

# The range of tree-indexed random walk

Jean-François LE GALL, Shen LIN

Université Paris-Sud

July 18, 2014

## Abstract

We provide asymptotics for the range  $R_n$  of a random walk on the  $d$ -dimensional lattice indexed by a random tree with  $n$  vertices. Using Kingman's subadditive ergodic theorem, we prove under general assumptions that  $n^{-1}R_n$  converges to a constant, and we give conditions ensuring that the limiting constant is strictly positive. On the other hand, in dimension 4 and in the case of a symmetric random walk with exponential moments, we prove that  $R_n$  grows like  $n/\log n$ . We apply our results to asymptotics for the range of branching random walk when the initial size of the population tends to infinity.

**Keywords.** Tree-indexed random walk, range, discrete snake, branching random walk, subadditive ergodic theorem.

**AMS 2010 Classification Numbers.** 60G50, 60J80

## 1 Introduction

The main goal of this work is to derive asymptotics for the number of distinct sites of the lattice visited by a tree-indexed random walk. Asymptotics for the range of an ordinary random walk on the  $d$ -dimensional lattice  $\mathbb{Z}^d$  have been studied extensively since the pioneering work of Dvoretzky and Erdős [4]. Consider for simplicity the case of a simple random walk on  $\mathbb{Z}^d$ , and, for every integer  $n \geq 1$ , let  $R_n$  be the number of distinct sites of  $\mathbb{Z}^d$  visited by the random walk up to time  $n$ . When  $d \geq 3$ , let  $q_d > 0$  be the probability that the random walk never returns to its starting point. Then,

- if  $d \geq 3$ ,

$$\frac{1}{n} R_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} q_d,$$

- if  $d = 2$ ,

$$\frac{\log n}{n} R_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \pi,$$

- if  $d = 1$ ,

$$n^{-1/2} R_n \xrightarrow[n \rightarrow \infty]{(d)} \sup_{0 \leq t \leq 1} B_t - \inf_{0 \leq t \leq 1} B_t,$$

where  $\xrightarrow{(d)}$  indicates convergence in distribution and  $(B_t)_{t \geq 0}$  is a standard linear Brownian motion. The cases  $d \geq 3$  and  $d = 2$  were obtained in [4], whereas the case  $d = 1$  is a very easy consequence of Donsker's invariance theorem (see e.g. [6]). The preceding asymptotics have

been extended to much more general random walks. In particular, for any random walk in  $\mathbb{Z}^d$ , an application of Kingman's subadditive ergodic theorem [10] shows that the quantity  $R_n/n$  converges a.s. to the probability that the random walk does not return to its starting point (which is positive if the random walk is transient). See also [6] for the almost sure convergence of the (suitably normalized) range of an arbitrary recurrent random walk in the plane, [5] for a central limit theorem for the range of transient random walk, [15] for a non-Gaussian central limit theorem in the plane and [19] for a general study of the range of random walks in the domain of attraction of a stable distribution.

In the present work, we discuss similar asymptotics for tree-indexed random walk. We consider (discrete) plane trees, which are rooted ordered trees that can be viewed as describing the genealogy of a population starting with one ancestor or root, which is usually denoted by the symbol  $\emptyset$ . Given such a tree  $\mathcal{T}$  and a probability measure  $\theta$  on  $\mathbb{Z}^d$ , we can consider the random walk with jump distribution  $\theta$  indexed by the tree  $\mathcal{T}$ . This means that we assign a (random) spatial location  $Z_{\mathcal{T}}(u) \in \mathbb{Z}^d$  to every vertex  $u$  of  $\mathcal{T}$ , in the following way. First, the spatial location  $Z_{\mathcal{T}}(\emptyset)$  of the root is the origin of  $\mathbb{Z}^d$ . Then, we assign independently to every edge  $e$  of the tree  $\mathcal{T}$  a random variable  $X_e$  distributed according to  $\theta$ , and we let the spatial location  $Z_{\mathcal{T}}(u)$  of the vertex  $u$  be the sum of the quantities  $X_e$  over all edges  $e$  belonging to the simple path from  $\emptyset$  to  $u$  in the tree. The number of distinct spatial locations is called the range of the tree-indexed random walk  $Z_{\mathcal{T}}$ .

Let us state a particular case of our results.

**Theorem 1.** *Let  $\theta$  be a probability distribution on  $\mathbb{Z}^d$ , which is symmetric and has finite support. Assume that  $\theta$  is not supported on a strict subgroup of  $\mathbb{Z}^d$ . For every integer  $n \geq 1$ , let  $\mathcal{T}_n$  be a random tree uniformly distributed over all plane trees with  $n$  vertices. Conditionally given  $\mathcal{T}_n$ , let  $Z_{\mathcal{T}_n}$  be a random walk with jump distribution  $\theta$  indexed by  $\mathcal{T}_n$ , and let  $\mathcal{R}_n$  stand for the range of  $Z_{\mathcal{T}_n}$ . Then,*

- if  $d \geq 5$ ,

$$\frac{1}{n} \mathcal{R}_n \xrightarrow[n \rightarrow \infty]{(P)} c_{\theta},$$

where  $c_{\theta} > 0$  is a constant depending on  $\theta$ , and  $\xrightarrow{(P)}$  indicates convergence in probability;

- if  $d = 4$ ,

$$\frac{\log n}{n} \mathcal{R}_n \xrightarrow[n \rightarrow \infty]{L^2} 8 \pi^2 \sigma^4,$$

where  $\sigma^2 = (\det M_{\theta})^{1/4}$ , with  $M_{\theta}$  denoting the covariance matrix of  $\theta$ ;

- if  $d \leq 3$ ,

$$n^{-d/4} \mathcal{R}_n \xrightarrow[n \rightarrow \infty]{(d)} a_{\theta} \lambda_d(\text{supp}(\mathcal{I})),$$

where  $a_{\theta} = 2^{d/4}(\det M_{\theta})^{1/2}$  is a constant depending on  $\theta$ , and  $\lambda_d(\text{supp}(\mathcal{I}))$  stands for the Lebesgue measure of the support of the random measure on  $\mathbb{R}^d$  known as ISE (Integrated Super-Brownian Excursion).

Notice the obvious analogy with the results for the range of (ordinary) random walk that were recalled above. At an intuitive level,  $\mathcal{R}_n$  is likely to be smaller than the range  $R_n$  of ordinary random walk, because one expects many more self-intersections in the tree-indexed case. This is reflected in the fact that the ‘‘critical dimension’’ is now  $d = 4$  instead of  $d = 2$ . In the same way as  $d = 2$  is critical for the recurrence of random walk on  $\mathbb{Z}^d$ , one may say that

$d = 4$  is critical for the recurrence of tree-indexed random walk, in the sense that for random walk indexed by a “typical” large tree of size  $n$ , the number of returns to the origin will grow logarithmically with  $n$ . Furthermore, one may notice that the set of all spatial locations of  $\mathcal{T}_n$  is contained in the ball of radius  $Cn^{1/4}$  centered at the origin, with a probability close to 1 if the constant  $C$  is sufficiently large (see Janson and Marckert [8] or Kesten [9] in a slightly different setting), so that the range  $\mathcal{R}_n$  is at most of order  $n^{d/4}$  in dimension  $d \leq 3$ . We finally mention that the limiting constant  $c_\theta$  in dimension  $d \geq 5$  can again be interpreted as a probability of no return to the origin for random walk indexed by a certain infinite random tree: See Section 2 below for more details.

Let us emphasize that asymptotics of the type of Theorem 1 hold in a much more general setting. Firstly, it is enough to assume that the jump distribution  $\theta$  is centered and has sufficiently high moments (a little more is needed when  $d = 4$ ). Our argument to get the case  $d \geq 5$  of Theorem 1 relies on an application of Kingman’s subadditive ergodic theorem, which gives the convergence of  $\frac{1}{n}\mathcal{R}_n$  to a (possibly vanishing) constant in any dimension  $d$ , without any moment assumption on  $\theta$ . Secondly, in all cases except the critical dimension  $d = 4$ , we can handle more general random trees. Our methods apply to Galton-Watson trees with an offspring distribution having mean one and finite variance, which are conditioned to have exactly  $n$  vertices. In the special case where the offspring distribution is geometric with parameter  $1/2$ , we recover uniformly distributed plane trees, but the setting of conditioned Galton-Watson trees includes other important “combinatorial trees” such as binary trees or Cayley trees (see e.g. [16]). Some of our results even hold for an offspring distribution with infinite variance in the domain of attraction of a stable distribution.

In the present work, we deal with the cases  $d \geq 5$  and  $d = 4$  of Theorem 1, and the extensions that have just been described. The companion paper [18] addresses the “subcritical” case  $d \leq 3$ , which involves different methods and is closely related to the invariance principles connecting branching random walk with super-Brownian motion.

Let us turn to a more precise description of our main results and of our methods. In Section 2 below, we discuss the convergence of  $\frac{1}{n}\mathcal{R}_n$  in a general setting. The basic ingredient of the proof is the introduction of a suitable probability measure on a certain set of infinite trees. Roughly speaking, for any offspring distribution  $\mu$  with mean one, we construct a random infinite tree consisting of an infinite “spine” and, for each node of the spine, of a random number of Galton-Watson trees with offspring distribution  $\mu$  that branch off the spine at this node. This construction is related to the infinite size-biased Galton-Watson tree (see [21] and references therein), with the difference that we consider only subtrees branching off the right side of the spine. For a more precise description, see subsection 2.3. The law of our infinite tree turns out to be invariant under a shift transformation, which basically involves re-rooting the tree at the first vertex (in lexicographical order) that does not belong to the spine. If we consider a random walk (with an arbitrary jump distribution  $\theta$ ) indexed by this infinite tree, the number of distinct locations of the random walk at the first  $n$  vertices of the infinite tree yields a subadditive process  $R_n$ , to which we can apply Kingman’s theorem in order to get the almost sure convergence of  $\frac{1}{n}R_n$  to a constant (Theorem 4). One then needs to discuss the positivity of the limiting constant, and this leads to conditions depending both on the offspring distribution  $\mu$  and on the jump distribution  $\theta$ . More precisely, we give a criterion (Proposition 5) involving the Green function of the random walk and the generating function of  $\mu$ , which ensures that the limiting constant is positive. In the case when  $\mu$  has finite variance and if the jump distribution  $\theta$  is centered (with sufficiently high moments), this criterion is satisfied if  $d \geq 5$ . The preceding line of reasoning is of course very similar to the classical application of Kingman’s theorem to the range of ordinary

random walk. In the present setting however, additional ingredients are needed to transfer the asymptotics from the case of the infinite random tree to a single Galton-Watson tree conditioned to have  $n$  vertices. At this point, we need to assume that the offspring distribution  $\mu$  has finite variance or is in the domain of attraction of a stable distribution, so that we can use known results [3] on the scaling limit of the height process associated with a sequence of Galton-Watson trees with offspring distribution  $\mu$ : Applying these results to the sequence of trees that branch off the spine of the infinite tree yields information about the “large” trees in the sequence, which is essentially what we need to cover the case of a single Galton-Watson tree conditioned to be large (Theorem 7). The case  $d \geq 5$  of Theorem 1 follows as a special case of the results in Section 2.

Section 3, which is the most technical part of the paper, is devoted to the proof of a generalized version of the case  $d = 4$  of Theorem 1 (Theorem 14). We restrict our attention to the case when the offspring distribution is geometric with parameter  $1/2$ , and we assume that the jump distribution  $\theta$  is symmetric with small exponential moments. While the symmetry assumption can presumably be weakened without too much additional work, the existence of exponential moments is used at a crucial point of our proof where we rely on the multidimensional extension of the celebrated Komlós-Major-Tusnády strong invariance principle. Our approach is based on the path-valued Markov chain called the discrete snake. In our setting, this process, which we denote by  $(W_n)_{n \geq 0}$ , takes values in the space of all infinite paths  $w : (-\infty, \zeta] \cap \mathbb{Z} \rightarrow \mathbb{Z}^d$ , where  $\zeta = \zeta(w) \in \mathbb{Z}$  is called the lifetime of  $w$ . If  $\zeta_n$  denotes the lifetime of  $W_n$ , the process  $(\zeta_n)_{n \geq 0}$  evolves like simple random walk on  $\mathbb{Z}$ . Furthermore, if  $\zeta_{n+1} = \zeta_n - 1$ , the path  $W_{n+1}$  is obtained by restricting  $W_n$  to the interval  $(-\infty, \zeta_n - 1] \cap \mathbb{Z}$ , whereas if  $\zeta_{n+1} = \zeta_n + 1$ , the path  $W_{n+1}$  is obtained by adding to  $W_n$  one step distributed according to  $\theta$ . We assume that the initial value  $W_0$  is just a path (indexed by negative times) of the random walk with jump distribution  $\theta$  started from the origin. Then the values of the discrete snake generate a random walk indexed by an infinite random tree, which corresponds, in the particular case of the geometric offspring distribution, to the construction developed in Section 2. Note however that, in contrast with Section 2, the Markovian properties of the discrete snake play a very important role in Section 3. A key estimate (Proposition 8) states that the probability that the “head of the discrete snake” (that is the process  $(W_k(\zeta_k))_{k \geq 0}$ ) does not return to the origin before time  $n$  behaves like  $c/\log n$  for a certain constant  $c$ . This is analogous to the well-known asymptotics for the probability that simple random walk in  $\mathbb{Z}^2$  does not come back to its starting point before time  $n$ , but the proof, which is developed in subsection 3.2, turns out to be much more involved in our setting. The main result of Section 3 (Theorem 14) gives the case  $d = 4$  of Theorem 1 under slightly more general assumptions. It would be of interest to extend this result to more general offspring distributions, but this would require a different approach.

Section 4 applies the preceding results to asymptotics for the range of a branching random walk in  $\mathbb{Z}^d$ ,  $d \geq 4$ , when the size of the initial population tends to infinity. This study is related to the recent work of Lalley and Zheng [12], who discuss the number of distinct sites occupied by a nearest neighbor branching random walk in  $\mathbb{Z}^d$  at a fixed time. Note that the genealogical structures of descendants of the different initial particles are described by independent Galton-Watson trees, which makes it possible to apply our results about the range of tree-indexed random walk. Still one needs to verify that points that are visited by the descendants of two distinct initial particles give a negligible contribution in the limit. The analogous problem for low dimensions  $d \leq 3$  is addressed in [18].

Let us finally mention that there are certain analogs of our results in a continuous setting. In particular, Delmas [2] considers the Lebesgue measure of the tubular neighborhood of radius  $\varepsilon$  of

the range of a super-Brownian measure in  $\mathbb{R}^d$ ,  $d \geq 4$ , and proves that this measure behaves like  $\varepsilon^{d-4}$  when  $d \geq 5$  and like  $|\log \varepsilon|^{-1}$  when  $d = 4$ . This is analogous to our results for branching random walk in Section 4. An exact analog of the cases  $d \geq 4$  in Theorem 1 is obtained by considering the Lebesgue measure of the tubular neighborhood of radius  $\varepsilon$  of the support of the so-called integrated super-Brownian excursion (ISE). Asymptotics for this quantity when  $\varepsilon \rightarrow 0$  are derived in [2, Corollary 2.4]. We do not know of any analog of Proposition 8 in the continuous setting. Abraham and Werner [1] discuss a very similar problem for the Brownian snake and super-Brownian motion in  $\mathbb{R}^d$ , but only when  $d \leq 3$ .

*Notation.* We use the notation  $\llbracket a, b \rrbracket := [a, b] \cap \mathbb{Z}$  for  $a, b \in \mathbb{Z}$ , with  $a \leq b$ . Similarly,  $\llbracket -\infty, a \rrbracket := (-\infty, a] \cap \mathbb{Z}$  for  $a \in \mathbb{Z}$ . For any finite set  $A$ ,  $\#A$  denotes the cardinality of  $A$ .

## 2 Linear growth of the range

### 2.1 Finite trees

We use the standard formalism for plane trees. We set

$$\mathcal{U} := \bigcup_{n=0}^{\infty} \mathbb{N}^n,$$

where  $\mathbb{N} = \{1, 2, \dots\}$  and  $\mathbb{N}^0 = \{\emptyset\}$ . If  $u = (u_1, \dots, u_n) \in \mathcal{U}$ , we set  $|u| = n$  (in particular  $|\emptyset| = 0$ ). We write  $\prec$  for the lexicographical order on  $\mathcal{U}$ , so that  $\emptyset \prec 1 \prec (1, 1) \prec 2$  for instance.

If  $u, v \in \mathcal{U}$ ,  $uv$  stands for the concatenation of  $u$  and  $v$ . In particular  $\emptyset u = u\emptyset = u$ . The genealogical (partial) order  $\ll$  is then defined by saying that  $u \ll v$  if and only if  $v = uw$  for some  $w \in \mathcal{U}$ .

A plane tree (also called rooted ordered tree)  $\mathcal{T}$  is a finite subset of  $\mathcal{U}$  such that the following holds:

- (i)  $\emptyset \in \mathcal{T}$ .
- (ii) If  $u = (u_1, \dots, u_n) \in \mathcal{T} \setminus \{\emptyset\}$  then  $\hat{u} := (u_1, \dots, u_{n-1}) \in \mathcal{T}$ .
- (iii) For every  $u = (u_1, \dots, u_n) \in \mathcal{T}$ , there exists an integer  $k_u(\mathcal{T}) \geq 0$  such that, for every  $j \in \mathbb{N}$ ,  $(u_1, \dots, u_n, j) \in \mathcal{T}$  if and only if  $1 \leq j \leq k_u(\mathcal{T})$ .

The notions of a child and a parent of a vertex of  $\mathcal{T}$  are defined in an obvious way. The quantity  $k_u(\mathcal{T})$  in (iii) is the number of children of  $u$  in  $\mathcal{T}$ . If  $u \in \mathcal{T}$ , we write  $[\mathcal{T}]_u = \{v \in \mathcal{U} : uv \in \mathcal{T}\}$ , which corresponds to the subtree of descendants of  $u$  in  $\mathcal{T}$ . We denote the set of all plane trees by  $\mathbb{T}_f$ .

Throughout this work, we consider a probability measure  $\mu$  on  $\mathbb{Z}_+$ , which is critical in the sense that

$$\sum_{k=0}^{\infty} k \mu(k) = 1.$$

We exclude the degenerate case where  $\mu(1) = 1$ . The law of the Galton-Watson tree with offspring distribution  $\mu$  is a probability measure on the space  $\mathbb{T}_f$ , which we denote by  $\Pi_\mu$  (see e.g. [16, Section 1]).

We also let  $\theta$  be a probability measure on  $\mathbb{Z}^d$ , which is adapted in the sense that it is not supported on a strict subgroup of  $\mathbb{Z}^d$ .

A ( $d$ -dimensional) spatial tree is a pair  $(\mathcal{T}, (z_u)_{u \in \mathcal{T}})$  where  $\mathcal{T} \in \mathbb{T}_f$  and  $z_u \in \mathbb{Z}^d$  for every  $u \in \mathcal{T}$ . Let  $\mathbb{T}_f^*$  be the set of all spatial trees. We write  $\Pi_{\mu, \theta}^*$  for the probability distribution on  $\mathbb{T}_f^*$  under which  $\mathcal{T}$  is distributed according to  $\Pi_\mu$  and, conditionally on  $\mathcal{T}$ , the “spatial locations”  $(z_u)_{u \in \mathcal{T}}$  are distributed as random walk indexed by  $\mathcal{T}$ , with jump distribution  $\theta$ , and started from 0 at the root  $\emptyset$  (see the definition given in Section 1). We then set

$$a_{\mu, \theta} := \Pi_{\mu, \theta}^*(z_u \neq 0, \forall u \in \mathcal{T} \setminus \{\emptyset\}),$$

and, for every  $y \in \mathbb{Z}^d$ ,

$$h_{\mu, \theta}(y) := \Pi_{\mu, \theta}^*(z_u \neq -y, \forall u \in \mathcal{T}).$$

Notice that  $a_{\mu, \theta} > 0$ , simply because with positive probability a tree distributed according to  $\Pi_\mu$  consists only of the root.

## 2.2 Infinite trees

We now introduce a certain class of infinite trees. Each tree in this class will consist of an infinite ray or spine starting from the root, and finite subtrees branching off every node of this infinite ray. We label the vertices of the infinite ray by nonpositive integers  $0, -1, -2, \dots$ . The reason for labelling the vertices of the spine by negative integers comes from the fact that  $-1$  is viewed as the parent of  $0$ ,  $-2$  as the parent of  $-1$ , and so on.

More precisely, we consider the set

$$\mathcal{V} := \mathbb{Z}_- \times \mathcal{U}$$

where  $\mathbb{Z}_- = \{0, -1, -2, \dots\}$ . For every  $j \in \mathbb{Z}_-$ , we identify the element  $(j, \emptyset)$  of  $\mathcal{V}$  with the integer  $j$ , and we thus view  $\mathbb{Z}_-$  as a subset of  $\mathcal{V}$ . We define the lexicographical order on  $\mathcal{V}$  as follows. If  $j, j' \in \mathbb{Z}_-$ , we have  $j \prec j'$  if and only if  $j \leq j'$ . If  $u \in \mathcal{U} \setminus \{\emptyset\}$ , we have always  $j' \prec (j, u)$ . If  $u, u' \in \mathcal{U} \setminus \{\emptyset\}$ , we have  $(j, u) \prec (j', u')$  if either  $j > j'$ , or  $j = j'$  and  $u \prec u'$ . The genealogical (partial) order  $\ll$  on  $\mathcal{V}$  is defined in an obvious way: in agreement with the preceding heuristic interpretation, the property  $j \ll j'$  for  $j, j' \in \mathbb{Z}_-$  holds if and only if  $j \leq j'$ .

Let  $\mathcal{T}$  be a subset of  $\mathcal{V}$  such that  $\mathbb{Z}_- \subset \mathcal{T}$ . For every  $j \in \mathbb{Z}_-$ , we set

$$\mathcal{T}_j := \{u \in \mathcal{U} : (j, u) \in \mathcal{T}\}.$$

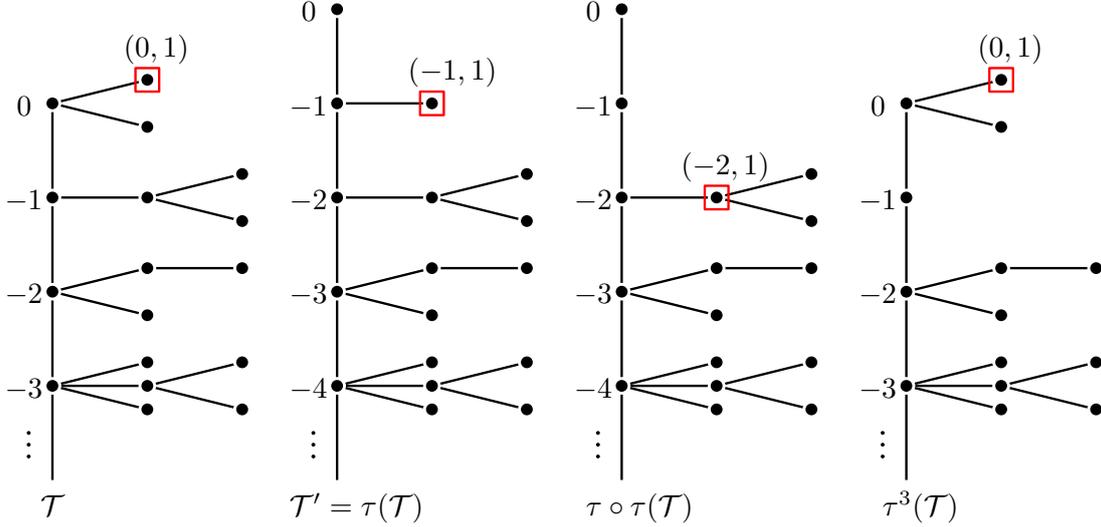
We say that  $\mathcal{T}$  is an *infinite tree* if, for every  $j \in \mathbb{Z}_-$ ,  $\mathcal{T}_j$  is a (finite) plane tree, and furthermore  $\mathcal{T} \setminus \mathbb{Z}_-$  is infinite. We write  $\mathbb{T}$  for the set of all infinite trees. By convention, the root of an infinite tree  $\mathcal{T}$  is the vertex  $0$ . Clearly,  $\mathcal{T}$  is determined by the collection  $(\mathcal{T}_j)_{j \in \mathbb{Z}_-}$ . Note that the lexicographical order of vertices corresponds to the order of visit when one “moves around” the tree in clockwise order, starting from the “bottom” of the spine and assuming that the “subtrees”  $\mathcal{T}_j$  are drawn on the right side of the spine, as in Fig. 1.

We next define a shift transformation  $\tau$  on the space  $\mathbb{T}$ . Starting from an infinite tree  $\mathcal{T}$ , its image  $\tau(\mathcal{T}) = \mathcal{T}'$  is obtained informally as follows. We look for the first vertex (in lexicographical order) of  $\mathcal{T} \setminus \mathbb{Z}_-$ . Call this vertex  $v$ . We then “re-root” the tree  $\mathcal{T}$  at  $v$  and, in the case when  $v$  is not a child of  $0$  (or equivalently if  $\mathcal{T}_0 = \{\emptyset\}$ ), we remove the vertices of the spine that are strict descendants of the parent of  $v$ .

For a more formal definition, let  $k \in \mathbb{Z}_-$  be the unique integer such that  $v \in \mathcal{T}_k$  (necessarily,  $v = (k, 1)$ ). Then,  $\mathcal{T}'$  is determined by requiring that:

- $\mathcal{T}'_j = \mathcal{T}_{j+k+1}$  if  $j \leq -2$ ;

- $\mathcal{T}'_0 = [\mathcal{T}_k]_1$ ;
- $\mathcal{T}'_{-1}$  is the unique plane tree such that there exists a bijection from  $\mathcal{T}_k \setminus \{u \in \mathcal{T}_k : 1 \ll u\}$  onto  $\mathcal{T}'_{-1}$  that preserves both the lexicographical order and the genealogical order.



**Figure 1:** The first 3 iterations of the shift transformation on an infinite tree  $\mathcal{T}$ . At each step, the marked vertex will become the new root after the shift.

Fig.1 explains the construction of  $\mathcal{T}'$  better than the formal definition.

### 2.3 The invariant measure on infinite trees

Let  $\mathbf{P}_\mu$  be the probability measure on  $\mathbb{T}$  that is determined by the following conditions. Under  $\mathbf{P}_\mu(d\mathcal{T})$ ,

- the trees  $\mathcal{T}_0, \mathcal{T}_{-1}, \mathcal{T}_{-2}, \dots$  are independent;
- $\mathcal{T}_0$  is distributed according to  $\Pi_\mu$ ;
- for every integer  $j \leq -1$ ,

$$\mathbf{P}_\mu(k_\emptyset(\mathcal{T}_j) = n) = \mu([n+1, \infty)),$$

for every  $n \geq 0$ ; furthermore, conditionally on  $k_\emptyset(\mathcal{T}_j) = n$ , the trees  $[\mathcal{T}_j]_1, [\mathcal{T}_j]_2, \dots, [\mathcal{T}_j]_n$  are independent and distributed according to  $\Pi_\mu$ .

Notice that  $\sum_{n \geq 0} \mu([n+1, \infty)) = 1$  due to the criticality of the probability measure  $\mu$ . The reason for introducing the probability measure  $\mathbf{P}_\mu$  comes from the next proposition.

**Proposition 2.** *The probability measure  $\mathbf{P}_\mu$  is invariant under the shift  $\tau$ .*

*Proof.* Suppose that  $\mathcal{T}$  is distributed according to  $\mathbf{P}_\mu$  and set  $\mathcal{T}' = \tau(\mathcal{T})$  as above. We need to verify that  $\mathcal{T}'$  is also distributed according to  $\mathbf{P}_\mu$ , or equivalently that the trees  $\mathcal{T}'_0, \mathcal{T}'_{-1}, \dots$  satisfy the same properties as  $\mathcal{T}_0, \mathcal{T}_{-1}, \dots$  above. The key point is to calculate the distribution

of  $(k_{\emptyset}(\mathcal{T}'_j), j \leq 0)$ . Fix an integer  $p \geq 1$ , and let  $n_0, n_1, \dots, n_p \in \mathbb{Z}_+$ . Also let  $k$  be the element of  $\mathbb{Z}_-$  determined as in the definition of  $\mathcal{T}' = \tau(\mathcal{T})$  at the end of subsection 2.2. The event

$$\{k = 0\} \cap \{k_{\emptyset}(\mathcal{T}'_0) = n_0, k_{\emptyset}(\mathcal{T}'_{-1}) = n_1, \dots, k_{\emptyset}(\mathcal{T}'_{-p}) = n_p\}$$

holds if and only if we have

$$k_{\emptyset}(\mathcal{T}_0) = n_1 + 1, k_1(\mathcal{T}_0) = n_0, k_{\emptyset}(\mathcal{T}_{-1}) = n_2, \dots, k_{\emptyset}(\mathcal{T}_{-p+1}) = n_p,$$

which occurs with probability

$$\mu(n_1 + 1)\mu(n_0)\mu([n_2 + 1, \infty)) \dots \mu([n_p + 1, \infty)).$$

Let  $\ell \in \mathbb{Z}_- \setminus \{0\}$ . Similarly, the event

$$\{k = \ell\} \cap \{k_{\emptyset}(\mathcal{T}'_0) = n_0, k_{\emptyset}(\mathcal{T}'_{-1}) = n_1, \dots, k_{\emptyset}(\mathcal{T}'_{-p}) = n_p\}$$

holds if and only if we have

$$k_{\emptyset}(\mathcal{T}_0) = 0, \dots, k_{\emptyset}(\mathcal{T}_{\ell+1}) = 0, k_{\emptyset}(\mathcal{T}_{\ell}) = n_1 + 1, k_1(\mathcal{T}_{\ell}) = n_0, k_{\emptyset}(\mathcal{T}_{\ell-1}) = n_2, \dots, k_{\emptyset}(\mathcal{T}_{\ell-p+1}) = n_p,$$

which occurs with probability

$$\mu(0)\mu([1, \infty))^{-\ell-1}\mu([n_1 + 2, \infty))\mu(n_0)\mu([n_2 + 1, \infty)) \dots \mu([n_p + 1, \infty)).$$

Summarizing, we see that the event

$$\{k_{\emptyset}(\mathcal{T}'_0) = n_0, k_{\emptyset}(\mathcal{T}'_{-1}) = n_1, \dots, k_{\emptyset}(\mathcal{T}'_{-p}) = n_p\}$$

has probability

$$\begin{aligned} & \mu(n_0)\mu(n_1 + 1)\mu([n_2 + 1, \infty)) \dots \mu([n_p + 1, \infty)) \\ & + \mu(n_0)\mu(0) \left( \sum_{\ell=-1}^{-\infty} \mu([1, \infty))^{-\ell-1} \right) \mu([n_1 + 2, \infty))\mu([n_2 + 1, \infty)) \dots \mu([n_p + 1, \infty)) \\ & = \mu(n_0)\mu([n_1 + 1, \infty))\mu([n_2 + 1, \infty)) \dots \mu([n_p + 1, \infty)) \end{aligned}$$

as desired. An immediate generalization of the preceding argument shows that, if  $\mathbf{t}_0$  and  $\mathbf{t}_{j,i}$ ,  $1 \leq j \leq p$ ,  $1 \leq i \leq n_j$  are given plane trees, the event

$$\{k_{\emptyset}(\mathcal{T}'_{-1}) = n_1, \dots, k_{\emptyset}(\mathcal{T}'_{-p}) = n_p\} \cap \{\mathcal{T}'_0 = \mathbf{t}_0\} \cap \left( \bigcap_{j=1}^p \left( \bigcap_{i=1}^{n_j} \{[\mathcal{T}'_{-j}]_i = \mathbf{t}_{j,i}\} \right) \right)$$

has probability

$$\mu([n_1 + 1, \infty))\mu([n_2 + 1, \infty)) \dots \mu([n_p + 1, \infty)) \times \Pi_{\mu}(\mathbf{t}_0) \times \prod_{j=1}^p \left( \prod_{i=1}^{n_j} \Pi_{\mu}(\mathbf{t}_{j,i}) \right).$$

This completes the proof. □

## 2.4 Random walk indexed by the infinite tree

Let  $\mathcal{T} \in \mathbb{T}$ . The definition of random walk indexed by  $\mathcal{T}$  requires some extra care because we need to specify the orientation of edges: The (oriented) edges of  $\mathcal{T}$  are all pairs  $(x, y)$  of elements of  $\mathcal{T}$  such that there exists  $j \in \mathbb{Z}_-$  such that

- either  $x = (j, u)$ ,  $y = (j, v)$ , where  $u, v \in \mathcal{T}_j$  and  $u$  is the parent of  $v$ ;
- or  $x = j - 1$ ,  $y = j$ .

See Fig. 2. We write  $\mathcal{E}(\mathcal{T})$  for the collection of all oriented edges of  $\mathcal{T}$ . The random walk indexed by  $\mathcal{T}$  is a collection  $(Z_{\mathcal{T}}(u))_{u \in \mathcal{T}}$  of random variables with values in  $\mathbb{Z}^d$ , such that  $Z_{\mathcal{T}}(0) = 0$  and the random variables  $(Z_{\mathcal{T}}(y) - Z_{\mathcal{T}}(x))_{(x,y) \in \mathcal{E}(\mathcal{T})}$  are independent and distributed according to  $\theta$ . Let  $P_{(\mathcal{T})}$  stand for the distribution of the collection  $(Z_{\mathcal{T}}(u))_{u \in \mathcal{T}}$ .

Let  $\mathbb{T}^*$  be the set of all pairs  $(\mathcal{T}, (z_u)_{u \in \mathcal{T}})$  where  $\mathcal{T} \in \mathbb{T}$  and  $z_u \in \mathbb{Z}^d$  for every  $u \in \mathcal{T}$ . We define a probability measure  $\mathbf{P}_{\mu, \theta}^*$  on  $\mathbb{T}^*$  by declaring that  $\mathbf{P}_{\mu, \theta}^*$  is the law of the random pair  $(\mathcal{T}, (\mathcal{Z}_u)_{u \in \mathcal{T}})$  where  $\mathcal{T}$  is distributed according to  $\mathbf{P}_{\mu}$  and conditionally on  $\mathcal{T} = \mathcal{T}$ ,  $(\mathcal{Z}_u)_{u \in \mathcal{T}}$  is distributed according to  $P_{(\mathcal{T})}$ .

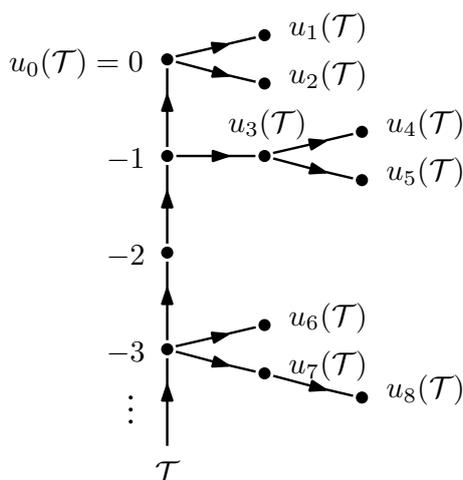
We next define a shift transformation  $\tau^*$  on  $\mathbb{T}^*$ . For  $(\mathcal{T}, (z_u)_{u \in \mathcal{T}}) \in \mathbb{T}^*$ , we set  $\tau^*(\mathcal{T}, (z_u)_{u \in \mathcal{T}}) = (\mathcal{T}', (z'_u)_{u \in \mathcal{T}'})$ , where  $\mathcal{T}' = \tau(\mathcal{T})$  and the spatial locations of vertices of  $\mathcal{T}'$  (which may be viewed as a subset of  $\mathcal{T}$ ) are obtained by shifting all original locations  $z_u$  so that the location of the root of  $\mathcal{T}'$  is again 0. More precisely, if  $k \in \mathbb{Z}_-$  is defined as above in the definition of  $\mathcal{T}' = \tau(\mathcal{T})$ , there is a unique bijection  $\phi_{\mathcal{T}}$  from  $\mathcal{T}'$  onto  $\mathcal{T} \setminus \{k+1, k+2, \dots, 0\}$  that maps 0 to  $(k, 1)$  and preserves both the lexicographical order and the genealogical order, and we set

$$z'_u = z_{\phi_{\mathcal{T}}(u)} - z_{\phi_{\mathcal{T}}(0)}$$

for every  $u \in \mathcal{T}'$ .

**Proposition 3.** *The probability measure  $\mathbf{P}_{\mu, \theta}^*$  is invariant under  $\tau^*$ .*

This is an easy consequence of Proposition 2 and the way the spatial positions are constructed. We leave the details to the reader.



**Figure 2:** The orientation of edges of  $\mathcal{T}$ , and the sequence  $u_0(\mathcal{T}), u_1(\mathcal{T}), u_2(\mathcal{T}), \dots$

Let  $\mathcal{T}^* = (\mathcal{T}, (z_u)_{u \in \mathcal{T}}) \in \mathbb{T}^*$ . We define a sequence  $(u_i(\mathcal{T}))_{i \geq 0}$  of elements of  $\mathcal{T}$  as follows. First,  $u_0(\mathcal{T}) = 0$  is the root of  $\mathcal{T}$ . Then  $u_1(\mathcal{T}), u_2(\mathcal{T}), \dots$  are all elements of  $\mathcal{T} \setminus \mathbb{Z}_-$  listed in lexicographical order (see Fig. 2). We set, for every integer  $n \geq 1$ ,

$$R_n(\mathcal{T}^*) := \#\{z_{u_0(\mathcal{T})}, z_{u_1(\mathcal{T})}, \dots, z_{u_{n-1}(\mathcal{T})}\}.$$

We let  $S = (S_k)_{k \geq 0}$  be a random walk in  $\mathbb{Z}^d$  with jump distribution  $\theta$ , which starts from  $x$  under the probability measure  $P_x$ , for every  $x \in \mathbb{Z}^d$ .

Recall the notation  $a_{\mu, \theta}$  and  $h_{\mu, \theta}$  introduced at the end of subsection 2.1.

**Theorem 4.** *We have*

$$\frac{R_n}{n} \xrightarrow[n \rightarrow \infty]{} c_{\mu, \theta}, \quad \mathbf{P}_{\mu, \theta}^* \text{-a.s.}$$

where the limiting constant  $c_{\mu, \theta} \in [0, 1]$  may be defined as

$$c_{\mu, \theta} = a_{\mu, \theta} E_0 \left[ \prod_{j=1}^{\infty} \Phi_{\mu, \theta}(-S_j) \right],$$

with

$$\Phi_{\mu, \theta}(x) = \sum_{k=0}^{\infty} \mu([k+1, \infty)) \left( \sum_{y \in \mathbb{Z}^d} \theta(y) h_{\mu, \theta}(x+y) \right)^k,$$

for every  $x \in \mathbb{Z}^d$ .

*Proof.* Set  $\tau_n^* = (\tau^*)^n$  for every integer  $n \geq 1$ . We claim that, for every  $n, m \geq 1$ ,

$$R_{n+m} \leq R_n + R_m \circ \tau_n^*.$$

Indeed,  $R_n(\mathcal{T}^*)$  is the number of distinct elements among  $z_{u_0(\mathcal{T})}, z_{u_1(\mathcal{T})}, \dots, z_{u_{n-1}(\mathcal{T})}$ , and similarly  $R_{n+m}(\mathcal{T}^*)$  is the number of distinct elements among  $z_{u_0(\mathcal{T})}, z_{u_1(\mathcal{T})}, \dots, z_{u_{n+m-1}(\mathcal{T})}$ . On the other hand, from the construction of the shift transformation, it is fairly easy to verify that  $R_m \circ \tau_n^*(\mathcal{T}^*)$  is the number of distinct elements among  $z_{u_n(\mathcal{T})}, z_{u_{n+1}(\mathcal{T})}, \dots, z_{u_{n+m-1}(\mathcal{T})}$ . The bound of the preceding display follows immediately.

Since  $1 \leq R_n \leq n$ , we can then apply Kingman's subadditive ergodic theorem to the sequence  $(R_n)_{n \geq 1}$ , and we get that  $R_n/n$  converges almost surely. The fact that the limit is constant is immediate from a simple zero-one law argument (we could also verify that  $\tau^*$  is ergodic). Furthermore, the limiting constant  $c_{\mu, \theta}$  is recovered by

$$c_{\mu, \theta} = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E}_{\mu, \theta}^*[R_n].$$

However, with the preceding notation,

$$\begin{aligned} \mathbf{E}_{\mu, \theta}^*[R_n] &= \mathbf{E}_{\mu, \theta}^* \left[ \sum_{i=0}^{n-1} \mathbf{1}_{\{z_{u_j} \neq z_{u_i}, \forall j \in \llbracket i+1, n-1 \rrbracket\}} \right] \\ &= \sum_{i=0}^{n-1} \mathbf{P}_{\mu, \theta}^*(z_{u_j} \neq z_{u_i}, \forall j \in \llbracket i+1, n-1 \rrbracket) \\ &= \sum_{i=0}^{n-1} \mathbf{P}_{\mu, \theta}^*(z_{u_j} \neq 0, \forall j \in \llbracket 1, n-i-1 \rrbracket) \end{aligned}$$

using the shift invariance in the last equality. It now follows that

$$c_{\mu,\theta} = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E}_{\mu,\theta}^*[R_n] = \mathbf{P}_{\mu,\theta}^*(z_{u_j} \neq 0, \forall j \geq 1),$$

and the right-hand side is easily computed in the form given in the theorem, using the definition of  $\mathbf{P}_{\mu,\theta}^*$ .  $\square$

Theorem 4 does not give much information when the limiting constant  $c_{\mu,\theta}$  is equal to 0. In the next proposition, we give sufficient conditions that ensure  $c_{\mu,\theta} > 0$ . We let  $g_\mu$  denote the generating function of  $\mu$ ,

$$g_\mu(r) := \sum_{k=0}^{\infty} \mu(k) r^k, \quad 0 \leq r \leq 1.$$

In the remaining part of this subsection, we assume that the random walk  $S$  is transient (it is not hard to see that  $c_{\mu,\theta} = 0$  if  $S$  is recurrent). We denote the Green function of  $S$  by  $G_\theta$ , that is

$$G_\theta(x) := E_0 \left[ \sum_{k=0}^{\infty} \mathbf{1}_{\{S_k=x\}} \right], \quad x \in \mathbb{Z}^d.$$

**Proposition 5.** (i) *The property  $c_{\mu,\theta} > 0$  holds if*

$$\prod_{j=1}^{\infty} \left( \frac{1 - g_\mu((1 - G_\theta(S_j))_+)}{G_\theta(S_j)} \right) > 0, \quad P_0\text{-a.s.}$$

(ii) *Suppose that the random walk  $S$  is centered and has finite moments of order  $(d-1) \vee 2$ . Then,*

- *if  $\mu$  has finite variance, then  $c_{\mu,\theta} > 0$  if  $d \geq 5$ .*
- *if  $\mu$  is in the domain of attraction of a stable distribution with index  $\alpha \in (1, 2)$ , then  $c_{\mu,\theta} > 0$  if  $d > \frac{2\alpha}{\alpha-1}$ .*

*Proof.* (i) We have already noticed that  $a_{\mu,\theta} > 0$ . We then observe that, for every  $r \in [0, 1)$ ,

$$\sum_{k=0}^{\infty} \mu([k+1, \infty)) r^k = \frac{1 - g_\mu(r)}{1 - r}. \quad (1)$$

Next we can get a lower bound on the function  $h_{\mu,\theta}(y)$  by saying that the probability for tree-indexed random walk to visit the point  $-y$  is bounded above by the expected value of the number of vertices at which the random walk sits at  $-y$ . Since  $\mu$  is critical, it follows that

$$h_{\mu,\theta}(y) \geq 1 - G_\theta(-y)$$

for every  $y \in \mathbb{Z}^d$ . Hence, for every  $x \in \mathbb{Z}^d$ ,

$$\sum_{y \in \mathbb{Z}^d} \theta(y) h_{\mu,\theta}(x+y) \geq 1 - \sum_{y \in \mathbb{Z}^d} \theta(y) G_\theta(-x-y).$$

However,

$$\sum_{y \in \mathbb{Z}^d} \theta(y) G_\theta(-x-y) = \sum_{y \in \mathbb{Z}^d} \theta(y) E_{x+y} \left[ \sum_{k=0}^{\infty} \mathbf{1}_{\{S_k=0\}} \right] = E_x \left[ \sum_{k=1}^{\infty} \mathbf{1}_{\{S_k=0\}} \right] \leq G_\theta(-x).$$

Consequently, using (1), we have, for all  $x$  such that  $G_\theta(-x) > 0$ ,

$$\Phi_{\mu,\theta}(x) \geq \frac{1 - g_\mu((1 - G_\theta(-x))_+)}{G_\theta(-x)}.$$

The assertion in (i) follows, noting that  $G_\theta(S_j) > 0$  for every  $j \geq 0$ ,  $P_0$ -a.s.

(ii) If  $S$  is centered with finite moments of order  $(d-1) \vee 2$ , then a standard bound for the Green function (see e.g. [15, Théorème 3.5]) gives the existence of a constant  $C_\theta$  such that, for every  $x \in \mathbb{Z}^d$ ,

$$G_\theta(x) \leq C_\theta |x|^{2-d} \quad (2)$$

(recall that we assume that  $S$  is transient, so that necessarily  $d \geq 3$  here).

Suppose first that  $\mu$  has a finite variance  $\sigma_\mu^2$ . Then,

$$g_\mu(1-s) = 1 - s + \frac{\sigma_\mu^2}{2} s^2 + o(s^2)$$

as  $s \rightarrow 0$ . Consequently,

$$\frac{1 - g_\mu(1-s)}{s} = 1 - \frac{\sigma_\mu^2}{2} s + o(s)$$

as  $s \rightarrow 0$ . By taking  $s = G_\theta(S_j)$ , we see that the condition in (i) will be satisfied if

$$\sum_{j=1}^{\infty} G_\theta(S_j) < \infty, \quad P_0\text{-a.s.}$$

However, using the local limit theorem and the preceding bound for  $G_\theta$ , it is an easy matter to verify that the property

$$E_0 \left[ \sum_{j=1}^{\infty} G_\theta(S_j) \right] < \infty$$

holds if  $d \geq 5$ . This gives the desired result when  $\mu$  has a finite variance.

Suppose now that  $\mu$  is in the domain of attraction of a stable distribution with index  $\alpha \in (1, 2)$ . Then the generating function of  $\mu$  must satisfy the property

$$g_\mu(1-s) = 1 - s + s^\alpha L(s)$$

where  $L$  is slowly varying as  $s \downarrow 0$  (see e.g. the discussion in [3, p.60]). By the same argument as above, we see that the condition in (i) will be satisfied if

$$\sum_{j=1}^{\infty} G_\theta(S_j)^{\alpha-1} L(G_\theta(S_j)) < \infty, \quad P_0\text{-a.s.}$$

and this holds if

$$(\alpha - 1)(d - 2)/2 > 1,$$

which completes the proof.  $\square$

**Remarks.** 1. The moment assumption in (ii) can be weakened a little: According to [13], the bound (2) holds provided the random walk  $S$  (is centered and) has moments of order  $(d-2+\varepsilon) \vee 2$  for some  $\varepsilon > 0$ . However moments of order  $d-2$  would not be sufficient for this bound.

2. Suppose that the random walk  $S$  satisfies the conditions in part (ii) of the proposition. If  $\mu$  has finite variance, it is not hard to verify that  $c_{\mu,\theta} = 0$  if  $d \leq 4$ . Let us briefly sketch the argument. It is enough to consider the case  $d = 4$ . Under the probability measure  $\Pi_{\mu,\theta}^*$ , write  $N_x$  for the number of vertices whose spatial location is equal to  $x$ . Then, if  $x \neq 0$ ,

$$\Pi_{\mu,\theta}^*[N_x] = G_\theta(x) \geq C'_\theta |x|^{-2}$$

for some constant  $C'_\theta > 0$ . On the other hand, standard arguments for Galton-Watson trees show that there exists a constant  $K_\mu$  such that

$$\Pi_{\mu,\theta}^*[(N_x)^2] \leq K_\mu \sum_{z \in \mathbb{Z}^4} G_\theta(z) G_\theta(x - z)^2.$$

Using (2) and simple calculations, we obtain the existence of a constant  $K'_{\mu,\theta}$  such that, for every  $x \in \mathbb{Z}^4$  with  $|x| \geq 2$ ,

$$\Pi_{\mu,\theta}^*[(N_x)^2] \leq K'_{\mu,\theta} |x|^{-2} \log |x|.$$

Hence, for every  $x \in \mathbb{Z}^4$  with  $|x| \geq 2$ ,

$$1 - h_{\mu,\theta}(x) = \Pi_{\mu,\theta}^*(N_{-x} \geq 1) \geq \frac{(\Pi_{\mu,\theta}^*[N_{-x}])^2}{\Pi_{\mu,\theta}^*[(N_{-x})^2]} \geq (C'_\theta)^2 (K'_{\mu,\theta})^{-1} |x|^{-2} (\log |x|)^{-1}.$$

The property  $c_{\mu,\theta} = 0$  now follows easily. In the next section, we will see (in a particular case) that the proper normalization factor for  $R_n$  is  $(\log n)/n$  when  $d = 4$ .

3. The paper [20] considers certain random walks with drift on  $\mathbb{Z}^d$ , for which the property  $c_{\mu,\theta} > 0$  holds if  $d \geq 4$  although the bound (2) no longer holds.

4. It is an interesting question whether the condition  $d > \frac{2\alpha}{\alpha-1}$  is also sharp when  $\mu$  is in the domain of attraction of a stable distribution of index  $\alpha$ . We will not discuss this problem here as our main interest lies in the case when  $\mu$  has finite variance.

## 2.5 Conditioned trees

Our goal is now to obtain an analog of the convergence of Theorem 4 for random walk indexed by a single Galton-Watson tree conditioned to be large. Recall from subsection 2.1 the notation  $\mathbb{T}_f^*$  for the set of all spatial trees. If  $\mathcal{T}^* = (\mathcal{T}, (z_u)_{u \in \mathcal{T}})$  is a spatial tree with at least  $n$  vertices, we keep the same notation  $R_n(\mathcal{T}^*)$  for the number of distinct points in the sequence  $z_{u_0}, z_{u_1}, \dots, z_{u_{n-1}}$ , where  $u_0, u_1, \dots, u_{n-1}$  are the vertices of  $\mathcal{T}$  listed in lexicographical order. Also recall from subsection 2.1 the definition of the probability measure  $\Pi_{\mu,\theta}^*$  on  $\mathbb{T}_f^*$ .

**Proposition 6.** *Assume that  $\mu$  has finite variance  $\sigma_\mu^2$ , or that  $\mu$  is in the domain of attraction of a stable distribution with index  $\alpha \in (1, 2)$ . For every  $n \geq 1$ , let  $\mathcal{T}_{(>n)}^*$  be a random spatial tree distributed according to the probability measure  $\Pi_{\mu,\theta}^*(\cdot \mid \#\mathcal{T} > n)$ . Then, for every  $a \in (0, 1]$ ,*

$$\frac{1}{n} R_{[an]}(\mathcal{T}_{(>n)}^*) \xrightarrow[n \rightarrow \infty]{} c_{\mu,\theta} a$$

*in probability.*

*Proof.* We first consider the case when  $\mu$  has finite variance  $\sigma_\mu^2$ . Let  $\mathcal{T}^* = (\mathcal{T}, (\mathcal{L}_u)_{u \in \mathcal{T}})$  be a  $\mathbb{T}^*$ -valued random variable distributed according to  $\mathbf{P}_{\mu, \theta}^*$  under the probability measure  $P$ . Recall the notation  $\mathcal{T}_j$ , for  $j \in \mathbb{Z}_-$ , introduced in subsection 2.2. By construction, the “subtrees”

$$\mathcal{T}_0, [\mathcal{T}_{-1}]_1, [\mathcal{T}_{-1}]_2, \dots, [\mathcal{T}_{-1}]_{k_\emptyset(\mathcal{T}_{-1})}, [\mathcal{T}_{-2}]_1, [\mathcal{T}_{-2}]_2, \dots, [\mathcal{T}_{-2}]_{k_\emptyset(\mathcal{T}_{-2})}, [\mathcal{T}_{-3}]_1, \dots$$

then form an infinite sequence of independent random trees distributed according to  $\Pi_\mu$ . To simplify notation we denote this sequence by  $\mathcal{T}_{(0)}, \mathcal{T}_{(1)}, \mathcal{T}_{(2)}, \dots$ . We then introduce the height process  $(H_k)_{k \geq 0}$  associated with this sequence of trees (see [16, Section 1]). This means that, for every  $j \geq 0$ , we first enumerate the vertices of  $\mathcal{T}_{(j)}$  in lexicographical order, then we concatenate the finite sequences obtained in this way to get an infinite sequence  $(v_k)_{k \geq 0}$  of elements in  $\mathcal{U}$ , and we finally set  $H_k := |v_k|$  for every  $k \geq 0$ . Note that the infinite sequence of vertices  $(v_k)_{k \geq 0}$  thus obtained is essentially the same as the sequence  $(u_k(\mathcal{T}))_{k \geq 0}$  introduced in subsection 2.4.

Then (see e.g. [16, Theorem 1.8]), we have the convergence in distribution

$$\left( \frac{1}{\sqrt{n}} H_{[nt]} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{(d)} \left( \frac{2}{\sigma_\mu} |\beta_t| \right)_{t \geq 0}, \quad (3)$$

where  $(\beta_t)_{t \geq 0}$  denotes a standard linear Brownian motion. Next, for every integer  $n \geq 1$ , set

$$k_n := \inf \{ k \geq 0 : \#\mathcal{T}_{(k)} > n \}.$$

Clearly, the tree  $\mathcal{T}_{(k_n)}$  is distributed according to  $\Pi_\mu(\cdot \mid \#\mathcal{T} > n)$ . Also set

$$d_{k_n} := \sum_{0 \leq j < k_n} \#\mathcal{T}_{(j)}.$$

Using the convergence (3), it is not hard to prove (see e.g. the proof of Theorem 5.1 in [17]) that

$$\frac{1}{n} d_{k_n} \xrightarrow[n \rightarrow \infty]{(d)} D_1, \quad (4)$$

where  $D_1$  denotes the initial time of the first excursion of  $\beta$  away from 0 with duration greater than 1.

By Theorem 4 and an obvious monotonicity argument, we have for every integer  $K > 0$ ,

$$\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq K} \left| \frac{1}{n} R_{[nt]}(\mathcal{T}^*) - c_{\mu, \theta} t \right| = 0, \quad \text{a.s.}$$

and it follows that

$$\lim_{n \rightarrow \infty} \left| \frac{1}{n} R_{(d_{k_n} + [an]) \wedge Kn}(\mathcal{T}^*) - \frac{1}{n} R_{d_{k_n} \wedge Kn}(\mathcal{T}^*) - c_{\mu, \theta} \left( \left( \frac{d_{k_n}}{n} + a \right) \wedge K - \frac{d_{k_n}}{n} \wedge K \right) \right| = 0, \quad \text{a.s.}$$

Since  $K$  can be chosen arbitrarily large, we deduce from the last convergence and (4) that we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left( R_{d_{k_n} + [an]}(\mathcal{T}^*) - R_{d_{k_n}}(\mathcal{T}^*) \right) = c_{\mu, \theta} a$$

in probability.

Let  $\mathcal{T}_{(k_n)}^*$  stand for the spatial tree obtained from  $\mathcal{T}_{(k_n)}$  by keeping the spatial positions induced by  $\mathcal{T}^*$ . Then, by construction, we have

$$R_{[an]}(\mathcal{T}_{(k_n)}^*) \geq R_{d_{k_n} + [an]}(\mathcal{T}^*) - R_{d_{k_n}}(\mathcal{T}^*).$$

Therefore, using the preceding convergence in probability, we obtain that, for every fixed  $\varepsilon > 0$ ,

$$P\left(R_{\lfloor an \rfloor}(\mathcal{T}_{(k_n)}^*) \geq (c_{\mu,\theta}a - \varepsilon)n\right) \xrightarrow{n \rightarrow \infty} 1. \quad (5)$$

We claim that we have also

$$P\left(R_{\lfloor an \rfloor}(\mathcal{T}_{(k_n)}^*) \leq (c_{\mu,\theta}a + \varepsilon)n\right) \xrightarrow{n \rightarrow \infty} 1. \quad (6)$$

To see this, we argue by contradiction and suppose that for all  $n$  belonging to a sequence  $(n_j)_{j \geq 1}$  converging to infinity, we have

$$P\left(R_{\lfloor an \rfloor}(\mathcal{T}_{(k_n)}^*) > (c_{\mu,\theta}a + \varepsilon)n\right) \geq \delta$$

for some  $\delta > 0$  independent of  $n$ . We suppose that  $c_{\mu,\theta} > 0$  (the case when  $c_{\mu,\theta} = 0$  is easier). We observe that, for every fixed  $n$ , the tree  $\mathcal{T}_{(k_n)}$  and the quantity  $R_{\lfloor an \rfloor}(\mathcal{T}_{(k_n)}^*)$  are independent of the random variable  $d_{k_n}$ . Notice that  $\mathcal{T}_{(k_n)}^*$  is *not* independent of  $d_{k_n}$ , because the value of  $d_{k_n}$  clearly influences the distribution of the spatial location of the root of  $\mathcal{T}_{(k_n)}$ . However, if we simultaneously translate all spatial locations of  $\mathcal{T}_{(k_n)}^*$  so that the new location of the root is 0, the new locations become independent of  $d_{k_n}$ , and the translation does not affect  $R_{\lfloor an \rfloor}(\mathcal{T}_{(k_n)}^*)$ . On the other hand, from the convergence in distribution (4), we can find  $\delta' > 0$  such that, for every sufficiently large  $n$ ,

$$P\left(d_{k_n} \leq \frac{\varepsilon}{2c_{\mu,\theta}} n\right) \geq \delta'.$$

Using the preceding independence property, we conclude that, for every sufficiently large  $n$  in the sequence  $(n_j)_{j \geq 1}$ ,

$$P\left(R_{\lfloor (\varepsilon n / 2c_{\mu,\theta}) + an \rfloor}(\mathcal{T}^*) \geq (c_{\mu,\theta}a + \varepsilon)n\right) \geq P\left(d_{k_n} \leq \frac{\varepsilon}{2c_{\mu,\theta}} n\right) P\left(R_{\lfloor an \rfloor}(\mathcal{T}_{(k_n)}^*) > (c_{\mu,\theta}a + \varepsilon)n\right) \geq \delta\delta'.$$

However Theorem 4 implies that

$$\frac{1}{n} R_{\lfloor (\varepsilon n / 2c_{\mu,\theta}) + an \rfloor}(\mathcal{T}^*) \xrightarrow{n \rightarrow \infty} c_{\mu,\theta} a + \frac{\varepsilon}{2}, \quad \text{a.s.}$$

and so we arrive at a contradiction, which completes the proof of (6).

By construction, the tree  $\mathcal{T}_{(k_n)}$  is distributed according to  $\Pi_\mu(\cdot \mid \#\mathcal{T} > n)$ , and if we shift all spatial locations of  $\mathcal{T}_{(k_n)}^*$  so that the new location of the root is 0, we get a random spatial tree distributed according to  $\Pi_{\mu,\theta}^*(\cdot \mid \#\mathcal{T} > n)$ . The convergence of the proposition thus follows from (5) and (6).

The proof in the case when  $\mu$  is in the domain of attraction of a stable distribution with index  $\alpha \in (1, 2)$  is essentially the same, noting that Theorems 2.3.1 and 2.3.2 in [3] give an analog of the convergence (3), where the role of reflected Brownian motion is played by the so-called height process associated with the stable Lévy process with index  $\alpha$ . We omit the details.  $\square$

We now would like to get a statement analogous to Proposition 6 for a tree conditioned to have a fixed number of vertices. This will follow from Proposition 6 by an absolute continuity argument. Before stating the result, we need to introduce some notation. Let  $\mathcal{G}$  be the smallest subgroup of  $\mathbb{Z}$  that contains the support of  $\mu$ . Plainly, the cardinality of the vertex set of a tree distributed according to  $\Pi_\mu$  belongs to  $1 + \mathcal{G}$ . On the other hand, for every sufficiently large

integer  $p \in 1 + \mathcal{G}$ , we have  $\Pi_\mu(\#\mathcal{T} = p) > 0$ , so that the definition of  $\Pi_\mu(\cdot \mid \#\mathcal{T} = p)$  makes sense.

If  $\mathcal{T}^* = (\mathcal{T}, (z_u)_{u \in \mathcal{T}})$  is a spatial tree, we write  $\mathcal{R}(\mathcal{T}^*)$  for the number of distinct elements in  $\{z_u : u \in \mathcal{T}\}$ .

**Theorem 7.** *Assume that  $\mu$  has finite variance  $\sigma_\mu^2$ , or that  $\mu$  is in the domain of attraction of a stable distribution with index  $\alpha \in (1, 2)$ . For every sufficiently large integer  $n \in \mathcal{G}$ , let  $\mathcal{T}_{(n)}^*$  be a random spatial tree distributed according to the probability measure  $\Pi_{\mu, \theta}^*(\cdot \mid \#\mathcal{T} = n + 1)$ . Then,*

$$\frac{1}{n} \mathcal{R}(\mathcal{T}_{(n)}^*) \xrightarrow{n \rightarrow \infty, n \in \mathcal{G}} c_{\mu, \theta}$$

in probability.

*Proof.* We assume in the proof that  $\mathcal{G} = \mathbb{Z}$ . Only minor modifications are needed to deal with the general case.

We first consider the case when  $\mu$  has finite variance  $\sigma_\mu^2$ . The arguments needed to derive Theorem 7 from Proposition 6 are then similar to the proof of Theorem 6.1 in [17]. The basic idea is as follows. For every  $a \in (0, 1)$ , the law under  $\Pi_\mu(\cdot \mid \#\mathcal{T} = n + 1)$  of the subtree obtained by keeping only the first  $\lfloor an \rfloor$  vertices of  $\mathcal{T}$  is absolutely continuous with respect to the law under  $\Pi_\mu(\cdot \mid \#\mathcal{T} > n)$  of the same subtree, with a density that is bounded independently of  $n$ . A similar property holds for spatial trees, and so we can use the convergence of Proposition 6, for a tree distributed according to  $\Pi_{\mu, \theta}^*(\cdot \mid \#\mathcal{T} > n)$ , to get a similar convergence for a tree distributed according to  $\Pi_{\mu, \theta}^*(\cdot \mid \#\mathcal{T} = n + 1)$ . Let us give some details for the sake of completeness.

As previously, we write  $u_0(\mathcal{T}), u_1(\mathcal{T}), \dots, u_{\#\mathcal{T}-1}(\mathcal{T})$  for the vertices of a plane tree  $\mathcal{T}$  listed in lexicographical order. The Lukasiewicz path of  $\mathcal{T}$  is then the finite sequence  $(X_\ell(\mathcal{T}), 0 \leq \ell \leq \#\mathcal{T})$ , which is defined inductively by

$$X_0(\mathcal{T}) = 0, \quad X_{\ell+1}(\mathcal{T}) - X_\ell(\mathcal{T}) = k_{u_\ell(\mathcal{T})}(\mathcal{T}) - 1, \quad \text{for every } 0 \leq \ell < \#\mathcal{T},$$

where we recall that, for every  $u \in \mathcal{T}$ ,  $k_u(\mathcal{T})$  is the number of children of  $u$  in  $\mathcal{T}$ . The tree  $\mathcal{T}$  is determined by its Lukasiewicz path. A key result (see e.g. [16, Section 1]) states that under  $\Pi_\mu(d\mathcal{T})$ , the Lukasiewicz path is distributed as a random walk on  $\mathbb{Z}$  with jump distribution  $\nu$  determined by  $\nu(j) = \mu(j + 1)$  for every  $j \geq -1$ , which starts from 0 and is stopped at the first time when it hits  $-1$  (in particular, the law of  $\#\mathcal{T}$  under  $\Pi_\mu(d\mathcal{T})$  coincides with the law of the latter hitting time). For notational convenience, we let  $(Y_k)_{k \geq 0}$  be a random walk on  $\mathbb{Z}$  with jump distribution  $\nu$ , which starts from  $j$  under the probability measure  $P_{(j)}$ , and we set

$$T := \inf\{k \geq 0 : Y_k = -1\}.$$

Next take  $n$  large enough so that  $\Pi_\mu(\#\mathcal{T} = n + 1) > 0$ . Fix  $a \in (0, 1)$ , and consider a tree  $\mathcal{T}$  such that  $\#\mathcal{T} > n$ . Then, the collection of vertices  $u_0(\mathcal{T}), \dots, u_{\lfloor an \rfloor}(\mathcal{T})$  forms a subtree of  $\mathcal{T}$  (because in the lexicographical order the parent of a vertex comes before this vertex), and we denote this subtree by  $\rho_{\lfloor an \rfloor}(\mathcal{T})$ . It is elementary to verify that  $\rho_{\lfloor an \rfloor}(\mathcal{T})$  is determined by the sequence  $(X_\ell(\mathcal{T}), 0 \leq \ell \leq \lfloor an \rfloor)$ . Let  $f$  be a bounded function on  $\mathbb{Z}^{\lfloor an \rfloor + 1}$ . Using the Markov property at time  $\lfloor an \rfloor$  for the random walk with jump distribution  $\nu$ , one verifies that

$$\begin{aligned} & \Pi_\mu \left[ f((X_k)_{0 \leq k \leq \lfloor an \rfloor}) \mid \#\mathcal{T} = n + 1 \right] \\ &= \frac{P_{(0)}(T > n)}{P_{(0)}(T = n + 1)} \Pi_\mu \left[ f((X_k)_{0 \leq k \leq \lfloor an \rfloor}) \frac{\psi_n(X_{\lfloor an \rfloor})}{\psi'_n(X_{\lfloor an \rfloor})} \mid \#\mathcal{T} > n \right] \end{aligned} \quad (7)$$

where, for every integer  $j \geq 0$ ,

$$\psi_n(j) = P_{(j)}(T = n + 1 - \lfloor an \rfloor), \quad \psi'_n(j) = P_{(j)}(T > n - \lfloor an \rfloor).$$

See [17, pp.742-743] for details of the derivation of (7). We now let  $n$  tend to infinity. Using Kemperman's formula (see e.g. Pitman [24, p.122]) and a standard local limit theorem, one easily checks that, for every  $c > 0$ ,

$$\lim_{n \rightarrow \infty} \left( \sup_{j \geq c\sqrt{n}} \left| \frac{P_{(0)}(T > n)}{P_{(0)}(T = n + 1)} \frac{\psi_n(j)}{\psi'_n(j)} - \Gamma_a\left(\frac{j}{\sigma_\mu \sqrt{n}}\right) \right| \right) = 0, \quad (8)$$

where for every  $x \geq 0$ ,

$$\Gamma_a(x) = \frac{2(2\pi(1-a)^3)^{-1/2} \exp(-x^2/2(1-a))}{\int_{1-a}^{\infty} ds (2\pi s^3)^{-1/2} \exp(-x^2/2s)}.$$

See again [17, pp.742-743] for details. Note that the function  $\Gamma_a$  is bounded over  $\mathbb{R}_+$ . Furthermore, from the local limit theorem again, it is easy to verify that

$$\lim_{c \downarrow 0} \limsup_{n \rightarrow \infty} \Pi_\mu(X_{\lfloor an \rfloor} \leq c\sqrt{n} \mid \#\mathcal{T} = n + 1) = 0, \quad \lim_{c \downarrow 0} \limsup_{n \rightarrow \infty} \Pi_\mu(X_{\lfloor an \rfloor} \leq c\sqrt{n} \mid \#\mathcal{T} > n) = 0. \quad (9)$$

(We take this opportunity to point out that the analogous statement in [17, p.743] is written incorrectly.) By combining (7), (8) and (9), we obtain that, for any uniformly bounded sequence of functions  $(f_n)_{n \geq 1}$  on  $\mathbb{Z}^{\lfloor an \rfloor + 1}$ , we have

$$\lim_{n \rightarrow \infty} \left| \Pi_\mu \left[ f_n((X_k)_{0 \leq k \leq \lfloor an \rfloor}) \mid \#\mathcal{T} = n + 1 \right] - \Pi_\mu \left[ f_n((X_k)_{0 \leq k \leq \lfloor an \rfloor}) \Gamma_a\left(\frac{X_{\lfloor an \rfloor}}{\sigma_\mu \sqrt{n}}\right) \mid \#\mathcal{T} > n \right] \right| = 0. \quad (10)$$

This convergence applies in particular to the case when, for every  $n$ ,  $f_n((X_k)_{0 \leq k \leq \lfloor an \rfloor})$  is a function of the tree  $\rho_{\lfloor an \rfloor}(\mathcal{T})$ . If we now replace  $\Pi_\mu$  by  $\Pi_{\mu, \theta}^*$ , the same convergence still holds, and we can even allow the function of the tree  $\rho_{\lfloor an \rfloor}(\mathcal{T})$  to depend also on the spatial locations of the vertices of  $\rho_{\lfloor an \rfloor}(\mathcal{T})$  (the point is that the conditional distribution of these spatial locations given the tree  $\mathcal{T}$  only depends on the subtree  $\rho_{\lfloor an \rfloor}(\mathcal{T})$ ). Consequently, if  $\varepsilon > 0$  is fixed, we have

$$\lim_{n \rightarrow \infty} \left| \Pi_{\mu, \theta}^* \left[ \mathbf{1}_{\{|R_{\lfloor an \rfloor} - c_{\mu, \theta} an| > \varepsilon n\}} \mid \#\mathcal{T} = n + 1 \right] - \Pi_{\mu, \theta}^* \left[ \mathbf{1}_{\{|R_{\lfloor an \rfloor} - c_{\mu, \theta} an| > \varepsilon n\}} \Gamma_a\left(\frac{X_{\lfloor an \rfloor}}{\sigma_\mu \sqrt{n}}\right) \mid \#\mathcal{T} > n \right] \right| = 0.$$

Recalling that the function  $\Gamma_a$  is bounded, and using Proposition 6, we now obtain that

$$\lim_{n \rightarrow \infty} \Pi_{\mu, \theta}^* \left( |R_{\lfloor an \rfloor} - c_{\mu, \theta} an| > \varepsilon n \mid \#\mathcal{T} = n + 1 \right) = 0.$$

Since  $0 \leq \mathcal{R}(\mathcal{T}^*) - R_{\lfloor an \rfloor}(\mathcal{T}^*) \leq n + 1 - \lfloor an \rfloor$ ,  $\Pi_{\mu, \theta}^*(\cdot \mid \#\mathcal{T} = n + 1)$ -a.s., and  $a$  can be chosen arbitrarily close to 1, the convergence in Theorem 7 follows.

Very similar arguments can be used in the case when  $\mu$  is in the domain of attraction of a stable distribution with index  $\alpha \in (1, 2)$ . We now refer to the proof of Lemma 3.3 in [11] for the exact analogs of the properties (7) – (10) used in the finite variance case. We leave the details to the reader.  $\square$

The case  $d \geq 5$  of Theorem 1 follows from Theorem 7 and Proposition 5, noting that when  $\mu$  is the critical geometric distribution, a tree distributed according to  $\Pi_\mu(\cdot \mid \#\mathcal{T} = n)$  is uniformly distributed over the set of all plane trees with  $n$  vertices (see e.g. [16, Section 1.5]).

### 3 The critical dimension

In this section, we discuss the dimension  $d = 4$ , which is critical in the case of random walks that are centered and have sufficiently high moments. We restrict our attention to the case when the offspring distribution is geometric with parameter  $1/2$ . Our main tool is the discrete snake, which is a path-valued Markov chain that can be used to generate the spatial positions of the tree-indexed random walk.

#### 3.1 Limit theorems

We now let  $\theta$  be a symmetric probability distribution on  $\mathbb{Z}^4$ . We assume that  $\theta$  has small exponential moments and is not supported on a strict subgroup of  $\mathbb{Z}^4$ . As previously, we write  $S = (S_k)_{k \geq 0}$  for the random walk in  $\mathbb{Z}^4$  with jump distribution  $\theta$ , and we now assume that  $S$  starts from 0 under the probability measure  $P$ . We will also assume for simplicity that the covariance matrix  $M_\theta$  of  $\theta$  is of the form  $\sigma^2 \text{Id}$ , where  $\text{Id}$  is the four-dimensional identity matrix and  $\sigma > 0$ . This isotropy condition can be removed, and the reader will easily check that all subsequent arguments remain valid for a non-isotropic random walk: the role of  $\sigma^2$  is then played by  $(\det M_\theta)^{1/4}$ .

We first introduce the free discrete snake associated with  $\theta$ . This is a Markov chain with values in the space  $\mathcal{W}$  that we now define. The space  $\mathcal{W}$  is the set of all semi-infinite discrete paths  $w = (w(k))_{k \in \llbracket -\infty, \zeta \rrbracket}$  with values in  $\mathbb{Z}^4$ . Here  $\zeta = \zeta(w) \in \mathbb{Z}$  is called the lifetime of  $w$ . We often write  $\widehat{w} = w(\zeta(w))$  for the endpoint of  $w$ .

If  $w \in \mathcal{W}$ , we let  $\bar{w}$  stand for the new path obtained by “erasing” the endpoint of  $w$ , namely  $\zeta(\bar{w}) = \zeta(w) - 1$  and  $\bar{w}(k) = w(k)$  for every  $k \in \llbracket -\infty, \zeta(w) - 1 \rrbracket$ . If  $x \in \mathbb{Z}^4$ , we let  $w \oplus x$  be the path obtained from  $w$  by “adding” the point  $x$  to  $w$ , namely  $\zeta(w \oplus x) = \zeta(w) + 1$ ,  $(w \oplus x)(k) = w(k)$  for every  $k \in \llbracket -\infty, \zeta(w) \rrbracket$  and  $(w \oplus x)(\zeta(w) + 1) = x$ .

The free discrete snake is the Markov chain  $(W_n)_{n \geq 0}$  in  $\mathcal{W}$  whose transition kernel is defined by

$$Q(w, dw') = \frac{1}{2} \delta_{\bar{w}}(dw') + \frac{1}{2} \sum_{x \in \mathbb{Z}^4} \theta(x) \delta_{w \oplus (\widehat{w} + x)}(dw').$$

We will write  $\zeta_n = \zeta(W_n)$  to simplify notation. It will also be convenient to write  $W_n^*$  for the path  $W_n$  shifted so that its endpoint is 0:  $W_n^*(k) = W_n(k) - \widehat{W}_n$  for every  $k \in \llbracket -\infty, \zeta_n \rrbracket$ .

If  $w \in \mathcal{W}$ ,  $\mathbb{P}_{(w)}$  will denote the probability measure under which the discrete snake  $W$  starts from  $w$ . For every integer  $m \in \mathbb{Z}$ , we also write  $\mathbb{P}_m$  for a probability measure under which  $\zeta_0 = m$  a.s. and the initial value  $W_0$  of the discrete snake is distributed as  $(-S_{m-k})_{k \in \llbracket -\infty, m \rrbracket}$  (since  $S$  is symmetric we could omit the minus sign here). We write  $\mathbb{P}$  for  $\mathbb{P}_0$ . As usual, the expectation under  $\mathbb{P}_m$ , resp. under  $\mathbb{P}$ , is denoted by  $\mathbb{E}_m$ , resp. by  $\mathbb{E}$ . Note that  $(\zeta_n)_{n \geq 0}$  is a simple random walk on  $\mathbb{Z}$  started from  $m$  under  $\mathbb{P}_m$ . We will use the notation

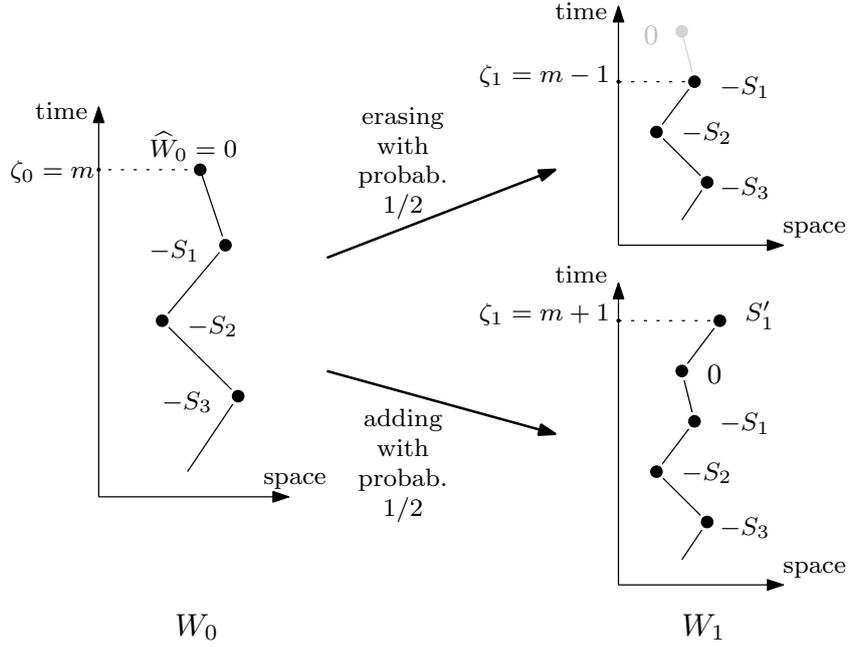
$$\tau_p := \inf\{n \geq 0 : \zeta_n = \zeta_0 - p\}$$

for every integer  $p \geq 0$ .

Furthermore, from the form of the transition kernel of the discrete snake, it is easy to verify that for every  $n \geq 0$ , for every integer  $\ell \in \mathbb{Z}$  such that  $\mathbb{P}_m(\zeta_n = \ell) > 0$ , the conditional distribution of  $W_n^*$  under  $\mathbb{P}_m(\cdot \mid \zeta_n = \ell)$ , coincides with the distribution of  $W_0$  under  $\mathbb{P}_\ell$ .

**Proposition 8.** *We have*

$$\lim_{n \rightarrow \infty} (\log n) \mathbb{P}(\widehat{W}_k \neq 0, \forall k \in \llbracket 1, n \rrbracket) = 4\pi^2 \sigma^4.$$



**Figure 3:** The discrete snake under  $\mathbb{P}_m$ . In this illustration,  $S'_1$  is an independent copy of  $S_1$ .

Furthermore,

$$\lim_{p \rightarrow \infty} (\log p) \mathbb{P}(\widehat{W}_k \neq 0, \forall k \in \llbracket 1, \tau_p \rrbracket) = 2\pi^2 \sigma^4.$$

The proof of Proposition 8 is given in subsection 3.2 below. Our first theorem is concerned with the range of the free snake.

**Theorem 9.** Set  $R_n := \#\{\widehat{W}_0, \widehat{W}_1, \dots, \widehat{W}_n\}$  for every integer  $n \geq 0$ . We have

$$\frac{\log n}{n} R_n \xrightarrow[n \rightarrow \infty]{L^2(\mathbb{P})} 4\pi^2 \sigma^4.$$

*Proof.* We first observe that

$$\mathbb{E}[R_n] = \mathbb{E}\left[\sum_{i=0}^n \mathbf{1}_{\{\widehat{W}_j \neq \widehat{W}_i, \forall j \in \llbracket i+1, n \rrbracket\}}\right] = \sum_{i=0}^n \mathbb{P}(\widehat{W}_j \neq \widehat{W}_i, \forall j \in \llbracket i+1, n \rrbracket).$$

Then, by applying the Markov property of the free snake, we have

$$\begin{aligned} \mathbb{E}[R_n] &= \sum_{i=0}^n \mathbb{E}\left[\mathbb{P}_{(W_i)}(\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, n-i \rrbracket)\right] \\ &= \sum_{i=0}^n \mathbb{E}\left[\mathbb{P}_{(W_i^*)}(\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, n-i \rrbracket)\right] \\ &= \sum_{i=0}^n \mathbb{P}(\widehat{W}_j \neq 0, \forall j \in \llbracket 1, n-i \rrbracket), \end{aligned}$$

where the second equality is easy by translation invariance, and the last one is a simple consequence of the remark before the statement of Proposition 8. Using now the result of Proposition 8, we get

$$\lim_{n \rightarrow \infty} \frac{\log n}{n} \mathbb{E}[R_n] = 4\pi^2 \sigma^4. \quad (11)$$

Let us turn to the second moment. We have similarly

$$\begin{aligned} \mathbb{E}[(R_n)^2] &= \mathbb{E} \left[ \sum_{i=0}^n \sum_{j=0}^n \mathbf{1}_{\{\widehat{W}_k \neq \widehat{W}_i, \forall k \in \llbracket i+1, n \rrbracket; \widehat{W}_\ell \neq \widehat{W}_j, \forall \ell \in \llbracket j+1, n \rrbracket\}} \right] \\ &= 2 \sum_{0 \leq i < j \leq n} \mathbb{P} \left( \widehat{W}_k \neq \widehat{W}_i, \forall k \in \llbracket i+1, n \rrbracket; \widehat{W}_\ell \neq \widehat{W}_j, \forall \ell \in \llbracket j+1, n \rrbracket \right) + \mathbb{E}[R_n] \\ &= 2 \sum_{0 \leq i < j \leq n} \mathbb{E} \left[ \mathbb{P}_{(W_i)}(\widehat{W}_k \neq \widehat{W}_0, \forall k \in \llbracket 1, n-i \rrbracket; \widehat{W}_\ell \neq \widehat{W}_{j-i}, \forall \ell \in \llbracket j-i+1, n-i \rrbracket) \right] \\ &\quad + \mathbb{E}[R_n] \\ &= 2 \sum_{0 \leq i < j \leq n} \mathbb{P} \left( \widehat{W}_k \neq 0, \forall k \in \llbracket 1, n-i \rrbracket; \widehat{W}_\ell \neq \widehat{W}_{j-i}, \forall \ell \in \llbracket j-i+1, n-i \rrbracket \right) + \mathbb{E}[R_n], \end{aligned}$$

where the last equality again follows from the observation preceding Proposition 8. Let us fix  $\alpha \in (0, 1/4)$  and define

$$\sigma_n := \inf \{k \geq 0 : \zeta_k \leq -n^{\frac{1}{2}-\alpha}\}.$$

By standard estimates, we have

$$\lim_{n \rightarrow \infty} (\log n)^2 \mathbb{P}(\sigma_n \leq n^{1-3\alpha} \text{ or } \sigma_n \geq n^{1-\alpha}) = 0.$$

Thus, using also (11),

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left( \frac{\log n}{n} \right)^2 \mathbb{E}[(R_n)^2] &= \limsup_{n \rightarrow \infty} 2 \left( \frac{\log n}{n} \right)^2 \sum_{0 \leq i < j \leq n} \mathbb{P} \left( \widehat{W}_k \neq 0, \forall k \in \llbracket 1, n-i \rrbracket; \right. \\ &\quad \left. \widehat{W}_\ell \neq \widehat{W}_{j-i}, \forall \ell \in \llbracket j-i+1, n-i \rrbracket; n^{1-3\alpha} \leq \sigma_n \leq n^{1-\alpha} \right). \end{aligned}$$

Clearly, in order to study the limsup in the right-hand side, we may restrict the sum to indices  $i$  and  $j$  such that  $j-i > n^{1-\alpha}$ . However, if  $0 \leq i < j \leq n$  are fixed such that  $j-i > n^{1-\alpha}$ ,

$$\begin{aligned} &\mathbb{P} \left( \widehat{W}_k \neq 0, \forall k \in \llbracket 1, n-i \rrbracket; \widehat{W}_\ell \neq \widehat{W}_{j-i}, \forall \ell \in \llbracket j-i+1, n-i \rrbracket; n^{1-3\alpha} \leq \sigma_n \leq n^{1-\alpha} \right) \\ &\leq \mathbb{P} \left( \widehat{W}_k \neq 0, \forall k \in \llbracket 1, \sigma_n \rrbracket; \widehat{W}_\ell \neq \widehat{W}_{j-i}, \forall \ell \in \llbracket j-i+1, n-i \rrbracket; n^{1-3\alpha} \leq \sigma_n \leq n^{1-\alpha} \right) \\ &= \mathbb{P} \left( \widehat{W}_k \neq 0, \forall k \in \llbracket 1, \sigma_n \rrbracket; n^{1-3\alpha} \leq \sigma_n \leq n^{1-\alpha} \right) \mathbb{P} \left( \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n-j \rrbracket \right). \end{aligned}$$

To derive the last equality, we use the strong Markov property at time  $\sigma_n$  and then, after conditioning on  $\sigma_n = m$ , the Markov property at time  $j-i-m$  for the free snake shifted at time  $\sigma_n$  and the observation preceding Proposition 8. Now obviously,

$$\mathbb{P} \left( \widehat{W}_k \neq 0, \forall k \in \llbracket 1, \sigma_n \rrbracket; n^{1-3\alpha} \leq \sigma_n \leq n^{1-\alpha} \right) \leq \mathbb{P} \left( \widehat{W}_k \neq 0, \forall k \in \llbracket 1, \lfloor n^{1-3\alpha} \rfloor \rrbracket \right),$$

and it follows that

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \left( \frac{\log n}{n} \right)^2 \mathbb{E}[(R_n)^2] \\
& \leq \limsup_{n \rightarrow \infty} 2 \left( \frac{\log n}{n} \right)^2 \sum_{\substack{0 \leq i < j \leq n \\ j-i > n^{1-\alpha}}} \mathbb{P}(\widehat{W}_k \neq 0, \forall k \in \llbracket 1, \lfloor n^{1-3\alpha} \rfloor \rrbracket) \mathbb{P}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n-j \rrbracket) \\
& = \frac{1}{1-3\alpha} (4\pi^2 \sigma^4)^2
\end{aligned}$$

by Proposition 8. Since  $\alpha$  can be chosen arbitrarily small, we get

$$\limsup_{n \rightarrow \infty} \left( \frac{\log n}{n} \right)^2 \mathbb{E}[(R_n)^2] \leq (4\pi^2 \sigma^4)^2. \quad (12)$$

Theorem 9 is an immediate consequence of (11) and (12).  $\square$

We now aim to prove a result similar to Theorem 9 for the “excursion” of the discrete snake. We set

$$T := \inf\{k \geq 0 : \zeta_k = -1\}.$$

For every integer  $n \geq 1$ , we let  $W^{(n)} = (W_k^{(n)})_{0 \leq k \leq 2n}$  be a process defined under  $\mathbb{P}$ , whose distribution coincides with the conditional distribution of  $(W_k)_{0 \leq k \leq 2n}$  knowing that  $T = 2n+1$ . To simplify notation, we write  $\zeta_k^{(n)} = \zeta(W_k^{(n)})$ . Note that  $(\zeta_k^{(n)})_{0 \leq k \leq 2n}$  is the contour function, also called depth-first walk, of a Galton-Watson tree with geometric offspring distribution of parameter  $1/2$ , conditioned to have  $n+1$  vertices (see e.g. [24, Chapter 6]). We have already noticed that the latter tree is uniformly distributed over plane trees with  $n+1$  vertices. From the form of the transition mechanism of the discrete snake, it then follows that  $\{\widehat{W}_k^{(n)}, 0 \leq k \leq 2n\}$  is distributed as the set of all spatial locations of a random walk with jump distribution  $\theta$  indexed by a uniform random plane tree with  $n+1$  vertices.

We will need two simple estimates that we gather in the next lemma.

**Lemma 10.** (i) *Let  $r \geq 1$  be an integer. There exists a constant  $C(r)$  such that, for every integers  $n \geq 1$  and  $m \geq 0$ ,*

$$\mathbb{E} \left[ (\#\{k \in \llbracket 0, 2n \rrbracket : \zeta_k^{(n)} = m\})^r \right] \leq C(r) (m+1)^r.$$

(ii) *Let  $\varepsilon > 0$ . Then, for every  $r > 0$ ,*

$$\mathbb{P} \left( \sup_{0 \leq k \leq 2n} \zeta_k^{(n)} > n^{\frac{1}{2} + \varepsilon} \right) = O(n^{-r})$$

as  $n \rightarrow \infty$ .

Part (i) of the lemma can be deduced from Theorem 1.13 in Janson [7] using the connection between  $\zeta^{(n)}$  and the critical geometric Galton-Watson tree (it is also possible to give a direct argument), while Part (ii) is standard. Notice that Part (i) of Lemma 10 implies

$$\mathbb{E} \left[ (\#\{k \in \llbracket 0, 2n \rrbracket : \zeta_k^{(n)} \leq n^{\frac{1}{2} - \frac{\alpha}{2}}\})^2 \right] = o\left(\left(\frac{n}{\log n}\right)^2\right)$$

as  $n \rightarrow \infty$ .

We will make a repeated use of Kemperman's formula for simple random walk (see [24, p.122] for a more general version): For every choice of the integers  $m, k$  such that  $k > m \geq 0$ ,

$$\mathbb{P}_m(T = k) = \frac{m+1}{k} \mathbb{P}_m(\zeta_k = -1) = \frac{m+1}{k} \mathbb{P}_0(\zeta_k = m+1). \quad (13)$$

Together with this formula, we will use the local limit theorem for simple random walk on  $\mathbb{Z}$ , which we state in the form found in Lawler and Limic [14, Proposition 2.5.3, Corollary 2.5.4]: As  $k \rightarrow \infty$ ,

$$\mathbb{P}_0(\zeta_k = m) = \sqrt{\frac{2}{\pi k}} \exp\left(-\frac{m^2}{2k}\right) \exp\left(O\left(\frac{1}{k} + \frac{m^4}{k^3}\right)\right) \quad (14)$$

uniformly over integers  $m$  such that  $|m| \leq k$  and  $k+m$  is even.

We fix  $\alpha \in (0, 1/4)$  and to simplify notation, we write  $p_n = \lfloor n^{\frac{1}{2}-\alpha} \rfloor$  for every integer  $n \geq 1$ . Recall the notation  $\tau_p = \inf\{n \geq 0 : \zeta_n = \zeta_0 - p\}$ .

**Lemma 11.** *If  $\eta > 0$  is sufficiently small, we have*

$$\lim_{n \rightarrow \infty} \left( \sup_{\substack{n^{1-\eta} \leq k \leq 2n \\ n^{\frac{1}{2}-\frac{\alpha}{2}} \leq m \leq n^{\frac{1}{2}+\eta}}} \left| (\log n) \mathbb{P}_m(\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, \tau_{p_n} \rrbracket \mid T = k) - \frac{4\pi^2 \sigma^4}{1-2\alpha} \right| \right) = 0,$$

where in the supremum we consider only integers  $m$  and  $k$  such that  $k+m$  is odd.

*Proof.* We first explain how to choose  $\eta$ . We set  $q_n = \lfloor n^{1-\frac{3\alpha}{2}} \rfloor$  and note that

$$\mathbb{P}_0(\zeta_{q_n} > n^{\frac{1}{2}-\frac{\alpha}{2}}) \leq \mathbb{P}_0(\zeta_{q_n} > q_n^{\frac{1}{2}+c(\alpha)})$$

where  $c(\alpha) = \frac{\alpha}{4-6\alpha} > 0$ . By a standard bound, the latter probability is bounded (for  $n$  large) by  $\exp(-n^\gamma)$ , where the constant  $\gamma = \gamma(\alpha) > 0$  only depends on  $\alpha$ . We fix  $\eta > 0$  such that  $3\eta < \gamma$  and  $\eta \in (0, \alpha/8)$ .

To simplify notation, we then set

$$\Delta_n := \{(m, k) : n^{\frac{1}{2}-\frac{\alpha}{2}} \leq m \leq n^{\frac{1}{2}+\eta}, n^{1-\eta} \leq k \leq 2n \text{ and } k+m \text{ is odd}\}.$$

Since  $p_n \sim n^{-\alpha/4} \sqrt{q_n}$ , standard estimates give, for every  $\delta \in (0, \frac{\alpha}{4})$ ,

$$\lim_{n \rightarrow \infty} n^\delta \mathbb{P}_0(\tau_{p_n} \geq q_n) = 0. \quad (15)$$

We claim that we have also, for every  $\delta \in (0, \frac{\alpha}{4})$ ,

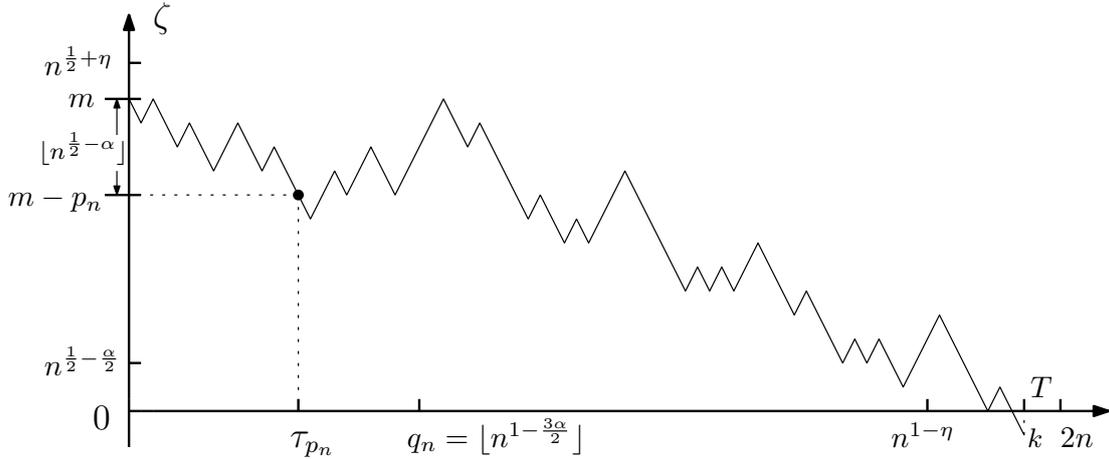
$$\lim_{n \rightarrow \infty} n^\delta \sup_{(m,k) \in \Delta_n} \mathbb{P}_m(\tau_{p_n} \geq q_n \mid T = k) = 0. \quad (16)$$

Let us postpone the proof of (16) and derive the estimate of the lemma.

Let us consider  $(m, k) \in \Delta_n$ . We have

$$\begin{aligned} & \mathbb{P}_m(\{\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, \tau_{p_n} \rrbracket\} \cap \{\tau_{p_n} \leq q_n\} \cap \{T = k\}) \\ &= \mathbb{E}_m \left[ \mathbf{1}_{\{\tau_{p_n} \leq q_n\}} \mathbf{1}_{\{\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, \tau_{p_n} \rrbracket\}} \mathbb{P}_{m-p_n}(T = k - \ell)_{\ell=\tau_{p_n}} \right], \end{aligned} \quad (17)$$

where we have used the strong Markov property at  $\tau_{p_n}$ . We now would like to say that the quantity  $\mathbb{P}_{m-p_n}(T = k - \ell)$ , evaluated at  $\ell = \tau_{p_n}$ , does not differ too much from  $\mathbb{P}_m(T = k)$



**Figure 4:** Illustration of the proof of Lemma 11

under our conditions on  $m, k$  and  $\tau_{p_n}$  (see Fig. 4 for an illustration). Let  $k'$  be an integer such that  $k - q_n \leq k' \leq k$  and  $k' + m - p_n$  is odd. By Kemperman's formula,

$$\mathbb{P}_{m-p_n}(T = k') = \frac{m - p_n + 1}{k'} \mathbb{P}_0(\zeta_{k'} = m - p_n + 1) \quad (18)$$

and by (14),

$$\mathbb{P}_0(\zeta_{k'} = m - p_n + 1) = \sqrt{\frac{2}{\pi k'}} \exp\left(-\frac{(m - p_n + 1)^2}{2k'}\right) \exp\left(O\left(\frac{1}{k'} + \frac{(m - p_n + 1)^4}{k'^3}\right)\right). \quad (19)$$

Next observe that

$$\begin{aligned} \left| \frac{(m - p_n + 1)^2}{2k'} - \frac{(m + 1)^2}{2k} \right| &\leq \frac{(m + 1)^2}{2} \left( \frac{1}{k'} - \frac{1}{k} \right) + \frac{(m + 1)^2 - (m - p_n + 1)^2}{2k'} \\ &\leq \frac{q_n(m + 1)^2}{kk'} + \frac{p_n(m + 1)}{k'}, \end{aligned}$$

which tends to 0 as  $n \rightarrow \infty$ , uniformly in  $m, k, k'$ . Comparing the estimate for  $\mathbb{P}_{m-p_n}(T = k')$  that follows from (18) and (19) with the similar estimate for  $\mathbb{P}_m(T = k)$  that follows from (13) and (14), we get

$$\lim_{n \rightarrow \infty} \left( \sup_{m, k, k'} \left| \frac{\mathbb{P}_{m-p_n}(T = k')}{\mathbb{P}_m(T = k)} - 1 \right| \right) = 0,$$

where the supremum is over all choices of  $(m, k, k')$  such that  $(m, k) \in \Delta_n$  and  $k'$  satisfies the preceding conditions. Using (17), we obtain that, for any fixed  $\delta > 0$ , we have for all sufficiently large  $n$ , for every  $(m, k) \in \Delta_n$ ,

$$\begin{aligned} &(1 - \delta) \mathbb{P}_m(\{\tau_{p_n} \leq q_n\} \cap \{\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, \tau_{p_n} \rrbracket\}) \\ &\leq \mathbb{P}_m(\{\tau_{p_n} \leq q_n\} \cap \{\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, \tau_{p_n} \rrbracket\} \mid T = k) \\ &\leq (1 + \delta) \mathbb{P}_m(\{\tau_{p_n} \leq q_n\} \cap \{\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, \tau_{p_n} \rrbracket\}). \end{aligned}$$

The quantity  $\mathbb{P}_m(\{\tau_{p_n} \leq q_n\} \cap \{\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, \tau_{p_n} \rrbracket\})$  does not depend on  $m \in \mathbb{Z}$ , and  $(\log n) \mathbb{P}_0(\tau_{p_n} > q_n)$  tends to 0 by (15). Using Proposition 8, we have thus

$$\lim_{n \rightarrow \infty} (\log n) \mathbb{P}_0(\{\tau_{p_n} \leq q_n\} \cap \{\widehat{W}_j \neq \widehat{W}_0, \forall j \in \llbracket 1, \tau_{p_n} \rrbracket\}) = \frac{4\pi^2 \sigma^4}{1 - 2\alpha}.$$

The estimate of the lemma follows from the preceding considerations and (16).

It remains to prove (16). If  $(m, k) \in \Delta_n$ , we have

$$\mathbb{P}_m(\tau_{p_n} \geq q_n \mid T = k) = \frac{\mathbb{P}_m(\{\tau_{p_n} \geq q_n\} \cap \{T = k\})}{\mathbb{P}_m(T = k)}.$$

Recall formula (13) for  $\mathbb{P}_m(T = k)$  and also note that by (14),

$$\mathbb{P}_0(\zeta_k = m + 1) = \sqrt{\frac{2}{\pi k}} \exp\left(-\frac{(m+1)^2}{2k}\right) \exp\left(O\left(\frac{1}{k} + \frac{m^4}{k^3}\right)\right), \quad (20)$$

when  $n \rightarrow \infty$ , uniformly in  $(m, k) \in \Delta_n$ . Notice that  $\frac{1}{k} + \frac{m^4}{k^3} \rightarrow 0$  as  $n \rightarrow \infty$ , uniformly in  $(m, k) \in \Delta_n$ , and that  $\frac{m^2}{2k} \leq n^{3\eta}$  if  $(m, k) \in \Delta_n$ . By our choice of  $\eta$ , it follows that

$$\mathbb{P}_m(\zeta_{q_n} > m + n^{\frac{1}{2} - \frac{\alpha}{2}} \mid T = k) \leq \frac{\mathbb{P}_m(\zeta_{q_n} > m + n^{\frac{1}{2} - \frac{\alpha}{2}})}{\mathbb{P}_m(T = k)} = \frac{k}{m+1} \frac{\mathbb{P}_0(\zeta_{q_n} > n^{\frac{1}{2} - \frac{\alpha}{2}})}{\mathbb{P}_0(\zeta_k = m+1)} = O\left(\frac{1}{n}\right) \quad (21)$$

as  $n \rightarrow \infty$ , uniformly in  $m$  and  $k$ .

On the other hand, by applying the Markov property at time  $q_n$ , we have

$$\mathbb{P}_m(\{\tau_{p_n} \geq q_n\} \cap \{\zeta_{q_n} \leq m + n^{\frac{1}{2} - \frac{\alpha}{2}}\} \cap \{T = k\}) = \mathbb{E}_m\left[\mathbf{1}_{\{\tau_{p_n} \geq q_n\} \cap \{\zeta_{q_n} \leq m + n^{\frac{1}{2} - \frac{\alpha}{2}}\}} \mathbb{P}_{\zeta_{q_n}}(T = k - q_n)\right].$$

On the event  $\{\tau_{p_n} \geq q_n\} \cap \{\zeta_{q_n} \leq m + n^{\frac{1}{2} - \frac{\alpha}{2}}\}$  we have  $m - p_n \leq \zeta_{q_n} \leq m + n^{\frac{1}{2} - \frac{\alpha}{2}}$ ,  $\mathbb{P}_m$  a.s. If  $m - p_n \leq m' \leq m + n^{\frac{1}{2} - \frac{\alpha}{2}}$  and  $m' + k - q_n$  is odd, using again Kemperman's formula, we have

$$\mathbb{P}_{m'}(T = k - q_n) = \frac{m' + 1}{k - q_n} \mathbb{P}_{m'}(\zeta_{k - q_n} = -1) = \frac{m' + 1}{k - q_n} \mathbb{P}_0(\zeta_{k - q_n} = m' + 1).$$

Furthermore, from (14),

$$\mathbb{P}_0(\zeta_{k - q_n} = m' + 1) = \sqrt{\frac{2}{\pi(k - q_n)}} \exp\left(-\frac{(m' + 1)^2}{2(k - q_n)}\right) \exp\left(O\left(\frac{1}{k} + \frac{m'^4}{k^3}\right)\right). \quad (22)$$

Now observe that

$$-\frac{(m' + 1)^2}{2(k - q_n)} + \frac{(m + 1)^2}{2k} \leq -\frac{(m' + 1)^2 - (m + 1)^2}{2k} = -\frac{(m' - m)(m' + m + 2)}{2k}$$

and the right-hand side tends to 0 as  $n \rightarrow \infty$ , uniformly in  $(m, k) \in \Delta_n$  and  $m'$  such that  $m - p_n \leq m' \leq m + n^{\frac{1}{2} - \frac{\alpha}{2}}$ . By comparing (20) and (22), noting that  $m' \leq 2m$  under our assumptions, we get

$$\limsup_{n \rightarrow \infty} \left( \sup_{m, k, m'} \frac{\mathbb{P}_{m'}(T = k - q_n)}{\mathbb{P}_m(T = k)} \right) \leq 2.$$

It follows that, for all  $n$  sufficiently large, we have, for every  $(m, k) \in \Delta_n$ ,

$$\frac{\mathbb{P}_m(\{\tau_{p_n} \geq q_n\} \cap \{\zeta_{q_n} \leq m + n^{\frac{1}{2} - \frac{\alpha}{2}}\} \cap \{T = k\})}{\mathbb{P}_m(T = k)} \leq 3 \mathbb{P}_m(\tau_{p_n} \geq q_n) = 3 \mathbb{P}_0(\tau_{p_n} \geq q_n).$$

Recalling (15), we have thus proved that, for every  $\delta \in (0, \frac{\alpha}{4})$ ,

$$\lim_{n \rightarrow \infty} n^\delta \sup_{(m, k) \in \Delta_n} \mathbb{P}_m(\{\tau_{p_n} \geq q_n\} \cap \{\zeta_{q_n} \leq m + n^{\frac{1}{2} - \frac{\alpha}{2}}\} \mid T = k) = 0,$$

and by combining this with (21), we get the desired estimate (16).  $\square$

We set, for every  $n \geq 1$ ,

$$R_n^\bullet := \#\{\widehat{W}_0^{(n)}, \widehat{W}_1^{(n)}, \dots, \widehat{W}_{2n}^{(n)}\}.$$

**Proposition 12.** *We have*

$$\limsup_{n \rightarrow \infty} \left(\frac{\log n}{n}\right)^2 \mathbb{E}[(R_n^\bullet)^2] \leq (8\pi^2\sigma^4)^2.$$

*Proof.* We note that

$$R_n^\bullet = \sum_{i=1}^{2n} \mathbf{1}_{\{\widehat{W}_\ell^{(n)} \neq \widehat{W}_i^{(n)}, \forall \ell \in \llbracket i+1, 2n \rrbracket\}}$$

and therefore

$$\mathbb{E}[(R_n^\bullet)^2] = \sum_{i,j=1}^{2n} \mathbb{P}(A_n(i,j)), \quad (23)$$

where  $A_n(i,j)$  is defined by

$$A_n(i,j) := \{\widehat{W}_\ell^{(n)} \neq \widehat{W}_i^{(n)}, \forall \ell \in \llbracket i+1, 2n \rrbracket\} \cap \{\widehat{W}_\ell^{(n)} \neq \widehat{W}_j^{(n)}, \forall \ell \in \llbracket j+1, 2n \rrbracket\}.$$

We fix  $\alpha \in (0, 1/4)$ , and define  $p_n$  and  $q_n$  for every  $n \geq 1$  as above. We also fix  $\eta > 0$  so that the conclusion of Lemma 11 holds.

In view of proving the proposition, we will use formula (23). In this formula, we can restrict our attention to values of  $i$  and  $j$  such that  $j - i > n^{1-\frac{\alpha}{2}}$  and  $j < 2n - n^{1-\eta}$  (or the same with  $i$  and  $j$  interchanged). Also, when bounding  $\mathbb{P}(A_n(i,j))$ , we may impose the additional constraint that  $n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_i^{(n)} \leq n^{\frac{1}{2}+\eta}$  and  $n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_j^{(n)} \leq n^{\frac{1}{2}+\eta}$ : Indeed, Lemma 10 readily shows that the event where either of these constraints is not satisfied will give a negligible contribution to the sum in (23).

Let us fix  $i, j \in \llbracket 1, 2n \rrbracket$  such that  $j - i > n^{1-\frac{\alpha}{2}}$  and  $j < 2n - n^{1-\eta}$ . By using the definition of  $W^{(n)}$  as a conditioned process and applying the Markov property at time  $i$ , we have

$$\begin{aligned} & \mathbb{P}(A_n(i,j) \cap \{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_i^{(n)} \leq n^{\frac{1}{2}+\eta}\} \cap \{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_j^{(n)} \leq n^{\frac{1}{2}+\eta}\}) \\ &= \frac{\mathbb{E}\left[\mathbf{1}_{\{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_i \leq n^{\frac{1}{2}+\eta}\}} \mathbf{1}_{\{T>i\}} \mathbb{E}_{\zeta_i}\left[\mathbf{1}_{\{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_{j-i} \leq n^{\frac{1}{2}+\eta}\}} \mathbf{1}_{A'_n(i,j)} \mathbf{1}_{\{T=2n+1-i\}}\right]\right]}{\mathbb{P}(T=2n+1)}, \end{aligned} \quad (24)$$

where

$$A'_n(i,j) := \{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, 2n-i \rrbracket\} \cap \{\widehat{W}_\ell \neq \widehat{W}_{j-i}, \forall \ell \in \llbracket j-i+1, 2n-i \rrbracket\}.$$

Setting  $r = j - i$ , we are thus led to bound

$$\mathbb{E}_m \left[ \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, k-1 \rrbracket\}} \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_r, \forall \ell \in \llbracket r+1, k-1 \rrbracket\}} \mathbf{1}_{\{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_r \leq n^{\frac{1}{2}+\eta}\}} \mathbf{1}_{\{T=k\}} \right], \quad (25)$$

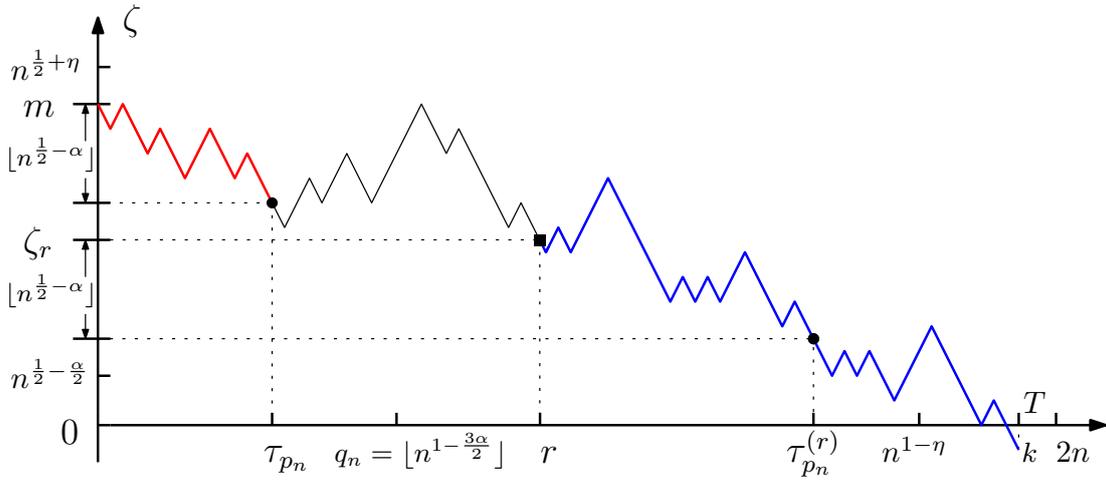
where  $n^{\frac{1}{2}-\frac{\alpha}{2}} \leq m \leq n^{\frac{1}{2}+\eta}$ ,  $r > n^{1-\frac{\alpha}{2}}$  and  $r + n^{1-\eta} < k \leq 2n$  (and moreover  $k + m$  needs to be odd). Recall the notation  $\tau_{p_n}$ , and set

$$\tau_{p_n}^{(r)} := \inf\{\ell \geq r : \zeta_\ell = \zeta_r - p_n\}.$$

Thanks to (16), we can also introduce the constraint  $\tau_{p_n} \leq q_n$  inside the expectation in (25), up to an error that is bounded above by  $\mathbb{P}_m(T = k) o(n^{-\delta})$  for some  $\delta > 0$  (here the term  $o(n^{-\delta})$  is uniform in  $m, r, k$  satisfying the preceding conditions). Furthermore, we get an upper bound by replacing the interval  $\llbracket 1, k-1 \rrbracket$ , resp.  $\llbracket r+1, k-1 \rrbracket$ , by  $\llbracket 1, \tau_{p_n} \rrbracket$ , resp.  $\llbracket r+1, \tau_{p_n}^{(r)} \rrbracket$ . Next, using the Markov property at time  $r$ , and noting that  $r > q_n$ , we have

$$\begin{aligned} & \mathbb{E}_m \left[ \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, \tau_{p_n} \rrbracket\}} \mathbf{1}_{\{\tau_{p_n} \leq q_n\}} \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_r, \forall \ell \in \llbracket r+1, \tau_{p_n}^{(r)} \rrbracket\}} \mathbf{1}_{\{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_r \leq n^{\frac{1}{2}+\eta}\}} \mathbf{1}_{\{T=k\}} \right] \\ &= \mathbb{E}_m \left[ \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, \tau_{p_n} \rrbracket\}} \mathbf{1}_{\{\tau_{p_n} \leq q_n\}} \mathbf{1}_{\{T>r\}} \mathbf{1}_{\{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_r \leq n^{\frac{1}{2}+\eta}\}} \right. \\ & \quad \left. \times \mathbb{E}_{(W_r)} \left[ \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, \tau_{p_n} \rrbracket\}} \mathbf{1}_{\{T=k-r\}} \right] \right]. \end{aligned}$$

See Fig. 5 for an illustration.



**Figure 5:** Illustration of the proof of Proposition 12

Then the key observation is the following. Let  $z_0 = m, z_1, \dots, z_r$  be a simple random walk trajectory over  $\llbracket 0, r \rrbracket$  such that  $0 \leq \min\{z_\ell : 0 \leq \ell \leq r\} \leq m - p_n$ . Then under  $\mathbb{P}_m$ , conditionally on the event  $\{\zeta_1 = z_1, \dots, \zeta_r = z_r\}$ , the path  $(W_r(z_r) - W_r(z_r - \ell))_{\ell \geq 0}$  is *independent* of the event  $\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, \tau_{p_n} \rrbracket\}$ , and distributed as  $(S_\ell)_{\ell \geq 0}$ . This property easily follows from the construction of the discrete snake.

Thanks to the latter observation, we may rewrite the right-hand side of the last display, after conditioning with respect to  $\zeta_1, \dots, \zeta_r$ , in the form

$$\mathbb{E}_m \left[ \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, \tau_{p_n} \rrbracket\}} \mathbf{1}_{\{\tau_{p_n} \leq q_n\}} \mathbf{1}_{\{T>r\}} \mathbf{1}_{\{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_r \leq n^{\frac{1}{2}+\eta}\}} \mathbb{E}_{\zeta_r} \left[ \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, \tau_{p_n} \rrbracket\}} \mathbf{1}_{\{T=k-r\}} \right] \right]. \quad (26)$$

Recall that  $k - r > n^{1-\eta}$ , and let  $\varepsilon > 0$ . It follows from Lemma 11 that, for  $n$  large enough, on the event  $\{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_r \leq n^{\frac{1}{2}+\eta}\}$ , the quantity

$$\mathbb{E}_{\zeta_r} \left[ \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, \tau_{p_n} \rrbracket\}} \mathbf{1}_{\{T=k-r\}} \right]$$

is bounded above by

$$\left( \frac{4\pi^2\sigma^4}{1-2\alpha} + \varepsilon \right) (\log n)^{-1} \mathbb{P}_{\zeta_r}(T = k - r).$$

Hence the quantity (26) is also bounded by

$$\begin{aligned} & \left( \frac{4\pi^2\sigma^4}{1-2\alpha} + \varepsilon \right) (\log n)^{-1} \mathbb{E}_m \left[ \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, \tau_{pn} \rrbracket\}} \mathbf{1}_{\{\tau_{pn} \leq q_n\}} \mathbf{1}_{\{T > r\}} \mathbf{1}_{\{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_r \leq n^{\frac{1}{2}+\eta}\}} \mathbb{P}_{\zeta_r}(T = k - r) \right] \\ & = \left( \frac{4\pi^2\sigma^4}{1-2\alpha} + \varepsilon \right) (\log n)^{-1} \mathbb{E}_m \left[ \mathbf{1}_{\{\widehat{W}_\ell \neq \widehat{W}_0, \forall \ell \in \llbracket 1, \tau_{pn} \rrbracket\}} \mathbf{1}_{\{\tau_{pn} \leq q_n\}} \mathbf{1}_{\{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_r \leq n^{\frac{1}{2}+\eta}\}} \mathbf{1}_{\{T=k\}} \right], \end{aligned}$$

again by the Markov property at time  $r$ . Finally, another application of Lemma 11 shows that the quantity in the last display is bounded above for  $n$  large by

$$\left( \frac{4\pi^2\sigma^4}{1-2\alpha} + \varepsilon \right)^2 (\log n)^{-2} \mathbb{P}_m(T = k).$$

Summarizing, we see that the quantity (25) is bounded above for  $n$  large by

$$\left( \left( \frac{4\pi^2\sigma^4}{1-2\alpha} + \varepsilon \right)^2 (\log n)^{-2} + o(n^{-\delta}) \right) \mathbb{P}_m(T = k).$$

Finally, from (24), we have for  $n$  large

$$\begin{aligned} & \mathbb{P}(A_n(i, j) \cap \{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_i^{(n)} \leq n^{\frac{1}{2}+\eta}\} \cap \{n^{\frac{1}{2}-\frac{\alpha}{2}} \leq \zeta_j^{(n)} \leq n^{\frac{1}{2}+\eta}\}) \\ & \leq \left( \left( \frac{4\pi^2\sigma^4}{1-2\alpha} + \varepsilon \right)^2 (\log n)^{-2} + o(n^{-\delta}) \right) \frac{\mathbb{E} \left[ \mathbf{1}_{\{T > i\}} \mathbb{P}_{\zeta_i}(T = 2n + 1 - i) \right]}{\mathbb{P}(T = 2n + 1)} \\ & = \left( \frac{4\pi^2\sigma^4}{1-2\alpha} + \varepsilon \right)^2 (\log n)^{-2} + o(n^{-\delta}), \end{aligned}$$

where the term  $o(n^{-\delta})$  is uniform in  $i$  and  $j$  satisfying the preceding conditions. The statement of the proposition follows by summing this bound over  $i$  and  $j$ .  $\square$

**Lemma 13.** *We have*

$$\liminf_{n \rightarrow \infty} \frac{\log n}{n} \mathbb{E}[R_n^\bullet] \geq 8\pi^2\sigma^4.$$

*Proof.* Let  $\delta > 0$  and  $\varepsilon \in (0, \frac{1}{2})$ . To simplify notation we write  $n(\varepsilon) = \lfloor 2(1-2\varepsilon)n \rfloor$  in this proof. We fix  $0 < a < b$  such that, if  $(\mathbf{e}_t)_{0 \leq t \leq 1}$  denotes a normalized Brownian excursion defined under the probability measure  $P$ , we have

$$P(\mathbf{e}_\varepsilon \notin (a, b)) = P(\mathbf{e}_{1-\varepsilon} \notin (a, b)) < \delta.$$

Since we know that the sequence of processes  $((2n)^{-1/2} \zeta_{\lfloor 2nt \rfloor}^{(n)})_{0 \leq t \leq 1}$  converges in distribution to  $(\mathbf{e}_t)_{0 \leq t \leq 1}$ , it follows that, for every sufficiently large  $n$ ,

$$\mathbb{P} \left( \zeta_{\lfloor 2n\varepsilon \rfloor}^{(n)} \notin [a\sqrt{2n}, b\sqrt{2n}] \text{ or } \zeta_{\lfloor 2n\varepsilon \rfloor + n(\varepsilon)}^{(n)} \notin [a\sqrt{2n}, b\sqrt{2n}] \right) \leq \delta. \quad (27)$$

Let  $\mu_\varepsilon^{(n)}$  denote the law of  $\zeta_{\lfloor 2n\varepsilon \rfloor}^{(n)}$ . If  $F_n$  is a nonnegative function on  $\mathbb{Z}^{n(\varepsilon)+1}$ , the Markov property

gives

$$\begin{aligned}
& \mathbb{E} \left[ \mathbf{1}_{\{a\sqrt{2n} \leq \zeta_{\lfloor 2n\varepsilon \rfloor}^{(n)} \leq b\sqrt{2n}\}} \mathbf{1}_{\{a\sqrt{2n} \leq \zeta_{\lfloor 2n\varepsilon \rfloor + n(\varepsilon)}^{(n)} \leq b\sqrt{2n}\}} F_n((\zeta_{\lfloor 2n\varepsilon \rfloor + k}^{(n)})_{0 \leq k \leq n(\varepsilon)}) \right] \\
&= \mathbb{E} \left[ \mathbf{1}_{\{a\sqrt{2n} \leq \zeta_{\lfloor 2n\varepsilon \rfloor}^{(n)} \leq b\sqrt{2n}\}} \mathbb{E}_{\zeta_{\lfloor 2n\varepsilon \rfloor}^{(n)}} \left[ \mathbf{1}_{\{a\sqrt{2n} \leq \zeta_{n(\varepsilon)} \leq b\sqrt{2n}\}} F_n((\zeta_k)_{0 \leq k \leq n(\varepsilon)}) \mid T = 2n + 1 - \lfloor 2n\varepsilon \rfloor \right] \right] \\
&= \sum_{a\sqrt{2n} \leq m \leq b\sqrt{2n}} \mu_\varepsilon^{(n)}(m) \frac{\mathbb{E}_m \left[ \mathbf{1}_{\{a\sqrt{2n} \leq \zeta_{n(\varepsilon)} \leq b\sqrt{2n}\}} F_n((\zeta_k)_{0 \leq k \leq n(\varepsilon)}) \mathbf{1}_{\{T = 2n + 1 - \lfloor 2n\varepsilon \rfloor\}} \right]}{\mathbb{P}_m(T = 2n + 1 - \lfloor 2n\varepsilon \rfloor)} \\
&= \sum_{a\sqrt{2n} \leq m \leq b\sqrt{2n}} \mu_\varepsilon^{(n)}(m) \frac{\mathbb{E}_m \left[ \mathbf{1}_{\{a\sqrt{2n} \leq \zeta_{n(\varepsilon)} \leq b\sqrt{2n}\}} F_n((\zeta_k)_{0 \leq k \leq n(\varepsilon)}) \mathbf{1}_{\{T > n(\varepsilon)\}} \mathbb{P}_{\zeta_{n(\varepsilon)}}(T = \tilde{n}(\varepsilon)) \right]}{\mathbb{P}_m(T = 2n + 1 - \lfloor 2n\varepsilon \rfloor)}
\end{aligned}$$

where  $\tilde{n}(\varepsilon) := 2n + 1 - \lfloor 2n\varepsilon \rfloor - n(\varepsilon)$ .

Let  $m, m' \in [a\sqrt{2n}, b\sqrt{2n}]$  be such that  $m + \lfloor 2n\varepsilon \rfloor$  and  $m' + \lfloor 2n\varepsilon \rfloor + n(\varepsilon)$  are even. By Kemperman's formula (13),

$$\frac{\mathbb{P}_{m'}(T = \tilde{n}(\varepsilon))}{\mathbb{P}_m(T = 2n + 1 - \lfloor 2n\varepsilon \rfloor)} = \frac{2n + 1 - \lfloor 2n\varepsilon \rfloor}{\tilde{n}(\varepsilon)} \frac{m' + 1}{m + 1} \frac{\mathbb{P}_0(\zeta_{\tilde{n}(\varepsilon)} = m' + 1)}{\mathbb{P}_0(\zeta_{2n+1-\lfloor 2n\varepsilon \rfloor} = m + 1)}$$

and using (14), we easily obtain that there exists a finite constant  $C(\varepsilon, a, b)$  such that, for every sufficiently large  $n$ , and every  $m, m'$  satisfying the above conditions,

$$\frac{\mathbb{P}_{m'}(T = \tilde{n}(\varepsilon))}{\mathbb{P}_m(T = 2n + 1 - \lfloor 2n\varepsilon \rfloor)} \leq C(\varepsilon, a, b).$$

We thus obtain that, for every large enough  $n$ ,

$$\begin{aligned}
& \mathbb{E} \left[ \mathbf{1}_{\{a\sqrt{2n} \leq \zeta_{\lfloor 2n\varepsilon \rfloor}^{(n)} \leq b\sqrt{2n}\}} \mathbf{1}_{\{a\sqrt{2n} \leq \zeta_{\lfloor 2n\varepsilon \rfloor + n(\varepsilon)}^{(n)} \leq b\sqrt{2n}\}} F_n((\zeta_{\lfloor 2n\varepsilon \rfloor + k}^{(n)})_{0 \leq k \leq n(\varepsilon)}) \right] \\
&\leq C(\varepsilon, a, b) \sum_{a\sqrt{2n} \leq m \leq b\sqrt{2n}} \mu_\varepsilon^{(n)}(m) \mathbb{E}_m \left[ \mathbf{1}_{\{a\sqrt{2n} \leq \zeta_{n(\varepsilon)} \leq b\sqrt{2n}\}} F_n((\zeta_k)_{0 \leq k \leq n(\varepsilon)}) \mathbf{1}_{\{T > n(\varepsilon)\}} \right] \\
&\leq C(\varepsilon, a, b) \sum_{a\sqrt{2n} \leq m \leq b\sqrt{2n}} \mu_\varepsilon^{(n)}(m) \mathbb{E}_m \left[ F_n((\zeta_k)_{0 \leq k \leq n(\varepsilon)}) \right].
\end{aligned}$$

Let  $G_n$  be a nonnegative measurable function on  $\mathcal{W}^{n(\varepsilon)+1}$ . The preceding bound remains valid if we replace  $F_n((\zeta_{\lfloor 2n\varepsilon \rfloor + k}^{(n)})_{0 \leq k \leq n(\varepsilon)})$  by  $G_n((W_{\lfloor 2n\varepsilon \rfloor + k}^{(n)})_{0 \leq k \leq n(\varepsilon)})$  in the left-hand side and  $F_n((\zeta_k)_{0 \leq k \leq n(\varepsilon)})$  by  $G_n((W_k)_{0 \leq k \leq n(\varepsilon)})$  in the right-hand side (just use the fact that the conditional distribution of  $W^{(n)}$  given  $\zeta^{(n)}$  is the same as the conditional distribution of  $W$  given  $\zeta$ ). In particular, if we let  $G_n(w_0, w_1, \dots, w_{n(\varepsilon)})$  be the indicator function of the set where

$$\left| \frac{\log n(\varepsilon)}{n(\varepsilon)} \#\{\hat{w}_0, \hat{w}_1, \dots, \hat{w}_{n(\varepsilon)}\} - 4\pi^2\sigma^4 \right| > \delta,$$

we obtain that

$$\begin{aligned}
& \mathbb{P} \left( \zeta_{\lfloor 2n\varepsilon \rfloor}^{(n)} \in [a\sqrt{2n}, b\sqrt{2n}], \zeta_{\lfloor 2n\varepsilon \rfloor + n(\varepsilon)}^{(n)} \in [a\sqrt{2n}, b\sqrt{2n}], \left| \frac{\log n(\varepsilon)}{n(\varepsilon)} R_n^{\bullet, \varepsilon} - 4\pi^2\sigma^4 \right| > \delta \right) \\
&\leq C(\varepsilon, a, b) \mathbb{P} \left( \left| \frac{\log n(\varepsilon)}{n(\varepsilon)} R_{n(\varepsilon)} - 4\pi^2\sigma^4 \right| > \delta \right), \tag{28}
\end{aligned}$$

where

$$R_n^{\bullet,\varepsilon} := \#\{\widehat{W}_{[2n\varepsilon]}^{(n)}, \widehat{W}_{[2n\varepsilon]+1}^{(n)}, \dots, \widehat{W}_{[2n\varepsilon]+n(\varepsilon)}^{(n)}\}.$$

Here we used the (obvious) fact that the distribution of  $R_n$  under  $\mathbb{P}_m$  does not depend on  $m$ .

By Theorem 9, the right-hand side of (28) tends to 0 as  $n \rightarrow \infty$ . Using also (27), we obtain that

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{\log n(\varepsilon)}{n(\varepsilon)} R_n^{\bullet,\varepsilon} - 4\pi^2\sigma^4\right| > \delta\right) \leq \delta.$$

Since  $R_n^\bullet \geq R_n^{\bullet,\varepsilon}$  and since both  $\delta$  and  $\varepsilon$  can be chosen arbitrarily small, the statement of the lemma follows.  $\square$

**Theorem 14.** *We have*

$$\frac{\log n}{n} R_n^\bullet \xrightarrow[n \rightarrow \infty]{L^2(\mathbb{P})} 8\pi^2\sigma^4.$$

*Proof.* By combining Proposition 12 and Lemma 13, we get that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \mathbb{E}\left[\left(\frac{\log n}{n} R_n^\bullet - 8\pi^2\sigma^4\right)^2\right] \\ & \leq \left(\limsup_{n \rightarrow \infty} \mathbb{E}\left[\left(\frac{\log n}{n} R_n^\bullet\right)^2\right]\right) - 16\pi^2\sigma^4 \left(\liminf_{n \rightarrow \infty} \mathbb{E}\left[\frac{\log n}{n} R_n^\bullet\right]\right) + (8\pi^2\sigma^4)^2 \leq 0, \end{aligned}$$

which gives the desired result.  $\square$

Theorem 14 and the remarks before Lemma 10 give the case  $d = 4$  of Theorem 1.

### 3.2 Proof of the main estimate

In this subsection, we prove Proposition 8, which was a key ingredient of the results of the previous subsection. We first recall some basic facts. For every  $x \in \mathbb{Z}^4$  and  $k \geq 0$ , we set

$$p_k(x) = P(S_k = x)$$

and we now denote the Green function of the random walk  $S$  by

$$G(x) = \sum_{k=0}^{\infty} p_k(x)$$

( $G = G_\theta$  in the notation of Section 2). A standard estimate (see e.g. [14, Chapter 4]) states that

$$\lim_{x \rightarrow \infty} |x|^2 G(x) = \frac{1}{2\pi^2\sigma^2}. \quad (29)$$

Let  $\mathbf{p}$  be the period of the random walk  $S$ . Since  $S$  is assumed to be symmetric, we have  $\mathbf{p} = 1$  or 2. Then from the local limit theorem (see e.g. [14, Chapter 2]), we have

$$\lim_{j \rightarrow \infty, j \in \mathbf{p}\mathbb{Z}} j^2 p_j(0) = \frac{\mathbf{P}}{4\pi^2\sigma^4}. \quad (30)$$

We state our first lemma.

**Lemma 15.** *We have  $\mathbb{P}(\widehat{W}_k = 0) = 0$  if  $k \notin \mathbf{p}\mathbb{Z}$ , and*

$$\lim_{k \rightarrow \infty, k \in \mathbf{p}\mathbb{Z}} k \mathbb{P}(\widehat{W}_k = 0) = \frac{\mathbf{P}}{4\pi^2\sigma^4}.$$

*Proof.* The first assertion is clear since  $\zeta_k$  is odd when  $k$  is odd,  $\mathbb{P}$  a.s. Then, for every integer  $k \geq 0$ , set

$$\underline{\zeta}_k = \min_{0 \leq j \leq k} \zeta_j,$$

and

$$X_k = \zeta_k - 2\underline{\zeta}_k.$$

From the construction of the discrete snake, and the fact that  $S$  is symmetric, the conditional distribution of  $\widehat{W}_k$  knowing that  $X_k = m$  is the law of  $S_m$ . Consequently,

$$\mathbb{P}(\widehat{W}_k = 0) = \sum_{m=0}^{\infty} \mathbb{P}(X_k = m) p_m(0). \quad (31)$$

Asymptotics for  $P(S_m = 0) = p_m(0)$  are given by (30). We then need to evaluate  $\mathbb{P}(X_k = m)$ . Set  $\widetilde{X}_k = 1 + X_k$  for every  $k \geq 0$ . The discrete version of Pitman's theorem (see [23, Lemma 3.1]) shows that, under the probability measure  $\mathbb{P}$ ,  $(\widetilde{X}_k)_{k \geq 0}$  is a Markov chain on  $\{1, 2, \dots\}$  with transition kernel  $Q$  given by  $Q(1, 2) = 1$  and for every  $j \geq 2$ ,

$$Q(j, j+1) = \frac{1}{2} \frac{j+1}{j}, \quad Q(j, j-1) = \frac{1}{2} \frac{j-1}{j}.$$

This Markov chain is also the discrete  $h$ -transform of simple random walk on  $\mathbb{Z}_+$  (killed upon hitting 0) corresponding to  $h(j) = j$ . Let  $(Y_k)_{k \geq 0}$  stand for a simple random walk on  $\mathbb{Z}$  that starts from  $\ell$  under the probability measure  $P_\ell$ , and let  $H_0 = \inf\{n \geq 0 : Y_n = 0\}$ . It follows from the preceding observations that, for every integer  $k \geq 1$  and every  $m \geq 1$  such that  $1 \leq m \leq k+1$  and  $k+m$  is odd,

$$\begin{aligned} \mathbb{P}(\widetilde{X}_k = m) &= m P_1(Y_k = m, H_0 > k) \\ &= m \left( P_0(Y_k = m-1) - P_0(Y_k = m+1) \right) \\ &= m \times 2^{-k} \left( \binom{k}{\frac{k+m-1}{2}} - \binom{k}{\frac{k+m+1}{2}} \right) \\ &= \frac{2m^2}{k+m+1} P_0(Y_k = m-1) \end{aligned}$$

Hence, for every  $m \geq 0$ ,

$$\mathbb{P}(X_k = m) = \frac{2(m+1)^2}{k+m+2} P_0(Y_k = m). \quad (32)$$

From (31) and (32), we get

$$\mathbb{P}(\widehat{W}_k = 0) = \sum_{m=0}^k \frac{2}{k+m+2} \left( (m+1)^2 p_m(0) \right) P_0(Y_k = m),$$

and the second assertion of the lemma follows using (30).  $\square$

In the next lemma, for every integer  $k \geq 0$ , we use the notation  $\widetilde{W}_k$  for the time-shifted path  $\widetilde{W}_k = (\widetilde{W}_k(j))_{j \leq 0}$ , where  $\widetilde{W}_k(j) := W_k(\zeta_k + j)$ , for every  $j \leq 0$ .

**Lemma 16.** *Let  $k \geq 1$  such that  $\mathbb{P}(\widehat{W}_k = 0) > 0$ . Under the conditional probability measure  $\mathbb{P}(\cdot \mid \widehat{W}_k = 0)$ , the two pairs  $(W_0, \widetilde{W}_k)$  and  $(\widetilde{W}_k, W_0)$  have the same distribution.*

*Proof.* Write  $\pi_k(i, j)$ ,  $i, j \geq 0$  for the joint distribution under  $\mathbb{P}$  of the pair

$$\left( - \min_{0 \leq \ell \leq k} \zeta_\ell, \zeta_k - \min_{0 \leq \ell \leq k} \zeta_\ell \right).$$

By an easy time-reversal argument, we have  $\pi_k(i, j) = \pi_k(j, i)$  for every  $i, j \geq 0$ . On the other hand, under  $\mathbb{P}$ , conditionally on

$$\left( - \min_{0 \leq \ell \leq k} \zeta_\ell, \zeta_k - \min_{0 \leq \ell \leq k} \zeta_\ell \right) = (i, j)$$

we have  $W_0(-i - \ell) = W_k(-i - \ell) = \widetilde{W}_k(-j - \ell)$  for every  $\ell \geq 0$ , and the two random paths

$$(W_0(-i + \ell) - W_0(-i))_{0 \leq \ell \leq i}$$

and

$$(\widetilde{W}_k(-j + \ell) - \widetilde{W}_k(-j))_{0 \leq \ell \leq j} = (W_k(-i + \ell) - W_0(-i))_{0 \leq \ell \leq j}$$

are independent and distributed as the random walk  $S$  stopped respectively at time  $i$  and at time  $j$ . Note that the event  $\{\widehat{W}_k = 0\}$  occurs if and only if the latter two paths have the same endpoint. The statement of the lemma easily follows from the preceding observations and the property  $\pi_k(i, j) = \pi_k(j, i)$ .  $\square$

Let us fix  $\eta \in (0, 1/4)$ . Thanks to Lemma 15, we may choose  $\delta > 0$  small enough so that, for every sufficiently large  $n$ ,

$$\sum_{k=\lfloor (1-\delta)n \rfloor}^n \mathbb{P}(\widehat{W}_k = 0) < \eta.$$

We then observe that

$$\begin{aligned} 1 &= \sum_{k=0}^n \mathbb{P}(\widehat{W}_k = 0; \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket k+1, n \rrbracket) \\ &= \sum_{k=0}^n \mathbb{E} \left[ \mathbf{1}_{\{\widehat{W}_k=0\}} \mathbb{P}_{(W_k)}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n-k \rrbracket) \right] \\ &= \sum_{k=0}^n \mathbb{E} \left[ \mathbf{1}_{\{\widehat{W}_k=0\}} \mathbb{P}_{(W_0)}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n-k \rrbracket) \right]. \end{aligned}$$

In the second equality, we applied the Markov property of the discrete snake at time  $k$ , and in the third one we used Lemma 16.

From the last equalities and our choice of  $\delta$ , it follows that, for  $n$  large,

$$\mathbb{E} \left[ \left( \sum_{k=0}^{\lfloor (1-\delta)n \rfloor} \mathbf{1}_{\{\widehat{W}_k=0\}} \right) \mathbb{P}_{(W_0)}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, \lfloor \delta n \rfloor \rrbracket) \right] \geq 1 - \eta.$$

Next fix  $\varepsilon \in (0, 1/2)$  and write  $n(\varepsilon) = \lfloor n^{\frac{1}{2}+\varepsilon} \rfloor$  to simplify notation. For every integer  $p \geq 1$ , there exists a constant  $C_{p,\varepsilon}$  such that, for every  $n \geq 1$ ,

$$\mathbb{P}(\tau_{n(\varepsilon)} \leq n) \leq C_{p,\varepsilon} n^{-p}.$$

Hence, we also get, for every sufficiently large  $n$ ,

$$\mathbb{E} \left[ \left( \sum_{k=0}^{\tau_{n(\varepsilon)}-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right) \mathbb{P}_{(W_0)} \left( \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, \lfloor \delta n \rfloor \rrbracket \right) \right] \geq 1 - 2\eta.$$

By conditioning with respect to  $W_0$ , we see that the left-hand side of the preceding display is equal to

$$\mathbb{E} \left[ \mathbb{E}_{(W_0)} \left[ \sum_{k=0}^{\tau_{n(\varepsilon)}-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right] \mathbb{P}_{(W_0)} \left( \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, \lfloor \delta n \rfloor \rrbracket \right) \right].$$

We now note that, for every integer  $m \geq 1$ ,

$$\mathbb{E}_{(W_0)} \left[ \sum_{k=0}^{\tau_m-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right] = 2 \sum_{j=0}^{m-1} G(-W_0(-j)) \quad (33)$$

(we could write  $G(W_0(-j))$  instead of  $G(-W_0(-j))$  because  $S$  is symmetric, but the preceding formula would hold also in the non-symmetric case). To derive formula (33), first consider the case  $m = 1$ . By a standard property of simple random walk, we have for every integer  $i \geq 0$ ,

$$\mathbb{E}_{(W_0)} \left[ \sum_{k=0}^{\tau_1-1} \mathbf{1}_{\{\zeta_k=i\}} \right] = 2.$$

Then using the conditional distribution of  $W$  given the lifetime process  $\zeta$ , we obtain

$$\begin{aligned} \mathbb{E}_{(W_0)} \left[ \sum_{k=0}^{\tau_1-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right] &= \sum_{i=0}^{\infty} \mathbb{E}_{(W_0)} \left[ \sum_{k=0}^{\tau_1-1} \mathbf{1}_{\{\zeta_k=i\}} \mathbf{1}_{\{\widehat{W}_k=0\}} \right] \\ &= \sum_{i=0}^{\infty} \mathbb{E}_{(W_0)} \left[ \sum_{k=0}^{\tau_1-1} \mathbf{1}_{\{\zeta_k=i\}} \right] p_i(-W_0(0)) \\ &= 2 G(-W_0(0)). \end{aligned}$$

(Of course here  $W_0(0) = 0$ , but the previous calculation holds independently of the value of  $W_0(0)$ .) The same argument shows that, for every  $j \in \llbracket 1, m-1 \rrbracket$ ,

$$\mathbb{E}_{(W_0)} \left[ \sum_{k=\tau_j}^{\tau_{j+1}-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right] = 2 G(-W_0(-j))$$

and formula (33) follows.

From (33) and the preceding considerations, we get that, for all sufficiently large  $n$ ,

$$2 \mathbb{E} \left[ \left( \sum_{j=0}^{n(\varepsilon)-1} G(-W_0(-j)) \right) \mathbb{P}_{(W_0)} \left( \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, \lfloor \delta n \rfloor \rrbracket \right) \right] \geq 1 - 2\eta. \quad (34)$$

Now recall that, under the probability measure  $\mathbb{P}$ ,  $(-W_0(-j))_{j \geq 0}$  has the same distribution as  $(S_j)_{j \geq 0}$ . At this point we need two other lemmas.

**Lemma 17.** *For every integer  $p \geq 1$ , there exists a constant  $C(p)$  such that, for every  $n \geq 2$ ,*

$$E \left[ \left( \sum_{j=0}^n G(S_j) \right)^p \right] \leq C(p) (\log n)^p.$$

*Proof.* We first observe that

$$E\left[\left(\sum_{j=0}^n G(S_j)\right)^p\right] = E\left[\left(\sum_{j=0}^n G(S_j) \mathbf{1}_{\{|S_j| \leq n\}}\right)^p\right] + o(1)$$

as  $n \rightarrow \infty$ , because the event where  $\sup\{|S_j| : 0 \leq j \leq n\} > n$  has a probability which decreases to 0 faster than any negative power of  $n$ . For every integer  $k \geq 1$  and  $x \in \mathbb{Z}^4$ , set

$$G_k(x) = \sum_{i=0}^k p_i(x).$$

Using (29) and the standard local limit theorem (see e.g. [14, Chapter 2]) one easily verifies that, for every sufficiently large  $n$ , for all  $x \in \mathbb{Z}^4$  such that  $|x| \leq n$ , the bound  $G_{n^3}(x) \geq \frac{1}{2}G(x)$  holds. Thanks to this observation, it is enough to bound

$$E\left[\left(\sum_{j=0}^n G_{n^3}(S_j) \mathbf{1}_{\{|S_j| \leq n\}}\right)^p\right].$$

However, if  $S'$  stands for another random walk with the same distribution as  $S$  but independent of  $S$ , we have

$$\sum_{j=0}^n G_{n^3}(S_j) = E\left[\sum_{j=0}^n \sum_{i=0}^{n^3} \mathbf{1}_{\{S_j=S'_i\}} \mid S\right],$$

and by Lemma 1 in Marcus and Rosen [22], we know that there exists a constant  $C'(p)$  such that, for every  $n \geq 2$ ,

$$E\left[\left(\sum_{j=0}^n \sum_{i=0}^{n^3} \mathbf{1}_{\{S_j=S'_i\}}\right)^p\right] \leq C'(p) (\log n)^p.$$

The desired bound follows since the conditional expectation is a contraction in  $L^p$ .  $\square$

**Lemma 18.** *For every  $\alpha > 0$ , there exists a constant  $C_\alpha$  such that, for every integer  $m \geq 2$ , we have*

$$P\left(\left|\sum_{k=0}^m G(S_k) - \frac{1}{4\pi^2\sigma^4} \log m\right| \geq \alpha \log m\right) \leq C_\alpha (\log m)^{-3/2}.$$

We postpone the proof of Lemma 18 and complete the proof of Proposition 8. An application of Hölder's inequality gives for  $p \geq 2$ ,

$$\begin{aligned} & E\left[\left(\sum_{j=0}^m G(S_j)\right) \mathbf{1}_{\{\sum_{j=0}^m G(S_j) \geq (\frac{1}{4\pi^2\sigma^4} + \alpha) \log m\}}\right] \\ & \leq E\left[\left(\sum_{j=0}^m G(S_j)\right)^p\right]^{1/p} P\left(\sum_{j=0}^m G(S_j) \geq (\frac{1}{4\pi^2\sigma^4} + \alpha) \log m\right)^{1/q} \\ & \leq C(p)^{1/p} \log m \times P\left(\left|\sum_{j=0}^m G(S_j) - \frac{1}{4\pi^2\sigma^4} \log m\right| \geq \alpha \log m\right)^{1/q} \end{aligned}$$

where  $\frac{1}{p} + \frac{1}{q} = 1$  and we used Lemma 17. Choosing  $p \geq 4$  and using Lemma 18, we obtain that

$$\lim_{m \rightarrow \infty} E\left[\left(\sum_{j=0}^m G(S_j)\right) \mathbf{1}_{\{\sum_{j=0}^m G(S_j) \geq (\frac{1}{4\pi^2\sigma^4} + \alpha) \log m\}}\right] = 0.$$

From (34) and the fact that  $(-W_0(-j))_{j \geq 0}$  has the same distribution as  $(S_j)_{j \geq 0}$ , we then get, for every sufficiently large  $n$ ,

$$\begin{aligned} & 2 \left( \frac{1}{4\pi^2\sigma^4} + \alpha \right) (\log n(\varepsilon)) \mathbb{P} \left( \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, \lfloor \delta n \rfloor \rrbracket \right) \\ & \geq 1 - 2\eta - 2 \mathbb{E} \left[ \left( \sum_{j=0}^{n(\varepsilon)-1} G(-W_0(-j)) \right) \mathbf{1}_{\left\{ \sum_{j=0}^{n(\varepsilon)-1} G(-W_0(-j)) \geq \left( \frac{1}{4\pi^2\sigma^4} + \alpha \right) \log n(\varepsilon) \right\}} \right] \\ & \geq 1 - 3\eta. \end{aligned}$$

Since  $\log n(\varepsilon) \leq (\frac{1}{2} + \varepsilon) \log n$ , the preceding bound implies that

$$\liminf_{n \rightarrow \infty} (\log n) \mathbb{P} \left( \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, \lfloor \delta n \rfloor \rrbracket \right) \geq \frac{1 - 3\eta}{1 + 2\varepsilon} \left( \frac{1}{4\pi^2\sigma^4} + \alpha \right)^{-1}.$$

Now note that the ratio  $\log \lfloor \delta n \rfloor / \log n$  tends to 1 as  $n \rightarrow \infty$ , and that  $\eta$ ,  $\varepsilon$  and  $\alpha$  can be chosen arbitrarily small. We conclude that

$$\liminf_{n \rightarrow \infty} (\log n) \mathbb{P} \left( \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n \rrbracket \right) \geq 4\pi^2\sigma^4.$$

The proof of the analogous result for the limsup behavior is similar. In the same way as we proceeded above, we arrive at the bound

$$\mathbb{E} \left[ \left( \sum_{k=0}^n \mathbf{1}_{\{\widehat{W}_k=0\}} \right) \mathbb{P}_{(W_0)} \left( \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n \rrbracket \right) \right] \leq 1.$$

At this point, we would like to replace the sum from  $k = 0$  to  $n$  by a sum from  $k = 0$  to  $\tau_{n'(\varepsilon)-1}$ , where  $n'(\varepsilon) = \lfloor n^{\frac{1}{2}-\varepsilon} \rfloor$  for some fixed  $\varepsilon \in (0, 1/2)$ . Simple arguments give the existence of a constant  $C'_\varepsilon$  such that, for every integer  $n \geq 1$ ,

$$\mathbb{P}(\tau_{n'(\varepsilon)} \geq n) \leq C'_\varepsilon n^{-\varepsilon/2}.$$

We can then write

$$1 \geq \mathbb{E} \left[ \left( \sum_{k=0}^{\tau_{n'(\varepsilon)}-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right) \mathbb{P}_{(W_0)} \left( \widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n \rrbracket \right) \right] - \mathbb{E} \left[ \mathbf{1}_{\{\tau_{n'(\varepsilon)} \geq n\}} \left( \sum_{k=0}^{\tau_{n'(\varepsilon)}-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right) \right],$$

and by the Cauchy-Schwarz inequality, we have

$$\mathbb{E} \left[ \mathbf{1}_{\{\tau_{n'(\varepsilon)} \geq n\}} \left( \sum_{k=0}^{\tau_{n'(\varepsilon)}-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right) \right] \leq (C'_\varepsilon n^{-\varepsilon/2})^{1/2} \times \mathbb{E} \left[ \left( \sum_{k=0}^{\tau_{n'(\varepsilon)}-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right)^2 \right]^{1/2}. \quad (35)$$

To bound the expectation in the right-hand side, one can verify that, for every integer  $m \geq 1$ ,

$$\mathbb{E} \left[ \left( \sum_{k=0}^{m-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right)^2 \middle| W_0 \right] \leq 4 \left( \sum_{j=0}^{m-1} G(-W_0(-j)) \right)^2 + 4 \sum_{j=0}^{m-1} \Phi(-W_0(-j))$$

where, for every  $x \in \mathbb{Z}^4$ ,

$$\Phi(x) := \sum_{y \in \mathbb{Z}^4} G(y) G(x - y)^2.$$

The proof of the latter bound is similar to that of (33) above, and we leave the details to the reader. One then checks from (29) that there exists a constant  $\tilde{C}$  such that

$$\Phi(x) \leq \tilde{C} (|x| \vee 1)^{-2} (1 + \log(|x| \vee 1)), \quad \text{for every } x \in \mathbb{Z}^4.$$

It easily follows that

$$\mathbb{E} \left[ \left( \sum_{k=0}^{\tau_m-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right)^2 \right] = O((\log m)^2)$$

as  $m \rightarrow \infty$ . Consequently the right-hand side of (35) tends to 0 as  $n \rightarrow \infty$  and if  $\eta > 0$  is fixed, we have, for all  $n$  sufficiently large,

$$\mathbb{E} \left[ \left( \sum_{k=0}^{\tau_{n'(\varepsilon)}-1} \mathbf{1}_{\{\widehat{W}_k=0\}} \right) \mathbb{P}_{(W_0)}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n \rrbracket) \right] \leq 1 + \eta.$$

Just as we obtained (34), we deduce from the latter bound that

$$2 \mathbb{E} \left[ \left( \sum_{j=0}^{n'(\varepsilon)-1} G(-W_0(-j)) \right) \mathbb{P}_{(W_0)}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n \rrbracket) \right] \leq 1 + \eta. \quad (36)$$

Then fix  $\alpha \in (0, (4\pi^2\sigma^4)^{-1})$ . It follows from (36) that

$$\begin{aligned} & 2 \left( \frac{1}{4\pi^2\sigma^4} - \alpha \right) (\log n'(\varepsilon)) \mathbb{E} \left[ \mathbf{1}_{\left\{ \sum_{j=0}^{n'(\varepsilon)-1} G(-W_0(-j)) \geq \left( \frac{1}{4\pi^2\sigma^4} - \alpha \right) \log n'(\varepsilon) \right\}} \mathbb{P}_{(W_0)}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n \rrbracket) \right] \\ & \leq 2 \mathbb{E} \left[ \left( \sum_{j=0}^{n'(\varepsilon)-1} G(-W_0(-j)) \right) \mathbb{P}_{(W_0)}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n \rrbracket) \right] \\ & \leq 1 + \eta. \end{aligned}$$

On the other hand,

$$(\log n'(\varepsilon)) \mathbb{P} \left( \sum_{j=0}^{n'(\varepsilon)-1} G(-W_0(-j)) < \left( \frac{1}{4\pi^2\sigma^4} - \alpha \right) \log n'(\varepsilon) \right) \xrightarrow{n \rightarrow \infty} 0$$

by Lemma 18. By combining the last two displays, we get

$$\limsup_{n \rightarrow \infty} 2 \left( \frac{1}{4\pi^2\sigma^4} - \alpha \right) (\log n'(\varepsilon)) \mathbb{P}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n \rrbracket) \leq 1 + \eta.$$

Since  $\eta$ ,  $\varepsilon$  and  $\alpha$  can be chosen arbitrarily small, we get

$$\limsup_{n \rightarrow \infty} (\log n) \mathbb{P}(\widehat{W}_\ell \neq 0, \forall \ell \in \llbracket 1, n \rrbracket) \leq 4\pi^2\sigma^4,$$

which completes the proof of the first assertion of Proposition 8. The second assertion is an easy consequence of the first one, noting that, for every  $\varepsilon > 0$ , both  $\mathbb{P}(\tau_p \geq p^{2+\varepsilon})$  and  $\mathbb{P}(\tau_p \leq p^{2-\varepsilon})$  are  $o((\log p)^{-1})$  as  $p \rightarrow \infty$ .  $\square$

*Proof of Lemma 18.* The general strategy of the proof is to derive an analogous result for Brownian motion in  $\mathbb{R}^4$ , and then to use a strong invariance principle to transfer this result to the random walk  $S$ .

We let  $B = (B_t)_{t \geq 0}$  be a four-dimensional Brownian motion started from 0 and set  $\rho_t = |B_t|$  for every  $t \geq 0$ , so that  $(\rho_t)_{t \geq 0}$  is a four-dimensional Bessel process started from 0. Here is the Brownian motion version of Lemma 18.

**Lemma 19.** *Let  $\varepsilon > 0$ . There exist two constants  $C(\varepsilon)$  and  $\beta(\varepsilon) > 0$  such that, for every  $t > r \geq 1$ ,*

$$P\left(\left|\int_r^t \frac{ds}{\rho_s^2} - \frac{1}{2} \log\left(\frac{t}{r}\right)\right| > \varepsilon \log\left(\frac{t}{r}\right)\right) \leq C(\varepsilon) \left(\frac{t}{r}\right)^{-\beta(\varepsilon)}.$$

Let us postpone the proof of Lemma 19. We fix  $\alpha > 0$  and consider an integer  $n \geq 1$ . By an extension due to Zaitsev [26] of the celebrated Komlós-Major-Tusnády strong invariance principle, we can construct on the same probability space the finite sequence  $(S_1, \dots, S_n)$  and the Brownian motion  $(B_t)_{t \geq 0}$ , in such a way that, for some constants  $c > 0, c' > 0$  and  $K > 0$  that do not depend on  $n$ , we have

$$E\left[\exp\left(c \max_{1 \leq k \leq n} |S_k - \sigma B_k|\right)\right] \leq K \exp(c' \log n).$$

It readily follows that we can find constants  $C > 0$  and  $a > 0$  (again independent of  $n$ ) such that

$$P\left(\max_{1 \leq k \leq n} |S_k - \sigma B_k| > C \log n\right) \leq K n^{-a}.$$

Let  $A > 2$  be a constant. Then

$$P\left(\inf_{t \geq (\log n)^4} \sigma |B_t| \leq AC \log n\right) = P\left(\inf_{t \geq 1} \sigma |B_t| \leq \frac{AC}{\log n}\right) = O((\log n)^{-2})$$

by an easy estimate. On the event

$$E_n := \left\{ \max_{1 \leq k \leq n} |S_k - \sigma B_k| \leq C \log n \right\} \cap \left\{ \inf_{t \geq (\log n)^4} \sigma |B_t| > AC \log n \right\}$$

we have, for every integer  $k$  such that  $(\log n)^4 \leq k \leq n$ ,

$$|S_k| \geq \sigma |B_k| - C \log n \geq (1 - \eta) \sigma |B_k|$$

and

$$|S_k| \leq \sigma |B_k| + C \log n \leq (1 + \eta) \sigma |B_k|$$

where  $\eta = 1/A$ . We now fix  $A$  so that  $\eta \in (0, \frac{1}{5})$  and  $5\eta < \pi^2 \sigma^4 \alpha / 2$ .

Recalling our estimate (29), we also see that (provided  $n$  is large enough) we have on the event  $E_n$ , for every integer  $k$  such that  $(\log n)^4 \leq k \leq n$ ,

$$(1 - 3\eta) \frac{1}{2\pi^2 \sigma^4} |B_k|^{-2} \leq G(S_k) \leq (1 + 3\eta) \frac{1}{2\pi^2 \sigma^4} |B_k|^{-2}.$$

Consequently, we have on the event  $E_n$ ,

$$(1 - 3\eta) \frac{1}{2\pi^2 \sigma^4} \sum_{k=\lceil (\log n)^4 \rceil}^n |B_k|^{-2} \leq \sum_{k=\lceil (\log n)^4 \rceil}^n G(S_k) \leq (1 + 3\eta) \frac{1}{2\pi^2 \sigma^4} \sum_{k=\lceil (\log n)^4 \rceil}^n |B_k|^{-2}.$$

The next step is to observe that

$$\sum_{k=\lceil (\log n)^4 \rceil}^n |B_k|^{-2}$$

is close to

$$\int_{\lceil (\log n)^4 \rceil}^{n+1} \frac{ds}{|B_s|^2}$$

up to a set of small probability. Indeed simple estimates show that, for any choice of  $\kappa > 0$ , we have

$$\sup_{0 \leq k \leq n} \sup_{k \leq s \leq k+1} |B_s - B_k| \leq \kappa \log n$$

outside of a set of probability  $O(n^{-1})$ . By choosing  $\kappa$  suitably, we then see that on the event

$$\tilde{E}_n := E_n \cap \left\{ \sup_{0 \leq k \leq n} \sup_{k \leq s \leq k+1} |B_s - B_k| \leq \kappa \log n \right\}$$

we have

$$(1 - \eta) \int_{\lceil (\log n)^4 \rceil}^{n+1} \frac{ds}{|B_s|^2} \leq \sum_{k=\lceil (\log n)^4 \rceil}^n |B_k|^{-2} \leq (1 + \eta) \int_{\lceil (\log n)^4 \rceil}^{n+1} \frac{ds}{|B_s|^2},$$

and consequently

$$(1 - 5\eta) \frac{1}{2\pi^2\sigma^4} \int_{\lceil (\log n)^4 \rceil}^{n+1} \frac{ds}{|B_s|^2} \leq \sum_{k=\lceil (\log n)^4 \rceil}^n G(S_k) \leq (1 + 5\eta) \frac{1}{2\pi^2\sigma^4} \int_{\lceil (\log n)^4 \rceil}^{n+1} \frac{ds}{|B_s|^2}. \quad (37)$$

We also need to bound the quantity

$$\sum_{k=0}^{\lceil (\log n)^4 \rceil - 1} G(S_k).$$

However, from Lemma 17 with  $p = 2$ , we immediately get that, for every integer  $m \geq 2$  and every  $h > 0$ ,

$$P\left(\sum_{k=0}^m G(S_k) \geq h\right) \leq \frac{C(2)(\log m)^2}{h^2}. \quad (38)$$

Finally,

$$\begin{aligned} & P\left(\left|\sum_{k=0}^n G(S_k) - \frac{1}{4\pi^2\sigma^4} \log n\right| \geq \alpha \log n\right) \\ & \leq P\left(\sum_{k=0}^{\lceil (\log n)^4 \rceil} G(S_k) \geq \frac{\alpha}{2} \log n\right) + P\left(\left|\sum_{k=\lceil (\log n)^4 \rceil}^n G(S_k) - \frac{1}{4\pi^2\sigma^4} \log n\right| \geq \frac{\alpha}{2} \log n\right). \end{aligned}$$

The first term in the right-hand side is  $O((\log n)^{-3/2})$  by (38). On the other hand, by (37), the second term is bounded by

$$P(\tilde{E}_n^c) + P\left(\left|\int_{\lceil (\log n)^4 \rceil}^{n+1} \frac{ds}{|B_s|^2} - \frac{1}{2} \log n\right| \geq \alpha' \log n\right)$$

where  $\alpha' = (\frac{1}{2}\pi^2\sigma^4\alpha) \wedge \frac{1}{4}$  is a constant independent of  $n$ , which satisfies

$$(1 + 5\eta)\left(\frac{1}{2} + \alpha'\right) \frac{1}{2\pi^2\sigma^4} < \frac{1}{4\pi^2\sigma^4} + \frac{\alpha}{2} \quad \text{and} \quad (1 - 5\eta)\left(\frac{1}{2} - \alpha'\right) \frac{1}{2\pi^2\sigma^4} > \frac{1}{4\pi^2\sigma^4} - \frac{\alpha}{2}.$$

(Here we use our choice of  $\eta$  such that  $5\eta < \pi^2\sigma^4\alpha/2$ .) From preceding estimates, we have  $P(\tilde{E}_n^c) = O((\log n)^{-2})$ . On the other hand, Lemma 19 implies that

$$P\left(\left|\int_{\lceil (\log n)^4 \rceil}^{n+1} \frac{ds}{|B_s|^2} - \frac{1}{2} \log n\right| \geq \alpha' \log n\right) = O(n^{-b})$$

for some  $b > 0$ . This completes the proof of Lemma 18.  $\square$

*Proof of Lemma 19.* By a scaling argument, it is enough to consider the case  $r = 1$ , and we consider only that case. For every integer  $k \geq 0$ , set

$$\gamma_k := \inf \{t \geq 0 : \rho_t = e^k\}$$

and

$$X_k := \int_{\gamma_k}^{\gamma_{k+1}} \frac{ds}{\rho_s^2}.$$

A scaling argument shows that the variables  $X_k$ ,  $k \geq 0$  are identically distributed. Moreover, the strong Markov property of the Bessel process implies that the variables  $X_k$ ,  $k \geq 0$  are independent. Furthermore, the absolute continuity relations between Bessel processes can be used to verify that these variables have small exponential moments. More precisely, using the explicit form of the density of the law over the time interval  $[0, t]$  of the four-dimensional Bessel process started at 1 with respect to Wiener measure (see question 3 in Exercise XI.1.22 of Revuz and Yor [25]), it is an easy exercise of martingale theory to verify that

$$E[e^{3X_0/8}] = E\left[\exp\left(\frac{3}{8} \int_{\gamma_0}^{\gamma_1} \frac{ds}{\rho_s^2}\right)\right] = \sqrt{e} < \infty.$$

Set

$$c_0 = E[X_0] = E[X_k]$$

for every  $k \geq 0$ . We can apply Cramér's large deviation theorem to the sequence  $(X_k)_{k \geq 0}$ . It follows that, for every  $\delta > 0$ , there exists a constant  $b(\delta) > 0$  such that for every sufficiently large  $n$ ,

$$P\left(\left|\int_{\gamma_0}^{\gamma_n} \frac{ds}{\rho_s^2} - c_0 n\right| > \delta n\right) \leq \exp(-b(\delta)n). \quad (39)$$

On the other hand, it is easy to verify that the variable

$$\int_1^{\gamma_0} \frac{ds}{\rho_s^2}$$

has exponential moments. Just use the above-mentioned argument involving the density of the law of the Bessel process to verify that

$$E\left[\exp\left(\frac{3}{8} \int_1^{\gamma_0} \frac{ds}{\rho_s^2}\right)\right] < \infty$$

(deal separately with the cases  $1 < \gamma_0$  and  $\gamma_0 < 1$ ). It then follows that, for every  $\delta > 0$ , and for all sufficiently large  $n$ ,

$$P\left(\int_1^{\gamma_0} \frac{ds}{\rho_s^2} > \delta n\right) \leq \exp(-b'(\delta)n)$$

with some constant  $b'(\delta) > 0$ . The same bound holds for the variable

$$\int_{e^{2m}}^{\gamma_m} \frac{ds}{\rho_s^2},$$

for any integer  $m \geq 0$ , since this variable has the same law as

$$\int_1^{\gamma_0} \frac{ds}{\rho_s^2}$$

by scaling.

By combining the latter facts with (39), we obtain that, for every  $\delta > 0$ , there exists a constant  $\tilde{b}(\delta) > 0$  such that, for every sufficiently large  $n$ ,

$$P\left(\left|\int_1^{e^{2n}} \frac{ds}{\rho_s^2} - c_0 n\right| > \delta n\right) \leq \exp(-\tilde{b}(\delta)n). \quad (40)$$

At this stage, we can identify the constant  $c_0$ , since the preceding arguments also show that

$$c_0 = \lim_{n \rightarrow \infty} \frac{1}{n} E\left[\int_1^{e^{2n}} \frac{ds}{\rho_s^2}\right] = 1$$

by a direct calculation of  $E[(\rho_s)^{-2}] = (2s)^{-1}$ . Once we know that  $c_0 = 1$ , the statement of Lemma 19 follows from (40) by elementary considerations: For every  $t \geq 1$ , choose  $n$  such that  $e^{2n} \leq t < e^{2(n+1)}$  and observe that

$$\left\{\int_1^t \frac{ds}{\rho_s^2} - \frac{1}{2} \log t > \varepsilon \log t\right\} \subseteq \left\{\int_1^{e^{2(n+1)}} \frac{ds}{\rho_s^2} - n > 2\varepsilon n\right\},$$

whereas

$$\left\{\int_1^t \frac{ds}{\rho_s^2} - \frac{1}{2} \log t < -\varepsilon \log t\right\} \subseteq \left\{\int_1^{e^{2n}} \frac{ds}{\rho_s^2} - n - 1 < -2\varepsilon n\right\}.$$

This completes the proof.  $\square$

## 4 The range of branching random walk

In this last section, we apply the preceding results to asymptotics for the range of branching random walk in  $\mathbb{Z}^d$ ,  $d \geq 4$ . We assume that the offspring distribution  $\mu$  is critical and has finite variance  $\sigma_\mu^2 > 0$ , and that the jump distribution  $\theta$  is centered and has finite moments of order  $d - 1$  (and as usual that  $\theta$  is not supported on a strict subgroup of  $\mathbb{Z}^d$ ).

Let  $M_p(\mathbb{Z}^d)$  stand for the set of all finite point measures on  $\mathbb{Z}^d$ . Let  $\mathcal{Z} = (\mathcal{Z}_n)_{n \geq 0}$  denote the (discrete time) branching random walk with jump distribution  $\theta$  and offspring distribution  $\mu$ . This is the Markov chain with values in  $M_p(\mathbb{Z}^d)$ , whose transition kernel  $\mathbf{Q}$  can be described as follows. If

$$\omega = \sum_{i=1}^p \delta_{x_i} \in M_p(\mathbb{Z}^d),$$

$\mathbf{Q}(\omega, \cdot)$  is the distribution of

$$\sum_{i=1}^p \sum_{j=1}^{\xi_i} \delta_{x_i + Y_{i,j}},$$

where  $\xi_1, \dots, \xi_p$  are independent and distributed according to  $\mu$  and, conditionally on  $(\xi_1, \dots, \xi_p)$ , the random variables  $Y_{i,j}$ ,  $1 \leq i \leq p$ ,  $1 \leq j \leq \xi_i$ , are independent and distributed according to  $\theta$ . More informally, each particle alive at time  $n$  is replaced at time  $n + 1$  by a number of offspring distributed according to  $\mu$ , and the spatial position of each of these offspring is obtained by adding a jump distributed according to  $\theta$  to the position of its parent.

The range of  $\mathcal{Z}$  is then defined by

$$\mathbf{R}(\mathcal{Z}) := \#\{x \in \mathbb{Z}^d : \exists n \geq 0, \mathcal{Z}_n(x) \geq 1\}.$$

We also write  $\mathbf{N}(\mathcal{Z})$  for the total progeny of  $\mathcal{Z}$ ,

$$\mathbf{N}(\mathcal{Z}) := \sum_{n=0}^{\infty} \langle \mathcal{Z}_n, 1 \rangle$$

where  $\langle \mathcal{Z}_n, 1 \rangle$  is the total mass of  $\mathcal{Z}_n$ . It is well known (and easy to prove using the Lukasiewicz path introduced in the proof of Theorem 7) that  $\mathbf{N}(\mathcal{Z})$  has the distribution of the hitting time of  $-\langle \mathcal{Z}_0, 1 \rangle$  by a random walk on  $\mathbb{Z}$  with jump distribution  $\nu(k) = \mu(k+1)$ , for  $k = -1, 0, 1, \dots$ , started from 0.

**Proposition 20.** *Suppose that  $d \geq 5$ . For every integer  $p \geq 1$ , let  $\mathcal{Z}^{(p)}$  be a branching random walk with jump distribution  $\theta$  and offspring distribution  $\mu$ , such that  $\langle \mathcal{Z}_0^{(p)}, 1 \rangle = p$ . Then,*

$$\lim_{p \rightarrow \infty} \frac{\mathbf{R}(\mathcal{Z}^{(p)})}{\mathbf{N}(\mathcal{Z}^{(p)})} = c_{\mu, \theta} \quad \text{in probability,}$$

where  $c_{\mu, \theta} > 0$  is the constant in Theorem 4. Consequently,

$$\frac{1}{p^2} \mathbf{R}(\mathcal{Z}^{(p)}) \xrightarrow[p \rightarrow \infty]{(d)} \frac{c_{\mu, \theta}}{\sigma_{\mu}^2} J,$$

where the positive random variable  $J$  has density  $(2\pi s^3)^{-1/2} \exp(-\frac{1}{2s})$  on  $(0, \infty)$ .

*Proof.* We may and will assume that there exists a sequence  $\mathcal{T}^1, \mathcal{T}^2, \dots$  of independent random trees distributed according to  $\Pi_{\mu}$ , such that, for every  $p \geq 1$ , the genealogy of  $\mathcal{Z}^{(p)}$  is coded by  $\mathcal{T}^1, \mathcal{T}^2, \dots, \mathcal{T}^p$ , meaning that  $\mathcal{T}^i$  is the genealogical tree of the descendants of the  $i$ -th initial particle of  $\mathcal{Z}^{(p)}$ , for every  $p \geq 1$  and  $i \in \{1, \dots, p\}$ . Notice that we have then

$$\mathbf{N}(\mathcal{Z}^{(p)}) = \#\mathcal{T}^1 + \dots + \#\mathcal{T}^p.$$

For every  $i \in \{1, \dots, p\}$ , we will write  $\mathcal{S}_i^{(p)}$  for the set of all spatial locations occupied by the particles of  $\mathcal{Z}^{(p)}$  that are descendants of the  $i$ -th initial particle. Note that the location of the  $i$ -th initial particle may depend on  $p$ . Clearly, we have

$$\mathbf{R}(\mathcal{Z}^{(p)}) \leq \#\mathcal{S}_1^{(p)} + \dots + \#\mathcal{S}_p^{(p)}. \quad (41)$$

Let  $(H_k)_{k \geq 0}$  be the height process associated with the sequence  $\mathcal{T}^1, \mathcal{T}^2, \dots$  (see the proof of Proposition 6). Then, as an easy consequence of (3), we have the joint convergence in distribution

$$\left( \left( \frac{1}{p} H_{\lfloor p^2 t \rfloor \wedge \mathbf{N}(\mathcal{Z}^{(p)})} \right)_{t \geq 0}, \frac{1}{p^2} \mathbf{N}(\mathcal{Z}^{(p)}) \right) \xrightarrow[p \rightarrow \infty]{(d)} \left( \left( \frac{2}{\sigma_{\mu}} |\beta_{t \wedge J_{1/\sigma_{\mu}}}| \right)_{t \geq 0}, J_{1/\sigma_{\mu}} \right), \quad (42)$$

where  $\beta$  is a standard linear Brownian motion, and for every  $s \geq 0$ ,  $J_s = \inf\{t \geq 0 : L_t^0(\beta) > s\}$ , where  $(L_t^0(\beta))_{t \geq 0}$  is the local time process of  $\beta$  at level 0. See [16, Section 1.4] for details of the derivation of (42).

Fix  $\varepsilon > 0$ . For  $\alpha \in (0, 1)$ , let  $i_{p,1}, i_{p,2}, \dots, i_{p,m_p}$  be all indices  $i \in \{1, \dots, p\}$  such that  $\#\mathcal{T}^i \geq \alpha p^2$ . It follows from (42) that, if  $\alpha$  has been chosen sufficiently small, the bound

$$\mathbf{N}(\mathcal{Z}^{(p)}) - (\#\mathcal{T}^{i_{p,1}} + \dots + \#\mathcal{T}^{i_{p,m_p}}) = \sum_{i \in \{1, \dots, p\} \setminus \{i_{p,1}, \dots, i_{p,m_p}\}} \#\mathcal{T}^i < \varepsilon p^2 \quad (43)$$

will hold with probability arbitrarily close to 1, uniformly for all sufficiently large  $p$ . On the other hand, it also follows from (42) that  $m_p$  converges in distribution as  $p \rightarrow \infty$  to a Poisson distribution with parameter  $\sigma_\mu^{-1} \sqrt{2/\pi\alpha}$  (here the quantity  $\sqrt{2/\pi\alpha}$  is the mass that the Itô excursion measure assigns to excursions of length greater than  $\alpha$ ). In particular, by choosing  $\alpha$  even smaller if necessary, we have  $P(m_p \geq 1) > 1 - \varepsilon$  for all  $p$  large enough. We now fix  $\alpha > 0$  so that the preceding properties hold for all  $p$  large enough.

Next we observe that, conditionally on  $m_p$ , the trees  $\mathcal{T}^{i_{p,1}}, \dots, \mathcal{T}^{i_{p,m_p}}$  are independent and distributed according to  $\Pi_\mu(\cdot \mid \#\mathcal{T} \geq \alpha p^2)$ . From Theorem 7, we now get that

$$P\left(\left|\frac{\#\mathcal{S}_{i_{p,1}}^{(p)} + \dots + \#\mathcal{S}_{i_{p,m_p}}^{(p)}}{\#\mathcal{T}^{i_{p,1}} + \dots + \#\mathcal{T}^{i_{p,m_p}}} - c_{\mu,\theta}\right| > \varepsilon \mid m_p \geq 1\right) \xrightarrow{p \rightarrow \infty} 0. \quad (44)$$

Then, on the one hand, we have from (41),

$$R(\mathcal{Z}^{(p)}) \leq \#\mathcal{S}_{i_{p,1}}^{(p)} + \dots + \#\mathcal{S}_{i_{p,m_p}}^{(p)} + \sum_{i \in \{1, \dots, p\} \setminus \{i_{p,1}, \dots, i_{p,m_p}\}} \#\mathcal{T}^i,$$

and on the other hand,

$$R(\mathcal{Z}^{(p)}) \geq \#\mathcal{S}_{i_{p,1}}^{(p)} + \dots + \#\mathcal{S}_{i_{p,m_p}}^{(p)} - \sum_{1 \leq k < \ell \leq m_p} \#(\mathcal{S}_{i_{p,k}}^{(p)} \cap \mathcal{S}_{i_{p,\ell}}^{(p)}).$$

Taking into account the bound (43) and the fact that  $p^{-2}\mathbf{N}(\mathcal{Z}^{(p)})$  converges in distribution to a positive random variable, we see that the first assertion of the proposition will follow from the last two bounds and (44), provided we can verify that

$$\frac{1}{p^2} \sum_{1 \leq k < \ell \leq m_p} \#(\mathcal{S}_{i_{p,k}}^{(p)} \cap \mathcal{S}_{i_{p,\ell}}^{(p)}) \xrightarrow{p \rightarrow \infty} 0. \quad (45)$$

Recall that  $m_p$  converges in distribution to a finite random variable. In order to establish (45), it is enough to verify that, if  $\mathcal{S}^{(p),1}$ , respectively  $\mathcal{S}^{(p),2}$ , is the set of points visited by a random walk indexed by a tree distributed according to  $\Pi_\mu(\cdot \mid \#\mathcal{T} \geq \alpha p^2)$ , with the spatial location of the root equal to  $x_1$ , resp. to  $x_2$ , and if  $\mathcal{S}^{(p),1}$  and  $\mathcal{S}^{(p),2}$  are independent, we have

$$\frac{1}{p^2} E\left[\#(\mathcal{S}^{(p),1} \cap \mathcal{S}^{(p),2})\right] \xrightarrow{p \rightarrow \infty} 0.$$

However,

$$E\left[\#(\mathcal{S}^{(p),1} \cap \mathcal{S}^{(p),2})\right] = \sum_{y \in \mathbb{Z}^d} P(y \in \mathcal{S}^{(p),1}) P(y \in \mathcal{S}^{(p),2}) \leq \sum_{y \in \mathbb{Z}^d} P(y \in \mathcal{S}^{(p),1})^2,$$

using the Cauchy-Schwarz inequality and translation invariance, which also allows us to take  $x_1 = 0$ . By a first moment argument, we have then

$$P(y \in \mathcal{S}^{(p),1}) \leq \frac{G_\theta(y)}{\Pi_\mu(\#\mathcal{T} \geq \alpha p^2)} \wedge 1 \leq (c_\mu^{-1} \sqrt{\alpha} p G_\theta(y)) \wedge 1,$$

where the constant  $c_\mu > 0$  depends only on  $\mu$ . Here we used the classical bound

$$\Pi_\mu(\#\mathcal{T} \geq k) \geq c_\mu k^{-1/2}, \quad k \geq 1,$$

which follows from the fact that the distribution of  $\#\mathcal{T}$  under  $\Pi_\mu$  coincides with the law of the first hitting time of  $-1$  by a random walk on  $\mathbb{Z}$  with jump distribution  $\nu$  started from 0 (see the proof of Theorem 7). Finally, we have

$$\frac{1}{p^2} E[\#(\mathcal{S}^{(p),1} \cap \mathcal{S}^{(p),2})] \leq \sum_{y \in \mathbb{Z}^d} (c_{(\mu)}^{-2} \alpha G_\theta(y)^2) \wedge \frac{1}{p^2}$$

and the right-hand side tends to 0 as  $p \rightarrow \infty$  by dominated convergence, noting that

$$\sum_{y \in \mathbb{Z}^d} G_\theta(y)^2 < \infty$$

by (2). This completes the proof of the first assertion of the proposition.

The second assertion follows from the first one and the convergence in distribution of  $p^{-2}\mathbf{N}(\mathcal{Z}^{(p)})$  to  $J_{1/\sigma_\mu}$ . Just note that  $J_{1/\sigma_\mu}$  has the same law as  $\sigma_\mu^{-2}J_1$  by scaling, and that  $J_1$  is distributed as the first hitting of 1 by a standard linear Brownian motion, whose density is as stated in the proposition.  $\square$

We now state the result corresponding to Proposition 20 in the critical dimension  $d = 4$ . As previously, we must restrict our attention to the geometric offspring distribution.

**Proposition 21.** *Suppose that  $d = 4$ , and that  $\mu$  is the critical geometric offspring distribution. Also assume that  $\theta$  is symmetric and has small exponential moments, and set  $\sigma^2 = (\det M_\theta)^{1/4}$ . For every integer  $p \geq 1$ , let  $\mathcal{Z}^{(p)}$  be a branching random walk with jump distribution  $\theta$  and offspring distribution  $\mu$ , such that  $\langle \mathcal{Z}_0^{(p)}, 1 \rangle = p$ . Then,*

$$\lim_{p \rightarrow \infty} \frac{(\log p) \mathbf{R}(\mathcal{Z}^{(p)})}{\mathbf{N}(\mathcal{Z}^{(p)})} = 8\pi^2 \sigma^4 \quad \text{in probability.}$$

Consequently,

$$\frac{\log p}{p^2} \mathbf{R}(\mathcal{Z}^{(p)}) \xrightarrow[p \rightarrow \infty]{(d)} 4\pi^2 \sigma^4 J,$$

where  $J$  is as in Proposition 20.

The proof of Proposition 21 goes along the same lines as that of Proposition 20, using now Theorem 14 instead of Theorem 7. A few minor modifications are needed, but we will leave the details to the reader.

*Acknowledgements.* The first author would like to thank Itai Benjamini for suggesting the study of the range of the discrete snake a few years ago. We are also indebted to an anonymous referee for a number of useful comments.

## References

- [1] R. ABRAHAM, W. WERNER, Avoiding probabilities for Brownian snakes and super-Brownian motion. *Electron. J. Probab.* 2 (1997), no. 3, 27 pp.
- [2] J.-F. DELMAS, Some properties of the range of super-Brownian motion. *Probab. Theory Rel. Fields* 114 (1999), 505–547.

- [3] T. DUQUESNE, J.-F. LE GALL, Random Trees, Lévy Processes and Spatial Branching Processes. *Astérisque* 281 (2002)
- [4] A. DVORETZKY, P. ERDÖS, Some problems on random walk in space. Proceedings Second Berkeley Symposium on Math. Statistics and Probability, 353-367. University of California Press, Berkeley 1951.
- [5] N.C. JAIN, W.E. PRUITT, The range of transient random walk. *J. Analyse Math.* 24 (1971), 369–393.
- [6] N.C. JAIN, W.E. PRUITT, The range of random walk. Proceedings Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. III: Probability theory, pp. 31–50. Univ. California Press, Berkeley, 1972.
- [7] S. JANSON, Random cutting and records in deterministic and random trees. *Random Structures Algorithms* 29 (2006), 139–179.
- [8] S. JANSON, J.-F. MARCKERT, Convergence of discrete snakes. *J. Theoret. Probab.* 18 (2005), 615–647.
- [9] H. KESTEN, Branching random walk with a critical branching part. *J. Theoret. Probab.* 8 (1995), 921–962.
- [10] J.F.C. KINGMAN, The ergodic theory of subadditive stochastic processes. *J. R. Stat. Soc. Ser. B* 30 (1968), 499–510.
- [11] I. KORTCHEMSKI, A simple proof of Duquesne’s theorem on contour processes of conditioned Galton-Watson trees. Séminaire de Probabilités XLV. Lecture Notes in Mathematics 2078, 537–558. Springer-Verlag, Berlin, 2013.
- [12] S. LALLEY, X. ZHENG, Occupation statistics of critical branching random walks in two or higher dimensions. *Ann. Probab.* 39 (2011), 327–368.
- [13] G.F. LAWLER, personal communication.
- [14] G.F. LAWLER, V. LIMIC, *Random Walk: A Modern Introduction*. Cambridge Studies in Advanced Mathematics 123. Cambridge University Press, Cambridge, 2010.
- [15] J.-F. LE GALL, Propriétés d’intersection des marches aléatoires I. *Comm. Math. Physics* 104 (1986), 471-507.
- [16] J.-F. LE GALL, Random trees and applications. *Probab. Surveys* 2 (2005), 245-311.
- [17] J.-F. LE GALL, Itô’s excursion theory and random trees. *Stoch. Process. Appl.* 120 (2010), 721-749.
- [18] J.-F. LE GALL, S. LIN, The range of tree-indexed random walk in low dimensions. *Ann. Probab.*, to appear.
- [19] J.-F. LE GALL, J. ROSEN, The range of stable random walks. *Ann. Probab.* 16 (1991), 650–705.
- [20] S. LIN, The range of tree-indexed random walk with drift. In preparation.

- [21] R. LYONS, R. PEMANTLE, Y. PERES, Conceptual proofs of  $L \log L$  criteria for mean behavior of branching processes. *Ann. Probab.* 23 (1995), 1125–1138.
- [22] M.B. MARCUS, J. ROSEN, Laws of the iterated logarithm for intersections of random walks on  $\mathbb{Z}^4$ . *Ann. Inst. H. Poincaré Probab. Stat.* 33 (1997), 37–63.
- [23] J.W. PITMAN, One-dimensional Brownian motion and the three-dimensional Bessel process, *Adv. Appl. Probab.* 7 (1975), 511–526.
- [24] J.W. PITMAN, *Combinatorial Stochastic Processes*. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. Lecture Notes in Mathematics 1875. Springer-Verlag, Berlin, 2006.
- [25] D. REVUZ, M. YOR, *Continuous Martingales and Brownian Motion*. Springer-Verlag, Berlin, 1991.
- [26] A.Y. ZAITSEV, Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM Probab. Stat.* 2 (1998), 41–108.