^{(\beta_1+\beta_2)/2} - \frac{c}{K}(\beta_1-\beta_2)N^{(\beta_1+\beta_2)/2}\Big\}. \qquad (1.49)$$

We will now take care of the issue (cf. below (1.31)) of additional hits during the time intervals $[\widetilde{R}_i, \mathcal{D}_{\tau_i}]$, $1 \le i \le \ell'$, of segments not hit previously. With this objective in mind we thus define, for $1 \le i \le \ell'$,

$$N_i = \sum_{j \notin \{j_1,\dots,j_i\}} 1\{H_{S_j} \circ \theta_{\widetilde{R}_i} < T_{\widetilde{B}} \circ \theta_{\widetilde{R}_i}\}, \qquad (1.50)$$

so that N_i is $\mathcal{F}_{\mathcal{D}_{\tau_i}}$-measurable (cf. (1.32)). Note that P-a.s, $X_{\widetilde{R}_i} \in S_{j_i}$ for $1 \le i \le \ell$ (cf. (1.31)), and by the strong Markov property at time $\widetilde{R}_i$, one finds

$$P[N_i \ge m \,|\, \mathcal{F}_{\widetilde{R}_i}] \le P_{X_{\widetilde{R}_i}}\Big[\sum_{1\le j\le \ell} 1\{H_{S_j} < T_{\widetilde{B}}\} \ge 1+m\Big] \le P_{X_{\widetilde{R}_i}}[V_m < T_{\widetilde{B}}], \qquad (1.51)$$

where $V_m, m \ge 1$, denote the successive times of visit of the walk to distinct segments S_j, $1 \le j \le \ell$.

It now follows from (1.20), (1.29) that for large N, when $z \in S$,

$$P_z[V_1 < T_{\widetilde{B}}] \le c(L+1)\sum_{k\ge 1}(kN^{\beta_1-\beta_2})^{-(d-2)} \le c_4 L N^{-(d-2)(\beta_1-\beta_2)} =: p. \qquad (1.52)$$

Coming back to (1.51) we thus see that by the strong Markov property, when N is large,

$$P[N_i \ge m \,|\, \mathcal{F}_{\widetilde{R}_i}] \le p^m \quad \text{for } m \ge 0, \qquad (1.53)$$

i.e. conditionally on $\mathcal{F}_{\widetilde{R}_i}$, N_i is stochastically dominated by a modified geometric distribution with success parameter p. Therefore when λ' is such that (cf. (1.52)) $e^{\lambda'}p < 1$, we find that for large N, λ' as above and $1 \le i \le \ell'$,

$$E[\exp\{\lambda' N_i\} \,|\, \mathcal{F}_{\widetilde{R}_i}] \le \sum_{m\ge 0}(1-p)p^m e^{\lambda' m} = \frac{1-p}{1-e^{\lambda'}p}, \qquad (1.54)$$

so that using induction and $\mathcal{F}_{\mathcal{D}_{\tau_{i-1}}} \subseteq \mathcal{F}_{\widetilde{R}_i}$ for $2 \le i \le \ell'$, we see in view of the measurability of N_i asserted below (1.50) that for large N,

$$E\Big[\exp\Big\{\sum_{1\le i\le \ell'} N_i\Big\}\Big] \le \Big(1 + \frac{(e-1)p}{1-ep}\Big)^{\ell'} \overset{(1.52)}{\le} \exp\{c\ell' p\}, \tag{1.55}$$

and hence by (1.45), the value of p in (1.52), and the fact that $d \ge 4$,

$$\begin{aligned} P\Big[\sum_{1\le i\le \ell'} N_i \ge \frac{\ell-\ell'}{2}\Big] &\le \exp\Big\{-\frac{N^{\beta_2}}{4} + c\ell' p\Big\} \\ &\le \exp\Big\{-\frac{N^{\beta_2}}{4} + cN^{\beta_2} L N^{-(d-3)(\beta_1-\beta_2)}\Big\} \le \exp\Big\{-\frac{N^{\beta_2}}{8}\Big\}. \end{aligned} \tag{1.56}$$

To conclude the proof of (1.28), we observe that by (1.31) for large N, on the event $\{\tau_{i-1} < k < \tau_i\}$, where $1 \le i \le \ell'$ and $\tau_0 = 0$ by convention, $X_n \notin \mathcal{S}_{i-1}$ for $\mathcal{R}_k \le n < \mathcal{R}_{k+1}$. As a result, on the event $\{\mathcal{R}_{k^*} > uN^d\} \cap \{\tau_{\ell'} > k^*\} \cap \{\sum_{1\le i\le\ell'} N_i < (\ell-\ell')/2\}$ at least $[(\ell-\ell')/2]$ segments S_i, $1 \le i \le \ell$, have not been visited by the walk up to time uN^d, so that when N is large the above event lies in the complement of the event that appears in (1.28). Collecting the bounds (1.11), (1.49), (1.56), we obtain (1.28). As already explained, this yields our claim (1.26), so that Theorem 1.2 is now proved. □

Remark 1.3. Concerning the large u regime, let us point out that when $d \ge 4$, given $K > 0$, the vacant set left by the walk at time uN^d typically for large N does not contain any segment of length $[K \log N]$ if u is chosen large enough. Indeed, when $8L \le N$, and $U \subseteq B$ is the segment $U = [0, L]e$ with $|e| = 1$, we obtain

$$\begin{aligned} P_x[H_U < T_{\widetilde{B}}] &\ge E_x\Big[\sum_{n=0}^{T_{\widetilde{B}}-1} 1\{X_n \in U\}\Big] \Big/ \sup_{y\in U} E_y\Big[\sum_{n=0}^{T_{\widetilde{B}}-1} 1\{X_n \in U\}\Big] \\ &\ge c\,\frac{L}{N^{d-2}} \quad \text{for any } x \in B, \end{aligned} \tag{1.57}$$

using the strong Markov property at time H_U for the first inequality, and bounds on the Green function of the walk killed when exiting $\widetilde{B}$ for the second inequality (see (1.20) and also (1.11) of [7]). From a straightforward modification of (1.24), we thus find that when $8L \le N$ and $u > 0$,

$$P[H_U > uN^d] \le c\exp\{-cuN^{d-2}\} + \Big(1 - c\frac{L}{N^{d-2}}\Big)_+^{k_*-1} \le c\exp\{-cuL\}, \tag{1.58}$$

using the value of k_* given below (1.12). Hence choosing $L = \big[c_*\frac{\log N}{u}\big]$ with c_* a large enough constant, we see that for $u > 0$,

$$\lim_N P\Big[X^c_{[0,uN^d]} \text{ contains some segment of length } \Big[c_*\frac{\log N}{u}\Big]\Big] = 0. \tag{1.59}$$

So for large N the vacant set typically does not contain segments of length $K \log N$ if u is chosen large enough. □

The theorem we have just proved will enter as a step when showing in the next section that the giant component we define, with overwhelming probability occurs in the regime of parameters we consider.

2. Exponential bound and giant component

In this section we derive an exponential bound on the probability that the walk covers certain subsets of E by times that are small multiples of N^d (cf. Theorem 2.1). This bound plays an important role in the construction of the giant component typically present in the vacant set left by the walk at such times. We also refer to Remark 2.4 where it is explained why some restrictions are needed on the class of sets to which the exponential bound applies.

We refer to (1.2) for the definition of $\mathcal{L}_m$, $1 \le m \le d$, and define for $1 \le m \le d$,

$$\begin{aligned}\mathcal{A}_m = \;&\text{the collection of non-empty subsets } A \text{ of } E \text{ such that } A \subseteq F\\ &\text{for some } F \in \mathcal{L}_m.\end{aligned} \tag{2.1}$$

Clearly $\mathcal{A}_m$ increases with m, and $\mathcal{A}_d$ is the collection of non-empty subsets of E. We will especially be interested in $\mathcal{A}_2$. We also recall the notation $q(\nu)$ in (0.10) and below (1.8). The next theorem contains the key exponential estimate.

Theorem 2.1. ($d \ge 4$, $1 \le m \le d-3$) *When* $\lambda > 0$ *is such that*

$$\chi := e^{2\lambda}\left(\frac{m}{d} + \left(1 - \frac{m}{d}\right)q(d-m)\right) < 1, \tag{2.2}$$

then for $u > 0$,

$$\limsup_N \sup_{A \in \mathcal{A}_m} |A|^{-1} \log\Big(E\Big[\exp\Big\{\lambda \sum_{x \in A} 1\{H_x \le uN^d\}\Big\}\Big]\Big) \le c_5 u \frac{e^{2\lambda} - 1}{1 - \chi}, \tag{2.3}$$

and there exist $N_1(d, m, \lambda) \ge 1$ *and* $u_1(d, m, \lambda) > 0$ *such that for* $N \ge N_1$,

$$P[X_{[0,u_1 N^d]} \supseteq A] \le \exp\{-\lambda |A|\} \quad \textit{for all } A \in \mathcal{A}_m. \tag{2.4}$$

We refer to Remark 2.4 below for an explanation on why some restriction on the class of subsets A that appear in (2.4) is needed.

Proof. We begin with the proof of (2.3). We consider $N \ge 1$, $u > 0$, $A \in \mathcal{A}_m$, $1 \le m \le d-3$. Roughly speaking, we chop the time interval $[0, [uN^d]]$ into successive intervals of length N^2, except maybe for the last one, and write, for $\lambda > 0$,

$$\begin{aligned}&E\Big[\exp\Big\{\lambda \sum_{x \in A} 1\{H_x \le uN^d\}\Big\}\Big]\\ &\qquad \le E\Big[\exp\Big\{\lambda \sum_{kN^2 \le uN^d} \sum_{x \in A} 1\{H_x < N^2\} \circ \theta_{kN^2}\Big\}\Big] \le \sqrt{a_1}\sqrt{a_2},\end{aligned} \tag{2.5}$$

where

$$a_1 := E\Big[\exp\Big\{2\lambda \sum_{k \text{ even}, \, kN^2 \le uN^d} \sum_{x \in A} 1\{H_x < N^2\} \circ \theta_{kN^2}\Big\}\Big],$$
$$a_2 := E\Big[\exp\Big\{2\lambda \sum_{k \text{ odd}, \, kN^2 \le uN^d} \sum_{x \in A} 1\{H_x < N^2\} \circ \theta_{kN^2}\Big\}\Big]. \tag{2.6}$$

We first bound a_1. To this end we define

$$k_0 = \max\{k \ge 0 : 2kN^2 \le uN^d\}, \tag{2.7}$$

$$\phi(z) = E_z\Big[\exp\Big\{2\lambda \sum_{x \in A} 1\{H_x < N^2\}\Big\}\Big] \; (\ge 1) \quad \text{for } z \in E. \tag{2.8}$$

Applying the strong Markov property at time H_A, we find

$$\phi(z) \le P_z[H_A \ge N^2] + E_z[H_A < N^2, \, \phi(X_{H_A})] \quad \text{for } z \in E. \tag{2.9}$$

By the simple Markov property applied at time $2k_0N^2$ and then at time $(2k_0 - 1)N^2$, we see that when $k_0 \ge 1$,

$$a_1 = E\Big[\exp\Big\{2\lambda \sum_{0 \le k < k_0} \sum_{x \in A} 1\{H_x < N^2\} \circ \theta_{2kN^2}\Big\} E_{X_{(2k_0-1)N^2}}[\phi(X_{N^2})]\Big]. \tag{2.10}$$

Note that for $z \in E$, one has

$$\begin{aligned} E_z[\phi(X_{N^2})] &\overset{(1.6)}{=} \sum_{y \in E} p_{N^2}(z, y)\phi(y) \\ &\overset{(2.9)}{\le} 1 + \sum_{y \in E} p_{N^2}(z, y) E_y[H_A < N^2, \, \phi(X_{H_A}) - 1] \\ &\le 1 + \frac{c}{N^d} \sum_{y \in E} \sum_{0 \le k < N^2} P_y[X_k \in A](\|\phi\|_\infty - 1) \\ &\le 1 + \frac{c}{N^{d-2}} |A| (\|\phi\|_\infty - 1) \le \exp\Big\{c \frac{|A|}{N^{d-2}}(\|\phi\|_\infty - 1)\Big\}, \end{aligned} \tag{2.11}$$

where in the third line we have used the fact that

$$\sup_{x, y \in E} N^d p_{N^2}(x, y) \le c, \tag{2.12}$$

as follows from standard upper bounds on the transition density of simple random walk on $\mathbb{Z}^d$ (cf. (2.4) of [9]). With an even simpler (and similar) argument we also have

$$E[\phi(X_0)] \le \exp\Big\{c \frac{|A|}{N^{d-2}}(\|\phi\|_\infty - 1)\Big\}. \tag{2.13}$$

Therefore using induction together with (2.11), and (2.13) to handle the term corresponding to $k = 0$ in (2.10), we see that

$$a_1 \le \exp\Big\{(k_0+1)c\,\frac{|A|}{N^{d-2}}(\|\phi\|_\infty - 1)\Big\}. \tag{2.14}$$

A similar bound holds for a_2, and by (2.5) we thus find

$$E\Big[\exp\Big\{\lambda \sum_{x\in A} 1\{H_x \le uN^d\}\Big\}\Big] \le \exp\Big\{c(uN^{d-2}+1)\frac{|A|}{N^{d-2}}(\|\phi\|_\infty - 1)\Big\}. \tag{2.15}$$

We will now seek an upper bound on $\|\phi\|_\infty$.

Lemma 2.2. $(d \ge 4,\ 1 \le m \le d-3,\ e^{2\lambda} m/d < 1,\ N \ge 2)$

$$\|\phi\|_\infty \le \frac{e^{2\lambda}}{1-e^{2\lambda}m/d}\Big(1-\frac{m}{d}\Big)(1+q_N(\|\phi\|_\infty - 1)), \tag{2.16}$$

where (with hopefully obvious notations)

$$q_N := P_{e_1}^{(\mathbb{Z}/N\mathbb{Z})^{d-m}}[H_0 < N^2]. \tag{2.17}$$

Proof. Consider $F \in \mathcal{L}_m$ such that $A \subseteq F$, and introduce (cf. (1.3))

$$R_F := H_F \circ \theta_{T_F} + T_F, \tag{2.18}$$

the return time to F. Since $A \subseteq F$, for $z \in E$ we find

$$\begin{aligned}
\phi(z) &\le E_z\Big[\exp\Big\{2\lambda\Big(T_F + 1\{R_F < N^2\}\Big(\sum_{x\in A} 1\{H_x < N^2\}\circ\theta_{R_F}\Big)\Big)\Big\}\Big] \\
&= E_z\Big[\exp\{2\lambda T_F\}\Big(1\{R_F \ge N^2\} + 1\{R_F < N^2\}\exp\Big\{2\lambda \sum_{x\in A} 1\{H_x < N^2\}\circ\theta_{R_F}\Big\}\Big)\Big] \\
&= E_z[\exp\{2\lambda T_F\}(1 + 1\{R_F < N^2\}(\phi(X_{R_F}) - 1))] \\
&\le E_z[\exp\{2\lambda T_F\}] + E_z[\exp\{2\lambda T_F\} P_{X_{T_F}}[H_F < N^2]](\|\phi\|_\infty - 1),
\end{aligned} \tag{2.19}$$

where we used the strong Markov property at time R_F in the third line. Considering the motion of X in the directions "transversal to F", we have

$$\text{for } z \in E,\ P_z\text{-a.s.,} \quad P_{X_{T_F}}[H_F < N^2] \le q_N. \tag{2.20}$$

When $z \in F$, T_F has geometric distribution with success probability $1 - m/d$, so that for λ as indicated above,

$$\begin{aligned}
&E_z[\exp\{2\lambda T_F\}] \\
&= \sum_{k\ge 1}\Big(1-\frac{m}{d}\Big)\Big(\frac{m}{d}\Big)^{k-1} e^{2\lambda k} = e^{2\lambda}\Big(1-\frac{m}{d}\Big)\Big(1-e^{2\lambda}\frac{m}{d}\Big)^{-1} \quad \text{for } z \in F,
\end{aligned} \tag{2.21}$$

whereas $T_F = 0$ P_z-a.s. when $z \notin F$. Hence coming back to the last line of (2.19), we obtain (2.16). □

In the next lemma we relate q_N of (2.17) to $q(d-m)$ (cf. (0.10)).

Lemma 2.3.

$$\limsup_N q_N \le q(d-m). \tag{2.22}$$

Proof. We denote by W the discrete cube image of $V := [-N/4, N/4]^{d-m} \cap \mathbb{Z}^{d-m}$ under the canonical projection onto $(\mathbb{Z}/N\mathbb{Z})^{d-m}$. We have

$$\begin{aligned} q_N &\le P_{e_1}^{(\mathbb{Z}/N\mathbb{Z})^{d-m}}[H_0 < T_W] + E_{e_1}^{(\mathbb{Z}/N\mathbb{Z})^{d-m}}[P_{X_{T_W}}^{(\mathbb{Z}/N\mathbb{Z})^{d-m}}[H_0 < N^2]] \\ &\le q(d-m) + \sup_{z \in \partial W} P_z^{(\mathbb{Z}/N\mathbb{Z})^{d-m}}[H_0 < N^2]. \end{aligned} \tag{2.23}$$

One has the classical upper bound (cf. for instance (2.4) of [9]),

$$P_x^{\mathbb{Z}^{d-m}}[X_k = y] \le \frac{c(m)}{k^{(d-m)/2}} \exp\left\{-c(m)\frac{|y-x|^2}{k}\right\} \quad \text{for } k \ge 1,\ x, y \in \mathbb{Z}^{d-m} \tag{2.24}$$

(using the convention concerning constants stated at the end of the Introduction). Hence for large N we obtain

$$\begin{aligned} &\sup_{z\in\partial W} P_z^{(\mathbb{Z}/N\mathbb{Z})^{d-m}}[H_0 < N^2] = \sup_{z \in \partial V} P_z^{\mathbb{Z}^{d-m}}[H_{N\mathbb{Z}^{d-m}} < N^2] \\ &\overset{(2.24)}{\le} \sup_{z\in\partial V} \sum_{y \in \mathbb{Z}^{d-m}} \sum_{2 \le k < N^2} \frac{c(m)}{k^{(d-m)/2}} \exp\left\{-c(m)\frac{|Ny-z|^2}{k}\right\} \\ &\le \sup_{z\in\partial V} \sum_{y\in\mathbb{Z}^{d-m}} \int_0^{N^2} \frac{c(m)}{s^{(d-m)/2}} \exp\left\{-c(m)\frac{|Ny-z|^2}{s}\right\} \\ &\le \sup_{w \in (\partial V)/N} c(m) N^{-(d-m-2)} \int_0^1 \sum_{y\in\mathbb{Z}^{d-m}} t^{-(d-m)/2} \exp\left\{-c(m)\frac{|y-w|^2}{t}\right\} dt. \end{aligned} \tag{2.25}$$

We can now split the sum under the integral, keeping on one hand $y \in \mathbb{Z}^{d-m}$ with $|y| \ge c(m)$, so that

$$|y-w|^2 \ge c(m)|y'|^2 \quad \text{for } y' \in y + [0,1]^{d-m} \text{ and } w \in (\partial V)/N (\subseteq [-1,1]^{d-m}),$$

and hence for $t \in (0,1]$, $w \in (\partial V)/N$,

$$\begin{aligned} &\sum_{|y| \ge c(m)} t^{-(d-m)/2} \exp\left\{-c(m)\frac{|y-w|^2}{t}\right\} \\ &\qquad \le \sum_{y\in\mathbb{Z}^{d-m}} \int_{y+[0,1]^{d-m}} t^{-(d-m)/2} \exp\left\{-c(m)\frac{|y'|^2}{t}\right\} dy' \le c(m). \end{aligned} \tag{2.26}$$

On the other hand, we consider the finitely many terms corresponding to $|y| < c(m)$. For these terms we also have, in view of the definition of V,

$$\inf\{|y-w|^2 : w \in (\partial V)/N,\ y \in \mathbb{Z}^{d-m}\} \geq c(m) > 0,$$

so that for $w \in (\partial V)/N$,

$$\int_0^1 \sum_{|y|<c(m)} t^{-(d-m)/2} \exp\left\{-c(m)\frac{|y-w|^2}{t}\right\} dt \leq c(m). \tag{2.27}$$

Thus coming back to the last line of (2.25) we find, for large N,

$$\sup_{z\in\partial W} P_z^{(\mathbb{Z}/N\mathbb{Z})^{d-m}}[H_0 < N^2] \leq c(m)N^{-(d-m-2)}, \tag{2.28}$$

and since $d - m - 2 \geq 1$ by assumption, letting N tend to infinity in (2.23) we find (2.22). □

From (2.16), (2.22), it follows by a straightforward computation that when $\lambda > 0$ satisfies (2.2),

$$\limsup_N \sup_{A\in\mathcal{A}_m} (\|\phi\|_\infty - 1) \leq \frac{e^{2\lambda}-1}{1-\chi}. \tag{2.29}$$

Coming back to (2.15), taking logarithms and dividing by $|A|$, the claim (2.3) readily follows.

We now turn to the proof of (2.4). We pick $\widetilde{\lambda}(d,m,\lambda) > \lambda$ and $\widetilde{q}(d,m,\lambda) > q$ so that

$$1 - e^{2\widetilde{\lambda}}\left(\frac{m}{d} + \left(1-\frac{m}{d}\right)\widetilde{q}\right) = \frac{1}{2}(1-\chi). \tag{2.30}$$

Applying (2.3) with $\widetilde{\lambda}$ (for which (2.2) holds) we see that for $u > 0$, $N \geq N_2(d,m,\lambda,u)$ and any $A \in \mathcal{A}_m$,

$$P[X_{[0,uN^d]} \supseteq A] \leq \exp\left\{-\widetilde{\lambda}|A| + cu\frac{e^{2\widetilde{\lambda}}-1}{1-e^{2\widetilde{\lambda}}(m/d+(1-m/d)\widetilde{q})}|A|\right\}. \tag{2.31}$$

Choosing $u = u_1(d,m,\lambda)$ small enough, and setting $N_1(d,m,\lambda) = N_2(d,m,\lambda,u_1)$, we obtain (2.4). □

Remark 2.4. 1) Let us mention that it is straightforward to argue in Lemma 2.3 that $\liminf_N q_N \geq q(d-m)$, so that (2.23) can be sharpened to

$$\lim_N q_N = q(d-m), \tag{2.32}$$

although we do not use this sharpened limiting result here.

2) As we now explain there is no exponential bound of type (2.4) valid uniformly for all $A \in \mathcal{A}_d$ (i.e. all non-empty subsets of E) when N is large, no matter how small $\lambda > 0$ is chosen. Indeed, when $\rho \in (0,1)$ and $A_L = \pi_E([-L,L]^d)$ with $L = [N^\rho]$, a calculation

qualitatively similar to that in Proposition 2.7, Chapter 3 of [16] (see in particular p. 114; the calculation in [16] is performed in a Brownian motion setting) shows that for large N, $T = [cL^d \log L]$, and for all $x \in A_L$,

$$\begin{aligned} P_x[X_{[0,T]} \supseteq A_L] &\geq P_x[X_{[0,T]} \supseteq A_L,\ T_{A_{2L}} > T] \geq \frac{1}{2} P_x[T_{A_{2L}} > T] \\ &\geq c \exp\left\{-\frac{c}{L^2} T\right\} \geq c \exp\left\{-c \frac{|A_L|}{L^2} \log L\right\}. \end{aligned} \tag{2.33}$$

Moreover by standard transition density estimates (cf. (2.4) of [9]), one has

$$\inf_{z \in E} P_z[H_{A_L} < N^2] \geq c \left(\frac{L}{N}\right)^{d-2},$$

so that using the Markov property at times kN^2, one finds, for large N,

$$P\left[H_{A_L} > \frac{u}{2} N^d\right] \leq \left(1 - c\left(\frac{L}{N}\right)^{d-2}\right)^{[\frac{u}{2} N^{d-2}]} \leq \frac{1}{2}. \tag{2.34}$$

As a result we see that for any $u > 0$ and $0 < \rho < 1$,

$$\liminf_N \, (|A_L|^{(d-2)/d} \log |A_L|)^{-1} \log P[X_{[0,uN^d]} \supseteq A_L] > -\infty, \tag{2.35}$$

and hence

$$\lim_N \sup_{A \in \mathcal{A}_d} |A|^{-1} \log P[X_{[0,uN^d]} \supseteq A] = 0. \tag{2.36}$$

This explains why some restriction on the class of subsets A entering (2.4) is needed.

We now turn to applications of Theorem 2.1 to the construction of the giant component in the vacant set left by the walk at times that are small multiples of N^d. We recall that $\star$-nearest neighbor paths have been defined at the beginning of Section 1, and write

$$\begin{aligned} a(n) = {} & \text{the cardinality of the collection of } \star\text{-nearest neighbor self-avoiding} \\ & \text{paths on } \mathbb{Z}^2, \text{ starting at the origin, with } n \text{ steps.} \end{aligned} \tag{2.37}$$

One has the easy upper bound

$$a(n) \leq 8 \cdot 7^{n-1} \quad \text{for } n \geq 1. \tag{2.38}$$

We now define, for $N \geq 1$, $K > 0$, $t \geq 0$, the event (cf. (1.2) for the notation)

$$\begin{aligned} \mathcal{U}_{K,t} = \{ & \text{for any } F \in \mathcal{L}_2, \text{ and connected subsets } O_1, O_2 \text{ of } F \setminus X_{[0,t]} \\ & \text{with } |\cdot|_\infty\text{-diameter at least } [K \log N],\ O_1 \text{ and } O_2 \text{ are in} \\ & \text{the same component of } F \setminus X_{[0,t]}\}. \end{aligned} \tag{2.39}$$

The above event will be useful in singling out the giant component. The next event will be convenient in the derivation of lower bounds on the relative volume of the giant component in Section 4. For $N \geq 1$, $K > 0$, $x \in E$, $t \geq 0$, we define, using the notation of the beginning of Section 1,

$$\mathcal{C}_{K,x,t} = \{\text{for some } F \in \mathcal{L}_2 \text{ with } x \in F, \text{ there is a nearest neighbor path in } F \setminus X_{[0,t]} \text{ from } x \text{ to } S(x, [K \log N])\}. \tag{2.40}$$

We can now state

Corollary 2.5. *There is a smallest* $d_0 \geq 5$ *such that*

$$\mu := 49\left(\frac{2}{d} + \left(1 - \frac{2}{d}\right) q(d-2)\right) < 1 \quad \text{for } d \geq d_0. \tag{2.41}$$

For $d \geq d_0$*, there is a constant* $c_0 > 0$ *(cf.* (2.47)*) such that*

$$\lim_N P[\mathcal{U}_{c_0, uN^d}] = 1 \quad \text{for small } u > 0, \tag{2.42}$$

$$\lim_{u \to 0} \liminf_N P[\mathcal{C}_{c_0, 0, uN^d}] = 1 \tag{2.43}$$

(and of course $P[\mathcal{C}_{c_0, x, uN^d}] = P[\mathcal{C}_{c_0, 0, uN^d}]$ *for all* $x \in E$*).*

Proof. One knows (cf. (5.4) in [14]) that $q(\cdot)$ has the asymptotic behavior

$$q(\nu) \sim (2\nu)^{-1} \quad \text{as } \nu \to \infty, \tag{2.44}$$

so that (2.41) straightforwardly follows. Now consider $d \geq d_0$, and choose $\lambda_0(d)$ such that

$$e^{\lambda_0} := 7\mu^{-1/4} \ (> 7), \quad \text{so that} \quad e^{2\lambda_0}\left(\frac{2}{d} + \left(1 - \frac{2}{d}\right) q(d-2)\right) < 1. \tag{2.45}$$

When N is large, on $\mathcal{U}^c_{K,uN^d}$ one can find $F \in \mathcal{L}_2$ and $O_1, O_2 \subseteq F \setminus X_{[0,uN^d]}$ that are distinct connected components of $F \setminus X_{[0,uN^d]}$ with $|\cdot|_\infty$-diameter at least $[K \log N]$. We can then introduce $\widehat{O}_i$, $i = 1, 2$, the inverse images of O_i under an "affine projection" of $\mathbb{Z}^2$ onto F. Considering separately the case when at least one of the $\widehat{O}_i$, $i = 1, 2$, has bounded components (necessarily of $|\cdot|_\infty$-diameter at least $[K \log N]$), or both $\widehat{O}_i$ have unbounded components, one can construct a $\star$-nearest neighbor self-avoiding path π with $[K \log N]$ steps in $\partial O_1 \cap F$ or $\partial O_2 \cap F$ ($\subseteq F \cap X_{[0,uN^d]}$) (see also Proposition 2.1, p. 387 in [11]). Therefore for $u < u_0 = u_1(d, m = 2, \lambda = \lambda_0)$ (cf. (2.4)), we have, writing A for the set of points visited by π,

$$\begin{aligned}
\limsup_N P[\mathcal{U}^c_{K,uN^d}] &\leq \limsup_N \sum_{F \in \mathcal{L}_2} \sum_\pi P[X_{[0,u_0 N^d]} \supseteq A] \\
&\overset{(2.4)}{\leq} \limsup_N \sum_{F \in \mathcal{L}_2} \sum_\pi \exp\{-\lambda_0 |A|\} \\
&\overset{(2.38)}{\leq} \limsup_N \sum_{F \in \mathcal{L}_2} 8N^2 7^{[K \log N]-1} e^{-\lambda_0 [K \log N]} \\
&\leq \limsup_N c N^d (7e^{-\lambda_0})^{[K \log N]},
\end{aligned} \tag{2.46}$$

where the sum over π is taken over to the collection of $\star$-nearest neighbor self-avoiding paths with values in F with $[K \log N]$ steps. By (2.44) we can thus choose

$$c_0 = 8d\left(\log \frac{1}{\mu}\right)^{-1}, \tag{2.47}$$

and find

$$\lim_N P[\mathcal{U}^c_{c_0,uN^d}] = 0 \quad \text{for } u < u_0, \tag{2.48}$$

from which (2.42) follows. We now turn to the proof of (2.43). Observe that for $u > 0$, $\ell \geq 1$ and large N, one has

$$\begin{aligned} P[\mathcal{C}^c_{c_0,0,uN^d}] &\leq P[X_{[0,uN^d]} \cap B_\infty(0,\ell) \neq \emptyset] \\ &\quad + P[X_{[0,uN^d]} \cap B_\infty(0,\ell) = \emptyset, \text{ and } \mathcal{C}^c_{c_0,0,uN^d}] \\ &\leq c\ell^d P[0 \in X_{[0,uN^d]}] + \sum_{F \in \mathcal{L}_2,\, 0 \in F} \sum_\pi P[X_{[0,uN^d]} \supseteq A], \end{aligned} \tag{2.49}$$

where we have used translation invariance in the last inequality, and the sum over π runs over $\star$-nearest neighbor self-avoiding paths with values in $F \cap (B_\infty(0, [c_0 \log N]) \setminus B_\infty(0,\ell))$, which disconnect 0 from $F \cap S(0, [c_0 \log N])$, and start on the positive half of the coordinate axis entering the definition of F with smallest label $i \in \{1, \dots, d\}$. As above A stands for the set of points visited by π. Summing over the different values $k \in [\ell + 1, [c_0 \log N]]$ of the coordinate of the starting point of π, we see that for small u and sufficiently large N,

$$\begin{aligned} \sum_{F \in \mathcal{L}_2,\, 0 \in F} \sum_\pi P[X_{[0,uN^d]} \supseteq A] &\leq c \sum_{k \geq \ell} \sum_{m \geq k} 7^m e^{-\lambda_0 m} \\ &= c \sum_{k \geq \ell} (7e^{-\lambda_0})^k (1 - 7e^{-\lambda_0})^{-1} = c(7e^{-\lambda_0})^\ell (1 - 7e^{-\lambda_0})^{-2}. \end{aligned} \tag{2.50}$$

Thus coming back to (2.49), we see from (1.13) that for $u > 0$, $\ell \geq 1$,

$$\limsup_N P[\mathcal{C}^c_{c_0,0,uN^d}] \leq c(1 - e^{-cu})\ell^d + c(7e^{-\lambda_0})^\ell (1 - 7e^{-\lambda_0})^{-2}. \tag{2.51}$$

Letting u tend to 0 and then ℓ to infinity we obtain (2.43). □

For $0 < \beta < 1$ and $t \geq 0$, we now introduce the events (cf. (0.3), (2.39))

$$\mathcal{G}_{\beta,t} = \mathcal{U}_{c_0,t} \cap \mathcal{V}_{c_0,\beta,t} \quad \text{(non-decreasing in } \beta). \tag{2.52}$$

The above events encode properties which enable us to single out a giant component. More precisely, with the notation of Corollary 2.5 we have:

Corollary 2.6. ($d \geq d_0$, $0 < \beta < 1$) *Assume $N \geq 2$ is large enough so that E has $|\cdot|_\infty$-diameter greater than $c_0 \log N$. For $t \geq 0$, on the event $\mathcal{G}_{\beta,t}$,* *there is a unique connected component in* $X^c_{[0,t]}$, *denoted by* O, *which contains connected sets* $A \in \mathcal{A}_2$ *with* $|\cdot|_\infty$*-diameter* $L_0 := [c_0 \log N]$ *(in particular a segment of length* L_0*),* (2.53)

for any $F \in \mathcal{L}_1$, $F \cap O$ *contains a segment of length* L_0, (2.54)

the N^β*-neighborhood of* O *coincides with* E. (2.55)

Moreover for any $x \in E$,

$$\text{on the event } \mathcal{G}_{\beta,t} \cap \mathcal{C}_{c_0,x,t},\ x \text{ belongs to } O. \tag{2.56}$$

Finally,

$$\lim_N P[\mathcal{G}_{\beta,uN^d}] = 1 \quad \text{for small } u > 0. \tag{2.57}$$

Proof. We begin with the proof of (2.53)–(2.55). By (0.3), we see that on $\mathcal{G}_{\beta,t}$,

$$\text{any } F \in \mathcal{L}_1 \text{ contains a segment of length } L_0 \text{ included in } X^c_{[0,t]}. \tag{2.58}$$

In particular given some $\widetilde{F} \in \mathcal{L}_2$, the above applies to all $F \in \mathcal{L}_1$ with $F \subseteq \widetilde{F}$. By (2.39) any two segments of length L_0 in $\widetilde{F} \setminus X_{[0,t]}$ belong to the same connected component of $F \setminus X_{[0,t]}$ (and hence of $X^c_{[0,t]}$). Now if $\widetilde{F}, \widetilde{F}_2 \in \mathcal{L}_2$,

$$\begin{aligned}&\text{when } \widetilde{F}_1 \cap \widetilde{F}_2 \in \mathcal{L}_1, \text{ all segments of length } L_0 \text{ in } (\widetilde{F}_1 \cup \widetilde{F}_2) \setminus X_{[0,t]}\\ &\text{are in the same connected component of } X^c_{[0,t]}.\end{aligned} \tag{2.59}$$

Then consider $y \in E$. We can find a nearest neighbor path $(y_i)_{0 \le i \le m}$ with $y_0 = 0$, $y_m = y$. Consider $\widetilde{F} \ni 0$ with $\widetilde{F} \in \mathcal{L}_2$. We can construct a sequence $\widetilde{F}_i \in \mathcal{L}_2$, $0 \le i \le m$, such that

$$\begin{aligned}&\widetilde{F}_0 = \widetilde{F},\ y_i \in \widetilde{F}_i \text{ for } 0 \le i \le m, \text{ and}\\ &\text{either } \widetilde{F}_{i-1} = \widetilde{F}_i \text{ or } \widetilde{F}_{i-1} \cap \widetilde{F}_i \in \mathcal{L}_1 \text{ for } 1 \le i \le m,\end{aligned} \tag{2.60}$$

as we now explain. If $y_1 \in \widetilde{F}_0$ $(= \widetilde{F})$, we set $\widetilde{F}_1 = \widetilde{F}_0$. Otherwise if $y_1 \notin \widetilde{F}_0$, we choose some canonical vector entering the definition of $\widetilde{F}_0$ and the canonical vector collinear to $y_1 - y_0$, and define $\widetilde{F}_1$ as passing through y_0 and generated by these two vectors. Clearly $y_1 \in \widetilde{F}_1$, and $\widetilde{F}_1 \cap \widetilde{F}_0 \in \mathcal{L}_1$. We then continue the construction by induction.

With a similar argument we also see that when $\widetilde{F}, \widetilde{F}' \in \mathcal{L}_2$ have a common point y in E, we can define $\widetilde{F}_i \in \mathcal{L}_2$, $0 \le i \le 2$, such that

$$\begin{aligned}&\widetilde{F}_0 = \widetilde{F},\ \widetilde{F}_2 = \widetilde{F}', \text{ with } y \in \widetilde{F}_i,\ 0 \le i \le 2, \text{ and}\\ &\text{either } \widetilde{F}_i = \widetilde{F}_{i-1} \text{ or } \widetilde{F}_i \cap \widetilde{F}_{i-1} \in \mathcal{L}_1 \text{ for } i = 1, 2.\end{aligned} \tag{2.61}$$

Combining (2.58)–(2.61), we see that on $\mathcal{G}_{\beta,t}$ all segments of length L_0 in $X^c_{[0,t]}$ belong to the same connected component of $X^c_{[0,t]}$. By (2.58) and the definitions (0.3), (2.39) of the events entering the definition of $\mathcal{G}_{\beta,t}$, (2.53)–(2.55) readily follow. The claim (2.56) is a direct consequence of (2.53) and (2.40). As for (2.57) it follows directly from (1.26) and (2.42). □

In the following, on the event $\mathcal{G}_{\beta,t}$ of (2.52), we will refer to the above uniquely defined connected component O as the *giant component*.

3. Excursions to small boxes in a large torus

The results of this section are preparatory for the next section, but also of independent interest. We investigate excursions of the random walk to small boxes in the large torus $E = (\mathbb{Z}/N\mathbb{Z})^d$ with $d \geq 3$. We consider two points x_1, x_2 in E at $|\cdot|_\infty$-distance of at least $2r + 3$, as well as closed $|\cdot|_\infty$-balls $C(x_i)$, $i = 1, 2$, with respective centers x_i and radius $L \leq r/10$. We are interested in suitably centered excursions of the walk from the time it first hits $C(x_1) \cup C(x_2)$ up to the last visit to $C(x_1) \cup C(x_2)$ before leaving the closed r-neighborhood of $\{x_1, x_2\}$, when the walk is conditioned to leave this r-neighborhood at some point w and start at a point u outside this r-neighborhood. Of course w determines whether the excursion lies in the neighborhood of x_1 or x_2, and we center the excursion around 0 by subtracting the relevant x_i (depending on w). As a limit model we consider the excursions of simple random walk on $\mathbb{Z}^d$ starting with the normalized harmonic measure viewed from infinity of a closed $|\cdot|_\infty$-ball C of radius L and center the origin up to the last visit of C. Our main thrust is to derive quantitative controls on the total variation norm between the centered excursions described above and the limit model just explained. Our main result appears in Theorem 3.1. Some of our calculations are similar in spirit to [6] (see in particular Lemma 2.3). However, apart from working in dimension $d \geq 3$ in place of $d = 2$, a feature of the results presented here is that they pin-point a limit model for the centered excursions.

We now introduce some notation. Throughout this section we assume that $d \geq 3$. We consider positive integers N, L, r such that

$$L \geq 1, \quad r \geq 10L, \quad N \geq 4r + 6. \tag{3.1}$$

For $x \in E$ we define (see the beginning of Section 1 for the notation)

$$C(x) = B(x, L), \quad \widetilde{C}(x) = B(x, r), \tag{3.2}$$

as well as the subsets of $\mathbb{Z}^d$,

$$C = B(0, L), \quad \widetilde{C} = B(0, r), \tag{3.3}$$

and tacitly identify $C(0)$ with C and $\widetilde{C}(0)$ with $\widetilde{C}$. We then consider two points in E,

$$x_1, x_2 \in E \quad \text{with} \quad |x_1 - x_2|_\infty \geq 2r + 3, \tag{3.4}$$

so that $\partial\widetilde{C}(x_1) \cap \partial\widetilde{C}(x_2) = \emptyset$. We then introduce the successive return times to $C(x_1) \cup C(x_2)$ and departures from $\widetilde{C}(x_1) \cup \widetilde{C}(x_2)$ (cf. (1.5)), which we denote by R_k, D_k, $k \geq 1$. In this section we will only need R_1, D_1. We also introduce the times of last visits to $C(x_1) \cup C(x_2)$ after R_k and prior to D_k:

$$L_k = \sup\{n \geq R_k : X_n \in C(x_1) \cup C(x_2),\ n < D_k\}, \quad k \geq 1 \tag{3.5}$$

(and for the sake of completeness L_k is defined as -1 when the above set is empty, an event which is P-negligible). In this section we only consider L_1. To describe the centered excursions that interest us, we introduce the canonical space

$$\begin{aligned} \mathcal{W} = &\text{ the space of finite nearest neighbor } \mathbb{Z}^d\text{-valued paths } \overline{w} = (\overline{w}_k)_{0 \leq k \leq T} \\ &\text{ with } |\overline{w}_0|_\infty = |\overline{w}_T|_\infty = L, \end{aligned} \tag{3.6}$$

denote by $Y_{\cdot}$ all the canonical processes on $\mathcal{W}$, and endow the countable space $\mathcal{W}$ with the σ-algebra $\mathcal{A}$ consisting of all subsets of $\mathcal{W}$. For $u \notin \widetilde{C}(x_1) \cup \widetilde{C}(x_2)$ and $w \in \partial \widetilde{C}(x_i)$, with $i = 1$ or 2, we define

$$Q_{u,w} = \text{the law on } \mathcal{W} \text{ of } (X_{R_1+k} - x_i)_{0 \le k \le L_1 - R_1} \text{ under } P_u[\cdot \mid X_{D_1} = w], \tag{3.7}$$

where it should be observed that the conditioning event $\{X_{D_1} = w\}$ has positive probability under P_u, and that $P_u[\cdot \mid X_{D_1} = w]$-a.s., $0 < R_1 < L_1 < \infty$ and $X_m \in \widetilde{C}(x_i)$ for $R_1 \le m \le L_1$, with i as above (3.7). So after identification of $\widetilde{C}$ with $\widetilde{C}(0)$, (3.7) is a meaningful definition.

We now turn to the construction of the limit model for these centered excursions. We first introduce the harmonic measure of C viewed from infinity and its mass, the capacity of C (cf. Chapter 2, §2 of [12]):

$$e_C(z) = \begin{cases} P_z[\widetilde{H}_C = \infty] & \text{if } z \in C \text{ (see (1.4) for the notation)}, \\ 0 & \text{if } z \notin C, \end{cases} \tag{3.8}$$

$$\text{cap}(C) = e_C(\mathbb{Z}^d), \tag{3.9}$$

$$\mu_C(z) = e_C(z)/\text{cap}(C), \tag{3.10}$$

which is the initial distribution of the limit law. We also define the time of last visit to C:

$$L_C = \sup\{n \ge 0 : X_n \in C\}, \tag{3.11}$$

with a similar convention as below (3.5) when the above set is empty, and introduce

$$Q = \text{the law on } \mathcal{W} \text{ of } (X_k)_{0 \le k \le L_C} \text{ under } P^{\mathbb{Z}^d}_{\mu_C}, \tag{3.12}$$

where $P^{\mathbb{Z}^d}_{\mu_C}$ stands for the law of simple random walk on $\mathbb{Z}^d$ with initial distribution μ_C. Note that for any $\overline{w} = (w_k)_{0 \le k \le T}$ in $\mathcal{W}$,

$$\begin{aligned} Q(Y = \overline{w}) &= E^{\mathbb{Z}^d}_{\mu_C}[X_k = \overline{w}_k, 0 \le k \le T, \text{ and } \widetilde{H}_C \circ \theta_T = \infty] \\ &= \text{cap}(C)^{-1} e_C(\overline{w}_0) P_{\overline{w}_0}[X_k = \overline{w}_k,\ 0 \le k \le T] e_C(\overline{w}_T), \end{aligned} \tag{3.13}$$

as a result of the simple Markov property and (3.8)–(3.10). We are now ready to state the main result of this section.

Theorem 3.1. ($d \ge 3$) *Assume that* (3.1), (3.4) *hold, and* $u \notin \bigcup_{i=1,2} \widetilde{C}(x_i)$, $w \in \bigcup_{i=1,2} \partial \widetilde{C}(x_i)$. *Then*

$$\| Q_{u,w} - Q \|_{TV} \le c \frac{L^2}{r}, \tag{3.14}$$

where for a signed measure ν *on* $\mathcal{W}$, $\|\nu\|_{TV} = \sum_{\overline{w} \in \mathcal{W}} |\nu(\overline{w})|$ *denotes the total variation of* ν.

Remark 3.2. It will be clear from the proof that the same result holds for collections x_i, $1 \le i \le M$, with $|x_i - x_j|_\infty \ge 2r + 3$ whenever $i \neq j$, $u \notin \bigcup_{1 \le i \le M} \widetilde{C}(x_i)$, $w \in \bigcup_{1 \le i \le M} \partial\widetilde{C}(x_i)$ (the $\partial\widetilde{C}(x_i)$, $1 \le i \le M$, are pairwise disjoint due to the above requirement), with L, r as in (3.1) and $N \ge M(2r+3)$. As will be clear from the proof below, the constant corresponding to (3.14) does not depend on M. For simplicity of notation we however restrict to the case $M = 2$.

Proof. We assume $w \in \partial\widetilde{C}(x_1)$ and consider $u \notin \widetilde{C}(x_1) \cup \widetilde{C}(x_2)$. The case where $w \in \partial\widetilde{C}(x_2)$ is treated analogously. Note that

$$Q_{u,w}(\widetilde{\mathcal{W}}) = 1, \quad \text{where} \tag{3.15}$$

$$\widetilde{\mathcal{W}} := \{\overline{w} = (\overline{w}_k)_{0 \le k \le T} \in \mathcal{W} : \overline{w}_k \in \widetilde{C} \text{ for } 0 \le k \le T\}, \tag{3.16}$$

and that for $\overline{w} \in \widetilde{\mathcal{W}}$,

$$Q_{u,w}(Y = \overline{w}) = A(\overline{w}) \Big/ \sum_{\overline{w}' \in \widetilde{\mathcal{W}}} A(\overline{w}'), \tag{3.17}$$

with the notation

$$\begin{aligned} A(\overline{w}) = P_u[&X_{R_1+k} = x_1 + \overline{w}_k,\ 0 \le k \le T, \\ &X_n \notin C(x_1) \text{ for } R_1 + T < n < D_1,\ X_{D_1} = w]. \end{aligned} \tag{3.18}$$

In what follows, when U is a subset of E (resp. $\mathbb{Z}^d$), we write $g_{E,U}(\cdot,\cdot)$ (resp. $g_{\mathbb{Z}^d,U}(\cdot,\cdot)$) to denote the Green function of the walk killed outside U, so that

$$g_{E,U}(x,y) = \sum_{k \ge 0} P_x[X_k = y,\ k < T_U], \qquad x, y \in E, \tag{3.19}$$

with a similar formula for $g_{\mathbb{Z}^d,U}$ where $P_x^{\mathbb{Z}^d}$ replaces P_x, and $x, y \in \mathbb{Z}^d$. We simply write $g_{\mathbb{Z}^d}(\cdot,\cdot)$ when $U = \mathbb{Z}^d$.

Summing over the values of the time of last visit to $(\bigcup_{i=1,2} \widetilde{C}(x_i))^c$ before D_1, we see that for $\overline{w} \in \widetilde{W}$,

$$A(\overline{w}) = \sum_{v,v'} g_{E,(\bigcup_{i=1,2} C(x_i))^c}(u,v) \frac{1}{2d} B_{v'}(\overline{w}), \tag{3.20}$$

where the above sum runs over $v \sim v'$ with $v \in \partial\widetilde{C}(x_1)$, $v' \in \widetilde{C}(x_1)$, with the notation

$$\begin{aligned} B_{v'}(\overline{w}) &= P_{v'}[R_1 + T < T_{\widetilde{C}(x_1)},\ X_{R_1+k} = \overline{w}_k + x_1,\ 0 \le k \le T, \\ &\qquad X_k \notin C(x_1) \text{ for } R_1 + T < k < D_1,\ X_{D_1} = w] \\ &= P_{\widehat{v}'}^{\mathbb{Z}^d}[H_C + T < T_{\widetilde{C}},\ X_{H_C+k} = \overline{w}_k,\ 0 \le k \le T, \\ &\qquad X_k \notin C \text{ for } H_C + T < k < T_{\widetilde{C}},\ X_{T_{\widetilde{C}}} = \widehat{w}], \end{aligned} \tag{3.21}$$

where $\widehat{z} = z - x_1$, using translation invariance and the identification of $\widetilde{C}(0)$ with $\widetilde{C}$. Summing over the values of the time of last visit to $\widetilde{C} \setminus C$ prior to H_C, we see that for $\overline{w} \in \widetilde{\mathcal{W}}$, v' as above,

$$\begin{aligned} B_{v'}(\overline{w}) &= \sum_{y'} g_{\mathbb{Z}^d, \widetilde{C}\setminus C}(\widehat{v}', y') \frac{1}{2d} P_{\overline{w}_0}[X_k = \overline{w}_k,\ 0 \le k \le T, \\ &\qquad\qquad X_k \notin C \text{ for } T < k < T_{\widetilde{C}},\ X_{T_{\widetilde{C}}} = \widehat{w}] \\ &= \sum_{y',z',w'} g_{\mathbb{Z}^d, \widetilde{C}\setminus C}(\widehat{v}', y') \frac{1}{2d} P_{\overline{w}_0}[X_k = \overline{w}_k,\ 0 \le k \le T] \frac{1}{2d} g_{\mathbb{Z}^d, \widetilde{C}\setminus C}(z', w') \frac{1}{2d} \end{aligned} \tag{3.22}$$

where y', z' run over the respective neighbors in C^c of $\overline{w}_0$ and $\overline{w}_T$, whereas w' runs over the neighbors in $\widetilde{C}$ of $\widehat{w}$, and we have used the simple Markov property at times $T+1$ and T, and summed over the values of the time of last visit to $\widetilde{C} \setminus C$ prior to the exit of $\widetilde{C}$ in $\widehat{w}$, to obtain the last expression. The next lemma contains a crucial decoupling estimate.

Lemma 3.3. $(d \ge 3,\ L \ge 1,\ 10L \le r)$ *For* $a \in \widetilde{C} \cap \partial(\widetilde{C}^c)$, $b \in \partial C$,

$$g_{\mathbb{Z}^d, \widetilde{C}\setminus C}(a, b) = P_b^{\mathbb{Z}^d}[T_{\widetilde{C}} < H_C]\, g_{\mathbb{Z}^d, \widetilde{C}}(a, 0)(1 + \psi_{a,b}), \tag{3.23}$$

where $\psi_{a,b}$ *is defined by this equality and*

$$|\psi_{a,b}| \le c_6 \frac{L^2}{r}. \tag{3.24}$$

Proof. For simplicity we write $g_U(\cdot,\cdot)$ and $g(\cdot,\cdot)$ in place of $g_{\mathbb{Z}^d,U}(\cdot,\cdot)$ and $g_{\mathbb{Z}^d}(\cdot,\cdot)$. Using the strong Markov property at time H_C, when $H_C < T_{\widetilde{C}}$, and the symmetry of the killed Green functions, one has

$$g_{\widetilde{C}}(a, b) = g_{\widetilde{C}\setminus C}(a, b) + E_b^{\mathbb{Z}^d}[g_{\widetilde{C}}(a, X_{H_C}),\ H_C < T_{\widetilde{C}}]. \tag{3.25}$$

Therefore we find

$$\begin{aligned} g_{\widetilde{C}\setminus C}(a, b) &= E_b^{\mathbb{Z}^d}[g_{\widetilde{C}}(a, b) - g_{\widetilde{C}}(a, X_{H_C}),\ H_C < T_{\widetilde{C}}] \\ &\quad + g_{\widetilde{C}}(a, b) P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}] \\ &= g_{\widetilde{C}}(a, 0) P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}] + (g_{\widetilde{C}}(a, b) - g_{\widetilde{C}}(a, 0)) P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}] \\ &\quad + E_b^{\mathbb{Z}^d}[g_{\widetilde{C}}(a, b) - g_{\widetilde{C}}(a, X_{H_C}), H_C < T_{\widetilde{C}}]. \end{aligned} \tag{3.26}$$

Note that $g_{\widetilde{C}}(a, \cdot)$ is a non-negative harmonic function on $\widetilde{C} \setminus \{a\}$. By the Harnack inequality (cf. Theorem 1.7.2 of [12], p. 42), and a standard covering argument (due to the fact that the cited theorem refers to Euclidean balls), we find

$$\sup_{|x|_\infty \le r/2} g_{\widetilde{C}}(a, x) \le c g_{\widetilde{C}}(a, 0) \le c g(a, 0). \tag{3.27}$$

Moreover by the gradient estimates in (a) of Theorem 1.7.1 of [12], p. 42, we see that

$$\sup_{|x|_\infty \le L, |e| \le 1} |g_{\widetilde{C}}(a, x+e) - g_{\widetilde{C}}(a, x)| \le \frac{c}{r} \sup_{|x|_\infty \le r/2} g_{\widetilde{C}}(a, x). \tag{3.28}$$

Combining (3.27), (3.28), we see that for all $a \in \widetilde{C} \cap \partial \widetilde{C}^c$,

$$\sup_{f \in C \cup \partial C} |g_{\widetilde{C}}(a, f) - g_{\widetilde{C}}(a, 0)| \le c \frac{L}{r} g_{\widetilde{C}}(a, 0). \tag{3.29}$$

Inserting this inequality in (3.26) we see that

$$\begin{aligned} & g_{\widetilde{C} \setminus C}(a, b) - g_{\widetilde{C}}(a, 0) P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}] =: R, \quad \text{with} \\ & |R| \le P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}] c \frac{L}{r} g_{\widetilde{C}}(a, 0) + c \frac{L}{r} g_{\widetilde{C}}(a, 0) \\ & \quad \le c \frac{L^2}{r} g_{\widetilde{C}}(a, 0) P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}], \end{aligned} \tag{3.30}$$

where in the last step we have used the lower bound

$$\begin{aligned} P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}] &\ge P_b^{\mathbb{Z}^d}[H_C > T_{2C}] \cdot \inf_{x \in (2C)^c} P_x^{\mathbb{Z}^d}[H_C = \infty] \\ &\ge c P_b^{\mathbb{Z}^d}[H_C > T_{2C}] \ge c P_1^{\mathbb{Z}}[H_0 > H_{L+1}] \ge \frac{c}{L}. \end{aligned}$$

Our claim (3.23), (3.24) now follows. □

We now continue the proof of Theorem 3.1. Note that by the strong Markov property applied at time H_C, and standard estimates on the Green function (cf. [12, p. 31]),

$$\begin{aligned} P_z^{\mathbb{Z}^d}[H_C < \infty] &\le \sum_{x \in C} g_{\mathbb{Z}^d}(z, x) \Big/ \inf_{y \in C} \sum_{x \in C} g_{\mathbb{Z}^d}(y, x) \\ &\le c \left(\frac{L}{r}\right)^{d-2} \quad \text{for } z \in \widetilde{C}^c. \end{aligned} \tag{3.31}$$

Also by similar estimates, and using if necessary the invariance principle to let the path move away, we see by (3.1) that

$$\sup_{z \in \widetilde{C}^c} P_z^{\mathbb{Z}^d}[H_C < \infty] \le c' < 1. \tag{3.32}$$

Hence for $b \in \partial C$, using the strong Markov property at time $T_{\widetilde{C}}$, we find that

$$\begin{aligned} 0 &\le P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}] - P_b^{\mathbb{Z}^d}[H_C = \infty] = P_b^{\mathbb{Z}^d}[H_C \circ \theta_{T_{\widetilde{C}}} < \infty, \ H_C > T_{\widetilde{C}}] \\ &\le P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}] c' \left(1 \wedge c \left(\frac{L}{r}\right)^{d-2}\right) \quad (\text{with } c' < 1). \end{aligned} \tag{3.33}$$

It thus follows that for $b \in \partial C$,

$$P_b^{\mathbb{Z}^d}[H_C > T_{\widetilde{C}}] = P_b^{\mathbb{Z}^d}[H_C = \infty](1+\epsilon_b) \quad \text{with}$$
$$0 \le \epsilon_b \le c\left(\frac{L}{r}\right)^{d-2} \le c_7 \frac{L^2}{r}. \tag{3.34}$$

We now assume for the time being (cf. (3.24), (3.34)) that

$$(c_6 + c_7)\frac{L^2}{r} \le \frac{1}{2}. \tag{3.35}$$

The case when (3.35) does not hold will be straightforwardly handled at the end of the proof. We then define, for $a \in \widetilde{C} \cap \partial(\widetilde{C}^c)$, $b \in \partial C$, with the notations of (3.23), (3.34),

$$\begin{aligned} &e^{\Gamma_{a,b}} = (1+\psi_{a,b})(1+\epsilon_b), \quad \text{so that} \\ &|\Gamma_{a,b}| \overset{(3.35)}{\le} c\frac{L^2}{r}, \quad \text{and} \quad g_{\mathbb{Z}^d,\widetilde{C}\setminus C}(a,b) = P_b^{\mathbb{Z}^d}[H_C = \infty] g_{\mathbb{Z}^d,\widetilde{C}}(a,0) e^{\Gamma_{a,b}}. \end{aligned} \tag{3.36}$$

Coming back to (3.20), (3.22), we see by (3.23), (3.34) that for $\overline{w} \in \widetilde{\mathcal{W}}$,

$$\begin{aligned} A(\overline{w}) = &\sum_{y',z'} \left(\frac{1}{2d}\right)^2 P_{y'}^{\mathbb{Z}^d}[H_C = \infty] P_{\overline{w}_0}^{\mathbb{Z}^d}[X_k = \overline{w}_k,\ 0 \le k \le T] P_{z'}^{\mathbb{Z}^d}[H_C = \infty] \\ &\times \left\{ \sum_{v,v',w'} \left(\frac{1}{2d}\right)^2 g_{E,(C(x_1)\cup C(x_2))^c}(u,v) g_{\mathbb{Z}^d,\widetilde{C}}(v',0) g_{\mathbb{Z}^d,\widetilde{C}}(w',0) e^{\Gamma_{v',y'}+\Gamma_{w',z'}} \right\}, \end{aligned} \tag{3.37}$$

where in the above sums y', z' run over C^c with $y' \sim \overline{w}_0$, $z' \sim \overline{w}_T$, v runs over $\partial\widetilde{C}(x_1)$, $v', w' \in \widetilde{C}$ with $v' \sim \widehat{v}$ $(= v - x_1)$, and $w' \sim \widehat{w} = w - x_1$. As a result we see that for $\overline{w}_1, \overline{w}_2 \in \widetilde{\mathcal{W}}$,

$$\frac{A(\overline{w}_1)}{A(\overline{w}_2)} = \frac{\widetilde{A}(\overline{w}_1)}{\widetilde{A}(\overline{w}_2)} e^{\Gamma(\overline{w}_1,\overline{w}_2)} \quad \text{with} \tag{3.38}$$

$$\begin{aligned} \widetilde{A}(\overline{w}) &= \sum_{y',z'} \left(\frac{1}{2d}\right)^2 P_{y'}^{\mathbb{Z}^d}[H_C = \infty] P_{\overline{w}_0}^{\mathbb{Z}^d}[X_k = \overline{w}_k,\ 0 \le k \le T] P_{z'}^{\mathbb{Z}^d}[H_C = \infty] \\ &\overset{(3.8)}{=} e_C(\overline{w}_0) P_{\overline{w}_0}^{\mathbb{Z}^d}[X_k = \overline{w}_k,\ 0 \le k \le T] e_C(\overline{w}_T) \\ &\overset{(3.13)}{=} \mathrm{cap}(C) Q(Y = \overline{w}) \quad \text{for } \overline{w} \in \widetilde{\mathcal{W}}, \end{aligned} \tag{3.39}$$

$$|\Gamma(\overline{w}_1, \overline{w}_2)| \le c\frac{L^2}{r}. \tag{3.40}$$

Inserting (3.38) into (3.17), we see that for $\overline{w} \in \widetilde{\mathcal{W}}$,

$$Q_{u,w}[Y = \overline{w}] = \frac{Q(Y = \overline{w})}{\sum_{\overline{w}' \in \widetilde{\mathcal{W}}} Q(Y = \overline{w}') e^{\Gamma(\overline{w}',\overline{w})}} = \frac{Q(Y = \overline{w})}{Q(\widetilde{\mathcal{W}})} e^{G(\overline{w})}, \tag{3.41}$$

where $|G(\overline{w})| \le cL^2/r$.

Note that by (3.12), (3.16), and the strong Markov property,

$$Q(\widetilde{\mathcal{W}}^c) = P^{\mathbb{Z}^d}_{\mu_C}[H_C \circ \theta_{T_{\widetilde{C}}} < \infty] \overset{(3.31),(3.32)}{\leq} c'\left(1 \wedge c\left(\frac{L}{r}\right)^{d-2}\right) \quad \text{with } c' < 1. \tag{3.42}$$

We thus find that

$$\begin{aligned} \|Q_{u,w} - Q\|_{TV} &= \sum_{\overline{w}\in\widetilde{\mathcal{W}}} |Q_{u,w}(Y = \overline{w}) - Q(Y = \overline{w})| + Q(\widetilde{\mathcal{W}}^c) \\ &= \sum_{\overline{w}\in\widetilde{\mathcal{W}}} \frac{Q(Y = \overline{w})}{Q(\widetilde{\mathcal{W}})} |\exp\{G(\overline{w})\} - 1 + Q(\widetilde{\mathcal{W}}^c)| + Q(\widetilde{\mathcal{W}}^c) \leq c\,\frac{L^2}{r}, \end{aligned} \tag{3.43}$$

using (3.41), (3.42). As a result we have proved (3.14) under (3.35). On the other hand, when (3.35) does not hold, $L^2/r \geq \frac{1}{2}(c_6 + c_7)^{-1}$, and

$$\|Q_{u,w} - Q\|_{TV} \leq 2 \leq 4(c_6 + c_7)\frac{L^2}{r},$$

so that adjusting the constant in (3.14) if necessary, we have completed the proof of Theorem 3.1. □

4. Volume estimate for the giant component

The main purpose of this section is to show that the giant component O in the vacant set left by the walk at time $t = uN^d$ (this component is well-defined on the event $\mathcal{G}_{\beta,t}$, cf. (2.53)) typically occupies a non-degenerate fraction of the volume of the torus E when N is large and u is chosen small. The statement (2.56) provides a local criterion, depending on the configuration of vacant sites left by the walk in a neighborhood of order const $\log N$ of a point $x \in E$, which ensures, when $\mathcal{G}_{\beta,t}$ occurs, that x belongs to O. By (2.57) this reduces the problem of proving the non-degeneracy of the volume of O to the question of showing that typically the asymptotic fraction of points x in E that fulfill the local condition $\mathcal{C}_{c_0,x,uN^d}$ is non-degenerate when u is small. By (2.4) this task is further reduced to the control on the variance of this quantity. It turns out that it is simpler to bound the variance of the fraction of points of E that satisfy a modified local condition where the fixed time $t = uN^d$ is replaced by a random time corresponding to the completion of const $u(\log N)^{2(d-2)}$ excursions of the walk to a neighborhood of order $(\log N)^2$ of the point (cf. (4.22)). The controls of Section 3 are then instrumental in bounding the variance of this modified ratio (cf. Proposition 4.2). Our main estimates on averages of suitable local functions are expressed in a general form (not specifically referring to (2.56)), and appear in Theorem 4.3, when $d \geq 3$. The applications to the vacant set, the giant component (when $d \geq d_0$, cf. Corollary 2.5), and the size of the largest ball in the vacant set are given in Corollaries 4.5, 4.6, 4.8.

We now begin with some additional notation. We consider $d \geq 3$, $L \geq 1$, $r \geq 10L$, $N \geq 10r$, $x \in E$, and recall the definition of $C(x) \subseteq \widetilde{C}(x)$ in (3.2). We consider some function ϕ, defined on the collection of subsets of $C(0)$:

$$\phi : A \subseteq C(0) \mapsto \phi(A) \in [0, 1]. \tag{4.1}$$

Typical examples to keep in mind are for instance

$$\phi_0(A) = 1\{0 \notin A\} \quad \text{for } A \subseteq C(0), \tag{4.2}$$
$$\phi_1(A) = 1\{\text{for some } F \in \mathcal{L}_2 \text{ with } 0 \in F,\ 0 \text{ is connected to } S(0, L) \text{ in } F \setminus A\}, \tag{4.3}$$

where we refer to (2.40) for the latter example. With ϕ as in (4.1), we then define, for $x \in E$ and $t \geq 0$,

$$h(x, t) = \phi((X_{[0,t]} \cap C(x)) - x). \tag{4.4}$$

Our chief task in this section consists in the derivation of appropriate lower bounds on ratios of the type

$$\Gamma_u = \frac{1}{N^d} \sum_{x \in E} h(x, uN^d) \quad \text{with } u > 0. \tag{4.5}$$

For $x \in E$ we introduce, in analogy to (1.9), (1.10),

$$B(x) = x + B \subseteq \widetilde{B}(x) = x + \widetilde{B} \quad (\text{so } C(x) \subsetneq \widetilde{C}(x) \subsetneq B(x) \subsetneq \widetilde{B}(x)) \tag{4.6}$$

as well as the successive returns to $B(x)$ and departures from $\widetilde{B}(x)$:

$$\mathcal{R}_k^x, \mathcal{D}_k^x, \quad k \geq 1. \tag{4.7}$$

We also consider (cf. (1.5)) the successive returns to $C(x)$ and departures from $\widetilde{C}(x)$:

$$R_k^x, D_k^x, \quad k \geq 1. \tag{4.8}$$

We begin with the following auxiliary result (note that r does not appear on the right-hand side of the inequalities):

Lemma 4.1. $(d \geq 3, L \geq 1, r \geq 10L, N \geq 10r)$ *There are constants $c_8 > c_9 > 0$ such that for $u > 0$, $x \in E$,*

$$P[R^x_{\ell^*(u)} \leq uN^d] \leq ce^{-cuL^{d-2}} \quad \textit{with} \quad \ell^*(u) = [c_8 uL^{d-2}], \tag{4.9}$$
$$P[D^x_{\ell_*(u)} \geq uN^d] \leq ce^{-cuL^{d-2}} \quad \textit{with} \quad \ell_*(u) = [c_9 uL^{d-2}]. \tag{4.10}$$

Proof. We begin with the proof of (4.9). We introduce, for $\ell \geq 1$,

$$Z^x_\ell = \sum_{m \geq 1} 1\{\mathcal{R}^x_\ell \leq R^x_m \leq \mathcal{D}^x_\ell\} = \sum_{m \geq 1} 1\{\mathcal{R}^x_\ell \leq D^x_m \leq \mathcal{D}^x_\ell\}. \tag{4.11}$$

By the strong Markov property at times D_m^x and $H_{\widetilde{C}(x)}$, we see that for $i \geq 0$, $\ell \geq 2$, P-a.s.,

$$\begin{aligned} P[Z_\ell^x > i \,|\, \mathcal{F}_{\mathcal{R}_\ell^x}] &= P_{X_{\mathcal{R}_\ell^x}}[R_{i+1}^x < T_{\widetilde{B}(x)}] \\ &\leq P_{X_{\mathcal{R}_\ell^x}}[H_{\widetilde{C}(x)} < T_{\widetilde{B}(x)}]\big(\sup_{|z-x|_\infty \in \{r, r+1\}} P_z[H_{C(x)} < T_{\widetilde{B}(x)}]\big)^{i+1} \\ &= P_{X_{\mathcal{R}_\ell^x}}[H_{\widetilde{C}(x)} < T_{\widetilde{B}(x)}]\big(\sup_{|z-x|_\infty = r} P_z[H_{C(x)} < T_{\widetilde{B}(x)}]\big)^{i+1}. \end{aligned} \tag{4.12}$$

Analogously we have, for $i \geq 0$,

$$P[Z_1^x > i] \leq P[H_C < T_{\widetilde{B}}]\big(\sup_{|z-x|_\infty = r} P_z[H_{C(x)} < T_{\widetilde{B}(x)}]\big)^i. \tag{4.13}$$

Using similar bounds to (3.31), (3.32), we find that for $\ell \geq 2$, $i \geq 0$, P-a.s.,

$$P[Z_\ell^x > i \,|\, \mathcal{F}_{\mathcal{R}_\ell^x}] \leq c' \wedge \left(c\left(\frac{r}{N}\right)^{d-2}\right) \cdot \left\{c' \wedge \left(c\left(\frac{L}{r}\right)^{d-2}\right)\right\}^{i+1} \quad \text{with } c' < 1.$$

Using the inequality

$$\begin{aligned} &P[H_{C(x)} < T_{\widetilde{B}(x)}] \\ &\leq E\Big[\sum_{k\geq 0} 1\{X_k \in C(x),\ k < T_{\widetilde{B}(x)}\}\Big]\Big/ \inf_{y \in C(x)} E_y\Big[\sum_{k \geq 0} 1\{X_k \in C(x),\ k < T_{\widetilde{B}(x)}\}\Big], \end{aligned} \tag{4.14}$$

a similar upper bound to (1.34) on the numerator and a lower bound of type cL^2 on the denominator with the help of the invariance principle, we find that

$$P[H_{C(x)} < T_{\widetilde{B}(x)}] \leq c\left(\frac{L}{N}\right)^{d-2},$$

and it is also straightforward to argue, by applying the invariance principle and similar arguments to the derivation of (3.32), that the above probability is bounded by some $c'' < 1$. Coming back to (4.12), (4.13), we thus see that

$$\begin{aligned} &P[Z_\ell^x > i \,|\, \mathcal{F}_{\mathcal{R}_\ell^x}] \leq p_0 p^{i+1}, \quad P[Z_1^x > i] \leq p_0 p^{i+1} \quad \text{for } i \geq 0,\ \ell \geq 2, \\ &\text{with } p_0 = c_{10} \wedge \left(c\left(\frac{r}{N}\right)^{d-2}\right),\ p = c_{10} \wedge \left(c\left(\frac{L}{r}\right)^{d-2}\right),\ c_{10} < 1. \end{aligned} \tag{4.15}$$

By stochastic domination, we thus see that for $\lambda > 0$ with $e^\lambda p < 1$, and $\ell \geq 2$,

$$\begin{aligned} &E[\exp\{\lambda Z_\ell^x\} \,|\, \mathcal{F}_{\mathcal{R}_\ell^x}] \leq 1 - p_0 p + \sum_{k \geq 1} e^{\lambda k} p_0 p^k (1-p) = 1 + p_0 p \frac{e^\lambda - 1}{1 - e^\lambda p}, \\ &E[\exp\{\lambda Z_1^x\}] \leq 1 + p_0 p \frac{e^\lambda - 1}{1 - e^\lambda p}. \end{aligned} \tag{4.16}$$

As a result we find that with the notation below (1.12),

$$P[Z_1 + \cdots + Z_{k^*} \geq n] \leq \exp\{-\lambda n\}\left(1 + p_0 p \frac{e^\lambda - 1}{1 - e^\lambda p}\right)^{k^*}$$
$$\leq \exp\left\{-\lambda n + k^* p_0 p \frac{e^\lambda - 1}{1 - e^\lambda p}\right\}. \tag{4.17}$$

Note that $k^* p_0 p \leq cuL^{d-2}$, and choosing λ so that $e^\lambda c_{10} = \frac{1}{2}(1 + c_{10})$ (recall $p \leq c_{10} < 1$), we thus obtain

$$P[R_n^x \leq uN^d] \leq P[\mathcal{R}_{k^*}^x \leq uN^d] + P[\mathcal{R}_{k^*}^x > uN^d \geq R_n^x]$$
$$\overset{(1.11),(4.17)}{\leq} c\exp\{-cuN^{d-2}\} + c\exp\{-\lambda n + cuL^{d-2}\}, \tag{4.18}$$

and (4.9) follows straightforwardly.

We now turn to the proof of (4.10). We use a bound from below on $P[H_{C(x)} < T_{\widetilde{B}(x)}]$ and $P_z[H_{C(x)} < T_{\widetilde{B}(x)}]$ with a similar right-hand side to that in (4.14), except that inf is replaced with sup, and in the case of the second probability E is replaced with E_z (see also (1.57)). Then by standard Green function estimates (see for instance (1.11) of [7]), we obtain

$$P[Z_\ell^x > 0 \,|\, \mathcal{F}_{\mathcal{R}_\ell^x}] \overset{P\text{-a.s.}}{\geq} c\left(\frac{L}{N}\right)^{d-2} \quad \text{for } \ell \geq 2, \quad P[Z_1^x > 0] \geq c\left(\frac{L}{N}\right)^{d-2}. \tag{4.19}$$

As a result we see that for $\lambda > 0$, $\ell \geq 2$,

$$E[\exp\{-\lambda Z_\ell^x\} \,|\, \mathcal{F}_{\mathcal{R}_\ell^x}] \leq 1 - (1 - e^{-\lambda})c\left(\frac{L}{N}\right)^{d-2},$$
$$E[\exp\{-\lambda Z_1^x\}] \leq 1 - (1 - e^{-\lambda})c\left(\frac{L}{N}\right)^{d-2},$$

so that for $n \geq 1$ (with the convention that the sum in the probability below vanishes when $k_* \leq 1$),

$$P[Z_1 + \cdots + Z_{(k_*-1)_+} < n] \leq \exp\left\{\lambda n - (k_* - 1)_+(1 - e^{-\lambda})c\left(\frac{L}{N}\right)^{d-2}\right\}, \tag{4.20}$$

where $(k_* - 1)_+ c(L/N)^{d-2} \geq cuL^{d-2} - c$ (cf. below (1.12)). We then see that

$$P[D_n^x \geq uN^d] \leq P[\mathcal{R}_{k_*}^x \geq uN^d] + P[\mathcal{R}_{k_*}^x < uN^d,\ D_n^x \geq \mathcal{R}_{k_*}^x]$$
$$\overset{(1.12)}{\leq} c\exp\{-cuN^{d-2}\} + P[Z_1 + \cdots + Z_{(k_*-1)_+} < n]$$
$$\overset{(4.20)}{\leq} c\exp\{-cuN^{d-2}\} + c\exp\{\lambda n - (1 - e^{-\lambda})cuL^{d-2}\}. \tag{4.21}$$

If we choose λ so that $e^{-\lambda} = 1/2$, the claim (4.10) follows straightforwardly. □

We now introduce a modification of Γ_u in (4.5), which is more convenient when bounding its variance. Namely, using (4.1), (4.4), and the notation from (4.9), we define

$$\widetilde{\Gamma}_u = \frac{1}{N^d} \sum_{x \in E} h(x, D^x_{\ell^*(u)}) \quad \text{for } u > 0. \tag{4.22}$$

Our main estimate on the variance of $\widetilde{\Gamma}_u$ comes in the next proposition. In what follows, var and cov denote the variance and covariance under P.

Proposition 4.2. *($d \geq 3$, $L \geq 1$, $N \geq 10r$, $r \geq 10L$, under* (4.1)*)*

$$\mathrm{var}(\widetilde{\Gamma}_u) \leq c\left(\left(\frac{r}{N}\right)^d + u\frac{L^d}{r}\right) \quad \textit{for } u > 0. \tag{4.23}$$

Proof. When $\ell^*(u) = 0$, with our conventions we see that $\widetilde{\Gamma}_u = N^{-d} \sum_{x \in E} \phi(C(x) \cap \{X_0\} - x)$, a non-random quantity as follows from translation invariance. The claim (4.23) is then trivially satisfied. We thus assume from now on that $\ell^*(u) \geq 1$. We then consider an integer r as in (4.23), and write

$$\begin{aligned}\mathrm{var}(\widetilde{\Gamma}_u) &= \frac{1}{N^{2d}} \sum_{x_1, x_2 \in E} \mathrm{cov}(h(x_1, D^{x_1}_{\ell^*(u)}), h(x_2, D^{x_2}_{\ell^*(u)})) \\ &\leq c\left(\frac{r}{N}\right)^d + \sup_{|x_1 - x_2|_\infty \geq 2r+3} |\mathrm{cov}(h(x_1, D^{x_1}_{\ell^*(u)}), h(x_2, D^{x_2}_{\ell^*(u)}))|. \end{aligned} \tag{4.24}$$

We recall the notations R_k, D_k, $k \geq 1$, introduced below (3.4), and write, for $i = 1, 2$,

$$\begin{aligned} n_1^{x_i} &= \inf\{k \geq 1 : X_{R_k} \in C(x_i)\}, \\ n_{j+1}^{x_i} &= \inf\{k > n_j^{x_i} : X_{R_k} \in C(x_i)\}, \quad j \geq 1, \end{aligned} \tag{4.25}$$

The relation between $R_k^{x_i}$, $D_k^{x_i}$, $k \geq 1$, for $i = 1, 2$, and R_k, D_k, $k \geq 1$, is the following: one has P-a.s.,

$$R_k^{x_i} = R_{n_k^{x_i}}, \quad D_k^{x_i} = D_{n_k^{x_i}} \quad \text{for } k \geq 1,\ i = 1, 2. \tag{4.26}$$

We then introduce the constant (cf. (4.9), (4.10))

$$c_{11} = c_8 / c_9 > 1. \tag{4.27}$$

We recall the definition (3.5) and denote by $e_k(\cdot)$ the P-a.s. well-defined centered excursion

$$e_k(m) = X_{R_k + m} - x_i, \quad 0 \leq m \leq L_k - R_k, \quad \text{on } \{X_{R_k} \in C(x_i)\},\ i = 1, 2. \tag{4.28}$$

We recall our tacit identification of $\widetilde{C}(0) \subseteq E$ with $\widetilde{C}$ in $\mathbb{Z}^d$ (see below (3.3)), so that P-a.s., $e_k(\cdot) \in \widetilde{\mathcal{W}} \subseteq \mathcal{W}$ (cf. (3.6), (3.16)). We also consider the k-th excursion to $C(x_i)$, after centering at the origin, which is also P-a.s. well-defined:

$$e_k^i(\cdot) = e_{n_k^{x_i}}(\cdot), \quad k \geq 1,\ i \in \{1, 2\}, \tag{4.29}$$

as well as its trace

$$\mathcal{S}_k^i = \operatorname{Im} e_k^i, \tag{4.30}$$

where for $\overline{w} = (\overline{w}_m)_{0 \le m \le T} \in \mathcal{W}$, $\operatorname{Im} \overline{w} = \{w_0, \dots, w_T\} \subseteq \mathbb{Z}^d$. With the above notation, we see that P-a.s.,

$$X_{[0, D^{x_i}_{\ell^*(u)}]} \cap C(x_i) - x_i = (\mathcal{S}_1^i \cup \dots \cup \mathcal{S}^i_{\ell^*(u)}) \cap C, \tag{4.31}$$

$$h(x_i, D^{x_i}_{\ell^*(u)}) = G(e_1^i, \dots, e^i_{\ell^*(u)}) \quad \text{for } i = 1, 2, \tag{4.32}$$

where G is the function from $\mathcal{W}^{\ell^*(u)}$ into [0, 1] defined by (cf. (4.1))

$$G(\overline{w}_1, \dots, \overline{w}_{\ell^*(u)}) = \phi((\operatorname{Im} \overline{w}_1 \cup \dots \cup \operatorname{Im} \overline{w}_{\ell^*(u)}) \cap C). \tag{4.33}$$

We now consider two [0, 1]-valued functions G_1, G_2 on $\mathcal{W}^{\ell^*(u)}$ (we are especially interested in the case $G_i = G$ or $G_i = 1$), and write

$$H_i = G_i(e_1^i, \dots, e^i_{\ell^*(u)}), \quad i = 1, 2. \tag{4.34}$$

We see that for $z \notin \widetilde{C}(x_1) \cup \widetilde{C}(x_2)$,

$$E_z[H_1 H_2] = \sum_{\mathcal{K}} E_z[H_1 H_2,\ A_{(\overline{k}^1, \overline{k}^2)}], \tag{4.35}$$

where $\mathcal{K}$ denotes the set of ordered pairs of $\ell^*(u)$-uples of integers

$$\begin{aligned} &(\overline{k}^1, \overline{k}^2) \text{ with } 1 \le \overline{k}_1^i < \dots < \overline{k}^i_{\ell^*(u)} \text{ for } i = 1, 2, \\ &\text{with all } \overline{k}^i_j \text{ distinct for } 1 \le j \le \ell^*(u),\ i = 1, 2, \end{aligned} \tag{4.36}$$

and for $(\overline{k}^1, \overline{k}^2) \in \mathcal{K}$, we write

$$A_{(\overline{k}^1, \overline{k}^2)} = \{n_m^{x_i} = \overline{k}^i_m \text{ for } 1 \le m \le \ell^*(u), i = 1, 2\}. \tag{4.37}$$

We introduce the σ-algebra

$$\mathcal{E} = \text{the } P\text{-completion of } \sigma(X_{D_k}, k \ge 1). \tag{4.38}$$

Note that $n_m^{x_i}$, $i \in \{1, 2\}$, $m \ge 1$, are $\mathcal{E}$-measurable, so that

$$A_{(\overline{k}^1, \overline{k}^2)} \in \mathcal{E} \quad \text{for any } (\overline{k}^1, \overline{k}^2) \in \mathcal{K}. \tag{4.39}$$

Using the strong Markov property at the times D_m, $m \le \overline{k} := \overline{k}^1_{\ell^*(u)} \vee \overline{k}^2_{\ell^*(u)}$, we see that for $(\overline{k}^1, \overline{k}^2) \in \mathcal{K}$, P-a.s. on $A_{(\overline{k}^1, \overline{k}^2)}$,

$$\begin{aligned} &E_z[H_1 H_2 \mid \mathcal{E}] \\ &= \int G_1(\overline{w}_{\overline{k}_1^1}, \dots, \overline{w}_{\overline{k}^1_{\ell^*(u)}}) G_2(\overline{w}_{\overline{k}_1^2}, \dots, \overline{w}_{\overline{k}^2_{\ell^*(u)}}) \prod_{m=1}^{\overline{k}} Q_{X_{D_{m-1}}, X_{D_m}}(d\overline{w}_m), \end{aligned} \tag{4.40}$$

where we have used the notation (3.7) and the convention $X_{D_0} = z$ when $m = 1$.

We can now find, for each $u \notin \widetilde{C}(x_1) \cup \widetilde{C}(x_2)$, $w \in \partial\widetilde{C}(x_1) \cup \partial\widetilde{C}(x_2)$, a coupling $\widetilde{Q}_{u,w}(d\overline{w}, d\overline{w}')$ on $\mathcal{W} \times \mathcal{W}$ such that (see (3.12))

$$\text{under the first (resp. second) canonical coordinate the image of } \widetilde{Q}_{u,w} \text{ is } Q_{u,w} \text{ (resp. } Q\text{)}, \tag{4.41}$$

$$\widetilde{Q}_{u,w}(\overline{w} \neq \overline{w}') = \frac{1}{2}\|Q_{u,w} - Q\|_{TV} \overset{(3.14)}{\leq} c\frac{L^2}{r}; \tag{4.42}$$

for the construction of $\widetilde{Q}_{u,w}$ see for instance Theorem 5.2, p. 19 of [13]. We thus see that for $(\overline{k}^1, \overline{k}^2) \in \mathcal{K}$, P-a.s. on $A_{(\overline{k}^1,\overline{k}^2)}$,

$$\begin{aligned}
&\Big|E_z[H_1 H_2 \,|\, \mathcal{E}] - \prod_{i=1}^{2} E^{Q^{\otimes \ell^*(u)}}[G_i]\Big| \\
&\quad = \Big| \int G_1(\overline{w}_{\overline{k}^1_1}, \dots, \overline{w}_{\overline{k}^1_{\ell^*(u)}}) G_2(\overline{w}_{\overline{k}^2_1}, \dots, \overline{w}_{\overline{k}^2_{\ell^*(u)}}) \\
&\qquad - G_1(\overline{w}'_{\overline{k}^1_1}, \dots, \overline{w}'_{\overline{k}^1_{\ell^*(u)}}) G_2(\overline{w}'_{\overline{k}^2_1}, \dots, \overline{w}'_{\overline{k}^2_{\ell^*(u)}}) \prod_{m=1}^{\overline{k}} \widetilde{Q}_{X_{D_{m-1}}, X_{D_m}}(d\overline{w}_m, d\overline{w}'_m)\Big| \\
&\quad \leq 2\ell^*(u) \sup_{m \in \{\overline{k}^i_j \,:\, i=1,2,\, 1 \leq j \leq \ell^*(u)\}} \widetilde{Q}_{X_{D_{m-1}}, X_{D_m}}(\overline{w}_m \neq \overline{w}'_m) \\
&\quad \overset{(4.42)}{\leq} c\ell^*(u)\frac{L^2}{r} \overset{(4.9)}{\leq} cu\frac{L^d}{r}.
\end{aligned} \tag{4.43}$$

Hence by (4.35), (4.39), we see that for $z \notin \widetilde{C}(x_1) \cup \widetilde{C}(x_2)$,

$$\Big|E_z[H_1 H_2] - \prod_{i=1}^{2} E^{Q^{\otimes \ell^*(u)}}[G_i]\Big| \leq cu\frac{L^d}{r},$$

and hence

$$\Big|E[H_1 H_2] - \prod_{i=1}^{2} E^{Q^{\otimes \ell^*(u)}}[G_i]\Big| \leq c\bigg(\Big(\frac{r}{N}\Big)^d + u\frac{L^d}{r}\bigg). \tag{4.44}$$

Choosing $G_i = G$ (cf. (4.33)) or $G_i = 1$, we see by (4.32) that the last term in the second line of (4.24) is smaller than $c((r/N)^d + uL^d/r)$. By (4.24), the claim (4.23) follows. □

If the function ϕ in (4.1) is decreasing, i.e. $\phi(A) \geq \phi(A')$ for $A \subseteq A' \subseteq C(0)$, then we can easily transfer controls on $\Gamma_{\cdot}$ from controls on $\widetilde{\Gamma}$.

Theorem 4.3. $(d \geq 3,\ L \geq 1,\ N \geq 100L)$ *Assume that* ϕ *in* (4.1) *is decreasing. Then for* $u > 0$, $s > 0$,

$$P[\Gamma_u < E[\Gamma_{c_{11}u}] - c\exp\{-cuL^{d-2}\} - s] \leq c\frac{\sigma^2_{u,L,N}}{s^2} + cN^d\exp\{-cuL^{d-2}\}, \tag{4.45}$$

and

$$P[\Gamma_u > E[\Gamma_{c_{11}^{-1}u}] + c\exp\{-cuL^{d-2}\} + s] \le c\frac{\sigma^2_{u,L,N}}{s^2} + cN^d\exp\{-cuL^{d-2}\}, \qquad (4.46)$$

where $c_{11} > 1$ *is defined in* (4.27) *and*

$$\sigma^2_{u,L,N} := \inf\left\{\left(\frac{r}{N}\right)^d + u\frac{L^d}{r} : 10L \le r \le \frac{N}{10}\right\}. \qquad (4.47)$$

Proof. Choose r as in (4.47) and define $\widetilde{\Gamma}_u$ as in (4.22). Since ϕ is decreasing, we see that

$$\begin{aligned} E[\Gamma_u] - E[\widetilde{\Gamma}_u] &= \frac{1}{N^d}\sum_{x\in E} E[h(x, uN^d) - h(x, D^x_{\ell^*(u)})] \\ &\ge -\frac{1}{N^d}\sum_{x\in E} P[D^x_{\ell^*(u)} < uN^d] \overset{(4.9)}{\ge} -ce^{-cuL^{d-2}}, \end{aligned}$$

and using the fact that $\ell^*(u/c_{11}) = \ell_*(u)$ (cf. (4.27), (4.9), (4.10)), we also have

$$\begin{aligned} E[\widetilde{\Gamma}_{u/c_{11}}] - E[\Gamma_u] &= \frac{1}{N^d}\sum_{x\in E} E[h(x, D^x_{\ell_*(u)}) - h(x, uN^d)] \\ &\ge -\frac{1}{N^d}\sum_{x\in E} P[D^x_{\ell_*(u)} \ge uN^d] \overset{(4.10)}{\ge} -ce^{-cuL^{d-2}}. \end{aligned}$$

As a result we find that

$$E[\Gamma_{c_{11}u}] - ce^{-cuL^{d-2}} \le E[\widetilde{\Gamma}_u] \le E[\Gamma_u] + ce^{-cuL^{d-2}} \quad \text{for } u > 0. \qquad (4.48)$$

In the same fashion we also find that for $u > 0$,

$$P[\Gamma_u < \widetilde{\Gamma}_u] \le cN^d e^{-cuL^{d-2}}, \qquad P[\widetilde{\Gamma}_{c_{11}^{-1}u} < \Gamma_u] \le cN^d e^{-cuL^{d-2}}. \qquad (4.49)$$

Hence using the first inequalities in (4.48) and (4.49), we find that for $u, s > 0$,

$$\begin{aligned} P[\Gamma_u < E[\Gamma_{c_{11}u}] - ce^{-cuL^{d-2}} - s] &\le P[\widetilde{\Gamma}_u < E[\widetilde{\Gamma}_u] - s] + cN^d e^{-cuL^{d-2}} \\ &\le \frac{\mathrm{var}(\widetilde{\Gamma}_u)}{s^2} + cN^d e^{-cuL^{d-2}}, \end{aligned}$$

and by (4.23), optimizing over r, the claim (4.45) follows. Using the rightmost inequalities of (4.49) and of (4.48), with $c_{11}^{-1}u$ in place of u, in the case of (4.48), we analogously obtain (4.46). □

Remark 4.4. In the applications we discuss below, we will choose $L = [(\log N)^2]$, so that for given $u > 0$ and $N \ge c(u)$,

$$\sigma^2_{u,L,N} \le cu^{\frac{d}{d+1}} L^{\frac{d^2}{d+1}} N^{-\frac{d}{d+1}} \le cu^{\frac{d}{d+1}} (\log N)^{\frac{2d^2}{d+1}} N^{-\frac{d}{d+1}}, \qquad (4.50)$$

as follows from a straightforward upper bound of the expression in (4.47).

We now turn to the first application of Theorem 4.3 that sharpens (1.13) to an estimate of the relative volume of the vacant set left by the walk at time uN^d.

Corollary 4.5. $(d \geq 3)$

$$\lim_N P[e^{-cu} \leq |E \setminus X_{[0,uN^d]}|/N^d \leq e^{-c'u}] = 1 \quad \textit{for } u > 0. \tag{4.51}$$

Proof. We choose $L = [(\log N)^2]$ and $\phi = \phi_0$ (cf. (4.2)), so that

$$\Gamma_u \overset{(4.5)}{=} \frac{1}{N^d} |E \setminus X_{[0,uN^d]}| \quad \text{for } u > 0, \tag{4.52}$$

and by translation invariance

$$E[\Gamma_u] = E[h(0, uN^d)] = P[0 \notin X_{[0,uN^d]}]. \tag{4.53}$$

Note that with the above choice for L, in view of (4.50), $\sigma_{u,L,N}$ and $N^d e^{-cuL^{d-2}}$ tend to 0 as N tends to infinity. If we choose for instance $s = \sqrt{\sigma_{u,L,N}}$, the claim (4.51) follows straightforwardly from (4.45), (4.46) and our estimates in (1.13) on $E[\Gamma_{c_{11}u}]$ and $E[\Gamma_{c_{11}^{-1}u}]$. □

We recall that on the event $\mathcal{G}_{\beta,t}$ defined in (2.52), the vacant set left by the walk at time t contains a well-defined unique giant component O (cf. (2.53)), and $\mathcal{G}_{\beta,uN^d}$ is typical under P for large N when $d \geq d_0$ and u is small (cf. (2.57)). As we will now see, in this regime O also typically occupies a non-degenerate fraction of the volume of E.

Corollary 4.6. $(d \geq d_0$, *cf.* (2.41)) *For any* $\beta, \gamma \in (0,1)$, *one has*

$$\lim_N P[\mathcal{G}_{\beta,uN^d} \cap \{|O|/N^d \geq \gamma\}] = 1 \quad \textit{for small } u > 0. \tag{4.54}$$

Proof. We choose $L = [(\log N)^2] \vee [c_0 \log N]$ (cf. Corollaries 2.5 and 2.6), and

$$\phi(A) = 1\{\text{for some } F \in \mathcal{L}_2 \text{ with } 0 \in F, 0 \text{ is connected to } S(0, L_0) \text{ in } F \setminus A\} \tag{4.55}$$

for any $A \subseteq C(0)$ $(= B(0, L))$, with L_0 as in (2.53). In this case for large N we have (cf. (2.40), (4.4), (4.5))

$$\Gamma_u = \frac{1}{N^d} \sum_{x \in E} 1_{\mathcal{C}_{c_0,x,uN^d}} \overset{(2.56)}{\leq} \frac{|O|}{N^d} \quad \text{on } \mathcal{G}_{\beta,uN^d}, \tag{4.56}$$

and by translation invariance we find

$$E[\Gamma_u] = P[\mathcal{C}_{c_0,0,uN^d}]. \tag{4.57}$$

As already mentioned below (4.53), $\sigma_{u,L,N}$ and $N^d e^{-cuL^{d-2}}$ tend to 0 as N tends to infinity. We can choose $s = \sqrt{\sigma_{u,L,N}}$ in (4.45), so that

$$\lim_N P[\Gamma_u \geq E[\Gamma_{c_{11}u}] - ce^{-cuL^{d-2}} - \sqrt{\sigma_{u,L,N}}] = 1.$$

The claim (4.54) then follows from (2.43), (2.57), and (4.56). □

Remark 4.7. 1) When $d \geq d_0$, the above corollary shows that for small $u > 0$, when N becomes large the giant component typically has non-degenerate volume in E. However, this does not rule out the existence of other components in the vacant set with non-degenerate volume. Note that by the definition of the giant O (cf. (2.53)) such components do not contain any connected sets $A \in \mathcal{A}_2$ of $|\cdot|_\infty$-diameter $L_0 = [c_0 \log N]$ and in particular any segment of length L_0.

2) When $d \geq 3$ and $u > 0$, the set visited by the walk up to time uN^d typically constitutes a giant component as well. Indeed, by Corollary 4.5 it typically occupies a non-degenerate fraction of the volume of E when N is large. Moreover, by a straightforward modification of (1.24) (see also Remark 1.3), we see that when $8\widetilde{L} \leq N$ and $u > 0$,

$$\begin{aligned} P[H_{B(0,\widetilde{L})} > uN^d] &\leq c\exp\{-cuN^{d-2}\} + (1 - c(\widetilde{L}/N)^{d-2})_+^{k_*-1} \\ &\leq c\exp\{-cu\widetilde{L}^{d-2}\}, \end{aligned} \tag{4.58}$$

using the definition of k_* below (1.12). In particular, choosing

$$\widetilde{L} = L_1 := \left[c_{12}\left(\frac{\log N}{u}\right)^{1/(d-2)}\right],$$

we find that

$$\lim_N P[\text{for some } x \text{ in } E,\ X_{[0,uN^d]} \cap B(x, L_1) = \emptyset] = 0 \quad \text{for all } u > 0. \tag{4.59}$$

So the set visited by the walk is ubiquitous as well, and typically comes within distance of order $(\log N)^{1/(d-2)}$ from any point of E. □

In fact, holes in the vacant set of order $(\log N)^{1/(d-2)}$ do occur as well. More precisely, consider the maximal radius of an $|\cdot|_\infty$-ball contained in the vacant set at time t:

$$\widehat{L}(t) = \sup\{m \geq 0 : \text{for some } x \text{ in } E,\ X_{[0,t]} \cap B(x,m) = \emptyset\}, \tag{4.60}$$

with the convention that $\widehat{L}(t) = 0$ when the set on the right-hand side of (4.60) is empty.

Corollary 4.8. $(d \geq 3)$ *There exists* c_{13} $(< c_{12})$ *such that*

$$\lim_N P[L_2 \leq \widehat{L}(uN^d) \leq L_1] = 1 \quad \textit{for } u > 0, \tag{4.61}$$

with L_1 *defined above* (4.59) *and* $L_2 = \left[c_{13}\left(\frac{\log N}{u}\right)^{1/(d-2)}\right]$.

Proof. In view of (4.59) we only need to prove the lower bound. The argument uses a variation on the proof of Corollary 4.5. For $\widetilde{L} \leq (\log N)^2$ and large N, by a straightforward modification of (1.25), we see that

$$\begin{aligned} P[H_{B(0,\widetilde{L})} > uN^d] &\geq -c\exp\{-cuN^{d-2}\} + c(1 - c(\widetilde{L}/N)^{d-2})^{k^*} \\ &\geq c\exp\{-cu\widetilde{L}^{d-2}\}, \end{aligned} \tag{4.62}$$

using the definition of k^* below (1.12). Proceeding as in Corollary 4.5, we then choose $L = [(\log N)^2]$, and the decreasing function

$$\phi : A \subseteq C(0) \mapsto \phi(A) = 1\{B(0, \widetilde{L}) \cap A = \emptyset\}.$$

With this choice we find that

$$\Gamma_u \stackrel{(4.5)}{=} \frac{1}{N^d} \sum_{x \in E} 1\{B(x, \widetilde{L}) \cap X_{[0,uN^d]} = \emptyset\}.$$

Setting $\widetilde{L} = \big[c_{13}\big(\frac{\log N}{u}\big)^{1/(d-2)}\big]$ with c_{13} small enough, we see by translation invariance that for large N,

$$E[\Gamma_{c_{11}u}] = P[H_{B(0,\widetilde{L})} > c_{11}uN^d] \geq N^{-1/6}. \tag{4.63}$$

We then choose $s = \sqrt{\sigma_{u,L,N}}$ in (4.45), and note by (4.50) that for large N, $s = \sqrt{\sigma_{u,L,N}}$ is much smaller than $N^{-1/6}$, and that $\sigma_{u,L,N}$ and $N^d e^{-cuL^{d-2}}$ tend to 0 as N tends to infinity. As a result we obtain

$$\lim_N P[\Gamma_u \leq \tfrac{1}{2} N^{-1/6}] = 0. \tag{4.64}$$

This is more than enough to prove the lower estimate in (4.61). This concludes the proof of Corollary 4.8. □

The above result exhibits a different asymptotic behavior from that of Bernoulli bond- (or site-) percolation on E, where for large N the largest $|\cdot|_\infty$-ball contained in a cluster typically has size of order $(\log N)^{1/d}$, which is much smaller than $(\log N)^{1/(d-2)}$.

Acknowledgements. We wish to thank Laurent Goergen and Roey Izkovsky for their simulations, as well as David Windisch for his comments on a previous version of this work. Alain-Sol Sznitman wants to thank Amir Dembo for helpful discussions.

References

[1] Aldous, D.: On the time taken by random walks on finite groups to visit every state. Z. Wahrsch. Verw. Gebiete **62**, 361–374 (1983) Zbl 0488.60011 MR 0688644

[2] Aldous, D., Fill, J.: Reversible Markov chains and random walks on graphs. http://www.stat.berkeley.edu/~aldous/RWG/book.html.

[3] Alon, N., Benjamini, I., Stacey, A.: Percolation on finite graphs and isoperimetric inequalities. Ann. Probab. **32**, 1727–1745 (2004) Zbl 1046.05071 MR 2073175

[4] Brummelhuis, M. J. A. M., Hilhorst, M. J.: Covering a finite lattice by a random walk. Phys. A **176**, 387–408 (1991) MR 1130067

[5] Dembo, A., Peres, Y., Rosen, J., Zeitouni, O.: Cover times for Brownian motion and random walks in two dimensions. Ann. of Math. **160**, 433–464 (2004) Zbl 1068.60018 MR 2123929

[6] Dembo, A., Peres, Y., Rosen, J., Zeitouni, O.: Late points for random walks in two dimensions. Ann. Probab. **34**, 219–263 (2006) Zbl 1100.60057 MR 2206347

[7] Dembo, A., Sznitman, A. S.: On the disconnection of a discrete cylinder by a random walk. Probab. Theory Related Fields **136**, 321–340 (2006) Zbl 1105.60029 MR 2240791

[8] Durrett, R.: Probability: Theory and Examples. Wadsworth and Brooks/Cole, Pacific Grove (1991) Zbl 0709.60002 MR 1068527

[9] Grigoryan, A., Telcs, A.: Sub-Gaussian estimates of heat kernels on infinite graphs. Duke Math. J. **109**, 451–510 (2001) Zbl 1010.35016 MR 1853353

[10] Heydenreich, M., van der Hofstadt, R.: Random graph asymptotics on high dimensional tori. Comm. Math. Phys. **270**, 335–358 (2007) Zbl pre05159687 MR 2276449

[11] Kesten, H.: Percolation Theory for Mathematicians. Birkhäuser, Basel (1982) Zbl 0522.60097 MR 0692943

[12] Lawler, G. F.: Intersections of Random Walks. Birkhäuser, Basel (1991) Zbl 0925.60078 MR 1117680

[13] Lindvall, T.: Lectures on the Coupling Method. Dover Publ., New York (1992) Zbl 0850.60019 MR 1180522

[14] Montroll, E. W.: Random walks in multidimensional spaces, especially on periodic lattices. J. Soc. Industr. Appl. Math. **4**, 241–260 (1956) Zbl 0080.35101 MR 0088110

[15] Sznitman, A. S.: How universal are asymptotics of disconnection times in discrete cylinders? Ann. Probab., in press

[16] Sznitman, A. S.: Brownian Motion, Obstacles and Random Media. Springer, Berlin (1998) Zbl 0973.60003 MR 1648164