DOUBLE JUMP PHASE TRANSITION IN A RANDOM SOLITON CELLULAR AUTOMATON

LIONEL LEVINE, HANBAEK LYU, AND JOHN PIKE

ABSTRACT. In this paper, we consider the soliton cellular automaton introduced in [21] with a random initial configuration. We give multiple constructions of a Young diagram describing various statistics of the system in terms of familiar objects like birth-and-death chains and Galton-Watson forests. Using these ideas, we establish limit theorems showing that if the first n boxes are occupied independently with probability $p \in (0,1)$, then the number of solitons is of order n for all p, and the length of the longest soliton is of order $\log n$ for $p < 1/2$, order $\sqrt{n}$ for $p = 1/2$, and order n for $p > 1/2$. Additionally, we uncover a condensation phenomenon in the supercritical regime: For each fixed $j \geq 1$, the top j soliton lengths have the same order as the longest for $p \leq 1/2$, whereas all but the longest have order at most $\log n$ for $p > 1/2$. As an application, we obtain scaling limits for the lengths of the k^{th} longest increasing and decreasing subsequences in a random stack-sortable permutation of length n in terms of random walks and Brownian excursions.

1. INTRODUCTION

In 1990, Takahashi and Satsuma proposed a $1+1$ dimensional cellular automaton of filter type called the *soliton cellular automaton*, also known as the *box-ball system* [14, 21]. It is defined as a discrete-time dynamical system $(X_s)_{s\in\mathbb{N}_0}$ whose states are binary sequences $X_s : \mathbb{N} \to \{0,1\}$ with finitely many 1's. We may think of the states as configurations of balls in boxes where box k contains a ball at stage s if $X_s(k) = 1$ and is empty if $X_s(k) = 0$. The update rule $X_s \mapsto X_{s+1}$ is defined as follows: At the beginning of stage s, each ball has been moved a total of s times. To reach stage $s+1$, successively move the leftmost ball which has been moved a total of s times to the first empty box on its right, continuing until all balls have been moved. Alternatively, at each stage $s \geq 0$ a 'ball carrier' starts at the origin and sweeps rightward to infinity. Each time she encounters an occupied box, she *pushes* the ball to the top of her stack. Each time she encounters an empty box and her stack is nonempty, she *pops* the topmost ball from her stack into the box. In keeping with this picture, we will refer to the stages of the box-ball system as *sweeps* henceforth.

Date: February 4, 2018.

2010 *Mathematics Subject Classification.* 37K40, 60F05.

Key words and phrases. Solitons, cellular automata, birth-and-death chains, Galton-Watson forest, condensation.

As a concrete example, the system initially having balls in boxes 2, 3, 5, 6, 7, 11 evolves through sweep $s = 3$ as

$$\begin{array}{r|cccccccccccccccccccccc} s=0 & 0&1&1&0&1&1&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&\dots \\ 1 & 0&0&0&1&0&0&0&1&1&1&0&1&1&0&0&0&0&0&0&0&0&0&\dots \\ 2 & 0&0&0&0&1&0&0&0&0&0&1&0&0&1&1&1&1&0&0&0&0&0&\dots \\ 3 & 0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&1&1&1&1&0&\dots \end{array}$$

In this model, a (non-interacting) *soliton of length* k is defined to be a string of k consecutive 1's followed by k consecutive 0's. During one sweep, such a soliton travels to the right at speed k. The physical interpretation is that of a traveling wave with velocity equal to its wavelength. If a k-soliton precedes a j-soliton with $j < k$, then the two will eventually collide, resulting in interference. The outcome depends on the congruence class of their initial distance modulo their relative speed, $k - j$. The case of three or more interacting solitons can be described similarly [21]. It is easy to see that since we have finitely many balls initially, after some finite time the system consists of non-interacting solitons whose lengths are nondecreasing from left to right. This final macrostate of the system can be encoded in a Young diagram having j^{th} column equal to the length of the j^{th} longest soliton.

In this paper, we start the soliton cellular automaton from a random initial configuration and study the limiting shape of the resulting Young diagram. We have two parameters, $n \in \mathbb{N}$ and $p \in (0,1)$. Let $X^{n,p}$ be a random coloring of $\mathbb{N}$ so that each site in $[1, n]$ is 1 with probability p and 0 with probability $1 - p$, independently of all others, and all sites in (n, ∞) are 0. Let $\Lambda^{n,p}$ be the corresponding random Young diagram and denote its i^{th} row and j^{th} column by $\rho_i(n)$ and $\lambda_j(n)$, respectively. (Thus $\lambda_j(n)$ gives the length of the j^{th} longest soliton and $\rho_i(n)$ the number of solitons of length at least i.) We are going to observe that each fixed row has order n for all values of p, but the column lengths vary drastically according to whether p is less than, equal to, or greater than 1/2. The asymptotics of the rows and columns of $\Lambda^{n,p}$ are summarized in the following table. For the precise meaning of the Landau notation employed, see Subsection 1.1

$i \geq 1, j \geq 2$ fixed	$\rho_i(n)$	$\lambda_j(n)$	$\lambda_1(n)$
Subcritical phase ($p < 1/2$)	$\Theta(n)$	$\Theta(\log n)$	$\Theta(\log n)$
Critical phase ($p = 1/2$)	$\Theta(n)$	$\Theta(\sqrt{n})$	$\Theta(\sqrt{n})$
Supercritical phase ($p > 1/2$)	$\Theta(n)$	$O(\log n)$	$\Theta(n)$

FIGURE 1. Double jump phase transition for the order of the longest j solitons (j fixed as $n \to \infty$) in the random box-ball system. All entries are up to constant factors that do not depend on n. In the sub- and supercritical phases the λ_j are concentrated, and the constant factor depends only on p (and not on j). In the critical phase the λ_j are not concentrated, and the constant factor depends on j. The implied constants may depend on indices i and j.

Erdős and Rényi coined the term *double jump* to describe the emergence of a giant component in the random graph $G(n, c/n)$ at $c = 1$. The phase transition in the random box-ball system is analogous to the phase transition in $G(n, p/n(1-p))$. There, it is well known that all connected components are of size $O(\log n)$ for $p < 1/2$; connected components of size $\Theta(n^{2/3})$ emerge at $p = 1/2$; and for $p > 1/2$, the largest component is of size $\Theta(n)$ while the rest have size $O(\log n)$.

1.1. **Landau notation.** We use $O(\cdot), \Omega(\cdot), \Theta(\cdot)$ in the sense of stochastic boundedness: Given $\{a_n\}_{n=1}^{\infty} \subseteq \mathbb{R}^+$ and a sequence $\{W_n\}_{n=1}^{\infty}$ of nonnegative random variables, we say that $W_n = O(a_n)$ if for every $\varepsilon > 0$, there is a $C \in (0,\infty)$ such that $\mathbb{P}\{W_n > Ca_n\} < \varepsilon$ for all n. We say that $W_n = \Omega(a_n)$ if for every $\varepsilon > 0$, there is a $c \in (0,\infty)$ such that $\mathbb{P}\{W_n < ca_n\} < \varepsilon$ for all n, and we say $W_n = \Theta(a_n)$ if $W_n = O(a_n)$ and $W_n = \Omega(a_n)$. All implied constants c, C may depend on p but not n.

1.2. **Main results.** We adopt the notation $\mathbb{R}^+ = [0,\infty)$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ throughout. Fix $p \in (0,1)$, and let $\xi_1, \xi_2, \ldots$ be an i.i.d. sequence with $\mathbb{P}\{\xi_1 = 1\} = p$ and $\mathbb{P}\{\xi_1 = -1\} = 1-p$. Define $X^p \in \{0,1\}^{\mathbb{N}}$ by

$$X^p(k) = \mathbf{1}\{\xi_k = 1\},$$

and for each $n \in \mathbb{N}$, set $X^{n,p} = X^p \mathbf{1}_{[1,n]}$. The interpretation is that $X^{n,p}$ corresponds to an arrangement of balls in boxes where boxes $1, \ldots, n$ are each occupied independently with probability p, and boxes $n+1, n+2, \ldots$ are empty.

For each fixed $n \geq 1$ and $p \in (0,1)$, we consider the box-ball system $(X_s)_{s\geq 0}$ with the random initial configuration $X_0 = X^{n,p}$. Recall that the soliton lengths in this system are denoted by $\lambda_1(n) \geq \lambda_2(n) \geq \ldots$. This information can be summarized by the Young diagram $\Lambda^{n,p}$ whose j^{th} column has length $\lambda_j(n)$. The length of its i^{th} row, $\rho_i(n)$, equals the number of solitons in the system having length at least i. In particular, $\rho_1(n)$ gives the total number of solitons.

Many properties of this Young diagram can be described in terms of the simple random walk $\{S_k\}_{k=0}^{\infty}$ defined by $S_0 = 0$ and $S_k = \xi_1 + \ldots + \xi_k$. Our first result shows that the j longest rows are of order n for any $p \in (0,1)$.

Theorem 1. *Let $X^{n,p}$ be as above. Then the following obtain.*

(i) *(SLLN for rows) Let $\varsigma = \inf\{k > 0 : S_k = 0\}$ be the first return time of S_k to 0. Then for any fixed $i \geq 1$,*

$$\frac{\rho_i(n)}{n} \to \mathbb{P}\Big\{\max_{0\leq k\leq \varsigma} S_k = i\Big\} > 0 \quad \text{a.s. as } n \to \infty.$$

(ii) *(CLT for the first row)*

$$\frac{\rho_1(n) - np(1-p)}{\sqrt{np(1-p)[1-3p(1-p)]}} \Rightarrow Z$$

where $Z \sim \mathcal{N}(0,1)$, the standard normal distribution.

Denote by $C(\mathbb{R})$ the space of continuous functions $f : \mathbb{R} \to \mathbb{R}$ endowed with the topology induced by the sup-norm, and let $C_0(\mathbb{R})$ be the subspace of $C(\mathbb{R})$ consisting of nonnegative compactly supported functions f such that $f \equiv 0$ on $(-\infty, 0]$. For any interval $I \subseteq \mathbb{R}$ containing 0, denote by $C(I)$ and $C_0(I)$ the space of restrictions $f|_I$ where $f \in C(\mathbb{R})$ and $f \in C_0(\mathbb{R})$, respectively. For $b \in I$, define the operator $\mathscr{E}_b : C(I) \to C(I)$ by

$$\mathscr{E}_b(f)(t) = f(t) - \min_{b\wedge t \leq s \leq b\vee t} f(s),$$

where $y \wedge z = \min(y,z)$ and $y \vee z = \max(y,z)$ for each $y,z \in \mathbb{R}$. We call b the *pivot* of $\mathscr{E}_b$. Define $\mathtt{m} : C_0(I) \to \mathbb{R}^+$ by $\mathtt{m}(g) = \sup\{x \in I : g(x) = \max(g)\}$, the location of the rightmost global maximum of g. Finally, define the *excursion operator* $\mathscr{E}$ on $C_0(I)$ by $\mathscr{E}(g) = \mathscr{E}_{\mathtt{m}(g)}(g)$. See Figure 6 for an illustration.

We now state the main result of the paper.

Theorem 2. *Let $X^{n,p}$ be as above. Then*

(i) *(Subcritical phase) For $p < 1/2$, set $\theta = (1-p)/p > 1$. The longest soliton length $\lambda_1(n)$ is concentrated around $\mu_n := \log_\theta\left(\frac{(1-2p)^2}{1-p}n\right)$ in the sense that for all $x \in \mathbb{R}$,*

$$\exp(-\theta^{-x}) \le \liminf_{n\to\infty} \mathbb{P}\left\{\lambda_1(n) \le x + \mu_n\right\} \le \limsup_{n\to\infty} \mathbb{P}\left\{\lambda_1(n) \le x + \mu_n\right\} \le \exp(-\theta^{-(x+1)}).$$

Furthermore, the sequence $\{\lambda_j(n) - \mu_n\}$ is tight for each $j \ge 1$, and $\lambda_j(n) = \Theta(\log n)$ for each fixed $j \ge 1$.

(ii) *(Critical phase) For $p = 1/2$, let $B = \{B_t\}_{0\le t\le 1}$ be a standard Brownian motion on $[0,1]$. Then for each fixed $j \ge 1$,*

$$n^{-1/2}[\lambda_1(n), \lambda_2(n), \dots, \lambda_\ell(n)] \Rightarrow [\max|B|, \max \mathscr{E}(|B|), \dots, \max \mathscr{E}^{j-1}(|B|)],$$

where $\Rightarrow$ denotes both weak and moment convergence. In particular, $\lambda_j(n) = \Theta(\sqrt{n})$ for each fixed $j \ge 1$.

(iii) *(Supercritical phase) For $p > 1/2$,*

$$\frac{\lambda_1(n) - (2p-1)n}{2\sqrt{p(1-p)n}} \Rightarrow Z \sim \mathcal{N}(0,1).$$

Furthermore, setting $\mu = p/(1-p) > 1$, we have that for any $\varepsilon > 0$, $c > 1$ and all sufficiently large n

$$\mathbb{P}\left\{\lambda_1(n) > (2p-1-\varepsilon)n)\right\} \ge 1 - cn^{-\varepsilon \log \mu/2},$$

and for any fixed $j \ge 2$,

$$\mathbb{P}\left\{\lambda_j(n) < (\varepsilon + 5/\log \mu)\log n\right\} \ge 1 - cn^{-\varepsilon \log \mu/8}.$$

We call the joint statement in Theorem 2 (iii) a *condensation phenomenon* because in the supercritical regime, a linear number of balls condense into the longest soliton while the next j longest solitons each have $O(\log n)$ balls with high probability.

Our main results have an interesting application in the context of random pattern avoiding permutations. For each $n \in \mathbb{N}$, let $\mathfrak{S}_n$ be the set of all permutations on $\{1,2,\cdots,n\}$. For any permutation $\sigma \in \mathfrak{S}_n$, denote by $\rho_1(\sigma)$ (resp., $\lambda_1(\sigma)$) the length of a longest increasing (resp., decreasing) subsequence in $\sigma(1), \sigma(2), \dots, \sigma(n)$. Given two permutations $\sigma \in \mathfrak{S}_n$ and $\tau \in \mathfrak{S}_k$ with $1 < k \le n$, we say that σ is τ*-avoiding* if no subsequence of σ has the same relative order as τ. Denote by $\mathfrak{S}_n^\tau$ the set of all τ-avoiding permutations in $\mathfrak{S}_n$. It is well known that $\sigma \in \mathfrak{S}_n$ is *stack-sortable* (resp., *stack-representable*) if and only if it is 231-avoiding (resp., 312-avoiding) [19]. Note that σ is 231-avoiding if and only if σ^{-1} is 312-avoiding.

In a classic work [19], Rotem studied properties of stack-sortable permutations chosen uniformly at random among all such permutations of a given length. He showed that if Σ^n is a permutation in $\mathfrak{S}_n^{231}$ chosen uniformly at random, then

$$\mathbb{E}[\rho_1(\Sigma^n)] = (n+1)/2, \quad \mathbb{E}[\lambda_1(\Sigma^n)] = \sqrt{\pi n} + O(1)$$

Our corollary is an extension of the above result both to the higher moments and to the k^{th} longest increasing and decreasing subsequences of Σ^n. Namely, for a given $\sigma \in \mathfrak{S}_n$, let $\sigma = \sigma_1, \sigma_2, \dots, \sigma_k$ be a sequence of permutations such that σ_{j+1} is obtained from σ_j by deleting a longest decreasing subsequence. Define $\lambda_j(\sigma) = \lambda_1(\sigma_j)$ for each $j \geq 1$, and define $\rho_i(\sigma)$ for each $i \geq 1$ similarly.

Corollary 3. *Let Σ^n be a uniformly chosen random stack-sortable permutation of length n.*

(i) *Suppose that $T_1^n, T_2^n, \cdots, T_i^n$ is a sequence of rooted trees where T_1^n is chosen uniformly at random among all rooted plane trees on $n+1$ nodes, and for $r \geq 1$, T_{r+1}^n is obtained from T_r^n by deleting all leaves. Then*

$$[\rho_1(\Sigma^n), \rho_2(\Sigma^n), \dots, \rho_i(\Sigma^n)] =_d [\#\text{ of leaves in } T_1^n, \#\text{ of leaves in } T_2^n, \dots, \#\text{ of leaves in } T_i^n].$$

Furthermore, let $\{S_k\}_{k=0}^{\infty}$ be a simple symmetric random walk with $S_0 = 0$ and let $\varsigma = \inf\{k > 0 : S_k = 0\}$ be the first return time of S_k to 0. Then for any fixed $i \geq 1$,

$$\frac{\rho_i(\Sigma^n)}{2n} \to \mathbb{P}\left\{ \max_{0 \leq k \leq \varsigma} S_k = i \right\} > 0 \quad a.s. \quad as \quad n \to \infty.$$

(ii) *Let $B^{ex} = (B_t^{ex})_{0 \leq t \leq 1}$ be a standard Brownian excursion on $[0,1]$. Then for each fixed $j \geq 1$,*

$$n^{-1/2}[\lambda_1(\Sigma^n), \lambda_2(\Sigma^n), \dots, \lambda_j(\Sigma^n)] \Rightarrow \sqrt{2}[\max B^{ex}, \max \mathscr{E}(B^{ex}), \dots, \max \mathscr{E}^{j-1}(B^{ex})],$$

where $\Rightarrow$ denotes both weak and moment convergence.

1.3. **Outline and organization.** Broadly speaking, we proceed by observing correspondences between various combinatorial objects related to box-ball configurations such as Motzkin paths, rooted forests, and 312-avoiding permutations (Figure 2). We can then interpret the rows and columns of the Young diagram associated with a box-ball configuration in terms of these objects (Figure 3). This allows us to reformulate the original soliton problem in other languages and vice versa.

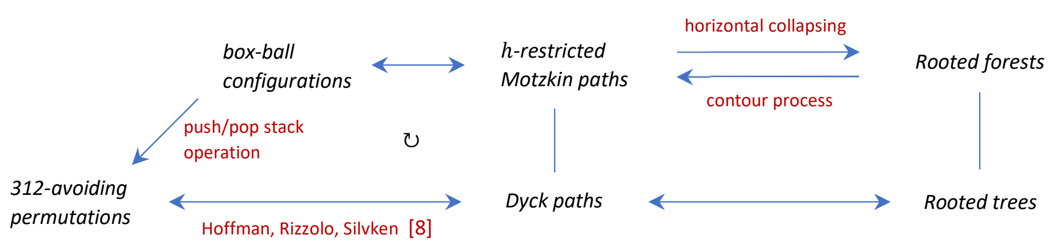

FIGURE 2. Correspondences and inclusions between six combinatorial objects. Objects in the same row are in bijective correspondence.

For us, Motzkin paths provide the most useful framework, especially in the random setting. This is because the random box-ball configuration $X^{n,p}$ can be viewed as the increment sequence of the first n steps of a simple random walk driven by the Bernoulli(p) measure. The corresponding random (h-restricted) Motzkin path is the same simple random walk except that downstrokes at height 0 are censored. The problem then essentially boils down to studying properties of the excursions of such censored random walks by using various limit theorems. The results for random Motzkin paths can then be translated back to solitons or permutations.

This paper is organized as follows: In Section 2, we describe relations between box-ball configurations, Motzkin paths, and rooted forests, and show how to construct the Young diagram from these objects. We state three important lemmas concerning Motzkin paths, their associated Young diagrams, and the 'column length functionals.' In Section 3, we discuss a correspondence between random box-ball configurations, a birth-and-death chain, and a Galton-Watson forest, and then prove Theorem 1. The proof of Theorem 2 is given in Sections 4, 5, and 6. In Section 7, we discuss a connection between box-ball configurations and pattern-avoiding permutations and prove Corollary 3. Finally, in Appendix A, we prove the three lemmas stated in Subsection 2.2.

	Box-ball configurations	Motzkin paths	Rooted forests	312-avoiding permutations
i th row length of Young diagram	Number of solitons of length $\geq i$	Number of subexcursions of height $\geq i$	Number of leaves after trimming leaves i times	Length of i th longest increasing subsequence
j th column length of Young diagram	Length of j th longest soliton	Maximum height after applying excursion operator j times	Maximum height after contracting longest path j times	Length of j th longest decreasing subsequence

FIGURE 3. Interpretation of rows and columns of the Young diagram associated with four combinatorial objects.

2. Constructing the time-invariant Young Diagram

In this section, we establish some important statements about the Young diagram which will be used crucially in later sections.

2.1. Motzkin paths.

We begin with a bijection between box-ball states and a class of lattice paths we call *h-restricted Motzkin*, a minor variant of the bijection with Dyck paths in [22]. Recall that a *Motzkin path of length ℓ* is a lattice path from $(0,0)$ to $(\ell,0)$ which never crosses below the x-axis and uses only $(1,1)$, $(1,-1)$, and $(1,0)$ steps (which we refer to as 'upstrokes,' 'downstrokes,' and 'h-strokes,' respectively). We call an infinite lattice path *Motzkin* if it is obtained by appending infinitely many h-strokes to a Motzkin path of finite length. Finally, we say that a Motzkin path *h-restricted* if its h-strokes occur only on the x-axis.

The bijection maps a (compactly supported) configuration $X : \mathbb{N} \to \{0,1\}$ to the h-restricted Motzkin path $\Gamma(X)$ defined by linear interpolation of its heights on $\mathbb{N}_0$, which are given recursively by $\Gamma(X)_0 = 0$ and

$$\Gamma(X)_{k+1} - \Gamma(X)_k = \begin{cases} +1 & \text{if } X(k+1) = 1 \\ -1 & \text{if } X(k+1) = 0 \text{ and } \Gamma(X)_k \geq 1 \\ 0 & \text{if } X(k+1) = 0 \text{ and } \Gamma(X)_k = 0 \end{cases}$$

for all $k \geq 0$. The inverse map from paths to configurations proceeds by writing a 0 for each downstroke or h-stroke and a 1 for each upstroke. See Figure 4 for an illustration.

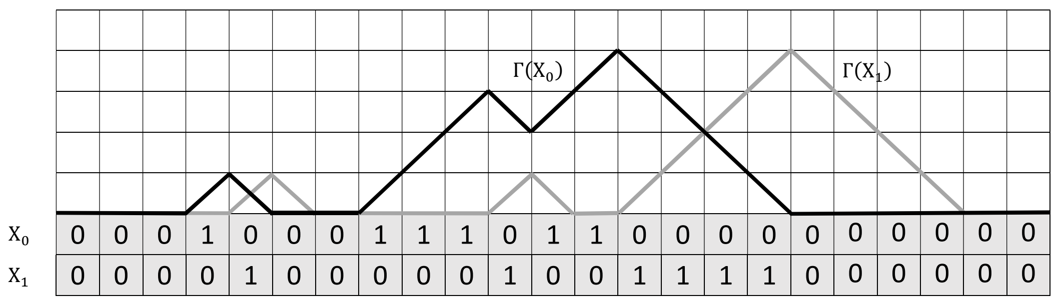

FIGURE 4. The top shaded row shows an initial box-ball configuration X_0, and the black path is $\Gamma(X_0)$. To update, balls are placed at downstrokes of $\Gamma(X_0)$, resulting in the configuration X_1 and the grey path $\Gamma(X_1)$.

The shape of this path tells us how to evolve the system by a single sweep: A ball is picked up at each upstroke and deposited at each downstroke. Specifically, label the balls $1, \dots, m$ from left to right. (This labeling applies only to states, not the system as a whole. In subsequent sweeps, the label of a particular ball may change.) Then the j^{th} upstroke occurs at the site where the carrier picks up the ball labeled j. The site at which she deposits ball j is determined by drawing a horizontal line from the center of the j^{th} upstroke to the first downstroke on its right. From this description, we see that the height of the path at any site equals the number of balls in the carrier's stack after she visits that site. When the sweep is completed, the new state of the system corresponds to the path formed by converting each downstroke to an upstroke and then uniquely completing the path so that it is h-restricted Motzkin (Figure 4). Formally, the box-ball state X_{s+1} is given in terms of the Motzkin path $\Gamma(X_s)$ by

$$X_{s+1}(k+1) = \mathbf{1}\{\Gamma(X_s)_{k+1} - \Gamma(X_s)_k = -1\} \tag{1}$$

where $\mathbf{1}$ is the indicator function.

2.2. **Hill-flattening and excursion operators.** We describe two methods of constructing a Young diagram $\Lambda(\Gamma)$ associated with a (not necessarily h-restricted) Motzkin path Γ. As usual, we denote the i^{th} row and j^{th} column by $\rho_i(\Gamma)$ and $\lambda_j(\Gamma)$, respectively.

First we give the row-wise construction using the *hill-flattening operator* $\mathscr{H}$ defined on the set of all Motzkin paths. To begin, we say that an interval $[a,b]$ with $a, b \in \mathbb{N}_0$ and $a \leq b$ is a *hill interval* of the Motzkin path Γ if for every $c \in [a,b]$, $\Gamma_{a-1} = \Gamma_c - 1 = \Gamma_{b+1}$. We write

$\mathscr{I}(\Gamma)$ for the collection of all hill intervals of Γ, and denote the number of hill intervals by $\rho(\Gamma) = |\mathscr{I}(\Gamma)|$. The hill-flattening operator $\mathscr{H}$ is then defined by

$$\mathscr{H}(\Gamma)_k = \begin{cases} \Gamma_k - 1 & \text{if } k \text{ is contained a hill interval of } \Gamma \\ \Gamma_k & \text{otherwise} \end{cases}$$

for $k \in \mathbb{N}_0$.

A *hill* of Γ is the graph of Γ over $[a-1, b+1]$ with $[a, b]$ a hill interval. Thus hills consist of a single upstroke, followed by zero or more h-strokes, followed by a single downstroke. Call a hill with no h-strokes a *peak*. Then the hill-flattening operator $\mathscr{H}$, when applied to Γ, *flattens* each hill of Γ by replacing the upstroke and downstroke with h-strokes and then lowering any intermediate h-strokes so that the path remains connected.

Note that each application of the hill-flattening operator decreases the maximum height of the Motzkin path by 1 and never increases the number of hills, so

$$\rho(\Gamma) \geq \rho(\mathscr{H}(\Gamma)) \geq \rho(\mathscr{H}^2(\Gamma)) \geq \cdots \geq \rho(\mathscr{H}^{\max\Gamma}(\Gamma)) = 0.$$

We define the Young diagram $\Lambda(\Gamma)$ associated to the Motzkin path Γ as having i^{th} row of length $\rho_i(\Gamma) = \rho(\mathscr{H}^{i-1}(\Gamma))$ for $1 \leq i \leq \max\Gamma$. Here repeated applications of $\mathscr{H}$ are denoted by $\mathscr{H}^{j+1}(f) = \mathscr{H}\big(\mathscr{H}^j(f)\big)$ with $\mathscr{H}^0$ the identity operator. In particular, given a box-ball configuration $X : \mathbb{N}_0 \to \{0, 1\}$ of finite support, we can construct the Young diagram $\Lambda(\Gamma(X))$. See Figure 5 for an illustration.

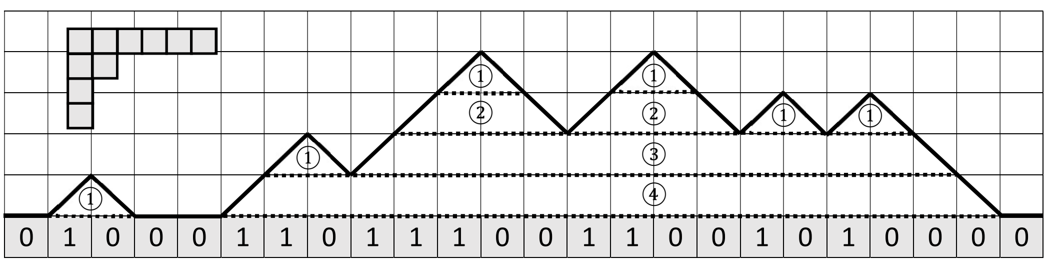

FIGURE 5. Construction of Young diagram via hill-flattening procedure applied to the Motzkin path associated with a box-ball configuration X. The bottom row is the configuration X and the black path is $\Gamma = \Gamma(X)$. Trapezoidal regions with label i are the hills of $\mathscr{H}^{i-1}(\Gamma)$, each of which becomes a distinct cell in the i^{th} row of $\Lambda(\Gamma)$. The resulting Young diagram $\Lambda(\Gamma)$ is depicted in the upper left corner.

Now consider a box-ball system $(X_s)_{s\geq 0}$ started from a configuration $X_0 : \mathbb{N}_0 \to \{0, 1\}$. The following lemma says that for each $s \geq 0$, the corresponding Young diagram $\Lambda(\Gamma(X_s))$ is independent of s and its column lengths correspond to the lengths of the solitons.

Lemma 2.1. $\Lambda(\Gamma(X_s)) = \Lambda(\Gamma(X_{s+1}))$ *for all* $s \geq 0$. *Moreover,* $\Lambda(\Gamma(X_0)) = \Lambda(X_0)$.

Next, we give the column-wise construction of $\Lambda(\Gamma)$. The key observation is that the j^{th} longest column length, which we denote by λ_j, is obtained by successively applying the excursion operator to Γ $j-1$ times and then taking a maximum.

Lemma 2.2. *Let Γ be a Motzkin path and let $\lambda_j(\Gamma)$ denote the length of the j^{th} column of $\Lambda(\Gamma)$. Then*

$$\lambda_j(\Gamma) = \max \mathscr{E}^{j-1}(\Gamma), \quad 1 \le j \le \rho(\Gamma).$$

In particular, if $(X_s)_{s\in\mathbb{N}_0}$ is any finitely supported box-ball system with initial configuration $X_0 : \mathbb{N} \to \{0,1\}$, then

$$\lambda_j(X_0) = \max \mathscr{E}^{j-1}(\Gamma(X_0)).$$

Lemma 2.2 gives the following column-wise construction of $\Lambda(\Gamma)$. Let $\mathtt{m} = \mathtt{m}(\Gamma)$ be the location of the rightmost global maximum of Γ, and set $\lambda_1(\Gamma) = \Gamma_{\mathtt{m}}$, the maximum height of Γ. To find $\lambda_2(\Gamma)$, one first computes $\mathscr{E}(\Gamma)$ by traversing Γ to the left and right of $\mathtt{m}$ as follows: Starting with height 0 at $\mathtt{m}$, move to the left, remaining at height 0 until the first local minimum, and then record the sequence of strokes until the original lattice path returns to the height of this minimum. Then repeat the process, staying at height 0 until encountering a local minimum and then recording the path of the second such excursion. Continue to the beginning of the path and then repeat the procedure moving to the right from $\mathtt{m}$. The resulting path precisely records all 'subexcursions' which are not subsumed by the maximum $(\mathtt{m}, \Gamma_{\mathtt{m}})$. $\lambda_2(\Gamma)$, the length of the second column of $\Lambda(\Gamma)$, is equal to the maximum of $\mathscr{E}(\Gamma)$. Continuing in this fashion gives $\lambda_j(\Gamma) = \max \mathscr{E}^{j-1}(\Gamma)$ for all $j \ge 1$.

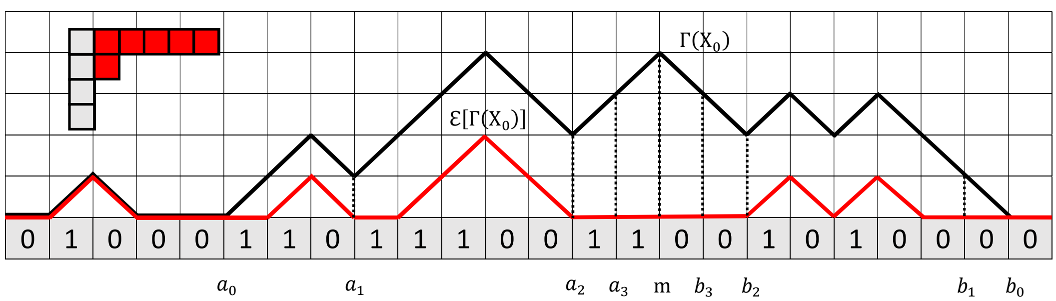

FIGURE 6. The black path is $\Gamma(X_0)$, the red path is $\mathscr{E}(\Gamma(X_0))$, and $\mathtt{m}$ is the location of the rightmost maximum of $\Gamma(X_0)$. The k^{th} box in the bottom row is $X_0(k)$, and the red Young diagram is constructed from the red path via hill flattening. The a and b terms are defined in the proof of Proposition A.2.

In light of Lemma 2.2, it is natural to call $\max \mathscr{E}^{j-1}$ the j^{th} *column length functional.* A crucial advantage of extracting the column length λ_j from the functional $\max \mathscr{E}^{j-1}$ is that this operation is continuous with respect to the topology of $C_0(\mathbb{R}^+)$, as stated in the lemma below. This enables us to take various scaling limits of the system.

Lemma 2.3. *For any interval $I \subseteq \mathbb{R}^+$, functions $f, g \in C_0(I)$, and $j \ge 1$,*

$$\left| \max \mathscr{E}^{j-1}(f) - \max \mathscr{E}^{j-1}(g) \right| \le 2\|f - g\|_\infty.$$

We relegate the proofs of Lemmas 2.1, 2.2, and 2.3 to Appendix A in order to maintain the flow of the paper.

Remark 2.4 (Depth process with drains)**.** In private communication with Jim Pitman, we learned that an operator equivalent to $\mathscr{E}_b$ was used in studying Brownian paths and continuum random trees. In our context, given a Motzkin path Γ, flip it upside down and consider it as a bucket filled to the top with water. Given $b \in \mathbb{R}^+$, put a hole at point $(b, -\Gamma(b))$. This will drain some of the water, and $-\mathscr{E}_b(\Gamma)(x)$ gives the water level at each $x \in \mathbb{R}^+$. For instance, the red path in Figure 6 can be obtained from the black one in this way with drain at $b = \mathtt{m}(\Gamma)$. A similar procedure can be defined with multiple drains. This operation was applied to Brownian paths to study, for example, the line-breaking construction of the continuum random tree in a Brownian excursion [1]; sampling bridges, meanders, and excursions at independent uniform times [15]; and developments in the tree setting with different metaphors such as "forest growth" and "bead crushing" [16, 17].

2.3. **Rooted forests.** In this subsection, we develop an alternative perspective for constructing the Young diagram from an associated rooted forest. The idea is to collapse a Motzkin path to a rooted forest by horizontal identification. Intuitively, one paints the underside of the graph of each excursion with glue and then compresses it horizontally to obtain a tree. Then the original Motzkin path can be viewed as the contour process (or Harris walk in the random setting) of the rooted forest so constructed. This point of view will be especially useful for thinking about the arguments in Section 6.

To begin, recall that a *rooted forest* is a sequence of vertex-disjoint trees $\{T_i\}_{i\geq 1}$ such that each T_i is rooted at a vertex $\mathtt{r}_i \in V(T_i)$. The *level* of a vertex $v \in T_i$ is defined as $\ell(v) = d(v, \mathtt{r}_i)$ where d is the graph distance. Given a Motzkin path Γ, we define a rooted forest $\mathfrak{F}(\Gamma)$ as follows: Let $G(\Gamma) = (V, E)$ be the graph with vertex set $V = \{(k, \Gamma_k)\}_{k\in\mathbb{N}_0} \subset \mathbb{N}_0^2$ and adjacency relation

$$(a, \Gamma_a) \overset{\text{adj}}{\sim} (b, \Gamma_b) \Longleftrightarrow |a - b| = 1 \text{ and } \Gamma_a, \Gamma_b \text{ not both } 0.$$

In words, $G(\Gamma)$ is obtained from Γ by removing the h-strokes at 0. Clearly each component of $G(\Gamma)$ is isomorphic to a path beginning and ending at height 0, and there are only finitely many such paths since Γ has finite support. Arranging the components from left to right so that their vertex labels are increasing, let P_i denote the i^{th} component from the left. Define an equivalence relation $\sim$ on the vertex set of $G(\Gamma)$ by

$$(a, \Gamma_a) \sim (b, \Gamma_b) \Longleftrightarrow \Gamma_a = \Gamma_b \leq \Gamma_s \text{ for all } t \in [a, b],$$

and write $T_i = P_i/\sim$ for the resulting rooted tree (see Figure 7). The rooted forest associated with Γ is $\mathfrak{F}(\Gamma) = \{T_i\}_{i\geq 1}$.

We can recover Γ from $\mathfrak{F}(\Gamma)$ by keeping track of the levels of the vertices explored in depth-first search. This exploration process begins at the root of T_1 and visits nodes from bottom to top and from left to right in such a way that it back-tracks to the parent of the current node only if there is no child left to visit. After exhausting all nodes in T_1, the particle moves to the second tree T_2, and so on.

More concretely, let $\iota : \mathbb{N}_0 \to V(\mathfrak{F})$ be the function which maps k to the location of the depth-first search at step k so that $\iota(0) = \mathtt{r}_1$, $\iota(k+1)$ is the leftmost unvisited child of $\iota(k)$ if such a child exists, and $\iota(k+1)$ is the parent of $\iota(k)$ if its children have all been visited (where the parent of $\mathtt{r}_i$ is taken to be $\mathtt{r}_{i+1}$). The depth-first-search ordering of the vertices of $\mathfrak{F}$ is given by $u \prec v$ if $\min\{k : \iota(k) = u\} < \min\{k : \iota(k) = v\}$. Finally, the *contour process* on

$\mathfrak{F}$ is the function $H(\mathfrak{F}) : \mathbb{N}_0 \to \mathbb{N}_0$ which maps k to the level of $\iota(k)$ in $\mathfrak{F}$. By construction, we have

$$H(\mathfrak{F})(k) = \text{level of } \iota(k) = \Gamma_k$$

for all $k \in \mathbb{N}_0$.

Now we discuss how to recover the Young diagram $\Lambda(\Gamma)$ from the corresponding rooted forest $\mathfrak{F}(\Gamma)$. In the previous subsection, we constructed the diagram from the Motzkin path via successive applications of the hill-flattening and excursion operators. In terms of rooted forests, these operators can be interpreted in terms of 'trimming' and 'lopping.' Namely, let Υ_0 be the collection of all rooted forests with finitely many vertices and consider the *trimming operator* $\mathscr{T} : \Upsilon_0 \to \Upsilon_0$ which deletes all leaves of the input forest. (See Figure 7.)

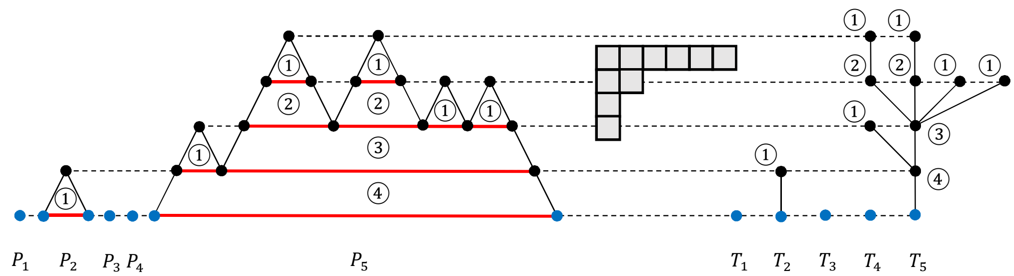

FIGURE 7. Rooted forest $\mathfrak{F}(X_0)$ corresponding to the box-ball configuration X_0 given in Figure 5. Each connected component of $G(\Gamma(X_0))$ (left) becomes a tree rooted at a blue node (right) by identifying vertices connected by the red horizontal lines. Flattening hills of $\Gamma(X_0)$ corresponds to trimming leaves from $\mathfrak{F}(X_0)$. Both procedures produce the Young diagram $\Lambda(X_0)$ shown in the middle.

Next, the *lopping operator* $\mathscr{L} : \Upsilon_0 \to \Upsilon_0$ is defined as follows: Given a rooted forest $\mathfrak{F} = \{T_i\} \in \Upsilon_0$, find the rightmost node of maximal level, say $v_{\mathtt{m}} \in V(T_k)$. Set $q = \iota(v_{\mathtt{m}})$ and let γ be the unique path from $\mathtt{r}_k$ to $v_{\mathtt{m}}$. Now let $\mathfrak{F}_1$ and $\mathfrak{F}_2$ be the rooted forests induced from $\mathfrak{F}$ such that $V(\mathfrak{F}_1) = \iota([1,q])$ and $V(\mathfrak{F}_2) = \iota([q,\infty))$. Then $\mathscr{L}(\mathfrak{F})$ is obtained by first deleting all edges contained in the copies of γ from $\mathfrak{F}_1$ and $\mathfrak{F}_2$, and then taking the union of the resulting rooted forests with components ordered according to the depth-first search. (See Figure 8.)

The following proposition shows that these operators are compatible with each other and gives a way to construct the Young diagram $\Lambda(\Gamma)$ from $\mathfrak{F}(\Gamma)$.

Proposition 2.5. *For each Motzkin path* Γ, *we have the following:*

(i) $\mathfrak{F}\big(\mathscr{H}(\Gamma)\big) = \mathscr{T}\big(\mathfrak{F}(\Gamma)\big)$.

(ii) $\mathfrak{F}\big(\mathscr{E}(\Gamma)\big) = \mathscr{L}\big(\mathfrak{F}(\Gamma)\big)$.

(iii) *For each* $1 \le i \le \max\Gamma$, $\rho_i =$ *# of leaves in* $\mathscr{T}^{i-1}(\mathfrak{F}(\Gamma))$.

(iv) *For each* $1 \le j \le \rho(\Gamma)$, $\lambda_j =$ *maximal level of nodes in* $\mathscr{L}^{j-1}(\mathfrak{F}(\Gamma))$.

Sketch of proof. For (i), note that leaves in the forest correspond to hills in the path, so applying $\mathscr{H}$ to Γ results in the forest obtained by applying $\mathscr{T}$ to $\mathfrak{F}(\Gamma)$. For (ii), observe that $\mathscr{E}$ only affects the rightmost excursion of maximum height in Γ, $\mathscr{L}$ only affects the

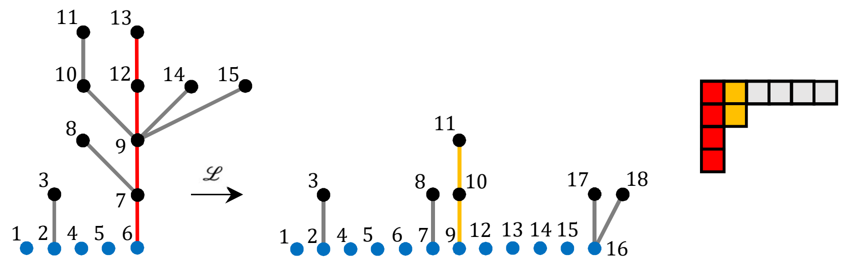

FIGURE 8. The rooted forest $\mathfrak{F} = \mathfrak{F}(\Gamma(X_0))$ on the left appeared in Figure 7, and the one on the right is $\mathscr{L}(\mathfrak{F})$. Numbers next to nodes indicate depth-first-search ordering and $q = 13$. Note that the maximum height of $\mathfrak{F}$ and $\mathscr{L}(\mathfrak{F})$ correspond to the first and second columns of $\Lambda(X_0)$, respectively.

rightmost tree of maximum height in $\mathfrak{F}(\Gamma)$, and the 'bushes' growing off of the 'trunk' of this tree correspond precisely to the subexcursions in the corresponding path component which are not subsumed by the maximum.

Now assertion (i) shows that $\mathfrak{F}(\mathscr{H}^{i-1}(\Gamma)) = \mathscr{T}^{i-1}(\mathfrak{F}(\Gamma))$ for all $1 \le i \le \max\Gamma$, and ρ_i is the number of hill intervals of $\mathscr{H}^{i-1}(\Gamma)$, which equals the number of leaves in $\mathfrak{F}(\mathscr{H}^{i-1}(\Gamma)) = \mathscr{T}^{i-1}(\mathfrak{F}(\Gamma))$, and (iii) follows. Finally, given a rooted forest $\mathfrak{F}$, denote by $\|\mathfrak{F}\|$ the maximal level of nodes in $\mathfrak{F}$. Then $\|\mathfrak{F}(\Gamma)\| = \max\Gamma$, so (ii) implies

$$\left\|\mathscr{L}^{i-1}\big(\mathfrak{F}(\Gamma)\big)\right\| = \left\|\mathfrak{F}\big(\mathscr{E}^{i-1}(\Gamma)\big)\right\| = \max\mathscr{E}^{i-1}(\Gamma) = \lambda_i(\Gamma). \qquad \square$$

We remark that Proposition 2.5 (iv) holds if we replace the lopping operator $\mathscr{L}$ by the much simpler one which simply contracts the rightmost longest path into a single root. However, for this contraction operator Proposition 2.5 (ii) no longer holds.

3. RANDOM BOX-BALL SYSTEM AND ASYMPTOTICS FOR THE ROWS

In this section, we describe random objects corresponding to the random box-ball system introduced in Subsection 1.2 and prove Theorem 1.

3.1. **Harris walks.** Fix $p \in (0,1)$, and let $\xi_1, \xi_2, \ldots$ be an i.i.d. sequence with $\mathbb{P}\{\xi_1 = 1\} = p$ and $\mathbb{P}\{\xi_1 = -1\} = 1 - p$. Let $X^p, X^{n,p} \in \{0,1\}^{\mathbb{N}}$ be as in Subsection 1.2, and let $\{S_k\}_{k=0}^{\infty}$ be the associated random walk, where $S_0 = 0$ and $S_k = \xi_1 + \cdots + \xi_n$. The *Harris walk* $\{H_k\}_{k=0}^{\infty}$ associated with X^p is defined by $H_0 = 0$ and $H_k = (H_{k-1} + \xi_k) \vee 0$ for $k \ge 1$. In other words, H_k is a simple random walk with increments ξ_j, except that downsteps at 0 are censored.

This defines an irreducible and aperiodic birth-and-death chain on $\mathbb{N}_0$ with transition probabilities $P(x, x+1) = p$, and $P(x, (x-1) \vee 0) = 1 - p$. One readily verifies that the chain is reversible with respect to the measure $\mu(x) = \theta^{-x}$ where $\theta = (1-p)/p$. Note that the sum $\sum_{k\ge1} \theta^k$ converges if and only if $p > 1/2$, so the chain is transient for these values of p and recurrent for $p \le 1/2$. It is null recurrent when $p = 1/2$ since then $\sum_{k\ge1} \theta^{-k} = \infty$, and it is positive recurrent for $p < 1/2$ as the latter sum converges in this case. (See [10] for background on recurrence criteria.) In the ergodic regime, $p < 1/2$, we can normalize μ to obtain the stationary distribution $\pi(x) = [(1-2p)/(1-p)]\theta^{-x}$.

Note that the random Motzkin path $\Gamma(X^{n,p})$ is given by the trajectory of the Harris walk up to time n, completed by appending downstrokes at the end until the height reaches 0 and appending h-strokes thereafter. More precisely, if we define $H:\mathbb{R}^+\to\mathbb{R}^+$ to be the linear interpolation of the Harris walk, $H(t)=H_{\lfloor t\rfloor}+(t-\lfloor t\rfloor)(H_{\lfloor t+1\rfloor}-H_{\lfloor t\rfloor})$, then we have

$$\Gamma(X^{n,p})(x)=H(x)\mathbf{1}_{[0,n]}+\max(0,H(n)-x+n)\mathbf{1}_{[n,\infty)}.$$

Moreover, an easy induction argument shows that for all $k\in\mathbb{N}_0$,

$$H_k=S_k-\min_{0\le r\le k}S_r.$$

Thus if $S:\mathbb{R}^+\to\mathbb{R}$ is the linear interpolation of the random walk $\{S\}_k^\infty$, then $H=\mathscr{E}_0(S)$. This observation also shows that, marginally, $H_k=_d\max_{0\le j\le k}S_j$.

3.2. **Galton-Watson forests.** Following the procedure outlined in Subsection 2.3, one can construct a random rooted forest $\mathfrak{F}(X^{n,p})=\mathfrak{F}\big(\Gamma(X^{n,p})\big)$ from the trajectory of the truncated Harris walk $\Gamma(X^{n,p})$, and it turns out that $\mathfrak{F}(X^{n,p})$ has the same law as the sub-forest of a Galton-Watson forest with mean offspring number $p/(1-p)$ consisting of the first n nodes revealed by depth-first search.

To be precise, let $\{\zeta_j^k\}_{j,k\ge1}$ be an array of i.i.d. $\mathbb{N}_0$-valued random variables, and define the sequence $\{Z_k\}_{k\ge0}$ by $Z_0=1$ and

$$Z_{k+1}=\begin{cases}\zeta_1^{k+1}+\cdots+\zeta_{Z_k}^{k+1} & \text{if } Z_k>0\\ 0 & \text{if } Z_k=0.\end{cases}$$

The interpretation is that Z_k is the population size in the k^{th} generation of a species in which individuals survive for a single generation and produce an i.i.d. number of offspring before dying. ζ_j^{k+1} is the number of offspring of the j^{th} individual in generation k, and the common law of the ζ's is called the *offspring distribution*. The family tree T for this population is known as a *Galton-Watson tree*. We will be interested in Galton-Watson trees with offspring distribution

$$\mathbb{P}\{\zeta_j^k=x\}=p^x(1-p),\quad x\in\mathbb{N}_0,$$

which is the number of independent Bern(p) trials preceding the first failure. Note that $\mathbb{E}[\zeta_j^k]=p/(1-p)$, so T is subcritical if $0<p<1/2$, critical if $p=1/2$, and supercritical if $1/2<p<1$. The law of a Galton-Watson tree with Geometric$(1-p)$ offspring distribution will be denoted by $\texttt{GWT}(p)$.

We call a sequence of i.i.d. Galton Watson trees $\mathfrak{F}_{GW}=\{T_i\}_{i\ge1}$ a *Galton-Watson forest*, and write $\texttt{GWF}(p)$ for the law of a forest of i.i.d. $\texttt{GWT}(p)$ trees. It is well known that for $0<p\le1/2$, each component T_i is finite with full probability, so the depth-first-search visits all nodes in the forest. However, for $p>1/2$, each component has a positive probability of being infinite, so almost surely there exists an index $I<\infty$ such that $|T_i|<\infty$ for all $i<I$ and $|T_I|=\infty$. Thus for $p>1/2$, the depth-first-search cannot pass beyond the leftmost infinite branch in T_I (see Figure 9).

Now let $\mathfrak{F}_p\sim\texttt{GWF}(p)$. Write $\mathfrak{F}_{n,p}$ for the vertex-induced subforest of $\mathfrak{F}_p$ on the set of nodes $\iota([1,n])\subseteq V(\mathfrak{F}_p)$ which are visited by the depth-first-search in the first n steps, and write $\texttt{GWF}(n,p)$ for the law of $\mathscr{F}_{n,p}$.

Proposition 3.1. $\mathfrak{F}(X^{n,p}) \sim \mathtt{GWF}(n,p)$.

Proof. Let $\Gamma = \Gamma(X^p)$ and $\mathfrak{F} = \mathfrak{F}(\Gamma)$. Denote by Z_v the number of children of node $v \in V(\mathfrak{F})$. We are going to show that Z_v's are i.i.d. and they have the law of the number of independent Bernoulli p trials before the first failure. This will imply that the Harris walk $\{H_k\}_{k=0}^{\infty}$ is distributed as the contour process of $\mathfrak{F}_p$. Then the relation between $\Gamma(X^{n,p})$ and H from the previous subsection yields the assertion.

Let $\mathfrak{F}(X^p) = \{T_i\}_{i\geq 1}$ and fix a node $v \in V(T_i)$ for some $i \geq 1$. Let P_i be the path component in $G(\Gamma)$ which is collapsed to T_i via the equivalence relation $\sim$. Note that the number of nodes in P_i which are identified with v equals the number of children of v. Let $x = (a_0, \Gamma_{a_0})$ be such a vertex of P_i with a_0 minimal. If $\Gamma_{a_0+1} - \Gamma_{a_0} = \xi_{a_0+1}$ is 1, then the depth-first search finds the first child of v; otherwise, v is childless and the search moves to its parent or to the root of next tree T_{i+1} depending on whether $\Gamma_{a_0} \geq 1$ or $\Gamma_{a_0} = 0$. If $\xi_{a_0+1} = 1$, then let $a_1 = \min\{k \geq a_0 : \Gamma_k = \Gamma_{a_0}\}$ be the first return time to level Γ_{a_0} after a_0. As before, the depth-first search finds the second child of v if and only if $\xi_{a_1+1} = 1$. Continuing thusly, we see that Z_v has a $\mathrm{Geom}(p)$ distribution, and the proof is complete. □

Proposition 3.1 allows us to describe the joint distribution of the first i rows or the first j columns in the random box-ball system started at $X^{n,p}$ in terms of Galton-Watson Forests.

Corollary 3.2. *Suppose that $\mathfrak{F} \sim \mathtt{GWF}(n,p)$. For each $i \geq 1$, let $\mathfrak{l}_i$ and $\mathfrak{h}_i$ be the number of leaves in $\mathscr{T}^{i-1}(\mathfrak{F})$ and the maximum height of $\mathscr{L}^{i-1}(\mathfrak{F})$, respectively. Then for any $1 \leq i \leq \max(\Gamma(X^{n,p}))$ and $1 \leq j \leq \rho(\Gamma(X^{n,p}))$, we have*

$$[\rho_1(n), \rho_2(n), \cdots, \rho_i(n)] =_d [\mathfrak{l}_1, \mathfrak{l}_2, \cdots, \mathfrak{l}_i]$$

and

$$[\lambda_1(n), \lambda_2(n), \cdots, \lambda_j(n)] =_d [\mathfrak{h}_1, \mathfrak{h}_2, \cdots, \mathfrak{h}_j].$$

3.3. **Asymptotics for the Rows.** In this subsection, we prove our first main result, Theorem 1. From the construction described in Subsection 2.2, we have that $\rho_1(n)$, the length of the first row of $\Lambda^{n,p}$, equals the number of peaks in $\Gamma(X^{n,p})$, which equals the number of 10 patterns in $X^{n,p}$. In general, $\rho_i(n)$ is the number of subexcursions of height i in the Harris walk $\{H_k\}_{k=0}^{n}$, and these can also be understood in terms of certain binary patterns in the initial configuration.

We begin with a proof of the $i = 1$ case of Theorem 1 using arguments from renewal theory. Strong laws for the other rows can be deduced similarly by considering analogous (delayed) renewal processes, but we will find it more convenient to pursue an alternative approach that will be of use in Section 7.

Proof of Theorem 1 for i = 1. First observe that the number of solitons in $X^{n,p}$ is equal to the number of 10 patterns, so $\rho_1(n) = \mathbf{1}\{\xi_n = 1\} + N_{10}(n)$ where $N_{10}(n)$ is the number of 10 patterns in the first n terms. Because of the scaling, it suffices to prove that $N_{10}(n) = \sum_{i=1}^{n-1} \mathbf{1}\{\xi_i = 1, \xi_{i+1} = -1\}$ satisfies the asserted limit theorems.

Now $N_{10}(n)$ counts occurrences of 'head, tail' patterns in a sequence of independent coin flips, which we can think of as a renewal process. Let T_{10} be distributed as the inter-event times in this process. Then the elementary renewal theorem gives $\mathbb{E}[N_{10}(n)]/n \to$

$1/\mathbb{E}[T_{10}]$. Since $\mathbb{E}[N_{10}(n)] = (n-1)p(1-p)$, it follows from the strong law for renewal processes that

$$\frac{N_{10}(n)}{n} \to \frac{1}{\mathbb{E}[T_{10}]} = p(1-p) \text{ a.s..}$$

Renewal theory also shows that $N_{10}(n)$ converges weakly to a standard normal random variable when appropriately normalized [3]. To compute the variance, we write $W_i = \mathbf{1}\{\xi_i = 1, \xi_{i+1} = -1\}$ and observe that $\mathbb{E}[W_i] = p(1-p)$, $\mathbb{E}[W_i W_{i+1}] = 0$, and $\mathbb{E}[W_i W_j] = p^2(1-p)^2$ when $|i-j| > 1$, hence

$$\mathbb{E}[N_{10}(n)^2] = \sum_{i=1}^{n-1} \mathbb{E}[W_i^2] + \sum_{|i-j|>1} \mathbb{E}[W_i W_j] = (n-1)p(1-p) + (n-2)(n-3)p^2(1-p)^2,$$

so

$$\operatorname{Var}\big(N_{10}(n)\big) = \mathbb{E}[N_{10}(n)^2] - \mathbb{E}[N_{10}(n)]^2 = (n-1)p(1-p) - (3n-5)p^2(1-p)^2.$$

The second part of the theorem follows upon invoking Slutsky's theorem to simplify the expression $(N_{10}(n) - \mathbb{E}[N_{10}(n)])/\operatorname{Var}(N_{10}(n))^{1/2}$. □

Remark 3.3. The normal convergence of $\rho_1(n)$ can also be established using Stein's method for sums of locally dependent random variables (see [2, Ch. 9]). Though this approach is more involved, it has the upshot of supplying a Berry-Esseen rate of order $O(n^{-1/2})$. One can show that a central limit theorem also holds for the other row lengths by a similar renewal theory argument, but the corresponding variance computations are not as straightforward.

To treat the $i > 1$ case, we need to establish some more notation and a useful lemma. To begin, let $\gamma : \mathbb{N}_0 \to \mathbb{Z}$ be any nearest neighbor lattice path (so that $|\gamma_{k+1} - \gamma_k| \in \{-1, 0, 1\}$ for all $k \in \mathbb{N}_0$). We say γ has a *subexcursion of height h* on the interval $[r, t]$ if $\gamma_r = \gamma_t < \gamma_s$ for all $r < s < t$ and $\max_{r<s<t} \gamma_s - \gamma_r = h$. In this case, we say the subexcursion begins at r and ends at s.

Let $\{S_k\}_{k=0}^{\infty}$ be the simple random walk with increment distribution $\mathbb{P}\{S_{k+1} - S_k = 1\} = 1 - \mathbb{P}\{S_{k+1} - S_k = -1\} = p$. For each $n \geq 1$, define

$$N_i(n) = \sum_{\ell=0}^{n} \mathbf{1}\{S_k \text{ has subexcursion of height } i \text{ beginning at } k = \ell\}. \tag{2}$$

The following lemma establishes a Chernoff-Hoeffding bound for $N_i(n)$.

Lemma 3.4. *Let $\varsigma = \inf\{k > 0 : S_k = 0\}$ be the first return time of S_k to zero. Fix $i \geq 1$, and set $\mu_i = \mathbb{P}\{\max_{0 \leq k \leq \varsigma} S_k = i\}$. Then there exist constants $C, D > 0$ such that for each $\varepsilon > 0$ and $n \geq 1$,*

$$\mathbb{P}\left\{\left|\frac{N_i(n)}{n} - \mu_i\right| > \varepsilon\right\} \leq C e^{-D\varepsilon^2 n}.$$

Proof. Linearity of expectation gives

$$\mathbb{E}[N_i(n)/n] = \mathbb{P}\big\{S_k \text{ has subexcursion of height } i \text{ beginning at } k = 0\big\} = \mu_i.$$

Observe that the indicator variables in the definition of $N_i(n)$ have negative association. Indeed, for a fixed interval J of integers, and fixed $i \geq 0$, consider the event

$$A(J) = \{\text{Random walk } S_k \text{ restricted to the interval } J \text{ is a subexcursion of height } i\ \}.$$

By the Markov property, the probability of an intersection of events $A(J_1), A(J_2), \cdots$ equals the product of the probabilities if the J_i's are disjoint, and 0 otherwise. This justifies the negative association claimed above. Now the assertion follows from Chernoff-Hoeffding bounds for negatively associated random variables (see, e.g., Proposition 5 of [6]). □

Proof of Theorem 1 for $\boldsymbol{i \geq 1}$. To treat $\rho_i(n)$ for $i > 1$, we need to consider subexcursions of height i in the truncated Harris walk $\{H_k\}_{k=0}^n$. The hill-flattening procedure produces a unique column of length at least h for each such subexcursion, so $\rho_i(n)$ is the number of height i subexcursions of H on $[0, n]$. Since the Harris walk H_k and the associated simple random walk S_k over $[0, n]$ share subexcursions of positive height, we may regard $\rho_i(n)$ as the number of subexcursions of S_k occuring on $[0, n]$. Furthermore, we can approximate $\rho_i(n)$ by $N_i(n)$ defined in (2) since the two only differ when H has a subexcursion of height at least i beginning at or after $n - i$, hence $|N_i(n) - \rho_i(n)| \leq 1$. Therefore, the assertion follows from Lemma 3.4 and the first Borel-Cantelli lemma. □

4. TOP SOLITON LENGTHS IN THE SUBCRITICAL REGIME

In this section, we fix $p \in (0, 1/2)$ and prove the following slightly stronger version of Theorem 2 (i):

Theorem 4.1. *Set $\theta = (1-p)/p$, $\sigma = (1-2p)/(1-p)$, and $\mu_n = \log_\theta\big((1-2p)\sigma n\big)$. Let $\lambda_j(n)$ be the length of the j^{th} longest soliton in the random box-ball system $X^{n,p}$. Then for any non-decreasing real sequence $\{x_n\}_{n\geq 1}$,*

$$\liminf_{n\to\infty} \exp\left(\theta^{-x_n}\right)\mathbb{P}\left\{\lambda_1(n) \leq x_n + \mu_n\right\} \geq 1,$$

and

$$\limsup_{n\to\infty} \exp\left(\theta^{-(x_n+1)}\right)\mathbb{P}\left\{\lambda_1(n) \leq x_n + \mu_n\right\} \leq 1.$$

Furthermore, for every fixed $j \geq 1$ and $x \in \mathbb{R}$,

$$\limsup_{n\to\infty} \mathbb{P}\left\{\lambda_j(n) \leq x + \mu_n\right\} \leq \exp\left(-\theta^{-(x+1)}\right)\sum_{k=0}^{j-1}\frac{\theta^{-kx}}{k!}.$$

Remark 4.2. The first part of Theorem 2 (i) is the special case where the sequence $\{x_n\}$ is constant and implies the tightness of $\{\lambda_1(n) - \mu_n\}$. Also, the last part of Theorem 4.1 shows that

$$\lim_{x\to-\infty}\liminf_{n\to\infty} \mathbb{P}(\lambda_j(n) - \mu_n > x) = 1 - \lim_{x\to-\infty}\limsup_{n\to\infty} \mathbb{P}(\lambda_j(n) - \mu_n \leq x) \geq 1.$$

On the other hand, $\lambda_j(n) \leq \lambda_1(n)$, so we have

$$\lim_{x\to\infty}\liminf_{n\to\infty} \mathbb{P}(\lambda_j(n) - \mu_n \leq x) \geq \lim_{x\to\infty}\liminf_{n\to\infty} \mathbb{P}(\lambda_1(n) - \mu_n \leq x) \geq 1,$$

hence $\{\lambda_j(n) - \mu_n\}$ is tight for each $j > 1$ as well. This implies $\lambda_j(n) = \Theta(\log n)$ for all $j \geq 1$. Thus Theorem 2 (i) follows from Theorem 4.1.

We will find the more general statement of Theorem 4.1 useful in Section 6. Roughly speaking, we proceed by showing that the Harris walk $\{H_k\}_{k=0}^{\infty}$ has $\Theta(n)$ excursions by time n. By relating the excursion heights to a gambler's ruin problem, we argue that their distribution has an exponential tail. Taking the maximum over the $\Theta(n)$ excursions shows that the law of $\lambda_1(n)$ is approximated by a Gumbel distribution after scaling appropriately. To treat the $j>1$ case, we appeal to the hill-flattening procedure described in Subsection 2.2.

To begin, set $\tau_1=0$ and for $k>1$, define $\tau_k=\inf\{j>\tau_{k-1}:H_j=0\}$ to be the time of the k^{th} visit to 0. Thus τ_k is the beginning of the k^{th} excursion above the x-axis, and τ_{k+1} is the end of the k^{th} such excursion. (In this section, if the random walk stays at 0, this counts as an excursion of height 0.) Let

$$h_k=\sup\{H_t:\tau_k<t\le\tau_{k+1}\}=\sup\left\{\sum_{i=\tau_k+1}^{t}\xi_i:\tau_k<t\le\tau_{k+1}\right\}\vee 0$$

be the maximum height of the k^{th} excursion. The strong Markov property ensures that $h_1,h_2,\ldots$ are i.i.d. $\mathbb{N}_0$-valued random variables. To compute their distribution function, $F(x)=\mathbb{P}\{h_1\le x\}$, we observe that $\mathbb{P}\{h_1=0\}=1-p$ and $\mathbb{P}\{h_1\le x\}=\mathbb{P}\{1\le h_1\le x\}+\mathbb{P}\{h_1=0\}$ for $x\ge 1$. In order for the event $\{1\le h_1\le x\}$ to occur, the random walk must begin with an upstep and then visit zero before visiting $x+1$. The latter occurs with the 'gambler's ruin' probability that a simple random walker, started at the origin and moving right with probability p, hits -1 before hitting x, which is given by $\big(\theta^x-1\big)/\big(\theta^x-\theta^{-1}\big)$ [7, Ch. 5.7]. Putting all of this together shows that $F(x)=(1-p)+p\big(\theta^x-1\big)/\big(\theta^x-\theta^{-1}\big)$ for all $x\in\mathbb{N}_0$. After a bit of rearranging, we get

$$F(x)=\left(1-\frac{1-2p}{\theta^{\lfloor x\rfloor+1}-1}\right)\mathbf{1}_{[0,\infty)}(x).$$

Now let $h_{1:m},\ldots,h_{m:m}$ denote the order statistics of $h_1,\cdots,h_m$ so that $h_{1:m}\ge\cdots\ge h_{m:m}$ and $\{h_{1:m},\ldots,h_{m:m}\}=\{h_1,\ldots,h_m\}$. Then

$$F_{j:m}(x):=\mathbb{P}\{h_{j:m}\le x\}=\sum_{k=0}^{j-1}\binom{m}{k}F(x)^{m-k}(1-F(x))^k,\quad j=1,\ldots,m.$$

In particular, the maximum $h_{1:m}$ has distribution function

$$F_{1:m}(x)=\left(1-\frac{1-2p}{\theta^{\lfloor x\rfloor+1}-1}\right)^m\mathbf{1}_{[0,\infty)}(x).$$

Write $M_n=\sup\{k:\tau_{k+1}\le n\}$ for the number of excursions completed by time n and let $r_n=\max\{\sum_{i=\tau_{M_n+1}}^{r}\xi_i:\tau_{M_n+1}\le r\le n\}$ be the maximum height attained after the last complete excursion. We are interested in the order statistics for $h_1,\ldots,h_{M_n},r_n$, which we denote by $h_1(n)\ge h_2(n)\ge\cdots\ge h_{M_n+1}(n)$. We begin by showing that M_n is sharply concentrated around its mean so that we can essentially treat it as a deterministic sequence.

Proposition 4.3. *If M_n is the number of excursions of H completed by time n, then*

$$\frac{M_n}{n}\to\frac{1-2p}{1-p}\ \text{a.s.}\quad\text{as } n\to\infty.$$

Proof. We may write $M_n = \sum_{k=1}^{n} \mathbf{1}\{H_k = 0\}$, the number of visits to 0 in $[1, n]$. Since the Harris chain is ergodic with stationary distribution $\pi(x) = [(1-2p)/(1-p)]\theta^{-x}$ for $p < 1/2$, we can apply the Markov chain ergodic theorem to obtain

$$\frac{M_n}{n} \to \pi(0) = \frac{1-2p}{1-p} \text{ a.s.} \qquad \square$$

The next ingredient in our argument is a simple stochastic monotonicity result.

Proposition 4.4. *Set $\sigma = (1-2p)/(1-p)$, $p \in (0, 1/2)$. For any real sequence $\{x_n\}_{n\geq 1}$ and any positive integer j, we have that for all $\varepsilon > 0$,*

$$\limsup_{n\to\infty} \mathbb{P}\left\{h_j(n) \leq x_n\right\} \leq \limsup_{n\to\infty} \mathbb{P}\left\{h_{j:\lfloor(\sigma-\varepsilon)n\rfloor} \leq x_n\right\}$$

and

$$\liminf_{n\to\infty} \mathbb{P}\{h_j(n) \leq x_n\} \geq \liminf_{n\to\infty} \mathbb{P}\left\{h_{j:\lceil(\sigma+\varepsilon)n\rceil} \leq x_n\right\}.$$

Proof. Define

$$N^-(n,\varepsilon) = \sup\{t : M_t \leq (\sigma-\varepsilon)n\}$$

and

$$N^+(n,\varepsilon) = \inf\{t : M_t \geq (\sigma+\varepsilon)n\}.$$

It follows from Proposition 4.3 that there is an a.s. finite N such that

$$\{h_1, \ldots, h_{M_{N^-(n,\varepsilon)}}\} \subseteq \{h_1, \ldots, h_{M_n+1}\} \subseteq \{h_1, \ldots, h_{M_{N^+(n,\varepsilon)}}\}$$

with probability one for all $n \geq N$. Because $r_n \leq h_{M_n+1}$, this means that, almost surely, $h_{j:M_{N^-(n,\varepsilon)}} \leq h_j(n) \leq h_{j:M_{N^+(n,\varepsilon)}}$ for n sufficiently large, hence

$$\limsup_{n\to\infty} \mathbb{P}\{h_j(n) \leq x_n\} \leq \limsup_{n\to\infty} \mathbb{P}\left\{h_{j:M_{N^-(n,\varepsilon)}} \leq x_n\right\}$$

and

$$\liminf_{n\to\infty} \mathbb{P}\{h_j(n) \leq x_n\} \geq \liminf_{n\to\infty} \mathbb{P}\left\{h_{j:M_{N^+(n,\varepsilon)}} \leq x_n\right\}.$$

The desired assertion follows by noting that $M_{N^-(n,\varepsilon)} = \lfloor(\sigma-\varepsilon)n\rfloor$ and $M_{N^+(n,\varepsilon)} = \lceil(\sigma+\varepsilon)n\rceil$ a.s. since 0 is a recurrent state of $\{H_k\}$ $\square$

With these results in hand, we are now in a position to prove the main results of this section.

Proof of Theorem 4.1. Fix $\varepsilon > 0$ and a non-decreasing sequence $\{x_n\}_{n\geq 1}$, and define $\mu_n = \log_\theta\left((1-2p)\sigma n\right)$. We first recall that for any deterministic sequence of integers $\{b_n\}$,

$$\mathbb{P}\{h_{1:b_n} \leq x_n + \mu_n\} = \left(1 - \frac{1-2p}{\theta^{\lfloor x_n+\mu_n\rfloor+1} - 1}\right)^{b_n} \mathbf{1}_{\{x_n+\mu_n\geq 0\}}.$$

Writing $\nu_n = (x_n + \mu_n) - \lfloor x_n + \mu_n \rfloor$, we have

$$\frac{1-2p}{\theta^{\lfloor x_n+\mu_n\rfloor+1} - 1} = \frac{1-2p}{\theta^{x_n+\mu_n}\theta^{1-\nu_n} - 1} = \frac{\theta^{-x_n}}{\sigma n \theta^{1-\nu_n} - (1-2p)^{-1}\theta^{-x_n}},$$

so, since $\theta > 1$ and $1 - \nu_n \in (0, 1]$, we see that

$$\frac{\theta^{-(x_n+1)}}{\sigma n - (1-2p)^{-1}\theta^{-(x_n+1)}} \leq \frac{1-2p}{\theta^{\lfloor x_n+\mu_n\rfloor+1} - 1} \leq \frac{\theta^{-x_n}}{\sigma n - (1-2p)^{-1}\theta^{-x_n}}. \tag{3}$$

Next, we claim that for any sequence $\{b_n\}_{n\geq 1}$ with $\lim_{n\to\infty} b_n/n = c > 0$ and any non-decreasing sequence $\{y_n\}_{n\geq 1}$, we have

$$\lim_{n\to\infty} \exp\left((c/\sigma)\theta^{-y_n}\right)\left(1-\frac{\theta^{-y_n}}{\sigma n-(1-2p)^{-1}\theta^{-y_n}}\right)^{b_n} = 1.$$

Indeed,

$$\begin{aligned}\log\left[\left(1-\frac{\theta^{-y_n}}{\sigma n-(1-2p)^{-1}\theta^{-y_n}}\right)^{b_n}\exp\left((c/\sigma)\theta^{-y_n}\right)\right]\\ = \theta^{-y_n}(b_n/\theta^{-y_n})\log\left(1-\frac{\theta^{-y_n}}{\sigma n-(1-2p)^{-1}\theta^{-y_n}}\right)+(c/\sigma)\theta^{-y_n}\\ = \theta^{-y_n}\left[\frac{c}{\sigma}+\frac{\log\left(1-\frac{\theta^{-y_n}}{\sigma n-(1-2p)^{-1}\theta^{-y_n}}\right)}{\theta^{-y_n}/b_n}\right].\end{aligned}$$

Since $\theta > 1$ and $\{y_n\}_{n\geq 1}$ is non-decreasing, θ^{-y_n} is bounded. The claim follows since a Taylor expansion of the log term shows that

$$\lim_{n\to\infty}\frac{\log\left(1-\frac{\theta^{-y_n}}{\sigma n-(1-2p)^{-1}\theta^{-y_n}}\right)}{\theta^{-y_n}/b_n} = -\frac{c}{\sigma}.$$

For the first assertion, Proposition 4.4 and the preceding claim with $y_n = x_n+1$, $b_n = \lfloor(\sigma-\varepsilon)n\rfloor$ show that

$$\begin{aligned}\limsup_{n\to\infty}\ \exp((1-\varepsilon/\sigma)\theta^{-(x_n+1)})\mathbb{P}\left\{h_1(n)\leq x_n+\mu_n\right\}\\ \leq \limsup_{n\to\infty}\ \exp((1-\varepsilon/\sigma)\theta^{-(x_n+1)})\mathbb{P}\left\{h_{1:\lfloor(\sigma-\varepsilon)n\rfloor}\leq x_n+\mu_n\right\}\\ \leq \limsup_{n\to\infty}\ \exp((1-\varepsilon/\sigma)\theta^{-(x_n+1)})\left(1-\frac{\theta^{-(x_n+1)}}{\sigma n-(1-2p)^{-1}\theta^{-(x_n+1)}}\right)^{b_n} = 1.\end{aligned}$$

Similarly, taking $y_n = x_n$ and $b_n' = \lceil(\sigma+\varepsilon)n\rceil$, gives

$$\begin{aligned}\liminf_{n\to\infty}\ \exp((1+\varepsilon/\sigma)\theta^{-x_n})\mathbb{P}\left\{h_1(n)\leq x_n+\mu_n\right\}\\ \geq \liminf_{n\to\infty}\ \exp((1+\varepsilon/\sigma)\theta^{-x_n})\mathbb{P}\left\{h_{1:\lceil(\sigma+\varepsilon)n\rceil}\leq x+\mu_n\right\}\\ \geq \liminf_{n\to\infty}\ \exp((1+\varepsilon/\sigma)\theta^{-x_n})\left(1-\frac{\theta^{-x_n}}{\sigma n-(1-2p)^{-1}\theta^{-x_n}}\right)^{b_n'} = 1.\end{aligned}$$

Letting $\varepsilon\searrow 0$ and noting that $h_1(n) = \lambda_1(n)$ completes the argument.

For the second assertion, let μ_n and b_n be as before, and fix $j\in\mathbb{N}$ and $x\in\mathbb{R}$. Recall that for all n large enough that $x\geq -\mu_n$, we have

$$\begin{aligned}\mathbb{P}\left\{h_{j:b_n}\leq x+\mu_n\right\} &= \sum_{k=0}^{j-1}\binom{b_n}{k}\left(1-\frac{1-2p}{\theta^{\lfloor x+\mu_n\rfloor+1}-1}\right)^{b_n-k}\left(\frac{1-2p}{\theta^{\lfloor x+\mu_n\rfloor+1}-1}\right)^k\\ &\leq \left(1-\frac{1-2p}{\theta^{\lfloor x+\mu_n\rfloor+1}-1}\right)^{b_n}\sum_{k=0}^{j-1}\frac{1}{k!}b_n^k\left(1-\frac{1-2p}{\theta^{\lfloor x+\mu_n\rfloor+1}-1}\right)^{-k}\left(\frac{1-2p}{\theta^{\lfloor x+\mu_n\rfloor+1}-1}\right)^k\end{aligned}$$

$$\leq \left(1 - \frac{1-2p}{\theta^{\lfloor x+\mu_n \rfloor+1}-1}\right)^{b_n} \sum_{k=0}^{j-1} \frac{(1-\varepsilon/\sigma)^k}{k!} \left(\theta^x - \frac{2(1-p)^2}{(1-2p)^2 n}\right)^{-k},$$

where the final inequality used the upper bound in (3). Accordingly, we have

$$\limsup_{n\to\infty} \mathbb{P}\left\{h_j(n) \leq x + \mu_n\right\} \leq \exp\left(-(1-\varepsilon/\sigma)\theta^{-(x+1)}\right) \sum_{k=0}^{j-1} \frac{\theta^{-kx}}{k!}(1-\varepsilon/\sigma)^k,$$

and since Lemma 2.2 shows that $\lambda_j(n) \geq h_j(n)$ for all $1 \leq j \leq M_n$, the result follows upon taking $\varepsilon \searrow 0$. □

5. TOP SOLITON LENGTHS AT CRITICALITY

In this section we observe that when $p = 1/2$, the (suitably scaled) Harris walk converges weakly to a reflected Brownian motion at the process level. This then enables us to deduce scaling limits for the top soliton lengths.

Theorem 5.1. *Let $\{B(t) : 0 \leq t \leq 1\}$ be a standard Brownian motion and define $H^n(t) = H(nt)/\sqrt{n}$ for $0 \leq t \leq 1$. Then for $p = 1/2$,*

$$\{H^n(t) : 0 \leq t \leq 1\} \Rightarrow \{|B(t)| : 0 \leq t \leq 1\} \text{ in } C([0,1]),$$

where $\Rightarrow$ denotes polynomial convergence. That is, if $F : C([0,1]) \to \infty$ is any continuous functional of polynomial growth (so there exists $r \geq 1$ such that $|F(\gamma)| \leq \|\gamma\|_\infty^r$ for all $\gamma \in C([0,1])$), then

$$\lim_{n\to\infty} \mathbb{E}[F(H^n)] = \mathbb{E}(F(|B|)).$$

Proof. By Theorem 9 in [5], we only need to show that H^n converges weakly to $|B|$. Recall from Subsection 1.2 that the linear interpolation of the $p = 1/2$ Harris walk is given by $H(t) = \mathscr{E}_0(S)(t) = S(t) - \min_{0\leq r\leq t} S(r)$ where S is the linear interpolation of symmetric simple random walk.

Donsker's Theorem shows that after scaling diffusively, $S(t)$ converges weakly to a standard Brownian motion in the space $C([0,1])$. That is, writing $S^n(t) = S(nt)/\sqrt{n}$, we have

$$\lim_{n\to\infty} \mathbb{E}[(F(S^n)] = \mathbb{E}[F(B)]$$

for every bounded and continuous functional $F : C([0,1]) \to \mathbb{R}$.

A direct computation shows that for any fixed $b \in [0,1]$, $\mathscr{E}_b$ is (2-Lipschitz) continuous and satisfies $\mathscr{E}_b(cf) = c\mathscr{E}_b(f)$ for all $b, c \geq 0$ (see Proposition A.5 (i) in Subsection A.3), so for every bounded and continuous $G : C([0,1]) \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}[G(H^n)] = \lim_{n\to\infty} \mathbb{E}[G(\mathscr{E}_0(S^n))] = \mathbb{E}[G(\mathscr{E}_0(B))],$$

hence H^n converges weakly to $\mathscr{E}_0(B)$. As

$$\mathscr{E}_0(B)(t) = B(t) - \min_{0\leq s\leq t} B(s) =_d -B(t) - \min_{0\leq s\leq t}(-B(s)) = \max_{0\leq s\leq t} B(s) - B(t),$$

Lévy's $M-B$ theorem (see [13, Ch. 2.3]) implies $\mathscr{E}_0(B) =_d |B|$ and the proof is complete. □

***Proof of Theorem 2* (ii).** First recall that the Motzkin path $\Gamma := \Gamma(X^{n,1/2})$ agrees with the Harris walk H on $[0,n]$, and has only downstrokes until it reaches height 0 on $[n,\infty)$, hence all of its peaks are contained in $[0,n]$. Recall also that the excursion operator deletes the peak at the rightmost maximum and preserves all the other peaks. Thus by Lemma 2.2, we have

$$n^{-1/2}\lambda_j(n) = n^{-1/2}\max \mathscr{E}^{j-1}(\Gamma) = n^{-1/2}\max_{[0,n]} \mathscr{E}^{j-1}(H|_{[0,n]}) = \max_{0\le t\le 1} \mathscr{E}^{j-1}(H^n).$$

Lemma 2.3 in Section 2 shows that the column length functionals $\max \mathscr{E}^{j-1} : C([0,1]) \to \mathbb{R}$, $j \ge 1$ are Lipschitz. In particular, they are functionals of polynomial growth. The weak and moment convergence assertions then follow from Theorem 5.1 by taking $F = \mathbf{1}\{f_j \le x\}$ and $F = f_j^m$ for $x \in [0,1]$ and $m \in \mathbb{N}$, where $f_j = \max \mathscr{E}^{j-1}$. A stronger version of the second part of the assertion is shown in Theorem 5.3 below. □

To establish the order of the other top soliton lengths, we appeal to known results about the marginal densities of the ranked maxima of $|B|$ over all excursions. To state our conclusions precisely, note that the continuity of B ensures that the random subset $\{t : B(t) \neq 0\}$ of $[0,1]$ is a countable union of maximal disjoint intervals, called the *excursion intervals* of B. We call an excursion interval (a,b) *complete* if $B(a) = B(b) = 0$, and *incomplete* otherwise. All of the excursion intervals are complete except possibly the last one $(g(t),1]$, where $g(t) = \sup\{0 \le t \le 1 : B(t) = 0\}$ is the last zero of B. Let

$$\mathtt{h}_1 \ge \mathtt{h}_2 \ge \cdots > 0$$

be the ranked sequence of values $\sup_{t\in(a,b)} |B_t|$ as (a,b) ranges over all excursion intervals of B. The marginal distributions of the ranked heights over excursions in the reflected Brownian bridge were first obtained by Pitman and Yor [18]. Lagnoux, Mercier, and Vallois [12] pointed out that the probability that the maximum of reflected Brownian motion is obtained during the last incomplete excursion is approximately 0.3069. Csaki and Hu [4] obtained the following explicit expressions for the marginal densities of ranked maxima of reflected Brownian motion over all excursions, including the final meander:

Theorem 5.2. *For each $j \ge 1$ and $y > 0$,*

$$\mathbb{P}\{\mathtt{h}_j \ge y\} = 2^{j+1}\sum_{k=0}^{\infty}(-1)^k\binom{k+j-1}{k}\Big(1-\Phi\big((2k+2j-1)y\big)\Big)$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Accordingly, Theorem 5.1 and Lemma 2.2 imply

Theorem 5.3. *At criticality, we have that for each $x > 0$*

$$\lim_{n\to\infty}\mathbb{P}\left\{\lambda_1(n) \le x\sqrt{n}\right\} = 1 - 2\sum_{k=0}^{\infty}(-1)^k(1-\Phi[(2k+1)x]).$$

Furthermore,

$$\limsup_{n\to\infty}\mathbb{P}\left\{\lambda_j(n) \le x\sqrt{n}\right\} \le 1 - 2^{j+1}\sum_{k=0}^{\infty}(-1)^k\binom{k+j-1}{k}(1-\Phi[(2k+2j-1)x]).$$

In particular, for any $j \ge 1$, $\lambda_j(n) = \Theta(\sqrt{n})$.

6. Top soliton lengths in the supercritical regime

In this section, we fix $p \in (1/2, 1)$ and prove Theorem 2 (iii). The intuition is the following. According to Proposition 3.1, the top soliton lengths are encoded in the first $N \sim \text{Binomial}(n, p)$ edges of a Galton-Watson forest $\mathfrak{F} = (T_i)_{i\geq 1} \sim \texttt{GWF}(p)$. Since the offspring distribution has mean $p/(1-p) > 1$ in the supercritical regime, the random index $I = \min\{i : |T_i| = \infty\}$ is almost surely finite. For n large, about np nodes of the infinite component T_I will be exposed by the Harris walk, which climbs up along the 'leftmost' infinite branch in T_I. Hence $\lambda_1(n)$ should behave like the maximum of a random walk with positive drift, and λ_2 is the maximum height of the first few finite components $T_1, \ldots, T_{I-1}$ together with the 'bushes' attached to the infinite branch in T_I. The $\lambda_1(n)$ assertion follows by approximating $\{H_k\}$ by $\{S_k\}$. To see that the probability that $\lambda_2(n) > c \log n$ is small for a suitable $c > 0$, we appeal to a duality argument: A backward Harris walk started at the last node will encounter a subcritical Galton-Watson forest, so its maximum height should be $\Theta(\log n)$ (see Figure 9).

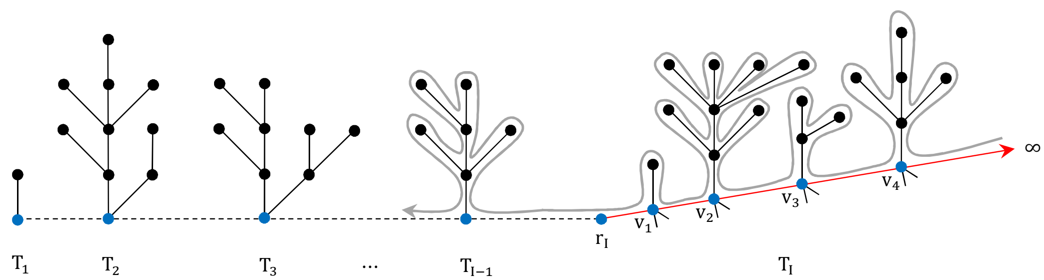

FIGURE 9. Supercritical Galton-Watson forest. T_I is the first infinite component and the red ray is the leftmost infinite branch in T_I on which the usual Harris walk climbs up. The grey contour is the backward Harris walk starting from the last vertex of level N, which encounters a subcritical Galton-Watson forest.

To make the above sketch rigorous, we introduce the notion of a dual configuration. Given a random box-ball configuration $X^{n,p}$, we define its *dual* as

$$\widehat{X}^{n,p}(k) = \left(1 - X^{n,p}(n-k+1)\right) \mathbf{1}\{1 \leq k \leq n\}.$$

Alternatively, if $X^{n,p}$ is defined in terms of $\xi_1, \xi_2, \ldots, \xi_n, 0, 0, \ldots$, then its dual is defined in terms of $-\xi_n, -\xi_{n-1}, \ldots, -\xi_1, 0, 0, \ldots$. For $p \in (1/2, 1)$, the dual configuration has the same law as the subcritical configuration $X^{n,1-p}$. Let $\widehat{\lambda}_j(n)$ be the length of the j^{th} longest soliton in the dual configuration. The key lemma in this section, Lemma 6.2, establishes that $\lambda_1(n)$ and $\lambda_j(n)$ can be approximated by $S_n = \xi_1 + \cdots + \xi_n$ and $\widehat{\lambda}_{j-1}(n)$, respectively.

Positive drift ensures that S and H are not too different, so the first claim seems reasonable since S should attain its maximum over $[0, n]$ near n. To explain why the second claim is true, let $\widehat{H} \in C_0(\mathbb{R}^+)$ be the Harris walk for the dual configuration so that $\widehat{\lambda}_1(n) = \max \widehat{H}$. Now H and $\widehat{H}$ are coupled in such a way that the latter is a time-reversal of $\mathscr{E}_n(S)$, which is approximated by $\mathscr{E}_n(H)$. Thus it all boils down to showing that the path $\mathscr{E}_n(H)$ pivoted at n is close to $\mathscr{E}(H) = \mathscr{E}_{\mathtt{m}}(H)$, pivoted at the actual location $\mathtt{m} = \mathtt{m}(H)$ of the rightmost maximum of H. But again the positive drift ensures that H attains its maximum near the end.

Continuity of the excursion operator can then be used to show that the two paths must be close to each other.

We begin with the observation that the maximum of the biased random walk $\{S_k\}_{k=0}^n$ is near S_n.

Proposition 6.1. *Fix $\varepsilon > 0$ and $\mu = p/(1-p) > 1$. Then for any $c > 1$*

$$\mathbb{P}\left\{\max_{1\le k\le n} S_k - S_n > (\varepsilon + 1/\log\mu)\log n\right\} \le c\mu^{-\varepsilon\log n}$$

for all sufficiently large n.

Proof. Let $\widetilde{X}^p$ be the subcritical box-ball configuration obtained by switching 0's to 1's and 1's to 0's in X^p, and denote $\widetilde{X}^{n,p} = \widetilde{X}^p\mathbf{1}_{[1,n]}$ for $n \in \mathbb{N}$. Note that $\widetilde{X}^{n,p}$ has the same law as $X^{n,1-p}$. For each $1 \le k \le n$, the associated random walk $\widetilde{S}_k$ and Harris walk $\widetilde{H}_k$ satisfy $\widetilde{S}_k = (-\xi_1) + \cdots + (-\xi_k) = -S_k$ and

$$\widetilde{H}_k = \widetilde{S}_k - \min_{1\le i\le k} \widetilde{S}_i = \max_{1\le i\le k} S_i - S_k.$$

Thus by Lemma 2.2, we have

$$\max_{1\le k\le n} S_k - S_n = \widetilde{H}_n \le \max_{1\le k\le n} \widetilde{H}_k = \widetilde{\lambda}_1(n),$$

where $\widetilde{\lambda}_1(n)$ is the longest soliton length in the subcritical configuration $\widetilde{X}^{n,p}$.

Now let p_n be the probability in the assertion and set $\kappa = (2p-1)^2/p < 1$. Then

$$\mathrm{p}_n \le 1 - \mathbb{P}\{\widetilde{\lambda}_1(n) \le \varepsilon\log n + \log_\mu(\kappa n)\}.$$

As Theorem 4.1 implies

$$\mathbb{P}\{\widetilde{\lambda}_1(n) \le \varepsilon\log n + \log_\mu(\kappa n)\} \ge e^{-c\mu^{-\varepsilon\log n}} \ge 1 - c\mu^{-\varepsilon\log n}$$

for all sufficiently large n, the assertion follows. □

The following lemma establishes our key observation about the duality between the supercritical and subcritical box-ball systems:

Lemma 6.2. *Fix $\varepsilon > 0$, $j \in \mathbb{N}$, and $\mu = p/(1-p) > 1$. Then for any $c > 1$*

$$\mathbb{P}\left\{|\lambda_1(n) - S_n| > (\varepsilon + 2/\log\mu)\log n\right\} \le c\mu^{-\frac{\varepsilon}{2}\log n}$$

and

$$\mathbb{P}\left\{\left|\lambda_{j+1}(n) - \widehat{\lambda}_j(n)\right| > (\varepsilon + 4/\log\mu)\log n\right\} \le c\mu^{-\frac{\varepsilon}{4}\log n}$$

for all sufficiently large n.

Proof. Fix $n \ge j$ and define random variables

$$R = \sup_{k\in\mathbb{N}}\left|\min_{1\le i\le k} S_i\right| \quad\text{and}\quad Q_n = \max_{1\le k\le n} S_k - S_n.$$

We first show that it suffices to establish the inequalities

$$\left|\lambda_1(n) - S_n\right| \le R + 2Q_n \quad\text{and}\quad \left|\lambda_{j+1}(n) - \widehat{\lambda}_j(n)\right| \le 2R + 4Q_n. \tag{4}$$

Indeed, by considering whether or not $R > \varepsilon\log(n)$, we get

$$\mathbb{P}\left\{R + 2Q_n > (\varepsilon + 2/\log\mu)\log n\right\} \le \mathbb{P}\left\{R > \varepsilon\log n\right\} + \mathbb{P}\left\{Q_n > (\varepsilon/2 + 1/\log\mu)\log n\right\}.$$

Casing out according to the value of ξ_1 shows that for any integer $k \geq 3$, $\mathbb{P}\{R \leq k\} = p\mathbb{P}\{R \leq k+1\} + (1-p)\mathbb{P}\{R \leq k-1\}$, hence

$$\mathbb{P}\{R = k\} = \mathbb{P}\{R \leq k\} - \mathbb{P}\{R \leq k-1\} = p\big(\mathbb{P}\{R \leq k+1\} - \mathbb{P}\{R \leq k-1\}\big) = p\big(\mathbb{P}\{R = k\} + \mathbb{P}\{R = k+1\}\big),$$

so $\mathbb{P}\{R = k+1\} = \mu^{-1}\mathbb{P}\{R = k\}$. It follows that $\mathbb{P}\{R > \ell\} = \mathbb{P}\{R = 3\}\mu^{2-\ell}/(1-\mu)$ for all $\ell \geq 3$. In particular,

$$\mathbb{P}\left\{R > \varepsilon \log n\right\} < \frac{p^3}{(1-p)^2(2p-1)}\mu^{-\varepsilon \log n}$$

for n large. Since Proposition 6.1 shows that

$$\mathbb{P}\left\{Q_n > (\varepsilon/2 + 1/\log \mu)\log n\right\} \leq \frac{c+1}{2}\mu^{-\frac{\varepsilon}{2}\log n}$$

as well, we see that it suffices to prove (4).

The first inequality in (4) follows from Lemma 2.2 and the triangle inequality upon observing that

$$\left|\max_{1\leq k\leq n} H_k - \max_{1\leq k\leq n} S_n\right| \leq \max_{1\leq k\leq n} |H_k - S_k| = \max_{1\leq k\leq n}\left|\min_{1\leq i\leq k} S_i\right| \leq \sup_{k\in\mathbb{N}}\left|\min_{1\leq i\leq k} S_i\right| = R.$$

To establish the second inequality, let $n^* := \mathtt{m}(S\mathbf{1}_{[0,n]})$ denote the rightmost maximum of S on $[0,n]$, and define the sequence of random variables $\{\check{S}_k\}_{0\leq k\leq n}$ by $\check{S}_k = S_k$ for all $k \neq n$ and $\check{S}_n = S_{n^*}$. As usual, let $\check{S}$ denote the linear interpolation of $\{\check{S}_k\}$. By construction, $\|\check{S} - S\|_\infty = Q_n$. Also, observe that $\mathscr{E}_n(S)(n) = 0 = \mathscr{E}(\check{S})(n)$, and for $0 \leq j < n$, writing $m_j = \min\{S_j, \dots, S_{n-1}\}$, we have $\mathscr{E}_n(S)(j) = S_j - \min\{m_j, S_n\}$ and $\mathscr{E}(\check{S})(j) = S_j - \min\{m_j, S_{n^*}\} = S_j - m_j$. If $\min\{m_j, S_n\} = S_n$, then $m_j = S_n + 1$. It follows that

$$\mathscr{E}_n(S)(j) = \mathscr{E}(\check{S})(j) + m_j - \min\{m_j, S_n\} = \mathscr{E}(\check{S})(j) + \mathbf{1}\{S_n < m_j\}.$$

Writing $\widehat{S}_k = -(S_n - S_{n-k})$ for the random walk associated with the dual configuration, we see that the Harris walk $\widehat{H}_k$ can be written as

$$\widehat{H}_k = (S_{n-k} - S_n) - \min_{0\leq j\leq k}(S_{n-j} - S_n) = S_{n-k} - \min_{n-k\leq i\leq n} S_i = \mathscr{E}(\check{S})(n-k) + \mathbf{1}\{S_n < m_{n-k}\}$$

for all $0 \leq k \leq n$. As $S_n < m_{n-k}$ implies $Q_n = \|\check{S} - S\|_\infty \geq 1$, we have $\left|\widehat{H}_k - \mathscr{E}(\check{S})(n-k)\right| \leq Q_n$ for all k. Since the functional $\max \mathscr{E}^{j-1}$ is invariant under time reversal, the above observation together with the Lemmas 2.2 and 2.3 yields

$$\left|\widehat{\lambda}_j(n) - \max \mathscr{E}^j(\check{S})\right| = \left|\max \mathscr{E}^{j-1}(\widehat{H}) - \max \mathscr{E}^j(\check{S})\right| \leq 2Q_n.$$

Finally, the triangle inequality, Lemma 2.2, and Lemma 2.3 give

$$\begin{aligned}\left|\lambda_{j+1}(n) - \widehat{\lambda}_j(n)\right| &\leq \left|\max \mathscr{E}^j(H) - \max \mathscr{E}^j(S)\right| + \left|\max \mathscr{E}^j(S) - \max \mathscr{E}^j(\check{S})\right| + 2Q_n \\ &\leq 2\|H - S\|_\infty + 2\|S - \check{S}\|_\infty + 2Q_n \leq 2R + 4Q_n.\end{aligned}$$

□

Now we are ready to complete the proof of Theorem 2.

***Proof of Theorem 2* (iii).** First, we may write

$$\frac{\lambda_1(n) - (2p-1)n}{2\sqrt{p(1-p)n}} = \frac{\lambda_1(n) - S_n}{2\sqrt{p(1-p)n}} + \frac{S_n - (2p-1)n}{2\sqrt{p(1-p)n}}.$$

Since the first term on the right-hand side converges to zero in probability by Lemma 6.2 and the second term converges in distribution to a standard normal by the usual central limit theorem, the first part of the assertion follows from Slutsky's theorem.

Next, Lemma 6.2 tells us that we can approximate $\lambda_1(n)$ by S_n and $\lambda_2(n)$ by $\widehat{\lambda}_1(n)$. For the former, we have

$$\mathbb{P}\left\{\lambda_1(n) < (2p-1-\varepsilon)n\right\} \le \mathbb{P}\left\{S_n < (2p-1-\varepsilon/2)n\right\} + \mathbb{P}\{|S_n - \lambda_1(n)| > \varepsilon n/2\}.$$

Since S_n is a random walk with mean $(2p-1)n$ and increments supported on $\{-1,1\}$, Hoeffding's inequality shows that the first term bounded above by $2e^{-\frac{\varepsilon^2 n}{8}}$, which is less than $\frac{c-1}{2}\mu^{-\frac{\varepsilon}{2}\log n}$ for n sufficiently large. Also, the second term is less than $\frac{c+1}{2}\mu^{-\frac{\varepsilon}{2}\log n}$ by the first part of Lemma 6.2.

We proceed similarly for the $\lambda_j(n)$ inequality. Since $\lambda_j(n) \le \lambda_2(n)$ for all $j \ge 2$, it suffices to show the assertion for $j = 2$. Breaking up the event in question according to the size of $\widehat{\lambda}_1(n)$, we can write

$$\begin{aligned}\mathbb{P}\left\{\lambda_2(n) > (\varepsilon + 5/\log\mu)\log n\right\} \le\ & \mathbb{P}\left\{\widehat{\lambda}_1(n) > (\varepsilon/2 + 1/\log\mu)\log n\right\} \\ & + \mathbb{P}\left\{|\lambda_2(n) - \widehat{\lambda}_1(n)| > (\varepsilon/2 + 4/\log\mu)\log n\right\}.\end{aligned}$$

Since $\widehat{\lambda}_1(n) =_d \widetilde{\lambda}_1(n)$, the proof of Proposition 6.1 shows that the first term is eventually bounded by $c\mu^{-\frac{\varepsilon}{2}\log n}$. Finally, Lemma 6.2 shows that the second term is at most $c\mu^{-\frac{\varepsilon}{8}\log n}$ for n large. □

7. Application to random stack-sortable permutations

In this section, we discuss some relations between box-ball systems and stack-sortable permutations and prove Corollary 3.

We begin by explaining (an equivalent version of) the construction of the time-invariant Young diagram introduced in [22], which was built upon a connection between box-ball configurations and stack-representable permutations. The first step is to map a box-ball configuration X_0 of m balls to a 312-avoiding permutation $\sigma = \sigma(X_0) \in \mathfrak{S}_m$ using the pushing and popping stack operations from [11, Ch. 2.2.1]. To do so, label the balls $1,\dots,m$ from left to right so that the i^{th} ball gets label i. Then the one-line notation for σ gives the left to right labels of the balls after a single update $X_0 \mapsto X_1$. That is, we push the symbol 1 onto an empty stack at the first ball and then, advancing to the right, pop the top of the stack off for storage at each empty box and push k onto the stack upon encountering the k^{th} ball. See Figure 10 for an illustration.

To get a Young diagram from this stack-representable permutation $\sigma(X_0)$, one applies the Robinson-Schensted (RS) algorithm (see, e.g., [20, Ch. 3.1]) to obtain a pair of standard Young tableaux, and records their common shape as $\mathtt{RS}(\sigma(X_0))$. Greene's theorem [8] relates the sum of the lengths of the first k rows (resp. columns) of the diagram of σ to the length of a longest k increasing (resp. k decreasing) subsequence in σ, and we show in Proposition 7.3 that more can be said about the individual row/columns lengths when σ is 312-avoiding.

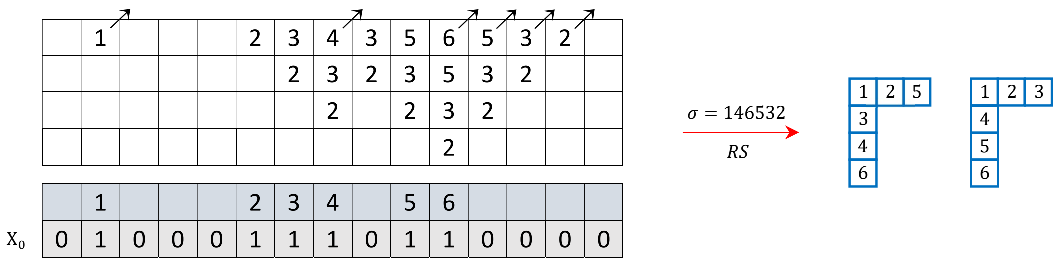

FIGURE 10. Construction of the 312-avoiding permutation corresponding to the box-ball environment in the bottom row via push-pop operations. The second row from the bottom indicates the labels of the balls, and the columns in the upper table give the contents of the right-sweeping stack. The resulting permutation is $\sigma = 146532$.

It was proven in [22] that $\mathtt{RS}(\sigma(X_s))$ is invariant in $s \geq 0$ and its j^{th} column length is the j^{th} longest soliton length in the system. Thus, by Lemma 2.1, this construction gives the same Young diagram which was obtained by hill-flattening operations applied to the Motzkin path.

Proposition 7.1. *Let $X_0 : \mathbb{N}_0 \to \{0,1\}$ be a finitely supported box-ball configuration. Then*

$$\mathtt{RS}(\sigma(X_0)) = \Lambda(X_0) = \Lambda(\Gamma(X_0)).$$

The following proposition shows that there is a bijection between 312-avoiding permutations of length n and Dyck paths of length $2n$ which 'factors through' box-ball configurations in a natural way.

Proposition 7.2. *Let $\mathfrak{S}_n^{312}$ be the set of all 312-avoiding permutations and let Dyck_{2n} be the set of all Dyck paths of length $2n$.*

(i) *There exists a bijection $\varphi : \text{Dyck}_{2n} \to \mathfrak{S}_n^{312}$.*

(ii) *For each $\tau \in \mathfrak{S}_n^{312}$ and $\Gamma \in \text{Dyck}_{2n}$ such that $\varphi(\Gamma) = \tau$, there is a box-ball configuration X_0 such that $\tau = \sigma(X_0)$ and $\Gamma = \Gamma(X_0)$.*

Recall that for a general permutation σ, Greene's theorem tells us that the sum of the first k row (resp., column) lengths of $\mathtt{RS}(\sigma)$ equals the length of the longest subsequence of σ obtained by taking a disjoint union of k increasing (resp., decreasing) subsequences. The next proposition shows that if σ is 312-avoiding, then we can in fact interpret the length of the k^{th} row (resp., column) of $\mathtt{RS}(\sigma)$ as the length of the k^{th} longest increasing (resp., decreasing) subsequence in σ. Proofs of Propositions 7.2 and 7.3 are given in Appendix A.4.

Proposition 7.3. *Let σ be a 312-avoiding permutation of length $n \geq 1$. For each $k \geq 1$, let ρ_k (resp., λ_k) denote the length of k^{th} row (resp., column) of $\mathtt{RS}(\sigma)$. Then ρ_k (resp., λ_k) equals the length of the k^{th} longest increasing (resp., decreasing) subsequence in σ.*

We are now able to prove Corollary 3 using similar ideas from the proof of Theorem 1 together with some known results on random Dyck paths and random walk excursions.

Proof of Corollary 3. Let Γ be a Dyck path of length $2n$ and let $\tau = \varphi(\Gamma)$ be the corresponding 312-avoiding permutation. Proposition 7.2 enables us to choose a box-ball configuration X_0 such that $\tau = \sigma(X_0)$ and $\Gamma = \Gamma(X_0)$, and Proposition 7.1 implies that $\mathtt{RS}(\tau) = \Lambda(\Gamma)$. If we denote by Σ^n and Γ^n uniformly random elements of $\mathfrak{S}_n^{312}$ and Dyck_{2n}, this yields

$$\mathtt{RS}(\Sigma^n) =_d \Lambda(\Gamma^n). \tag{5}$$

Now the contour process described in Subsection 2.3 gives a bijection between Dyck paths of length $2n$ and rooted plane trees with $n+1$ nodes, so the first part of (i) follows from (5) and Proposition 2.5.

Part (ii) also follows easily from known results. Indeed, it is well known that under diffusive scaling the random walk excursion converges weakly to a standard Brownian excursion [1]. Moreover, by [5, Theorem. 9], the convergence is also polynomial in the sense of Theorem 5.1). Thus (ii) follows from (5) and Lemmas 2.2 and 2.3.

Lastly, we establish the strong law for $\rho_i(\Gamma^n)$ stated in the second part of (i). To begin, fix $i \geq 1$, and let $(S_k)_{k\geq 0}$ be a simple symmetric random walk with $S_0 = 0$. We may view the uniformly random Dyck path Γ^n of length $2n$ as the trajectory of S_k over the interval $[0,2n]$ conditioned to stay non-negative and satisfy $S_{2n} = 0$. By (5) and the hill-flattening procedure, $\rho_i(\Gamma^n)$ equals the number of subexcursions of Γ^n of height i. Let $N_i(n)$ and μ_i be as in (2) and Lemma 3.4. Then $\rho_i(\Gamma^n) \leq N_i(2n) \leq \rho_i(\Gamma^n)+1$, so for all $n \geq 1$ and $\varepsilon < 1/2n$,

$$\begin{aligned}\mathbb{P}\left\{\left|\frac{\rho_i(\Gamma^n)}{2n} - \mu_i\right| > 2\varepsilon\right\} &\leq \mathbb{P}\left\{\left|\frac{N_i(2n)}{2n} - \mu_i\right| > \varepsilon \,\middle|\, S_k \text{ is a Dyck path over } [0,2n]\right\} \\ &\leq \frac{\mathbb{P}\{|N_i(2n)/2n - \mu_i| > \varepsilon\}}{\mathbb{P}\left\{S_k \text{ is a Dyck path over } [0,2n]\right\}}.\end{aligned}$$

It is well known that the number of Dyck paths of length $2n$ is the n^{th} Catalan number $\frac{1}{n+1}\binom{2n}{n}$, so by Stirling's approximation, $\mathbb{P}\left\{S_k \text{ is a Dyck path over } [0,2n]\right\} \sim n^{-3/2}/\sqrt{\pi}$. Now by the Chernoff-Hoeffding bound (Lemma 3.4) with $\varepsilon = \varepsilon(n) = \sqrt{3\log n/2Dn} \searrow 0$, we get

$$\mathbb{P}\left\{\left|\frac{\rho_i(\Gamma^n)}{2n} - \mu_i\right| > 2\varepsilon(n)\right\} = O(n^{-3/2}).$$

In particular, these probabilities are summable, so Borel-Cantelli I implies $\rho_i(\Gamma^n)/2n \to \mu_i$ a.s. as $n \to \infty$. This shows the assertion. □

Appendix A. Proofs of Combinatorial Lemmas

In this appendix, we provide proofs of Lemmas 2.1, 2.2, and 2.3, and Propositions 7.3, which we have assumed in the earlier sections.

A.1. **Time invariance of the Young diagram.** Our proof of Lemma 2.1 is similar to the argument from [22], which is formulated in terms of Dyck words intead of Motzkin paths. The argument is simplified by Proposition A.1.

To begin, recall that given a box-ball configuration X_s of finite support, the associated lattice path $\Gamma(X_s)$ is constructed by reading X_s from left to right: Starting at height 0, increase by 1 every time a 1 is encountered, decrease by 1 whenever a 0 is encountered at positive height, and remain at height 0 otherwise. A simple but useful observation is that reading X_s from right to left produces the lattice path $\Gamma(X_{s-1})$. More precisely, let $(X_s)_{t\geq 0}$ be a box-ball system started from a finitely supported configuration X_0. For each $t \geq 0$, let $r_s = \max\{k \geq 0 : X_s(k) = 1\}$ be the location of the rightmost 1 at time s. Construct a (backward) lattice path $\overleftarrow{\Gamma}(X_s) : \mathbb{N}_0 \to \mathbb{N}_0$ by $\overleftarrow{\Gamma}(X_s)_k = 0$ for $k \geq r_s$ and

$$\overleftarrow{\Gamma}(X_s)_k = \begin{cases} \overleftarrow{\Gamma}(X_s)_{k+1} + 1 & \text{if } X_s(k+1) = 1 \\ \overleftarrow{\Gamma}(X_s)_{k+1} - 1 & \text{if } X_s(k+1) = 0 \text{ and } \overleftarrow{\Gamma}(X_s)_{k+1} \geq 1 \\ 0 & \text{if } \overleftarrow{\Gamma}(X_s)_{k+1} = X_s(k+1) = 0 \end{cases}$$

for $0 \leq k < r_s$. See Figure A.1 for an illustration. In this appendix, we denote the ordinary lattice path Γ by $\vec{\Gamma}$ to emphasize the reading direction.

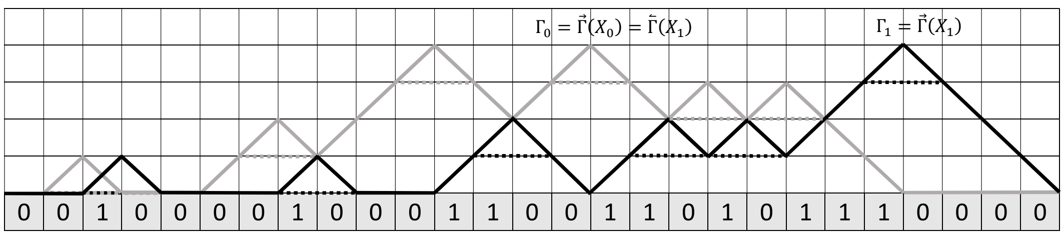

FIGURE A.1. The environment is X_1 where X_0 is the environment given in Figure 5. The black path is $\vec{\Gamma}(X_1)$ and the grey path is $\overleftarrow{\Gamma}(X_1)$. Notice that the latter coincides with the black path in Figure 5.

Proposition A.1. *For all $t \geq 0$,*

$$\overleftarrow{\Gamma}(X_{s+1}) = \vec{\Gamma}(X_s).$$

Proof. Fix $t \geq 0$, and observe that both paths are 0 on $[r_{t+1}, \infty)$, so the assertion holds on this interval. Now suppose the paths agree on $[k+1, \infty)$ for some $k < r_{t+1}$. We must show that $\overleftarrow{\Gamma}(X_{s+1})_k = \vec{\Gamma}(X_s)_k$.

The definition of the box-ball dynamics shows that $X_{s+1}(k+1) = 1$ if and only if $\vec{\Gamma}(X_s)_k - 1 = \vec{\Gamma}(X_s)_{k+1}$, hence

$$\begin{aligned} \vec{\Gamma}(X_s)_k - \vec{\Gamma}(X_s)_{k+1} = 1 \quad &\Longleftrightarrow \quad X_{s+1}(k+1) = 1 \\ &\Longleftrightarrow \quad \overleftarrow{\Gamma}(X_{s+1})_k - \overleftarrow{\Gamma}(X_{s+1})_{k+1} = 1. \end{aligned}$$

The induction hypothesis implies

$$\begin{aligned} \vec{\Gamma}(X_s)_k = \vec{\Gamma}(X_s)_{k+1} &\Longleftrightarrow X_{s+1}(k+1) = 0 \text{ and } \vec{\Gamma}(X_s)_{k+1} = 0 \\ &\Longleftrightarrow X_{s+1}(k+1) = 0 \text{ and } \overleftarrow{\Gamma}(X_{s+1})_{k+1} = 0 \\ &\Longleftrightarrow \overleftarrow{\Gamma}(X_{s+1})_k = \overleftarrow{\Gamma}(X_{s+1})_{k+1} \end{aligned}$$

and

$$\begin{aligned} \vec{\Gamma}(X_s)_k - \vec{\Gamma}(X_s)_{k+1} = -1 &\Longleftrightarrow X_{s+1}(k+1) = 1 \text{ and } \vec{\Gamma}(X_s)_{k+1} \geq 1 \\ &\Longleftrightarrow X_{s+1}(k+1) = 1 \text{ and } \overleftarrow{\Gamma}(X_{s+1})_{k+1} \geq 1 \end{aligned}$$

$$\Longleftrightarrow \overleftarrow{\Gamma}(X_{s+1})_k - \overleftarrow{\Gamma}(X_{s+1})_{k+1} = -1.$$

This establishes the assertion. □

To facilitate the proof of Lemma 2.1, it is convenient to reformulate the procedure for building Young diagrams row by row: Rather than flattening hills, we can *contract* peaks by deleting the upstroke-downstroke pair and then identifying the endpoints so that the path remains connected. The number of hills after flattening is the same as the number of peaks after contracting, so everything is exactly same as before. The advantage here is that if one begins with an h-restricted Motzkin path, then the hills are always peaks and the Motzkin paths are always h-restricted. Moreover, the contraction operation can be understood in terms of the environment as deleting 1 0 patterns.

***Proof of Lemma** 2.1.* The second part of the assertion clearly holds for all stable box-ball configurations $X_0 : \mathbb{N} \to \{0,1\}$ of finite support. Since the system always stabilizes, the second part follows from the time invariance as state in the first part.

Now let $(X_s)_{s\geq 0}$ be as before. To show the time invariance of $\Lambda(X_s)$, recall that the construction of $\Lambda(X_s)$ begins by counting the number of peaks in the path corresponding $X_s = X_s^{(0)}$. This is equal to the number of 1 0 patterns, which is equal to the number of 1-strings, which is equal to the number of 0 1 patterns. The length of the first row of $\Lambda(X_s)$ is given by this number. The peaks are then contracted by deleting the 1 0 patterns from X_s to obtain $X_s^{(1)}$ and the process is repeated with $\Gamma(X_s^{(1)})$. At each step, the 1-strings are counted, the diagram is updated, and the 1 0 patterns are deleted, continuing until the path consists only of h-strokes.

The key insights are that the number of 1 strings is the same regardless of whether the environment is read from left to right or conversely, and that the number of 1-strings after 1 0 patterns are deleted is the same as the number of 1 strings after 0 1 patterns are deleted. In the first case, each 1 string either decreases in length by 1 (possibly dissapearing), or it merges with the string on its right. In the second, each string either decreases in length by 1 or merges with the string on its left.

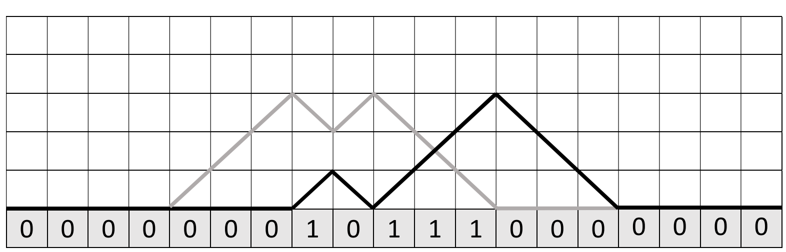

FIGURE A.2. The environment is formed by deleting either 1 0 patterns or 0 1 patterns from the environment in Figure A.1. The corresponding left-right (black) and right-left (grey) lattice paths have the same number of hills as the flattened paths in Figure A.1.

Now for any fixed $t \geq 0$, $\vec{\Gamma}(X_s)$ and $\vec{\Gamma}(X_{s+1})$ can be read off from X_{s+1} by proceeding from right to left and from left to right, respectively. The update rule for the former is to count 1-strings and then delete 1 0 patterns, and the update rule for the latter is to count 1-strings and then delete 0 1 patterns. By the previous observations, both result in the same final Young diagram.

At this point, it remains only to show that soliton lengths are given by the column lengths of the Young diagram $\Lambda(X_0)$. To see that this is so, observe that the path $\Gamma(X_\tau)$, which corresponds to the first stable configuration, consists of a series of single peaks of nondecreasing height, each as tall as the length of the associated soliton. As each flattening step reduces the height of the peaks by 1, we see that the number of rows of $\Lambda(X_\tau)$ having length at least ℓ corresponds to the number of solitons of length at least ℓ. Therefore, the columns of $\Lambda(X_\tau)$ encode the soliton lengths, so the same is true of $\Lambda(X_0)$ by invariance. □

A.2. **Extracting column lengths with excursion operators.** In this subsection, we prove Lemma 2.2. The key observation is that the hill-flattening and excursion operators commute on the space of Motzkin paths.

To begin, we need to establish a couple of technical results. First, for any interval $I \subseteq \mathbb{R}$ and function $f \in C_0(I)$, we denote by $\operatorname{supp}(f)$ the closure of the set $\{x \in I : f(x) > 0\}$, which is a finite disjoint union of closed intervals. Accordingly, we may write $\operatorname{supp}(f) = \bigsqcup_{i=1}^{n} [c_i, d_i]$, where $d_i < c_j$ if $i < j$. We call $J_i := [c_i, d_i]$ the i^{th} *excursion interval* of f.

Proposition A.2. *Fix a Motzkin path* Γ *and let* $x \in \mathbb{N}$ *be contained in a hill interval* I_x *of* Γ. *Denote* $\operatorname{supp}(\mathscr{E}_x(\Gamma)) = \bigsqcup_{i=1}^{n} J_i$ *as above. Then* $\Gamma - \mathscr{E}_x(\Gamma)$ *is constant on each* J_i. *In addition,* $\mathscr{I}(\mathscr{E}_x(\Gamma)) = \mathscr{I}(\Gamma) \setminus \{I_x\}$ *and* $\max \mathscr{E}^{j-1}(\Gamma) \geq 1$ *for all* $1 \leq j \leq \rho(\Gamma)$.

Proof. To establish the first part, write $M = \Gamma_x \geq 0$, and define integers $a_0 < a_1 < \cdots < a_{M-1} < a_M = x = b_M < b_{M-1} < \cdots < b_1 < b_0$ by

$$a_i = \max\{k \leq x : \Gamma_k = i\}, \quad b_i = \min\{k \geq x : \Gamma_k = i\}$$

for each $0 \leq i \leq M$. In words, they are the first locations where Γ has height i when moving to the left and right from x (see Figure 6). To simplify notation, we set $a_{-1} = 0$ and $b_{-1} = \infty$. Now $\Gamma_y - \mathscr{E}_x(\Gamma)_y = \min\left\{\Gamma_z : x \wedge y \leq z \leq x \vee y\right\}$, so on $\mathbb{N}_0$

$$\Gamma - \mathscr{E}_x(\Gamma) = \sum_{i=0}^{M-1} i\left(\mathbf{1}_{(a_{i-1}, a_i]} + \mathbf{1}_{[b_i, b_{i-1})}\right) + M\mathbf{1}_{(a_{M-1}, b_{M-1})}.$$

It follows that $\mathscr{E}_x(\Gamma)$ vanishes at the a_i's and b_i's, and differs from Γ by a constant on (a_{M-1}, b_{M-1}) and each interval of the form $(a_{i-1}, a_i]$ or $[b_i, b_{i-1})$, $0 \leq i \leq M-1$. J_i is the i^{th} such interval (from left to right) where $\mathscr{E}_x(\Gamma)$ is not constant. This shows the first part of the assertion.

The preceding argument also implies that $\mathscr{I}(\mathscr{E}_x(\Gamma)) \subseteq \mathscr{I}(\Gamma)$. In addition, $\mathscr{E}_x(\Gamma) = 0$ on $[a_{M-1}, b_{M-1}]$ and $I_x \subseteq (a_{M-1}, b_{M-1})$, so I_x is not a hill interval of $\mathscr{E}_x(\Gamma)$. Finally, the definition of the a and b terms ensures that if $J \in \mathscr{I}(\Gamma) \setminus \{I_x\}$, then either $J \subseteq (a_{i-1}, a_i]$ or $J \subseteq [b_i, b_{i-1})$ for some $0 \leq i \leq M-1$. Since $\mathscr{E}_x(\Gamma)$ is a vertical translate of Γ on these intervals, J must be a hill interval of $\mathscr{E}_x(\Gamma)$. This shows $\mathscr{I}(\mathscr{E}_x(\Gamma)) = \mathscr{I}(\Gamma)$.

Lastly, taking $x = \mathtt{m}$ in the first part gives $\mathscr{I}(\mathscr{E}(\Gamma)) = \mathscr{I}(\Gamma) \setminus \{I_{\mathtt{m}}\}$, and the second part of the second assertion follows from the first since each application of $\mathscr{E}$ removes a single hill interval and the height of a Motzkin path is at least one while hill intervals remain. □

Proposition A.3. *For any interval $I \subseteq \mathbb{R}^+$, $f \in C_0(I)$, $x, y \in I$, if f is constant on the interval $[x, y] \subseteq I$, then $\mathscr{E}_x(f) = \mathscr{E}_y(f)$.*

Proof. Casing out according to whether $t < x$, $x \le t \le y$, or $t > y$ shows that

$$\min_{t\wedge x\le s\le t\vee x} f(s) = \min_{t\wedge y\le s\le t\vee y} f(s)$$

□

Proposition A.4. *For any Motzkin path Γ and any $x \in \mathbb{N}$ contained in a hill interval of Γ, $\mathscr{E}_x \circ \mathscr{H}(\Gamma) = \mathscr{H} \circ \mathscr{E}_x(\Gamma)$. In particular, $\mathscr{E} \circ \mathscr{H}(\Gamma) = \mathscr{H} \circ \mathscr{E}(\Gamma)$.*

Proof. Let $\mathtt{m} = \mathtt{m}(\Gamma)$ and $\mathtt{m}^* = \mathtt{m}(\mathscr{H}(\Gamma))$. Note that $\mathtt{m} < \mathtt{m}^*$ and that $\mathscr{H}(\Gamma)$ is constant on $[\mathtt{m}, \mathtt{m}*]$. Thus by Proposition A.3 with $I = \mathbb{R}^+$, it suffices to prove the first part. To this end, we first note that for any $k \in \mathbb{N}_0$,

$$\min_{k\wedge x\le y\le k\vee x} \Gamma_y - \min_{k\wedge x\le y\le k\vee x} \mathscr{H}(\Gamma)_y = \mathbf{1}\{k \in I_x\}.$$

Indeed, $\mathscr{H}(\Gamma) = \Gamma - 1$ on I_x, so the left-hand side is 1 for all $k \in I_x$. Now fix $k \notin I_x$, and let x_* be the location of the leftmost minimum of Γ over the interval $[k \wedge x, k \vee x]$. Then x_* is an integer which is not contained in any hill interval of Γ, so $\mathscr{H}(\Gamma)_{x_*} = \Gamma_{x_*}$. Moreover, x_* minimizes $\mathscr{H}(\Gamma)$ on $[k \wedge x, k \vee x]$ since the only integer points with $\mathscr{H}(\Gamma)_y < \Gamma_y$ are those contained in a hill interval of Γ, in which case $\Gamma_y \ge \Gamma_{x_*} + 1$. This shows that the left-hand side is 0 for $k \notin I_x$ as desired.

In conjunction with Proposition A.2, we have

$$\begin{aligned}
\mathscr{E}_x(\mathscr{H}(\Gamma))_k &= \mathscr{H}(\Gamma)_k - \min_{k\wedge x\le y\le k\vee x} \mathscr{H}(\Gamma)_y \\
&= \mathscr{H}(\Gamma)_k - \min_{k\wedge x\le y\le k\vee x} \Gamma_y + \mathbf{1}\{k \in I_x\} \\
&= \begin{cases} \mathscr{E}_x(\Gamma)_k - 1 & \text{if } k \in \mathscr{I}(\Gamma) \setminus \{I_x\} \\ \mathscr{E}_x(\Gamma)_k & \text{otherwise} \end{cases} = \mathscr{H}(\mathscr{E}_x(\Gamma))_k.
\end{aligned}$$

□

Now we prove Lemma 2.2.

***Proof of Lemma** 2.2.* Let Γ be a Motzkin path and write λ_j for the length of the j^{th} column of $\Lambda(\Gamma)$ for each $1 \le j \le \rho(\Gamma)$. We show

$$\lambda_j = \max \mathscr{E}^{j-1}(\Gamma).$$

by induction on $\max \Gamma$. If the maximum is zero, then the assertion is trivial, so we may assume that it holds for all Motzkin paths with maximum less than $M \in \mathbb{N}$. Now fix a path Γ with $\max \Gamma = M$. The inductive hypothesis implies that the assertion holds for $\mathscr{H}(\Gamma)$ since it has maximum $M - 1 \ge 0$. Moreover, $\Lambda(\mathscr{H}(\Gamma))$ is obtained by deleting the first row of $\Lambda(\Gamma)$. Thus by Proposition A.4, we have

$$\begin{aligned}
\lambda_j - 1 &= \max \mathscr{E}^{j-1}(\mathscr{H}(\Gamma)) \\
&= \max \mathscr{H}(\mathscr{E}^{j-1}(\Gamma))
\end{aligned}$$

$$= \max \mathscr{E}^{j-1}(\Gamma) - 1,$$

where the final equality used the second part of Proposition A.2 to ensure $\max \mathscr{E}^{j-1}(\Gamma) \geq 1$ for any $1 \leq j \leq \rho(\Gamma)$. □

A.3. **Regularity of the column length functionals.** In this subsection we prove Lemma 2.3, establishing Lipschitz continuity of the 'column length functionals' $\max \mathscr{E}^{j-1}(\cdot)$. The general strategy is to show that the column length functionals satisfy a Lipschitz condition on Motzkin paths and then extend the result to arbitrary functions in $C_0(\mathbb{R}^+)$ by an approximation argument. We begin by establishing some preparatory results.

Proposition A.5.

(i) *Fix an interval $I \subseteq \mathbb{R}^+$, a point $b \in I$, and functions $f, g \in C_0(I)$. Then*

$$\|\mathscr{E}_b(f) - \mathscr{E}_b(g)\|_\infty \leq 2\|f - g\|_\infty.$$

(ii) *For any $f, g \in C_0(\mathbb{R}^+)$ whose graphs are Motzkin paths,*

$$\|\mathscr{H}(f) - \mathscr{H}(g)\|_\infty \leq \|f - g\|_\infty.$$

Proof. For (i), the triangle inequality gives

$$\|\mathscr{E}_b(f) - \mathscr{E}_b(g)\|_\infty \leq \|f - g\|_\infty + \left| \min_{[t\wedge b, t\vee b]} f - \min_{[t\wedge b, t\vee b]} g \right| \leq 2\|f - g\|_\infty$$

since the minima of two functions over a given interval can differ by no more than their maximum difference over the interval.

For (ii), observe that the maximum distance between Motzkin paths is necessarily $\mathbb{N}_0$-valued and the claim is clearly true if $f = g$, so we may assume that $\left\|\mathscr{H}(f) - \mathscr{H}(g)\right\|_\infty \geq 1$. Let

$$x^* = \max\left\{x \in \mathbb{N} : |\mathscr{H}(f)_x - \mathscr{H}(g)_x| = \|\mathscr{H}(f) - \mathscr{H}(g)\|_\infty\right\},$$

and assume without loss of generality that $\mathscr{H}(f)_{x^*} > \mathscr{H}(g)_{x^*}$. If x^* is not in a hill interval of g, then $g(x^*) = \mathscr{H}(g)_{x^*} < \mathscr{H}(f)_{x^*} \leq f(x^*)$, so

$$\left\|\mathscr{H}(f) - \mathscr{H}(g)\right\|_\infty = \left|\mathscr{H}(f)_{x^*} - \mathscr{H}(g)_{x^*}\right| \leq \left|f(x^*) - g(x^*)\right| \leq \left\|f - g\right\|_\infty.$$

If x^* is in a hill interval of both f and g, then

$$\left\|\mathscr{H}(f) - \mathscr{H}(g)\right\|_\infty = \left|\mathscr{H}(f)_{x^*} - \mathscr{H}(g)_{x^*}\right| = \left|\left(f(x^*) - 1\right) - \left(g(x^*) - 1\right)\right| \leq \left\|f - g\right\|_\infty.$$

Finally, suppose that x^* is in a hill interval $[a, b]$ of g but is not in any hill interval of f. Then g is constant on $[a, b]$, so our choice of x^* implies that $f(x^*) \geq f(y)$ for all $y \in [a, b]$. By considering whether or not $x^* < b$, we see that we must have $f(x^* + 1) = f(x^*) - 1$. A similar consideration of whether $f(x^*) = f(y)$ for all $a \leq y \leq x^*$ leads to the contradiction that x^* is in a hill interval of f. This proves the assertion. □

To state our next result, we say that a function $\varphi : \mathbb{R} \to \mathbb{R}$ is an *affine scaling* if $\varphi(x) = ax + b$ for some $a > 0$, $b \in \mathbb{R}$. The set of all affine scalings forms a group under composition. Given $f \in C_0(\mathbb{R})$ and an affine scaling φ, we write $\varphi^*(f)$ for the function $f \circ \varphi$. A function $\Gamma : \mathbb{Z} \to \mathbb{N}_0$ is an *extended Motzkin path* if $\Gamma(n) = 0$ for all $n \leq 0$ and $\Gamma|_{\mathbb{N}_0}$ is a Motzkin path.

Proposition A.6. *For any $f_1, f_2 \in C_0(\mathbb{R})$ which are not identically zero and any $\varepsilon > 0$, there exist affine scalings φ, ψ and extended Motzkin paths Γ_1, Γ_2 such that $\psi(0) = 0$ and for $i = 1,2$, the function $\bar{f}_i = \psi \circ \varphi^*(\Gamma_i) \in C_0(\mathbb{R})$ satisfies*

$$\|f_i - \bar{f}_i\|_\infty < \varepsilon \quad \textit{and} \quad \mathtt{m}(\bar{f}_i) = \mathtt{m}(f_i).$$

Proof. By hypothesis, $\mathtt{m}(f_1), \mathtt{m}(f_2) \in (0,\infty)$. Also, the f_i's are uniformly continuous, so there is some $\delta > 0$ such that $|x - y| < \delta$ implies $|f_1(x) - f_1(y)|, |f_2(x) - f_2(y)| < \varepsilon/4$. Set $\mathtt{s} = |\mathtt{m}(f_1) - \mathtt{m}(f_2)| + \mathbf{1}\{\mathtt{m}(f_1) = \mathtt{m}(f_2)\}$ and choose N large enough that $\Delta := \mathtt{s}/2^N < \delta$. Define the lattice

$$\mathfrak{L} = \{\mathtt{m}(f_1) + k\Delta\}_{k \in \mathbb{Z}}.$$

Note that $\mathtt{m}(f_1), \mathtt{m}(f_2) \in \mathfrak{L}$. Set $a = 2\Delta/\varepsilon$, $\mathfrak{L}^+ = \mathfrak{L} \cap [0,\infty)$, and let ℓ_0 denote the smallest element of $\mathfrak{L}^+$. Observe that $0 \le \ell_0 < \Delta$ by construction.

For $i = 1,2$, define the function $\gamma_i : \mathfrak{L} \to \mathfrak{L}^+$ by

$$\gamma_i(\ell) = \begin{cases} \ell_0 & \text{if } \ell \le \ell_0 \\ \Delta \lceil (a f_i(\ell))/\Delta \rceil + \ell_0 & \text{otherwise.} \end{cases}$$

Note that $a f_i$ changes by no more than $\Delta/2$ when the argument changes by no more than Δ. In conjunction with the fact that $f_i \equiv 0$ on $(-\infty, 0]$, $f_i \ge 0$, and $\ell_0 \in [0,\Delta)$, this implies that γ_i is an extended Motzkin path on $\mathfrak{L}$. That is, $\gamma_i(\ell) = \ell_0$ for all $\ell \in \mathfrak{L} \cap (-\infty, \ell_0]$ and for each $\ell, \ell' \in \mathfrak{L}$ with $|\ell - \ell'| = \Delta$, we have $\gamma_i(\ell) \ge \ell_0$ and $|\gamma_i(\ell) - \gamma_i(\ell')| \in \{0, \Delta\}$.

Let $\varphi(x) = \Delta \cdot x + \ell_0$. Then φ is an affine scaling which maps $\mathbb{Z}$ bijectively to $\mathfrak{L}$. Also define the affine scaling $\sigma(x) = (x - \ell_0)/a$. By a slight abuse of notation, we will henceforth let γ_i denote its extension to $\mathbb{R}$ by linear interpolation. Let $\Gamma_i \in C_0(\mathbb{R})$ be the extended Motzkin path defined by $\Gamma_i = \varphi^{-1} \circ \gamma_i \circ \varphi$. Now define

$$\bar{f}_i = (\gamma_i - \ell_0)/a = \sigma \circ \varphi \circ \Gamma_i \circ \varphi^{-1} = \psi \circ \varphi^*(\Gamma_i)$$

where $\psi = \sigma \circ \varphi$. Then $\psi(0) = \sigma(\ell_0) = 0$ and $\mathtt{m}(\bar{f}_i) = \mathtt{m}(\gamma_i) = \mathtt{m}(f_i)$. For $x \in \mathfrak{L}$, a direct computation gives $|f_i(x) - \bar{f}_i(x)| < \varepsilon/2$. For $x \notin \mathfrak{L}$, writing ℓ_x for the nearest lattice point to x gives

$$|f_i(x) - \bar{f}_i(x)| \le |f_i(x) - f_i(\ell_x)| + |f_i(\ell_x) - \bar{f}_i(\ell_x)| + |\bar{f}_i(\ell_x) - \bar{f}_i(x)| < \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\Delta}{2a} = \varepsilon,$$

hence $\|f_i - \bar{f}_i\|_\infty < \varepsilon$ as desired. □

We are now ready to prove Lemma 2.3.

Proof of Lemma 2.3. Fix $j \ge 1$. To begin, we observe that it is enough to show the assertion for $I = \mathbb{R}$. Indeed, for any $I \subseteq \mathbb{R}$ and any $h \in C_0(I)$, we can define a function $\tilde{h} \in C_0(\mathbb{R})$ which equals h on I and drops linearly to zero on $[b, b+1]$ where b is the right-most boundary point of I. This construction ensures that $\max \mathscr{E}^{j-1}(h) = \max \mathscr{E}^{j-1}(\tilde{h})$ and $\|h_1 - h_2\|_\infty = \|\tilde{h}_1 - \tilde{h}_2\|_\infty$.

Next we show that the result holds if the graphs of f and g are (extended) Motzkin paths by induction on $m = \max \mathscr{E}^{j-1}(f) + \max \mathscr{E}^{j-1}(g)$. The assertion is trivial when $j = 1$ or $m = 0$. If $\max \mathscr{E}^{j-1}(f) \ge 1$ and $\max \mathscr{E}^{j-1}(g) = 0$, write $\mathtt{m}_j := \mathtt{m}(\mathscr{E}^{j-1}(f))$. Let $J = [a, b]$

be the excursion interval of $\mathscr{E}^{j-1}(\Gamma)$ which contains $\mathtt{m}_j$. By Proposition A.2, $\Gamma - \mathscr{E}^{j-1}(\Gamma)$ is constant on J_i. Hence we get

$$\max \mathscr{E}^{j-1}(f) = \mathscr{E}^{j-1}(f)(\mathtt{m}_j) - \mathscr{E}^{j-1}(f)(a) = f(\mathtt{m}_j) - f(a).$$

This yields

$$\begin{aligned}\max \mathscr{E}^{j-1}(f) = f(\mathtt{m}_j) - f(a) &\le f(\mathtt{m}_j) - f(a) + \left|g(\mathtt{m}_j) - g(a)\right| \\ &\le \left|f(\mathtt{m}_j) - g(\mathtt{m}_j)\right| + \left|f(a) - g(a)\right| \le 2\|f-g\|_\infty,\end{aligned}$$

By symmetry, the result also holds when $m \ge 1$ and $\max \mathscr{E}^{j-1}(f) = 0$, so we may assume that $\max \mathscr{E}^{j-1}(f), \max \mathscr{E}^{j-1}(g) \ge 1$. As the maxima are necessarily attained on hill intervals, Proposition A.4, the inductive hypothesis, and part (ii) of Proposition A.5 imply

$$\begin{aligned}\left|\max \mathscr{E}^{j-1}(f) - \max \mathscr{E}^{j-1}(g)\right| &= \left|\max \mathscr{H} \circ \mathscr{E}^{j-1}(f) - \max \mathscr{H} \circ \mathscr{E}^{j-1}(g)\right| \\ &= \left|\max \mathscr{E}^{j-1} \circ \mathscr{H}(f) - \max \mathscr{E}^{j-1} \circ \mathscr{H}(g)\right| \\ &\le 2\left\|\mathscr{H}(f) - \mathscr{H}(g)\right\|_\infty \le 2\left\|f-g\right\|_\infty.\end{aligned}$$

This completes the proof for Motzkin paths.

Now we show the assertion for $f, g \in C_0(\mathbb{R})$ by induction on $j \ge 1$. The base case is tautological. For the inductive step, choose $\psi, \varphi, \Gamma_1, \Gamma_2, \bar{f}, \bar{g}$ as in Proposition A.6 with $f_1 = f, f_2 = g$. Then by the choice of $\bar{f}$, Proposition A.4, the induction hypothesis, and Proposition A.5 (i), we have

$$\begin{aligned}\left|\max \mathscr{E}^k(f) - \max \mathscr{E}^k(\bar{f})\right| &= \left|\max \mathscr{E}^{k-1} \circ \mathscr{E}_{\mathtt{m}(f)}(f) - \max \mathscr{E}^{k-1} \circ \mathscr{E}_{\mathtt{m}(f)}(\bar{f})\right| \\ &\le \|\mathscr{E}_{\mathtt{m}(f)}(f) - \mathscr{E}_{\mathtt{m}(f)}(\bar{f})\|_\infty \\ &\le 2\|f - \bar{f}\|_\infty < 2\varepsilon,\end{aligned}$$

and similarly for g. Also, since $\psi(0) = 0$, the triangle inequality gives

$$\begin{aligned}\psi(\|\Gamma_1 - \Gamma_2\|_\infty) &= \psi(\|\varphi^*\Gamma_1 - \varphi^*\Gamma_2\|_\infty) \\ &= \|\psi \circ \varphi^*\Gamma_1 - \psi \circ \varphi^*\Gamma_2\|_\infty \\ &= \|\bar{f} - \bar{g}\|_\infty < 4\varepsilon + \|f-g\|_\infty.\end{aligned}$$

Lastly, observe that the functional $\max \mathscr{E}^k$ satisfies

$$\max \mathscr{E}^k(\bar{f}_i) = \psi \circ \max \mathscr{E}^k(\Gamma_i).$$

Thus in conjunction with the assertion for the Motzkin paths, we obtain

$$\begin{aligned}\left|\max \mathscr{E}^k(f) - \max \mathscr{E}^k(g)\right| &< 4\varepsilon + \left|\max \mathscr{E}^k(\bar{f}) - \max \mathscr{E}^k(\bar{g})\right| \\ &\le 4\varepsilon + \psi\left(\left|\max \mathscr{E}^k(\Gamma_1) - \max \mathscr{E}^k(\Gamma_2)\right|\right) \\ &\le 4\varepsilon + \psi\left(\|\Gamma_1 - \Gamma_2\|_\infty\right) < 8\varepsilon + \|f-g\|_\infty.\end{aligned}$$

Letting $\varepsilon \searrow 0$ completes the inductive step and the proof. □

A.4. **Statistics of 312-avoiding permutations.** In this subsection, we provide proofs of Propositions 7.2 and 7.3.

Denote by $\mathfrak{S}_n^{231}$ the set of all 231-avoiding permutations of length n. Recall that a permutation σ is 312-avoiding iff its inverse σ^{-1} is 231-avoiding. In the proof of Proposition 7.2, we will make use of a bijection between Dyck paths of length $2n$ and 231-avoiding permutations of length n, which was used by Hoffman, Rizzolo, and Silvken in their recent work [9] studying random 231-avoiding permutations in terms of random walks and Brownian excursions.

For a given h-restricted Motzkin path Γ, we define a permutation $\sigma(\Gamma)$ as follows: Let v_k be the location of the k^{th} upstroke of Γ. (Thus if $\Gamma = \Gamma(X_0)$, then v_k is the location of the k^{th} ball in X_0). Then we define a 231-avoiding permutation $\sigma(\Gamma)$ by

$$\sigma(\Gamma)(k) = k + \frac{1}{2}\sup\left\{r \geq 0 : \Gamma_{v_k+r} \geq \Gamma_{v_k}\right\} + 1 - \Gamma_{v_k}. \tag{A.1}$$

When restricted to Dyck paths, this map $\Gamma \mapsto \sigma(\Gamma)$ is shown to be a bijection between Dyck_{2n} and $\mathfrak{S}_n^{231}$ in [9].

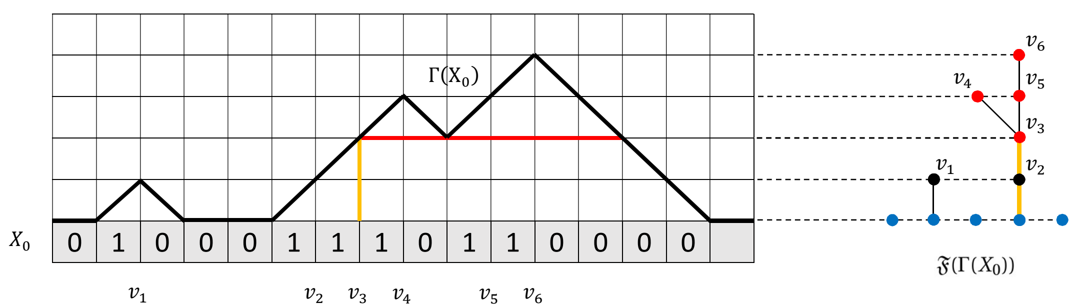

FIGURE A.3. Construction of the 231-avoiding permutation $\sigma^{-1} = 165243$ directly from the corresponding Motzkin path $\Gamma(X_0)$ and rooted forest $\mathfrak{F}(\Gamma(X_0))$. On the left, the lengths of the red and orange paths correspond to the supremum and Γ_{v_k} terms in (A.1) for $k = 3$. On the right, the subtree rooted at v_3 consists of the four red nodes and the level of v_3 is the number of edges in the orange path. Thus (A.1) and (A.2) each show that $\sigma^{-1}(3) = 5$.

Remark A.7. For a given rooted forest $\mathfrak{F}$, a permutation $\sigma(\mathfrak{F})$ can be defined similarly: Let v_k be the k^{th} non-root node in $\mathfrak{F}$ according to the depth-first order and define

$$\sigma(\mathfrak{F})(k) := k + \#\{\text{nodes in the subtree of } \mathfrak{F} \text{ rooted at } v_k\} - (\text{level of } v_k \text{ in } \mathfrak{F}). \tag{A.2}$$

This map naturally commutes with the map (A.1). Namely, let Γ be the h-restricted Motzkin path which is a contour process of $\mathfrak{F}$. Then

$$\sigma(\Gamma) = \sigma(\mathfrak{F}).$$

See Figure A.3 for an illustration.

Proof of Proposition 7.2. We define a map $\varphi : \text{Dyck}_{2n} \to \mathfrak{S}_n^{312}$ by the compositions of the following maps:

$$\Gamma \mapsto \sigma(\Gamma) \mapsto \sigma(\Gamma)^{-1},$$

where the first map is given by (A.1). Then being a composition of two bijections, φ is a bijection from Dyck_{2n} to $\mathfrak{S}_n^{312}$. This shows (i).

To show (ii), fix $\Gamma \in \text{Dyck}_{2n}$ and let X_0 be the box-ball configuration obtained from Γ by

$$X_0(i) = \mathbf{1}(\Gamma(i+1) - \Gamma(i) = 1)$$

for all $i \geq 0$. It then suffices to show that

$$\sigma(X_0) = \sigma(\Gamma)^{-1}. \tag{A.3}$$

To this end, label the balls $1, \dots, n$ from left to right, and recall the push-pop stack construction $X_0 \mapsto \sigma(X_0)$ described in Section 7. Fix a label $1 \leq k \leq n$. We are going to track the trajectory of ball k during the push-pop stack construction. Using the notation from equation (A.1), ball k is at site v_k. Note that Γ_{v_k} equals the number of balls in the stack after the ball k is pushed into the stack. Hence the number of balls which have been popped out in previous steps equals $k - \Gamma_{v_k}$. Next, while the stack sweeps sites to the right of v_k, balls with larger labels will be pushed in and popped out until ball k is finally deposited. This happens precisely when Γ first hits height $\Gamma_{v_k} - 1$ after location v_k. Accordingly, the number of balls that are deposited during the period when ball k is in the stack equals the height of the subexcursion of Γ started at v_k, which equals to half of the duration of this excursion. Thus

$$\#\text{ balls popped out before ball } k = k - \Gamma_{v_k} + \frac{1}{2}\sup\left\{r \geq 0 : \Gamma_{v_k + r} \geq \Gamma_{v_k}\right\}.$$

Therefore, $\sigma(\Gamma)(k)$, which is one more than the above quantity, is the position of k in $\sigma(X_0)$ as desired. □

Proof of Proposition 7.3. By induction on the length of the permutation, we suppose that the assertion holds for all 312-avoiding permutation of length less than n for some $n \geq 2$, and fix a 312-avoiding permutation τ of length n. Using Proposition 7.2, choose a box-ball configuration X_0 and a Dyck path Γ such that $\tau = \sigma(X_0)$ and $\Gamma = \Gamma(X_0)$.

By Greene's theorem ([8]), we know that the length of the first row of $\mathtt{RS}(\tau)$ equals the length of a longest increasing subsequence in τ. Since $\mathtt{RS}(\tau) = \Lambda(\Gamma)$, we see that the length of the longest increasing subsequence of τ equals the number of peaks in Γ.

Let X_0' be the box-ball configuration obtained from X_0 by deleting all 10 patterns from X_0, as in the proof of Lemma 2.1, and let $\Gamma' = \Gamma(X_0')$ and $\tau' = \sigma(X_0')$ be the h-restricted Motzkin path and 312-avoiding permutation constructed from X_0' (see the commutative diagram (A.4)). It is easy to see that Γ' can be directly obtained from Γ by first applying the hill-flattening operator $\mathscr{H}$ and then contracting new h-strokes which are not at height 0. On the other hand, observe that τ' can be obtained from τ by deleting a longest increasing subsequence.

$$\begin{array}{ccccc} \tau & \longleftarrow & X_0 & \longrightarrow & \Gamma \\ \downarrow & & \downarrow & & \downarrow \\ \tau' & \longleftarrow & X_0' & \longrightarrow & \Gamma' \end{array} \tag{A.4}$$

To see this (we refer the readers to Figure A.3), let L be the number of 10 patterns in X_0, which is the same as the number of peaks in Γ. By the observation at the end of previous paragraph, we know that L equals the length of a longest increasing subsequence in

τ. When reading X_0 from left to right, the i^{th} 10 pattern corresponds to the i^{th} deposite of ball with label $\tau(i)$; then the stack-construction ensures that $\tau(1), \tau(2), \cdots, \tau(L)$ is an increasing sequence. Hence this sequence is *a* longest increasing subsequence of τ.

To complete the argument, recall that $\Lambda(\Gamma')$ is obtained from $\Lambda(\Gamma)$ by deleting the first row. Since $\mathtt{RS}(\tau) = \Lambda(\Gamma)$ and $\mathtt{RS}(\tau') = \Lambda(\Gamma')$ by Proposition 7.1, we have that $\mathtt{RS}(\tau')$ is obtained from $\mathtt{RS}(\tau)$ by deleting its first row. Since τ' can be obtained from τ by deleting a longest increasing subsequence, the inductive hypothesis applied to τ' completes the proof. □

Acknowledgments

The first author thanks Karthik Karnik, Thomas Lam, Pavlo Pylyavskyy, and Mikaeel Yunus for inspiring conversations. The second author thanks Yuval Peres and Thomas Lam for helpful discussions.

References

[1] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[2] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao. *Normal approximation by Stein's method.* Probability and its Applications (New York). Springer, Heidelberg, 2011.

[3] David Roxbee Cox. *Renewal theory.* Methuen & Co. Ltd., London; John Wiley & Sons, Inc., New York, 1962.

[4] Endre Csáki and Yueyun Hu. Lengths and heights of random walk excursions. In *Discrete random walks (Paris, 2003)*, Discrete Math. Theor. Comput. Sci. Proc., AC, pages 45–52. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2003.

[5] Michael Drmota. Stochastic analysis of tree-like data structures. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 460(2041):271–307, 2004. Stochastic analysis with applications to mathematical finance.

[6] Devdatt P Dubhashi and Desh Ranjan. Balls and bins: A study in negative dependence. *BRICS Report Series*, 3(25), 1996.

[7] Rick Durrett. *Probability: theory and examples.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[8] Curtis Greene. An extension of schensted's theorem. In *Young Tableaux in Combinatorics, Invariant Theory, and Algebra*, pages 39–50. Elsevier, 1982.

[9] Christopher Hoffman, Douglas Rizzolo, and Erik Slivken. Pattern-avoiding permutations and Brownian excursion, part II: fixed points. *Probab. Theory Related Fields*, 169(1-2):377–424, 2017.

[10] Samuel Karlin and James McGregor. The classification of birth and death processes. *Trans. Amer. Math. Soc.*, 86:366–400, 1957.

[11] Donald E. Knuth. *The art of computer programming. Vol. 1.* Addison-Wesley, Reading, MA, third edition, 1997. Fundamental algorithms.

[12] Agnès Lagnoux, Sabine Mercier, and Pierre Vallois. Probability that the maximum of the reflected Brownian motion over a finite interval $[0, t]$ is achieved by its last zero before t. *Electron. Commun. Probab.*, 20:no. 62, 9, 2015.

[13] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, 2010. With an appendix by Oded Schramm and Wendelin Werner.

[14] Atsushi Nagai, Daisuke Takahashi, and Tetsuji Tokihiro. Soliton cellular automaton, Toda molecule equation and sorting algorithm. *Phys. Lett. A*, 255(4-6):265–271, 1999.
[15] Jim Pitman. Brownian motion, bridge, excursion, and meander characterized by sampling at independent uniform times. *Electron. J. Probab.*, 4:no. 11, 33, 1999.
[16] Jim Pitman and Matthias Winkel. Growth of the brownian forest. *Ann. of Probab.*, 33(6):2188–2211, 2005.
[17] Jim Pitman and Matthias Winkel. Regenerative tree growth: Markovian embedding of fragmenters, bifurcators, and bead splitting processes. *Ann. of Probab.*, 43(5):2611–2646, 2015.
[18] Jim Pitman and Marc Yor. On the distribution of ranked heights of excursions of a Brownian bridge. *Ann. Probab.*, 29(1):361–384, 2001.
[19] Doron Rotem. Stack sortable permutations. *Discrete Math.*, 33(2):185–196, 1981.
[20] Bruce E. Sagan. *The symmetric group*, volume 203 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2001. Representations, combinatorial algorithms, and symmetric functions.
[21] Daisuke Takahashi and Junkichi Satsuma. A soliton cellular automaton. *J. Phys. Soc. Japan*, 59(10):3514–3519, 1990.
[22] Makoto Torii, Daisuke Takahashi, and Junkichi Satsuma. Combinatorial representation of invariants of a soliton cellular automaton. *Phys. D*, 92(3-4):209–220, 1996.

LIONEL LEVINE, DEPARTMENT OF MATHEMATICS, CORNELL UNIVERSITY, ITHACA, NY 14853.
E-mail address: levine@math.cornell.edu

HANBAEK LYU, DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, COLUMBUS, OH 43210.
E-mail address: yu.1242@osu.edu

JOHN PIKE, DEPARTMENT OF MATHEMATICS, BRIDGEWATER STATE UNIVERSITY, BRIDGEWATER, MA 02324.
E-mail address: john.pike@bridgew.edu