



**HAL**  
open science

## Boundary of the Range I: Typical Behavior

Amine Asselah, Bruno Schapira

► **To cite this version:**

| Amine Asselah, Bruno Schapira. Boundary of the Range I: Typical Behavior. 2015. hal-01171542v1

**HAL Id: hal-01171542**

**<https://hal.science/hal-01171542v1>**

Preprint submitted on 4 Jul 2015 (v1), last revised 5 Jul 2016 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Boundary of the Range I: Typical Behavior.

AMINE ASSELAH\* & BRUNO SCHAPIRA†

## Abstract

We study the boundary of the range for the simple random walk on  $\mathbb{Z}^d$  in the transient regime  $d \geq 3$ . We show that sizes of the range and its boundary differ mainly by a martingale. As a consequence, we obtain a bound on the variance of order  $n \log n$  in dimension three. We also establish a central limit theorem in dimension four and larger.

## 1 Introduction

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

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

where  $x \sim y$  means that  $|x_1 - y_1| + \dots + |x_d - y_d| = 1$ . Our interest was triggered by a recent paper of Berestycki and Yadin [BY] 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, 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 [BKYY], 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 [Ok1] 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*

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

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

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

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 an independent copy.

---

\*LAMA, UPEC & IMéRA E-mail: amine.asselah@u-pec.fr

†Aix-Marseille Université, CNRS, Centrale Marseille, I2M, UMR 7373, 13453 Marseille, France. E-mail: bruno.schapira@univ-amu.fr

The thrust of our study is to show that for a transient random walk, range and boundary of the range are highly correlated objects: when one of them shrinks, so does the other. We present two ways to appreciate their similar nature. First, the sizes of a range-like set and the boundary of the range differ roughly by a martingale. To make this 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  $(S_k + V_0) \cap \mathcal{R}_{k-1}^c = S_k + V$ . 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  $n$  integer*

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

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

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

Jain and Pruitt [JP] 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 1.3.** *There are positive constants  $C_d$  such that*

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

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

**Proposition 1.4.** *Assume that  $d = 3$ , and let  $V$  be a nonempty subset of  $V_0$ . There is a positive constant  $C$ , such that*

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

Then, a useful corollary of Propositions 1.2, 1.3 and 1.4 is the corresponding bound for the variance of the boundary of the range in dimension 3.

**Theorem 1.5.** *Assume that  $d = 3$ . Then, there is a positive constant  $C$ , such that*

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

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

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

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

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

Equality (1.8) is the basis of Le Gall celebrated paper [LG] 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 [C], 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  $\Lambda^+ = \Lambda + \overline{V}_0$ , 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 1.6.** *For any integers  $n < m$*

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

with

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

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

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

We have the following bounds for expectation and variance of the boundary of the range.

**Proposition 1.7.** *When dimension is larger than or equal to three, there are constants  $C_d$  such that for any integer  $n$*

$$\frac{C_d \psi_d(n)}{n} \geq \frac{E[|\partial\mathcal{R}_n|]}{n} - \lim_{k \rightarrow \infty} \frac{E[|\partial\mathcal{R}_k|]}{k} \geq 0. \quad (1.12)$$

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

$$\left| \frac{\text{Var}(|\partial\mathcal{R}_n|)}{n} - \lim_{k \rightarrow \infty} \frac{\text{Var}(|\partial\mathcal{R}_k|)}{k} \right| \leq \frac{C_d \sqrt{n} \psi_d(n)}{n}, \quad (1.13)$$

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

**Remark 1.8.** 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 [Ok1] a large deviation principle for the upper tail (the probability that the boundary be larger than its mean), and in [Ok2] 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 [AS], 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 Section 2, we fix notation, recall known results on the Green's function, and prove a result about covering a finite subset. In Section 3, we establish the Martingale decomposition of Proposition 1.2 and prove Proposition 1.3. In Section 4, we present the dyadic decomposition for the boundary of the range and deduce Proposition 1.7, using Le Gall's argument. Finally in the Appendix, we prove Proposition 1.4.

## 2 Notation and Prerequisites

For any  $y, z \in \mathbb{Z}^d$ , we denote by  $\|z - y\|$  the graph distance between  $y$  and  $z$ , which is also the  $\ell_1$ -norm of  $(z - y)$ . 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  $x$ . For  $\Lambda$  a subset of  $\mathbb{Z}^d$  we define the hitting time of  $\Lambda$  as

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

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  $k \geq 2$ ,

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

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

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

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

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

and for  $n \geq 0$ ,

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

It is well known that for  $\psi_d$  defined in (1.11), we have, for some constant  $C > 0$

$$\sum_{z \in \mathbb{Z}^d} G_n^2(0, z) \leq C\psi_d(n). \quad (2.3)$$

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

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

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

$$G_A(u, v) = G_A(v, u) \quad \text{for all } u, v \in \mathbb{Z}^d,$$

and that  $G$  is also invariant by translation of the coordinates:  $G(u, v) = G(0, v - u)$ . Now a well-known bound, used extensively in this paper, is the following (see [LL, Theorem 4.3.1]):

$$G(0, z) = \mathcal{O} \left( \|z\|^{2-d} \right). \quad (2.4)$$

We can now state the main result of this Section.

**Lemma 2.1.** *Let  $\Lambda$  be a 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  $y \sim y'$ ,*

$$\mathbb{P}_y(\Lambda \subset \mathcal{R}_\infty) - \mathbb{P}_{y'}(\Lambda \subset \mathcal{R}_\infty) = c(\Lambda) \frac{\langle y' - y, y - z \rangle}{\|y - z\|^d} + \mathcal{O}\left(\frac{1}{\|y - z\|^d}\right). \quad (2.5)$$

Moreover,

$$c(\Lambda) = \frac{1}{dv_d} \sum_{\substack{x \in \Lambda, x' \notin \Lambda \\ x' \sim x}} \mathbb{P}_{x'}(H_\Lambda = \infty) \mathbb{P}_x(\Lambda \subset \mathcal{R}_\infty),$$

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

*Proof.* First, since  $\Lambda$  is finite, we can always assume that  $y$  and  $y'$  do not belong to  $\Lambda$ . Now by a first entry decomposition

$$\mathbb{P}_y(\Lambda \subset \mathcal{R}_\infty) = \sum_{x \in \Lambda} \mathbb{P}_y(S_{H_\Lambda} = x, H_\Lambda < \infty) \mathbb{P}_x(\Lambda \subset \mathcal{R}_\infty). \quad (2.6)$$

Next, fix  $x \in \Lambda$  and transform the harmonic measure into the restricted Green's function (see for instance [LL, 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 [LL, Proposition 4.6.2]) that

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

Therefore,

$$\begin{aligned} \mathbb{P}_y(S_{H_\Lambda} = x) - \mathbb{P}_{y'}(S_{H_\Lambda} = x) &= \frac{1}{2d} \sum_{v \in \Lambda^c, v \sim x} (G(v, y) - G(v, y')) \\ &\quad - \frac{1}{2d} \sum_{v \in \Lambda^c, v \sim x} \mathbb{E}_v \left[ \mathbb{1}\{H_\Lambda < \infty\} (G(S_{H_\Lambda}, y) - G(S_{H_\Lambda}, y')) \right]. \end{aligned} \quad (2.7)$$

Now, we have ([LL, Corollary 4.3.3]) the expansion for any  $z \in \Lambda$ ,

$$G(z, y) - G(z, y') = \frac{2}{v_d} \frac{\langle y' - y, y - z \rangle}{\|y - z\|^d} + \mathcal{O}\left(\frac{1}{\|y - z\|^d}\right). \quad (2.8)$$

Combining (2.6), (2.7) and (2.8) we obtain the result (2.5).  $\square$

### 3 Martingale Decomposition

In this Section, we establish Proposition 1.2, as well as Proposition 1.3 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

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

and

$$J_{k,V} = \mathbb{I}\{(S_k + V) \not\subseteq \{S_j, j \geq k\}\}.$$

Then for  $n \geq 1$ , define

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

and

$$M_{n,V} = \sum_{k=0}^{n-1} I_{k,V} \left( \mathbb{E}[J_{k,V} | \mathcal{F}_n] - \mathbb{E}[J_{k,V} | \mathcal{F}_k] \right),$$

with  $(\mathcal{F}_k, k \geq 0)$  the standard filtration of the walk. Note that by construction  $(M_{n,V}, n \geq 0)$  is a martingale. Moreover, the Markov property gives

$$\mathbb{E}[J_{k,V} | \mathcal{F}_k] = \mathbb{E}[J_{k,V}] = \mathbb{P}(V \not\subseteq \mathcal{R}_\infty) = \rho_V,$$

for all  $k \geq 0$ , where  $\mathcal{R}_\infty$  is the range of the whole path of the walk. In particular the second sum in the definition of  $M_{n,V}$  is precisely equal to  $\rho_V |\mathcal{R}_{n-1,V}|$ . Now for any  $k \leq n$ , write

$$J_{k,n,V} = \mathbb{I}\{(S_k + V) \not\subseteq \{S_k, \dots, S_n\}\}.$$

Then define

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

and

$$\mathcal{E}_{n,V} := \sum_{k=0}^{n-1} I_{k,V} \{J_{k,n,V} - \mathbb{E}[J_{k,V} | \mathcal{F}_n]\}.$$

With these definitions we get

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

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 an *error term*

$$\mathcal{E}_n = \sum_{V \subset V_0} \mathcal{E}_{n,V}.$$

Since  $\partial \mathcal{R}_n$  is the disjoint union of  $\partial \mathcal{R}_{n,V}$ 's over nonempty  $V \subset V_0$ , we obtain the first part (1.4) of Proposition 1.2.

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

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

Then by using the invariance of the walk by time inversion, we get

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

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

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

The desired result follows by using Cauchy-Schwarz.

### 3.2 Variance of the Martingale

We establish here Proposition 1.3. Let us notice here 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 1.3 is a direct consequence of the following result.

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

The proof of (3.4) is divided in three steps. The first step brings us to a decomposition of  $\Delta M_n$  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  $\Delta M_n$  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  $k \leq n$

$$\mathbb{E}[J_{k,V} | \mathcal{F}_n] = \sum_{V \cap V' \neq \emptyset} I_{k,n,V'} \mathbb{P}_{S_n - S_k} (V \cap V' \not\subseteq \mathcal{R}_\infty).$$

Note that  $I_{k,n,V'} \neq I_{k,n+1,V'}$  means that  $S_{n+1}$  and  $S_k$  are neighbors. However, the number of  $k$  indices such that  $S_{n+1}$  and  $S_k$  are neighbors and  $I_{k,V} = 1$  is at most  $2d$ . 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

$$\Delta M_n = \sum_{k=0}^{n-1} \sum_{V \cap V' \neq \emptyset} I_{k,V} I_{k,n,V'} \left\{ \mathbb{P}_{S_{n+1} - S_k} (V \cap V' \not\subseteq \mathcal{R}_\infty) - \mathbb{P}_{S_n - S_k} (V \cap V' \not\subseteq \mathcal{R}_\infty) \right\} + r_n, \quad (3.5)$$



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

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

with

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

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

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

Then by assuming for instance that  $\|z\| \leq \|z'\|$  (and  $z \neq z'$ ), and by using (2.4) we obtain

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

Therefore

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

Next we can for instance divide the last sum into two parts and we immediately get:

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

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

$$\tilde{I}_{k,n,U} = \sum_{V \cap V' \sim U} I_{k,V} I_{k,n,V'},$$

and

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

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

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

Then observe that for any  $U$ ,

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

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  $\tilde{I}_{k,n,U}$  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

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

where

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

Note that  $\partial \tilde{\mathcal{R}}_{n,U}$  reads also

$$\partial \tilde{\mathcal{R}}_{n,U} = \{S_k : k \leq n-1, (S_k + V_0) \cap \mathcal{R}_n^c \sim U\}.$$

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

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

Step 3. First note that

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

In dimension 4 or larger, (3.13) can be established as follows. First by using Cauchy-Schwarz inequality we get

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

for all  $z, z'$ . Then by using again the standard bound on Green's functions as in (3.9), we get

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

as desired. We consider now the more difficult case  $d = 3$ . Actually, 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  $z \in \partial \tilde{\mathcal{R}}_{n,U}$ , then in particular  $\|z\| \leq n$  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

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

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

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

Thus it only remains to bound the sum in (3.14) restricted to the  $z$  and  $z'$  satisfying  $\|z'\| \geq 2\|z\|$ . To this end observe that if  $z'$  is visited first,

$$\begin{aligned} & \sum_{1 \leq 2\|z\| \leq \|z'\| \leq n} \frac{|\langle z, z' \rangle|}{(1 + \|z\|^d)(1 + \|z'\|^d)} \mathbb{P}(H_{z'+V_0} < H_z < \infty) \\ &= \mathcal{O} \left( \sum_{1 \leq 2\|z\| \leq \|z'\| \leq n} \|z\|^{1-d} \|z'\|^{3-2d} \|z - z'\|^{2-d} \right) \\ &= \mathcal{O} \left( \sum_{1 \leq 2\|z\| \leq \|z'\| \leq n} \|z\|^{1-d} \|z'\|^{5-3d} \right) \\ &= \mathcal{O}(\log n), \end{aligned}$$

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 can also add in all the probabilities the events  $H_z < H_{z'+V_0}$ . Now for any  $z$ , call

$$\tau_z := \inf\{n \geq H_z : S_n \notin B(z, \|z\|/2)\},$$

and

$$\sigma_z := \inf\{n \geq \tau_z : S_n \in z + V_0\}.$$

Then define the event

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

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

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

Therefore, similar computations as above, show that in (3.14), we can replace the events  $z, z' \in \partial\tilde{\mathcal{R}}_{n,U}$  respectively by  $E_{z,n,U}$  and  $E_{z',n,U}$ . Finally a last simplification of the problem is to remove the indices  $n$ . More precisely define

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

The easiest part is to replace  $E_{z,n,U}$  by  $E_{z,U}$ , since it suffices to observe that

$$E_{z,n,U} \cap E_{z,U}^c \cap E_{z',n,U} \cap \{H_z < H_{z'+V_0}\} = \emptyset.$$

The analogue for  $z'$  is more delicate, but one can argue as follows. Recall that we assume  $1 \leq \|z\| \leq \|z'\|/2$ , and let  $k \leq n$ . By using (2.4) and (2.2), we get for a constant  $C > 0$  (and

recalling also the definition (2.1)) for some positive constants  $c$  and  $C$ ,

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

where for the second and third lines we used the Strong Markov Property, and Cauchy-Schwarz and (3.3) for the fourth one. In particular

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

Then it follows by using again the Markov property, that

$$\begin{aligned}
\mathbb{P}(E_{z,U}, E_{z',n,U} \cap E_{z',U}^c, H_z < H_{z'+V_0}) &\leq \mathbb{P}(E_{z',n,U} \cap E_{z',U}^c, H_z < H_{z'+V_0}) \\
&= \sum_{k \leq n} \mathbb{P}(H_z = n - k) \mathbb{P}_z(E_{z',k,U} \cap E_{z',U}^c) \\
&= \mathcal{O} \left( \frac{1}{\|z\|\|z'\|^2} \right).
\end{aligned}$$

Thus indeed we can replace the event  $z' \in \widetilde{\partial\mathcal{R}}_{n,U}$  by  $E_{z',U}$  in (3.14). Now in the remaining sum, we gather together the pairs  $(z, z')$  and  $(z, -z')$ , and we get, using Cauchy-Schwarz again,

$$\begin{aligned}
&\left| \sum_{1 \leq 2\|z\| \leq \|z'\| \leq n} \frac{\langle z, z' \rangle}{(1 + \|z\|^d)(1 + \|z'\|^d)} \mathbb{P}(E_{z,U}, E_{z',U}, H_z < H_{z'+V_0}) \right| \tag{3.16} \\
&\leq \sum_{1 \leq 2\|z\| \leq \|z'\| \leq n} \frac{2}{\|z\|^{d-1}\|z'\|^{d-1}} \left| \mathbb{P}(E_{z,U}, E_{z',U}, H_z < H_{z'+V_0}) - \mathbb{P}(E_{z,U}, E_{-z',U}, H_z < H_{-z'+V_0}) \right|.
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{P}(E_{z,U}, E_{z',U}, H_z < H_{z'+V_0}) \\
&= \sum_{y \in \overline{\partial}B(z, \|z\|/2)} \mathbb{P}(E_{z,U}^*, H_z < H_{z'+V_0}, S_{\tau_z} = y) \mathbb{P}_y(H_{z+V_0} = \infty, E_{z',U}), \tag{3.17}
\end{aligned}$$

where by  $\overline{\partial}B(z, \|z\|/2)$  we denote the set of points which are not in  $B(z, \|z\|/2)$  but have a neighbor in this ball, 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  $\|z'\| \geq 2\|z\|$ , by using again (2.4) we get

$$\mathbb{P}_y(H_{z+V_0} = \infty, E_{z',U}) = \mathbb{P}_y(E_{z',U}) - \mathcal{O} \left( \frac{1}{\|z\|^{d-2}\|z'\|^{d-2}} \right). \tag{3.18}$$

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  $1 \leq \|z\| \leq \|z'\|/2$  and  $y \in \bar{\partial}B(z, \|z\|/2)$  we get

$$\mathbb{P}_y(E_{z',U}) = \mathbb{P}_0(E_{z',U}) + \mathcal{O}\left(\frac{\|y\|}{\|z' - y\|^{d-1}}\right) = \mathbb{P}_0(E_{z',U}) + \mathcal{O}\left(\frac{\|z\|}{\|z'\|^{d-1}}\right). \quad (3.19)$$

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 (3.16), (3.17), (3.18) and (3.19), we obtain (3.13) and conclude the proof of (3.4).

## 4 On Le Gall's decomposition

In this Section, we establish Proposition 1.6 and 1.7.

### 4.1 Mean and Variance

Inequality (1.9) holds since

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

Define

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

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.

$$0 \geq \mathbb{E}[X(0, m)] - (\mathbb{E}[X(0, n)] + \mathbb{E}[X(n, m)]) \geq -\mathbb{E}[Z(0, n, m)]. \quad (4.1)$$

Combining (1.9) and (4.1), we obtain our key (and simple) estimates

$$|\bar{X}(0, m) - (\bar{X}(0, n) + \bar{X}(n, m))| \leq \max(Z(0, n, m), \mathbb{E}[Z(0, n, m)]). \quad (4.2)$$

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

$$\|\bar{X}(0, m)\|_p - \|\bar{X}(0, n) + \bar{X}(n, m)\|_p \leq \|Z(0, n, m)\|_p + \|Z(0, n, m)\|_1. \quad (4.3)$$

The deviation from linearity of the centered  $p$ -th moment will then depend on the  $p$ -th moment of  $Z(0, n, m)$ . We invoke now Hammersley's Lemma [HA], which extends the classical subadditivity argument in a useful manner.

**Lemma 4.1.** [Hammersley [HA]] *Let  $(a_n)$ ,  $(b_n)$ , and  $(b'_n)$  be sequences such that*

$$a_n + a_m - b'_m \leq a_{m+n} \leq a_n + a_m + b_m \quad \text{for all } m \text{ and } n, \quad (4.4)$$

*$(b_n)$  and  $(b'_n)$  are positive and non-decreasing, and*

$$\sum_{n>0} \frac{b_n + b'_n}{n(n+1)} < \infty. \quad (4.5)$$

*Then, the limit of  $a_n/n$  exists, and*

$$-\frac{b'_n}{n} + 4 \sum_{k>2n} \frac{b'_k}{k(k+1)} \geq \frac{a_n}{n} - \lim_{k \rightarrow \infty} \frac{a_k}{k} \geq \frac{b_n}{n} - 4 \sum_{k>2n} \frac{b_k}{k(k+1)}. \quad (4.6)$$

A first consequence is the following moment estimate.

**Lemma 4.2.** *For any integer  $k$ , there is a constant  $C_k$  such that for any  $n, m$  integers,*

$$E[Z^k(0, n, n+m)] \leq C_k (\psi_d^k(n) \psi_d^k(m))^{1/2}. \quad (4.7)$$

Recall that  $\psi_d$  is defined in (1.11).

*Proof.* Observe that  $Z(0, n, n+m)$  is bounded as follows.

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

We now take the expectation of the  $k$ -th power, and use the independence of  $\overleftarrow{\mathcal{R}}^+(0, n)$  and  $\mathcal{R}(n, n+m)$

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

which concludes the proof.  $\square$

As an immediate consequence of (4.1) and Lemma 4.2, we obtain for any  $n, m \in \mathbb{N}$ ,

$$\mathbb{E}[|\partial \mathcal{R}_n|] + \mathbb{E}[|\partial \mathcal{R}_m|] - \max(\psi_d(n), \psi_d(m)) \leq \mathbb{E}[|\partial \mathcal{R}_{n+m}|] \leq \mathbb{E}[|\partial \mathcal{R}_n|] + \mathbb{E}[|\partial \mathcal{R}_m|]. \quad (4.8)$$

The inequalities of (4.8) and Hammersley's Lemma imply that the limit of  $\mathbb{E}[|\partial \mathcal{R}_n|]/n$  exists and it yields (1.12) of Proposition 1.7.

**Variance of  $X(0, n)$ .** Lemma 4.2 with  $k = 2$  yields for  $n, m$

$$(\mathbb{E}[\overline{X}^2(0, n+m)])^{1/2} \leq \left( \mathbb{E}[\overline{X}^2(0, n)] + \mathbb{E}[\overline{X}^2(0, m)] \right)^{1/2} + \sqrt{\psi_d(n) \psi_d(m)}. \quad (4.9)$$

We use (4.9) to bootstrap Hammersley's Lemma. First, we deduce the following rough inequality. For  $\epsilon > 0$ ,

$$\text{Var}(X(0, 2n)) \leq (1 + \epsilon) 2 \text{Var}(X(0, n)) + \left(1 + \frac{1}{\epsilon}\right) \psi_d^2(n). \quad (4.10)$$

We iterate this inequality  $L$  times with  $L$  the integer part of  $\log n / \log 2$  to obtain for some constant  $C$  independent of  $n, L$

$$\begin{aligned} \text{Var}(X(0, n)) &\leq C(1 + \epsilon)^L 2^L + C'_d \left(1 + \frac{1}{\epsilon}\right) \sum_{k=1}^L (1 + \epsilon)^{k-1} 2^{k-1} \psi_d^2\left(\frac{n}{2^k}\right) \\ &\leq C n \log^2(n) \quad \text{when we choose} \quad \epsilon = \frac{1}{L}. \end{aligned} \quad (4.11)$$

Second, we use the rough bound of (4.11) as a priori bound for the right hand side of (4.9). Now, note that in dimension larger than three, we do fulfill the hypotheses of Hammersley's Lemma, which in turn produces the improved bound  $\text{Var}(|\partial\mathcal{R}_n|) \leq Cn$ , and then again used as an a priori bound in (4.9) it yields (1.13) of Proposition 1.7.

## 4.2 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  $X_k^{(l)} = X((k-1)n/2^l, kn/2^l)$ . Let also  $Z_k^{(l)}$  be the functional of the two strands obtained by dividing the  $k$ -th strand after  $l-1$  divisions. Thus, after  $L$  divisions, with  $2^L < n$ , we get

$$\sum_{i=1}^{2^L} X_i^{(L)} - \sum_{l=1}^L \sum_{i=1}^{2^{l-1}} Z_i^{(l)} \leq X(0, n) \leq \sum_{i=1}^{2^L} X_i^{(L)}.$$

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

$$\overline{X}(0, n) = \sum_{i=1}^{2^L} (X_i^{(L)} - \mathbb{E}[X_i^{(L)}]) - \mathcal{E}(n),$$

and (1.9) implies that

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

Note that we can choose  $L$  growing to infinity with  $n$ , but such that  $\mathbb{E}[\mathcal{E}(n)]/\sqrt{n}$  goes to 0.

We now use a bright remark of Le Gall [LG, Remark (iii) p.503]. In order to show that Lindeberg's condition for the triangular array  $\{X_i^{(L)}, i = 1, \dots, 2^L\}$  is satisfied, we estimate the fourth moment of  $X(0, n)$  using the previous decomposition and a sub-additivity argument. Using (4.3) with  $p = 4$ , we have

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

Take  $m = n$ , to obtain

$$\begin{aligned} \|\overline{X}(0, 2n)\|_4 &\leq \left( 2\mathbb{E}[\overline{X}^4(0, n)] + 6\mathbb{E}[\overline{X}^2(0, n)]^2 \right)^{1/4} + \psi_d(n) \\ &\leq 2^{1/4} \|\overline{X}(0, n)\|_4 + 6^{1/4} \|\overline{X}(0, n)\|_2 + \psi_d(n). \end{aligned}$$

Define  $A_n = \|\overline{X}(0, n)\|_4 / (\|\overline{X}(0, n)\|_2)$ , then

$$A_{2n} \leq \left( \frac{2^{1/4}}{2^{1/2}} A_n + \frac{6^{1/4}}{2^{1/2}} \right) \frac{2^{1/2} \|\overline{X}(0, n)\|_2}{\|\overline{X}(0, 2n)\|_2} + \frac{\psi_d(n)}{\|\overline{X}(0, 2n)\|_2}.$$

The inequality (4.2) is valid in any dimension. It is however only in dimension larger than three that we have a lower bound for the variance. Note that (4.2) yields

$$|\overline{X}(0, n) + \overline{X}(n, m)| \leq |\overline{X}(0, m)| + Z(0, n, m).$$

With  $m = 2n$ , we obtain

$$\left( 2\mathbb{E}[\overline{X}^2(0, n)] \right)^{1/2} \leq \|\overline{X}(0, 2n)\|_2 + \psi_d(n).$$

Thus,

$$\frac{2^{1/2} \|\overline{X}(0, n)\|_2}{\|\overline{X}(0, 2n)\|_2} \leq 1 + \frac{\psi_d(n)}{\|\overline{X}(0, 2n)\|_2}.$$

This implies that

$$A_{2n} \leq \frac{1 + o(1)}{2^{1/4}} A_n + \frac{6^{1/4}}{2^{1/2}} (1 + o(1)).$$

and therefore  $(A_n)$  is bounded. We deduce that Lindeberg's condition for triangular arrays is satisfied (see [F] page 530, or [HH]), and as a consequence the Central Limit Theorem holds.

## Appendix

### A Estimates on Ranges

In this section, we prove Proposition 1.4. We first introduce some other range-like sets allowing us to use the approach of Jain and Pruitt [JP]. Recall that the sets  $\mathcal{R}_{n,V}$  are disjoint, and for  $U \subset V_0$ , define

$$\overline{\mathcal{R}}_{n,U} := \bigcup_{V \supset U} \mathcal{R}_{n,V}. \tag{A.1}$$

Next for  $U \subset V_0$ , define

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

The definition (A.1) yields

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

and this relation can be inverted as follows:

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

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

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



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

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

where

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

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

$$\underline{\mathcal{R}}_{n,V} = \sum_{i=0}^{n-1} Z_i^n, \quad (\text{A.2})$$

and note that since the increments are symmetric and independent,  $\overline{\mathcal{R}}_{n,V}$  and  $\underline{\mathcal{R}}_{n,V}$  are equal in law. Now, equality (A.2) reads as

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

and thus by Cauchy-Schwarz,

$$\text{Var}(\underline{\mathcal{R}}_{n,V}) \leq 2 \sum_{i=1}^{n-1} \text{Var}(Z_i) + 4 \sum_{j=1}^{n-1} \sum_{i=0}^{j-1} \text{Cov}(Z_i, Z_j) + 4 \sum_{j=1}^{n-1} \sum_{i=0}^j \mathbb{E}(W_i^n W_j^n). \quad (\text{A.3})$$

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

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

where for the second equality we just used the Markov property and translation invariance of the walk. Therefore

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

Then Lemma 4 of [JP] shows that

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

and thus

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

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

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

On the other hand, similarly as above,

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

Since in addition,

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

and

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

we deduce that

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

with

$$b_V(x) := \mathbb{P}_x(H_{\bar{V}} < \infty) \mathbb{P}_x(H_{x+\bar{V}} = \infty) - \mathbb{P}_x(H_{\bar{V}} < \infty, H_{x+\bar{V}} = \infty). \quad (\text{A.5})$$

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

**Lemma A.1.** *For any  $V \subset V_0$ , and  $x \notin \bar{V}$ ,*

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

with

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

Moreover,

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

Assuming this lemma for a moment, we get

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

from which we deduce that

$$\sum_{j=0}^{n-1} a_j = \begin{cases} \mathcal{O}(n \log n) & \text{if } d = 3 \\ \mathcal{O}(n) & \text{if } d \geq 4. \end{cases} \quad (\text{A.6})$$

Then Proposition 1.4 follows from (A.3), (A.4) and (A.6).

*Proof of Lemma A.1.* Note first that

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

Moreover,

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

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 [LL, Corollary 4.3.3]).  $\square$

*Remark.* By adapting the argument in [JP] we could also prove that in dimension 3,  $\text{Var}(|\underline{\mathcal{R}}_{n,V}|) \sim \sigma^2 n \log n$ , 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.

**Acknowledgements.** We would like to thank Gregory Maillard for discussions at an early stage of this work. A.A. received support of the A\*MIDEX grant (ANR-11-IDEX-0001-02) funded by the French Government "Investissements d'Avenir" program.

## References

- [AS] **A. Asselah and B. Schapira**, *Boundary of the Range II: Lower Tail Large Deviations*.
- [BKYY] **I. Benjamini, G. Kozma, A. Yadin, A. Yehudayoff**, *Entropy of random walk range*, Ann. Inst. Henri Poincaré Probab. Stat. 46 (2010), 1080–1092.
- [BY] **N. Berestycki and A. Yadin**, *Condensation of random walks and the Wulff crystal*, arXiv:1305.0139.
- [C] **X. Chen**, *Random Walk Intersections, Large Deviations and Related Topics*. Mathematical Surveys and Monographs, Vol 157, 2009, AMS.
- [DE] **A. Dvoretzky, P. Erdős**, *Some problems on random walk in space*, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, pp. 353–367. University of California Press, Berkeley and Los Angeles, (1951).
- [F] **W. Feller**, *An introduction to probability theory and its applications, II*, New York, Wiley, 1961.

- [HSC] **W. Hebisch, L. Saloff-Coste**, *Gaussian estimates for Markov chains and random walks on groups*, Ann. Probab. 21, (1993), 673–709.
- [HH] **P. Hall, C. C. Heyde**, *Martingale limit theory and its application*, Probability and Mathematical Statistics. Academic Press, Inc., New York-London, (1980), xii+308 pp.
- [HA] **J. M. Hammersley**, *Generalization of the Fundamental Theorem on Subadditive Functions*, Proc. Cambridge Phil. Society, 58, (1962), 235–238.
- [JP] **N. C. Jain, W. E. Pruitt**, *The range of transient random walk*, J. Analyze Math. 24 (1971), 369–393.
- [LL] **G. F. Lawler and V. Limic**, *Random walk: a modern introduction*. Cambridge Studies in Advanced Mathematics, 123. Cambridge University Press, Cambridge, (2010). xii+364 pp.
- [LG] **J.-F. Le Gall**, *Propriétés d'intersection des marches aléatoires. I. Convergence vers le temps local d'intersection*, Comm. Math. Phys. 104 (1986), 471–507.
- [Ok1] **I. Okada** *The inner boundary of random walk range*, to appear in J. Math. Soc. Japan.
- [Ok2] **I. Okada** *Frequently visited site of the inner boundary of simple random walk range*, arXiv:1409.8368.