Probab. Theory Relat. Fields
DOI 10.1007/s00440-016-0722-4

Boundary of the range of transient random walk

Amine Asselah[1] · **Bruno Schapira**[2]

Received: 10 July 2015 / Revised: 29 April 2016

Abstract We study the boundary of the range of simple random walk on $\mathbb{Z}^d$ in the transient case $d \geq 3$. We show that volumes of the range and its boundary differ mainly by a martingale. As a consequence, we obtain an upper bound on the variance of order $n \log n$ in dimension three. We also establish a Central Limit Theorem in dimension four and larger.

Mathematics Subject Classification 60F05 · 60G50

1 Introduction

Let $(S_n,\ n \geq 0)$ be a simple random walk on $\mathbb{Z}^d$. Its range $\mathcal{R}_n = \{S_0, \ldots, S_n\}$ is a familiar object of Probability Theory since Dvoretzky and Erdös' influential paper [6]. The object of interest in this paper is the boundary of the range

$$\partial\mathcal{R}_n = \{x \in \mathcal{R}_n :\ \text{there exists } y \sim x \text{ with } y \notin \mathcal{R}_n\}, \tag{1.1}$$

where $x \sim y$ means that x and y are at (graph) distance one. Our interest was triggered by a recent paper of Berestycki and Yadin [3] which proposes a model of hydrophobic polymer in an aqueous solvent, consisting of tilting the law of a simple random walk by $\exp(-\beta|\partial\mathcal{R}_n|)$. One interprets the range as the space occupied by the polymer,

✉ Bruno Schapira
bruno.schapira@univ-amu.fr

Amine Asselah
amine.asselah@u-pec.fr

1 LAMA, UPEC & IMéRA, Créteil, France

2 Aix-Marseille Université, CNRS, Centrale Marseille, I2M, UMR 7373, 13453 Marseille, France

Published online: 10 June 2016

and its complement as the space occupied by the solvent. Hydrophobic means that the monomers dislike the solvent, and the polymer tries to minimize the *boundary of the range*. The Gibbs' weight tends to minimize contacts between the monomers and the solvent, and the steric effect has been forgotten to make the model mathematically tractable. Besides its physical appeal, the model gives a central role to the boundary of the range, an object which remained mainly in the shadow until recently. To our knowledge it first appeared in the study of the entropy of the range of a simple random walk [2], with the conclusion that in dimension two or larger, the entropy of the range scales like the size of the boundary of the range. Recently, Okada [13] has established a law of large numbers for the boundary of the range for a transient random walk, and has obtained bounds on its expectation in dimension two.

Theorem 1.1 (Okada) *Consider a simple random walk in dimension $d = 2$. Then*

$$\frac{\pi^2}{2} \leq \lim_{n\to\infty} \frac{\mathbb{E}[|\partial\mathcal{R}_n|]}{n/\log^2(n)} \leq 2\pi^2, \tag{1.2}$$

where part of the result is that the limit exists. Moreover, when $d \geq 3$, almost surely

$$\lim_{n\to\infty} \frac{|\partial\mathcal{R}_n|}{n} = \mathbb{P}(\{z : z \sim 0\} \not\subset \mathcal{R}_\infty \cup \widetilde{\mathcal{R}}_\infty,\ H_0 = \infty), \tag{1.3}$$

where $\mathcal{R}_\infty$ is the range of a random walk in an infinite time horizon, and H_0 is the hitting time of 0, whereas quantities with tilde correspond to those of an independent copy of the random walk.

The range of a random walk has nice properties: (i) it is an increasing function of time, (ii) the event that S_k belongs to $\mathcal{R}_n$ for $k \leq n$ is $\sigma(S_0, \ldots, S_k)$-measurable, (iii) the volume of the range $\mathcal{R}_n$ is the union of the collection of sub-ranges $\{S_k, k \in I\}$ as I runs over a partition of $[0, n]$. A little thought shows that the boundary of the range shares none of these properties, making its study more difficult. The thrust of our study is to show that for a transient random walk, range and boundary of the range are nonetheless correlated objects. Indeed, we present two ways to appreciate their similar nature. On one hand the sizes of the boundary of the range and some range-like sets defined below (the $\mathcal{R}_{n,V}$) differ mainly by a martingale. On the other hand, we show that the boundary of the range, as the range itself, can be analyzed through a dyadic decomposition of the path. To make the first statement precise, we need more notation. Let $V_0 = \{z : z \sim 0\}$, be the neighbors of the origin, and for any nonempty subset V of V_0, let $\mathcal{R}_{n,V}$ be the set of sites of $\mathbb{Z}^d$ whose first visit occurs at some time $k \leq n$, and such that $(S_k + V_0) \cap \mathcal{R}_{k-1}^c = S_k + V$. In particular, $\mathcal{R}_{n,V}$ behaves like the range in the sense that properties (i)–(ii) listed above do hold, and as we will see below, their variance can be bounded using the same kind of techniques as for the range.

Note also that $\mathcal{R}_n$ is the disjoint union of the $\mathcal{R}_{n,V}$, with V subset of V_0. We are now ready for our first observation.

Proposition 1.2 *There is a martingale* $(M_n,\ n \in \mathbb{N})$*, adapted to the standard filtration such that for any positive integer* n*,*

$$|\partial \mathcal{R}_n| = \sum_{V \subset V_0} \rho_V \, |\mathcal{R}_{n-1,V}| + M_n + \mathcal{E}_n, \tag{1.4}$$

with $\rho_\emptyset = 0$ *and for any non-empty* V *in* V_0

$$\rho_V = \mathbb{P}(V \not\subset \mathcal{R}_\infty) \quad \textit{and} \quad \mathbb{E}\left(\mathcal{E}_n^2\right) = \begin{cases} \mathcal{O}(n) & \textit{if } d = 3 \\ \mathcal{O}(\log^3(n)) & \textit{if } d = 4 \\ \mathcal{O}(1) & \textit{if } d \geq 5. \end{cases} \tag{1.5}$$

Remark 1.3 The decomposition (1.4) is simply Doob's decomposition of the adapted process $|\partial \mathcal{R}_n| - \mathcal{E}_n$, as we see more precisely in Sect. 3. The key observation however is that the increasing process (in Doob's decomposition) behaves like the range.

Jain and Pruitt [9] have established a Central Limit Theorem for the range in dimension three with a variance scaling like $n \log n$. Proposition 1.2 makes us expect that the boundary of the range has a similar behavior. Indeed, we establish the following estimate on the mean square of the martingale. This estimate is delicate, uses precise Green's function asymptotics, and the symmetry of the walk. It is our main technical contribution.

Proposition 1.4 *There are positive constants* $\{C_d, d \geq 3\}$*, such that*

$$Var\,(M_n) \leq \begin{cases} C_3\, n \log n & \textit{if } d = 3 \\ C_d\, n & \textit{if } d \geq 4. \end{cases}$$

Also, following the approach of Jain and Pruitt [9], we establish the following estimate on the range-like object $\mathcal{R}_{n,V}$.

Proposition 1.5 *Assume that* $d = 3$*, and let* V *be a nonempty subset of* V_0*. There is a positive constant* C*, such that*

$$Var\,(|\mathcal{R}_{n,V}|) \leq C\, n \log n. \tag{1.6}$$

Then, a useful corollary of Propositions 1.2, 1.4 and 1.5 is the corresponding bound for the variance of the boundary of the range in dimension 3.

Theorem 1.6 *Assume that* $d = 3$*. Then, there is a positive constant* C*, such that*

$$Var\,(|\partial \mathcal{R}_n|) \leq C\, n \log n. \tag{1.7}$$

Remark 1.7 Using the approach of Jain and Pruitt [9], it is not clear how to obtain a Central Limit Theorem for $\mathcal{R}_{n,V}$ (see Remark 5.5 of the Appendix).

Now the boundary of the range has a decomposition similar to the classical Le Gall's decomposition [11] in terms of intersection of independent ranges. This decomposition, though simple, requires more notation to be presented. For integers n, m let $\mathcal{R}(n, n+m) = \{S_k - S_n\}_{n \le k \le n+m}$, with the shorthand notation $\mathcal{R}_n = \mathcal{R}(0, n)$, and note that

$$\mathcal{R}(0, n+m) = \mathcal{R}(0, n) \cup (S_n + \mathcal{R}(n, n+m)).$$

Observe that $\overleftarrow{\mathcal{R}}(0, n) := -S_n + \mathcal{R}(0, n)$ and $\mathcal{R}(n, n+m)$ are independent and that by the symmetry of the walk $\overleftarrow{\mathcal{R}}(0, n)$ (resp. $\mathcal{R}(n, n+m)$) has the same law as $\mathcal{R}(0, n)$ (resp. $\mathcal{R}(0, m)$): it corresponds to the range of a walk seen backward from position S_n. Finally, note the well known decomposition

$$|\mathcal{R}(0, n+m)| = |\mathcal{R}(0, n)| + |\mathcal{R}(n, n+m)| - |\overleftarrow{\mathcal{R}}(0, n) \cap \mathcal{R}(n, n+m)|. \quad (1.8)$$

Equality (1.8) is the basis of Le Gall's celebrated paper [11] on the range of recurrent random walk. It is also a key ingredient in most work on self-intersection of random walks (see the book of Chen [4], for many references).

To write a relation as useful as (1.8) for the boundary of the range, we introduce more notation. For $\Lambda \subset \mathbb{Z}^d$, we denote $\Lambda^+ = \Lambda + \overline{V}_0$, with $\overline{V}_0 = V_0 \cup \{0\}$, and we define its boundary as

$$\partial\Lambda = \{z \in \Lambda \ : \ \exists y \in \Lambda^c \text{ with } y \sim z\}.$$

Now, our simple observation is as follows.

Proposition 1.8 *For any integers n, m*

$$0 \ge |\partial\mathcal{R}(0, n+m)| - (|\partial\mathcal{R}(0, n)| + |\partial\mathcal{R}(n, n+m)|) \ge -Z(n, m), \quad (1.9)$$

with

$$Z(n, m) = |\overleftarrow{\mathcal{R}}(0, n) \cap \mathcal{R}^+(n, n+m)| + |\overleftarrow{\mathcal{R}}^+(0, n) \cap \mathcal{R}(n, n+m)|. \quad (1.10)$$

We focus now on consequences of this simple decomposition. For $d \ge 3$, we define functions $n \mapsto \psi_d(n)$, with the following dimension depending growth

$$\psi_3(n) = \sqrt{n}, \qquad \psi_4(n) = \log n, \qquad \text{and for } d > 4, \qquad \psi_d(n) = 1. \quad (1.11)$$

An essential step for a Central Limit Theorem, is to establish a linear lower bound on the variance. Our bounds hold in dimension three and larger.

Proposition 1.9 *Assume that $d \ge 3$. There are positive constants $\{c_d, \ d \ge 3\}$, such that*

$$Var\,(|\partial\mathcal{R}_n|) \ \ge c_d\, n. \quad (1.12)$$

The idea behind the linear lower bound (1.12) is to show that there is a *clock process* whose fluctuations are normal (on a scale square root of the time elapsed), and which is independent of the boundary of the range process. Thus, typical fluctuations of the clock process, provoke a time change *at constant* boundary of the range. Note that in dimension 3, this technique does not allow to obtain a lower bound of order $n \log n$, matching our upper bound (see also Remark 5.5 for some additional comment on this).

We now formulate our main Theorem.

Theorem 1.10 *When dimension is larger than or equal to three, there are constants* $\{C_d, d \geq 3\}$*, such that for any positive integer* n

$$\frac{C_d \psi_d(n)}{n} \geq \frac{\mathbb{E}[|\partial \mathcal{R}_n|]}{n} - \lim_{k\to\infty} \frac{\mathbb{E}[|\partial \mathcal{R}_k|]}{k} \geq 0. \tag{1.13}$$

Assume now that the dimension is four or larger. Then, the limit of $Var(|\partial \mathcal{R}_n|)/n$ *exists, is positive, and for all* $n \geq 1$*,*

$$\left| \frac{Var(|\partial \mathcal{R}_n|)}{n} - \lim_{k\to\infty} \frac{Var(|\partial \mathcal{R}_k|)}{k} \right| \leq \frac{C_d \sqrt{n} \psi_d(n)}{n}. \tag{1.14}$$

Moreover, a standard Central Limit Theorem holds for $|\partial \mathcal{R}_n|$*.*

Remark 1.11 We have stated our results for the simple random walk, but they hold, with similar proofs, for walks with symmetric and finitely supported increments.

Okada obtains also in [13] a large deviation principle for the upper tail (the probability that the boundary be larger than its mean), and in [14] he studies the most frequently visited sites of the boundary, and proves results analogous to what is known for the range.

In a companion paper [1], we obtain large deviations for the lower tail, and provide applications to phase transition for a properly normalized Berestycki-Yadin's polymer model.

The rest of the paper is organized as follows. In Sect. 2, we fix notation, recall known results on the Green's function, and prove a result about covering a finite subset. In Sect. 3, we establish the Martingale decomposition of Proposition 1.2 and prove Proposition 1.4. We prove Proposition 1.9 in Sect. 4. In Sect. 5, we present the dyadic decomposition for the boundary of the range and deduce Theorem 1.10, using Le Gall's argument. Finally in the Appendix, we prove Proposition 1.5.

2 Notation and prerequisites

For any $y, z \in \mathbb{Z}^d$, we denote by $\|z - y\|$ the Euclidean norm between y and z, and by $\langle y, z \rangle$ the corresponding scalar product. Then for any $r > 0$ we denote by $B(z, r)$ the ball of radius r centered at z:

$$B(z, r) := \{y \in \mathbb{Z}^d \ : \ \|z - y\| \leq r\}.$$

For $x \in \mathbb{Z}^d$, we let $\mathbb{P}_x$ be the law of the random walk starting from x, and denote its standard filtration by $(\mathcal{F}_k, k \geq 0)$. For Λ a subset of $\mathbb{Z}^d$ we define the hitting time of Λ as

$$H_\Lambda := \inf\{n \geq 1 \ : \ S_n \in \Lambda\},$$

that we abbreviate in H_x when Λ is reduced to a single point x. Note that in this definition we use the convention to consider only times larger than or equal to one. At some point it will also be convenient to consider a shifted version, so we also define for $k \geq 0$,

$$H_\Lambda^{(k)} := \inf\{n \geq k \ : \ S_n \in \Lambda\}. \tag{2.1}$$

We will need bounds on the heat kernel, so let us recall a standard result:

$$\mathbb{P}(S_n = z) \leq C \, \frac{1}{n^{d/2}} \, \exp(-c\|z\|^2/n) \quad \text{for all } z \text{ and } n \geq 1, \tag{2.2}$$

for some positive constants c and C (see for instance [7]). Now we recall also the definition and some basic properties of Green's function. For $u, v \in \mathbb{Z}^d$, the Green's function is

$$G(u, v) = \mathbb{E}_u \left[\sum_{n \geq 0} \mathbb{1}\{S_n = v\} \right] = \mathbb{P}_u[H_v < \infty] \cdot G(0, 0),$$

and we use extensively the well-known bound (see [10, Theorem 4.3.1]):

$$G(0, z) = \mathcal{O}\left(\frac{1}{1 + \|z\|^{d-2}} \right). \tag{2.3}$$

We also consider Green's function restricted to a set $A \subset \mathbb{Z}^d$, which for $u, v \in A$ is defined by

$$G_A(u, v) = \mathbb{E}_u \left[\sum_{n=0}^{H_{A^c}-1} \mathbb{1}\{S_n = v\} \right].$$

We recall that G_A is symmetric (see [10, Lemma 4.6.1]):

$$G_A(u, v) = G_A(v, u) \quad \text{for all } u, v \in A,$$

and that G is also invariant by translation of the coordinates: $G(u, v) = G(0, v - u)$. Also, for $n \geq 0$,

$$G_n(u, v) = \mathbb{E}_u \left[\sum_{k=0}^{n} \mathbb{1}\{S_n = v\} \right]. \tag{2.4}$$

It is well known (use (2.3) and Theorem 3.6 of [11]) that for ψ_d defined in (1.11), we have, for some positive constants $\{C_d, d \geq 3\}$

$$\sum_{z \in \mathbb{Z}^d} G_n^2(0, z) \leq C_d \, \psi_d(n). \tag{2.5}$$

We can now state the main result of this section.

Lemma 2.1 *Let Λ be a fixed finite subset of $\mathbb{Z}^d$, and fix $z \in \Lambda$. Then, there is a constant $c(\Lambda)$, such that for any two neighboring sites $y \sim y'$,*

$$\mathbb{P}_y(\Lambda \subset \mathcal{R}_\infty) - \mathbb{P}_{y'}(\Lambda \subset \mathcal{R}_\infty) = c(\Lambda) \frac{\langle y' - y, y - z \rangle}{\|y - z\|^d} + \mathcal{O}\left(\frac{1}{\|y - z\|^d}\right). \tag{2.6}$$

Moreover,

$$c(\Lambda) = \frac{1}{d v_d} \sum_{x \in \Lambda} \sum_{v \notin \Lambda} \mathbb{1}_{\{v \sim x\}} \mathbb{P}_v(H_\Lambda = \infty) \, \mathbb{P}_x(\Lambda \subset \mathcal{R}_\infty),$$

where v_d denote the volume of the unit ball in $\mathbb{R}^d$.

Proof First, since Λ is finite, and (2.6) is an asymptotic result, we can always assume that y and y' do not belong to Λ. Now by a first entry decomposition

$$\mathbb{P}_y(\Lambda \subset \mathcal{R}_\infty) = \sum_{x \in \Lambda} \mathbb{P}_y(S_{H_\Lambda} = x, \ H_\Lambda < \infty) \mathbb{P}_x(\Lambda \subset \mathcal{R}_\infty). \tag{2.7}$$

Next, fix $x \in \Lambda$ and transform the harmonic measure into the restricted Green's function (see for instance [10, Lemma 6.3.6]):

$$\mathbb{P}_y(S_{H_\Lambda} = x, \ H_\Lambda < \infty) = \frac{1}{2d} \sum_{v \in \Lambda^c, \, v \sim x} G_{\Lambda^c}(y, v) = \frac{1}{2d} \sum_{v \in \Lambda^c, \, v \sim x} G_{\Lambda^c}(v, y).$$

Note also (see [10, Proposition 4.6.2]) that

$$G_{\Lambda^c}(v, y) = G(v, y) - \mathbb{E}_v[\mathbb{1}\{H_\Lambda < \infty\} \, G(S_{H_\Lambda}, y)].$$

Therefore,

$$\begin{aligned} &\mathbb{P}_y(S_{H_\Lambda} = x, H_\Lambda < \infty) - \mathbb{P}_{y'}(S_{H_\Lambda} = x, H_\Lambda < \infty) \\ &= \frac{1}{2d} \sum_{v \in \Lambda^c, \, v \sim x} (G(v, y) - G(v, y')) \\ &\quad - \frac{1}{2d} \sum_{v \in \Lambda^c, \, v \sim x} \mathbb{E}_v[\mathbb{1}\{H_\Lambda < \infty\}(G(S_{H_\Lambda}, y) - G(S_{H_\Lambda}, y'))]. \end{aligned} \tag{2.8}$$

Now, since Λ is finite, we have [10, Corollary 4.3.3] the expansion for any $z' \in \Lambda^+$ (recall that z is a given site in Λ),

$$G(z', y) - G(z', y') = \frac{2}{v_d} \frac{\langle y' - y, y - z \rangle}{\|y - z\|^d} + \mathcal{O}\left(\frac{1}{\|y - z\|^d}\right). \tag{2.9}$$

Combining (2.7), (2.8) and (2.9) we obtain the result (2.6). □

3 Martingale decomposition

In this section, we establish Proposition 1.2, as well as Proposition 1.4 dealing with the variance of the martingale.

3.1 Definition of the martingale and proof of Proposition 1.2

For V nonempty subset of V_0 and $k \geq 0$, let

$$I_{k,V} = \mathbb{1}\{S_k \notin \mathcal{R}_{k-1} \text{ and } (S_k + V_0) \cap \mathcal{R}_k^c = S_k + V\},$$

and

$$J_{k,V} = \mathbb{1}\{(S_k + V) \nsubseteq \{S_j,\ j \geq k\}\}.$$

Then for $n \geq 1$, define

$$J_{k,n,V} = \mathbb{1}\{(S_k + V) \nsubseteq \{S_k, \ldots, S_n\}\},$$

and

$$\partial\mathcal{R}_{n,V} = \{S_k \,:\, I_{k,V} J_{k,n,V} = 1,\ k \leq n\}. \tag{3.1}$$

Note that $\partial\mathcal{R}_n$ is the disjoint union of the $\partial\mathcal{R}_{n,V}$, for V non empty subset of V_0. Now instead of looking at $\sum_{k \leq n} I_{k,V} J_{k,n,V}$ (which is equal to $|\partial\mathcal{R}_{n,V}|$), we look at

$$Y_{n,V} = \sum_{k=0}^{n-1} I_{k,V} J_{k,V}.$$

However, since $Y_{n,V}$ is not adapted to $\mathcal{F}_n$, we consider

$$X_{n,V} = \mathbb{E}[Y_{n,V} \mid \mathcal{F}_n],$$

and think of $X_{n,V}$ as a good approximation for $|\partial\mathcal{R}_{n,V}|$. So we define an error term as

$$\mathcal{E}_{n,V} := |\partial\mathcal{R}_{n,V}| - X_{n,V}.$$

Now the Doob decomposition of the adapted process $X_{n,V}$ reads as $X_{n,V} = M_{n,V} + A_{n,V}$, with $M_{n,V}$ a martingale and $A_{n,V}$ a predictable process. Since

$$X_{n,V} = \sum_{k=0}^{n-1} I_{k,V}\mathbb{E}[J_{k,V} \mid \mathcal{F}_n],$$

we have

$$A_{n,V} = \sum_{k=0}^{n-1} I_{k,V}\mathbb{E}[J_{k,V} \mid \mathcal{F}_k].$$

Moreover, the Markov property also gives

$$\mathbb{E}[J_{k,V} \mid \mathcal{F}_k] = \mathbb{E}[J_{k,V}] = \mathbb{P}(V \nsubseteq \mathcal{R}_\infty) = \rho_V,$$

for any $k \geq 0$. Therefore,

$$|\partial\mathcal{R}_{n,V}| = M_{n,V} + \rho_V|\mathcal{R}_{n-1,V}| + \mathcal{E}_{n,V} \quad \text{for all } V \subset V_0,$$

where we defined for $m \geq 0$,

$$\mathcal{R}_{m,V} = \{S_k \,:\, I_{k,V} = 1,\ k \leq m\}.$$

Summing up $M_{n,V}$ over nonempty subsets of V_0 we obtain another martingale $M_n = \sum_{V\subset V_0} M_{n,V}$, and the error term $\mathcal{E}_n = \sum_{V\subset V_0} \mathcal{E}_{n,V}$, and since $|\partial\mathcal{R}_n|$ is also the sum over nonempty subsets of the $|\partial\mathcal{R}_{n,V}|$, we obtain the first part of Proposition 1.2, namely Eq. (1.4).

Now we prove (1.5). First note that for any $k \leq n-1$,

$$\left|J_{k,n,V} - \mathbb{E}[J_{k,V} \mid \mathcal{F}_n]\right| \leq \mathbb{P}_{S_n}(H_{S_k+V_0} < \infty) = \mathcal{O}\left(\frac{1}{1+\|S_n - S_k\|^{d-2}}\right),$$

using (2.3) for the last equality. Then by using invariance of the walk by time inversion, we get

$$\mathbb{E}[\mathcal{E}_{n,V}^2] = \mathcal{O}\left(\sum_{k,k'\leq n} \mathbb{E}\left[\frac{1}{(1+\|S_k\|^{d-2})(1+\|S_{k'}\|^{d-2})}\right]\right). \tag{3.2}$$

Moreover, by using the heat kernel bound (2.2), we arrive at

$$\mathbb{E}\left[\frac{1}{(1+\|S_k\|^{d-2})^2}\right] = \begin{cases} \mathcal{O}(1/k) & \text{if } d = 3 \\ \mathcal{O}((\log k)/k^2) & \text{if } d = 4 \\ \mathcal{O}(k^{-d/2}) & \text{if } d \geq 5. \end{cases} \tag{3.3}$$

The desired result follows by using Cauchy–Schwarz.

3.2 Variance of the martingale

We establish here Proposition 1.4. Let us notice that our proof works for M_n only, and not for all the $M_{n,V}$'s. If we set for $n \geq 0$,

$$\Delta M_n = M_{n+1} - M_n,$$

then, Proposition 1.4 is a direct consequence of the following result.

$$\mathbb{E}[(\Delta M_n)^2] = \begin{cases} \mathcal{O}(\log n) & \text{if } d = 3 \\ \mathcal{O}(1) & \text{if } d \geq 4. \end{cases} \tag{3.4}$$

The proof of (3.4) is divided in three steps. The first step brings us to a decomposition of ΔM_n as a finite combination of simpler terms (3.6), plus a rest whose L^2-norm we show is negligible. In the second step, we observe that when we gather together some terms (3.10), their L^2-norm takes a particularly nice form (3.11). Finally in the third step we use these formula and work on it to get the right bound.

Step 1 In this step, we just use the Markov property to write ΔM_n in a nicer way, up to some error term, which is bounded by a deterministic constant. Before that, we introduce some more notation. For $k \leq n$, set

$$I_{k,n,V} = \mathbb{1}\{(S_k + V_0) \cap \{S_k, \dots, S_n\}^c = S_k + V\}.$$

The Markov property and the translation invariance of the walk show that for all $k \leq n$

$$\mathbb{E}[J_{k,V} \mid \mathcal{F}_n] = \sum_{V' \subset V_0} \mathbb{1}_{\{V \cap V' \neq \emptyset\}} I_{k,n,V'} \, \mathbb{P}_{S_n - S_k}(V \cap V' \nsubseteq \mathcal{R}_\infty).$$

Note that $I_{k,n,V'} \neq I_{k,n+1,V'}$ implies that S_{n+1} and S_k are neighbors. However, the number of indices k such that S_{n+1} and S_k are neighbors and $I_{k,V} = 1$ is at most $2d$, since by definition of $I_{k,V}$ we only count the first visits to neighbors of S_{n+1}. Therefore the number of indices k satisfying $I_{k,V} \neq 0$ and $I_{k,n,V'} \neq I_{k,n+1,V'}$, for some V', is bounded by $2d$. As a consequence, by using also that terms in the sum defining $M_{n,V}$ are bounded in absolute value by 1, we get

$$\begin{aligned}
\Delta M_{n,V} &= \sum_{k=0}^{n-1} I_{k,V} (\mathbb{E}[J_{k,V} \mid \mathcal{F}_{n+1}] - \mathbb{E}[J_{k,V} \mid \mathcal{F}_n]) + I_{n,V} \mathbb{E}[J_{n,V} \mid \mathcal{F}_{n+1}] \\
&= \sum_{k=0}^{n-1} \sum_{V' \subset V_0} \mathbb{1}_{\{V \cap V' \neq \emptyset\}} I_{k,V} \{ I_{k,n+1,V'} \mathbb{P}_{S_{n+1} - S_k}(V \cap V' \nsubseteq \mathcal{R}_\infty) \\
&\quad - I_{k,n,V'} \mathbb{P}_{S_n - S_k}(V \cap V' \nsubseteq \mathcal{R}_\infty) \} + I_{n,V} \mathbb{E}[J_{n,V} \mid \mathcal{F}_{n+1}] \\
&= \sum_{k=0}^{n-1} \sum_{V \cap V' \neq \emptyset} I_{k,V} I_{k,n,V'} \{ \mathbb{P}_{S_{n+1} - S_k}(V \cap V' \nsubseteq \mathcal{R}_\infty) \\
&\quad - \mathbb{P}_{S_n - S_k}(V \cap V' \nsubseteq \mathcal{R}_\infty) \} + r_{n,V},
\end{aligned}$$

with $|r_{n,V}| \le 2d+1$. Summing up over V, we get

$$\Delta M_n = \sum_{k=0}^{n-1} \sum_{V \cap V' \neq \emptyset} I_{k,V} I_{k,n,V'} \{ \mathbb{P}_{S_{n+1}-S_k}(V \cap V' \nsubseteq \mathcal{R}_\infty) - \mathbb{P}_{S_n - S_k}(V \cap V' \nsubseteq \mathcal{R}_\infty) \} + r_n, \tag{3.5}$$

with $|r_n| \le 2^d(2d+1)$. Lemma 2.1 is designed to deal with the right hand side of (3.5), with the result that

$$\Delta M_n = \sum_{k=0}^{n-1} \sum_{V \cap V' \neq \emptyset} c(V \cap V') I_{k,V} I_{k,n,V'} \frac{\langle S_{n+1} - S_n, S_n - S_k \rangle}{1 + \|S_n - S_k\|^d} + \mathcal{O}(B_n), \tag{3.6}$$

with

$$B_n = \sum_{z \in \partial \mathcal{R}_n} \frac{1}{1 + \|S_n - z\|^d}. \tag{3.7}$$

Step 2 The term B_n of (3.7) can be bounded as follows. By using first the invariance of the law of the walk by time inversion, we can replace the term $S_n - z$ by z. Then we write

$$\begin{aligned} \mathbb{E}[B_n^2] &= \mathbb{E}\left[\left(\sum_{z \in \partial \mathcal{R}_n} \frac{1}{1 + \|z\|^d} \right)^2 \right] \\ &= \sum_{z, z' \in \mathbb{Z}^d} \frac{1}{(1 + \|z\|^d)(1 + \|z'\|^d)} \mathbb{P}(z \in \partial \mathcal{R}_n,\ z' \in \partial \mathcal{R}_n). \end{aligned} \tag{3.8}$$

Then by assuming for instance that $\|z\| \le \|z'\|$ (and $z \neq z'$), and by using (2.3) we obtain

$$\begin{aligned} \mathbb{P}(z \in \partial \mathcal{R}_n,\ z' \in \partial \mathcal{R}_n) &\le \mathbb{P}(H_z < \infty,\ H_{z'} < \infty) \\ &\le 2G(0,z)G(z,z') = \mathcal{O}\left(\frac{1}{1 + \|z\|^{d-2} \|z' - z\|^{d-2}} \right). \end{aligned} \tag{3.9}$$

Therefore

$$\mathbb{E}[B_n^2] = \mathcal{O}\left(\sum_{1 \le \|z\| < \|z'\|} \frac{1}{\|z\|^{2d-2} \|z'\|^d \|z' - z\|^{d-2}} \right).$$

Next, we divide the last sum into two parts:

$$\begin{aligned}
&\sum_{1\le\|z\|<\|z'\|} \frac{1}{\|z\|^{2d-2}\|z'\|^d\|z'-z\|^{d-2}} \\
&= \sum_{1\le\|z\|<\|z'\|\le 2\|z\|} \frac{1}{\|z\|^{2d-2}\|z'\|^d\|z'-z\|^{d-2}} + \sum_{1\le 2\|z\|<\|z'\|} \frac{1}{\|z\|^{2d-2}\|z'\|^d\|z'-z\|^{d-2}} \\
&= \mathcal{O}\left(\sum_{1\le\|z\|<\|z'\|\le 2\|z\|} \frac{1}{\|z\|^{3d-2}\|z'-z\|^{d-2}} + \sum_{1\le 2\|z\|<\|z'\|} \frac{1}{\|z\|^{2d-2}\|z'\|^{2d-2}} \right) = \mathcal{O}(1).
\end{aligned}$$

Now it remains to bound the main term in (3.6). For two nonempty subsets U and U' of V_0, write $U \sim U'$, if there exists an isometry of $\mathbb{Z}^d$ sending U onto U'. This of course defines an equivalence relation on the subsets of V_0, and for any representative U of an equivalence class, we define

$$\widetilde{I}_{k,n,U} = \sum_{V\cap V'\sim U} I_{k,V}\, I_{k,n,V'} = \mathbb{1}\{V_0 \cap (\mathcal{R}_n^c - S_k) \sim U\}$$

and

$$H_{n,U} = \sum_{k=0}^{n-1} \widetilde{I}_{k,n,U}\, \frac{\langle S_{n+1}-S_n, S_n - S_k\rangle}{1+\|S_n-S_k\|^d}.$$

Note that since the function $c(\cdot)$ is invariant under isometry, we can rewrite the main term in (3.6) as

$$\sum_U c(U)\, H_{n,U}. \tag{3.10}$$

Then observe that for any U,

$$\mathbb{E}[H_{n,U}^2 \mid \mathcal{F}_n] = \left\| \sum_{k=0}^{n-1} \widetilde{I}_{k,n,U}\, \frac{S_n - S_k}{1+\|S_n-S_k\|^d} \right\|^2.$$

Moreover, since the law of the walk is invariant under time inversion, and since for any path $S_0, \dots, S_n$, and any k, the indicator $\widetilde{I}_{k,n,U}$ is equal to 1 if and only if it is also equal to 1 for the reversed path $S_n, \dots, S_0$, we get

$$\mathbb{E}[H_{n,U}^2] = \mathbb{E}[\|\mathcal{H}_{n,U}\|^2], \tag{3.11}$$

where

$$\mathcal{H}_{n,U} := \sum_{z\in\partial\widetilde{\mathcal{R}}_{n,U}} \frac{z}{1+\|z\|^d}, \quad \text{with} \quad \partial\widetilde{\mathcal{R}}_{n,U} := \{S_k \,:\, \widetilde{I}_{k,n,U} = 1,\ k \le n-1\}. \tag{3.12}$$

Therefore, we only need to prove that for any U,

$$\mathbb{E}[\|\mathcal{H}_{n,U}\|^2] = \begin{cases} \mathcal{O}(\log n) & \text{if } d = 3 \\ \mathcal{O}(1) & \text{if } d \geq 4. \end{cases} \tag{3.13}$$

Step 3 First note that

$$\mathbb{E}[\|\mathcal{H}_{n,U}\|^2] = \sum_{z,z' \in \mathbb{Z}^d} \frac{\langle z, z' \rangle}{(1 + \|z\|^d)(1 + \|z'\|^d)} \, \mathbb{P}(z \in \partial \widetilde{\mathcal{R}}_{n,U},\ z' \in \partial \widetilde{\mathcal{R}}_{n,U}). \tag{3.14}$$

In dimension 4 or larger, (3.13) can be established as follows. First Cauchy–Schwarz inequality gives for all z, z'

$$|\langle z, z' \rangle| \ \leq \ \|z\| \, \|z'\|,$$

Then, by using again the standard bound on Green's functions, that is (3.9), we get the desired bound

$$\mathbb{E}[\|\mathcal{H}_{n,U}\|^2] = \mathcal{O}\left(\sum_{1 \leq \|z\| < \|z'\|} \|z\|^{3-2d} \, \|z'\|^{1-d} \, \|z - z'\|^{2-d} \right) = \mathcal{O}(1).$$

We consider now the case $d = 3$. Since it might be interesting to see what changes in dimension 3, we keep the notation d in all formula as long as possible. Note that if $z \in \partial \widetilde{\mathcal{R}}_{n,U}$, then $\|z\| \leq n$ and H_z is finite. Therefore the restriction of the sum in (3.14) to the set of z, z' satisfying $\|z\| \leq \|z'\| \leq 2\|z\|$ is bounded in absolute value by

$$\sum_z \sum_{z'} \mathbb{1}_{\{\|z\| \leq \|z'\| \leq 2\|z\| \leq 2n\}} \frac{2\|z\|^2}{(1 + \|z\|^d)^2} \, \mathbb{P}(H_z < \infty,\ H_{z'} < \infty). \tag{3.15}$$

Moreover, as we have already recalled, for any $z \neq z'$, with $\|z\| \leq \|z'\|$,

$$\mathbb{P}(H_z < \infty,\ H_{z'} < \infty) \leq 2G(0, z)G(z, z') = \mathcal{O}\left(\frac{1}{\|z\|^{d-2} \|z - z'\|^{d-2}} \right).$$

Therefore, the sum in (3.15) is bounded above (up to some constant) by

$$\begin{aligned} \sum_z \sum_{z'} \mathbb{1}_{\{1 \leq \|z\| < \|z'\| \leq 2\|z\| \leq 2n\}} \|z\|^{4-3d} \|z - z'\|^{2-d} &= \mathcal{O}\left(\sum_{1 \leq \|z\| \leq n} \|z\|^{6-3d} \right) \\ &= \mathcal{O}(\log n). \end{aligned}$$

It remains to bound the sum in (3.14) restricted to the z and z' satisfying $\|z'\| \geq 2\|z\|$. To this end observe that the price of visiting z' first is too high. Indeed,

$$\begin{aligned}
&\sum_z \sum_{z'} \mathbb{1}_{\{1\leq 2\|z\|\leq\|z'\|\leq n\}} \frac{\left|\langle z, z'\rangle\right|}{(1+\|z\|^d)(1+\|z'\|^d)} \mathbb{P}(H_{z'} < H_z < \infty) \\
&= \mathcal{O}\left(\sum_z \sum_{z'} \mathbb{1}_{\{1\leq 2\|z\|\leq\|z'\|\leq n\}} \|z\|^{1-d} \|z'\|^{3-2d} \|z-z'\|^{2-d}\right) \\
&= \mathcal{O}\left(\sum_z \sum_{z'} \mathbb{1}_{\{1\leq 2\|z\|\leq\|z'\|\leq n\}} \|z\|^{1-d} \|z'\|^{5-3d}\right) \\
&= \mathcal{O}(\log n),
\end{aligned}$$

where for the first equality we used in particular Cauchy–Schwarz inequality and again the standard bound on Green's function, and for the second one, we used that when $\|z'\| \geq 2\|z\|$, we have $\|z'\| \asymp \|z - z'\|$. Thus in (3.14) we consider the events $\{H_z < H_{z'}\}$. We now refine this argument in saying that after H_{z+V_0}, and after having left the ball $B(z, \|z\|/2)$, it cost too much to return to $z + V_0$ (and the same fact holds for z'). Formally, for any z, define

$$\tau_z := \inf\{k \geq H_{z+V_0} \ : \ S_k \notin B(z, \|z\|/2)\},$$

and

$$\sigma_z := \inf\{k \geq \tau_z \ : \ S_k \in z + V_0\}.$$

Then define the event

$$E_{z,n,U} := \{z \in \partial\widetilde{\mathcal{R}}_{n,U}\} \cap \{\sigma_z = \infty\}.$$

Observe next that if $1 \leq \|z\| \leq \|z'\|/2$,

$$\mathbb{P}(z \in \partial\widetilde{\mathcal{R}}_{n,U},\ z' \in \partial\widetilde{\mathcal{R}}_{n,U},\ (E_{z,n,U} \cap E_{z',n,U})^c) = \mathcal{O}\left(\frac{1}{\|z\|^{2d-4}\|z'\|^{d-2}}\right).$$

Therefore, similar computations as above, show that in (3.14), we can replace the event $\{z \text{ and } z' \in \partial\widetilde{\mathcal{R}}_{n,U}\}$ by $E_{z,n,U} \cap E_{z',n,U}$. So at this point it remains to bound the (absolute value of the) sum

$$\sum_{\|z'\|\geq 2\|z\|} \frac{\langle z, z'\rangle}{(1+\|z\|^d)(1+\|z'\|^d)} \mathbb{P}(H_z < H_{z'},\ E_{z,n,U},\ E_{z',n,U}). \tag{3.16}$$

We now eliminate the time n-dependence in $E_{z,n,U}$ and $E_{z',n,U}$ by replacing these events respectively by $E_{z,U}$ and $E_{z',U}$ defined as

$$E_{z,U} := \{H_z < \infty\} \cap \{V_0 \cap \{S_0 - z, \ldots, S_{\tau_z} - z\}^c \sim U\} \cap \{\sigma_z = \infty\}.$$

Note that when $\tau_z \le n$ we have $E_{z,U} = E_{z,n,U}$. This latter relation holds in particular when z is visited before z', and z' is visited before time n. Therefore one has

$$E_{z,n,U} \cap E_{z',n,U} \cap \{H_z < H_{z'}\} = E_{z,U} \cap E_{z',n,U} \cap \{H_z < H_{z'}\}.$$

In other words in (3.16) one can replace the event $E_{z,n,U}$ by $E_{z,U}$. We want now to do the same for z', but the argument is a bit more delicate. First define the symmetric difference of two sets A and B as $A \,\Delta\, B = (A \cap B^c) \cup (A^c \cap B)$. Recall that we assume $1 \le \|z\| \le \|z'\|/2$. Let now $k \le n$. By using (2.2) and (2.3), we get for some positive constants c and C (recall also the definition (2.1)),

$$\begin{aligned}
\mathbb{P}_z(E_{z',k,U} \,\Delta\, E_{z',U}) &\le \mathbb{P}_z\left(H_{z'} \le k \le H_{z'+V_0}^{(k+1)} < \infty\right) \\
&\le C\, \mathbb{E}_z\left[\mathbb{1}_{\{H_{z'} \le k\}} \frac{1}{1 + \|S_k - z'\|^{d-2}}\right] \\
&\le C \sum_{i=1}^{k} \mathbb{P}_z(S_i = z')\, \mathbb{E}\left(\frac{1}{1 + \|S_{k-i}\|^{d-2}}\right) \\
&\le C \sum_{i=1}^{k} \frac{e^{-c\|z'\|^2/i}}{i\sqrt{i}} \frac{1}{1 + \sqrt{k-i}},
\end{aligned}$$

where for the second and third lines we used the Strong Markov Property, and Cauchy–Schwarz and (3.3) for the fourth one. The last sum above can be bounded by first separating it into two sums, one with indices i smaller than $k/2$, and the other sum over indices $i \ge k/2$. Then using a comparison with an integral for the first sum, one can see that

$$\mathbb{P}_z(E_{z',k,U} \,\Delta\, E_{z',U}) \le C \frac{1}{\|z'\|\sqrt{k}}\, e^{-c\|z'\|^2/(2k)}.$$

In particular

$$\sup_{k \ge 1} \mathbb{P}_z(E_{z',k,U} \,\Delta\, E_{z',U}) = \mathcal{O}\left(\frac{1}{\|z'\|^2}\right).$$

Then it follows by using again the Markov property and (2.3), that

$$\begin{aligned}
\mathbb{P}(E_{z,U},\, E_{z',n,U} \,\Delta\, E_{z',U},\, H_z < H_{z'}) &\le \mathbb{P}(E_{z',n,U} \,\Delta\, E_{z',U},\, H_z < H_{z'}) \\
&= \sum_{k \le n} \mathbb{P}(H_z = n - k)\, \mathbb{P}_z(E_{z',k,U} \,\Delta\, E_{z',U}) \\
&\le \mathbb{P}(H_z \le n) \sup_{k \ge 1} \mathbb{P}_z(E_{z',k,U} \,\Delta\, E_{z',U}) \\
&= \mathcal{O}\left(\frac{1}{\|z\|^{d-2}} \times \frac{1}{\|z'\|^2}\right).
\end{aligned}$$

In conclusion, one can indeed replace the event $E_{z',n,U}$ by $E_{z',U}$ in (3.16). Now in the remaining sum, we gather together the pairs (z, z') and $(z, -z')$, and we get, using Cauchy–Schwarz again,

$$\begin{aligned}\Bigg|\sum_{1\le 2\|z\|\le\|z'\|\le n}\frac{\langle z,z'\rangle}{(1+\|z\|^d)(1+\|z'\|^d)}\,\mathbb{P}(E_{z,U},\,E_{z',U},\,H_z<H_{z'})\Bigg|\\ \le\sum_{1\le 2\|z\|\le\|z'\|\le n}\frac{2}{\|z\|^{d-1}\|z'\|^{d-1}}\Big|\mathbb{P}(E_{z,U},\,E_{z',U},\,H_z<H_{z'})\\ -\mathbb{P}(E_{z,U},\,E_{-z',U},\,H_z<H_{-z'})\Big|.\end{aligned}\tag{3.17}$$

Then for any $1\le\|z\|\le\|z'\|/2$, we have

$$\begin{aligned}&\mathbb{P}(E_{z,U},\,E_{z',U},\,H_z<H_{z'})\\ &=\sum_{y\in\overline{\partial}B(z,\|z\|/2)}\mathbb{P}(E^*_{z,U},\,H_z<H_{z'},\,S_{\tau_z}=y)\,\mathbb{P}_y(H_{z+V_0}=\infty,\,E_{z',U}),\end{aligned}\tag{3.18}$$

where by $\overline{\partial}B(z,\|z\|/2)$ we denote the external boundary of $B(z,\|z\|/2)$, and where

$$E^*_{z,U}:=\{V_0\cap\{S_0-z,\dots,S_{\tau_z}-z\}^c\sim U\}.$$

Now for any $y\in\overline{\partial}B(z,\|z\|/2)$, and $\|z'\|\ge 2\|z\|$, by using again (2.3) we get

$$\mathbb{P}_y(H_{z+V_0}=\infty,\,E_{z',U})=\mathbb{P}_y(E_{z',U})-\mathcal{O}\left(\frac{1}{\|z\|^{d-2}\|z'\|^{d-2}}\right).\tag{3.19}$$

Moreover, the same argument as in the proof of Lemma 2.1, shows that if y and y' are neighbors,

$$\mathbb{P}_y(E_{z',U})=\mathbb{P}_{y'}(E_{z',U})+\mathcal{O}\left(\frac{1}{\|z'-y\|^{d-1}}\right).$$

Therefore if $1\le\|z\|\le\|z'\|/2$ and $y\in\overline{\partial}B(z,\|z\|/2)$ we get

$$\mathbb{P}_y(E_{z',U})=\mathbb{P}_0(E_{z',U})+\mathcal{O}\left(\frac{\|y\|}{\|z'-y\|^{d-1}}\right)=\mathbb{P}_0(E_{z',U})+\mathcal{O}\left(\frac{\|z\|}{\|z'\|^{d-1}}\right).\tag{3.20}$$

On the other hand, by symmetry, for any U,

$$\mathbb{P}_0(E_{z',U})=\mathbb{P}_0(E_{-z',U}).$$

By combining this with (3.17), (3.18), (3.19) and (3.20), we obtain (3.13) and conclude the proof of (3.4).

4 Lower bound on the variance

In this section, we prove Proposition 1.9. The proof is inspired by the proof of Theorem 4.11 in [12], where the authors use lazyness of the walk. Here, since the walk we consider is not lazy, we use instead the notion of double backtracks. We say that the simple random walk makes a double backtrack at time n, when $S_{n+1} = S_{n-1}$ and $S_{n+2} = S_n$. When this happens the range (and its boundary) remain constant during steps $\{n+1, n+2\}$. With this observation in mind, a lower bound on the variance is obtained as we decompose the simple random walk into two independent processes: a clock process counting the number of double-backtracks (at even times), and a trajectory without double-backtrack (at even times).

4.1 Clock process

We construct by induction a no-double backtrack walk $(\widetilde{S}_n,\ n \in \mathbb{N})$. First, $\widetilde{S}_0 = 0$, and $\widetilde{S}_1$ and $\widetilde{S}_2 - \widetilde{S}_1$ are chosen uniformly at random among the elements of V_0 (the set of neighbors of the origin). Next, assume that $\widetilde{S}_k$ has been defined for all $k \le 2n$, for some $n \ge 1$. Let $\mathcal{N}_2 = \{(x, y) :\ x \sim 0 \text{ and } y \sim x\}$ and choose (X, Y) uniformly at random in $\mathcal{N}_2 \backslash \{(\widetilde{S}_{2n-1} - \widetilde{S}_{2n}, 0)\}$. Then set

$$\widetilde{S}_{2n+1} = \widetilde{S}_{2n} + X \quad \text{and} \quad \widetilde{S}_{2n+2} = \widetilde{S}_{2n} + Y. \tag{4.1}$$

Thus, the walk $\widetilde{S}$ makes no double-backtrack at even times. Note that by sampling uniformly in the whole of $\mathcal{N}_2$ we would have generated a simple random walk (SRW). Now, to build a SRW out of $\widetilde{S}$, it is enough to add at each even time a geometric number of double-backtracks. The geometric law is given by

$$\mathbb{P}(\xi = k) = (1-p)p^k \quad \text{for all } k \ge 0, \tag{4.2}$$

with $p = 1/(2d)^2$. Note that the mean of ξ is equal to $p/(1-p)$. Now, consider a sequence $(\xi_n,\ n \ge 1)$ of i.i.d. random variables distributed like ξ and independent of $\widetilde{S}$. Then define

$$\widetilde{N}_0 = \widetilde{N}_1 = 0 \quad \text{and} \quad \widetilde{N}_k := \sum_{i=1}^{[k/2]} \xi_i \quad \text{for } k \ge 2. \tag{4.3}$$

A SRW can be built out from $\widetilde{S}$ and $\widetilde{N}$ as follows. First, $S_i = \widetilde{S}_i$ for $i = 0, 1, 2$. Then, for any integer $k \ge 1$

$$S_{2i-1} = \widetilde{S}_{2k-1} \quad \text{and} \quad S_{2i} = \widetilde{S}_{2k} \quad \text{for all } i \in [k + \widetilde{N}_{2(k-1)}, k + \widetilde{N}_{2k}].$$

This implies that if $\widetilde{\mathcal{R}}$ is the range of $\widetilde{S}$ and $\partial\widetilde{\mathcal{R}}$ its boundary, then for any integer k

$$\mathcal{R}_{k+2\widetilde{N}_k} = \widetilde{\mathcal{R}}_k \qquad \text{and} \qquad \partial\mathcal{R}_{k+2\widetilde{N}_k} = \partial\widetilde{\mathcal{R}}_k. \tag{4.4}$$

4.2 A law of large numbers and some consequences

Recall that Okada [13] proved a law of large numbers for $|\partial\mathcal{R}_n|$, see (1.3), and call ν_d the limit of $|\partial\mathcal{R}_n|/n$. Since $\widetilde{N}_n/n$ also converges almost surely toward $p/[2(1-p)]$, we deduce from (4.4) that

$$\frac{|\partial\widetilde{\mathcal{R}}_n|}{n} \longrightarrow \frac{\nu_d}{1-p} \quad \text{almost surely.} \tag{4.5}$$

Let us show now another useful property. We claim that for any $\alpha > 0$,

$$\lim_{r\to\infty} \mathbb{P}(|(\mathcal{R}'_\infty)^+ \cap \widetilde{\mathcal{R}}_r^+| \geq \alpha r) = 0, \tag{4.6}$$

where $\mathcal{R}'_\infty$ is the total range of another simple random walk independent of $\widetilde{\mathcal{R}}$. To see this recall that the process $(\mathcal{R}_n)$ is increasing, and therefore using (4.4) one deduce

$$\mathbb{P}(|(\mathcal{R}'_\infty)^+ \cap \widetilde{\mathcal{R}}_r^+| \geq \alpha r) \leq \mathbb{P}\left(\widetilde{N}_r \geq \frac{p}{1-p} r\right) + \mathbb{P}(|(\mathcal{R}'_\infty)^+ \cap \mathcal{R}^+_{Cr}| \geq \alpha r),$$

with $C = 2p/(1-p) + 1$. The first term on the right-hand side goes to 0, in virtue of the law of large numbers satisfied by $\widetilde{N}$, and the second one also as can be seen using Markov's inequality and the estimate:

$$\mathbb{E}[|(\mathcal{R}'_\infty)^+ \cap \mathcal{R}^+_{Cr}|] \leq \sum_{x,y\in\overline{V}_0} \sum_{z\in\mathbb{Z}^d} G(0, z+x) G_{Cr}(0, z+y) = \mathcal{O}(\sqrt{r}\log r),$$

which follows from (2.3) and [11, Theorem 3.6].

A consequence of (4.6) is the following. Define $c = \nu_d/[2(1-p)]$. We have that for k large enough, any $t \geq 1$, and $r \geq \sqrt{k}$

$$\mathbb{P}(|\partial\widetilde{\mathcal{R}}_k| \geq t) \geq \frac{1}{2} \quad \Longrightarrow \quad \mathbb{P}(|\partial\widetilde{\mathcal{R}}_{k+r}| \geq t + cr) \geq \frac{1}{4}, \tag{4.7}$$

and also

$$\mathbb{P}(|\partial\widetilde{\mathcal{R}}_k| \leq t) \geq \frac{1}{2} \quad \Longrightarrow \quad \mathbb{P}(|\partial\widetilde{\mathcal{R}}_{k-r}| \leq t - cr) \geq \frac{1}{4}. \tag{4.8}$$

To see this first note that the set-inequality (1.9) holds as well for $\widetilde{\mathcal{R}}$. Hence, with evident notation

$$|\partial\widetilde{\mathcal{R}}_{k+r}| \geq |\partial\widetilde{\mathcal{R}}_r| + |\partial\widetilde{\mathcal{R}}(r, k+r)| - 2|\widetilde{\mathcal{R}}^+(r, r+k) \cap \widetilde{\mathcal{R}}_r^+|. \tag{4.9}$$

Now observe that the last intersection term is stochastically dominated by $|(\mathcal{R}'_\infty)^+ \cap \widetilde{\mathcal{R}}_r^+|$, with $\mathcal{R}'_\infty$ a copy of $\mathcal{R}_\infty$, independent of $\widetilde{\mathcal{R}}_r^+$. Therefore, (4.5), (4.6) and (4.9) immediately give (4.7) and (4.8).

4.3 Lower bound

First, by using (1.13), there is a positive constant $C_0 > 2d$, such that

$$|\mathbb{E}[|\partial\mathcal{R}_n|] - v_d n| \le C_0\sqrt{n} \quad \text{for all } n \ge 1. \tag{4.10}$$

Take k_n to be the integer part of $(1-p)n$. We have either of the two possibilities

$$\text{(i)} \quad \mathbb{P}(|\partial\widetilde{\mathcal{R}}_{k_n}| \le v_d n) \ge \frac{1}{2} \quad \text{or} \quad \text{(ii)} \quad \mathbb{P}(|\partial\widetilde{\mathcal{R}}_{k_n}| \ge v_d n) \ge \frac{1}{2}. \tag{4.11}$$

Assume for instance that (i) holds, and note that (ii) would be treated symmetrically. Define, $i_n = [(1-p)(n - A\sqrt{n})]$, with $A = 3C_0/(c(1-p))$, and note that using (4.8)

$$\mathbb{P}(|\partial\widetilde{\mathcal{R}}_{i_n}| \le v_d n - 3C_0\sqrt{n})) \ge \frac{1}{4}, \tag{4.12}$$

for n large enough. Now set

$$\mathcal{B}_n = \left\{ \frac{2\widetilde{N}_{i_n} - 2\mathbb{E}[\widetilde{N}_{i_n}]}{\sqrt{n}} \in [A+1, A+2] \right\}.$$

Note that there is a constant $c_A > 0$, such that for all n large enough

$$\mathbb{P}(\mathcal{B}_n) \ge c_A. \tag{4.13}$$

Moreover, by construction,

$$\mathcal{B}_n \subset \{i_n + 2\widetilde{N}_{i_n} \in [n, n + 3\sqrt{n}]\}. \tag{4.14}$$

Now using the independence of $\widetilde{N}$ and $\partial\widetilde{\mathcal{R}}$, (4.4), (4.12), (4.13) and (4.14), we deduce that

$$\mathbb{P}(\exists m \in \{0, \dots, 3\sqrt{n}\} \ : \ |\partial\mathcal{R}_{n+m}| \le v_d n - 3C_0\sqrt{n}) \ge \frac{c_A}{4}.$$

Then one can use the deterministic bound:

$$|\partial\mathcal{R}_n| \le |\partial\mathcal{R}_{n+m}| + 2dm,$$

which holds for all $n \ge 1$ and $m \ge 0$. This gives

$$\mathbb{P}(|\partial\mathcal{R}_n| \le v_d n - 2C_0\sqrt{n}) \ge \frac{c_A}{4},$$

which implies that $\mathrm{Var}(|\partial\mathcal{R}_n|)/n \ge C_0^2 c_A/4 > 0$, using (4.10).

5 On Le Gall's decomposition

In this section, we establish Proposition 1.8 and Theorem 1.10.

5.1 Mean and variance

Inequality (1.9) holds since

$$z \in \partial\mathcal{R}(0,n)\backslash(S_n + \mathcal{R}(n,n+m))^+ \cup \partial(S_n + \mathcal{R}(n,n+m))\backslash\mathcal{R}^+(0,n)$$
$$\Longrightarrow z \in \partial\mathcal{R}(0,n+m).$$

Define

$$X(i,j) = |\partial\mathcal{R}(i,j)| \quad \text{and} \quad \overline{X}(i,j) = X(i,j) - \mathbb{E}[X(i,j)].$$

Observe that in (1.9) the deviation from linearity is written in terms of an intersection of two independent range-like sets. This emphasizes the similarity between range and boundary of the range. Now (1.9) implies the same inequalities for the expectation.

$$0 \geq \mathbb{E}[X(0,n+m)] - (\mathbb{E}[X(0,n)] + \mathbb{E}[X(n,n+m)]) \geq -\mathbb{E}[Z(n,m)]. \tag{5.1}$$

Combining (1.9) and (5.1), we obtain our key (and simple) estimates

$$|\overline{X}(0,n+m) - (\overline{X}(0,n) + \overline{X}(n,n+m))| \leq \max(Z(n,m), \mathbb{E}[Z(n,m)]). \tag{5.2}$$

If $\|X\|_p = (\mathbb{E}[X^p])^{1/p}$, then using the triangle inequality, we obtain for any $p > 0$,

$$|\|\overline{X}(0,n+m)\|_p - \|\overline{X}(0,n) + \overline{X}(n,n+m)\|_p| \leq \|Z(n,m)\|_p + \|Z(n,m)\|_1. \tag{5.3}$$

The deviation from linearity of the centered p-th moment will then depend on the p-th moment of $Z(n,m)$. We invoke now Hammersley's Lemma [8], which extends the classical subadditivity argument in a useful manner.

Lemma 5.1 (Hammersley) *Let (a_n), (b_n), and (b'_n) be sequences such that*

$$a_n + a_m - b'_{n+m} \leq a_{m+n} \leq a_n + a_m + b_{n+m} \quad \text{for all } m \text{ and } n. \tag{5.4}$$

Assume also that the sequences (b_n) and (b'_n) are positive and non-decreasing, and satisfy

$$\sum_{n>0} \frac{b_n + b'_n}{n(n+1)} < \infty. \tag{5.5}$$

Then, the limit of a_n/n exists, and

$$-\frac{b'_n}{n}+4\sum_{k>2n}\frac{b'_k}{k(k+1)}\geq\frac{a_n}{n}-\lim_{k\to\infty}\frac{a_k}{k}\geq+\frac{b_n}{n}-4\sum_{k>2n}\frac{b_k}{k(k+1)}. \tag{5.6}$$

We obtain now the following moment estimate.

Lemma 5.2 *For any integer k, there is a constant C_k such that for any n, m integers,*

$$\mathbb{E}[Z^k(n,m)]\leq C_k(\psi_d^k(n)\psi_d^k(m))^{1/2}. \tag{5.7}$$

Recall that ψ_d is defined in (1.11).

Proof Observe that $Z(n,m)$ is bounded as follows.

$$\begin{aligned}
Z(n,m)&\leq 2\sum_{z\in\mathbb{Z}^d}\mathbb{1}\{z\in\mathcal{R}^+(0,n)\cap(S_n+\mathcal{R}^+(n,n+m))\}\\
&\leq 2\sum_{z\in\mathbb{Z}^d}\mathbb{1}\{z\in\overleftarrow{\mathcal{R}}^+(0,n)\cap\mathcal{R}^+(n,n+m)\}\\
&\leq 2\sum_{x\in\overline{V}_0}\sum_{y\in\overline{V}_0}\sum_{z\in\mathbb{Z}^d}\mathbb{1}\{z+x\in\overleftarrow{\mathcal{R}}(0,n),\ z+y\in\mathcal{R}(n,n+m)\}.
\end{aligned}$$

We now take the expectation of the k-th power, and use the independence of $\overleftarrow{\mathcal{R}}(0,n)$ and $\mathcal{R}(n,n+m)$. Then (recalling the definition (2.4) of G_n and using (2.5) for the last inequality)

$$\begin{aligned}
\mathbb{E}[Z^k(n,m)]&\leq 2^k\sum_{x_1,y_1\in\overline{V}_0}\cdots\sum_{x_k,y_k\in\overline{V}_0}\sum_{z_1,\dots,z_k}\mathbb{E}\Bigg[\prod_{i=1}^k\mathbb{1}\{z_i+x_i\in\overleftarrow{\mathcal{R}}(0,n)\}\\
&\qquad\times\mathbb{1}\{z_i+y_i\in\mathcal{R}(n,n+m)\}\Bigg]\\
&\leq 2^k\sum_{x_1,y_1\in\overline{V}_0}\dots\sum_{x_k,y_k\in\overline{V}_0}\sum_{z_1,\dots,z_k}\mathbb{P}(H_{z_i+x_i}<n,\ \forall i=1,\dots,k)\\
&\qquad\times\mathbb{P}(H_{z_i+y_i}<m,\ \forall i=1,\dots,k)\\
&\leq 2^k|\overline{V}_0|^{2k}\Bigg(\sum_{z_1,\dots,z_k}\mathbb{P}(H_{z_i}<n,\ \forall i=1,\dots,k)^2\Bigg)^{1/2}\\
&\qquad\times\Bigg(\sum_{z_1,\dots,z_k}\mathbb{P}(H_{z_i}<m,\ \forall i=1,\dots,k)^2\Bigg)^{1/2}\\
&\leq 2^k|\overline{V}_0|^{2k}k!\Bigg(\sum_{z_1,\dots,z_k}G_n^2(0,z_1)\dots G_n^2(z_{k-1},z_k)\Bigg)^{1/2}
\end{aligned}$$

$$\times \left(\sum_{z_1,\dots,z_k} G_m^2(0,z_1)\dots G_m^2(z_{k-1},z_k) \right)^{1/2}$$
$$\leq C_k(\psi_d^k(n)\psi_d^k(m))^{1/2},$$

which concludes the proof. □

Henceforth, and for simplicity, we think of ψ_d of (1.11) rather as $\psi_3(n) = \mathcal{O}(\sqrt{n})$, $\psi_4(n) = \mathcal{O}(\log(n))$ and for $d \geq 5$, $\psi_d(n) = \mathcal{O}(1)$ (in other words, we aggregate in ψ_d innocuous constants). As an immediate consequence of (5.1) and Lemma 5.2, we obtain for any $n, m \in \mathbb{N}$,

$$\mathbb{E}[|\partial\mathcal{R}_n|]+\mathbb{E}[|\partial\mathcal{R}_m|]-\max(\psi_d(n),\psi_d(m)) \leq \mathbb{E}[|\partial\mathcal{R}_{n+m}|] \leq \mathbb{E}[|\partial\mathcal{R}_n|]+\mathbb{E}[|\partial\mathcal{R}_m|]. \tag{5.8}$$

The inequalities of (5.8) and Hammersley's Lemma imply that the limit of $\mathbb{E}[|\partial\mathcal{R}_n|]/n$ exists and it yields (1.13) of Theorem 1.10.

Variance of $X(0,n)$. Let us write (5.3) for $p = 2$

$$|\|\overline{X}(0,n+m)\|_2 - \|\overline{X}(0,n) + \overline{X}(n,n+m)\|_2| \leq 2\|Z(n,m)\|_2. \tag{5.9}$$

Now, the independence of $\overline{X}(0,n)$ and $\overline{X}(n,n+m)$ gives

$$\|\overline{X}(0,n) + \overline{X}(n,n+m)\|_2^2 = \|\overline{X}(0,n)\|_2^2 + \|\overline{X}(0,m)\|_2^2. \tag{5.10}$$

By taking squares on both sides of (5.9) and using (5.10), we obtain

$$\begin{aligned}\|\overline{X}(0,n+m)\|_2^2 \leq{}& \|\overline{X}(0,n)\|_2^2 + \|\overline{X}(0,m)\|_2^2\\ &+ 4\|\overline{X}(0,n) + \overline{X}(n,n+m)\|_2\|Z(n,m)\|_2\\ &+ 4\|Z(n,m)\|_2^2,\end{aligned} \tag{5.11}$$

and

$$\begin{aligned}\|\overline{X}(0,n)\|_2^2 + \|\overline{X}(0,m)\|_2^2 \leq{}& \|\overline{X}(0,n+m)\|_2^2 + 4\|\overline{X}(0,n+m)\|_2\|Z(n,m)\|_2\\ &+ 4\|Z(n,m)\|_2^2.\end{aligned} \tag{5.12}$$

Now define for $\ell \geq 1$,

$$A_\ell := \sup_{2^\ell < i \leq 2^{\ell+1}} \|\overline{X}(0,i)\|_2^2.$$

Next, using (5.10), (5.11) and Lemma 5.2 with $k = 2$, we deduce that for any $\ell \geq 1$ and $\varepsilon > 0$ (using also the inequality $2ab \leq \varepsilon a^2 + b^2/\varepsilon$),

$$A_{\ell+1} \leq (1+\varepsilon)2A_\ell + \left(1 + \frac{1}{\varepsilon}\right)\psi_d^2(2^\ell). \tag{5.13}$$

We iterate this inequality L times to obtain for some constant C independent of L

$$A_L \le C(1+\varepsilon)^L 2^L + C\left(1+\frac{1}{\varepsilon}\right)\sum_{\ell=1}^{L}(1+\varepsilon)^{\ell-1}2^{\ell-1}\psi_d^2(2^{L-\ell})$$
$$\le C\,L^2\,2^L \quad \text{when we choose} \quad \varepsilon = \frac{1}{L}. \tag{5.14}$$

Then we use the rough bound of (5.14) as an a priori bound for the upper and lower bounds respectively (5.11) and (5.12) for the sequence $a_n = \mathrm{Var}(X(0,n))$, in order to apply Hammersley's Lemma with $b_n = b'_n = \sqrt{n}\log n \times \psi_d(n)$. In dimension four or more we do fulfill the hypotheses of Hammersley's Lemma, which in turn produces the improved bound $\mathrm{Var}(|\partial\mathcal{R}_n|) \le Cn$, and then again we can use Hammersley's Lemma with a smaller $b_n = b'_n = \sqrt{n}\psi_d(n)$ which eventually yields (1.14) of Theorem 1.10. The fact that the limit of the normalized variance is positive follows from Proposition 1.9.

5.2 Central limit theorem

The principle of Le Gall's decomposition is to repeat dividing each strand into smaller and smaller pieces producing independent boundaries of *shifted* ranges. For two reals s,t let $[s],[t]$ be their integer parts and define $X(s,t) = X([s],[t])$. For ℓ and k integer, let $X_{k,n}^{(\ell)} = X((k-1)n/2^\ell, kn/2^\ell)$. Let also $Z_{k,n}^{(\ell)}$ be the functional of the two strands obtained by dividing the k-th strand after $\ell-1$ divisions. In other words, as in (1.10) (but without translating here) let

$$Z_{k,n}^{(\ell)} = |U \cap \widetilde{U}^+| + |U^+ \cap \widetilde{U}|,$$

with

$$U := \left\{ S_{\left[(k-1)\frac{n}{2^\ell}\right]}, \dots, S_{\left[(2k-1)\frac{n}{2^{\ell+1}}\right]} \right\}, \quad \text{and} \quad \widetilde{U} := \left\{ S_{\left[(2k-1)\frac{n}{2^{\ell+1}}\right]}, \dots, S_{\left[k\frac{n}{2^\ell}\right]} \right\}.$$

Thus, after L divisions, with $2^L \le n$, we get

$$\sum_{i=1}^{2^L} X_{i,n}^{(L)} - \sum_{\ell=1}^{L}\sum_{i=1}^{2^{\ell-1}} Z_{i,n}^{(\ell)} \le X(0,n) \le \sum_{i=1}^{2^L} X_{i,n}^{(L)}.$$

The key point is that $\{X_{i,n}^{(L)},\ i=1,\dots,2^L\}$ are independent, and have the same law as $X(0,n/2^L)$ or $X(0,n/2^L+1)$. Now, we define the (nonnegative) error term $\mathcal{E}(n)$ as

$$X(0,n) = \sum_{i=1}^{2^L} X_{i,n}^{(L)} - \mathcal{E}(n),$$

and (1.9) and (1.13) imply that

$$\mathbb{E}[\mathcal{E}(n)] \leq \sum_{\ell=1}^{L} 2^{\ell} \psi_d\left(\frac{n}{2^{\ell}}\right).$$

Note that, in dimension $d \geq 4$, we can choose L growing to infinity with n, and such that $\mathbb{E}[\mathcal{E}(n)]/\sqrt{n}$ goes to 0: for instance $2^L = \sqrt{n}/\log^2(n)$. Therefore, for such choice of L, it suffices to prove the Central Limit Theorem for the sum $\sum_{i=1}^{2^L} X_{i,n}^{(L)}$. Our strategy is to apply the Lindeberg–Feller triangular array Theorem, that we recall for convenience (see for instance [5, Theorem 3.4.5] for a proof).

Theorem 5.3 (Lindeberg–Feller) *For each integer N let $(X_{N,i} : 1 \leq i \leq N)$ be a collection of independent random variables with zero mean. Suppose that the following two conditions are satisfied*

(i) $\sum_{i=1}^{N} \mathbb{E}[X_{N,i}^2] \to \sigma^2 > 0$ *as* $N \to \infty$ *and*
(ii) $\sum_{i=1}^{N} \mathbb{E}[(X_{N,i})^2 \mathbb{1}\{|X_{N,i}| > \varepsilon\}] \to 0$ *as* $N \to \infty$ *for all* $\varepsilon > 0$.
Then, $S_N = X_{N,1} + \ldots + X_{N,N} \Longrightarrow \sigma\mathcal{N}(0,1)$ *as* $N \to \infty$.

We apply Lindeberg-Feller's Theorem with $N = 2^L$ and $X_{N,i} = \overline{X}_{i,n}^{L}/\sqrt{n}$. The condition (i) was proved in the previous subsection. The condition (ii) is usually called Lindeberg's condition. To check (ii), we estimate the fourth moment of $\overline{X}(0,n)$, and as was noticed by Le Gall [11, Remark (iii) p. 503], this is achieved using the previous decomposition and a sub-additivity argument. More precisely, using (5.3) with $p = 4$, we have

$$\begin{aligned}\|\overline{X}(0,n+m)\|_4 \leq\ & ((\mathbb{E}[\overline{X}^4(0,n)] + 6\mathbb{E}[\overline{X}^2(0,n)]\mathbb{E}[\overline{X}^2(0,m)] + \mathbb{E}[\overline{X}^4(0,m)]))^{1/4} \\ & + \sqrt{\psi_d(n)\psi_d(m)}.\end{aligned}$$

Thus, if we define for $\ell \geq 1$,

$$A'_{\ell} := \sup_{2^{\ell} < i \leq 2^{\ell+1}} \|\overline{X}(0,i)\|_4,$$

we obtain (using also that $(a+b)^{1/4} \leq a^{1/4} + b^{1/4}$ for any a, b),

$$\begin{aligned}A'_{\ell+1} &\leq (2(A'_{\ell})^4 + 6A_{\ell}^2)^{1/4} + \psi_d(2^{\ell}) \\ &\leq 2^{1/4} A'_{\ell} + 6^{1/4} A_{\ell}^{1/2} + \psi_d(2^{\ell}).\end{aligned}$$

Define then $A''_{\ell} = \sup_{2^{\ell} < i \leq 2^{\ell+1}} \|\overline{X}(0,i)\|_4 / 2^{\ell/2}$, and recall that $A_{\ell} \leq C_d 2^{\ell}$, for some constant $C_d > 0$, in dimension four and larger. Therefore if $d \geq 4$,

$$A''_{\ell+1} \leq \frac{2^{1/4}}{2^{1/2}} A''_{\ell} + 6^{1/4} C_d + \frac{\psi_d(2^{\ell})}{2^{(\ell+1)/2}}.$$

This recursive inequality implies that (A''_ℓ) is bounded, as well as $n \mapsto \|\overline{X}(0,n)\|_4^2/n$. We then deduce that Lindeberg's condition is satisfied, and the Central Limit Theorem holds for $X(0,n)$.

Acknowledgments We would like to thank Gregory Maillard for discussions at an early stage of this work. We thank Pierre Mathieu for mentioning that we omitted to show that the limiting term in (1.14) was nonzero, and Perla Sousi for mentioning a few other inaccuracies. Finally, we thank an anonymous referee for his very careful reading, and his numerous corrections and suggestions which greatly improved the presentation. A.A. received support of the A*MIDEX Grant (ANR-11-IDEX-0001-02) funded by the French Government "Investissements d'Avenir" program.

Appendix: Estimates on ranges

In this section, we prove Proposition 1.5. We first introduce some other range-like sets allowing us to use the approach of Jain and Pruitt [9]. Recall that the sets $\mathcal{R}_{n,V}$ are disjoint, and for $U \subset V_0$, define

$$\overline{\mathcal{R}}_{n,U} := \bigcup_{V \supset U} \mathcal{R}_{n,V} = \{S_k \,:\, S_k \notin \mathcal{R}_{k-1} \text{ and } S_i \notin (S_k + U),\ i \le k-1,\ 1 \le k \le n\}. \tag{5.15}$$

Next for $U \subset V_0$, define

$$\alpha(U) = |\overline{\mathcal{R}}_{n,U}| - \mathbb{E}(|\overline{\mathcal{R}}_{n,U}|) \qquad \text{and} \qquad \beta(U) = |\mathcal{R}_{n,U}| - \mathbb{E}(|\mathcal{R}_{n,U}|).$$

The definition (5.15) yields

$$\alpha(U) = \sum_{V \supset U} \beta(V),$$

and this relation is inverted as follows:

$$\beta(V) = \sum_{U \supset V} (-1)^{|U \setminus V|} \alpha(U).$$

As a consequence, for $V \subset V_0$,

$$\mathrm{Var}(|\mathcal{R}_{n,V}|) = \mathbb{E}(\beta^2(V)) \le 2^{|V_0 \setminus V|} \sum_{U \supset V} \mathbb{E}(\alpha^2(U)).$$

We will see below that each $\overline{\mathcal{R}}_{n,V}$ has the same law as a range-like functional that Jain and Pruitt analyze by using a last passage decomposition, after introducing some new variables. But let us give more details now. So first, we fix some $V \subset V_0$, and for $n \in \mathbb{N}$, set $Z_n^n = 1$, and

$$\begin{aligned} Z_i &= \mathbf{1}(\{S_{i+k} \notin (S_i + \overline{V}) \quad \forall k \geq 1\}) && \forall i \in \mathbb{N}, \\ Z_i^n &= \mathbf{1}(\{S_{i+k} \notin (S_i + \overline{V}) \quad \forall k = 1, \ldots, n-i\}) && \forall i < n \\ W_i^n &= Z_i^n - Z_i && \forall i \leq n, \end{aligned}$$

where

$$\overline{V} = V \cup \{0\}.$$

A key point in this decomposition is that Z_n and Z_i^n are independent. Now, define

$$\underline{\mathcal{R}}_{n,V} = \{S_k : \ S_i \notin S_k + \overline{V}, \ n \geq i > k, \ 0 \leq k < n\}, \quad \text{and} \quad |\underline{\mathcal{R}}_{n,V}| = \sum_{i=0}^{n-1} Z_i^n. \tag{5.16}$$

Since the increments are symmetric and independent, $|\overline{\mathcal{R}}_{n,V}|$ and $|\underline{\mathcal{R}}_{n,V}|$ are equal in law. Now, equality (5.16) reads as

$$|\underline{\mathcal{R}}_{n,V}| = \sum_{i=0}^{n-1} Z_i + \sum_{i=0}^{n-1} W_i^n.$$

Now using that $\mathrm{Var}(|\overline{\mathcal{R}}_{n,V}|) \leq \mathbb{E}[(|\overline{\mathcal{R}}_{n,V}| - \sum_{i \leq n-1} \mathbb{E}[Z_i])^2]$, and that $(a+b)^2 \leq 2(a^2+b^2)$ we obtain

$$\mathrm{Var}(|\underline{\mathcal{R}}_{n,V}|) \leq 2 \sum_{i=1}^{n-1} \mathrm{Var}(Z_i) + 4 \sum_{j=1}^{n-1} \sum_{i=0}^{j-1} \mathrm{Cov}(Z_i, Z_j) + 4 \sum_{j=1}^{n-1} \sum_{i=0}^{j} \mathbb{E}(W_i^n W_j^n). \tag{5.17}$$

Next for $i < j < n$, we have (recall the definition (2.1))

$$\begin{aligned} \mathbb{E}\left(W_i^n \, W_j^n\right) &= \mathbb{P}\left(n < H_{S_i+\overline{V}}^{(i+1)} < \infty, \ n < H_{S_j+\overline{V}}^{(j+1)} < \infty\right) \\ &= \sum_{x \notin \overline{V}} \mathbb{P}(S_{j-i} = x, \ H_{\overline{V}} > j-i) \, \mathbb{P}_x(n-j < H_{\overline{V}} \\ &\qquad < \infty, \ n-j < H_{x+\overline{V}} < \infty) \\ &\leq \sum_{x \notin V} \mathbb{P}(S_{j-i} = x, \ H_{\overline{V}} > j-i) \, \mathbb{P}_x(n-j < H_{\overline{V}} \\ &\qquad < \infty, \ n-j < H_{x+\overline{V}} < \infty), \end{aligned}$$

where for the second equality we just used the Markov property and translation invariance of the walk. The last inequality is written to cover as well the case $i = j$. Therefore,

$$\sum_{i=0}^{j} \mathbb{E}\left(W_i^n W_j^n\right) \le \sum_{x\notin V} G_j(0,x)\,\mathbb{P}_x(H_{\overline{V}} < \infty,\ n-j < H_{x+\overline{V}} < \infty)$$
$$\le \sum_{y,z\in\overline{V}} \sum_{x\notin V} G_j(0,x)\,\mathbb{P}_x(H_y < \infty,\ n-j < H_{x+z} < \infty).$$

Then Lemma 4 of [9] shows that

$$\sum_{i=0}^{j} \mathbb{E}\left(W_i^n W_j^n\right) = \begin{cases} \mathcal{O}\left(\sqrt{\frac{j}{n-j}}\right) & \text{if } d=3 \\ \mathcal{O}\left(\frac{\log j}{n-j}\right) & \text{if } d=4 \\ \mathcal{O}\left((n-j)^{1-d/2}\right) & \text{if } d\ge 5, \end{cases}$$

and thus

$$\sum_{j=1}^{n-1}\sum_{i=0}^{j} \mathbb{E}\left(W_i^n W_j^n\right) = \begin{cases} \mathcal{O}(n) & \text{if } d=3 \\ \mathcal{O}((\log n)^2) & \text{if } d=4 \\ \mathcal{O}(1) & \text{if } d\ge 5. \end{cases} \tag{5.18}$$

Now, for $i < j < n$, by using that Z_i^j and Z_j are independent, we get

$$\mathrm{Cov}(Z_i, Z_j) = -\mathrm{Cov}(W_i^j, Z_j).$$

On the other hand, assuming $i < j \le n$,

$$\mathbb{E}(W_i^j Z_j) = \mathbb{P}\left(j < H_{S_i+\overline{V}}^{(i+1)} < \infty,\ H_{S_j+\overline{V}}^{(j+1)} = \infty\right)$$
$$= \sum_{x\notin\overline{V}} \mathbb{P}(S_{j-i} = x,\ H_{\overline{V}} > j-i)\,\mathbb{P}_x(H_{\overline{V}} < \infty,\ H_{x+\overline{V}} = \infty).$$

Since in addition,

$$\mathbb{E}(Z_j) = \mathbb{P}_x(H_{x+\overline{V}} = \infty) \quad \text{for all } x,$$

and

$$\mathbb{E}(W_i^j) = \sum_{x\notin\overline{V}} \mathbb{P}(S_{j-i} = x,\ H_{\overline{V}} > j-i)\,\mathbb{P}_x(H_{\overline{V}} < \infty),$$

we deduce that

$$\mathrm{Cov}(Z_i, Z_j) = \sum_{x\notin\overline{V}} \mathbb{P}(S_{j-i} = x,\ H_{\overline{V}} > j-i)\,b_V(x),$$

with

$$b_V(x) := \mathbb{P}_x(H_{\overline{V}} < \infty)\,\mathbb{P}_x(H_{x+\overline{V}} = \infty) - \mathbb{P}_x(H_{\overline{V}} < \infty,\ H_{x+\overline{V}} = \infty). \tag{5.19}$$

Now we need the following equivalent of Lemma 5 of [9].

Lemma 5.4 *For any $V \subset V_0$, and $x \notin \overline{V}$,*

$$b_V(x) = \mathbb{P}_x(H_{\overline{V}} < H_{x+\overline{V}} < \infty)\mathbb{P}_x(H_{\overline{V}} = \infty) + \mathcal{E}(x, V),$$

with

$$\mathcal{E}(x, V) := \sum_{z \in x+\overline{V}} \mathbb{P}_x(S_{H_{x+\overline{V}}} = z,\ H_{x+\overline{V}} < H_{\overline{V}})(\mathbb{P}_z(H_{\overline{V}} < \infty) - \mathbb{P}_x(H_{\overline{V}} < \infty)).$$

Moreover,

$$|\mathcal{E}(x, V)| = \mathcal{O}\left(\frac{1}{\|x\|^{d-1}}\right).$$

Assuming this lemma for a moment, we get

$$\begin{aligned} a_j &= \sum_{i=0}^{j-1} \mathrm{Cov}(Z_i, Z_j) = \sum_{i=0}^{j-1} \sum_{x \notin \overline{V}} \mathbb{P}(S_{j-i} = x,\ H_{\overline{V}} > j - i)\, b_V(x) \\ &= \mathcal{O}\left(\sum_{x \notin \overline{V}} \frac{G_j(0, x)}{\|x\|^{d-1}}\right) = \mathcal{O}\left(\sum_{1 \le \|x\| \le j} \frac{1}{\|x\|^{2d-3}}\right) \\ &= \begin{cases} \mathcal{O}(\log j) & \text{if } d = 3 \\ \mathcal{O}(1) & \text{if } d \ge 4, \end{cases} \end{aligned}$$

from which we deduce that

$$\sum_{j=0}^{n-1} a_j = \begin{cases} \mathcal{O}(n \log n) & \text{if } d = 3 \\ \mathcal{O}(n) & \text{if } d \ge 4. \end{cases} \tag{5.20}$$

Then Proposition 1.5 follows from (5.17), (5.18) and (5.20).

Proof of Lemma 5.4. Note first that

$$\begin{aligned} b_V(x) &= \mathbb{P}_x(H_{\overline{V}} < \infty,\ H_{x+\overline{V}} < \infty) - \mathbb{P}_x(H_{\overline{V}} < \infty)\mathbb{P}_x(H_{x+\overline{V}} < \infty) \\ &= \mathbb{P}_x(H_{\overline{V}} < H_{x+\overline{V}} < \infty) + \mathbb{P}_x(H_{x+\overline{V}} < H_{\overline{V}} < \infty) \\ &\quad - \mathbb{P}_x(H_{\overline{V}} < \infty)\mathbb{P}_x(H_{x+\overline{V}} < \infty). \end{aligned}$$

Moreover,

$$\begin{aligned}\mathbb{P}_x(H_{x+\overline{V}} < H_{\overline{V}} < \infty) &= \sum_{z\in x+\overline{V}} \mathbb{P}_x(S_{H_{x+\overline{V}}} = z,\ H_{x+\overline{V}} < H_{\overline{V}})\mathbb{P}_z(H_{\overline{V}} < \infty)\\ &= \mathbb{P}_x(H_{x+\overline{V}} < H_{\overline{V}})\mathbb{P}_x(H_{\overline{V}} < \infty) + \mathcal{E}(x, V).\end{aligned}$$

The first assertion of the lemma follows. The last assertion is then a direct consequence of standard asymptotics on the gradient of the Green's function (see for instance [10, Corollary 4.3.3]). □

Remark 5.5 By adapting the argument in [9] we could also prove that in dimension 3, $\mathrm{Var}(|\underline{\mathcal{R}}_{n,V}|) \sim \sigma^2 n \log n$, for some constant $\sigma > 0$, and then obtain a central limit theorem for this modified range. However it is not clear how to deduce from it an analogous result for $|\mathcal{R}_{n,V}|$, which would be useful in view of a potential application to the boundary of the range.

References

1. Asselah, A., Schapira, B.: Moderate deviations for the range of a transient walk: path concentration. arXiv:1601.03957
2. Benjamini, I., Kozma, G., Yadin, A., Yehudayoff, A.: Entropy of random walk range. Ann. Inst. Henri Poincaré Probab. Stat. **46**, 1080–1092 (2010)
3. Berestycki, N., Yadin, A.: Condensation of random walks and the Wulff crystal. arXiv:1305.0139
4. Chen, X.: Random walk intersections, large deviations and related topics. In: Mathematical Surveys and Monographs, vol. 157. AMS, New York (2009)
5. Durrett, R.: Probability: Theory and Examples, vol. 4. Cambridge University Press, Cambridge (2010)
6. Dvoretzky, A., Erdós, P.: Some problems on random walk in space. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 353–367. University of California Press, Berkeley (1951)
7. Hebisch, W., Saloff-Coste, L.: Gaussian estimates for Markov chains and random walks on groups. Ann. Probab. **21**, 673–709 (1993)
8. Hammersley, J.M.: Generalization of the fundamental theorem on subadditive functions. Proc. Cambr. Phil. Soc. **58**, 235–238 (1962)
9. Jain, N.C., Pruitt, W.E.: The range of transient random walk. J. Analyze Math. **24**, 369–393 (1971)
10. Lawler, G.F., Limic, V.: Random walk: a modern introduction. In: Cambridge Studies in Advanced Mathematics, vol. 123. Cambridge University Press, Cambridge (2010)
11. Le Gall, J.-F.: Propriétés d'intersection des marches aléatoires. I. Convergence vers le temps local d'intersection. Commun. Math. Phys. **104**, 471–507 (1986)
12. Mathieu, P., Sisto, A.: Deviation inequalities and CLT for random walks on acylindrically hyperbolic groups. arXiv:1411.7865
13. Okada, I.: The inner boundary of random walk range. J. Math. Soc. Japan **68**, 939–959 (2016)
14. Okada, I.: Frequently visited site of the inner boundary of simple random walk range. arXiv:1409.8368