Abstract. We describe scaling limits of recurrent excited random walks (ERWs) on \( \mathbb{Z} \) in i.i.d. cookie environments with a bounded number of cookies per site. We allow both positive and negative excitations. It is known that ERW is recurrent if and only if the expected total drift per site, \( \delta \), belongs to the interval \([-1, 1]\). We show that if \(|\delta| < 1\) then the diffusively scaled ERW under the averaged measure converges to a \((\delta, -\delta)\)-perturbed Brownian motion. In the boundary case, \(|\delta| = 1\), the space scaling has to be adjusted by an extra logarithmic term, and the weak limit of ERW happens to be a constant multiple of the running maximum of the standard Brownian motion, a transient process.

1. Introduction and main results

Given an arbitrary positive integer \( M \) let
\[
\Omega_M := \left\{ (\omega_z(i))_{i \in \mathbb{N}} \mid \omega_z(i) \in [0, 1], \text{ for } i \in \{1, 2, \ldots, M\} \right\}
\]
and \( \omega_z(i) = 1/2 \), for \( i > M, \ z \in \mathbb{Z} \).

An element of \( \Omega_M \) is called a cookie environment. For each \( z \in \mathbb{Z} \), the sequence \( \{\omega_z(i)\}_{i \in \mathbb{N}} \) can be thought of as a pile of cookies at site \( z \). The number \( \omega_z(i) \) represents the transition probability from \( z \) to \( z + 1 \) of a nearest-neighbor random walk upon the \( i \)-th visit to \( z \). If \( \omega_z(i) \geq 1/2 \) (resp. \( \omega_z(i) < 1/2 \)) the corresponding cookie is called non-negative (resp. negative).

Let \( \mathbb{P} \) be a probability measure on \( \Omega_M \), which satisfies the following two conditions:

(A1) Independence: the sequence \( (\omega_z(\cdot))_{z \in \mathbb{Z}} \) is i.i.d. under \( \mathbb{P} \);

(A2) Non-degeneracy: \( \mathbb{E} \left[ \prod_{i=1}^{M} \omega_0(i) \right] > 0 \) and \( \mathbb{E} \left[ \prod_{i=1}^{M} (1 - \omega_0(i)) \right] > 0 \).

For \( x \in \mathbb{Z} \) and \( \omega \in \Omega_M \) consider an integer valued process \( X := (X_j), \ j \geq 0, \) on some probability space \( (\mathcal{X}, \mathcal{F}, \mathbb{P}_{x,\omega}) \), which \( \mathbb{P}_{x,\omega}\text{-a.s.} \) satisfies \( \mathbb{P}_{x,\omega}(X_0 = x) = 1 \) and
\[
P_{x,\omega}(X_{n+1} = X_n + 1 \mid \mathcal{F}_n) = 1 - P_{x,\omega}(X_{n+1} = X_n - 1 \mid \mathcal{F}_n) = \omega_{X_n}(L_{X_n}(n)),
\]
where \( \mathcal{F}_n \subset \mathcal{F}, \ n \geq 0, \) is the natural filtration of \( X \) and \( L_m(n) := \sum_{j=0}^{n} \mathbb{1}_{\{X_j = m\}} \) is the number of visits to site \( m \) by \( X \) up to time \( n \). Informally speaking, upon
each visit to a site the walker eats the topmost cookie from the pile at that site and makes one step to the right or to the left with probabilities prescribed by this cookie. Since \( \omega_z(i) = \frac{1}{2} \) for all \( i > M \), the walker will make unbiased steps from \( z \) starting from the \((M + 1)\)-th visit to \( z \). The consumption of a cookie \( \omega_z(i) \) induces a drift of size \( 2\omega_z(i) - 1 \) with respect to \( P_{x,\omega} \). Let \( \delta \) be the expected total drift per site, i.e.

\[
\delta := \mathbb{E} \left[ \sum_{i \geq 1} (2\omega_0(i) - 1) \right] = \mathbb{E} \left[ \sum_{i=1}^{M} (2\omega_0(i) - 1) \right].
\]

The parameter \( \delta \) plays a key role in the classification of the asymptotic behavior of the walk. For a fixed \( \omega \in \Omega \) the measure \( P_{\omega,x} \) is called quenched. The averaged measure \( P_x \) is obtained by averaging over environments, i.e. \( P_x(\cdot) := \mathbb{E}(P_{x,\omega}(\cdot)) \).

There is an obvious symmetry between positive and negative cookies: if the environment \((\omega_z)_z \in \mathbb{Z} \) is replaced by \((\omega'_z)_z \in \mathbb{Z} \) where \( \omega'_z(i) = 1 - \omega_z(i) \), for all \( i \in \mathbb{N}, z \in \mathbb{Z} \), then \( X' \), the ERW corresponding to the new environment, satisfies \( X' \overset{d}{=} -X \), where \( \overset{d}{=} \) denotes the equality in distribution. Thus, it is sufficient to consider only non-negative \( \delta \) (this, of course, allows both negative and positive cookies), and we shall always assume this to be the case.

ERW on \( \mathbb{Z} \) in a non-negative cookie environment and its natural extension to \( \mathbb{Z}^d \) (when there is a direction in \( \mathbb{R}^d \) such that the projection of a drift induced by every cookie on that direction is non-negative) were considered previously by many authors (see, for example, [4], [21], [22], [2], [3], [17] [5], [9], [16], and references therein).

Our model allows both positive and negative cookies but restricts their number per site to \( M \). This model was studied in [14], [15], [19]. It is known that the process is recurrent, i.e. for \( \mathbb{P}\)-a.e. \( \omega \) it returns to the starting point infinitely often, if and only if \( \delta \leq 1 \) ([14]). For transient (i.e. not recurrent) ERW, there is a rich variety of limit laws under \( P_0 \) (see [15]).

In this paper we study scaling limits of recurrent ERW under \( P_0 \). The functional limit theorem for recurrent ERW in stationary ergodic non-negative cookie environments on strips \( \mathbb{Z} \times (\mathbb{Z}/L\mathbb{Z}) \), \( L \in \mathbb{N} \), under the quenched measure was proven in [9]. Our results deal only with i.i.d. environments on \( \mathbb{Z} \) with bounded number of cookies per site but remove the non-negativity assumption on the cookies. We are also able to treat the boundary case \( \delta = 1 \). Extensions of these results and results of [15] to strips or \( \mathbb{Z}^d \), \( d > 1 \), or the “boundary” case for the model treated in [9] are still open problems.

To state our results we need to define the candidates for limiting processes. Let \( D([0, \infty)) \) be the Skorokhod space of càdlàg functions on \([0, \infty)\) and denote by \( \overset{J}{\to} \) the weak convergence in the standard \((J_1)\) Skorokhod topology on \( D([0, \infty)) \). Unless stated otherwise, all processes start at the origin at time 0. Let \( B = (B(t)), \ t \geq 0, \) denote a standard Brownian motion and \( X_{\alpha,\beta} = (X_{\alpha,\beta}(t)), \ t \geq 0, \)
be an \((\alpha, \beta)\)-perturbed Brownian motion, i.e. the solution of the equation

\[
X_{\alpha,\beta}(t) = B(t) + \alpha \sup_{s \leq t} X_{\alpha,\beta}(s) + \beta \inf_{s \leq t} X_{\alpha,\beta}(s),
\]

Equation (2) has a pathwise unique solution if \((\alpha, \beta) \in (-\infty, 1) \times (-\infty, 1)\) that is adapted to the filtration of \(B\) ([18], [7]) and a.s. continuous. Now we can state the results of our paper.

**Theorem 1** (Non-boundary case). If \(\delta \in [0, 1)\) then

\[
\frac{X[n]}{\sqrt{n}} \overset{D}{\to} X_{\delta, -\delta}(\cdot) \quad \text{as} \quad n \to \infty.
\]

We note that the first random walk models converging to perturbed Brownian Motion were studied in [8, 20].

**Theorem 2** (Boundary case). Let \(\delta = 1\) and \(B^*(t) := \max_{s \leq t} B(s), t \geq 0\). Then there exists a constant \(D\) such that

\[
\frac{X[n]}{D \sqrt{n \log n}} \overset{D}{\to} B^*(\cdot) \quad \text{as} \quad n \to \infty.
\]

Observe that for \(\delta = 1\) the limiting process is transient while the original process is recurrent. To prove Theorem 2 we consider the process \(S_j := \max_{0 \leq i \leq j} X_i, j \geq 0\), and show that after rescaling it converges to the running maximum of Brownian motion. The stated result then comes from the fact that with an overwhelming probability the maximum amount of “backtracking” of \(X_j\) from \(S_j\) for \(j \leq [Tn]\) is of order \(\sqrt{n}\), which is negligible on the scale \(\sqrt{n \log n}\) (see Lemma 10).

2. Notation and preliminaries

Assume that \(\delta \geq 0\) and \(X_0 = 0\). Let \(T_x = \inf\{j \geq 0 : X_j = x\}\) be the first hitting time of \(x \in \mathbb{Z}\). Set

\[
S_n = \max_{k \leq n} X_k, \quad I_n = \min_{k \leq n} X_k, \quad R_n = S_n - I_n + 1, \quad n \geq 0.
\]

At first, we recall the connection with branching processes exploited in [2], [3], [14], and [15].

For \(n \in \mathbb{N}\) and \(0 \leq k \leq n\) define

\[
D_{n,k} = \sum_{j=0}^{T_n-1} 1\{X_j = k, X_{j+1} = k-1\},
\]

the number of jumps from \(k\) to \(k-1\) before time \(T_n\). Then

\[
T_n = n + 2 \sum_{k \leq n} D_{n,k} = n + 2 \sum_{0 \leq k \leq n} D_{n,k} + 2 \sum_{k < 0} D_{n,k}.
\]
Consider the “backward” process \( (D_{n,n}, D_{n,n-1}, \ldots, D_{n,0}) \). Obviously, \( D_{n,n} = 0 \) for every \( n \in \mathbb{N} \). Moreover, given \( D_{n,n}, D_{n,n-1}, \ldots, D_{n,k+1} \), we can write

\[
D_{n,k} = \sum_{j=1}^{D_{n,k+1}} \text{(\# of jumps from } k \text{ to } k - 1 \text{ between the } (j - 1)-\text{th and } j\text{-th jump from } k \text{ to } k + 1 \text{ before time } T_n), \quad k = 0, 1, \ldots, n - 1.
\]

Here we used the observation that the number of jumps from \( k \) to \( k + 1 \) before time \( T_n \) is equal to \( D_{n,k+1} + 1 \) for all \( 0 \leq k \leq n - 1 \). It follows from the definition that \( (D_{n,n}, D_{n,n-1}, \ldots, D_{n,0}) \) is a Markov process. Moreover, it can be recast as a branching process with migration (see [14], Section 3, as well as [15], Section 2).

Let \( V := (V_k) \), \( k \geq 0 \), be the process such that \( V_0 = 0 \) and

\[
(V_0, V_1, \ldots, V_n) \overset{d}{=} (D_{n,n}, D_{n,n-1}, \ldots, D_{n,0}) \quad \text{for all } n \in \mathbb{N}.
\]

Denote by \( \sigma \in [1, \infty] \) and \( \Sigma \in [0, \infty] \) respectively the lifetime and the total progeny over the lifetime of \( V \), i.e. \( \sigma = \inf\{k > 0 : V_k = 0\} \), \( \Sigma = \sum_{k=0}^{\sigma-1} V_k \). The probability measure that corresponds to \( V \) will be denoted by \( P^V_0 \). The following result will be used several times throughout the paper.

**Theorem 3** ([15], Theorems 2.1 and 2.2). Let \( \delta > 0 \). Then

\[
\lim_{n \to \infty} n^{\delta} P^V_0 (\sigma > n) = C_1 \in (0, \infty);
\]

\[
\lim_{n \to \infty} n^{\delta} P^V_0 (\Sigma > n^2) = C_2 \in (0, \infty).
\]

We shall need to consider \( V \) over many lifetimes. Let \( \sigma_0 = 0 \), \( \Sigma_0 = 0 \),

\[
\sigma_i = \inf\{k > \sigma_{i-1} : V_k = 0\}, \quad \Sigma_i = \sum_{k=\sigma_{i-1}}^{\sigma_i} V_k, \quad i \in \mathbb{N}.
\]

Then \( (\sigma_i - \sigma_{i-1}, \Sigma_i)_{i \in \mathbb{N}} \) are i.i.d. under \( P^V_0 \), \( (\sigma_i - \sigma_{i-1}, \Sigma_i) \overset{d}{=} (\sigma, \Sigma), \quad i \in \mathbb{N} \).

**3. Non-boundary case: two useful lemmas**

Let \( \delta \in [0, 1) \). First of all, we show that by time \( n \) the walker consumes almost all the drift between \( I_n \) and \( S_n \).

**Lemma 4.** Assume that \( \delta \in [0, 1) \). Given \( \gamma_1 > \delta \), there exist \( \gamma_2 > 0 \) and \( \theta \in (0, 1) \) such that for all \( 1 \leq \ell \leq n \)

\[
P_0 \left( \sum_{m=n-\ell}^{n-1} \mathbb{1}_{\{L_m(T_n) < M\}} > \ell^{\gamma_1} \right) \leq \theta \ell^{\gamma_2} \quad \text{and}
\]

\[
P_0 \left( \sum_{m=-(n-1)}^{-(n-\ell)} \mathbb{1}_{\{L_m(T_{n-1}) < M\}} > \ell^{\gamma_1} \right) \leq \theta \ell^{\gamma_2}.
\]
Proof. We shall start with (7) and use the connection with branching processes. Since the event we are interested in depends only on the environment and the behavior of the walk on \( \{ n - \ell, n - \ell + 1, \ldots \} \), we may assume without loss of generality that the process starts at \( n - \ell \) and, thus, by translation invariance consider only the case \( \ell = n \).

Let \( L_k^V(n) = \sum_{j=0}^n \indic{V_j = k} \). We have

\[
(9) \quad P_0^V \left( \sum_{m=0}^{n-1} \indic{L_m(T_n) < M} > n^{\gamma_1} \right) \leq P_0^V \left( \sum_{m=0}^{n} \indic{D_{n,m} < M} > n^{\gamma_1} \right) = P_0^V \left( \sum_{m=0}^{n} \indic{V_m < M} > n^{\gamma_1} \right)
\]

\[
\leq M \max_{0 \leq k < M} P_0^V \left( \sum_{m=0}^{n} \indic{V_m = k} > \frac{n^{\gamma_1}}{M} \right) = M \max_{0 \leq k < M} P_0^V \left( L_k^V(n) > \frac{n^{\gamma_1}}{M} \right).
\]

At first, consider the case \( \delta \in (0, 1) \). Let \( k = 0 \). Then (see (4) and (6)) for all sufficiently large \( n \) we get

\[
P_0^V \left( L_0^V(n) > \frac{n^{\gamma_1}}{M} \right) \leq \prod_{i=1}^{\left\lfloor n^{\gamma_1}/M \right\rfloor} P_0^V (\sigma_i - \sigma_{i-1} \leq n) \leq \left( 1 - \frac{C_1}{2n^\delta} \right)^{\left\lfloor n^{\gamma_1}/M \right\rfloor}.
\]

Since \( \gamma_1 > \delta \), this implies the desired estimate for \( k = 0 \).

Let \( k \in \{1, 2, \ldots, M - 1\} \). Then for any \( \varepsilon > 0 \)

\[
P_0^V \left( L_k^V(n) > \frac{n^{\gamma_1}}{M} \right) = P_0^V \left( L_k^V(n) > \frac{n^{\gamma_1}}{M}, L_0^V(n) > \frac{\varepsilon n^{\gamma_1}}{2M} \right) + P_0^V \left( L_k^V(n) > \frac{n^{\gamma_1}}{M}, L_0^V(n) \leq \frac{\varepsilon n^{\gamma_1}}{2M} \right)
\]

\[
\leq P_0^V \left( L_0^V(n) > \frac{\varepsilon n^{\gamma_1}}{2M} \right) + P_0^V \left( L_0^V(n) \leq \frac{\varepsilon n^{\gamma_1}}{2M} \mid L_k^V(n) > \frac{n^{\gamma_1}}{M} \right).
\]

We only need to estimate the last term. Notice that by (A2) there is \( \varepsilon > 0 \) such that \( P_0^V(V_{j+1} = 0 \mid V_j = k) \geq \varepsilon \) for all \( k \in \{1, 2, \ldots, M - 1\} \) and \( j \in \mathbb{N} \). Therefore, the last term is bounded above by the probability that in at least \( \left\lfloor n^{\gamma_1}/M \right\rfloor \) independent Bernoulli trials with probability of success in each trial of at least \( \varepsilon \) there are at most \( \left\lfloor \varepsilon n^{\gamma_1}/(2M) \right\rfloor \) successes. This probability is bounded above by \( \exp(-c n^{\gamma_1}/M) \) for some positive \( c = c(\varepsilon) \). This completes the proof of (7) for \( \delta > 0 \).

If \( \delta = 0 \) we modify the environment by increasing slightly the drift (to the right) in the first cookie at each site. Let \( \tilde{V} \) be the branching process corresponding to the modified environment. There is a natural coupling between \( V \) and \( \tilde{V} \) such that \( \tilde{V}_j \leq V_j, j \in \{0, 1, \ldots, n\} \). Accordingly,

\[
\sum_{j=0}^{n} \indic{V_j < M} \leq \sum_{j=0}^{n} \indic{\tilde{V}_j < M}.
\]
and (7) for $\delta = 0$ follows from the result for $\delta > 0$ and the second line of (9).

Next after replacing $X$ by $-X$ proving (8) reduces to proving (7) for $\delta \leq 0$ and $\gamma_1 > 0$. As above, the result for $\delta \leq 0$ can be deduced from the result for $\delta \in (0, \gamma_1)$ by coupling of the corresponding branching processes. □

Next we show that $\sqrt{n}$ is a correct scaling in Theorem 1.

**Lemma 5.** Assume that $\delta \in [0, 1)$. There exists $\theta \in (0, 1)$ such that for all $L > 0$ and $\ell \in \mathbb{N} \cup \{0\}$ and $n \in \mathbb{N}$

$$P_0 \left( T_{\ell+n} - T_\ell \leq \frac{n^2}{L} \right) \leq \theta \sqrt{\ell}, \quad P_0 \left( T_{-\ell-n} - T_{-\ell} \leq \frac{n^2}{L} \right) \leq \theta \sqrt{\ell}.$$

**Proof.** We shall prove the first inequality for $\delta \in (0, 1)$. The case $\delta = 0$ and the second inequality are handled in exactly the same way as in the proof of Lemma 4.

Since $T_{n+\ell} - T_\ell \geq \sum_{k=\ell}^{n+\ell} D_{n+\ell,k} \stackrel{d}{=} \sum_{j=0}^{n} V_j$, it is enough to show that

$$P_0^V \left( \sum_{j=0}^{n} V_j \leq \frac{n^2}{L} \right) \leq \theta \sqrt{\ell}.$$

Notice that by the Markov property and the stochastic monotonicity of $V$ in the initial number of particles

$$P_0^V \left( \sum_{j=0}^{m+k} V_j \leq n \right) \leq P_0^V \left( \sum_{j=m+1}^{m+k} V_j \leq n \right) \left( P_0^V \left( \sum_{j=0}^{m} V_j \leq n \right) \right) \leq P_0^V \left( \sum_{j=0}^{n} V_j \leq n \right).$$

Suppose that we can show that there exist $K, n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$P_0^V \left( \sum_{j=0}^{Kn/n} V_j \leq \frac{n^2}{L} \right) \leq \frac{1}{2}. \quad (11)$$

Then using (10) and (11) we get that for all $L > 4K^2$ and $n \geq \sqrt{\ell} n_0$

$$P_0^V \left( \sum_{j=0}^{n} V_j \leq \frac{n^2}{L} \right) \leq \left( P_0^V \left( \sum_{j=0}^{2Kn/n} V_j \leq \frac{n^2}{L} \right) \right)^{[\sqrt{\ell}/(2K)]}$$

$$\leq \left( P_0^V \left( \sum_{j=0}^{2Kn/\sqrt{\ell}} V_j \leq \frac{1}{4} \left( \frac{n}{\sqrt{\ell}} \right)^2 \right) \right)^{[\sqrt{\ell}/(2K)]} \leq \left( \left( \frac{1}{2} \right)^{1/(4K)} \right)^{\sqrt{\ell}},$$

and we are done.

To prove (11), we observe that due to (4) the sequence $\sigma_m/m^{1/4}, \ m \in \mathbb{N},$ has a limiting distribution ([10], Theorem 3.7.2) and, thus, if $K$ is large then
\[ P_0(\sigma_{[\sqrt{Kn}]} > Kn) \leq 1/4 \] for all large enough \( n \). We conclude that there is an \( n_0 \in \mathbb{N} \) such that for all \( n \geq n_0 \)

\[
P_0^V \left( \sum_{j=0}^{Kn} V_j \leq n^2 \right) \leq P_0 \left( \sum_{j=0}^{\sigma_{[\sqrt{Kn}]} \leq Kn} V_j \leq n^2 \right) + \frac{1}{4} \]

\[
\leq P_0 \left( \sum_{i=1}^{[\sqrt{Kn}]} \Sigma_i \leq n^2 \right) + \frac{1}{4} \leq P_0 \left( \sum_{i=1}^{[\sqrt{Kn}]} \Sigma_i \leq n^2 \right) + \frac{1}{4} \left( 1 - \frac{C_2}{2n^\delta} \right) + \frac{1}{4}.
\]

This immediately gives (11) if \( K \) is chosen sufficiently large. \( \square \)

4. Non-boundary case: Proof of Theorem 1

Let \( \Delta_n = X_{n+1} - X_n \) and

\[ B_n = \sum_{k=0}^{n-1} (\Delta_k - E_0,\omega(\Delta_k|\mathcal{F}_k)), \quad C_n = \sum_{k=0}^{n-1} E_0,\omega(\Delta_k|\mathcal{F}_k). \]

Then \( X_n = B_n + C_n \), where \( (B_n), n \geq 0 \) is a martingale. Consider the following sequences of rescaled continuous time càdlàg processes

\[ X^{(n)}(t) := \frac{X_{[nt]}}{\sqrt{n}}, \quad B^{(n)}(t) := \frac{B_{[nt]}}{\sqrt{n}}, \quad C^{(n)}(t) := \frac{C_{[nt]}}{\sqrt{n}}, \quad t \geq 0. \]

Theorem 1 is an easy consequence of the following three lemmas, the first of which holds for the quenched and the last two for the averaged measures.

Lemma 6. Let \( B \) be a standard Brownian motion. Then \( B^{(n)} \xrightarrow{\mathcal{D}} B \) as \( n \to \infty \) for \( \mathbb{P} \)-a.e. \( \omega \).

Lemma 7. For each \( t \geq 0 \) and \( \varepsilon > 0 \)

\[ P_0 \left( \sup_{k \leq nt} \frac{|C_k - \delta R_k|}{\sqrt{n}} > \varepsilon \right) \to 0. \]

Lemma 8. The sequence \( X^{(n)}, n \geq 1 \), is tight in \( D([0,\infty)) \). Moreover, if \( X \) is a limit point of this sequence and \( P \) is the corresponding measure on \( D([0,\infty)) \) then \( P(X \in C([0,\infty])) = 1 \).

Proof of Theorem 1 assuming Lemmas 6–8. Since \( X^{(n)}, n \geq 1 \), is tight and \( B^{(n)} \xrightarrow{\mathcal{D}} B \) as \( n \to \infty \), the sequence \( C^{(n)}, n \geq 1 \), as the difference of two tight sequences is also tight. We can assume by choosing a subsequence that \( X^{(n)} \xrightarrow{\mathcal{D}} X \), where \( X \) is continuous by Lemma 8. The mapping \( x(\cdot) \mapsto r^x(\cdot) := \sup_{s \leq t} x(s) - \inf_{s \leq t} x(s) \) is continuous on \( C([0,t]) \). Therefore, by the continuous mapping theorem

\[ r^{X^{(n)}}(\cdot) = \frac{R_{[nt]}}{\sqrt{n}} \xrightarrow{\mathcal{D}} r^X(\cdot). \]
The tightness of $C^{(n)}$, $n \geq 1$, (13), Lemma 7, and the "convergence together" result ([6], Theorem 3.1) imply that $C^{(n)} \xrightarrow{d} \delta r^X$ as $n \to \infty$.

Now we have a vector-valued sequence of processes $(X^{(n)}, B^{(n)}, C^{(n)})$, $n \geq 1$, that is tight. Therefore, along a subsequence, this 3-dimensional process converges to $(X, B, \delta r^X)$. Since $X^{(n)} = B^{(n)} + C^{(n)}$, we get that $X = B + \delta r^X$. □

We shall conclude this section with proofs of Lemmas 6–8.

**Proof of Lemma 6.** We shall use the functional limit theorem for martingale differences ([6], Theorem 18.2). Let $\xi_{nk} = n^{-1/2}(\Delta_{k-1} - E_{0,\omega}(\Delta_{k-1}|\mathcal{F}_{k-1}))$, $k, n \in \mathbb{N}$. Due to rescaling and the fact that ERW moves in unit steps, it is obvious that the Lindeberg condition,

$$\sum_{k \leq nt} E_{0,\omega}[\xi_{nk}^2 1_{\{\xi_{nk} \geq \varepsilon\}}] \to 0 \quad \text{as } n \to \infty \text{ for every } t \geq 0 \text{ and } \varepsilon > 0,$$

is satisfied. Thus, we just have to show the convergence of the quadratic variation process, i.e. for $P$-a.e. $\omega$ for each $t \geq 0$

$$\sum_{k \leq nt} E_{0,\omega}(\xi_{nk}^2|\mathcal{F}_{k-1}) = \frac{[nt]}{n} - \frac{1}{n} \sum_{k \leq nt} (E_{0,\omega}(\Delta_{k-1}|\mathcal{F}_{k-1}))^2 \to t$$

as $n \to \infty$. Since

$$0 \leq \frac{1}{n} \sum_{k \leq nt} (E_{0,\omega}(\Delta_{k-1}|\mathcal{F}_{k-1}))^2 \leq \frac{M}{n} R_{[nt]},$$

it is enough to prove that $P_{0,\omega}(R_{[nt]} > \varepsilon n) \to 0$ a.s. for each $\varepsilon > 0$. We have

$$P_{0,\omega}(R_{[nt]} > \varepsilon n) \leq P_{0,\omega}(T_{[en/3]} \leq nt) + P_{0,\omega}(T_{-[en/3]} \leq nt) =: f_{n,\varepsilon}(\omega, t).$$

By Fubini’s theorem and Lemma 5,

$$\mathbb{E}\left(\sum_{n=1}^{\infty} f_{n,\varepsilon}(\omega, t)\right) = \sum_{n=1}^{\infty} \mathbb{E}f_{n,\varepsilon}(\omega, t) = \sum_{n=1}^{\infty} \left(P_{0}(T_{[en/3]} \leq nt) + P_{0}(T_{-[en/3]} \leq nt)\right) < \infty.$$

This implies that $f_{n,\varepsilon}(\omega, t) \to 0$ a.s. as $n \to \infty$ and completes the proof. □

**Proof of Lemma 7.** Let $d_m = \sum_{i=1}^{M} (2\omega_m(i) - 1)$ be the total drift stored at site $m$, $m \in \mathbb{Z}$. Then

$$C_k - \delta R_k = \sum_{m=L_k}^{S_k} (d_m - \delta) - \sum_{m=L_k}^{S_k} 1_{\{L_m(k) < M\}} \sum_{j=L_m(k)+1}^{M} (2\omega_m(j) - 1).$$
By Lemma 5, given $\nu > 0$, we can choose $K$ sufficiently large so that $P_0(R_{[nt]} > K\sqrt{n}) < \nu/2$ for all $n \in \mathbb{N}$. We have

\begin{equation}
0 \left( \sup_{k \leq nt} \frac{|C_k - \delta R_k|}{\sqrt{n}} > \varepsilon \right) \leq P_0 \left( \max_{k \leq nt} \frac{\left| \sum_{m=I_k}^{d_m - \delta} S_k \right|}{R_k} \frac{R_k}{\sqrt{n}} > \frac{\varepsilon}{2}, R_{[nt]} \leq K \right) + P_0 \left( \frac{M}{\sqrt{n}} \sum_{m=I_{[nt]}}^{S_{[nt]}} \mathbb{1}_{\{L_m([nt]) < M\}} > \frac{\varepsilon}{2}, R_{[nt]} \leq K \right) + \frac{\nu}{2}.
\end{equation}

By the strong law of large numbers $\lim_{(a+b) \to -\infty} (a + b)^{-1} \sum_{m=-a}^{b} (d_m - \delta) = 0$ (\textsc{P}-a.s.). Therefore, for \textsc{P}-a.e. $\omega$ there is an $r(\omega) \in \mathbb{N}$ such that $R_k^{-1} \left| \sum_{m=I_k}^{d_m - \delta} S_k \right| \leq \varepsilon/(2K)$ whenever $R_k \geq r(\omega)$, and the first term in the right-hand side of (15) does not exceed

\[ P_0 \left( \frac{2(M + 1)r(\omega)}{\sqrt{n}} > \frac{\varepsilon}{2}, R_{[nt]} \leq K \right) \leq \mathbb{E} \left( P_{0,\omega} \left( r(\omega) > \frac{\varepsilon \sqrt{n}}{4(M + 1)} \right) \right) \to 0 \text{ as } n \to \infty. \]

Thus, we only need to estimate the second term in the right-hand side of (15).

Divide the interval $[I_{[nt]}, S_{[nt]}]$ into subintervals of length $n^{1/4}$. By Lemma 4, given $\gamma_1 \in (\delta, 1)$, with probability at least $1 - \theta^{n^{1/4}} K n^{1/4}$ all subintervals except the two extreme ones have at most $n^{\gamma_1/4}$ points which are visited less than $M$ times. Hence, for $n$ sufficiently large

\[ P_0 \left( \frac{M}{\sqrt{n}} \sum_{m=I_{[nt]}}^{S_{[nt]}} \mathbb{1}_{\{L_m([nt]) < M\}} > \frac{\varepsilon}{2}, R_{[nt]} \leq K \right) \leq P_0 \left( \sum_{m=I_{[nt]}}^{S_{[nt]}} \mathbb{1}_{\{L_m([nt]) < M\}} > n^{(1+\gamma_1)/4} + 2n^{1/4}, R_{[nt]} \leq K \right) \leq \theta^{n^{1/4}} K n^{1/4}, \]

and the proof is complete. \hfill \Box

\textbf{Proof of Lemma 8.} The idea of the proof is the following. If $X(n)$ has large fluctuations then either $B(n)$ has large fluctuations or $C(n)$ has large fluctuations. $B(n)$ is unlikely to have large fluctuations, since it converges to the Brownian motion. By Lemma 4, $C_n$ can have large fluctuations only if $S_n$ increases or $I_n$ decreases. However by Lemma 5 neither $I_n$ nor $S_n$ can change too quickly. Let us give the details.

To prove both statements of Lemma 8 it is enough to show that there exists $C_3, \alpha > 0$ such that for all $\ell \in \mathbb{N}$ and sufficiently large $n$, $n > 2^\ell$

\[ P_0(\cup_{k < 2^\ell} \Omega_{n,k,\ell}) \leq C_3 2^{-\alpha \ell}, \]
where
\[ \Omega_{n,k,\ell} = \left\{ \left| X^{(n)} \left( \frac{k + 1}{2^\ell} \right) - X^{(n)} \left( \frac{k}{2^\ell} \right) \right| > 2^{-\ell/8} \right\} \]
(see e.g. the last paragraph in the proof of Lemma 1 in [12], Chapter III, Section 5).

Let
\[ m_1 := \left\lfloor \frac{kn}{2^\ell} \right\rfloor, \quad m_2 := \left\lfloor \frac{(k + 1)n}{2^\ell} \right\rfloor, \quad J := \frac{1}{4} n^{1/2} 2^{-\ell/8}. \]

Then
\[ \Omega_{n,k,\ell} = \{ |X_{m_2} - X_{m_1}| > 4J \} = \Omega^+_n \cup \Omega^-_{n,k,\ell}, \]
where
\[ \Omega^+_n = \{ X_{m_2} > X_{m_1} + 4J \}, \quad \Omega^-_{n,k,\ell} = \{ X_{m_2} < X_{m_1} - 4J \}. \]

We shall deal with \( \Omega^+_n \) and leave \( \Omega^-_{n,k,\ell} \) to the reader. Consider two cases.

1. \( S_{m_1} \geq X_{m_1} + 2J \). Let \( B^+_m = C^+_m = 0 \), and
\[ B^+_j = \sum_{k=m_1}^{j-1} (\Delta_k - E_0,\omega(\Delta_k|\mathcal{F}_k)) \mathbb{1}_{\{X_{m_1} \leq X_k \leq X_{m_1} + 2J\}}, \]
\[ C^+_j = \sum_{k=m_1}^{j-1} E_0,\omega(\Delta_k|\mathcal{F}_k) \mathbb{1}_{\{X_{m_1} \leq X_k \leq X_{m_1} + 2J\}}, \quad m_1 < j \leq m_2. \]

Define \( \tau := \inf \{ m > m_1 : X_m > X_{m_1} + 2J \} \). Then
\[ \Omega^+_{n,k,\ell} \subset \{ \tau < m_2 \} = \{ \tau < m_2, \sum_{k=m_1}^{\tau-1} \Delta_k \mathbb{1}_{\{X_{m_1} \leq X_k \leq X_{m_1} + 2J\}} = 2J + 1 \}
\[ \subset \{ \tau < m_2 \} \cap (\{ B^+_\tau \geq J \} \cup \{ C^+_\tau \geq J \}) \subset \Omega^+_{n,k,\ell} \cup \Omega^C_{n,k,\ell}, \]
where
\[ \Omega^+_{n,k,\ell} = \left\{ \sup_{m_1 \leq j \leq m_2} B^+_j \geq J \right\} \quad \text{and} \quad \Omega^C_{n,k,\ell} = \left\{ \sup_{m_1 \leq j \leq m_2} C^+_j \geq J \right\}. \]

Since the quadratic variation of \( B^+ \) grows at most linearly, the maximal inequality for martingales and Burkholder-Davis-Gundy inequality ([13], Theorem 2.11, \( p = 4 \)) imply that
\[ P_0,\omega(\Omega^+_{n,k,\ell}) \leq \frac{C (m_2 - m_1)^2}{J^4} \leq C' 2^{-3\ell/2}. \]

Hence, \( P_0 (\bigcup_{k<2^\ell} \Omega^+_{n,k,\ell}) \leq C' 2^{-\ell/2} \). Next, since \( S_{m_1} \geq X_{m_1} + 2J \), we have that
\[ \max_{m_1 \leq j \leq m_2} C^+_j \leq M \sum_{k \in [X_{m_1}, X_{m_1} + 2J]} \mathbb{1}_{\{ L_k(T_{X_{m_1}} \vee T_{X_{m_1} + [2J]} < M) \}}. \]
and, therefore,

\[ P_0 \left( \mathcal{O}_{n,k,\ell}^{C^+} \right) \leq P_0 \left( \sum_{k \in \mathcal{X}_{m_1} \cup \mathcal{X}_{m_1 + 2J}} \mathbb{1}_{\{L_k(T_{X_{m_1} + 2J}) < M\}} \geq \frac{J}{M} \right) \]

\[ \leq 2m_1 \max_{-m_1 \leq x \leq m_1 - 2J} P_0 \left( \sum_{k \in [x,x+2J]} \mathbb{1}_{\{L_k(T_{x+2J}) < M\}} \geq \frac{J}{M} \right) \leq 2m_1 \theta^{[J/(2M)]^{\gamma_2}}. \]

The last inequality is due to Lemma 4. Recalling the definitions of \( m_1 \) and \( J \) (see (16)) we conclude that for each \( \alpha > 0 \) there are \( n_1 > 2^\ell \) and \( c > 0 \) such that for all \( n \geq n_1 \)

\[ P_0 \left( \bigcup_{k < 2^\ell} \mathcal{O}_{n,k,\ell}^{C^+} \right) \leq 2^\ell (2m_1) \theta^{[J/(2M)]^{\gamma_2}} \leq 2^{-c_0^\gamma_2/4}. \]

This completes the analysis of case (1).

(2) \( S_{m_1} < X_{m_1} + 2J \). In this case, \( \mathcal{O}_{n,k,\ell}^+ \subset \{T_{S_{m_1} + [2J]} - T_{S_{m_1} + 1} \leq m_2 - m_1\} \). There exists \( m \in \mathbb{N} \) such that \( S_{m_1 + 1} - m \leq m \leq m + 1 \). Let \( \tilde{\ell} = \ell/8 \). Using Lemma 5, we can find \( K > 1 \) such that \( P_0 (S_n > K \sqrt{n}) < 2^{-\tilde{\ell}} \) for all sufficiently large \( n \). On the set \( \{S_n \leq K \sqrt{n}\} \) we need to estimate

\[ P_0 \left( \bigcup_{m < 2^\ell + 3K} \mathcal{O}_{n,m,\ell}^+ \right), \text{ where } \mathcal{O}_{n,m,\ell}^+ = \{T_{(m+1)[J]} - T_m [J] \leq m_2 - m_1\}. \]

Since \( m_2 - m_1 \leq C[J]^2 / 2^{\tilde{\ell}} \) for some constant \( C > 0 \), Lemma 5 imply that there is \( c > 0 \) such that and all sufficiently large \( n \)

\[ P_0 \left( \bigcup_{m < 2^\ell + 3K} \mathcal{O}_{n,m,\ell}^+ \right) \leq \sum_{m < 2^\ell + 3K} P_0 \left( \mathcal{O}_{n,m,\ell}^+ \right) \leq 2^\ell + 3K \theta^{c2^{3\tilde{\ell}}}. \]

This completes the analysis of case (2) establishing Lemma 8. \( \square \)


Let \( \delta = 1 \). For \( t \geq 0 \) set

\[ T^{(n)}(x) := \frac{T_{[nx]}}{n^2/\log^2 n}, \quad X^{(n)}(t) := \frac{X_{[nt]}}{\sqrt{n/\log n}}, \quad S^{(n)}(t) := \frac{S_{[nt]}}{\sqrt{n/\log n}}. \]

Let \( \Sigma_j, j \geq 0 \) be i.i.d. positive integer-valued random variables defined in (6). They satisfy (5) with \( \delta = 1 \) and by ([11], Chapter 9, Section 6) for some constant \( a > 0 \)

\[ \sum_{j=0}^{[n]} \frac{\Sigma_j}{n^2} \Rightarrow aH(\cdot) \text{ as } n \to \infty, \]

where \( H := (H(x)), x \geq 0, \) is a stable subordinator with index 1/2. More precisely,

\[ H(x) = \inf \{t \geq 0 : B(t) = x\}. \]

We shall need the following two lemmas.
Lemma 9. The finite dimensional distributions of $T^{(n)}$ converge to those of $cH$, where $c > 0$ is a constant and $H$ is given by (18).

Lemma 10. For every $\varepsilon > 0$, $T > 0$

$$\lim_{n \to \infty} P_0 \left( \sup_{0 \leq t \leq T} (S^{(n)}(t) - X^{(n)}(t)) > \varepsilon \right) = 0.$$ 

Theorem 2 is an easy consequence of these lemmas.

Proof of Theorem 2. Lemma 9 implies that the finite dimensional distributions of the process $S^{(n)}$ converge to those of $DB^*$, where $D > 0$ is a constant. Since the trajectories of $S^{(n)}$ are monotone and the limiting process $B^*$ is continuous, we conclude that $S^{(n)}$ converges weakly to $DB^*$ in the (locally) uniform topology (see [1], Corollary 1.3 and Remark (e) on p. 588). Finally, by Lemma 10 for each $T > 0$

$$\sup_{0 \leq t \leq T} (S^{(n)}(t) - X^{(n)}(t)) \to 0$$

in $P_0$ probability. By the “converging together” theorem ([6], Theorem 3.1) we conclude that $X^{(n)}$ converges weakly to $DB^*$ in the (locally) uniform topology. □

Proof of Lemma 9. Let $k \in \mathbb{N}$ and $0 = x_0 < x_1 < \cdots < x_k$. We have to show that for any $0 = t_0 < t_1 < t_2 < \cdots < t_k$

$$P_0(T^{(n)}(x_k) - T^{(n)}(x_i) \leq t_{k-i}, \forall i = 0, 1, 2, \ldots, k - 1)$$

$$\to P(T(x_k) - T(x_i) \leq t_{k-i}, \forall i = 0, 1, \ldots, k - 1), \text{ as } n \to \infty,$$

where $T(\cdot) = cH(\cdot)$ for some $c > 0$.

At time $T_{[nx_i]}$ consider the structure of the corresponding branching process as we look back from $[nx_k]$. Notice that $D_{[nx_i],j} \leq D_{[nx_k],j}$ for $i \leq k$ and all $j$. This simple observation will allow us to get bounds on $T_{[nx_i]}$, $i = 1, 2, \ldots, k - 1$, in terms of the structure of downcrossings at time $T_{[nx_k]}$. This means that we can use the same copy of the branching process $V$ to draw conclusions about all hitting times $T_{[nx_i]}$, $i = 1, 2, \ldots, k$.

We shall use notation (6) and let $N^{(0)} = 0$,

$$N^{(k-i)} = \min \{ m \in \mathbb{N} : \sigma_m \geq [nx_k] - [nx_i] \}, \quad i = 0, 1, 2, \ldots, k - 1.$$ 

Since

$$2 \sum_{j=1}^{N^{(k-i)}-1} \sigma_j \leq T_{[nx_k]} - T_{[nx_i]} \leq nx_k - nx_i + 2 \sum_{j=1}^{N^{(k-i)}} \Sigma_j,$$

we have

$$P_0(T^{(n)}(x_k) - T^{(n)}(x_i) \leq t_{k-i}, \forall i = 0, 1, 2, \ldots, k - 1)$$

$$\leq P \left( 2 \sum_{j=1}^{N^{(k-i)}-1} \Sigma_j \leq n^2 t_{k-i} / \log^2 n, \forall i = 0, 1, 2, \ldots, k - 1 \right).$$
and 

\( P_0(T^{(n)}(x_k) - T^{(n)}(x_i) \leq t_{k-i}, \ \forall i = 0, 1, 2, \ldots, k-1) \)

\[ \geq P \left[ n x_k - n x_i + 2 \sum_{j=1}^{N(k-i)} \sigma_j \leq n^2 t_{k-i}/\log^2 n, \ \forall i = 0, 1, 2, \ldots, k-1 \right]. \]

Next we provide some control on \( N^{(k-i)} \), \( i = 0, 1, \ldots, k-1 \), and on the maximal lifetime over \( [nx_k] \) generations. Theorem 3 and [10, Theorem 3.7.2] imply that \( \sigma_n/(n \log n) \to b^{-1} \) for some positive constant \( b \). From this it is easily seen that

\[ \min\{m \in \mathbb{N} : \sigma_m > n\} \]

\[ \rightarrow \frac{nb}{\log n} \]

as \( n \to \infty \).

Recalling our definition of \( N^{(k-i)} \) we get that for every \( \varepsilon, \nu > 0 \) there is \( n_0 \) such that for all \( n \geq n_0 \)

\[ P \left( 1 - \nu \leq \frac{N^{(k-i)}}{\bar{N}^{(k-i)}} \leq 1 + \nu, \ i = 0, \ldots, k-1 \right) > 1 - \varepsilon, \]

where \( \bar{N}^{(k-i)} = b(x_k - x_i)n/\log n \). In particular, for \( C = (1 + \nu)bx_k \) we have that

\[ P \left( N^{(k)} \leq \frac{C n}{\log n} \right) > 1 - \varepsilon. \]

Define \( \lambda_n = (\log n)^{-1/2} \) (any sequence \( \lambda_n, n \in \mathbb{N} \), such that \( \lambda_n \to 0 \) and \( \lambda_n \log n \to \infty \) will work) and notice that by Theorem 3 there is \( n_1 \) such that for all \( n \geq n_1 \)

\[ P \left( \max_{1 \leq i \leq C n/\log n} (\sigma_i - \sigma_{i-1}) \leq n \lambda_n \right) \geq \left( 1 - \frac{2C_1}{n \lambda_n} \right) \] \[ \rightarrow \frac{C n}{\log n} \]

Thus, on a set \( \Omega_\varepsilon \) of measure at least \( 1 - 2\varepsilon \) for all \( n \geq n_0 \vee n_1 \) the number of lifetimes of the branching process \( V \) covering \([nx_k] - [nx_i] \) generations, \( i = 0, 1, 2, \ldots, k-1 \), is well controlled and the maximal lifetime over \([nx_k] \) generations does not exceed \( n \lambda_n \). In particular, on \( \Omega_\varepsilon \), the number of lifetimes in any interval \([nx_i], [nx_{i+1}]\), \( i = 0, 1, \ldots, k-1 \), goes to infinity as \( n \to \infty \).

Finally, on \( \Omega_\varepsilon \) we get from (19) and (17) that

\[ P_0(T^{(n)}(x_k) - T^{(n)}(x_i) \leq t_{k-i}, \ \forall i = 0, 1, 2, \ldots, k-1) \]

\[ \leq P \left( 2 \sum_{j=1}^{(1-\nu)N^{(k-1)}} \sigma_j \leq n^2 t_{k-i}/\log^2 n, \ \forall i = 0, 1, 2, \ldots, k-1 \right) \]

\[ = P \left( \frac{\sum_{j=1}^{(1-\nu)N^{(k-1)}} \sigma_j}{(1-\nu)n/\log n} \leq \frac{t_{k-i}}{2(1-\nu)^2}, \ \forall i = 0, 1, 2, \ldots, k-1 \right) \]

\[ \rightarrow P(aH(b(x_k - x_i))) \leq (1 - \nu)^{-2}t_{k-i}/2 \ \forall i = 0, 1, 2, \ldots, k-1) \]

\[ = P(2ab^2(H(x_k) - H(x_i)) \leq t_{k-i}(1 - \nu)^{-2}) \ \forall i = 0, 1, 2, \ldots, k-1). \]
The lower bound is shown starting from (20) in exactly the same way. Letting $\nu \to 0$ and then $\epsilon \to 0$ we obtain the statement of the lemma with $T(\cdot) = 2ab^2H(\cdot) =: cH(\cdot)$.

Proof of Lemma 10. Without loss of generality we can consider $t \in [0,1]$. Fix some $\nu > 0$. We have

\begin{equation}
(22) \quad P_0 \left( \sup_{0 \leq t \leq 1} (S^{(n)}(t) - X^{(n)}(t)) > \epsilon \right) \leq P_0(S_n \geq K \sqrt{n \ln n}) + \nu \\
P_0 \left( \max_{0 \leq m \leq n} (S_m - X_m) > \epsilon \sqrt{n \ln n} \right) \leq P_0(S_n \geq K \sqrt{n \ln n}) + \nu.
\end{equation}

By Lemma 9 we can find $K > 0$ such that for all large $n$

\[ P_0(S_n \geq K \sqrt{n \ln n}) \leq P_0(T_{K\sqrt{n \ln n}} \leq n) < \nu. \]

To estimate the last term in (22) we shall use properties of the branching process $V$. Let $N = \min\{m \in \mathbb{N} : \sigma_m > K \sqrt{n \ln n}\}$. Then the last term in (22) is bounded by

\begin{align*}
P_0^V \left( \max_{i \leq N} (\sigma_i - \sigma_{i-1}) \geq \epsilon \sqrt{n \ln n} \right) &\leq \nu + P_0^V \left( \max_{i \leq C \sqrt{n}} (\sigma_i - \sigma_{i-1}) \geq \epsilon \sqrt{n \ln n} \right),
\end{align*}

for some large $C$ and all sufficiently large $n$. Finally, from (4) we conclude that for all large enough $n$ the last probability does not exceed

\[ 1 - \left(1 - \frac{2C_1}{\epsilon \sqrt{n \ln n}}\right)^{C \sqrt{n}} < \nu. \]

This completes the proof. □

Acknowledgments. The authors are grateful to the Fields Institute for Research in Mathematical Sciences for support and hospitality. D. Dolgopyat was partially supported by the NSF grant DMS 0854982. E. Kosygina was partially supported by a Collaboration Grant for Mathematicians (Simons Foundation) and the PSC CUNY Award # 64603-00 42. The authors also thank the anonymous referee for suggestions, which helped to improve the exposition.

References

[8] B. Davis (1996). Weak limits of perturbed random walks and the equation \(Y_t = B_t + \alpha \sup\{Y_s: s \leq t\} + \beta \inf\{Y_s: s \leq t\}\). Ann. Probab. 24 2007–2023.

DEPARTMENT OF MATHEMATICS
University of Maryland
4417 Mathematics Building
College Park, MD 20742, USA
dmitry@math.umd.edu

DEPARTMENT OF MATHEMATICS
Baruch College, Box B6-230
One Bernard Baruch Way
New York, NY 10010, USA
elena.kosygina@baruch.cuny.edu