Random Sorting Networks

Omer Angel Alexander E. Holroyd Dan Romik
Bálint Virág

September 16, 2006

Abstract

A sorting network is a shortest path from $12\cdots n$ to $n\cdots 21$ in the Cayley graph of S_n generated by nearest-neighbour swaps. We prove that for a uniform random sorting network, as $n\to\infty$ the space-time process of swaps converges to the product of semicircle law and Lebesgue measure. We conjecture that the trajectories of individual particles converge to random sine curves, while the permutation matrix at half-time converges to the projected surface measure of the 2-sphere. We prove that, in the limit, the trajectories are Hölder-1/2 continuous, while the support of the permutation matrix lies within a certain octagon. A key tool is a connection with random Young tableaux.

1 Introduction

Let $\mathcal{S}_n$ be the symmetric group of all permutations $\sigma=(\sigma(1),\ldots,\sigma(n))$ on $\{1,\ldots,n\}$, with composition given by $(\sigma\tau)(i):=\sigma(\tau(i))$. For $1\le s\le n-1$ denote the adjacent transposition or **swap** at location s by $\tau_s:=(\,s\;\;s{+}1\,)=(1,2,\ldots,s{+}1,s,\ldots,n)\in\mathcal{S}_n$. Denote the **identity** id $:=(1,2,\ldots,n)$ and the **reverse permutation** $\rho:=(n,\ldots,2,1)$. An n-element **sorting network** is a sequence $\omega=(s_1,\ldots,s_N)$ such that

$$\tau_{s_1}\tau_{s_2}\cdots\tau_{s_N}=\rho$$

Key words: sorting network, random sorting, reduced word, maximal chain in the weak Bruhat order, Young tableau, permutahedron.
2000 Mathematics Subject Classifications: 60C05; 05E10; 68P10
Funded in part by NSERC discovery grants (AEH and BV); NSF-FRG grant 0244479 (DR); a Connaught grant, a Canada Research Chair and a Sloan Fellowship (BV); a PIMS postdoctoral fellowship (OA); and by MSRI, PIMS and BIRS (all authors).

Figure 1: Selected particle trajectories for a uniformly chosen 2000-element sorting network.

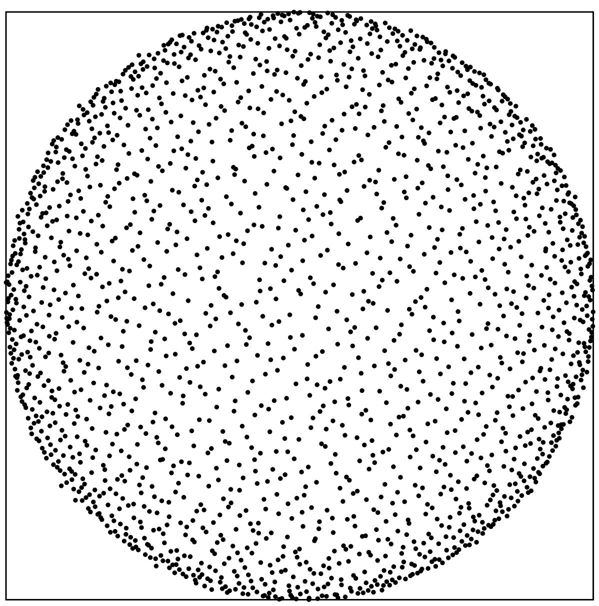

Figure 2: The permutation matrix of the half-time configuration $\sigma_{N/2}$ for a uniformly chosen 2000-element sorting network.

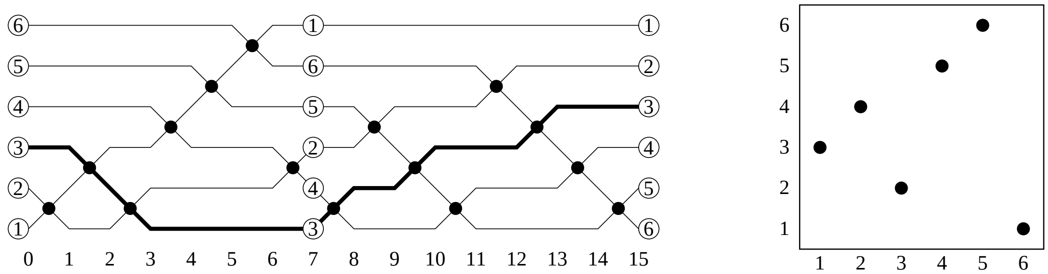

Figure 3: *Left:* the "wiring diagram" of the 6-element sorting network $\omega = (1, 2, 1, 3, 4, 5, 2, 1, 3, 2, 1, 4, 3, 2, 1)$. The swap process is shown by the black discs. The trajectory of particle 3 is highlighted. *Right:* the graph (or permutation matrix) of the configuration $\sigma_7 = (3, 4, 2, 5, 6, 1)$ of ω at time 7.

where

$$N := \binom{n}{2}.$$

(It is easily verified that N is the minimum possible length of a sequence of swaps whose composition is ρ, while ρ is the unique permutation for which this minimum length is maximized.) For $1 \le k \le N$ we refer to $s_k = s_k(\omega)$ as the kth **swap location**, and we call the permutation $\sigma_k = \sigma_k(\omega) := \tau_{s_1} \cdots \tau_{s_k}$ the **configuration at time** k. We call $\sigma_k^{-1}(i)$ the **location of particle** i at time k, and we call the function $k \mapsto \sigma_k^{-1}(i)$ the **trajectory** of particle i. See Figures 1, 2 and 3 for some illustrations.

Let Ω_n be the set of all n-element sorting networks, and let $\mathbb{P}_{\mathrm{U}} = \mathbb{P}_{\mathrm{U}}^n$ denote the uniform probability measure on Ω_n (assigning probability $1/\#\Omega_n$ to each $\omega \in \Omega_n$). We refer to a random sorting network chosen according to $\mathbb{P}_{\mathrm{U}}$ as a **uniform sorting network (USN)**.

Our first results concern the swap locations.

Theorem 1 (Stationarity and semicircle law)**.** *Let ω_n be a uniform n-element sorting network.*

(i) The random sequence $(s_1, \ldots, s_N)$ of swap locations is stationary; that is $(s_1, \ldots, s_{N-1})$ and $(s_2, \ldots, s_N)$ are equal in law under $\mathbb{P}_U$.

(ii) The first swap location s_1 satisfies the convergence in distribution

$$2s_1(\omega_n)/n - 1 \Longrightarrow Z \qquad \text{as} \quad n \to \infty$$

where Z is a random variable with semicircle law; that is with probability density function $\frac{2}{\pi}\sqrt{1-y^2}$ for $y \in (-1, 1)$.

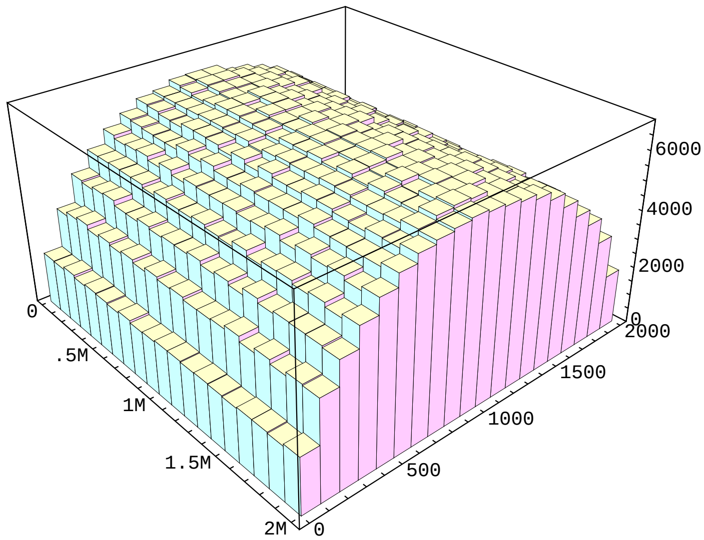

Figure 4: A histogram of the swap process for a uniformly chosen 2000-element sorting network. (The height of each column represents the number of swaps in the corresponding space-time window.)

In fact we can compute the exact distribution of s_1 for each n; see Proposition 9. In addition we establish the following "law of large numbers" for the swap locations. For an n-element sorting network ω, define the **scaled swap process** $\eta = \eta(\omega)$ to be the measure

$$\eta := \frac{1}{N} \sum_{k=1}^{N} \delta\Big(\frac{k}{N}, \ \frac{2s_k}{n} - 1\Big),$$

where $\delta(x, y)$ is the point measure at (x, y) on $\mathbb{R}^2$. Figure 4 is a histogram of η for a uniform 2000-element sorting network. Denote the semicircle measure by $\mathfrak{semi}(dy) := \frac{2}{\pi}\sqrt{1-y^2}\mathbf{1}_{y\in(-1,1)}\, dy$, and Lebesgue measure on $[0, 1]$ by $\mathfrak{Leb}(dx) := \mathbf{1}_{x\in[0,1]}\, dx$.

Theorem 2 (Law of large numbers)**.** *Let ω_n be a uniform n-element sorting network. The scaled swap process η satisfies*

$$\eta(\omega_n) \Longrightarrow \mathfrak{Leb} \times \mathfrak{semi} \qquad \text{as} \quad n \to \infty.$$

Here $\Longrightarrow$ denotes convergence in distribution of random measures in the vague topology on Borel measures on $\mathbb{R}^2$, and the right side denotes the deterministic product measure.

For a sorting network ω, define the **scaled trajectory** $T_i(t) = T_i(t,\omega)$ of particle i by

$$T_i(t) := 2\sigma_{tN}^{-1}(i)/n - 1$$

when tN is an integer, and by linear interpolation for other $t \in [0,1]$.

Theorem 3 (Hölder trajectories)**.** *Let ω_n be a uniform n-element sorting network.*

(i) For any $\varepsilon > 0$, the scaled trajectories satisfy

$$\mathbb{P}_U^n\Big(\forall i,s,t:\ \big|T_i(t) - T_i(s)\big| \le \sqrt{8}\big|t-s\big|^{1/2} + \varepsilon\Big) \to 1 \qquad \text{as} \quad n \to \infty.$$

(ii) Let $T_{i(n)}$ be the scaled trajectory of an arbitrarily chosen particle $i(n) = i(n,\omega_n)$ in ω_n. Then the random sequence $\{T_{i(n)}\}_{n=1}^{\infty}$ has subsequential limits in distribution with respect to uniform convergence of functions, and any subsequential limit is supported on Hölder$(\sqrt{8}, \frac{1}{2})$ *continuous paths.*

For a uniform sorting network, the particle configuration σ_k at a given time is a random permutation. We prove the following bounds on its distribution.

Theorem 4 (Octagon bounds)**.** *Let ω_n be a uniform n-element sorting network. For any $\varepsilon > 0$ we have*

$$\mathbb{P}_U^n\left(\forall k,i:\ \begin{array}{c} \big|\sigma_k(i) - i\big| < d_k + \varepsilon n, \quad \text{and} \\ \big|\sigma_k(i) - (n-i)\big| < d_{N-k} + \varepsilon n \end{array}\right) \to 1 \qquad \text{as} \quad n \to \infty,$$

where $d_k := n\sqrt{\frac{k}{N}\left(2 - \frac{k}{N}\right)}$.

Theorem 4 states that for each t, all the 1's in the permutation matrix of the configuration $\sigma_{\lfloor tN \rfloor}$ lie within a certain octagon asymptotically almost surely; see Figure 5.

Results of [5] and [12] give rise to an efficient algorithm for exactly sampling a uniform sorting network (specifically, see Theorems 8 and 13 in this article). The resulting simulations, together with heuristic arguments, have led us to striking conjectures about the asymptotic behaviour of the uniform sorting network.

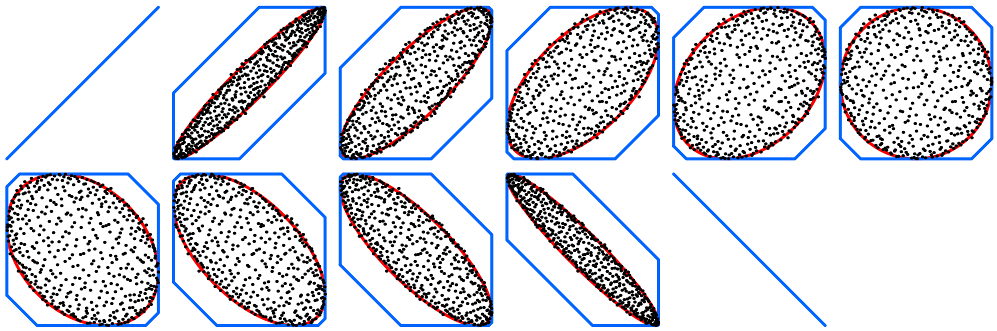

Figure 5: Graphs of the configurations at times $0, \frac{N}{10}, \frac{2N}{10}, \ldots, N$ for a uniformly chosen 500-element sorting network. Also shown are the asymptotic "octagon bounds" of Theorem 4, and the conjectural asymptotic "ellipse bounds" implied by Conjecture 2.

Figure 1 illustrates some trajectories for a uniform 2000-element sorting network. We conjecture that as $n \to \infty$, all particle trajectories converge to sine curves of random amplitudes and phases.

Conjecture 1 (Sine trajectories)**.** *Let ω_n be an n-element uniform sorting network and let T_i be the scaled trajectory of particle i. For each n there exist random variables $(A_i^n)_{i=1}^n, (\Theta_i^n)_{i=1}^n$ such that for all $\varepsilon > 0$,*

$$\mathbb{P}_U^n\Big(\max_{i\in[1,n]}\max_{t\in[0,1]}\big|T_i(t,\omega_n) - A_i^n \sin(\pi t + \Theta_i^n)\big| > \varepsilon\Big) \to 0 \qquad \textit{as} \quad n \to \infty.$$

Figures 2 and 5 illustrate the graphs $\{(i, \sigma_k(i)) : i \in [1, n]\}$ (i.e. the locations of 1's in the permutation matrix) of some configurations from uniform sorting networks. We conjecture that as $n \to \infty$ the graphs asymptotically concentrate in a family of ellipses, with a certain particle density in the interior of the ellipse. Define the **scaled configuration** $\mu_t = \mu_t(\omega)$ at time t by

$$\mu_t := \frac{1}{n}\sum_{i=1}^{n} \delta\Big(\frac{2i}{n} - 1\ ,\ \frac{2\sigma_{\lfloor tN\rfloor}(i)}{n} - 1\Big). \tag{1}$$

We define the **Archimedes measure** with parameter $t \in (0, 1)$ by

$$\mathfrak{Arch}_t(dx \times dy) := \frac{1}{2\pi}\sqrt{\Big[\sin^2(\pi t) + 2xy\cos(\pi t) - x^2 - y^2\Big]^{-1} \vee 0}\ \ dx\, dy.$$

Conjecture 2 (Archimedes configurations)**.** *Let ω_n be an n-element uniform sorting network. For all $t \in (0,1)$, the scaled configuration at time t satisfies*

$$\mu_t(\omega_n) \Longrightarrow \mathfrak{Arch}_t \qquad \text{as} \quad n \to \infty.$$

Here $\Longrightarrow$ denotes convergence in distribution in the vague topology for random Borel measures on $\mathbb{R}^2$.

In the case $t = 1/2$, the measure $\mathfrak{Arch}_{1/2}$ has density $1/\big(2\pi\sqrt{1-x^2-y^2}\big)$ on the circular disc $x^2 + y^2 < 1$. This is the unique circularly symmetric measure whose linear projections are uniform. It may be obtained by projecting surface area measure on the 2-sphere in $\mathbb{R}^3$ onto $\mathbb{R}^2$—that this gives a measure with the aforementioned property follows from the observation of Archimedes that the surface area of a sphere between two horizontal planes equals the corresponding area of a circumscribed vertical cylinder. (The claimed uniqueness follows from uniqueness of the characteristic function, [13, Theorem 5.3]). For general t the measure $\mathfrak{Arch}_t$ is obtained from $\mathfrak{Arch}_{1/2}$ by the linear transformation $(x, y) \mapsto (x, x\cos(\pi t) + y \sin(\pi t))$, and is supported on the interior of an ellipse—see Figure 5.

Conjectures 1 and 2 (and more) are implied by a very natural conjecture about the geometry of uniform sorting networks. The **permutahedron** is the natural embedding of the Cayley graph $(\mathcal{S}_n, (\tau_i)_{i=1}^{n-1})$ in Euclidean space in which we assign the permutation $\sigma \in \mathcal{S}_n$ to the point

$$\sigma^{-1} = (\sigma^{-1}(1), \ldots, \sigma^{-1}(n)) \in \mathbb{R}^n.$$

For all $\sigma \in \mathcal{S}_n$, clearly σ^{-1} lies on the $(n-2)$-sphere

$$\mathbb{S}_n := \Big\{z \in \mathbb{R}^n : \textstyle\sum_{i=1}^n z_i = \frac{n(n+1)}{2}\Big\} \cap \Big\{z \in \mathbb{R}^n : \textstyle\sum_{i=1}^n z_i^2 = \frac{n(n+1)(2n+1)}{6}\Big\},$$

while id^{-1} and ρ^{-1} are antipodal points on $\mathbb{S}_n$. Furthermore each edge of the Cayley graph has Euclidean length $\|\sigma^{-1} - (\sigma\tau_i)^{-1}\|_2 = \sqrt{2}$. See Figure 6 for an illustration of the case $n = 4$ (where $\mathbb{S}_4$ is a 2-sphere). A sorting network corresponds to a shortest path from id^{-1} to ρ^{-1} in the Cayley graph. It is natural to guess that such a path might typically be close to a **great circle** of $\mathbb{S}_n$; that is, a Euclidean circle in $\mathbb{R}^n$ having the same centre and radius as $\mathbb{S}_n$. We show that, if a sorting network lies close to some great circle, then its trajectories are approximately sine curves, its particle configurations approximate Archimedes measure, and its swap locations are approximately governed by the semicircle law.

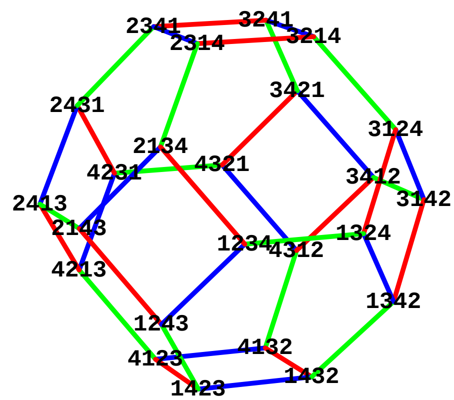

Figure 6: The permutahedron for $n = 4$.

Theorem 5 (Great circles)**.** *Suppose for each n that ω_n is a (non-random) n-element sorting network, and suppose that there is a sequence of great circles $c_n \subset \mathbb{S}_n$ such that*

$$d_\infty(\omega_n, c_n) = o(n) \quad \text{as} \quad n \to \infty,$$

(with distance defined as $d_\infty(\omega, c) := \max_{i\in[1,N]} \inf_{z\in c} \|\sigma_i^{-1} - z\|_\infty$). Then:

(i) there exist a_i^n, θ_i^n such that the scaled trajectories satisfy

$$\max_{i\in[1,N]} \max_{t\in[0,1]} \Big| T_i(t, \omega_n) - a_i^n \sin(\pi t + \theta_i^n) \Big| \to 0 \qquad \text{as} \quad n \to \infty;$$

(ii) for all $t \in (0, 1)$, the scaled configuration satisfies the vague convergence

$$\mu_t(\omega_n) \Longrightarrow \mathfrak{Arch}_t \qquad \text{as} \quad n \to \infty;$$

(iii) the scaled swap process satisfies the vague convergence

$$\eta(\omega_n) \Longrightarrow \mathfrak{Leb} \times \mathfrak{semi} \qquad \text{as} \quad n \to \infty.$$

We conjecture that, asymptotically almost surely as $n \to \infty$, the uniform sorting network does indeed lie close to a great circle on the permutahedron.

Conjecture 3 (Great circles). *Let ω_n be an n-element uniform sorting network. For each n there exists a random great circle $C_n \subset \mathbb{S}_n$ such that*

$$d_\infty(\omega_n, C_n) = o(n) \quad \textit{in probability as } n \to \infty.$$

Simulations provide overwhelming numerical evidence in support of Conjecture 3. Indeed, the evidence suggests that for the optimum great circle, typically $d_\infty(\omega_n, C_n) \approx \text{const} \times n^\alpha$, where $\alpha \approx 1/2$. For example, an exact simulation of a 10000-element uniform sorting network gave $d_\infty(\omega_{10000}, c) \leq 159$ for a certain great circle c. If Conjecture 3 holds then, by Theorem 5, Conjectures 1 and 2 follow, as well as the result in Theorem 2. The fact that Theorem 2 does indeed hold thus provides some further circumstantial evidence for Conjecture 3. It is interesting that the proofs of Theorem 2 and Theorem 5(iii) use entirely different methods.

In addition to Theorem 4, we note that certain other special permutations may be shown to have asymptotically much lower probability than others. Since the number of permutations is $n!$, at any given time step $k \in [0, N]$ there must exist some permutation which is visited with probability at least $1/n! \geq \exp[-n \log n]$. However, some permutations are much less likely, as illustrated by the following.

Example 6. *For n even, let $h = N/2 - n/4$, and consider the permutation $\psi := \left(\frac{n}{2}, \frac{n}{2} - 1, \ldots, 1,\ \ n, n-1, \ldots, \frac{n}{2} + 1\right)$. The probability that the uniform sorting network passes through particle configuration ψ equals*

$$\mathbb{P}^n_U(\sigma_h = \psi) = \exp\left[-\tfrac{\log 2}{4} n^2 + O(n)\right] \qquad \textit{as} \quad n \to \infty.$$

(This will be verified in Section 3.)

Remarks

History and connections. Sorting networks were first considered by Stanley [21], who proved the remarkable formula

$$\#\Omega_n = \frac{\binom{n}{2}!}{1^n 3^{n-1} 5^{n-2} \cdots (2n-3)^1}. \tag{2}$$

Another breakthrough was achieved by Edelman and Greene [5], who obtained a *bijective* proof of (2). (A related approach to the enumeration of sorting networks was independently developed by Lascoux and Schützenberger;

see [17], [10, p. 94–95].) The Edelman-Greene bijection is between the set Ω_n of sorting networks and the set of all staircase-shape standard Young tableaux of size n. This bijection will be an important ingredient for our results; we describe it in Section 4. See [7, 10, 11, 18, 20] for further background.

Sorting networks are of interest in computer science, since they can be interpreted as networks of comparators capable of sorting any sequence into descending order; see [16, Exer. 5.3.4.36–38]. There is also a connection with change-ringing (English-style church bell ringing); for background see [23] and the references therein.

About the proofs. The proof of Theorem 1(i) is very simple, and the proof of (ii) is straightforward given the results of [5]. Similar computations appear in [20]. Our proofs of Theorems 2, 3 and 4 are more involved, and depend on results from [19] on limiting profiles for random Young tableaux. A key tool is an extension of the result in [19] from square tableaux to staircase tableaux; see Section 5. Theorem 3 is a straightforward consequence of Theorem 4 together with Theorem 1(i). The proof of Theorem 5 employs geometric arguments, and relies on the characterization of $\mathfrak{Arch}_{1/2}$ as the unique measure all of whose linear projections are uniform on $[-1, 1]$.

Simulations. As remarked above, simulation evidence strongly supports Conjecture 3. The measurement $d_\infty(\omega_{10000}, c) \leq 159$ was obtained by using an exact simulation of a 10000-element USN ω_{10000}, and calculating the maximum L^∞ distance from the configuration σ_k^{-1} at time k to a point moving at constant angular speed around the great circle c that passes through σ_0^{-1} and $\sigma_{N/2}^{-1}$. In contrast, applying the same procedure to the "bubble sort" network $\omega = (1, 2, \ldots, n,\ 1, 2, \ldots, n-1,\ \ldots,\ 1, 2,\ 1)$ gives for $n = 10000$ a distance of approximately 9997. It is also easy to see that the condition $d(\omega_n, c_n) = o(n)$ does not hold for every sequence of sorting networks ω_n. For example, it does not hold for any sequence of sorting networks which pass through the permutation ψ in Example 6, since the configuration at time $\lfloor N/2 \rfloor \approx h$ cannot satisfy the condition in Theorem 5(ii).

A particularly striking illustration of Conjectures 1–3 results from plotting the graph of the permutation $\sigma_k^{-1}\sigma_{k+N/2}$, and then viewing the animation as k varies. Stationarity (Theorem 1(i)) implies that at any given time the picture will resemble Figure 2, while at time $k = N/2$ the initial picture will have been exactly rotated by $\pi/2$. In fact (for large n) the points appear to

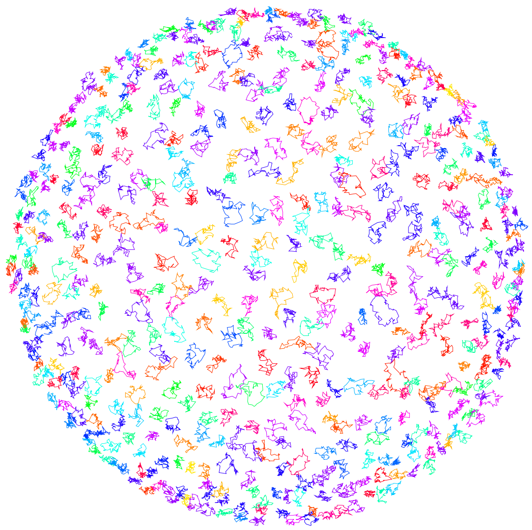

Figure 7: The evolution of the permutation graph of a sliding window, modulo uniform rotation, for a uniformly chosen 500-element sorting network.

rotate all at the same constant angular speed. To further illustrate this we may simultaneously rotate the entire picture by the (uniformly changing) angle $-\pi k/N$, and plot the resulting paths of the moving points as k increases from 0 to $N/2$. This is shown in Figure 7. The observation that each path is localized is a manifestation of Conjectures 1 and 3.

Further works. In forthcoming articles [1, 2, 3] we study several closely related issues. In [3] we prove further bounds on the configurations σ_k in the USN. In [2] we study the *local* structure of the swap process. In [1] we study another natural probability measure on sorting networks, in which at every step, a swap location is chosen uniformly from among those locations where the two particles are in increasing order. It turns out that this model can be analyzed in detail via the theory of exclusion processes. Its behaviour is very different from that of the USN, but it has the property, apparently shared by the USN (see Conjecture 1), that asymptotically each particle initially moves at a well-defined randomly chosen speed, and continues on a trajectory which is deterministic given this initial choice.

Stretchable sorting networks. The following is one way to generate a sorting network. Consider a set of n points in general position in $\mathbb{R}^2$, and label them $1, \ldots, n$ in order of increasing x-coordinate. Now rotate the set of points by an angle θ. For all but finitely many θ, listing the labels of the points in order of increasing x-coordinate gives a permutation in $\mathcal{S}_n$. And if we increase θ continuously from 0 to π, these permutations yield the sequence of configurations for a sorting network. Not all sorting networks can be obtained in this way; in fact those which can are exactly those whose wiring diagram may be drawn in the plane so that all the trajectories are straight lines; such networks are called *stretchable*—see [11] for details. (The smallest non-stretchable network, unique up to symmetries, is the 5-element example $\omega = (1, 3, 4, 2, 1, 3, 4, 2, 1, 3)$.) In the proof of Theorem 5 we will see that the assumption of that theorem implies that the sorting network is approximated by a stretchable network obtained by rotating a set of points in $\mathbb{R}^2$ which approximate the Archimedes measure $\mathfrak{Arch}_{1/2}$.

Consider an n-element USN, and choose m out of the n particles uniformly at random, independently of the USN. If we observe only the relative order of these m particles then we obtain a random m-element sorting network. If Conjecture 3 holds then it may be deduced that, as $n \to \infty$ with m fixed, the distribution of this sorting network converges to a measure whose support is exactly the set of stretchable m-element networks. This follows from the proofs in Section 8.

Gallery. For more simulation pictures, see the gallery `http://www.math.ubc.ca/~holroyd/sort`.

2 Preliminaries

In this section we present some definitions and basic results.

Proof of Theorem 1(i). If $\omega = (s_1, \ldots, s_N)$ is any sorting network then it is easily seen that

$$\omega' := (s_2, \ldots, s_N, n - s_1)$$

is also a sorting network, and furthermore that the map $\omega \mapsto \omega'$ is a bijection from Ω_n to Ω_n. The result follows immediately. $\square$

We note also that

$$(s_1, \ldots, s_N) \mapsto (s_N, \ldots, s_1) \tag{3}$$

and

$$(s_1, \ldots, s_N) \mapsto (n - s_1, \ldots, n - s_N) \tag{4}$$

are bijections from Ω_n to Ω_n, so the measure $\mathbb{P}^n_{\mathrm{U}}$ has the corresponding symmetries.

For a permutation $\sigma \in \mathcal{S}_n$, denote the **inversion number**

$$\mathrm{inv}(\sigma) = \#\Big\{(i,j) :\ 1 \le i < j \le n \text{ and } \sigma(i) > \sigma(j)\Big\}.$$

It is straightforward to see that $\mathrm{inv}(\sigma)$ is the graph-theoretic distance from the identity to σ in the Cayley graph of $\mathcal{S}_n$ generated by the swaps $\{\tau_1, \ldots, \tau_{n-1}\}$. Hence in any sorting network we have $\mathrm{inv}(\sigma_k) = k$ for all k.

3 Young tableaux

Young tableaux are a central tool in our proofs; we start by introducing some standard notation and facts. Let $N \in \mathbb{N} := \{1, 2, \ldots\}$. A **partition of** N is a sequence $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ of positive integers such that $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_k$ and $N = \sum_i \lambda_i$. We denote $|\lambda| := N$. We identify each partition λ with its associated **Young diagram**, which is the set $\{(i,j) \in \mathbb{N}^2 \ :\ 1 \le i \le k,\ 1 \le j \le \lambda_i\}$. Traditionally each element (i,j) (called a **cell**) in the diagram is drawn as a square, in the coordinate system with $(1,1)$ at the top-left and $(1,2)$ to its right. We denote the set of partitions of N by $\mathrm{Par}(N)$.

Two Young diagrams will play a central role: the $n \times n$ **square diagram** $(n, n, \ldots, n)$, which we denote by $\square_n$, and the **staircase diagram** $(n-1, n-2, \ldots, 1)$, which we denote by $\triangle_n$.

If $\lambda \in \mathrm{Par}(N)$, let $\lambda' = (\lambda'_1, \lambda'_2, \ldots, \lambda'_d)$ denote the **conjugate partition** to λ, where $d = \lambda_1$ and $\lambda'_i = \#\{1 \le j \le k \ :\ \lambda_j \ge i\}$. The conjugate partition corresponds to the Young diagram obtained by reflecting the Young diagram of λ along the northwest-southeast diagonal.

A **Young tableau** of shape λ, where $\lambda \in \mathrm{Par}(N)$, is an assignment of positive integers, called **entries**, to the cells of λ such that every row and column of the diagram contain increasing sequences of numbers. A **standard Young tableau (SYT)** is a Young tableau in which the numbers assigned to all the cells are $1, 2, \ldots, N$. See Figure 8. We denote the set of SYT of shape λ by $\mathrm{SYT}(\lambda)$, and we denote $d(\lambda) = \#\mathrm{SYT}(\lambda)$ (sometimes called the *dimension* of λ in representation-theoretic contexts), the number of standard Young tableaux of shape λ. Frame, Robinson and Thrall [9], [16, Sec. 1.5.4] proved the following formula for $d(\lambda)$.

1	2	3	4
2	4	5	6
13			

1	2	4	7
3	5	6	8
9			

Figure 8: The Young diagram $(4,4,1)$, a Young tableau and a standard Young tableau.

Theorem 7 (Hook formula; Frame, Robinson and Thrall)**.** *For each cell $(i,j) \in \lambda$ let $h_{i,j}(\lambda) := \lambda_i - j + \lambda'_j - i + 1$ be the* **hook number** *of (i,j) in λ. Then*

$$d(\lambda) = \frac{|\lambda|!}{\prod_{(i,j)\in\lambda} h_{i,j}(\lambda)}.$$

For two Young diagrams μ, λ write $\mu \nearrow \lambda$ ("μ increases to λ") to mean that λ can be obtained from μ by the addition of one cell. The **Young lattice** is the directed graph whose vertex set is $\cup_{N=0}^{\infty} \mathrm{Par}(N)$ and whose edges are all the pairs (μ, λ) with $\mu \nearrow \lambda$. Standard Young tableaux of shape λ are in bijection with paths in the Young lattice leading from the empty diagram $\emptyset$ to λ: to the path $\emptyset = \lambda_0 \nearrow \lambda_1 \nearrow \lambda_2 \nearrow \cdots \nearrow \lambda_N = \lambda$ we attach the SYT which records the order in which new cells were added to the diagrams along the path, i.e., the unique tableau $T = (t_{i,j})_{(i,j)\in\lambda}$ such that for all $0 \le k \le N$ we have that

$$\lambda_k = \{(i,j) \in \lambda : t_{i,j} \le k\}.$$

We call T the **recording tableau** of the increasing sequence of diagrams $(\lambda_k)_{0\le k\le N}$.

As an illustration of the use of the hook formula for sorting networks, we verify the claim of Example 6. We will use the fact (see [8, p. 135]) that

$$J(n) := \prod_{j=1}^{n} j^j = \exp\left[\frac{n^2+n}{2}\log n + O(n)\right] \qquad \text{as} \quad n \to \infty.$$

Thus, using the hook formula and Stirling's formula we can compute

$$\begin{aligned} d(\triangle_n) &= \frac{\binom{n}{2}!}{1^{n-1}3^{n-2}5^{n-3}\cdots(2n-3)^1} = \frac{\binom{n}{2}!\left(\frac{J(2n-2)}{2^{n(n-1)}J(n-1)^2}\right)^{1/2}}{\left(\frac{(2n-2)!}{2^{n-1}(n-1)!}\right)^{n-\frac{1}{2}}} \\ &= \exp\left[\frac{n^2-n}{2}\log n + (\tfrac{1}{4} - \log 2)n^2 + O(n)\right]; \end{aligned} \tag{5}$$

$$\begin{aligned} d(\square_n) &= \frac{(n^2)!}{J(n)(n+1)^{n-1}(n+2)^{n-2}\cdots(2n-1)^1} \\ &= \frac{(n^2)!\frac{J(2n-1)}{J(n)}}{J(n)\left[\frac{(2n-1)!}{n!}\right]^{2n}} = \exp\Big[n^2\log n + (\tfrac{1}{2}-\log 4)n^2 + O(n)\Big]. \end{aligned} \quad (6)$$

By (5) and (2) we have $\#\Omega_n = d(\triangle_n)$ (also see Section 4 below).

For a permutation $\nu \in S_n$, a **partial sorting network** (also called a reduced word) of ν is a sequence $(s_1, s_2, \ldots, s_k)$ such that $\nu = \tau_{s_1}\tau_{s_2}\cdots\tau_{s_k}$ and $k = \mathrm{inv}(\nu)$. Let $R(\nu)$ denote the number of partial sorting networks of ν. In general, evaluation of $R(\nu)$ is a deep problem—see e.g. [10, 18].

Let ν be any permutation, and let $k = \mathrm{inv}(\nu)$. Then $\omega = (s_1, s_2, \ldots, s_N)$ is a sorting network passing through configuration ν if and only if $(s_1, \ldots, s_k)$ is a partial sorting network for ν and $(s_{k+1}, \ldots, s_N)$ is a partial sorting network for $\nu^{-1}\rho$. Hence the probability that the USN passes through ν equals

$$\mathbb{P}_{\mathrm{U}}(\sigma_k = \nu) = \frac{R(\nu)R(\nu^{-1}\rho)}{R(\rho)}. \quad (7)$$

Proof of Example 6. In the case of ψ, we can compute the factors in (7) above explicitly. We have $\mathrm{inv}(\psi) = h$. Firstly, $R(\psi)$ is equal to $\binom{h}{h/2}d(\triangle_{n/2})^2$, since to get from id to ψ one must reverse the particles $1, \ldots, \frac{n}{2}$, and independently reverse the particles $\frac{n}{2}+1, \ldots, n$, with $\binom{h}{h/2}$ choices for the order in which to intersperse the left- and the right-half swaps.

Secondly, we claim that the number of partial sorting networks of $\psi^{-1}\rho = (\frac{n}{2}+1, \ldots, n,\ 1, \ldots, \frac{n}{2})$ is equal to $d(\square_{n/2})$. This is because, given such a partial sorting network $(s_1, s_2, \ldots, s_{N-h})$, we can construct a standard Young tableau of shape $\square_{n/2}$ whose ith row lists the times $k_1 < k_2 < \ldots < k_{n/2}$ at which particle i moved, and it is easy to see that this map is a bijection from the set of partial sorting networks of $\psi^{-1}\rho$ onto $\mathrm{SYT}(\square_{n/2})$. Thus we have:

$$\begin{aligned} R(\psi) &= \tbinom{h}{h/2}d(\triangle_{n/2})^2 = 2^{n^2/4+O(n)}d(\triangle_{n/2})^2; \\ R(\psi^{-1}\rho) &= d(\square_{n/2}); \\ R(\rho) &= d(\triangle_n). \end{aligned}$$

An application of the asymptotics (5) and (6) for the number of tableaux together with (7) verifies the claim of Example 6. Interestingly, the leading terms in $n^2\log n$ cancel in the exponent. ∎

$$\begin{array}{rrrr} \mathbf{1} & 2 & 3 & 9 \\ \mathbf{4} & 5 & 10 & \\ \mathbf{6} & \mathbf{11} & 12 & \\ 7 & \mathbf{13} & \mathbf{15} & \\ 8 & 14 & & \end{array} \xrightarrow{\text{slide}} \begin{array}{rrrr} \mathbf{0} & 2 & 3 & 9 \\ \mathbf{1} & 5 & 10 & \\ \mathbf{4} & \mathbf{6} & 12 & \\ 7 & \mathbf{11} & \mathbf{13} & \\ 8 & 14 & & \end{array} \xrightarrow{\text{increment}} \begin{array}{rrrr} 1 & 3 & 4 & 10 \\ 2 & 6 & 11 & \\ 5 & 7 & 13 & \\ 8 & 12 & 14 & \\ 9 & 15 & & \end{array}$$

Figure 9: The sliding sequence and the Schützenberger operator. Shown are: (a) A tableau T. In bold is the sliding sequence (obtained by starting from the maximum entry and repeatedly passing to the larger of the entries above and left); (b) the tableau obtained by sliding the entries down along the sliding sequence; (c) the tableau $\Phi(T)$.

4 The Edelman-Greene bijection

Stanley, who proved (2), noticed that by the hook formula the right-hand side of (2) is equal to $d(\triangle_n)$, the number of staircase shape standard Young tableaux of order n. Later, Edelman and Greene [5] found an explicit bijection between $\mathrm{SYT}(\triangle_n)$ and Ω_n. This bijection will play an important part in what follows, so we describe it and its inverse now.

Given a standard Young tableau $T \in \mathrm{SYT}(\lambda)$, where $N = |\lambda|$, denote by $(i_{\max}(T), j_{\max}(T))$ the coordinates of the cell containing the maximum entry N in T.

Define the **Schützenberger operator** $\Phi : \mathrm{SYT}(\lambda) \to \mathrm{SYT}(\lambda)$ as follows. Start with a tableau $T = (t_{i,j})_{(i,j)\in\lambda}$. Construct the **sliding sequence** of cells $c_0, c_1, \ldots, c_d \in \lambda$, where $c_0 = (i_{\max}(T), j_{\max}(T))$ and $c_d = (1,1)$, by the requirements that $c_r - c_{r+1} = (1,0)$ or $(0,1)$ for all $0 \le r \le d-1$, and $c_r - c_{r+1} = (1,0)$ if and only if $t_{c_r-(1,0)} > t_{c_r-(0,1)}$ (where we adopt the notational convention that for a cell (i,j) with either of i,j being non-positive we have $t_{i,j} = -\infty$). Then the tableau $\Phi(T) = (t'_{i,j})_{(i,j)}$ is defined by setting $t'_{c_r} = t_{c_{r+1}} + 1$ for $0 \le r \le d-1$, $t'_{1,1}$=1, and $t'_{i,j} = t_{i,j} + 1$ for all other cells $(i,j) \in \lambda$. The definition is illustrated in Figure 9. It is easy to see that Φ is a bijection of $\mathrm{SYT}(\lambda)$ onto itself.

Definition. *The Edelman-Greene bijection $EG : SYT(\triangle_n) \to \Omega_n$ is defined by*

$$EG(T) = \Big(j_{\max}(\Phi^{N-k}(T)) \Big)_{k=1,\ldots,N}$$

where as before $N = \binom{n}{2}$, and Φ^k denotes the kth iterate of Φ.

$$
\left.\begin{array}{l}
\begin{array}{l}1\end{array} \to \begin{array}{l}1\,2\end{array} \to \begin{array}{l}1\,2\\2\end{array} \to \begin{array}{l}1\,2\,3\\2\end{array} \to \begin{array}{l}1\,2\,3\,4\\2\end{array} \to \begin{array}{l}1\,2\,3\,4\,5\\2\end{array} \to \begin{array}{l}1\,2\,3\,4\,5\\2\,3\end{array} \to \begin{array}{l}1\,2\,3\,4\,5\\2\,3\\3\end{array} \\[2ex]
\to \begin{array}{l}1\,2\,3\,4\,5\\2\,3\,4\\3\end{array} \to \begin{array}{l}1\,2\,3\,4\,5\\2\,3\,4\\3\,4\end{array} \to \begin{array}{l}1\,2\,3\,4\,5\\2\,3\,4\\3\,4\\4\end{array} \to \begin{array}{l}1\,2\,3\,4\,5\\2\,3\,4\,5\\3\,4\\4\end{array} \to \begin{array}{l}1\,2\,3\,4\,5\\2\,3\,4\,5\\3\,4\,5\\4\end{array} \to \begin{array}{l}1\,2\,3\,4\,5\\2\,3\,4\,5\\3\,4\,5\\4\,5\end{array} \\[2ex]
\to \begin{array}{l}1\,2\,3\,4\,5\\2\,3\,4\,5\\3\,4\,5\\4\,5\\5\end{array}
\end{array}\right\} \xrightarrow[\text{tableau}]{\text{recording}} \begin{array}{llll}1 & 2 & 4 & 5\;6\\3 & 7 & 9 & 12\\8 & 10 & 13 & \\11 & 14 & & \\15 & & & \end{array}
$$

Figure 10: Computation of $\mathrm{EG}^{-1}(1,2,1,3,4,5,2,1,3,2,1,4,3,2,1)$.

It is far from obvious that the map EG is a bijection to Ω_n, nor what its inverse looks like. It turns out that the inverse may be described in terms of a Young tableau construction algorithm which is a modification of the RSK algorithm (see [22, Ch. 7.11]). Given a sorting network $\omega = (s_1, s_2, \ldots, s_N) \in \Omega_n$, we construct a sequence of (non-standard) Young tableaux $T_0, T_1, \ldots, T_N$ whose shapes $\emptyset = \lambda_0, \lambda_1, \ldots, \lambda_N = \triangle_N$ form an increasing sequence of diagrams, i.e., $\lambda_i \nearrow \lambda_{i+1}$. To get T_{i+1} from T_i, apply the following **insertion algorithm** to the input (T_i, s_{i+1}).

Insertion algorithm. *Given a Young tableau $T = (t_{i,j})_{(i,j)\in\lambda}$ of shape λ and a positive number u, construct a new tableau $T' = (t'_{i,j})$ whose shape is the union of λ with one new cell, as follows.*

Step 1. (Initialize). Set $k \leftarrow 1$ and $q \leftarrow u$. Set $t'_{i,j} \leftarrow t_{i,j}$ for all $(i,j) \in \mathbb{N}^2$, with the convention that $t_{i,j} = \infty$ for a cell $(i,j) \in \mathbb{N}^2 \setminus \lambda$.

Step 2. (Find next bumping cell). Set ℓ to be the least positive integer j such that $t_{k,j} \geq q$. Set $t'_{k,\ell} \leftarrow q$. If $q = t_{k\ell}$, set $q \leftarrow q+1$, otherwise set $q \leftarrow t_{k\ell}$. Set $k \leftarrow k+1$.

Step 3. If $q = \infty$, terminate and return the enlarged tableau T'. Otherwise return to Step 2.

Definition. *The inverse Edelman-Greene bijection $EG^{-1} : \Omega_n \to SYT(\triangle_n)$ is defined by setting $EG^{-1}(\omega)$ to be the recording tableau of the sequence of Young diagrams $\lambda_0 \nearrow \lambda_1 \nearrow \cdots \nearrow \lambda_N = \triangle_n$ constructed above.*

Figure 10 shows EG^{-1} applied to the sorting network of Figure 3. The following theorem justifies these definitions. The proof can be found in [5]; see also [7].

Theorem 8 (Edelman and Greene)**.** *The map EG is a bijection from $SYT(\triangle_n)$ to Ω_n, and the map EG^{-1} is its inverse.*

As a first application of the Edelman-Greene bijection, we prove an exact formula for the distribution of the first swap location $s_1 = s_1(\omega_n)$ of a uniform n-element sorting network ω_n, and use it to prove Theorem 1(ii).

Proposition 9 (Swap distribution)**.** *If ω_n is a uniform n-element sorting network, then*

$$\mathbb{P}_U^n(s_1 = r) = \frac{1}{N} \cdot \frac{\big(3 \cdot 5 \cdot 7 \cdots (2r-1)\big)\big(3 \cdot 5 \cdots (2(n-r)-1)\big)}{\big(2 \cdot 4 \cdot 6 \cdots (2r-2)\big)\big(2 \cdot 4 \cdots (2(n-r)-2)\big)}. \tag{8}$$

Proof. Let $1 \le r \le n-1$. By the definition of EG, the sorting networks $\omega = (s_1, s_2, \ldots, s_N) \in \Omega_n$ for which $s_1 = r$ are exactly the ones for which the standard Young tableau $\mathrm{EG}^{-1}(\omega)$ has its maximum entry in the cell $(n-r, r)$. Since EG is a bijection, the number of such ω's is the number of SYTs of shape $\triangle_n \setminus \{(n-r, r)\}$. Thus

$$\mathbb{P}_{\mathrm{U}}^n(s_1 = r) = \frac{d(\triangle_n \setminus \{(n-r,r)\})}{d(\triangle_n)}.$$

Write this using the hook formula, Theorem 7, to yield the result. □

Proof of Theorem 1(ii). Denote $a_{n,r} = \mathbb{P}_{\mathrm{U}}^n(s_1 = r), 1 \le r \le n-1$. Observe that by Proposition 9 we have

$$\begin{aligned} a_{n,r} &= \frac{2}{n(n-1)} \cdot \frac{(2r)(2r)!}{2^{2r}(r!)^2} \cdot \frac{2(n-r)(2(n-r))!}{2^{2(n-r)}((n-r)!)^2} \\ &= \frac{8\sqrt{r(n-r)}}{\pi n(n-1)} \cdot \frac{\sqrt{\pi r}\binom{2r}{r}}{2^{2r}} \cdot \frac{\sqrt{\pi(n-r)}\binom{2(n-r)}{n-r}}{2^{2(n-r)}} \end{aligned}$$

Therefore, using Stirling's formula in its explicit form $1 \le m!(2\pi m)^{-1/2}\left(\frac{e}{m}\right)^m \le 1 + \frac{1}{12m-1}$ (an immediate consequence of [6, Eq (9.15), p. 54]), we get that

$$\begin{aligned} \left(1 - \frac{1}{6r}\right)\left(1 - \frac{1}{6(n-r)}\right) &\le \frac{a_{n,r}}{\frac{8}{\pi n(n-1)}\sqrt{r(n-r)}} \\ &\le \left(1 + \frac{1}{24r-1}\right)\left(1 + \frac{1}{24(n-r)-1}\right). \end{aligned}$$

This implies easily that for all $-1 < a < b < 1$ we have

$$\mathbb{P}^n_{\mathrm{U}}\left(a \leq \frac{2s_1}{n} - 1 \leq b\right) = \sum_{\frac{n}{2}(a+1) \leq r \leq \frac{n}{2}(b+1)} a_{n,r} \xrightarrow[n\to\infty]{} \frac{2}{\pi}\int_a^b \sqrt{1-t^2}dt,$$

as required. $\square$

5 Limit profile for staircase tableaux

For any Young diagram λ we write $\mathbb{P}_\lambda$ for the uniform measure on the set $SYT(\lambda)$ of standard Young tableaux. It is natural to consider the limiting behaviour of a random tableau of distribution $\mathbb{P}_{\lambda_n}$ for a sequence of diagrams (λ_n) of a given shape and increasing size. For general shape, the problem of rigorously determining the complete limiting profile is open (see, however [14, 15] and [4, Theorem 1.5.1]). An exception is the square diagram $\square_n$, where the problem was solved by Pittel and Romik [19]. In this section we use their result to derive a solution for the staircase diagram $\triangle_n$.

We start by stating the main result from [19]. It will be convenient to use the following coordinate system. If (i,j) is a cell of $\square_n$, then its rotated (and scaled) coordinates are

$$u = u(i,j) := \frac{i-j}{n}; \qquad v = v(i,j) := \frac{i+j}{n},$$

(note that this differs from the coordinate system in [19] by a factor of $\sqrt{2}$).

We define the following functions, which will describe the limiting profile. For $\alpha \in [0,2]$ the function $h_\alpha : [-\sqrt{\alpha(2-\alpha)}, \sqrt{\alpha(2-\alpha)}] \to [0,1]$ is defined by

$$h_\alpha(u) := \frac{2}{\pi}\left[u \arctan\left(\frac{u}{R}\right) + \arctan^{-1} R\right] \quad \text{where} \quad R = \frac{\sqrt{\alpha(2-\alpha) - u^2}}{1-\alpha}, \tag{9}$$

for $\alpha \in [0,1]$ (where $\tan^{-1}\infty := \pi/2$ giving $h_1 \equiv 1$), and by

$$h_{2-\alpha}(u) := 2 - h_\alpha(u)$$

for $\alpha \in (1,2]$. The curve $v = h_\alpha(u)$ will approximate the level-$(\alpha n^2/2)$ contour of the tableau; Figure 11 shows some of these curves. The function $L : [0,1] \times [0,1] \to [0,2]$ is defined implicitly by

$$L\left(\frac{u+v}{2}, \frac{v-u}{2}\right) = \alpha \iff h_\alpha(u) = v.$$

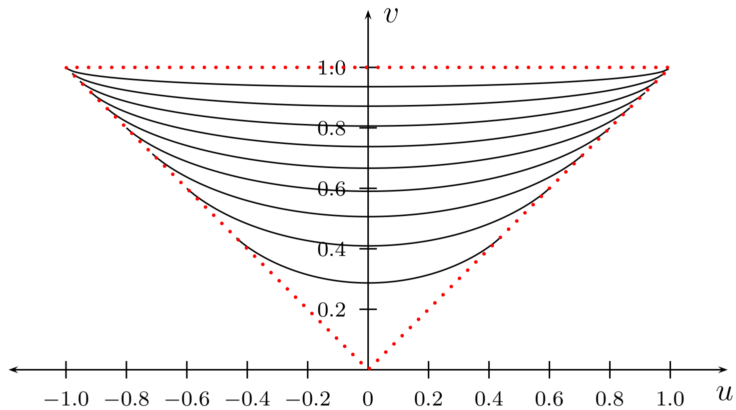

Figure 11: The curves $v = h_\alpha(u)$ for $\alpha = 0.1, 0.2, 0.3, \ldots, 0.9$, bounded between the graphs of $v = |u|$ and $v = 1$.

The following result [19, Theorem 1(i)]) gives the limiting profile for uniform square Young tableaux.

Theorem 10 (Limit profile for square tableaux; Pittel and Romik)**.** *Let $\mathbb{P}_{\square_n}$ be the uniform measure on Young tableaux $(s_{i,j})_{(i,j)\in\square_n} \in SYT(\square_n)$. For any $\varepsilon > 0$,*

$$\mathbb{P}_{\square_n}\left(\max_{(i,j)\in\square_n}\left|\frac{2s_{i,j}}{n^2} - L\left(\frac{i}{n}, \frac{j}{n}\right)\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0.$$

We shall deduce the following analogous result for the limit profile of a staircase tableau, where the function L is the same as above.

Theorem 11 (Limit profile for staircase tableaux)**.** *Let $\mathbb{P}_{\triangle_n}$ be the uniform measure on Young tableaux $(t_{i,j})_{(i,j)\in\triangle_n} \in SYT(\triangle_n)$. For any $\varepsilon > 0$,*

$$\mathbb{P}_{\triangle_n}\left(\max_{(i,j)\in\triangle_n}\left|\frac{2t_{i,j}}{n^2} - L\left(\frac{i}{n}, \frac{j}{n}\right)\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0.$$

Thus the limit profile for the staircase tableau is the same as that for half of the square tableau. Other results in [19] give explicit bounds on deviations from the limit profile, but only in the interior of the square. These estimates may be translated to staircase tableaux as well. However, uniform convergence in probability is sufficient for our purposes. It is important that Theorem 11 includes the boundary of the diagram.

Our main tool in proving the above is the following general result concerning continuity of random tableaux in the shape. For Young diagrams λ, μ we write $\lambda \subseteq \mu$ if this relation holds for λ, μ as subsets of $\mathbb{N}^2$.

Theorem 12 (Coupling)**.** *Let $\lambda \subseteq \mu$ be a pair of Young diagrams. There exists a coupling of the measures $\mathbb{P}_\mu$ on $S = (s_{i,j}) \in SYT(\mu)$ and $\mathbb{P}_\lambda$ on $T = (t_{i,j}) \in SYT(\lambda)$ such that for all $(i,j) \in \lambda$*

$$s_{i,j} \leq t_{i,j} + |\mu \setminus \lambda|.$$

To prove Theorem 12 we will make use of an algorithm from [12] for sampling from $\mathbb{P}_\lambda$. First note that, in order to simulate a tableau with distribution $\mathbb{P}_\lambda$, it suffices to be able to choose the location $c_{\max} = (i_{\max}, j_{\max})$ of the maximum entry $|\lambda|$ with the correct distribution. For then, after inserting this entry, we may iteratively apply the same algorithm to the smaller diagram $\lambda \setminus \{c_{\max}\}$ to locate the second largest entry, and so on.

For a cell $(i,j) \in \lambda$, define its **hook** to be the set

$$H_{(i,j)}(\lambda) := \big\{(k,j) \in \lambda : k \geq i\big\} \cup \big\{(i,k) \in \lambda : k \geq j\big\}.$$

The location $c_{\max}$ of the maximum entry may be simulated using the following **hook walk** algorithm from [12].

Hook walk algorithm. *Given a Young diagram λ, choose a random sequence of cells $c_0, \ldots, c_r$ iteratively as follows.*

Step 1. Choose a cell c_0 uniformly at random from λ.

Step 2. Given that cells $c_0, \ldots, c_{k-1}$ have been chosen, choose c_k uniformly at random from the hook $H_{c_{k-1}}(\lambda)$.

Step 3. Repeat Step 2 until we obtain a cell c_r with $\#H_{c_r}(\lambda) = 1$, then stop.

Theorem 13 (Hook walk; Greene, Nijenhuis and Wilf)**.** *The random final cell c_r constructed by the hook walk has the same distribution as $c_{\max}$ under $\mathbb{P}_\lambda$.*

Lemma 14 (Domination)**.** *Assume $\lambda \subseteq \mu$, and let $\lambda' := \lambda \setminus \{c_{\max}(\lambda)\}$ and $\mu' := \mu \setminus \{c_{\max}(\mu)\}$ be the random Young diagrams obtained by removing the largest entry in the respective uniform standard Young tableaux. Then we have the stochastic domination $\lambda' \subseteq_{\mathrm{st}} \mu'$.*

Proof. Consider the hook walk applied to λ and μ. It is enough to couple the two hook walks so that either they stop at the same cell, or the walk in μ stops at a cell in $\mu \setminus \lambda$. This will hold provided the two walks coincide until the first time the one in μ enters $\mu \setminus \lambda$. And this can be achieved as follows. Run the hook walk in μ according to the usual rules. Let the walk in λ be identical to that in μ while the latter is in λ. If and when the walk in μ jumps to a cell in $\mu \setminus \lambda$, continue the walk in λ according to the usual rules for λ using an independent source of randomness. It is easy to see that this gives the correct hook walk terminating probabilities for $c_{\max}(\lambda)$. (A key observation is that a uniformly chosen element in the hook $H_{(i,j)}(\mu)$ conditioned to be in $H_{(i,j)}(\lambda)$ is distributed uniformly in $H_{(i,j)}(\lambda)$). $\square$

Proof of Theorem 12. Construct the random tableaux S, T iteratively by first choosing the maximum entry in each, then the second largest, and so on. Do this using the hook walks, and at each stage couple the two hook walks according to Lemma 14, so that the remaining unfilled Young diagrams are always ordered. Let $m = |\mu \setminus \lambda|$. At the step when k is entered into λ, say at location (i, j), the entry $k + m$ is entered into μ, and all subsequent entries entered into μ are $\leq k+m$. By the ordering property one of those subsequent entries (possibly $k + m$) will be at the cell (i, j). Therefore $s_{i,j} \leq k + m = t_{i,j} + m$. $\square$

Proof of Theorem 11. Fix $\varepsilon > 0$, and consider a random square tableau $S = (s_{i,j})$ with law $\mathbb{P}_{\square_n}$. Let μ be the random Young diagram obtained by removing the cells with entries greater than $(1/2 + \varepsilon)n^2$ from $\square_n$,and define the event $A_n = \{\triangle_n \subseteq \mu\}$. From Theorem 10 we find that $\mathbb{P}(A_n) \to 1$, because $L \equiv 1$ along the diagonal.

Note that, conditional on μ, the tableau obtained by restricting S to μ has law $\mathbb{P}_\mu$. Also let $T = (t_{i,j})$ have law $\mathbb{P}_{\triangle_n}$. Theorem 12 implies that S and T can be coupled so that on the event A_n we have $s_{i,j} \leq t_{i,j} + n^2\varepsilon$ for $(i, j) \in \triangle_n$. Thus by the square diagram result, Theorem 10:

$$\mathbb{P}_{\triangle_n}\left(\max_{(i,j)\in\triangle_n}\left[-\frac{2t_{i,j}}{n^2} + L\left(\frac{i}{n}, \frac{j}{n}\right)\right] > 3\varepsilon\right) \leq$$
$$\mathbb{P}_{\square_n}\left(\max_{(i,j)\in\square_n}\left[-\frac{2s_{i,j}}{n^2} + L\left(\frac{i}{n}, \frac{j}{n}\right)\right] > \varepsilon\right) + \mathbb{P}(A_n^c) \xrightarrow[n\to\infty]{} 0,$$

as required. The corresponding upper bound is similar: instead we begin by removing the entries greater than $(1/2 - \varepsilon)n^2$. $\square$

We extract the following consequences of Theorem 11 for use in the later proofs.

Corollary 15 (First row)**.** *For a staircase tableau* $T = (t_{i,j})$ *let* $R_k = R_k(T) := \max\{j : t_{1,j} \le k\}$ *be the number entries* $\le k$ *in the first row. For any* $\varepsilon > 0$ *we have*

$$\mathbb{P}_{\triangle_n}\big(\max_k |R_k - d_k| > \varepsilon n\big) \xrightarrow[n\to\infty]{} 0$$

where $d_k := n\sqrt{\frac{k}{N}\left(2 - \frac{k}{N}\right)}$.

Proof. Theorem 11 easily implies that for any δ, with probability tending to 1 as $n \to \infty$,

$$\max_j \left| \frac{t_{1,j}}{N} - L\left(0, \frac{j}{n}\right) \right| < \delta. \tag{10}$$

From the definition of L, the map $x \mapsto L(0,x)$ is continuous and strictly monotone, and satisfies $L(0, \sqrt{\alpha(2-\alpha)}) = \alpha$ for $\alpha \in [0,1]$. Therefore given any $\epsilon > 0$, we can choose $\delta > 0$ such that for all $x, \alpha \in [0,1]$,

$$|L(0,x) - \alpha| < \delta \text{ implies } \left|x - \sqrt{\alpha(2-\alpha)}\right| < \varepsilon.$$

We deduce that on the event (10) we have

$$\max_j |j - d_{t_{1,j}}| < \varepsilon n.$$

Since $k \mapsto d_k$ is strictly monotone this implies that $\max_k |R_k - d_k| < \varepsilon n$. $\square$

Corollary 16 (Contours)**.** *Fix some* $\alpha \in [0,1]$*, and let* $\mathcal{H}_\alpha = \mathcal{H}_\alpha(T)$ *be the set of entries in a staircase tableau* T *in the cells where* $v > h_\alpha(u)$*. For any* $\varepsilon > 0$ *we have the following bound on the symmetric difference:*

$$\mathbb{P}_{\triangle_n}\Big[\#\big(\mathcal{H}_\alpha \,\Delta\, \{\lceil \alpha N\rceil, \dots, N\}\big) > \varepsilon N\Big] \xrightarrow[n\to\infty]{} 0.$$

Proof. Fix $\varepsilon > 0$. By the continuity and strict monotonicity of the function L, we may choose $\delta > 0$ such that the area of the region $D := \{(u,v) : h_{\alpha-2\delta}(u) \le v \le h_{\alpha+2\delta}(u)\}$ is at most ε. Then for n sufficiently large, on the event

$$\max_{(i,j)\in\triangle_n} \left| \frac{2t_{i,j}}{n^2} - L\left(\frac{i}{n}, \frac{j}{n}\right) \right| \le \delta,$$

we have that all entries in the symmetric difference $\mathcal{H}_\alpha \,\Delta\, \{\lceil \alpha N\rceil, \dots, N\}$ lie in D, so the result follows from Theorem 11. $\square$

6 Law of large numbers

This section contains the proof of Theorem 2. Recall the semicircle measure $\mathfrak{semi}(dx) = \frac{2}{\pi}\sqrt{1-x^2}\ \mathbf{1}_{x\in(-1,1)}\,dx$. Fix some interval $[a,b] \subset (-1,1)$, and define for $0 \le s \le t \le 1$ and a sorting network ω:

$$S_{s,t}(\omega) := \#\Big\{sN \le k < tN \ :\ \tfrac{2}{n}s_k(\omega) - 1 \in [a,b]\Big\}.$$

We will deduce Theorem 2 from the following.

Lemma 17. *Fix an interval $[a,b] \subset (-1,1)$. For any δ small enough (depending on a,b),*

$$\mathbb{P}_U^n\Big(\big|S_{0,\delta} - \delta N \mathfrak{semi}[a,b]\big| > 8N\delta^2\Big) \xrightarrow[n\to\infty]{} 0.$$

Proof of Theorem 2. It suffices to prove that for any $[a,b] \subseteq [-1,1]$ and for any $\varepsilon > 0$ and $0 \le s < t \le 1$ we have

$$\mathbb{P}_{\mathrm{U}}^n\left[\left|\tfrac{1}{N}S_{s,t} - (t-s)\mathfrak{semi}[a,b]\right| > \varepsilon\right] \xrightarrow[n\to\infty]{} 0. \tag{11}$$

Since the total number of swaps is deterministically N, it is enough to prove this in the case $[a,b] \subset (-1,1)$. We deduce this from Lemma 17 as follows. Fix some positive integer m to be chosen later, split the time interval $[s,t)$ into m smaller intervals of length $\delta := \frac{t-s}{m}$, and define the events

$$B_k := \left\{\left|S_{s+k\delta,s+(k+1)\delta}(\omega) - \delta N \mathfrak{semi}[a,b]\right| > 9N\delta^2\right\}, \qquad 0 \le k < m.$$

Let $B = \bigcup B_k$. By stationarity of the swap location process (Theorem 1(i)), each of the random variables $S_{s+k\delta,s+(k+1)\delta}$ is within ± 1 of a random variable having the same law as $S_{0,\delta}$. Hence by Lemma 17,

$$\mathbb{P}_{\mathrm{U}}^n(B) \le \sum_{k=0}^{m-1} \mathbb{P}_{\mathrm{U}}^n(B_k) \xrightarrow[n\to\infty]{} 0.$$

If B does not occur then the quantity

$$S_{s,t} = \sum_{k=0}^{m-1} S_{s+k\delta,s+(k+1)\delta}$$

satisfies

$$\left|S_{s,t} - m\delta N \mathfrak{semi}[a,b]\right| \le 9mN\delta^2;$$

i.e., since $m\delta = (t-s)$

$$\left|\tfrac{1}{N}S_{s,t} - (t-s)\mathfrak{semi}[a,b]\right| \le 9(t-s)\delta \le 9\delta.$$

Now (11) follows by setting m large enough that $9\delta < \varepsilon$. $\square$

Proof of Lemma 17. By stationarity of the swap process, the first δN swaps and the last δN swaps have the same law. The idea of the proof is now as follows. From the Edelman-Greene bijection we see that the last δN swaps are determined by the locations of the δN largest entries in the staircase shaped Young tableau T corresponding to ω. By Corollary 16, the set of these locations is almost deterministic, which will imply our claim.

For a Young tableau T, consider $j_{\max}(\Phi^k T)$. To find it we start with the element $N-k$ in T, and perform k iterations of Φ. At each iteration, the entry increases by 1, and possibly moves one square towards the diagonal. If it started close to the diagonal, it can only hit the diagonal in a limited region. In particular, if $N-k$ started in region A of Figure 12 then necessarily $\left(\frac{2}{n}j_{\max}(\Phi^k T) - 1\right) \in [a,b]$. Similarly, if it started in either of the regions labelled C then it will not exit through that interval. If $N-k$ started in the region marked B, then whether or not it exits in the interval $[a,b]$ depends on locations of other entries in the tableau.

Let $\omega = EG(T)$, and let $A_\delta(T)$ be the number of entries greater than $(1-\delta)N$ in region A of T, and similarly define $B_\delta(T)$ with region B. We find

$$0 \le S_{0,\delta}(\omega) - A_\delta(T) \le B_\delta(T). \tag{12}$$

To prove the lemma we show that with probability tending to 1, for a uniformly random tableau T, $B_\delta(T)$ and $\left|A_\delta(T) - \delta N \mathfrak{semi}[a,b]\right|$ are both of order $\delta^2 N$. Here we use Corollary 16, with ε of the corollary equal to δ^2. Consider the tableau T in the (u,v) co-ordinate system of Section 5 as shown in Figure 12. Let H be the region where $v > h_{1-\delta}(u)$ for the function defined in (9) (the shaded region in Figure 12), and let $\mathcal{H}$ be the set of entries of T in H.

Corollary 16 states that with probability tending to 1,

$$\#\Big(\mathcal{H} \,\Delta\, \{\lceil N-\delta N\rceil, \dots, N\}\Big) \le \delta^2 N.$$

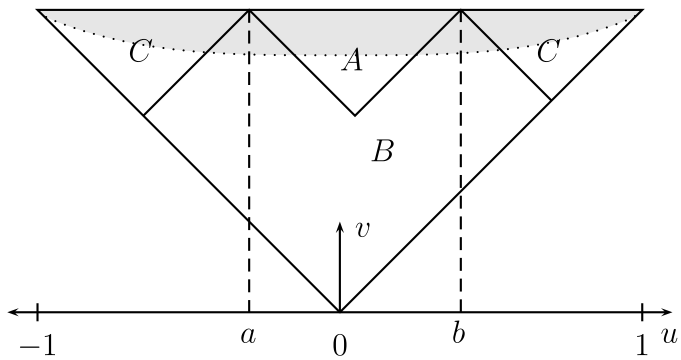

Figure 12: A Young tableau in the (u, v) co-ordinate system. Entries in region A can only exit through the interval $[a, b]$, while entries in C cannot, and entries in B may or may not. The region where $v > h_{1-\delta}(u)$ is shaded—this is the typical location of the entries greater than $(1-\delta)N$.

This implies that with probability tending to 1 we have

$$\Big|A_\delta(T) - N\mathcal{L}(H \cap A)\Big| \le \delta^2 N, \qquad \Big|B_\delta(T) - N\mathcal{L}(H \cap B)\Big| \le \delta^2 N,$$

where $\mathcal{L}$ denotes area of sets in the (u, v) plane. Applying these to (12) we find that (with probability tending to 1)

$$\Big|S_{0,\delta}(\omega) - N\mathcal{L}(H \cap A)\Big| \le \big(2\delta^2 + \mathcal{L}(H \cap B)\big)\, N \tag{13}$$

So, we need to estimate the areas of $H \cap A$ and $H \cap B$.

We use the Taylor expansion of $h_{1-\delta}(u)$ around $\delta = 0$, which is

$$h_{1-\delta}(u) = 1 - \tfrac{2}{\pi}\sqrt{1-u^2}\,\delta + O(\delta^3) \quad \text{as } \delta \to 0, \text{uniformly in } u \in [a, b]$$

(see [19, formula (7), p. 13]). To estimate $\mathcal{L}(H \cap B)$ note that the side length of each of the two "triangles" comprising $H \cap B$ is of order δ. Indeed, each is contained in a square with diagonal $\frac{4}{\pi}\delta$, and so

$$\mathcal{L}(H \cap B) \le \frac{16}{\pi^2}\delta^2 < 2\delta^2. \tag{14}$$

It remains to estimate $\mathcal{L}(H \cap A)$. The fact that $\frac{\partial^2}{\partial \delta^2} h_{1-\delta}(u)\Big|_{\delta=0} \equiv 0$ implies that for δ small enough (depending on a, b), for any $u \in [a, b]$ the error term in the Taylor expansion is at most δ^2. Consequently, the area of $H \cap A$ can be estimated by integrating:

$$\left| \mathcal{L}(H \cap A) - \delta \int_a^b \tfrac{2}{\pi}\sqrt{1-u^2}du \right| \le (b-a)\delta^2 + \mathcal{L}(H \cap B) \le 4\delta^2. \qquad (15)$$

(The $\mathcal{L}(H \cap B)$ term comes from the truncation near a and b.)

The result follows by applying (14) and (15) to (13). $\square$

7 Octagon and Hölder bounds

In this section we prove Theorems 3 and 4.

Lemma 18. *Let $\mathcal{R}_k = \mathcal{R}_k(\omega) = (\lambda_k)_1$ be the length of the first row of the Young diagram λ_k created by the first k steps of the EG^{-1} algorithm from a sorting network ω. Then $\sigma_k^{-1}(i) - i \le \mathcal{R}_k$ for all i, k.*

Proof. The first row of the recording and insertion tableaux during the EG^{-1} algorithm behave exactly the same way as during the celebrated RSK algorithm. It is well known (see [22]) that the RSK algorithm applied to any sequence of numbers creates a tableau whose first row has length given by the longest increasing subsequence (check this on Figure 10).

Here we only need the upper bound, which for completeness we verify here. Observe that the entries of the first row cannot increase as the EG^{-1} algorithm proceeds, as they only change through bumping which replaces an element by something less or equal. So given any increasing subsequence $a_1, \dots, a_\ell$, we know that each a_i has to be inserted to the right of where a_{i-1} was inserted. This shows that at any step k the length of the first row is at least the length of the longest increasing subsequence of the input so far.

Now by time k the position of particle i has changed by $\eta = \sigma_k^{-1}(i) - i$. If $\eta \le 0$ then the required statement is vacuously true, and if $\eta > 0$ then this implies that swaps at positions $i, \dots, i+\eta-1$ have appeared in this order (with possibly other swaps in between). Thus $\eta \le \mathcal{R}_k$, as required. $\square$

Proof of Theorem 4. We use Lemma 18 and Corollary 15. With the notation there, the Edelman-Greene bijection (Theorem 8) shows that if $\omega = EG(T)$ then $\mathcal{R}_k(\omega) = R_k(T)$. Therefore Corollary 15 gives for any $\varepsilon > 0$,

$$\mathbb{P}_{\mathrm{U}}^n \left(\forall k : \mathcal{R}_k < d_k + \varepsilon n \right) \xrightarrow[n\to\infty]{} 1,$$

and by Lemma 18 we deduce

$$\mathbb{P}_{\mathrm{U}}^{n}\left(\forall j,k : \sigma_k^{-1}(j) - j < d_k + \varepsilon n\right) \xrightarrow[n\to\infty]{} 1.$$

Since σ_k is a permutation this is equivalent to

$$\mathbb{P}_{\mathrm{U}}^{n}\left(\forall i,k : i - \sigma_k(i) < d_k + \varepsilon n\right) \xrightarrow[n\to\infty]{} 1.$$

The symmetries (3) and (4) now imply the other three required bounds. ☐

Proof of Theorem 3. Part (i). Fix $\varepsilon > 0$. Consider the event

$$E = \left\{\forall i, \, \forall\, 0 \le j < k \le N \, : \, |\sigma_j^{-1}(i) - \sigma_k^{-1}(i)| \le n\sqrt{\frac{2}{N}}(k-j)^{1/2} + \frac{\varepsilon n}{2}\right\}$$

We will show that

$$\mathbb{P}_{\mathrm{U}}^{n} E \xrightarrow[n\to\infty]{} 1. \tag{16}$$

This will be enough, since the effect of linear interpolation is negligible (that is proving (16) for all $\varepsilon > 0$ implies the required statement for all $\varepsilon > 0$). Denote $M = \lfloor \frac{\varepsilon^2 N}{128} \rfloor$ and $K = \lfloor \frac{128}{\varepsilon^2} \rfloor$. For each integer $0 \le v \le K$, denote the event

$$E_v = \left\{\forall i,k \, : \, |\sigma_k^{-1}(i) - \sigma_{vM}^{-1}(i)| \le \frac{\varepsilon}{8}n + n\left(\frac{2|k - vM|}{N}\right)^{1/2}\right\}.$$

By Theorem 4 together with stationarity, Theorem 1(i) we get $\mathbb{P}_{\mathrm{U}}(E_v) \to 1$. Since the number of these events is fixed, we deduce

$$\mathbb{P}_{\mathrm{U}}\left[\bigcap_{0\le v\le K} E_v\right] \xrightarrow[n\to\infty]{} 1.$$

We claim that if $\omega \in \bigcap_{0\le v\le K} E_v$ then $\omega \in E$. For each $0 \le j < k \le N$ consider two cases. First, it is possible that there is some $0 \le v \le K$ such that $vM \le j < k < (v+1)M$. In this case, $\omega \in E_v$ implies that for all i

$$|\sigma_j^{-1}(i) - \sigma_{vM}^{-1}(i)| \le \frac{\varepsilon}{8}n + n\left(\frac{2|j - vM|}{N}\right)^{1/2} \le \frac{\varepsilon}{8}n + \frac{\varepsilon}{8}n = \frac{\varepsilon}{4}n, \tag{17}$$

where we used the fact that $j - vM < \frac{\varepsilon^2 N}{128}$. Similarly,

$$|\sigma_k^{-1}(i) - \sigma_{vM}^{-1}(i)| \le \frac{\varepsilon}{4}n,$$

and therefore
$$|\sigma_j^{-1}(i) - \sigma_k^{-1}(i)| \le \frac{\varepsilon}{4}n + \frac{\varepsilon}{4}n = \frac{\varepsilon}{2}n.$$

The second possibility is that for some $0 < v \le K$ we have that $(v-1)M < j < vM \le k$. In that case, (17) is still true, and furthermore since $\omega \in E_v$ and $k - vM < k - j$ we get

$$|\sigma_k^{-1}(i) - \sigma_{vM}^{-1}(i)| \le \frac{\varepsilon}{8}n + n\left(\frac{2|k-vM|}{N}\right)^{1/2} \le \frac{\varepsilon}{8}n + n\sqrt{\frac{2}{N}}(k-j)^{1/2}.$$

Combining this with (17) gives

$$|\sigma_j^{-1}(i) - \sigma_k^{-1}(i)| \le \frac{\varepsilon}{2}n + n\sqrt{\frac{2}{N}}(k-j)^{1/2},$$

as claimed.

Part (ii). For some fixed sequence $\varepsilon_n \to 0$, consider the set A_n of continuous functions $T : [0,1] \to [-1,1]$ satisfying

$$\forall t, s \in [0,1] \ : \ |T(t) - T(s)| \le \sqrt{8}|t-s| + \varepsilon_n.$$

By part (i) we can choose $\varepsilon_n \to 0$ so that $\mathbb{P}(T_{i(n)} \in A_n) \to 1$.

Let $w(T,h) = \sup\{|T(t) - T(s)| : |t-s| \le h\}$. It follows from $\mathbb{P}(T_{i(n)} \in A_n) \to 1$ that
$$\lim_{h\to 0}\limsup_{n\to\infty} \mathbb{E}(w(T_{i(n)}, h) \wedge 1) = 0.$$
By [13, Theorem 16.5] we have tightness of the random sequence $T_{i(n)}$ under this condition (note that the target space $[-1,1]$ is compact). This establishes the existence of subsequential limits.

Now if we have a weakly convergent subsequence $T_{i(n(j))}$, then it must have the same limit as the conditioned random variables $\widetilde{T}^j \stackrel{d}{:=} (T_{i(n(j))} \mid T_{i(n(j))} \in A_{n(j)})$. We may realize the sequence $\{\widetilde{T}^j\}$ on the same probability space so that $\widetilde{T}^j \to T$ a.s. [13, Theorem 4.30]. We conclude by observing that any limit of deterministic paths $T^j \in A_{n(j)}$ is Hölder$(\sqrt{8}, \frac{1}{2})$. $\square$

8 Great circles

In this section we prove Theorem 5. The idea is as follows. If a sorting network lies close to a great circle then its trajectories are close to sine curves up to some time change. Equivalently, it is close to a stretchable network obtained by rotating a set of points as in the remark in the introduction. This set of points must have roughly uniform one-dimensional projections in all directions, so its empirical measure must be close to $\mathfrak{Arch}_{1/2}$. Finally, since the inversion number of the resulting configurations is close to linear in the angle of rotation, the time change mentioned above must be linear.

Here are the details. Denote the centre of $\mathbb{S}_n$ by $\mathbf{c} = \left(\frac{n+1}{2}, \dots, \frac{n+1}{2}\right)$ and the radius by $R = \sqrt{\frac{n^3-n}{12}}$. Given the circle c_n we may choose a pair of orthogonal vectors $\mathbf{u}, \mathbf{v}$ of length R so that the circle has the representation $c_n = \{c_n(\theta)\}_{\theta\in\mathbb{R}}$ where

$$c_n(\theta) = \mathbf{c} + \mathbf{u}\cos\theta + \mathbf{v}\sin\theta.$$

For $k \in \{0, \dots, N\}$, define a sequence θ_k (up to addition of multiples of 2π) by

$$\theta_k = \arg\min_\theta \|\sigma_k^{-1} - c_n(\theta)\|_\infty.$$

Thus $c_n(\theta_k)$ is the point of c_n closest in L^∞ to σ_k^{-1}. W.log. we may choose $\mathbf{u}$ so that $\theta_0 = 0$ (this leaves us two possibilities for $\mathbf{v}$). For other k, the angle θ_k is uniquely determined inductively by requiring $|\theta_{k+1} - \theta_k| < \pi$. By symmetry, $\theta_N = (2k+1)\pi$ for some integer k (and we will see that in fact $k = 0$).

Fix some $\varepsilon > 0$. The condition on ω_n implies that for n large enough (depending on ε),

$$\left\|\sigma_k^{-1} - c_n(\theta_k)\right\|_\infty \le \varepsilon n \quad \text{for all } k. \tag{18}$$

Since $\|\sigma_k^{-1} - \sigma_{k+1}^{-1}\|_\infty = 1$, this implies that

$$\|c_n(\theta_{k+1}) - c_n(\theta_k)\|_\infty \le 1 + 2\varepsilon n.$$

Since $R \approx n^{3/2}$, simple geometry implies that for n large enough we have

$$|\theta_{k+1} - \theta_k| \le 2\arcsin\left(\frac{(1+2\varepsilon n)\sqrt{n}}{2R}\right) \le 8\varepsilon,$$

for all k (the $\sqrt{n}$ term comes from passing from the L^∞ norm to the L^2 norm). Thus $\{\theta_k\}$ does not change too quickly. In particular, there must be some k so that either $|\theta_k - \pi/2| \le 4\varepsilon$, or $|\theta_k + \pi/2| \le 4\varepsilon$. We can negate $\mathbf{v}$, so w.log. assume the former is the case.

Considering the ith coordinate in (18), one finds that

$$\left|\sigma_k^{-1}(i) - \left(\tfrac{n+1}{2} + u_i \cos\theta_k + v_i \sin\theta_k\right)\right| \le \varepsilon n. \tag{19}$$

We would like to show that the sorting network is approximated by motion along the circle with constant speed, i.e. that $\theta_k \approx \pi k/N$. If that were the case, part (i) of Theorem 5 would follow. As it is, (19) only implies that the paths are approximately sine curves up to a time change. The key point here is that the same time change applies to all particles.

Define a probability measure ν_n on $\mathbb{R}^2$ by

$$\nu_n = \frac{1}{n}\sum_{i=1}^{n} \delta\left(\tfrac{2}{n}u_i, \tfrac{2}{n}v_i\right),$$

where $\delta(x,y)$ is the delta measure at (x,y). Thus ν_n is the empirical measure for the (rescaled) coordinates of $\mathbf{u}$ and $\mathbf{v}$.

Lemma 19. *With the above notations we have the vague convergence $\nu_n \Longrightarrow \mathfrak{Arch}_{1/2}$.*

Proof. We first claim that ν_n is supported inside the disc of radius 2. Indeed, the vector $\mathbf{c}+\mathbf{u}$ approximates the identity permutation, and so (by (18)) all entries of $\frac{2}{n}\mathbf{u}$ are in $[-1-3\varepsilon, 1+3\varepsilon]$. For $\mathbf{v}$, note that there is some k so that $|\theta_k - \pi/2| < 4\varepsilon$, we find that $\mathbf{c}+\mathbf{v}$ approximates σ_k^{-1}, with some additional error from $\mathbf{u}\cos\theta_k$. Thus the coordinates of $\frac{2}{n}\mathbf{v}$ are all in $[-1-6\varepsilon, 1+6\varepsilon]$.

We use the continuity theorem for the multi-dimensional characteristic function, (see e.g. [13, Theorem 5.3]). Thus it suffices to prove pointwise convergence of the characteristic function of ν_n to the characteristic function of $\mathfrak{Arch}_{1/2}$. This in turn will be deduced from considering the one-dimensional projections of ν_n.

More precisely, (19) says that for n large enough, for all k and i,

$$\left|\left(\tfrac{2}{n}u_i\cos\theta_k + \tfrac{2}{n}v_i\sin\theta_k\right) - \left(\tfrac{2}{n}\sigma_k^{-1}(i) - 1\right)\right| \le 3\varepsilon.$$

This states that the projection of ν_n in direction θ_k can be coupled to the empirical measure of a permutation scaled to $[-1,1]$ so that they differ by at most 2ε. But the scaled empirical measure of a permutation consists of equal point masses along an arithmetic progression, and does not depend on the permutation. Let $F(x) := 0 \vee \frac{x+1}{2} \wedge 1$ be the distribution function of uniform measure on $[-1,1]$. If $P_\theta(a,b) = a\cos\theta + b\sin\theta$ denotes projection on a line in direction θ, then we deduce that for n large enough, for any $x \in \mathbb{R}$,

$$\Big|(P_{\theta_k}\nu_n)(-\infty,x] - F(x)\Big| \le 3\varepsilon + \frac{1}{n} \le 4\varepsilon. \tag{20}$$

For an arbitrary angle θ, there is necessarily some k such that either $|\theta_k - \theta| < 4\varepsilon$, or the same holds for $\theta+\pi$. Fix $x \in [-1,1]$. Note that for two angles ϕ, ψ, and any $z \in \mathbb{R}^2$ we have $|P_\phi z - P_\psi z| \le |z||\phi - \psi|$. Since ν_n is supported inside the disc of radius 2, $P_\theta \nu_n$ is close to $P_{\theta_k}\nu_n$, and so

$$\Big|(P_\theta\nu_n)(-\infty,x] - (P_{\theta_k}\nu_n)(-\infty,x]\Big| \le 8\varepsilon. \tag{21}$$

Combining (20) and (21) gives that for n large enough (depending on ε), for all θ and $x \in [-1,1]$,

$$\Big|(P_\theta\nu_n)(-\infty,x] - F(x)\Big| \le 12\varepsilon.$$

By monotonicity of cumulative distribution funcitons, the same bound holds for $1 \le |x| \le 2$. Since the support of ν_n is bounded in the disc of radius 2, for x outside $[-2,2]$ we have the stronger identity $(P_\theta\nu_n)(-\infty,x] = F(x)$.

We wish to compare the characteristic functions ϕ and ϕ_n of $\mathfrak{Arch}_{1/2}$ and ν_n respectively. Note that for any θ, the measure $P_\theta\mathfrak{Arch}_{1/2}$ is the uniform measure on $[-1,1]$. We have that

$$(\phi - \phi_n)(r\cos\theta, r\sin\theta) = \int_{\mathbb{R}} e^{irx} P_\theta(\nu_n - \mathfrak{Arch}_{1/2})(dx).$$

Integrating by parts gives

$$\begin{aligned}
\Big|(\phi - \phi_n)(r\cos\theta, t\sin\theta)\Big| &\le \int_{\mathbb{R}} \Big| ire^{irx}\big[P_\theta(\nu_n - \mathfrak{Arch}_{1/2})\big](-\infty,x]\Big| dx \\
&\le \int_{-2}^{2} 2\Big|\big[P_\theta(\nu_n - \mathfrak{Arch}_{1/2})\big](-\infty,x]\Big| dx \\
&\le 96\varepsilon.
\end{aligned}$$

Since ε can be arbitrarily small, this proves pointwise convergence of the characteristic functions, and therefore convergence of ν_n to $\mathfrak{Arch}_{1/2}$. $\square$

Lemma 20. *With the above notation,*

$$\max_k \left|\theta_k - \frac{k\pi}{N}\right| \xrightarrow[n\to\infty]{} 0$$

Proof. Fix some θ, and let P_θ denote the projection on a line in direction θ, so that

$$P_\theta(x,y) = x\cos\theta + y\sin\theta,$$

and consider the permutation $\rho_n(\theta)$ derived from $\mathbf{u}, \mathbf{v}$ by arranging $i \in [1,n]$ in increasing order of $P_\theta(u_i, v_i)$. We first estimate the inversion number $\operatorname{inv}(\rho_n(\theta))$. Define

$$A(\theta) = \left\{ ((x,y),(x',y')) \in (\mathbb{R}^2)^2 \ : \ \begin{array}{c} x < x', \\ P_\theta(x,y) > P_\theta(x',y') \end{array} \right\}.$$

We have

$$\frac{1}{N}\operatorname{inv}(\rho_n(\theta)) = \iint \mathbf{1}_{A(\theta)}\, d\nu_n\, d\nu_n,$$

which is a continuous functional of ν_n. Since $\nu_n \Longrightarrow \mathfrak{Arch}_{1/2}$, this implies

$$\frac{1}{N}\operatorname{inv}(\rho_n(\theta)) \xrightarrow[n\to\infty]{} \iint \mathbf{1}_{A(\theta)}\, d\mathfrak{Arch}_{1/2}\, d\mathfrak{Arch}_{1/2} = \frac{\theta}{\pi} \tag{22}$$

To check the last equality, note that the integral is the probability that the x-projections of two points chosen independently from $\mathfrak{Arch}_{1/2}$ change order after rotation by at most θ. By rotational invariance the angle of the line between two such points is uniform on $[0,\pi]$.

Equation (22) holds for any fixed θ. However, $\operatorname{inv}(\rho_n(\theta))$ is increasing in $\theta \in [0,\pi]$, and consequently,

$$\frac{1}{N}\operatorname{inv}(\rho_n(\theta)) \xrightarrow[n\to\infty]{} \frac{\theta}{\pi} \quad \text{uniformly in } \theta \in [0,\pi]. \tag{23}$$

Comparing (19) for i,j we find that for any $\varepsilon > 0$, for n large enough we have

$$\sigma_k^{-1}(i) - \sigma_k^{-1}(j) = (u_i - u_j)\cos\theta_k + (v_i - v_j)\sin\theta_k + \delta \tag{24}$$

with $|\delta| < 2\varepsilon n$ for all i,j. Say a pair (i,j) is **θ-uncertain** if

$$|(u_i - u_j)\cos\theta_k + (v_i - v_j)\sin\theta_k| \le 2\varepsilon.$$

So, for any $i < j$ we have that $\sigma_k^{-1}(i) > \sigma_k^{-1}(j)$ if and only if $(\rho_n(\theta_k))(i) > (\rho_n(\theta_k))(j)$, unless (i,j) is θ_k-uncertain. Recall that any such pair $i < j$ with $\sigma_k^{-1}(i) > \sigma_k^{-1}(j)$ contributes 1 to the number of inversions of σ_k^{-1} (hence, if (i,j) is not θ_k-uncertain, also to $\operatorname{inv}(\rho_n(\theta_k))$). Consequently, $\operatorname{inv}(\rho_n(\theta_k))$ differs from $\operatorname{inv}(\sigma_k^{-1}) = k$ by at most the number of θ_k-uncertain pairs.

It remains to bound the number of θ-uncertain pairs. Fix $\varepsilon > 0$, and consider the set $S_{4\varepsilon}$ of all strips of width 4ε in $\mathbb{R}^2$. Since $\nu_n \Longrightarrow \mathfrak{Arch}_{1/2}$, we have

$$\limsup_{n\to\infty} \sup_{A\in S_{4\varepsilon}} \nu_n(A) \le 2\varepsilon$$

because $\mathfrak{Arch}_{1/2}(A) \le 2\varepsilon$ for any such strip. This implies that for large n for any i there are at most $2\varepsilon n$ values of j such that (i,j) is θ-uncertain for some θ. In summary, for n large enough, depending only on ε, and any θ, the total number of θ-uncertain points is at most $2\varepsilon n^2$.

Combining (24) and the above discussion we find that for large n

$$|\operatorname{inv}(\sigma_k) - \operatorname{inv}(\rho_n(\theta_k))| \le 2\varepsilon n^2,$$

uniformly in θ. Since $\operatorname{inv}(\sigma_k) = k$, combining with (23) yields the result. □

Proof of Theorem 5. Combining Lemma 20 and (19) gives part (i):

$$\max_{i,t} \left| \left(\tfrac{2}{n}\sigma_{[tN]}(i) - 1\right) - \tfrac{2}{n}(u_i \cos(\pi t) + v_i \sin(\pi t)) \right| \xrightarrow[n\to\infty]{} 0. \tag{25}$$

In particular, we have $2u_i/n = 2i/n - 1 + o(1)$. By inserting (25) into the definition of μ_t in (1), we find that $\mu_t(\omega_n)$ is close to $R_t\nu_n$, where R_t is the linear map $R_t(x,y) = (x, x\cos(\pi t) + y\sin(\pi t))$, in the sense the two measures can be coupled with maximal distance tending to 0. Since $\nu_n \Longrightarrow \mathfrak{Arch}_{1/2}$, this implies $\mu_t(\omega_n) \Longrightarrow \mathfrak{Arch}_t$, which is (ii).

Next, we prove (iii). To sample from the scaled swap process $\eta(\omega_n)$ one may choose uniformly a pair of particles i,j and consider the time and location of their swap. Consider the pair of points $z_i = \frac{2}{n}(u_i, v_i)$ and $z_j = \frac{2}{n}(u_j, v_j)$. If i,j are swapped at step k of the network, then $\sigma_k^{-1}(i) - \sigma_k^{-1}(j) = 1$, so by (24) we have for n large enough that $|P_{\theta_k}(z_i) - P_{\theta_k}(z_j)| \le 2\varepsilon$, and by (19) the scaled location of the swap is given to within 2ε by $P_{\theta_k} z_i$. Thus for any i,j, the time of the (i,j) swap is given by the angle of a certain line, and the location of the swap by the distance of the line from the origin, where this line passes within distance ε of both z_i and z_j. Thus, unless z_i, z_j are sufficiently close, the location of the two points approximately determines the time and place of the swap.

Specifically, for any pair z, z', as $\varepsilon \to 0$ the set of possible times converges to a single time, and the set of possible locations converges to a single location. Since $\nu_n \Longrightarrow \mathfrak{Arch}_{1/2}$, it follows that $\eta(\omega_n)$ converges to the measure resulting from applying the same operation to $\mathfrak{Arch}_{1/2}$.

Let $z, z' \in \mathbb{R}^2$ be chosen independently with law $\mathfrak{Arch}_{1/2}$. Let $\theta \in [0, \pi]$ be the angle that the line through them makes with the positive y-axis, and let $r := z_1 \cos\theta + z_2 \sin\theta$ be its signed distance from the origin. It remains to prove that θ and r are independent, θ is uniform in $[0, \pi]$ and r has law $\mathfrak{semi}$. Independence and uniformity of θ are clear by rotational symmetry of $\mathfrak{Arch}_{1/2}$. Finally, to calculate the distribution of r we introduce some further variables. Let $\widehat{z}, \widehat{z}'$ be z, z' rotated by $-\theta$ and let $\widehat{z} = (r, y)$ and $\widehat{z}' = (r, y')$ be their coordinates. Let $w = y/\sqrt{1-r^2}$ and $w' = y'/\sqrt{1-r^2}$. Thus we have

$$\begin{aligned} z_1 &= r\cos\theta - w\sqrt{1-r^2}\sin\theta \\ z_2 &= r\sin\theta + w\sqrt{1-r^2}\cos\theta \\ z_1' &= r\cos\theta - w'\sqrt{1-r^2}\sin\theta \\ z_2' &= r\sin\theta + w'\sqrt{1-r^2}\cos\theta. \end{aligned}$$

We can compute the probability density function of r using the Jacobian of the transformation $(z_1, z_2, z_1', z_2') \mapsto (r, \theta, w, w')$; after some straightforward manipulation we obtain

$$\int_0^\pi \int_{-1}^1 \int_{-1}^1 \left| \frac{\partial(z_1, z_2, z_1', z_2')}{\partial(r, \theta, w, w')} \right| \frac{1}{2\pi\sqrt{1-\|z\|_2^2}} \, \frac{1}{2\pi\sqrt{1-\|z'\|_2^2}} \, dw\, dw'\, d\theta \\ = \tfrac{2}{\pi}\sqrt{1-r^2}$$

as required. □

Acknowledgements

We thank Nathanaël Berestycki, Alex Gamburd, Alan Hammond, Pawel Hitczenko, Martin Kassabov, Rick Kenyon, Scott Sheffield, David Wilson and Doron Zeilberger for many valuable conversations. This work has benefitted greatly from the superb resources provided by a meeting at BIRS (Banff, Canada) and by the 2005 programme in Probability, Algorithms and Statistical Physics at MSRI (Berkeley, USA).

References

[1] O. Angel, A. E. Holroyd, and D. Romik. Directed random walk on the permutahedron. In preparation.

[2] O. Angel, A. E. Holroyd, and B. Virag. The local limit of the uniform sorting network. In preparation.

[3] O. Angel, M. Kassabov, A. E. Holroyd, D. Romik, and B. Virag. Bounds on the uniform sorting network. In preparation.

[4] P. Biane. Representations of symmetric groups and free probability. *Adv. Math.*, 138(1):126–181, 1998.

[5] P. Edelman and C. Greene. Balanced tableaux. *Adv. in Math.*, 63(1):42–99, 1987.

[6] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[7] S. Felsner. The skeleton of a reduced word and a correspondence of Edelman and Greene. *Electron. J. Combin.*, 8(1):Research Paper 10, 21 pp. (electronic), 2001.

[8] S. R. Finch. *Mathematical constants*, volume 94 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2003.

[9] J. S. Frame, G. B. Robinson, and R. M. Thrall. The hook graphs of the symmetric groups. *Canadian J. Math.*, 6:316–324, 1954.

[10] A. Garsia. The saga of reduced factorizations of elements of the symmetric group. Preprint, `http://www.math.ucsd.edu/~garsia/recentpapers/saga.pdf`.

[11] J. E. Goodman and J. O'Rourke, editors. *Handbook of discrete and computational geometry*. Discrete Mathematics and its Applications (Boca Raton). Chapman & Hall/CRC, Boca Raton, FL, second edition, 2004.

[12] C. Greene, A. Nijenhuis, and H. S. Wilf. A probabilistic proof of a formula for the number of Young tableaux of a given shape. *Adv. in Math.*, 31(1):104–109, 1979.

[13] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[14] S. Kerov. A differential model for the growth of Young diagrams. In *Proceedings of the St. Petersburg Mathematical Society, Vol. IV*, volume 188 of *Amer. Math. Soc. Transl. Ser. 2*, pages 111–130, Providence, RI, 1999. Amer. Math. Soc.

[15] S. V. Kerov. Transition probabilities of continual Young diagrams and the Markov moment problem. *Funktsional. Anal. i Prilozhen.*, 27(2):32–49, 96, 1993.

[16] D. E. Knuth. *The Art of Computer Programming, Vol. 3: Sorting and Searching.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973. Addison-Wesley Series in Computer Science and Information Processing.

[17] A. Lascoux and M.-P. Schützenberger. Structure de Hopf de l'anneau de cohomologie et de l'anneau de Grothendieck d'une variété de drapeaux. *C. R. Acad. Sci. Paris Sér. I Math.*, 295(11):629–633, 1982.

[18] D. P. Little. Combinatorial aspects of the Lascoux-Schützenberger tree. *Adv. Math.*, 174(2):236–253, 2003.

[19] B. Pittel and D. Romik. Limit shapes for random square young tableaux. *Adv. Appl. Math.*, to appear.

[20] V. Reiner. Note on the expected number of Yang-Baxter moves applicable to reduced decompositions. *European J. Combin.*, 26(6):1019–1021, 2005.

[21] R. P. Stanley. On the number of reduced decompositions of elements of Coxeter groups. *European J. Combin.*, 5(4):359–372, 1984.

[22] R. P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

[23] A. T. White. Fabian Stedman: the first group theorist? *Amer. Math. Monthly*, 103(9):771–778, 1996.

Omer Angel: `angel@utstat.toronto.edu`
Department of Statistics, University of Toronto,
100 St George St., Toronto ON M5S 3G3, Canada.

Alexander E. Holroyd: `holroyd@math.ubc.ca`
Department of Mathematics, University of British Columbia,
121-1984 Mathematics Rd., Vancouver BC V6T 1Z2, Canada.

Dan Romik: `romik@stat.berkeley.edu`
Department of Statistics, University of California,
367 Evans Hall, Berkeley CA 94720-3860, USA.

Bálint Virág: `balint@math.toronto.edu`
Department of Mathematics, University of Toronto,
40 St George St., Toronto ON M5S 2E4, Canada.