

FLUCTUATIONS OF SCHENSTED ROW INSERTION

BY MIKOŁAJ MARCINIAK^{1,a}  AND PIOTR ŚNIADY^{2,b} 

¹*Interdisciplinary Doctoral School “Academia Copernicana”, Faculty of Mathematics and Computer Science, Nicolaus Copernicus University in Toruń, marciniak@int.pl*

²*Institute of Mathematics, Polish Academy of Sciences, psniady@impan.pl*

We investigate asymptotic probabilistic phenomena arising from the application of the Schensted row insertion algorithm, a key component of the Robinson–Schensted–Knuth (RSK) correspondence, to random inputs. Our analysis centers on a random tableau T with a given shape λ , which may itself be random or deterministic. We examine the stochastic properties of the position of the new box created when inserting a deterministic entry into T . Specifically, we focus on the fluctuations of this position around its expected value as the size of the Young diagram λ approaches infinity. Our findings reveal that these fluctuations are asymptotically Gaussian, with the mean and variance expressed in terms of Kerov’s transition measure of the diagram λ . An important application of this analysis is the RSK algorithm applied to a finite, long sequence of independent, identically distributed random variables. While there remains a gap in the reasoning for this case, we present an explicit conjecture regarding its behavior.

CONTENTS

1. Teaser: New conjectures related to RSK algorithm applied to random input	2628
1.1. Basic definitions	2628
1.1.1. Young diagrams and tableaux	2628
1.1.2. The Schensted row insertion	2629
1.1.3. The Robinson–Schensted–Knuth algorithm	2630
1.2. Context and motivations: RSK applied to random input	2630
1.2.1. Plancherel measure	2630
1.2.2. Extremal characters of $\mathfrak{S}_\infty$	2630
1.2.3. The key motivation: RSK as an isomorphism of dynamical systems	2631
1.3. The main problem: Position of the new box	2632
1.3.1. Notations	2632
1.3.2. The main problem: Relationship between the value of the new entry and the position of the new box	2632
1.4. The limit shape and its parametrization	2633
1.4.1. Scaling of Young diagrams	2634
1.5. The Logan–Shepp–Vershik–Kerov limit shape	2634
1.6. Position of the new box for an unspecified new entry	2634
1.7. The natural parametrization of the limit curve	2635
1.8. The first main conjecture: The fluctuations of Schensted row insertion	2635
1.9. Donsker’s functional central limit theorem for jeu de taquin trajectories	2637
1.10. Second class particles	2638
2. Main results and structure of the paper	2639
2.1. Poissonized tableaux: A generalization	2639
2.1.1. Plancherel tableau, the key example	2639
2.1.2. Poissonized tableaux of a given shape	2640
2.1.3. Additional randomness of the shape	2640

Received October 2024; revised September 2025.

MSC2020 subject classifications. Primary 60C05; secondary 60F05, 05E10, 20C30, 60K35, 82C22.

Key words and phrases. Robinson–Schensted–Knuth correspondence, Schensted row insertion, random Young tableaux, limit shape, jeu de taquin, second class particles.

2.1.4. Challenges in applying results to random Plancherel tableaux	2640
2.2. Assumptions of the main results. Challenges in describing large Young diagrams	2640
2.2.1. Method 1: Continual diagrams and the Russian convention	2640
2.2.2. Method 2: Kerov's transition measure	2640
2.2.3. Assumptions of the main results. Navigating between the two viewpoints	2641
2.3. Conclusions of the main results	2641
2.4. Proof strategy for the asymptotic Gaussianity of the new box position	2641
2.5. Content of the paper	2642
3. Preliminaries: Transition measure and continual diagrams	2643
3.1. Plancherel growth process	2643
3.2. Transition measure of a Young diagram	2643
3.3. Continual diagrams	2644
3.4. Transition measure for continual diagrams	2644
3.5. Transition measure as a homeomorphism	2645
3.6. The metric d_{XY} on the set of continual diagrams	2645
4. The general form of RSK insertion fluctuations	2646
4.1. Asymptotic determinism of Schensted insertion	2646
4.2. The interaction energy	2647
4.3. The main theorem: General form of the fluctuations	2648
4.4. Explicit examples for Theorem 4.3	2649
4.4.1. Example: Staircase tableaux	2649
4.4.2. Almost isosceles triangle tableaux	2650
4.4.3. More general almost triangle tableaux	2651
4.4.4. Where to look for more concrete examples?	2653
4.5. Towards Conjecture 1.5, part 1	2653
4.5.1. General strategy	2653
4.5.2. Problematic Assumption (A2)	2653
4.5.3. Towards Conjecture 4.7	2654
4.5.4. Transition measure for the limit shape and its interaction energy	2655
4.5.5. Conclusion	2656
5. A version of the main theorem with weaker assumptions	2656
5.1. Motivations	2656
5.1.1. Tension between local and global	2656
5.1.2. Transition measure as the language of the proof	2657
5.2. Notations	2657
5.3. The main theorem, alternative version	2657
5.4. Towards Conjecture 1.7, part 2	2658
6. Implication between two versions of the main theorem	2659
6.1. Assumptions of Theorem 4.3 imply Assumption (B1)	2659
6.2. Assumption (B2). The lower bound	2659
6.3. Bounding the interaction energy from above	2660
6.3.1. The expected value of $\mathcal{E}_v(\mathbf{u}_n)$ as a double integral. Decomposition of the integration domain	2661
6.3.2. Close to the singularity	2662
6.3.3. Extremely close to the singularity	2665
6.3.4. Both variables on the same side of z_0 , but not too close	2665
6.3.5. The main contribution to the integral (away from the singularity)	2666
6.3.6. One variable very close to z_0 , the other far away	2667
6.3.7. Conclusion of the proof of Proposition 6.2	2667
6.4. Proposition 6.2 implies the upper bound in Assumption (B2)	2667
6.5. Towards Assumption (B3). Cauchy transform in a random point is not too singular	2668
6.5.1. Conditioning	2669
6.5.2. Split the domain of integration	2669
6.5.3. Conclusion of the proof of Lemma 6.10	2672
6.6. Assumptions of Theorem 4.3 imply Assumption (B3)	2672
7. The first tool: The cumulative function of a tableau	2672
7.1. Cumulative function of a tableau	2672
7.2. Wider perspective: Cumulative function of a random tableau	2673
7.2.1. Plancherel-distributed Poissonized tableaux	2673

- 7.2.2. Future research directions 2675
- 7.3. Single-point evaluation of the cumulative function 2675
 - 7.3.1. Cumulants of a probability distribution 2675
 - 7.3.2. Cumulants of the cumulative function 2675
- 8. The second tool: The double cumulative function of a Young diagram 2676
 - 8.1. Definition 2676
 - 8.2. Partitioning of the domain 2676
 - 8.3. Properties for fixed u 2677
 - 8.4. Properties for fixed z 2677
 - 8.5. Navigating the two viewpoints 2678
 - 8.6. Examples 2678
 - 8.6.1. The example for $\lambda = (3, 1)$ 2678
 - 8.6.2. Staircase tableaux 2678
- 9. First-order asymptotics of the double cumulative function 2678
- 10. Fine asymptotics of the double cumulative function 2680
 - 10.1. New coordinate system 2681
 - 10.2. The Gaussian profile of the double cumulative function 2681
 - 10.3. Case study: Large staircase tableaux 2681
 - 10.4. Normalized cumulative function 2682
 - 10.5. Randomized probability distribution of $X_{n,q}$ 2682
 - 10.5.1. Randomized approach 2683
 - 10.5.2. Formal framework 2683
 - 10.5.3. Two spacial coordinate systems 2683
 - 10.5.4. Conditional moments and conditional cumulants 2683
 - 10.5.5. Finite number of moments is enough 2684
 - 10.5.6. Proof of Proposition 10.4 2684
 - 10.6. Conditioning on a bounded interval 2684
 - 10.7. Conditional CDF of $X_{n,Q}$ 2685
 - 10.8. The consequences of monotonicity of the double cumulative function 2685
- 11. Proofs of the main results 2685
 - 11.1. The main idea 2686
 - 11.2. First-order asymptotics for Schensted insertion 2686
 - 11.2.1. Kerov transition measure formulation 2686
 - 11.2.2. Proof of Theorem 4.1 2686
 - 11.3. Gaussianity of insertion fluctuations: The proof 2687
- Acknowledgments 2687
- Funding 2687
- References 2687

1. Teaser: New conjectures related to RSK algorithm applied to random input. This paper is quite extensive; to engage the reader, we begin with a teaser: two new conjectures (Conjecture 1.5 and Conjecture 1.7) concerning the Robinson–Schensted–Knuth (RSK) algorithm applied to random input.

1.1. *Basic definitions.* We begin by reviewing some fundamental combinatorial concepts. For a more comprehensive treatment of this topic, we refer the reader to the book of Fulton [8].

1.1.1. *Young diagrams and tableaux.* A Young diagram is a finite collection of boxes arranged in the positive quadrant, aligned to the left and bottom edges. This arrangement is known as the French convention (see Figure 1a). To each Young diagram with ℓ rows, we associate an integer partition $\lambda = (\lambda_1, \dots, \lambda_\ell)$, where λ_j denotes the number of boxes in the j th row, counting from bottom to top. We identify a Young diagram with its corresponding partition λ and denote the total number of boxes by $|\lambda| = \lambda_1 + \dots + \lambda_\ell$.

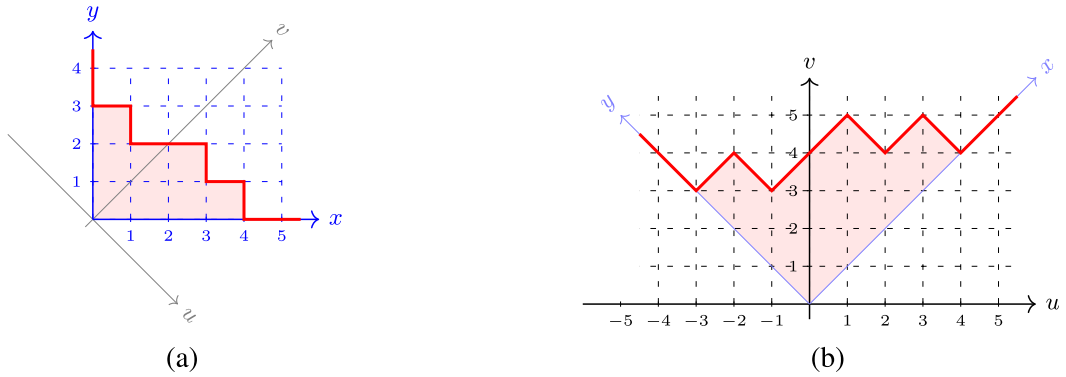


FIG. 1. The Young diagram $(4, 3, 1)$ is depicted in two conventions: (a) the French convention, (b) the Russian convention. In both representations, the solid red line illustrates the diagram's profile. The coordinate systems are as follows: (x, y) for the French convention, and (u, v) for the Russian convention.

For asymptotic problems, it is convenient to draw Young diagrams using the *Russian convention* (see Figure 1b). This corresponds to the coordinate system (u, v) , which relates to the usual French Cartesian coordinates as follows:

$$(1.1) \quad u = x - y, \quad v = x + y.$$

A *tableau* (also known as *semistandard tableau*) is a filling of the boxes of a Young diagram with numbers; we require that the entries should be weakly increasing in each row (from left to right) and strictly increasing in each column (from bottom to top). An example is given in Figure 2a. We say that a tableau T of shape λ is a *standard Young tableau* if it contains only entries from the set $\{1, 2, \dots, |\lambda|\}$ and each element is used exactly once.

1.1.2. The Schensted row insertion. The *Schensted row insertion* is an algorithm that takes a tableau T and a number z as input. The process begins by inserting z into the first (bottom) row of T , following these rules:

- z is placed in the leftmost box containing an entry strictly larger than z .
- If no such box exists, z is appended to the end of the row in a new box, and the algorithm terminates.
- If z displaces an existing entry z' , this z' is “bumped” to the second row.
- The process repeats with z' being inserted into the second row, following the same rules.
- This continues until a number is inserted into an empty box.

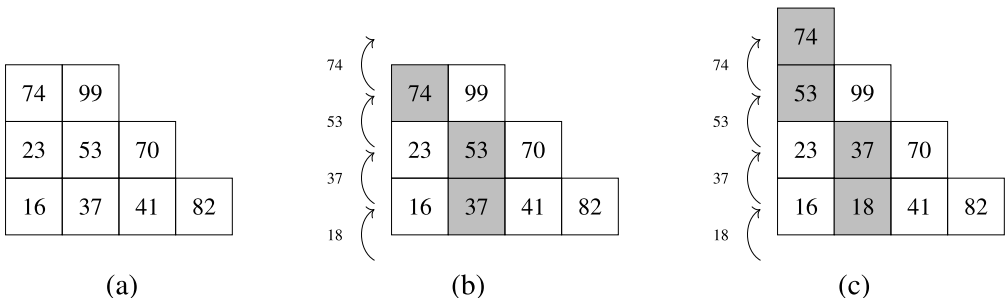


FIG. 2. (a) The original tableau T . (b) The highlighted boxes indicate the bumping route for the Schensted insertion $T \leftarrow 18$. The numbers next to the arrows represent the bumped entries. (c) The resulting tableau after the Schensted insertion $T \leftarrow 18$.

The resulting tableau is denoted as $T \leftarrow z$, see Figures 2b and 2c for an example. The sequence of boxes whose contents change during this process is called *the bumping route*.

Schensted insertion is a key component of the Robinson–Schensted–Knuth algorithm (RSK), see below.

1.1.3. *The Robinson–Schensted–Knuth algorithm.* This article considers a simplified version of the Robinson–Schensted–Knuth algorithm (RSK), which is more accurately described as the Robinson–Schensted algorithm. However, we retain the RSK acronym due to its widespread recognition. The RSK algorithm maps a finite sequence $w = (w_1, \dots, w_n)$ to a pair of tableaux: the insertion tableau $P(w)$ and the recording tableau $Q(w)$.

The insertion tableau is defined as:

(1.2)
$$P(w) = (((\emptyset \leftarrow w_1) \leftarrow w_2) \leftarrow \dots) \leftarrow w_n.$$

This represents the result of iteratively applying Schensted insertion to the entries of w , beginning with an empty tableau $\emptyset$.

The recording tableau $Q(w)$ is a standard Young tableau with the same shape as $P(w)$. Each entry in $Q(w)$ corresponds to the iteration number in (1.2) when that box was first filled. In other words, $Q(w)$ records the order in which the entries of the insertion tableau were populated. Both $P(w)$ and $Q(w)$ share a common shape, denoted as $\text{RSK}(w)$, which we refer to as *the RSK shape associated with w* .

The RSK algorithm is a fundamental tool in algebraic combinatorics and representation theory, particularly in relation to Littlewood–Richardson coefficients (see [8, 28]).

1.2. *Context and motivations: RSK applied to random input.* A fruitful area of research is investigating the Robinson–Schensted–Knuth algorithm applied to random input. We shall review some selected highlights of this field.

1.2.1. *Plancherel measure.* The simplest example involves applying RSK to a finite sequence $w = (w_1, \dots, w_n)$ of i.i.d. random variables uniformly distributed on the unit interval $[0, 1]$. The resulting probability distribution of $\text{RSK}(w)$ is the celebrated *Plancherel measure* Pl_n on Young diagrams with n boxes. This measure arises naturally in the context of decomposing the left regular representation of the symmetric group $\mathfrak{S}_n$ into irreducible components. For a Young diagram λ with n boxes, the Plancherel measure is defined as:

$$\text{Pl}_n(\lambda) = \frac{(f^\lambda)^2}{n!},$$

where f^λ denotes the number of standard Young tableaux of shape λ , see [23], Section 1.8, A remarkable result related to the probability distribution of $\text{RSK}(w)$ is the solution to *the Ulam–Hammersley problem*. This solution [1, 20] reveals a surprising connection with *the Tracy–Widom distribution* [30], which originates in random matrix theory. The Tracy–Widom distribution describes the asymptotic behavior of the largest eigenvalue in certain random matrix ensembles and has found applications in various fields, including growth models.

For a comprehensive introduction to these topics, we recommend the book by Romik [23].

1.2.2. *Extremal characters of $\mathfrak{S}_\infty$.* Consider an infinite sequence of i.i.d. random variables $w_1, w_2, \dots$ (possibly with a more complex probability distribution including atoms). Let

(1.3)
$$\lambda^{(n)} = \text{RSK}(w_1, \dots, w_n)$$

denote the RSK shape corresponding to the first n variables. The resulting sequence of Young diagrams

$$(1.4) \quad \emptyset = \lambda^{(0)} \nearrow \lambda^{(1)} \nearrow \dots$$

forms a Markov random walk on the Young graph, where vertices are Young diagrams and edges connect diagrams differing by one box. This random walk (1.4) has some additional convenient properties, which are out of scope of the current paper.

Vershik and Kerov [15, 32] demonstrated that classifying such random walks on the Young graph with these additional properties (a problem intersecting probability theory, harmonic analysis on the Young graph, and ergodic theory) is equivalent to identifying *the extremal characters of the infinite symmetric group* $\mathfrak{S}_\infty$. This approach led to a novel, conceptual proof of Thoma's classification of such characters [29]. As a consequence, there exists a bijection between the extremal characters of $\mathfrak{S}_\infty$ and the probability laws of the random variables (w_n) . The RSK algorithm provides an explicit method for generating the corresponding random walk (1.4).

1.2.3. *The key motivation: RSK as an isomorphism of dynamical systems.* Consider the recording tableau for an infinite sequence $w_1, w_2, \dots$:

$$Q_\infty(w_1, w_2, \dots) := \lim_{n \rightarrow \infty} Q(w_1, \dots, w_n).$$

This infinite standard tableau fills a subset of the upper-right quarterplane, where each natural number appears exactly once, with rows and columns increasing. Each entry in $Q_\infty(w_1, w_2, \dots)$ represents the iteration at which the corresponding box became nonempty during the infinite sequence of row insertions:

$$((\emptyset \leftarrow w_1) \leftarrow w_2) \leftarrow \dots.$$

Thus, Q_∞ encodes the infinite path (1.4) in the Young graph.

The connection between RSK and the Plancherel measure can be formulated as follows: Q_∞ is a homomorphism between two probability spaces:

- the product space $[0, 1]^\infty$ with the product Lebesgue measure (representing a sequence of i.i.d. uniform random variables on $[0, 1]$), and
- the set of infinite standard Young tableaux with the Plancherel measure (fundamental in the harmonic analysis of $\mathfrak{S}_\infty$).

Each space admits a natural measure-preserving transformation: the one-sided shift and *the jeu de taquin transformation*, respectively. This leads to our central question.

PROBLEM 1.1 (Key motivation). *Is Q_∞ an isomorphism of measure-preserving dynamical systems? If so, how can we construct its inverse?*

The answer is nontrivial because finite RSK requires both the recording tableau $Q(w)$ and the insertion tableau $P(w)$ to recover w , but $P(w)$ is unavailable in the infinite case. An affirmative answer would illuminate aspects of harmonic analysis on $\mathfrak{S}_\infty$, such as the ergodicity of jeu de taquin.

Romik and the second named author [24] investigated this problem, and we present their findings below. Spoiler alert: Problem 1.1 has a positive resolution.

1.3. *The main problem: Position of the new box.*

1.3.1. *Notations.* For a finite sequence $w = (w_1, \dots, w_{n+1})$, we denote by

$$\text{Ins}(w_1, \dots, w_n; w_{n+1}) = (x_n, y_n)$$

the coordinates of the last box which was inserted to the Young diagram by the RSK algorithm applied to the sequence w . In other words, it is the box containing the biggest number in the recording tableau $Q(w)$. Above, (x_n, y_n) refer to the Cartesian coordinates of this box in the French convention, that is, x_n is the number of the column and y_n is the number of the row. By

$$\text{u-Ins}(w_1, \dots, w_n; w_{n+1}) = x_n - y_n$$

we denote the u -coordinate of the aforementioned box.

For later use, given a tableau T and a real number z we denote by $\text{Ins}(T; z)$ the coordinates of the new box which was created by the Schensted row insertion $T \leftarrow z$; in other words it is the unique box of the skew diagram

$$\text{shape}(T \leftarrow z) / \text{shape } T.$$

The quantity $\text{u-Ins}(T; z)$ is defined in an analogous way as the u -coordinate of $\text{Ins}(T; z)$.

1.3.2. *The main problem: Relationship between the value of the new entry and the position of the new box.* In this paper, we focus on a fundamental case where $w_1, w_2, \dots$ is a sequence of i.i.d. random variables, each uniformly distributed on the unit interval $[0, 1]$. Romik and the second named author [24] observed that constructing the inverse map to Q_∞ (and consequently, providing a positive answer to Problem 1.1) requires addressing the following question about the finite version of RSK applied to such a random input.

PROBLEM 1.2 (The main problem). *Let $w = (w_1, w_2, \dots)$ be a sequence of i.i.d. random variables, each uniformly distributed on $[0, 1]$. We seek to understand the relationship between:*

- *The value of the new entry w_{n+1} , and*
- *The position of the corresponding newly created box:*

$$(1.5) \qquad (x_n, y_n) = \text{Ins}(w_1, \dots, w_n; w_{n+1}).$$

Our interest lies in the asymptotic behavior of this probabilistic relationship as $n \rightarrow \infty$.

The RSK algorithm’s dynamics can be visualized by replacing each entry (x, y) in the recording tableau $Q(w) = (Q_{xy})$ with the corresponding element from the input sequence $w = (w_1, \dots, w_n)$ that led to its creation. The resulting matrix $(w_{Q_{xy}})$ is termed the *responsibility matrix* (see Figure 3b). For improved readability in Figure 3, we sample integers from $[0, 100]$ instead of using real numbers.

Following Pittel and Romik’s approach [21], Section 1.1, we can represent the responsibility matrix geometrically as a three-dimensional array of cuboids above the plane $\mathbb{R}^2 \times \{0\}$. Each cuboid’s height corresponds to $w_{Q_{xy}}$, with a unit square $[x - 1, x] \times [y - 1, y] \times \{0\}$ as its base. Alternatively, we can interpret the function $(x, y) \mapsto w_{Q_{xy}}$ as the graph of a (discontinuous) surface forming the upper envelope of this array.

By rescaling the base unit squares to side length $\frac{1}{\sqrt{n}}$ (where n is the length of sequence w), we normalize the total base area to 1 (see Figure 4). In both Figure 3b and Figure 4, we use *layer tinting* to indicate elevation, that is, the values of the responsibility matrix.

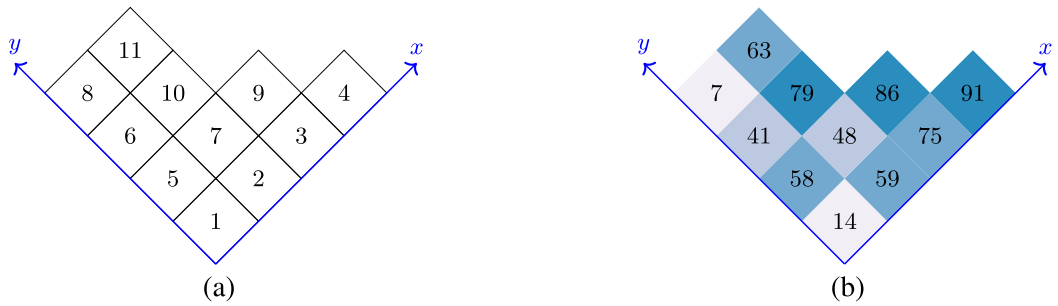


FIG. 3. (a) Recording tableau $Q_{xy} = Q(w)$ (depicted using Russian convention) corresponding to the sequence $w = (14, 59, 75, 91, 58, 41, 48, 7, 86, 79, 63)$ selected from the interval $J = [0, 100]$. (b) Responsibility matrix $(w_{Q_{xy}})$ derived by substituting each entry of the recording tableau with its corresponding value from sequence w . The background employs a four-color layer tinting scheme, representing different quarters of the interval J : $[0, 25]$ (near-white), $(25, 50]$ (beige), $(50, 75]$ (blue), and $(75, 100]$ (dark blue).

Monte Carlo simulations, such as the one depicted in Figure 4, suggest that the value of the new entry w_{n+1} in Problem 1.2 approximately determines the ray (a half-line originating from the coordinate system's origin) on which the new box $\text{Ins}(w_1, \dots, w_n; w_{n+1})$ will appear. This behavior is observed with high probability as $n \rightarrow \infty$. Smaller values of the new entry ($w_{n+1} \approx 0$) tend to correspond to rays closer to the Oy axis (the nearly white area in Figure 4), while larger values ($w_{n+1} \approx 1$) typically align with rays nearer to the Ox axis (the dark blue area in Figure 4).

1.4. *The limit shape and its parametrization.* To address Problem 1.2, we begin with a crucial observation: the newly created box $\text{Ins}(w_1, \dots, w_n; w_{n+1})$ must be situated in one of the concave corners of the Young diagram $\lambda^{(n)} = \text{RSK}(w_1, \dots, w_n)$. Consequently, under-

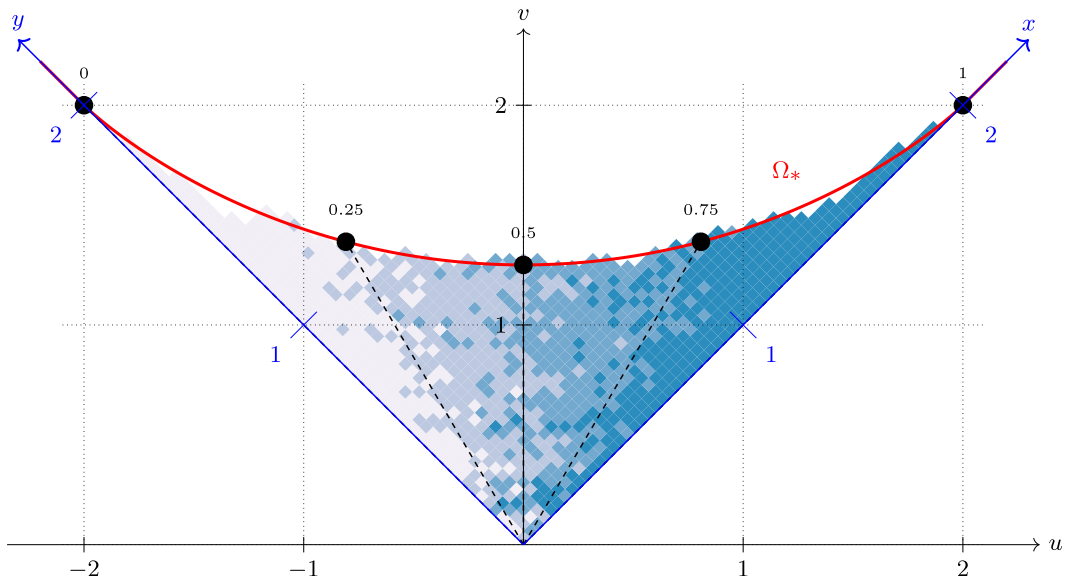


FIG. 4. Analogue of Figure 3b for a sequence $w = (w_1, \dots, w_{1,000})$ comprising $n = 1,000$ i.i.d. random variables following the uniform distribution $U(0, 1)$ on the unit interval $I = [0, 1]$. The layer tinting represents the values of the responsibility matrix $(w_{Q_{xy}})$, with four color-coded intervals: $[0, 1/4]$ (near-white), $(1/4, 1/2]$ (beige), $(1/2, 3/4]$ (blue), and $(3/4, 1]$ (dark blue). The red solid line depicts the Logan–Shepp–Vershik–Kerov limit curve Ω_* . Five black dots indicate its natural parametrization, dividing the area between the curve and the Oxy axes into four curvilinear triangles of equal areas.

standing the asymptotic behavior of the random Young diagram $\lambda^{(n)}$ becomes paramount for our analysis. Fortunately, the probability distribution of such an RSK shape is well understood: it follows the Plancherel measure on Young diagrams with n boxes. The limit shape for this measure has been extensively studied and characterized. We will review this limit shape in the following discussion. For a pedagogical introduction to this topic, we refer to [23], Chapter 1.

1.4.1. *Scaling of Young diagrams.* The boundary of a Young diagram λ is called its *profile*, see Figure 1a. In the Russian coordinate system the profile can be seen as the plot of the function $\omega_\lambda : \mathbb{R} \rightarrow \mathbb{R}_+$, see Figure 1b.

If $c > 0$ is a positive number, the output $c\lambda \subset \mathbb{R}_+^2$ of a homogeneous dilation with scale c applied to the Young diagram λ , viewed as a subset of the quarterplane, might no longer be a Young diagram. Nevertheless, its profile is still well defined as:

(1.6)
$$\omega_{c\lambda}(u) = c\omega_\lambda\left(\frac{u}{c}\right) \quad \text{for } u \in \mathbb{R}.$$

1.5. *The Logan–Shepp–Vershik–Kerov limit shape.* Logan and Shepp [16] as well as Vershik and Kerov [33] independently proved a law of large numbers for the shapes of Young diagrams: as the number of boxes $n \rightarrow \infty$ tends to infinity, the scaled-down profile of a Plancherel-distributed random Young diagram converges in probability to an explicit limit curve (see Figure 4 for an illustration). The limit shape is given by the function $\Omega_* : \mathbb{R} \rightarrow [0, \infty)$ defined as:

(1.7)
$$\Omega_*(u) = \begin{cases} \frac{2}{\pi} \left[u \arcsin\left(\frac{u}{2}\right) + \sqrt{4 - u^2} \right] & \text{if } -2 \leq u \leq 2, \\ |u| & \text{otherwise.} \end{cases}$$

THEOREM 1.3 (Limit shape of Plancherel-distributed Young diagrams [16, 33]). *Let $\lambda^{(n)}$ be a random Young diagram distributed according to the Plancherel measure on Young diagrams with n boxes. Then we have convergence in the supremum norm in probability:*

$$\sup_{u \in \mathbb{R}} \left| \omega_{\frac{1}{\sqrt{n}}\lambda^{(n)}}(u) - \Omega_*(u) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

The rate of convergence is quite fast, and precise results about the magnitude of local fluctuations have been established [5].

1.6. *Position of the new box for an unspecified new entry.* In the context of Problem 1.2, let us momentarily assume that we do not know the value of the new entry w_{n+1} . The complete answer to Problem 1.2 would then be the probability distribution of the new box (1.5). This distribution coincides with that of the box containing the entry $n + 1$ in a large Plancherel-distributed random standard Young tableau. Pittel and Romik [21], Section 1.2, studied an analogous problem for a different class of random tableaux.

As expected, the probability distribution of this vector (after rescaling, and in Russian coordinates)

$$(u_n, v_n) = \left(\frac{x_n - y_n}{\sqrt{n}}, \frac{x_n + y_n}{\sqrt{n}} \right)$$

converges, as $n \rightarrow \infty$, to a certain probability measure μ supported on the limit curve Ω_* . To uniquely specify this measure, it suffices to find the limit distribution for the u -coordinates, that is, for the random variables (u_n) .

It turns out that the sequence (u_n) converges in distribution to the *semicircle measure* on the interval $[-2, 2]$ (see [13] and [24], Theorem 3.2). The density of this measure is given by

$$(1.8) \quad f_{\text{SC}}(u) = \frac{1}{2\pi} \sqrt{4 - u^2} \quad \text{for } u \in [-2, 2].$$

We denote by $F_{\text{SC}} : [-2, 2] \rightarrow [0, 1]$ the cumulative distribution function (CDF) of this semicircle law, given by

$$F_{\text{SC}}(u) = \frac{1}{2\pi} \int_{-2}^u \sqrt{4 - w^2} dw = \frac{1}{2} + \frac{u\sqrt{4 - u^2}}{4\pi} + \frac{\arcsin(\frac{u}{2})}{\pi} \quad \text{for } u \in [-2, 2].$$

1.7. *The natural parametrization of the limit curve.* The following two *RSK-trigonometric functions* are defined for $z \in [0, 1]$:

$$\begin{aligned} \text{RSKcos } z &= F_{\text{SC}}^{-1}(z), \\ \text{RSKsin } z &= \Omega_*(F_{\text{SC}}^{-1}(z)). \end{aligned}$$

The map

$$(1.9) \quad [0, 1] \ni z \mapsto (\text{RSKcos } z, \text{RSKsin } z) \in \mathbb{R}^2$$

is a two-dimensional analogue of the *quantile function* (the inverse of the CDF) in the context of the limit measure μ . This map provides a convenient parametrization of the limit curve Ω_* , as illustrated by the five black dots in Figure 4. It is a natural analogue of the parametrization of the unit circle by the angle

$$z \mapsto (\cos z, \sin z)$$

as well as the parametrization of the hyperbola by the hyperbolic angle

$$z \mapsto (\cosh z, \sinh z),$$

since in all three cases the area of the curvilinear triangle between the ray, the curve, and the y -axis (or x -axis) is proportional to the value of the parameter z .

1.8. *The first main conjecture: The fluctuations of Schensted row insertion.* To address Problem 1.2, it is most convenient to condition on the random variable w_{n+1} taking a fixed value $z \in [0, 1]$. The following result by Romik and the second named author provides a first-order asymptotic description.

THEOREM 1.4 (Asymptotic determinism of RSK insertion [24], Theorem 5.1). *Let $z \in [0, 1]$ be fixed. We denote by*

$$(1.10) \quad (u_0, v_0) = (\text{RSKcos } z, \text{RSKsin } z)$$

the corresponding point on the limit curve Ω_ via (1.9).*

Let $w_1, w_2, \dots$ be a sequence of i.i.d. random variables with the uniform distribution $U(0, 1)$ on the unit interval. We denote by

$$(x_n, y_n) = \text{Ins}(w_1, \dots, w_n; z)$$

the position of the box created by the insertion of z .

Then we have the following convergence in probability:

$$(1.11) \quad \left(\frac{x_n - y_n}{\sqrt{n}}, \frac{x_n + y_n}{\sqrt{n}} \right) \xrightarrow[n \rightarrow \infty]{P} (u_0, v_0).$$

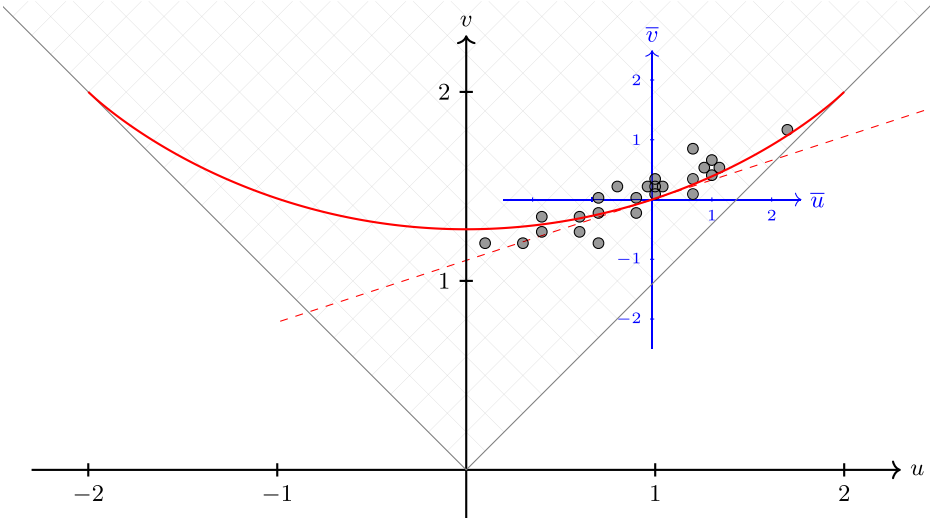


FIG. 5. Grey circles represent simulated (rescaled) positions of the new box $\text{Ins}(w_1, \dots, w_n; z)$ for an initial sequence $w_1, \dots, w_n$ of i.i.d. $U(0, 1)$ random variables with length $n = 100$ and new entry $z = 0.8$. The (u, v) coordinates of the circles correspond to the quantities in the law of large numbers (Theorem 1.4). The grid indicates the actual size of the boxes. The solid red curve depicts the Logan–Shepp–Vershik–Kerov limit shape, with the red dashed line tangent to this curve at the point corresponding to z in the natural parametrization. The coordinates of the circles in the blue $(\bar{u}, \bar{v})$ coordinate system correspond to the quantities in Conjecture 1.5.

This relationship between the value of the new entry being inserted and the location of the corresponding new box was visualized in Figure 4 by the layer tinting. Theorem 1.4 was illustrated by a Monte Carlo simulation in Figure 5. One of the results of the current paper is a generalization of this result in Theorem 4.1.

We conjecture that the ultimate answer to Problem 1.2 is provided by the following description of the fluctuations of the new box.

CONJECTURE 1.5 (The first main conjecture). *We keep the notation from Theorem 1.4. Then the scaled sequence of random points*

$$\begin{aligned} (1.12) \quad & \sqrt[4]{n} \left[\left(\frac{x_n - y_n}{\sqrt{n}}, \frac{x_n + y_n}{\sqrt{n}} \right) - (u_0, v_0) \right] \\ &= \left(\frac{x_n - y_n}{\sqrt[4]{n}} - \sqrt[4]{n} u_0, \frac{x_n + y_n}{\sqrt[4]{n}} - \sqrt[4]{n} v_0 \right) \in \mathbb{R}^2 \end{aligned}$$

converges in distribution to a centered, degenerate Gaussian distribution on the plane which is supported on the line

$$(1.13) \quad \{(u, v) : v = \Omega'_*(u_0)u\}$$

that is parallel to the tangent line to the limit curve Ω_* in u_0 .

This Gaussian distribution is uniquely determined by the limit measure for the u -coordinates

$$\sqrt[4]{n} \left[\frac{x_n - y_n}{\sqrt{n}} - u_0 \right] \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N(0, \sigma_{u_0}^2)$$

which is the centered Gaussian distribution with the variance

$$(1.14) \quad \sigma_{u_0}^2 = \frac{\pi}{3} \sqrt{4 - u_0^2}.$$

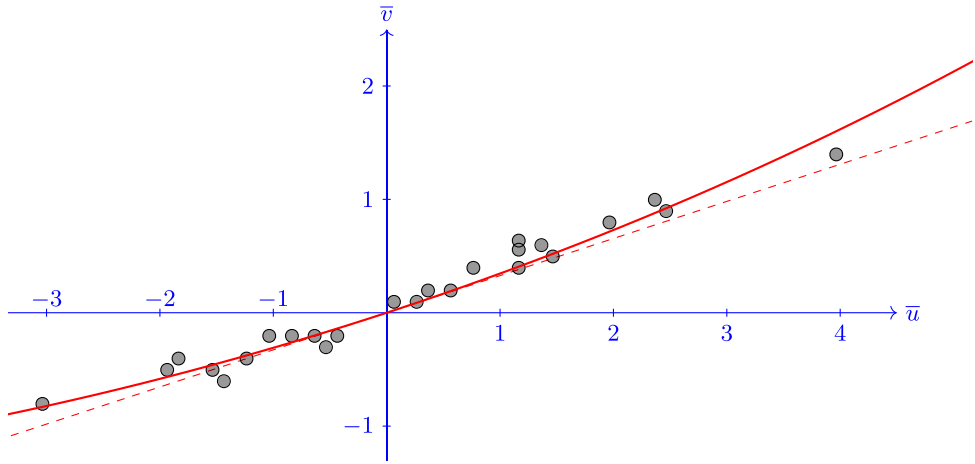


FIG. 6. Analogue of Figure 5 for sequence length $n = 10^4$ and $z = 0.8$. The image is zoomed to focus on the $(\bar{u}, \bar{v})$ coordinate system. For improved visibility, the grid showing the actual size of the boxes has been omitted.

This conjecture is visualized by Monte Carlo simulations in Figures 5 and 6. The blue coordinate system $(\bar{u}, \bar{v})$

$$(\bar{u}, \bar{v}) = \sqrt[4]{n} \left[\left(\frac{u}{\sqrt{n}}, \frac{v}{\sqrt{n}} \right) - (u_0, v_0) \right]$$

corresponds to the quantities which appear in the n th random point (1.12).

The formula for the variance (1.14) may raise questions about its conceptual interpretation. This motivates further reading, as Conjecture 1.5 appears to be a special case of the more general Theorem 4.3, though establishing a rigorous reduction would require additional arguments.

REMARK 1.6. It would be valuable to compare Conjecture 1.5 with the Monte Carlo experiments conducted by Vassiliev, Duzhin, and Kuzhin [31]. However, it is important to note that some of their results appear to be based on corrupted computer data. For instance, the plots in [31], Figure 8, lack the expected axial symmetry with respect to the $W = 0$ axis and the central symmetry around $W = 0$ and $\mu = \frac{1}{2}$. These symmetries should be present due to the inherent properties of the RSK correspondence. Specifically, replacing the numbers $z, w_1, w_2, \dots$ with $1 - z, 1 - w_1, 1 - w_2, \dots$ (using the notation from Conjecture 1.5) should result in the transposition of the corresponding RSK diagram's shape. This property is a well-established feature of the RSK algorithm and should be reflected in any accurate simulation results.

1.9. Donsker's functional central limit theorem for jeu de taquin trajectories. Given an infinite standard Young tableau T , we view it in the Russian coordinate system (see Figure 7). We define an infinite lattice path, referred to as the *jeu de taquin path* of tableau T , as follows. Starting from the corner box of the tableau, we travel in unit steps northwest or northeast. At each step, we choose the direction among the two in which the entry in the tableau is smaller. This process continues indefinitely, creating the jeu de taquin path. For an integer $n \geq 1$, we denote by $j_n = j_n(T) = (u_n, v_n) \in \mathbb{Z}^2$ the last box in the jeu de taquin path of T which contains a number $\leq n$. We refer to the sequence (j_n) as the *jeu de taquin path in the lazy parametrization* (in the Russian coordinate system). Figure 7 illustrates these concepts.

We conjecture that the following analogue of Donsker's functional central limit theorem holds true for jeu de taquin paths. This claim is a refinement of a result of Romik and the second named author [24], Theorem 5.2.

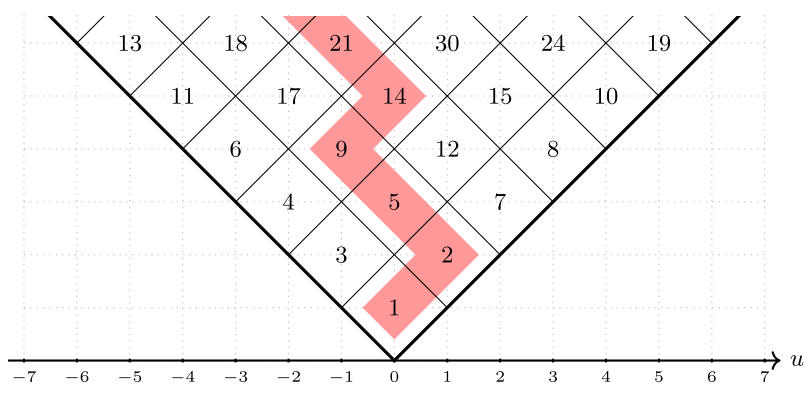


FIG. 7. A part of an infinite standard Young tableau drawn in the Russian coordinate system. The highlighted boxes form the beginning of the jeu de taquin path.

CONJECTURE 1.7 (The second main conjecture). *Let $z \in [0, 1]$ be fixed and let (u_0, v_0) given by (1.10) be the corresponding point on the limit curve. Let $w_1, w_2, \dots$ be a sequence of i.i.d. random variables with the uniform distribution $U(0, 1)$ on the unit interval. Let*

$$T = Q_\infty(1 - z, w_1, w_2, \dots)$$

be the random infinite recording tableau and let $j_1, j_2, \dots$ be the jeu de taquin path in the tableau T in the lazy parametrization.

As $c \rightarrow \infty$, the random function

$$(1.15) \quad \mathbb{R}_+ \ni t \mapsto \sqrt[4]{c} \left[\frac{1}{\sqrt{c}} j_{\lceil ct^2 \rceil} - (tu_0, tv_0) \right] \in \mathbb{R}^2$$

converges in distribution to the two-dimensional Brownian motion (U_t, V_t) that is supported on the line (1.13) and for which the covariance of the u -coordinates is given by

$$\mathbb{E} U_t U_s = \min(t, s) \sigma_{u_0}^2,$$

where the constant σ_{u_0} is given by (1.14).

A special case of Conjecture 1.5 for $z = \frac{1}{2}$ was conjectured by Wojtyniak [35] (with a slightly different form of the covariance) based on extensive Monte Carlo simulations.

Conjecture 1.7 would imply Conjecture 1.5; indeed, one of the consequences of Conjecture 1.7 is the information about the limit distribution of the random function (1.15) evaluated at the fixed time $t = 1$. Fairly standard methods allow to relate this probability distribution to the one from Conjecture 1.5, see [24], Section 5.2. Figure 8 shows numerical simulations of several jeu de taquin paths together with their asymptotes, illustrating the kind of fluctuations around a straight line that Conjecture 1.7 is meant to describe.

1.10. *Second class particles.* With the notations of Conjecture 1.7, if z is assumed to be random with the uniform distribution $U(0, 1)$, then the recording tableau T is a Plancherel-distributed infinite standard Young tableau. Thanks to Rost’s map [25], this can be seen as the *Plancherel-TASEP interacting particle system* [24], Section 7. In this context, the u -coordinate u_n of the jeu de taquin box j_n can be interpreted as the position of the *second class particle* in this interacting particle system at time n . Thus, part of Conjecture 1.7 can be reformulated as follows.

CONJECTURE 1.8. *Consider the Plancherel-TASEP interacting particle system and let u_n be the position of the second class particle at time n . There exists a random variable*

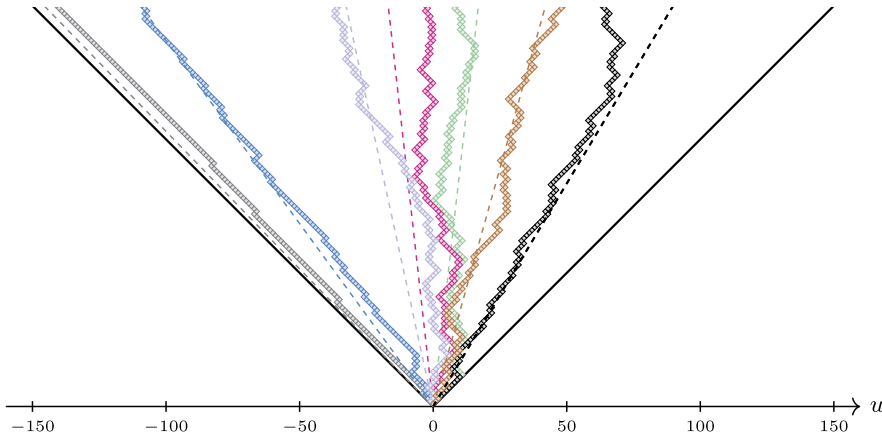


FIG. 8. Several simulated paths of jeu de taquin and (dashed lines) their asymptotes. Figure excerpted from [24].

V with the semicircle distribution μ_{SC} on the interval $[-2, 2]$ such that the random function

$$\mathbb{R}_+ \ni t \mapsto \frac{\sqrt[4]{c}}{\sigma_V} \left[\frac{u_{\lfloor ct^2 \rfloor}}{\sqrt{c}} - Vt \right]$$

converges in distribution to the standard Brownian motion $B(t)$ as $c \rightarrow \infty$.

We suspect that in the above conjecture $V = F_{\text{SC}}^{-1}(z)$ can be defined as the appropriate quantile of the semicircle distribution.

Heuristically, this conjecture suggests that asymptotically the second class particles follow a drift with the random velocity V plus a (rescaled) Brownian motion, so that the approximate equality

$$u_{\lfloor t^2 \rfloor} \approx Vt + \sigma_V B(t)$$

holds true for $t \rightarrow \infty$. In particular, the fluctuations are of order $t^{\frac{1}{2}}$. It would be interesting to explore the links of this conjecture with analogous results available for the usual TASEP model and the competition interface [7], Section 2.3, [22]. It *seems* that the corresponding fluctuations are superdiffusive, of order $t^{\frac{2}{3}}$; however, we are not aware of a definitive, rigorous treatment of this topic in the literature.

2. Main results and structure of the paper.

2.1. Poissonized tableaux: A generalization.

2.1.1. Plancherel tableau, the key example. In Section 1, we introduced Theorem 1.4 and Conjecture 1.5, which focused on the Schensted row insertion $T \leftarrow z$ for a specific random tableau $T = P(w_1, \dots, w_n)$. This tableau, generated by applying the RSK algorithm to a sequence of i.i.d. random variables with the uniform $U(0, 1)$ distribution, is referred to as a *Plancherel-distributed random Poissonized tableau with n boxes*. While significant, this is just one example of a broader class known as *Poissonized tableaux of a given shape*, which we will explore in detail.

2.1.2. *Poissonized tableaux of a given shape.* Poissonized tableaux [9] are a generalization of standard Young tableaux where entries are real numbers from the unit interval $[0, 1]$. We denote the set of Poissonized tableaux with shape λ as $\mathcal{T}^\lambda$.

Each Poissonized tableau in $\mathcal{T}^\lambda$ can be represented as a point in the unit cube $[0, 1]^n$, where n is the number of boxes in λ . The increasing row and column constraints define a set of inequalities, making $\mathcal{T}^\lambda$ identifiable with a convex polytope in $[0, 1]^n$. This polytope has positive volume, allowing us to equip it with a uniform probability measure. Thus, it makes sense to speak about a *uniformly random Poissonized tableau with shape λ* .

2.1.3. *Additional randomness of the shape.* This class becomes even broader if we allow λ itself to be a random variable. In this case, the *uniformly random Poissonized tableau with shape λ* results from a two-step sampling process: first, we sample λ with some prescribed probability distribution, and then we sample a Poissonized tableau with this shape.

As we already mentioned, a notable example within this broader class is the aforementioned random Plancherel tableau which arises as the insertion tableau corresponding to a sequence of i.i.d. $U(0, 1)$ random variables, see [18], proof of Lemma 5.1. This connection to the RSK algorithm provides a bridge between Poissonized tableaux and classical combinatorial constructions.

This framework offers significant flexibility in selecting the desired level of randomness, ranging from λ being a deterministic Young diagram to more randomized examples. This flexibility will prove advantageous in subsequent discussions. For instance, we will utilize some results from our recent work [18], which concern the setup in which λ is deterministic. The main results of the current paper address scenarios where λ is either deterministic or random.

2.1.4. *Challenges in applying results to random Plancherel tableaux.* The setup of random Plancherel tableaux in Conjecture 1.5 fits within the general framework of Theorems 5.1 and 4.3. However, an insufficient understanding of the level of randomness in the shape of λ creates a gap that prevents directly applying these theorems to prove Conjecture 1.5.

The primary challenge lies in the specific rate of convergence required by Theorems 5.1 and 4.3 for the tableau shape to converge towards the limit shape. We are uncertain whether Plancherel-distributed random Young diagrams achieve this rate, which hinders the direct application of these theorems to prove Conjecture 1.5. See Section 4.5 for more details.

2.2. *Assumptions of the main results. Challenges in describing large Young diagrams.* Describing large Young diagrams in asymptotic problems poses some challenges. In the following, we will examine two standout methods developed to address these issues.

2.2.1. *Method 1: Continual diagrams and the Russian convention.* The first method involves using *the Russian convention* (see Section 1.1.1) and the *profiles / continual diagrams* (see Sections 1.4.1 and 3.3). This approach provides an intuitive understanding of the Young diagram shape as well as its convergence rates in asymptotic problems. We used it to state the results and conjectures in Section 1.

2.2.2. *Method 2: Kerov's transition measure.* The second method utilizes Kerov's *transition measure*, a foundational concept introduced by Sergey Kerov [13, 14]. This measure represents the shape of a Young diagram as a centered probability measure on the real line, see Section 3.2. This very natural concept is directly linked to *the Plancherel growth process*, a fundamental tool for harmonic analysis on the Young graph. Additionally, the moments of the transition measure are related to the irreducible characters of the symmetric group [3]. However, the bijection between continual diagrams and centered probability measures is complex and nonlocal. Moreover, the transition measure does not provide a direct intuitive understanding of the rate of convergence for Young diagrams.

TABLE 1
Main results classified by conclusion and assumption type

Assumption: description of Young diagram λ	Conclusion	
	First-order asymptotics	Gaussian fluctuations
via continual diagrams	Theorem 4.1	Theorem 4.3
via Kerov’s transition measure	Theorem 11.1	Theorem 5.1

2.2.3. *Assumptions of the main results. Navigating between the two viewpoints.* The assumptions of our main results can be expressed using either method, as summarized in Table 1. The transition measure approach (Theorems 5.1 and 11.1) offers the most natural expression for our findings. However, recognizing that some readers may find the transition measure less intuitive, we also present results using continual diagrams (Theorems 4.1 and 4.3). This dual presentation ensures accessibility to a broader audience, accommodating diverse conceptual preferences.

The key distinctions between these methods are noteworthy:

- Theorems based on Kerov’s transition measure have weaker assumptions, potentially allowing for broader applications. These theorems are also proven directly, without any detours.
- Conversely, theorems formulated using continual diagrams require stronger, more intuitive assumptions, which can be advantageous for certain analyses, albeit with a potentially narrower range of applications. Additionally, their proofs are not direct; initially, we must reduce them to their counterparts formulated in the language of Kerov’s transition measure.

By presenting our findings through both lenses, we aim to provide a comprehensive understanding of the problem.

2.3. *Conclusions of the main results.* Table 1 summarizes the four main results of this paper, categorized by assumption type and conclusion. Among these findings, we particularly highlight Theorem 4.3 as our key contribution. The results fall into two categories:

- *First-order asymptotics (Theorems 4.1 and 11.1):* These results, analogous to the law of large numbers, demonstrate that the rescaled position of the new box converges in probability to a specific point on the plane and generalize the findings of Romik and the second named author (Theorem 1.4). These results require relatively weak assumptions about the convergence of the tableau’s shape towards a limit shape.
- *Gaussian fluctuations (Theorems 5.1 and 4.3):* Similar to the central limit theorem, these findings show that the rescaled fluctuation of the new box around the first-order approximation converges to a Gaussian distribution. These results align with Conjecture 1.5. These results have more stringent requirements, necessitating either the tableau shape or the CDF of the corresponding transition measure to converge towards the limit at a prescribed rate.

2.4. *Proof strategy for the asymptotic Gaussianity of the new box position.* The crux of our paper lies in proving Theorem 5.1, which establishes the asymptotic Gaussianity of the u -coordinate for the new box created during Schensted row insertion.

Our approach builds upon our previous work [18], centering on the *cumulative function* $u \mapsto F_T(u)$ as our primary analytical tool. This function establishes a relationship between the position of a newly created box and the value of the inserted number for a given tableau T , see Section 7.1. In our earlier analysis, we treated $F_T(u)$ as a threshold, differentiating between small inserted numbers (yielding new boxes to the left of u) and large inserted numbers (resulting in new boxes to the right of u).

A crucial insight for the current paper is the inverse relationship between the insertion function $z \mapsto \text{u-Ins}(T; z)$ and the cumulative function $u \mapsto F_T(u)$. This relationship parallels that of a probability distribution's CDF and its corresponding quantile function. Consequently, to determine the u -coordinate $\text{u-Ins}(T; z)$ of the new box for a specific inserted value z , we must comprehend the threshold value $F_T(u)$ across the entire spectrum of u -coordinates.

An additional layer of complexity arises from the randomness of tableau T (specifically, a uniformly random Poissonized tableau of fixed shape λ), rendering the corresponding cumulative function F_T random as well. To address this, we introduce the *double cumulative function* $\mathcal{F}_\lambda(u, z)$, a two-dimensional analogue of the usual CDF from the probability theory, see Section 8. For a given diagram λ , the value of $\mathcal{F}_\lambda(u, z)$ represents the probability that the plot of the cumulative function $u \mapsto F_T(u)$ lies southeast of the point (u, z) . This construct integrates information about the new box's position, the inserted number, and the associated probability, providing a comprehensive analytical framework for our study.

Fixing one argument of the double cumulative function yields interesting observations:

- For a fixed u -coordinate u_0 , the map $z \mapsto \mathcal{F}_\lambda(u_0, z)$ coincides with the cumulative distribution function of the threshold value $F_T(u_0)$. Our previous work [18] established explicit upper bounds on its cumulants, allowing us to prove the asymptotic Gaussianity of $F_T(u_0)$. Consequently, we can approximate $\mathcal{F}_\lambda(u, z)$ using the CDF of the standard normal distribution.
- For a fixed inserted number z , the map $u \mapsto 1 - \mathcal{F}_\lambda(u, z)$ is the CDF of the u -coordinate $\text{u-Ins}(z)$ of the new box created during insertion. This is precisely the random variable we need to analyze. The asymptotics of the double cumulative function found in the previous step provides the necessary information to conclude the Gaussianity of $\text{u-Ins}(z)$ which concludes the proof.

2.5. Content of the paper. This paper is structured as follows:

In Section 3, we introduce the necessary vocabulary for our results, particularly focusing on *Kerov's transition measure* of a Young diagram, which will play an important role.

Section 4 presents the two main results (Theorems 4.1 and 4.3) with the assumptions expressed in the language of continual diagrams.

In Section 5, we introduce Theorem 5.1, which reformulates Theorem 4.3 under weaker assumptions and serves as an intermediate step in its proof. The two theorems differ in their approach to the assumptions:

- Theorem 4.3 is expressed in terms of the shape of the Young diagrams, providing a more concrete geometric interpretation.
- Theorem 5.1 is formulated in terms of the transition measures of the diagrams, offering a more general probabilistic perspective.

This distinction allows Theorem 5.1 to serve as a bridge between the geometric and probabilistic aspects of the problem. It facilitates the proof of Theorem 4.3 while potentially extending its applicability to a broader class of random structures.

Section 6 presents the proof that the main result in the second formulation (Theorem 5.1) implies the main result in the first formulation (Theorem 4.3).

In Section 7, we revisit the *cumulative function* $u \mapsto F_T(u)$, a concept we first introduced in our recent work [18]. In Lemma 7.1, we summarize the key findings from that paper. While the original work presented a wealth of results, Lemma 7.1 focuses specifically on the aspects that are essential for our current context.

In Section 8, we introduce our primary tool: the *double cumulative function* $\mathcal{F}_\lambda$. This function provides comprehensive insights into the relationship between the number inserted

into a random Poissonized tableau of shape λ , the location of the new box, and the associated probabilities.

In Section 9, we present and prove Proposition 9.1. This result offers a first-order approximation for the asymptotic behavior of the double cumulative function, serving as a pivotal step toward establishing the determinism of the Schensted insertion, as detailed in Theorem 4.1.

In Section 10, we establish Theorem 10.1, which elucidates the fine asymptotic behavior of the double cumulative function $\mathcal{F}_\lambda$. This result is intricately connected to the CDF of the standard normal distribution and plays a crucial role in the proof of Theorem 5.1.

Finally, Section 11 culminates in the demonstration of our principal findings: Theorem 4.1 and Theorem 5.1. Here, we leverage the tools developed in the preceding sections to achieve our conclusions.

3. Preliminaries: Transition measure and continual diagrams.

3.1. *Plancherel growth process.* Let $w_1, w_2, \dots$ be a sequence of i.i.d. random variables with the uniform distribution $U(0, 1)$. We define:

$$\lambda^{(n)} := \text{RSK}(w_1, \dots, w_n).$$

The resulting random sequence of Young diagrams:

$$(3.1) \quad \emptyset = \lambda^{(0)} \nearrow \lambda^{(1)} \nearrow \lambda^{(2)} \nearrow \dots$$

is called the *Plancherel growth process*. This process forms a Markov chain with specific transition probabilities which will be discussed below. The construction described above is a special case of the more general setup discussed earlier in Section 1.2.2.

3.2. *Transition measure of a Young diagram.* For a given Young diagram λ with n boxes we denote by $\mathfrak{x}_0 < \dots < \mathfrak{x}_\mathbb{L}$ the u -coordinates of its concave corners and by $\mathfrak{y}_1 < \dots < \mathfrak{y}_\mathbb{L}$ the u -coordinates of its convex corners, see Figure 9. The *Cauchy transform* of λ is defined as the rational function (see [13] and [14], Chapter 4, Section 1):

$$(3.2) \quad \mathbf{G}_\lambda(z) = \frac{(z - \mathfrak{y}_1) \cdots (z - \mathfrak{y}_\mathbb{L})}{(z - \mathfrak{x}_0) \cdots (z - \mathfrak{x}_\mathbb{L})}.$$

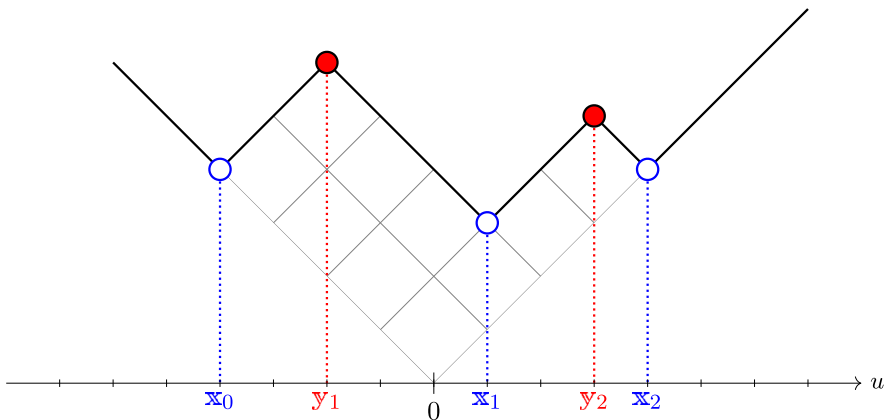


FIG. 9. Concave corners (empty, blue) and convex corners (filled, red) of a Young diagram $(4, 2, 2, 2)$ and their u -coordinates.

Note that in the work of Kerov this function is called *the generating function* of λ . The Cauchy transform can be written uniquely as a sum of simple fractions:

$$\mathbf{G}_\lambda(z) = \sum_{0 \leq i \leq \mathbb{L}} \frac{p_i}{z - \mathbb{x}_i}$$

with the coefficients $p_0, \dots, p_{\mathbb{L}} > 0$ such that $p_0 + \dots + p_{\mathbb{L}} = 1$. We define *the transition measure* of λ as the discrete measure

$$\mu_\lambda = p_0 \delta_{\mathbb{x}_0} + \dots + p_{\mathbb{L}} \delta_{\mathbb{x}_{\mathbb{L}}};$$

in this way

$$\mathbf{G}_\lambda(z) = \int_{\mathbb{R}} \frac{1}{z - x} \, \mathrm{d}\mu_\lambda(x)$$

is indeed the Cauchy transform of μ_λ .

Kerov proved that the transition probabilities of the Markov chain (3.1) are encoded by the transition measure, see [13] and [14], Chapter 4, Section 1. More specifically, the conditional probability that the new box will have the u -coordinate equal to $\mathbb{x}_i$ is given by the corresponding atom of the transition measure

(3.3)
$$\mathbb{P}[u(\lambda^{(n+1)})/\lambda^{(n)} = \mathbb{x}_i | \lambda^{(n)} = \lambda] = p_i.$$

We denote by

(3.4)
$$K_\lambda(u) = \mu_\lambda((-\infty, u]) \quad \text{for } u \in \mathbb{R}$$

the CDF of μ_λ . To keep the notation short, we will call it *the cumulative function* of λ .

3.3. Continual diagrams. We say that $\omega: \mathbb{R} \rightarrow \mathbb{R}_+$ is a *continual diagram* (see [13] and [14], Chapter 4, Section 1) if the following two conditions are satisfied:

- $|\omega(u_1) - \omega(u_2)| \leq |u_1 - u_2|$ holds true for any $u_1, u_2 \in \mathbb{R}$,
- $\omega(u) = |u|$ for sufficiently large $|u|$.

In the literature, such continual diagrams are also referred to as *generalized diagrams*, *continuous diagrams*, or simply *diagrams*.

Notably, an important class of examples arises from the profiles ω_λ associated with the usual Young diagrams. Another important class of examples arises from the *rescaled* profiles

(3.5)
$$\omega_{\frac{1}{\sqrt{n}}\lambda^{(n)}},$$

where $\lambda^{(n)}$ is a Young diagram with n boxes, see Section 1.4.1. The motivation for the scaling factor in (3.5) lies in the observation that if each box of $\lambda^{(n)}$ is drawn as a square with the side $\frac{1}{\sqrt{n}}$ then the total area of the boxes is equal to 1, irrespective of the Young diagram size n . In the current paper, the continual diagrams will serve as limit objects toward which a sequence of (rescaled) Young diagrams of the form (3.5) can converge.

3.4. Transition measure for continual diagrams. Let $\omega = \omega_\lambda$ be a profile of a Young diagram λ ; we will use the notations from Figure 9. The second derivative ω'' is well defined as a Schwartz distribution and can be identified with a signed measure on the real line. The positive part of this measure is supported in the concave corners $\mathbb{x}_0, \dots, \mathbb{x}_{\mathbb{L}}$ while the negative part of this measure is supported in the convex corners $\mathbb{y}_1, \dots, \mathbb{y}_{\mathbb{L}}$, with each atom having

equal weight 2. An application of logarithm transforms the product on the right-hand side of (3.2) into a sum. It was observed by Kerov that it can be conveniently written in the form

$$(3.6) \quad \begin{aligned} \log[z\mathbf{G}_\omega(z)] &= - \int_{-\infty}^{\infty} \log(z-w) \left(\frac{\omega(w) - |w|}{2} \right)'' dw \\ &= - \int_{-\infty}^{\infty} \frac{1}{z-w} \left(\frac{\omega(w) - |w|}{2} \right)' dw, \end{aligned}$$

where z is a complex variable.

We can drop the assumption that $\omega = \omega_\lambda$ is a profile of a Young diagram and use (3.6) to define the Cauchy transform $\mathbf{G}_\omega$ for an arbitrary continual diagram ω . The transition measure μ_ω is then defined from the Cauchy transform $\mathbf{G}_\omega$ via Stieltjes inversion formula. The cumulative function K_ω of a continual diagram ω is defined analogously as in (3.4) as the CDF of the corresponding transition measure

$$K_\omega(u) = \mu_\omega((-\infty, u]) \quad \text{for } u \in \mathbb{R}.$$

3.5. Transition measure as a homeomorphism. For $C > 0$, we say that a continual diagram ω is supported on the interval $[-C, C]$ if $\omega(u) = |u|$ for each real number u such that $|u| \geq C$. For example, for a rescaled Young diagram $\omega = \omega_{c\lambda}$ this holds true if and only if the rescaled first row, as well as rescaled first column are shorter than C :

$$c\lambda_1 \leq C, \quad c\lambda'_1 \leq C.$$

The following result is due to Kerov (see [13] and [14], Chapter 4, Section 1).

PROPOSITION 3.1. *For each constant $C > 0$, the map $\omega \mapsto \mu_\omega$, which associates a continual diagram with its transition measure, is a homeomorphism between the following two topological spaces:*

- *The set of continual diagrams supported on $[-C, C]$, with the topology given by the supremum distance.*
- *The set of centered probability measures supported on the interval $[-C, C]$, with the weak topology of probability measures.*

This convenient homeomorphism allows us to change the perspective and to parameterize the set of (continual) diagrams not by their profiles, but via their transition measure. This viewpoint has proven to be very fruitful for problems in asymptotic representation theory; see [3, 6, 12, 26] for some examples.

3.6. The metric d_{XY} on the set of continual diagrams. While the supremum distance provides a useful topology for the set of continual diagrams (see, e.g., Proposition 3.1), some problems benefit from a stronger topology. For this purpose, we introduce the metric d_{XY} , recently proposed by the second named author [27]. This metric offers a more nuanced approach, essentially measuring the supremum distance along both the x -coordinate and the y -coordinate separately. Heuristically, we can think of d_{XY} as capturing the maximum discrepancy between two diagrams when viewed from both horizontal and vertical perspectives. This approach provides a finer-grained comparison between continual diagrams, potentially revealing subtle differences that the standard supremum distance might overlook. The formal definition of the metric d_{XY} is detailed below.

For the purposes of this section, we will use the following notational shorthand. For $x, y \geq 0$ and a continual diagram $\omega: \mathbb{R} \rightarrow \mathbb{R}_+$ we will write $(x, y) \in \omega$ if the point (x, y) belongs to the graph of ω drawn in the French coordinate system or, equivalently, if

$$\omega(x - y) = x + y,$$

cf. (1.1). Additionally, for $x \geq 0$ we denote by

$$\Pi_Y^\omega(x) = \{y : (x, y) \in \omega\} \subseteq \mathbb{R}_+$$

the projection on the y -axis of the intersection of the plot of ω with the line having a specified x -coordinate. One can verify that $\Pi_Y^\omega(x)$ is a nonempty closed set. Similarly, for $y \geq 0$ we denote by

$$\Pi_X^\omega(y) = \{x : (x, y) \in \omega\} \subseteq \mathbb{R}_+$$

the projection on the x -axis.

If ω_1, ω_2 are continual diagrams, we define their y -distance

$$d_Y(\omega_1, \omega_2) = \sup_{x \geq 0} d_H(\Pi_Y^{\omega_1}(x), \Pi_Y^{\omega_2}(x))$$

as the supremum (over all choices of the x -coordinate) of the *Hausdorff distance* d_H between their y -projections. Heuristically, the metric d_Y is a way to quantify the discrepancy between continual diagrams along the y -coordinate. In a similar manner, we define the x -distance between continual diagrams as

$$d_X(\omega_1, \omega_2) = \sup_{y \geq 0} d_H(\Pi_X^{\omega_1}(y), \Pi_X^{\omega_2}(y)).$$

REMARK 3.2. Note that in the special case when $\omega_1 = \omega_\lambda$ and $\omega_2 = \omega_\mu$ are profiles of the Young diagrams $\lambda = (\lambda_1, \lambda_2, \dots)$ and $\mu = (\mu_1, \mu_2, \dots)$ this distance has a simple interpretation

$$d_X(\omega_\lambda, \omega_\mu) = \max_i |\lambda_i - \mu_i|$$

as the supremum distance between the parts of the partitions. Similarly, the y -distance $d_Y(\omega_\lambda, \omega_\mu)$ is the supremum distance for the parts of the conjugate partitions λ' and μ' .

Finally, we define the xy -distance between ω_1 and ω_2 as the maximum of the x - and the y -distance:

$$d_{XY}(\omega_1, \omega_2) = \max(d_X(\omega_1, \omega_2), d_Y(\omega_1, \omega_2)).$$

4. The general form of RSK insertion fluctuations.

4.1. *Asymptotic determinism of Schensted insertion.* We start with the following generalization of the aforementioned result of Romik and the second named author (Theorem 1.4).

THEOREM 4.1 (Asymptotic determinism of Schensted insertion). Assume that

- For each $n \geq 1$ we are given a random Young diagram $\lambda^{(n)}$; we denote by

$$\omega_n = \omega_{\frac{1}{\sqrt{n}}\lambda^{(n)}}$$

its rescaled profile.

- There is a continual diagram Ω such that the convergence in probability

$$\omega_n(u) \xrightarrow[n \rightarrow \infty]{P} \Omega(u)$$

holds true for each $u \in \mathbb{R}$.

- There is a constant $C > 0$ such that the probability of the event that the rescaled diagram ω_n is supported on the interval $[-C, C]$ converges to 1, as $n \rightarrow \infty$.
- A real number $0 < z < 1$ is such that the quantile function K_Ω^{-1} of the transition measure μ_Ω is continuous in z .

Let $T^{(n)}$ be a uniformly random Poissonized tableau of shape $\lambda^{(n)}$. Then the rescaled u -coordinate of the new box created by the insertion of z

$$\frac{1}{\sqrt{n}} \text{u-Ins}(T^{(n)}; z) \xrightarrow[n \rightarrow \infty]{P} K_\Omega^{-1}(z)$$

converges in probability to the corresponding quantile of the transition measure μ_Ω .

In particular, Theorem 4.1 combined with the classic result of Logan, Shepp, Vershik and Kerov (Theorem 1.3, “the asymptotic shape of random Plancherel-distributed Young diagrams”) provides a conceptually new proof of the result by Romik and the second named author (Theorem 1.4, “the asymptotic determinism of RSK insertion for i.i.d. random variables”).

The proof of Theorem 4.1 is postponed to Section 11.2.2.

REMARK 4.2. To avoid potential confusion, we clarify that throughout this paper—in Theorem 4.1 above, as well as in Theorems 4.3, 5.1, and 11.1 in other sections—the parameter n is a “size” variable that grows roughly linearly with the number of boxes. While in some specific situations n may equal the actual number of boxes, these theorems do not require the Young diagram $\lambda^{(n)}$ to contain exactly n boxes, nor do they assume that the ratio $\frac{|\lambda^{(n)}|}{n}$ converges to 1.

4.2. *The interaction energy.* If μ is a probability measure on the real line and $u_0 \in \mathbb{R}$, we define

$$(4.1) \quad \mathcal{E}_\mu(u_0) = \iint_{\{(z_1, z_2): z_1 \leq u_0 < z_2\}} \frac{1}{z_2 - z_1} d\mu(z_1) d\mu(z_2)$$

whenever this double integral is finite. This quantity will play an important role in the statement of our main result.

Let us interpret μ as the distribution of the electrostatic charge along a one-dimensional rod. If we split this rod at the point u_0 , the two parts will act on one another with the electrostatic force. The integral (4.1) can be interpreted as the *interaction energy* between these two parts.

We now consider an alternative universe in which the space is two-dimensional so that the electrostatic force decays as the inverse of the distance between the charges. It follows that the double integral $\mathcal{E}_\mu(u_0)$ is equal to the *electrostatic force* between the two parts of the rod.

If we separate the two parts of the rod by an additional distance l , the aforementioned electrostatic force is the derivative of the total energy of the system with respect to the variable l . It is worth pointing out that in our context *the total energy* should be understood as the logarithmic energy which is ubiquitous in random matrix theory [2], Voiculescu’s free entropy [34], as well as in asymptotic representation theory [23], page 46, quantity $Q(h)$. For this reason, we suspect that the interaction energy $\mathcal{E}_\mu(u_0)$ can be found also in the context of random matrix theory.

Suppose that the support of the measure μ is contained in some interval $[u_{\min}, u_{\max}]$. We choose an arbitrary real number $u_{-\infty}$ such that $u_{-\infty} < u_{\min}$ and consider a contour $\mathcal{C}$ on the complex plane shown in Figure 10. This contour was chosen in such a way that it crosses the real line in two points, namely, $u_{-\infty}$ and u_0 .

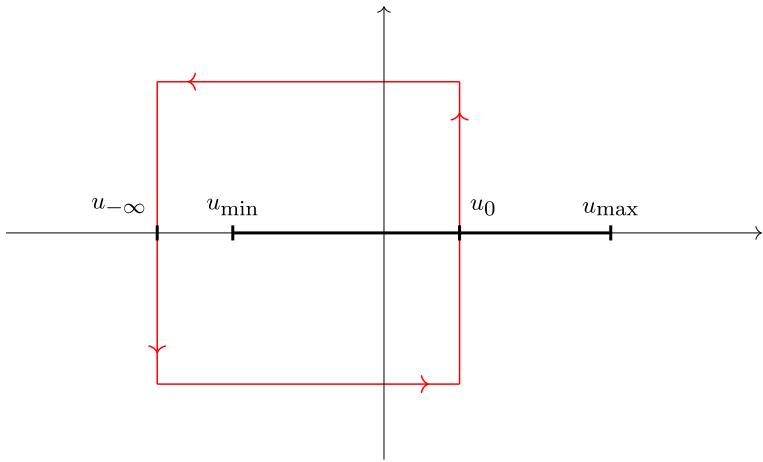


FIG. 10. The contour C on the complex plane.

In the special case when the support of μ is a finite set and u_0 does not belong to the support of μ we may apply Cauchy’s residue theorem and the interaction energy is equal to the contour integral

(4.2)
$$\mathcal{E}_{\mu\Omega}(u_0) = -\frac{1}{4\pi i} \oint_C [\mathbf{G}_\mu(z)]^2 \, dz,$$

where

$$\mathbf{G}_\mu(z) = \int_{\mathbb{R}} \frac{1}{z - x} \, d\mu(x)$$

is the Cauchy transform of μ . The formula (4.2) remains true for general measures μ as long as the Cauchy transform is sufficiently regular in the neighborhood of u_0 , however the justification of its validity is more technically involved. In practical applications the formula (4.2) is more convenient than (4.1).

4.3. The main theorem: General form of the fluctuations.

THEOREM 4.3 (The main theorem; fluctuations of Schensted insertion are Gaussian). Assume that:

(A1) For each $n \geq 1$ we are given a random Young diagram $\lambda^{(n)}$; we denote by

(4.3)
$$\omega_n = \omega_{\frac{1}{\sqrt{n}}\lambda^{(n)}}$$

its rescaled profile.

(A2) There is a continual diagram Ω which is the limit of ω_n (with respect to the metric d_{XY}), furthermore the rate of convergence is such that

$$\sqrt[4]{n} \log n \cdot d_{XY}(\omega_n, \Omega) \xrightarrow[n \rightarrow \infty]{P} 0.$$

(A3) There are constants $a < u_0 < b$ such that the transition measure μ_Ω restricted to the interval $[a, b]$ is absolutely continuous, with a density given by a continuous, strictly positive function $f > 0$. We denote

$$z_0 = K_\Omega(u_0), \quad \mathcal{E}_0 = \mathcal{E}_{\mu_\Omega}(u_0), \quad f_0 = f(u_0).$$

(A4) The restriction $\Omega: [a, b] \rightarrow \mathbb{R}$ of the continual diagram Ω to the interval $[a, b]$ is a contraction (i.e., it is a K -Lipschitz function for some constant $K < 1$).

Let $T^{(n)}$ be a uniformly random Poissonized tableau of shape $\lambda^{(n)}$. Then the fluctuations of the newly created box around the first-order approximation

$$(4.4) \quad \sqrt[4]{n} \left[\frac{\text{u-Ins}(T^{(n)}; z_0)}{\sqrt{n}} - u_0 \right] \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N\left(0, \frac{\varepsilon_0}{f_0^2}\right)$$

converge to a centered Gaussian measure with the variance $\frac{\varepsilon_0}{f_0^2}$.

The proof is quite involved, the first step is to reduce this theorem to another result (Theorem 5.1). We postpone the details to Section 6.

4.4. Explicit examples for Theorem 4.3. The objective of this section is to provide tangible examples that demonstrate the applicability of Theorem 4.3. While many theoretical examples exist, we specifically seek to move beyond abstract objects by focusing on cases where the essential components—the sequence of Young diagrams, the limit continual diagram, its transition measure, and the interaction energy—are all computationally tractable.

4.4.1. Example: Staircase tableaux. In Theorem 4.3, we will pass to the subsequence of triangular numbers

$$n = n_N = 1 + \cdots + N = \frac{N(N+1)}{2}.$$

We consider the sequence of deterministic Young diagrams $(\lambda^{(n_N)})$, where

$$(4.5) \quad \lambda^{(n_N)} = (N, N-1, \dots, 3, 2, 1)$$

is a *staircase Young diagram* with n_N boxes. We define the rescaled diagram ω_n via (4.3).

Let $\Omega: \mathbb{R} \rightarrow \mathbb{R}_+$ be the *isosceles triangle diagram* given by

$$\Omega(u) = \begin{cases} \sqrt{2} & \text{for } |u| \leq \sqrt{2}, \\ |u| & \text{otherwise.} \end{cases}$$

In the following, we will verify that the assumptions of Theorem 4.3 are satisfied for this configuration and we will calculate the missing data: the density of the transition measure and its interaction energy.

The rate of convergence of the rescaled diagrams towards the limit is very fast:

$$d_{XY}(\omega_n, \Omega) = O\left(\frac{1}{N}\right) = O\left(\frac{1}{\sqrt{n}}\right) \ll o\left(\frac{1}{\sqrt[4]{n} \log n}\right),$$

which ensures that Assumption (A2) in Theorem 4.3 holds. The example of staircase Young diagrams and the rate of convergence of their transition measures μ_{ω_n} to the limit μ_Ω were studied in more detail in the recent work of the second named author [27].

Since the second derivative Ω'' is a Schwartz distribution which can be identified with a very simple measure supported in $\pm\sqrt{2}$, a straightforward calculation based on (3.6) shows that the corresponding Cauchy transform is given by

$$(4.6) \quad \mathbf{G}_{\text{AS}}(z) = \frac{1}{\sqrt{z^2 - 2}} \quad \text{for } z \in \mathbb{C} \setminus I,$$

where $I = [-\sqrt{2}, \sqrt{2}]$. It is easy to verify (e.g., by the Stieltjes inversion formula) that the corresponding transition measure $\mu_\Omega = \mu_{\text{AS}}$ is the arcsine distribution which is supported on the interval I , with the density

$$f_{\text{AS}}(u) = \frac{1}{\pi \sqrt{2 - u^2}} \quad \text{for } u \in I$$

and the CDF

$$K_{\Omega}(u) = F_{\text{AS}}(u) = \frac{1}{2} + \frac{1}{\pi} \arcsin \frac{u}{\sqrt{2}} \quad \text{for } u \in I.$$

The formula (4.2) combined with (4.6) implies that if $-\sqrt{2} < u_0 < \sqrt{2}$ then the value of $(-2)\mathcal{E}_{\text{AS}}(u_0)$ is equal to the residue of the rational function $z \mapsto \frac{1}{z^2-2}$ in $z = -\sqrt{2}$. In this way, we get that the interaction energy of the arcsine measure is given by the piecewise constant function

(4.7)
$$\mathcal{E}_{\text{AS}}(u) = \begin{cases} 0 & \text{if } u < -\sqrt{2}, \\ \frac{1}{4\sqrt{2}} & \text{if } -\sqrt{2} < u < \sqrt{2}, \\ 0 & \text{if } u > \sqrt{2}. \end{cases}$$

COROLLARY 4.4 (Schensted insertion into staircase tableaux). *Let $0 < z_0 < 1$ be fixed and let*

$$u_0 = F_{\text{AS}}^{-1}(z_0) = -\sqrt{2} \cos \pi z_0$$

be the corresponding quantile of the arcsine distribution. For each integer n of the form $n = n_N$ let $T^{(n)}$ be a uniformly random Poissonized tableau of the staircase shape $\lambda^{(n_N)}$.

Then the sequence of random variables

$$\sqrt[4]{n} \left[\frac{1}{\sqrt{n}} \text{u-Ins}(T^{(n)}; z_0) - u_0 \right] \xrightarrow{\mathcal{D}} N(0, \sigma_{\text{AS}}^2(u_0))$$

converges in distribution to the centered Gaussian measure with the variance

$$\sigma_{\text{AS}}^2(u_0) = \frac{\pi^2}{4\sqrt{2}}(2 - u_0^2).$$

PROOF. We will verify that the assumptions of Theorem 4.3 are satisfied. The relationship between u_0 and z_0 is precisely as required in Assumption (A3). Assumption (A2) was already verified above.

For any numbers a and b with $-\sqrt{2} < a < u_0 < b < \sqrt{2}$, the density f_{AS} restricted to $[a, b]$ is continuous and strictly positive. Moreover, the restriction of Ω to $[a, b]$ is 0-Lipschitz (hence a contraction), satisfying Assumption (A4).

The result follows from Theorem 4.3, with the variance computed from the interaction energy (4.7) and the density of the arcsine distribution. $\square$

We will revisit this example in Section 8.6 and Section 10.3.

4.4.2. *Almost isosceles triangle tableaux.* We revisit and generalize the above example by defining the auxiliary functions

$$\begin{aligned} Y(x) &= \max(\sqrt{2} - x, 0), \\ X(y) &= \max(\sqrt{2} - y, 0) \end{aligned}$$

which provide the parametrization of the continual diagram Ω in the French coordinate system (x, y) and its transpose (y, x) , respectively.

Unlike the previous case, there is no need to restrict to a subsequence. Consider any sequence of (deterministic or random) Young diagrams $(\lambda^{(n)})$ such that the rescaled rows converge to the shape specified by the isosceles triangle diagram at the specific rate:

$$(4.8) \quad \max_{i \geq 1} \frac{\log n}{\sqrt[4]{n}} \left| \lambda_i^{(n)} - \sqrt{n} X \left(\frac{i}{\sqrt{n}} \right) \right| \xrightarrow[n \rightarrow \infty]{P} 0,$$

and similarly for the columns:

$$(4.9) \quad \max_{i \geq 1} \frac{\log n}{\sqrt[4]{n}} \left| (\lambda^{(n)})'_i - \sqrt{n} Y \left(\frac{i}{\sqrt{n}} \right) \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

It is straightforward to verify that these assumptions ensure condition (A2) is satisfied, making Theorem 4.3 applicable. Fortunately, all calculations for the continual isosceles triangle diagram Ω —including the transition measure, its CDF, and the interaction energy—were already completed in Section 4.4.1 and remain valid here.

COROLLARY 4.5 (Schensted insertion into almost isosceles triangle tableaux). *Under the above assumptions, Corollary 4.4 remains valid when $T^{(n)}$ is a uniformly random Poissonized tableau of shape $\lambda^{(n)}$.*

This result demonstrates the power of Theorem 4.3: unlike the staircase case, we do not require Young diagrams to have a very specific shape. The corollary applies to any diagram sequence satisfying the convergence conditions above.

For comparison, staircase diagrams enjoy explicit transition measure formulas [27], making results like Theorem 5.1 directly applicable. For the broader class of diagrams studied here, such explicit calculations are not available—Theorem 5.1 cannot be applied directly—highlighting the significance of Theorem 4.3.

4.4.3. More general almost triangle tableaux. We extend the example from Section 4.4.2 even more. We fix real parameters a and b with $|a| < 1$ and $b > 0$, and consider the continual triangle diagram

$$\Omega(u) = \max(au + b, |u|).$$

Note that $a = 0$ and $b = \sqrt{2}$ recovers the continual isosceles triangle diagram from Section 4.4.1. In what follows, we compute the Cauchy transform, transition measure, CDF, and interaction energy for this continual diagram. The calculations follow those in Section 4.4.1 closely, though the results are less elegant. Since the methods are analogous, we provide abbreviated explanations.

4.4.3.1. Critical points and Cauchy transform. The critical points of the triangle diagram Ω , which delineate the different linear segments, are

$$u_1 = -\frac{b}{1+a}, \quad u_2 = \frac{b}{1-a}.$$

The Cauchy transform of Ω takes the form

$$\mathbf{G}(z) = (z - u_1)^{-\frac{1+a}{2}} (z - u_2)^{-\frac{1-a}{2}} \quad \text{for } z \in \mathbb{C} \setminus I,$$

where $I = [u_1, u_2]$.

4.4.3.2. *Transition measure and CDF.* The corresponding transition measure μ_Ω is a scaled and centered beta distribution with shape parameters

$$\alpha = \frac{1-a}{2}, \qquad \beta = \frac{1+a}{2},$$

supported on the interval I . Its probability density function is given by

$$f(u) = \frac{1}{\pi} \cos \frac{a\pi}{2} (u-u_1)^{-\frac{1+a}{2}} (u_2-u)^{-\frac{1-a}{2}} \quad \text{for } u \in I.$$

The corresponding cumulative distribution function K_Ω can be expressed explicitly in terms of the regularized incomplete beta function:

$$K_\Omega(u) = I_{\frac{u-u_1}{u_2-u_1}}\left(\frac{1-a}{2}, \frac{1+a}{2}\right) \quad \text{for } u \in I.$$

4.4.3.3. *Interaction energy.* Computing the interaction energy via (4.2) is more challenging since it involves a contour integral of the function

$$[\mathbf{G}(z)]^2 = (z-u_1)^{-1-a} (z-u_2)^{-1+a} \quad \text{for } z \in \mathbb{C} \setminus I,$$

which is not analytic on the interval I except when $a = 0$.

For $a < 0$, we choose the contour $\mathcal{C}$ from Figure 10 as a thin rectangle very close to the real axis, ultimately taking the limit as the rectangle’s height approaches zero. The interaction energy evaluates to

$$\mathcal{E}(u) = -\frac{1}{2\pi} \sin \pi a \cdot \int_{u_1}^u (z-u_1)^{-1-a} (u_2-z)^{-1+a} \, dz \quad \text{for } u \in I.$$

Using the substitution $t = \frac{z-u_1}{u_2-u_1}$, this can be expressed as

$$\mathcal{E}(u) = -\frac{1}{2\pi} \sin \pi a \cdot \frac{1}{u_2-u_1} \cdot B\left(\frac{u-u_1}{u_2-u_1}; -a, a\right),$$

where $B(x; p, q)$ denotes the incomplete beta function. Note that this expression involves a negative second parameter $a < 0$, which requires careful interpretation beyond the standard definition.

For $a > 0$, this derivation fails due to the excessive singularity at u_1 . Instead, one must use a contour that extends rightward from u_0 rather than leftward. We omit these details here.

4.4.3.4. *Application to random almost triangle Young tableaux.* Let $(\lambda^{(n)})$ be a random or deterministic sequence satisfying the convergence conditions (4.8), (4.9) from Section 4.4.2, where the auxiliary functions are adapted to parameterize the triangle diagram Ω :

$$\begin{aligned} X(y) &= \max\left(\frac{b-(1+a)y}{1-a}, 0\right), \\ Y(x) &= \max\left(\frac{b-(1-a)x}{1+a}, 0\right). \end{aligned}$$

For a concrete example, one may take $\lambda^{(n)}$ to be the deterministic diagram with rows given by

$$\lambda_i^{(n)} = \left\lfloor \sqrt{n} X\left(\frac{i}{\sqrt{n}}\right) \right\rfloor,$$

which provides the most straightforward discretization of the diagram Ω .

Theorem 4.3 is applicable in this setting and gives the following concrete statement.

COROLLARY 4.6 (Schensted insertion into almost triangle tableaux). Assume $-1 < a < 0$ and let $(\lambda^{(n)})$ be a sequence satisfying the above assumptions. Let $u_0 \in (u_1, u_2)$ and define

$$z_0 = I_{\frac{u_0 - u_1}{u_2 - u_1}} \left(\frac{1 - a}{2}, \frac{1 + a}{2} \right) \in (0, 1).$$

Let $T^{(n)}$ be a uniformly random Poissonized tableau of shape $\lambda^{(n)}$. Then the fluctuations of the newly created box around the first-order approximation satisfy

$$\sqrt[4]{n} \left[\frac{\mathbf{u}\text{-Ins}(T^{(n)}; z_0)}{\sqrt{n}} - u_0 \right] \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N(0, \sigma^2),$$

where the variance is given by

$$\sigma^2 = -\pi \sin \frac{\pi a}{2} \cdot (u_0 - u_1)^{1+a} (u_2 - u_0)^{1-a} \cdot \frac{1}{u_2 - u_1} \cdot B \left(\frac{u_0 - u_1}{u_2 - u_1}; -a, a \right).$$

4.4.4. *Where to look for more concrete examples?* Several additional natural candidates for the continual diagram Ω arise for which explicit calculations are feasible, allowing Theorem 4.3 to take concrete form.

The most immediate example is the Logan–Shepp–Vershik–Kerov limit shape Ω_* from Section 1.5. For sequences of Young diagrams converging to this continual diagram at the appropriate rate, our framework applies directly. Moreover, all necessary quantities—the density of the measure, CDF, and interaction energy—are explicitly known. This case will be developed in detail later, in Section 4.5.3. While that section primarily addresses Conjecture 1.5 for Plancherel random diagrams, the same methods apply to any sequence satisfying Assumption (A2), yielding a concrete realization of our results.

Another promising family consists of the continual diagrams introduced by Biane [4] as limit shapes arising from the RSK correspondence applied to random sequences of letters sampled from a finite alphabet. These diagrams have explicit parametrizations and their transition measures are explicitly given by free Poisson distributions, making them particularly tractable for concrete calculations with a comparable level of difficulty as above.

More generally, any continual diagram with an explicitly known transition measure that avoids exotic special functions would serve as a suitable test case for our methods.

4.5. Towards Conjecture 1.5, part 1.

4.5.1. *General strategy.* One might attempt to prove Conjecture 1.5 by applying Theorem 4.3 with $\lambda^{(n)}$ being a Plancherel-distributed random Young diagram with n boxes and $\Omega = \Omega_*$ being the Logan–Shepp–Vershik–Kerov limit shape from Section 1.5.

4.5.2. *Problematic Assumption (A2).* Regretfully, we do not know whether Assumption (A2) is satisfied in this context. Some naive Monte Carlo simulations reveal that for diagram sizes n in the range 10^2 to 10^6 , spanning several orders of magnitude, the mean distance

$$(4.10) \quad \mathbb{E} d_{XY}(\omega_n, \Omega_*) \approx C n^\alpha$$

between the rescaled Plancherel-distributed random continual diagram ω_n and the limit shape Ω_* can be well approximated by a power law with the exponent

$$(4.11) \quad \alpha \approx -0.32$$

and the multiplicative constant $C \approx 2$. For this reason, we dare to state the following conjecture which is essentially a reformulation of the condition (A2).

CONJECTURE 4.7. *There exists an exponent $\alpha \leq -\frac{1}{4}$ with the property that if $\omega_n = \omega \frac{1}{\sqrt{n}} \lambda^{(n)}$ is the rescaled profile of a Plancherel-distributed random Young diagram $\lambda^{(n)}$ with n boxes then*

$$n^{-\alpha} d_{XY}(\omega_n, \Omega_*) \xrightarrow[n \rightarrow \infty]{P} 0.$$

4.5.3. *Towards Conjecture 4.7.* Let $X = X(y)$ denote the parametrization of the limit curve Ω_* (given by (1.7)) in the French coordinate system. This function associates each $y \geq 0$ with its corresponding x -coordinate. By convention, $X(0) = 2$ so that $X : [0, \infty) \rightarrow [0, 2]$. By a minor adjustment of Remark 3.2 it follows that the x -distance $d_X(\omega_n, \Omega_*)$ is the supremum distance between the function X and the piecewise constant function which to the French y coordinate associates the corresponding x -coordinate of the right envelope of the rescaled diagram $\lambda^{(n)}$. As a consequence,

$$(4.12) \quad d_X(\omega_n, \Omega_*) = \max \left(\max_{i \geq 1} \left| \frac{1}{\sqrt{n}} \lambda_i^{(n)} - X \left(\frac{i-1}{\sqrt{n}} \right) \right|, \max_{i \geq 1} \left| \frac{1}{\sqrt{n}} \lambda_i^{(n)} - X \left(\frac{i}{\sqrt{n}} \right) \right| \right).$$

The results of Baik, Deift and Johansson [1] and Okounkov [20] imply that each finite family of (shifted and rescaled) lengths of the bottom rows

$$n^{-\frac{1}{6}} (\lambda_i^{(n)} - 2\sqrt{n}) = n^{\frac{1}{3}} \left(\frac{1}{\sqrt{n}} \lambda_i^{(n)} - 2 \right)$$

indexed by $i \in \{1, \dots, i_{\max}\}$ converges in distribution to a specific nondegenerate probability distribution (which happens to be related to the largest $i_{\max}$ eigenvalues of a large GUE random matrix). As a consequence, each individual component of the iterated maximum (4.12) is a random variable which asymptotically is of order *at least* $n^{-\frac{1}{3}}$. By the symmetry of the problem, an analogous analysis can be performed for the y -distance $d_Y(\omega_n, \Omega_*)$ which gives a lower bound $\alpha \geq -\frac{1}{3}$ for the exponent in the conjectural asymptotic formula (4.10). It also shows that Conjecture 4.7 cannot be true for $\alpha \leq -\frac{1}{3}$.

Interestingly, this theoretical lower bound $\alpha \geq -\frac{1}{3}$ is very close to the value (4.11) based on the Monte Carlo experiments. This could be an indication that the ‘*main contribution*’ to the maximum in (4.12) comes from a ‘*small*’ (finite, or growing very slowly with n) number of components there which correspond to the bottom rows of the Young diagram. Monte Carlo experiments related to Plancherel-distributed random Young diagrams with $n = 10^6$ boxes give some support to this heuristic statement. Regretfully, any rigorous results related to Conjecture 4.7 with $\alpha \approx -\frac{1}{3}$ are currently beyond our reach.

The seminal results of Logan and Shepp [16] and Vershik and Kerov [33] can be interpreted as a rudimentary version of Conjecture 4.7 with exponent $\alpha = 0$. To explore potential improvements, we revisit their original proofs.

Using Romik’s notation [23], we aim to refine [23], Theorem 1.26, by replacing the fixed positive constant $\epsilon > 0$ with a sequence (ϵ_n) defined as:

$$\epsilon_n = n^\alpha$$

where $\alpha < 0$. Our goal is to maintain the theorem’s conclusion (that “the probability of a specific event converges to 1”) while minimizing α . This modification must be propagated throughout the entire supply chain of supporting proofs.

However, a critical component of this chain, [23], Theorem 1.20, relies on the condition:

$$-\epsilon_n^2 n + C\sqrt{n} + O(\log n \sqrt{n}) \xrightarrow[n \rightarrow \infty]{} -\infty$$

for some constant $C > 0$. This condition holds only if ϵ_n does not decay too rapidly, specifically when $\alpha > -\frac{1}{4}$.

Consequently, we can prove that Conjecture 4.7 holds for $\alpha > -\frac{1}{4}$. While this represents an improvement over the original result, it falls short of our target $\alpha = -\frac{1}{4}$, highlighting the limitations of this approach.

We shall discuss some alternative versions of the Assumption (A2) as well as some other paths towards Conjecture 1.7 later in Section 5.4.

4.5.4. Transition measure for the limit shape and its interaction energy. Our quest towards Conjecture 1.5 via the strategy outlined in Section 4.5.1 involves also identification of the transition measure μ_{Ω_*} of the limit shape Ω_* . By the result of Kerov [13] it is the semicircle measure μ_{SC} with the density f_{SC} given by (1.8); it is no coincidence that this measure already appeared in Section 1.6. The following result provides an explicit formula for its interaction energy.

PROPOSITION 4.8. *For the semicircle measure μ_{SC} the corresponding interaction energy is given by*

$$\mathcal{E}_{\text{SC}}(u) = \begin{cases} \frac{1}{12\pi}(4-u^2)^{\frac{3}{2}} & \text{for } -2 \leq u \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

PROOF. The method of the contour integral considered in Section 4.2 is applicable; we leave the details for the interested reader. In the following, we provide an alternative method.

For $\epsilon \in \mathbb{R}$, we define the regularized version of $\mathcal{E}_{\text{SC}}(u)$ given by

$$\mathcal{E}_{\epsilon}(u) := \iint_{-2 < z_1 < u < z_2 < 2} \Re \frac{1}{z_2 - z_1 + i\epsilon} f_{\text{SC}}(z_1) f_{\text{SC}}(z_2) \, dz_1 \, dz_2;$$

our goal is to evaluate $\mathcal{E}_0(u) = \mathcal{E}_{\text{SC}}(u)$. Note that for real numbers $z_2 > z_1$

$$\mathbb{R}_+ \ni \epsilon \mapsto \Re \frac{1}{z_2 - z_1 + i\epsilon} = \frac{z_2 - z_1}{(z_2 - z_1)^2 + \epsilon^2}$$

is a decreasing, positive, continuous function. By Lebesgue monotone convergence theorem it follows that

$$F_0(u) = \lim_{\epsilon \rightarrow 0} F_{\epsilon}(u).$$

In order to evaluate $F_{\epsilon}(u)$ we notice that for $\epsilon > 0$ the derivative of $\mathcal{E}_{\epsilon}$ is given by

$$\mathcal{E}'_{\epsilon}(u)$$

$$\begin{aligned} &= \int_{u < z_2 < 2} \Re \frac{1}{z_2 - u + i\epsilon} f_{\text{SC}}(u) f_{\text{SC}}(z_2) \, dz_2 - \int_{-2 < z_1 < u} \Re \frac{1}{u - z_1 + i\epsilon} f_{\text{SC}}(z_1) f_{\text{SC}}(u) \, dz_1 \\ &= f_{\text{SC}}(u) \Re \int_{-2 < z < 2} \frac{1}{z - (u + i\epsilon)} f_{\text{SC}}(z) \, dz = f_{\text{SC}}(u) \Re \mathbf{G}_{\text{SC}}(u + i\epsilon); \end{aligned}$$

above we used the fact that the integral on the right-hand side is the Cauchy transform of the semicircular distribution

$$\mathbf{G}_{\text{SC}}(w) = \int_{-2 < z < 2} \frac{1}{z - w} f_{\text{SC}}(z) \, dz = \frac{w - \sqrt{w^2 - 4}}{2},$$

evaluated at $w = i + i\epsilon$, see [19], Section 3.1.

For $-2 \leq u \leq 2$ the principal value of the Cauchy integral is given by

$$\mathbf{G}_{\text{SC}}(u) = \lim_{\epsilon \rightarrow 0} \Re \mathbf{G}_{\text{SC}}(u + i\epsilon) = -\frac{u}{2}$$

and the convergence is uniform over the interval $[-2, 2]$. We proved in this way that

(4.13)
$$\mathcal{E}'_{\text{SC}}(u) = f_{\text{SC}}(u)\mathbf{G}_{\text{SC}}(u).$$

It follows that

$$\mathcal{E}_{\text{SC}}(u) = \int_{-2}^u f_{\text{SC}}(w) \frac{-w}{2} \, dw = \frac{1}{12\pi} (4 - u^2)^{\frac{3}{2}} \quad \text{for } -2 \leq u \leq 2.$$

The above calculation can be generalized to other measures than the semicircular law for which the Cauchy transform is sufficiently regular. Be warned, that for the example of the arc-sine law from Section 4.4.1 the analogue of (4.13) implies that

$$\mathcal{E}'_{\text{AS}}(u) = 0$$

which, technically speaking, is correct but not very helpful for finding the exact form of the formula (4.7). $\square$

4.5.5. *Conclusion.* One can now easily verify that, apart from the condition (A2), the remaining assumptions of Theorem 4.3 are fulfilled for the setup outlined in Section 4.5.1. Our choice of the variance (1.14) in Conjecture 1.5 is equal to $\frac{\mathcal{E}_{\text{SC}}}{(f_{\text{SC}})^2}$ and it was based on the the variance on the right-hand side of (4.4).

5. A version of the main theorem with weaker assumptions. This section focuses on Theorem 5.1, which serves as both a stronger version of the main theorem (Theorem 4.3) and an essential intermediate step in its proof. We begin by discussing the motivations behind this result.

5.1. *Motivations.* Assumptions (A2), (A4) of the main result (Theorem 4.3) were formulated in the language that involves the ‘*shape*’ of the continual diagrams (ω_n) as well as the ‘*shape*’ of their limit Ω . This language provides a very convenient intuitive meaning to the assumptions, but it also has some disadvantages which we discuss below.

5.1.1. *Tension between local and global.* First, Assumption (A2) may be overly restrictive for certain applications, as discussed in Section 4.5. It would be advantageous to replace it with a weaker condition. The key issue is that the distance $d_{XY}(\omega_n, \Omega)$ can be viewed as a form of the supremum distance, which is influenced by the ‘*worst behaving*’ part of the continual diagram ω_n . This influence persists even if the problematic part constitutes a relatively ‘*small*’ fraction of the Young diagram profile or is located ‘*far away*’ from where the new box is expected to appear. This phenomenon is evident in Plancherel-distributed random Young diagrams, where fluctuations at the edge are significantly larger than those in the bulk.

For this reason, it would be desirable to replace (A2) by some assumptions that fall into the following two types.

- The assumptions of the first type should be a quite restrictive about the ‘*local fluctuations*’ of the continual diagrams (ω_n) in the neighborhood of the point u_0 where we expect the new box to appear.
- The assumptions of the second type should provide quite rough information about the overall global shape of the diagrams (ω_n) .

As we shall see, each assumption of Theorem 5.1 is expressed in terms of the transition measure of ω_n . One can argue that each of them is a mixture of these two types. Indeed, on one hand, each of these assumptions is very sensitive to the local behavior of the transition measure of ω_n (which itself is most sensitive to the local behavior of the continual diagram ω_n). On the other hand, the transition measure is defined in a very ‘*nonlocal*’ way so that its behavior around u_0 is influenced also by very ‘*distant*’ areas of ω_n .

We believe that the assumptions of Theorem 5.1 strike the right balance between the local and the global conditions that allows to apply the main result in a wider class of examples.

5.1.2. *Transition measure as the language of the proof.* Second, the vast majority of the proof of Theorem 4.3 is formulated purely in the language of the transition measures of the continual diagrams (ω_n) and not in the language that involves their shapes. In fact, the first step of the proof (Proposition 6.1) is to translate the assumptions of Theorem 4.3 to the language of the transition measures and to never look back again.

5.2. *Notations.* If μ is a probability measure on the real line, $u_0 \in \mathbb{R}$ and $\epsilon \geq 0$ we define the regularized interaction energy

$$\mathcal{E}_{\epsilon; \mu}(u_0) = \iint_{\{(z_1, z_2): z_1 \leq u_0 < z_2\}} \frac{1}{z_2 - z_1 + \epsilon} d\mu(z_1) d\mu(z_2).$$

Note that for $\epsilon = 0$, we recover the usual interaction energy which was introduced in Section 4.2.

For a probability measure μ on the real line, we define

$$(5.1) \quad \mathbf{G}_\mu^+(z) = \int_{\mathbb{R}} \frac{1}{|z - x| + 1} d\mu(x)$$

as a (slightly regularized) version of the Cauchy transform with the absolute value of the original kernel. For a Young diagram λ , we will use the simplified notation

$$\mathbf{G}_\lambda^+ = \mathbf{G}_{\mu_\lambda}^+.$$

Recall that K_λ is the cumulative function of the diagram λ , see Equation (3.4).

5.3. *The main theorem, alternative version.*

THEOREM 5.1 (The main theorem with alternative, weaker assumptions). *Let $u_0 \in \mathbb{R}$, and $0 < z_0 < 1$, and $f_0 > 0$, and $\mathcal{E}_0 > 0$ be fixed. For each integer $n \geq 1$ let $\lambda^{(n)}$ be a random Young diagram. We denote by*

$$\omega_n = \omega_{\frac{1}{\sqrt{n}}\lambda^{(n)}}$$

the corresponding rescaled random continual diagram.

Let $\mathbf{Q}$ be a standard Gaussian variable which is independent from the random diagrams $(\lambda^{(n)})$. We denote

$$(5.2) \quad \mathbf{u}_n = u_0 + \frac{1}{\sqrt[4]{n}} \mathbf{Q}.$$

We assume that:

(B1) ('the cumulative functions can be approximated locally by an affine function')

$$\sqrt[4]{n} [K_{\omega_n}(\mathbf{u}_n) - z_0 - f_0 \cdot (\mathbf{u}_n - u_0)] \xrightarrow[n \rightarrow \infty]{P} 0,$$

(B2) ('the regularized interaction energy can be locally approximated by $\mathcal{E}_0$ ')

$$\mathcal{E}_{\frac{1}{\sqrt{n}}; \mu_{\omega_n}}(\mathbf{u}_n)$$

(B3) ('the integrand within Cauchy transform is not too singular')

$$(5.3) \quad n^{\frac{3}{8}} \mathbf{G}_{\lambda^{(n)}}^+(\sqrt{n} \mathbf{u}_n) \xrightarrow[n \rightarrow \infty]{P} 0.$$

Let $T^{(n)}$ be a uniformly random Poissonized tableau of shape $\lambda^{(n)}$. Then the fluctuations of the newly created box around the first-order approximation

(5.4)
$$\sqrt[4]{n}\left[\frac{\mathbf{u}\text{-Ins}(T^{(n)}; z_0)}{\sqrt{n}} - u_0\right] \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N\left(0, \frac{\mathcal{E}_0}{f_0^2}\right)$$

converge to a centered Gaussian measure with the variance $\frac{\mathcal{E}_0}{f_0^2}$.

The proof is postponed to Section 11.3.

REMARK 5.2. Theorem 5.1 makes Assumptions (B1)–(B3) in terms of the random variable $\mathbf{Q}$. A reader might initially assume that these assumptions hold true for all possible values of $\mathbf{Q} \in \mathbb{R}$. However, it is crucial to understand that the theorem’s assumptions are much weaker than that. The key point is that the Assumptions (B1)–(B3) need not hold for every single value of $\mathbf{Q}$. Instead, the theorem allows for the existence of ‘bad’ values for which these assumptions may not be satisfied. By considering $\mathbf{Q}$ as a random variable and requiring convergence in probability, the theorem provides a convenient framework to state that such ‘bad’ values do not dominate the overall behavior.

REMARK 5.3. The reader may question the assumption that the random variable $\mathbf{Q}$ follows the Gaussian distribution. It is important to note that there is flexibility in this assumption, and the probability distribution of $\mathbf{Q}$ can be modified. The key requirement is that the distribution of $\mathbf{Q}$ is absolutely continuous, with a density that is strictly positive over the entire real line. If the result holds true for one probability distribution satisfying this property, it will hold for any other distribution with the same property. The details of this generalization are left for the reader to explore.

However, the proof of the implication from Theorem 4.3 to Theorem 5.1 is more straightforward if the density converges to zero in $\pm\infty$. This property motivated the choice of the Gaussian distribution.

5.4. *Towards Conjecture 1.7, part 2.* Building upon our discussion in Section 4.5, we now revisit our goal of proving Conjecture 1.7. Our previous attempt relied on Theorem 4.3. In this section, we present a refined approach using Theorem 5.1, which offers the advantage of weaker assumptions.

Theorem 5.1 presents two key assumptions, among which assumption (B1) emerges as the most challenging. This critical condition requires that the CDF of the transition measure for a random Plancherel-distributed Young diagram exhibits fluctuations within a prescribed, sufficiently small rate.

Our investigation into this condition has yielded promising insights. Monte Carlo experiments, as illustrated by the red thin curves in Figures 12 and 13, suggest that this condition is indeed satisfied. Furthermore, analogous results in random matrix theory, particularly those of Gustavsson [10], provide additional support for this hypothesis. Based on these observations, we propose a stronger conjecture.

CONJECTURE 5.4. Let $\lambda^{(n)}$ be a random Young diagram with n boxes distributed according to the Plancherel measure. For each $-2 < u < 2$ and each $\alpha < \frac{1}{2}$

$$n^\alpha \left[K_{\lambda^{(n)}}(\sqrt{n}u) - F_{\text{SC}}(u) \right] \xrightarrow[n \rightarrow \infty]{P} 0.$$

This conjecture encompasses and extends beyond the requirements of Assumption (B1). The case $\alpha = 0$ corresponds to the established weak convergence of probability measures [24], Theorem 3.2, while Assumption (B1) aligns with the case $\alpha = \frac{1}{4}$.

While Conjecture 5.4 remains beyond our current reach, it offers a tantalizing glimpse into the deeper structure of Plancherel-distributed Young diagrams and their transition measures.

6. Implication between two versions of the main theorem. Probably the best way of giving an intuitive meaning of the assumptions of Theorem 5.1 is to establish a link to the main result (Theorem 4.3). This link is provided by the following proposition, which also serves as an intermediate step for the proof of the main result.

PROPOSITION 6.1. *Theorem 5.1 implies Theorem 4.3.*

PROOF. Both theorems share the same conclusion. Therefore, to establish the implication, it suffices to show that the assumptions of Theorem 4.3 entail those of Theorem 5.1. We begin by assuming that the hypotheses of Theorem 4.3 hold. The subsequent sections are dedicated to verifying under these conditions each assumption of Theorem 5.1 in turn. $\square$

6.1. *Assumptions of Theorem 4.3 imply Assumption (B1).* Let us choose the constants a_0 and b_0 in such a way that

$$a < a_0 < u_0 < b_0 < b.$$

The assumptions of Theorem 4.3 were fine-tuned in such a way that a recent result of the second named author [27], Theorem 1.4, is applicable with the exponent $\alpha = \frac{1}{4}$; as a consequence

$$(6.1) \quad \sqrt[4]{n} \sup_{z \in [a_0, b_0]} |K_{\lambda^{(n)}}(\sqrt{n}z) - K_{\Omega}(z)| \xrightarrow{n \rightarrow \infty} 0$$

and Assumption (B1) follows easily from the Taylor expansion of the cumulative function:

$$K_{\Omega}\left(u_0 + \frac{c}{\sqrt[4]{n}}\right) = z_0 + \frac{f_0 c}{\sqrt[4]{n}} + o\left(\frac{c}{\sqrt[4]{n}}\right).$$

6.2. *Assumption (B2). The lower bound.* For a given probability measure μ on the real line, the interaction energy $\mathcal{E}_{\mu}(u_0)$ (cf. (4.1)) can be viewed as an integral with respect to the product measure $\mu \times \mu$ of the function

$$\phi(z_1, z_2) = \frac{1}{z_2 - z_1} \mathbb{1}_{(-\infty, u_0]}(z_1) \mathbb{1}_{(u_0, \infty)}(z_2)$$

which has a singularity at $z_1 = z_2 = u_0$. Our strategy is to approximate this function by some more regular functions.

Let us fix $\epsilon > 0$. By Assumption (A3) about the local behavior of the transition measure μ_{Ω} there exists $\delta > 0$ and a nonnegative continuous function $g: \mathbb{R}^2 \rightarrow \mathbb{R}_+$ with the following three properties:

- *upper bound:*

$$g \leq \phi,$$

- *support is away from the singularity:*

$$\text{if } g(z_1, z_2) > 0 \text{ then } z_1 < u_0 - \delta \text{ and } z_2 > u_0 + \delta,$$

- the total mass is almost maximal:

$$\iint g(z_1, z_2) \, d\mu_\Omega(z_1) \, d\mu_\Omega(z_2) > \mathcal{E}_{\mu_\Omega}(u_0) - \epsilon.$$

Our objective in this subsection is to demonstrate that the sequence of probabilities

$$(6.2) \qquad \mathbb{P}\big[\mathcal{E}_{\frac{1}{\sqrt{n}}; \mu_{\omega_n}}(\mathbf{u}_n) > \mathcal{E}_0 - \epsilon\big] \xrightarrow{n \rightarrow \infty} 1$$

converges to 1. To achieve this, we employ a conditioning argument on the event

$$\mathbf{u}_n \in [u_0 - \delta, u_0 + \delta].$$

Since the probability of the complement of this event converges to zero as $n \rightarrow \infty$, we can assume without loss of generality that $\mathbf{u}_n \in [u_0 - \delta, u_0 + \delta]$ holds almost surely.

By comparing the analogue of the kernel ϕ for the regularized interaction energy with the kernel g , we obtain the inequality

$$(6.3) \qquad \mathcal{E}_{\frac{1}{\sqrt{n}}; \mu_{\omega_n}}(\mathbf{u}_n) \geq \frac{2\delta}{2\delta + \frac{1}{\sqrt{n}}} \iint g(z_1, z_2) \, d\mu_{\omega_n}(z_1) \, d\mu_{\omega_n}(z_2).$$

Note that the numerical factor on the right-hand side converges to 1 as $n \rightarrow \infty$.

Kerov’s results ([13] and [14], Chapter 4, Section 1) imply that Assumption (A2) is sufficient to conclude that the sequence of random measures (μ_{ω_n}) converges to μ_Ω in the weak topology of probability measures, in probability. Consequently, the double integral on the right-hand side of (6.3) converges in probability:

$$(6.4) \qquad \iint g(z_1, z_2) \, d\mu_{\omega_n}(z_1) \, d\mu_{\omega_n}(z_2) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \iint g(z_1, z_2) \, d\mu_\Omega(z_1) \, d\mu_\Omega(z_2).$$

Combining (6.3) and (6.4), we conclude that (6.2) holds true, completing the proof of the lower bound in (B2). The proof of the upper bound is postponed to Section 6.4.

6.3. *Bounding the interaction energy from above.* We present an intermediate result crucial for proving that the upper bound in Assumption (B2) holds.

PROPOSITION 6.2. *Let constants $a < u_0 < b$ be such that the probability measure μ restricted to the interval $[a, b]$ is absolutely continuous, with a density given by a bounded function which is bounded below by a positive constant. We denote*

$$z_0 = F_\mu(u_0), \qquad \mathcal{E}_0 = \mathcal{E}_\mu(u_0).$$

For any $\epsilon > 0$ and any open interval J containing z_0 , there exists an open neighborhood G of μ in the topology of weak convergence of probability measures, along with a natural number $n_0 \in \mathbb{N}$ such that the following holds:

For every $n \geq n_0$ and for any probability measure ν on the real line that satisfies both of the following conditions:

- global proximity condition (“ $\nu \approx \mu$ ”):

$$(6.5) \qquad \nu \in G,$$

- local refinement condition (“the CDFs of ν and μ are locally very close”):

$$(6.6) \qquad |F_\nu(u) - F_\mu(u)| < \frac{1}{\sqrt[4]{n}} \quad \text{for each } u \in J,$$

the following bound is satisfied:

$$(6.7) \qquad \mathbb{E}[\mathcal{E}_\nu(\mathbf{u}_n)] < \mathcal{E}_0 + \epsilon.$$

The remaining part of this subsection is devoted to its proof.

6.3.1. *The expected value of $\mathcal{E}_v(\mathbf{u}_n)$ as a double integral. Decomposition of the integration domain.* The expected value on the left-hand side of (6.7) can be written as:

$$(6.8) \quad \mathbb{E}[\mathcal{E}_v(\mathbf{u}_n)] = \iint_{z_1 < z_2} \frac{\mathbb{P}[\mathbf{u}_n \in [z_1, z_2]]}{z_2 - z_1} d\nu(z_1) d\nu(z_2).$$

Let I be an open interval containing z_0 , such that its closure $\bar{I}$ is contained within J . The necessity of introducing another open interval, in addition to J , will become clearer in the proof of Lemma 6.3.

We will divide the integration region in the integral on the right-hand side of (6.8) into a number of (nondisjoint) parts, and analyze the contribution of each part separately. This division follows the following scheme which forms a tree:

- The small macroscopic region where $(z_1, z_2) \in I \times I$ and we have very good control over the CDF of ν . This region splits further:
 - the part where

$$z_2 - z_1 \geq \frac{\log n}{\sqrt[4]{n}}$$

and the denominator on the right-hand side of (6.8) is well separated from the singularity. We will examine the contribution of this area in Lemmas 6.3 and 6.5. This scenario represents one of the two cases where the limit contribution is nonzero for $n \rightarrow \infty$. However, it is important to note that this contribution approaches zero in the iterated limit, as $n \rightarrow \infty$ and the length of the interval I tends to zero.

- the part where

$$0 < z_2 - z_1 < \frac{\log n}{\sqrt[4]{n}}$$

and the coordinates are close to each other. This part splits into three further (nonexclusive) possibilities:

- * Both coordinates are not too far from z_0 :

$$(6.9) \quad |z_1 - z_0|, |z_2 - z_0| \leq \frac{1}{\sqrt[8]{n} \log n}.$$

We will investigate the contribution of this area in Lemma 6.6 and we will show that it converges to zero.

- * Both coordinates are on the same side of z_0 and well separated from z_0 , that is, one of the following two cases holds true:

$$(6.10) \quad z_1, z_2 > z_0 + \frac{\log n}{\sqrt[4]{n}} \quad \text{or} \quad z_1, z_2 < z_0 - \frac{\log n}{\sqrt[4]{n}}.$$

We will investigate the contribution of the first case in Lemma 6.7 and show that it converges to zero. The analysis for the second case is analogous and will be omitted.

- * The third option, which occurs when neither condition (6.9) nor (6.10) is satisfied, is not possible for sufficiently large n for which

$$\frac{2 \log n}{\sqrt[4]{n}} < \frac{1}{\sqrt[8]{n} \log n}.$$

Indeed, let us say that it is z_2 that is far from z_0 so that (6.9) is false. The case when z_2 is on the left-hand side of z_0 , so that:

$$z_1 < z_2 < z_0 - \frac{1}{\sqrt[8]{n} \log n}$$

is not possible as it would contradict that the condition (6.10) is not satisfied.

Thus, z_2 is on the right-hand side of z_0 and

$$z_2 > z_0 + \frac{1}{\sqrt[8]{n} \log n}.$$

Since condition (6.10) is not satisfied, it follows that

$$z_1 \leq z_0 + \frac{\log n}{\sqrt[4]{n}}.$$

This leads to a contradiction:

$$\frac{\log n}{\sqrt[4]{n}} > z_2 - z_1 > \left(z_0 + \frac{1}{\sqrt[8]{n} \log n} \right) - \left(z_0 + \frac{\log n}{\sqrt[4]{n}} \right) \gg \frac{\log n}{\sqrt[4]{n}}$$

for sufficiently large values of n . Therefore, this case is not possible for n above a certain threshold.

- The complement of $I \times I$, that is, $(z_1, z_2) \notin I \times I$. This region splits further:
 - the part where the coordinates are on the opposite sides of z_0 :

$$z_1 < z_0 < z_2.$$

We will investigate the contribution of this area in Lemma 6.8. This is the region which provides the main asymptotic contribution.

- and the one where the coordinates are the same side of z_0 , say

$$z_0 < z_1, z_2$$

(the other case when $z_1, z_2 < z_0$ follows a similar pattern and we skip it). There are two additional cases:

- * z_1 and z_2 are well separated from z_0 and the first inequality in (6.10) holds true. As we already mentioned, Lemma 6.7 is applicable in this case; it shows that the contribution of this area converges to zero.
- * z_1 is very close to z_0 while z_2 is very far:

$$z_0 < z_1 < z_0 + \frac{\log n}{\sqrt[4]{n}} \quad \text{and} \quad z_2 \notin I.$$

We will study the contribution of this area in Lemma 6.9 and we will show that it converges to zero. The other case, when $z_1, z_2 < z_0$ is analogous and we skip it.

We will now explore the specific contributions of these cases before presenting the culmination of our proof in Section 6.3.7.

6.3.2. Close to the singularity. In this section, we analyze the integral on the left-hand side of (6.11). We will do it in two steps. First, in Lemma 6.3 we bound this integral over the product measure $\nu \times \nu$ by a similar integral over the more regular measure $\mu \times \mu$.

6.3.2.1. Conversion of the integral to $\mu \times \mu$.

LEMMA 6.3. *Let μ be a probability measure on the real line, and let constants $a < u_0 < b$ be such that μ restricted to the interval $[a, b]$ is absolutely continuous, with a density given by a function which is bounded below by a positive constant.*

For any pair of open intervals I and J containing z_0 , where the closure of I is contained within J and F_μ is Lipschitz continuous on J , there exists a constant $C > 0$ and a natural number $n_0 \in \mathbb{N}$ such that the following holds.

For every $n \geq n_0$ and for any probability measure ν on the real line that satisfies the local refinement condition (6.6) the following bound is satisfied:

$$\begin{aligned}
 & \iint_{\substack{\{(z_1, z_2): \\ z_2 - z_1 > \log nn^{-\frac{1}{4}}, \\ (z_1, z_2) \in I \times I\}}} \frac{\mathbb{P}[\mathbf{u}_n \in [z_1, z_2]]}{z_2 - z_1} d\nu(z_1) d\nu(z_2) \\
 (6.11) \quad & \leq \iint_{\substack{\{(z'_1, z'_2): \\ z'_2 - z'_1 > (\log n - 2C)n^{-\frac{1}{4}}, \\ (z'_1, z'_2) \in J \times J\}}} \frac{\mathbb{P}[\mathbf{u}_n \in [z'_1 - Cn^{-\frac{1}{4}}, z'_2 + Cn^{-\frac{1}{4}}]]}{z'_2 - z'_1 - 2Cn^{-\frac{1}{4}}} d\mu(z'_1) d\mu(z'_2).
 \end{aligned}$$

PROOF. Let I' be an open interval such that the following sequence of inclusions is valid:

$$(6.12) \quad z_0 \in I \subset \bar{I} \subset I' \subset \bar{I}' \subset J.$$

From the local refinement condition (6.6), it follows that the image $F_\nu(I)$ is contained within an $o(1)$ -neighborhood of $F_\mu(I)$ as $n \rightarrow \infty$. Given that F_μ is a homeomorphism on the interval J , we can conclude that

$$\overline{F_\mu(I)} \subset F_\mu(I').$$

Thus, there exists an integer n_0 such that for all sufficiently large $n \geq n_0$, and for any measure ν satisfying the local refinement condition (6.6), the following inclusion holds for the corresponding quantiles:

$$F_\nu(I) \subseteq F_\mu(I').$$

For each $i \in \{1, 2\}$, the integral on the left-hand side of (6.11) over the variable $z_i \in I$ with respect to the measure ν can be transformed using a change of variables by setting $z_i = F_\nu^{-1}(t_i)$. The variable t_i has a natural interpretation as the quantile of the measure ν . This allows us to express the integral in terms of the uniform measure over the variable t_i , where $t_i \in F_\nu(I)$.

An analogous idea can be applied to the integral on the right-hand side of (6.11) over the variable $z'_i \in J$ with respect to the measure μ . By combining these two changes of the variables we can express the variable on the left-hand side

$$z_i = (F_\nu^{-1} \circ F_\mu)(z'_i)$$

in terms of the variable on the right-hand side, over the set

$$(F_\mu^{-1} \circ F_\nu^{-1})(I) \subseteq J.$$

We apply Lemma 6.4. With the notations of that result, our claim holds true for

$$C := \frac{1}{f_{\min}}.$$

□

LEMMA 6.4. Under the assumptions of Lemma 6.3 and (6.12), we denote by $f_{\min} > 0$ any lower bound on the density of the measure μ on the interval J . If n is sufficiently large, then for any $z' \in I'$ and any probability measure ν that satisfies the local refinement condition (6.6), the following bound holds:

$$|(F_\nu^{-1} \circ F_\mu)(z') - z'| \leq \frac{1}{f_{\min} \sqrt[4]{n}}.$$

PROOF. Let $z = (F_v^{-1} \circ F_\mu)(z')$. Due to the definition of the quantile function, particularly at discontinuities, it follows that

$$(6.13) \qquad F_\mu(z') \in [F_v(z^-), F_v(z)],$$

with the usual convention that $F_v(z^-) = \lim_{\xi \rightarrow z^-} F_v(\xi)$ denotes the left limit of CDF.

We define:

$$\begin{aligned} z_+ &:= z + \frac{|F_\mu(z) - F_v(z)|}{f_{\min}}, \\ z_- &:= z - \frac{|F_\mu(z) - F_v(z^-)|}{f_{\min}}. \end{aligned}$$

There exists n_0 such that for all sufficiently large $n \geq n_0$, both z_- and z_+ are guaranteed to remain within the interval J by the local refinement condition (6.6). Since F_μ locally grows at least linearly with slope $f_{\min}$, we have:

$$F_\mu(z_+) \geq F_\mu(z) + |F_\mu(z) - F_v(z)| \geq F_v(z) \geq F_\mu(z'),$$

where the last inequality follows by (6.13). As F_μ is strictly increasing on J , we conclude that

$$z' \leq z_+.$$

The argument for the inequality $z_- \leq z'$ follows similarly. Consequently, as $z_- \leq z \leq z_+$, we establish that:

$$|z - z'| \leq \frac{1}{f_{\min}} \max(|F_\mu(z) - F_v(z)|, |F_\mu(z) - F_v(z^-)|).$$

This completes the proof by the local refinement condition (6.6). $\square$

6.3.2.2. *Estimate for the integral with respect to $\mu \times \mu$.*

LEMMA 6.5. *Under the assumptions of Lemma 6.3, let us further suppose that the density of the measure μ , when restricted to the interval J , is bounded above by $f_{\max}$. Given these conditions, we can establish an upper bound for the right-hand side of (6.11) by:*

$$f_{\max}^2 |J| + o(1).$$

PROOF. We begin by applying a change of variables to the integral on the right-hand side of (6.11). Let us integrate over the variables:

$$z'_1, \quad t := z'_2 - z'_1.$$

To obtain an upper bound, we extend the integration domain as follows: we integrate z'_1 over the whole real line $\mathbb{R}$, and we integrate t over the interval

$$\left[\frac{\log n - 2C}{\sqrt[4]{n}}, |J| \right].$$

For a fixed value of $t = z'_2 - z'_1$, the integral of the numerator simplifies to:

$$\int \mathbb{P}[\mathbf{u}_n \in [z'_1 + a, z'_1 + b]] \, dz'_1 = b - a.$$

Notably, this result is independent of the specific probability distribution of $\mathbf{u}_n$.

Consequently, we can bound the right-hand side of (6.11) by:

$$\begin{aligned} & f_{\max}^2 \int_{(\log n - 2C)n^{-\frac{1}{4}}}^{|J|} \frac{t + 2Cn^{-\frac{1}{4}}}{t - 2Cn^{-\frac{1}{4}}} dt \\ & \leq f_{\max}^2 \left(|J| + \frac{4C}{\sqrt[4]{n}} \int_{(\log n - 4C)n^{-\frac{1}{4}}}^{|J| + 2Cn^{-\frac{1}{4}}} \frac{1}{u} du \right) = f_{\max}^2 |J| + o(1), \end{aligned}$$

which completes the proof. $\square$

6.3.3. Extremely close to the singularity.

LEMMA 6.6. *Let μ be as in Proposition 6.2. There exists a sequence (C_n) which converges to zero such that for any measure ν which fulfills the local refinement condition (6.6) the following bound holds true:*

$$\iint_{\substack{z_1 < z_2, \\ |z_1 - z_0|, |z_2 - z_0| \leq \frac{1}{\sqrt[8]{n} \log n}}} \frac{\mathbb{P}[\mathbf{u}_n \in [z_1, z_2]]}{z_2 - z_1} d\nu(z_1) d\nu(z_2) < C_n.$$

PROOF. The integral is bounded by the product of some upper bound on the integrand with the measure of the domain of integration.

Observe that the integrand is bounded above by the density of the random variable $\mathbf{u}_n$; this density is of order $O(\sqrt[4]{n})$.

The measure of the domain of integration bounded by the square of:

$$\nu\left(\left[z_0 - \frac{1}{\sqrt[8]{n} \log n}, z_0 + \frac{1}{\sqrt[8]{n} \log n}\right]\right) \leq \mu\left(\left[z_0 - \frac{1}{\sqrt[8]{n} \log n}, z_0 + \frac{1}{\sqrt[8]{n} \log n}\right]\right) + \frac{2}{\sqrt[4]{n}}.$$

Clearly, the right-hand side is of order at most $o(\frac{1}{\sqrt[4]{n}})$, so its square is of order $o(\frac{1}{\sqrt[4]{n}})$.

The whole product is therefore of order $o(1)$, as required. $\square$

6.3.4. Both variables on the same side of z_0 , but not too close.

LEMMA 6.7. *There exists a sequence (C_n) which converges to zero, such that for any integer $n \geq 1$ and any probability measure ν on the real line the following bound is true:*

$$\iint_{\substack{\{(z_1, z_2): \\ z_2 > z_1 > z_0 + n^{-\frac{1}{4}} \log n\}}} \frac{\mathbb{P}[\mathbf{u}_n \in [z_1, z_2]]}{z_2 - z_1} d\nu(z_1) d\nu(z_2) \leq C_n.$$

PROOF. The integrand is bounded above by the supremum of the probability density of $\mathbf{u}_n$ on the interval

$$\left[\frac{\log n}{\sqrt[4]{n}}, \infty\right)$$

which is the same as the product of $\sqrt[4]{n}$ by the supremum of the probability density of $\mathbf{Q}$ on the interval

$$[\log n, \infty).$$

A direct calculation involving the explicit formula for the density of the normal distribution concludes the proof. $\square$

6.3.5. *The main contribution to the integral (away from the singularity).* The numerator in (6.8) is trivially bounded by 1. Consequently, the following result offers an upper bound for the corresponding part of the integral in (6.8).

LEMMA 6.8. *Let μ be a probability measure on the real line and $u_0 \in \mathbb{R}$. We denote*

$$z_0 = F_\mu(u_0), \quad \mathcal{E}_0 = \mathcal{E}_\mu(u_0).$$

For any $\epsilon > 0$ and any open interval I containing z_0 , there exists an open neighborhood G of μ in the topology of weak convergence of probability measures such that for any probability measure $\nu \in G$ the following bound holds:

$$(6.14) \quad \iint_{\substack{\{(z_1, z_2): \\ z_1 \leq z_0 \leq z_2, \\ (z_1, z_2) \notin I \times I\}}} \frac{1}{z_2 - z_1} d\nu(z_1) d\nu(z_2) < \mathcal{E}_0 + \epsilon.$$

PROOF. The integral on the left-hand side of (6.14) involves a bounded, positive integrand integrated over a closed set. This integral can be rewritten as an integral over $\mathbb{R}^2$ of the original integrand

$$\mathbb{R}^2 \ni (z_1, z_2) \mapsto \frac{1}{z_2 - z_1}$$

multiplied by an indicator function of a closed subset of $\mathbb{R}^2$. This product is the pointwise limit of a weakly decreasing sequence (f_n) of continuous, uniformly bounded functions $f_n : \mathbb{R}^2 \rightarrow \mathbb{R}_+$. Consequently, the left-hand side of (6.14) is bounded above by

$$(6.15) \quad \iint_{\mathbb{R}^2} f_n(z_1, z_2) d\nu(z_1) d\nu(z_2).$$

Let $n_0 \geq 1$ be a fixed integer to be specified later. Consider the integral (6.15) for $n := n_0$ as ν converges to μ in the weak topology of probability measures. Since the integral is a continuous functional over the integrands, there exists a neighborhood G of μ such that for any $\nu \in G$,

$$(6.16) \quad \iint_{\mathbb{R}^2} f_{n_0}(z_1, z_2) d\nu(z_1) d\nu(z_2) \leq \frac{\epsilon}{2} + \iint_{\mathbb{R}^2} f_{n_0}(z_1, z_2) d\mu(z_1) d\mu(z_2).$$

Moreover,

$$\lim_{n \rightarrow \infty} \iint_{\mathbb{R}^2} f_n(z_1, z_2) d\mu(z_1) d\mu(z_2) = \iint_{\substack{\{(z_1, z_2): \\ z_1 \leq z_0 \leq z_2, \\ (z_1, z_2) \notin I \times I\}}} \frac{1}{z_2 - z_1} d\mu(z_1) d\mu(z_2) \leq \mathcal{E}_0,$$

allowing us to choose n_0 such that

$$(6.17) \quad \iint_{\mathbb{R}^2} f_{n_0}(z_1, z_2) d\mu(z_1) d\mu(z_2) < \mathcal{E}_0 + \frac{\epsilon}{2}.$$

Combining (6.15), (6.16), and (6.17) completes the proof. $\square$

6.3.6. One variable very close to z_0 , the other far away.

LEMMA 6.9. *Let μ be as in Proposition 6.2. For each $\epsilon > 0$ there exists n_0 and an open neighborhood G of the measure μ such that for any $n \geq n_0$ and for any probability measure $\nu \in G$ on the real line, the following bound holds true:*

$$(6.18) \quad \iint_{\substack{\{(z_1, z_2): \\ z_1 < z_2, \\ |z_1 - z_0| < n^{-\frac{1}{4}} \log n, \\ z_2 \notin I\}}} \frac{1}{z_2 - z_1} d\nu(z_1) d\nu(z_2) \leq \epsilon.$$

PROOF. Note that when n is sufficiently large, the integrand is uniformly bounded above by a constant which is independent of n . Our strategy is to find the neighborhood G in such a way that the measure of the integration area is smaller than an arbitrary $\epsilon' > 0$.

Let $c > 0$ be sufficiently small that

$$\mu([z_0, z_0 + c]) < \epsilon'.$$

There exists a neighborhood G of μ such that for any $\nu \in G$ the following bound is valid:

$$\nu([z_0, z_0 + c]) < \epsilon'.$$

When n is sufficiently large then

$$\frac{\log n}{\sqrt[4]{n}} < c$$

which completes the proof. $\square$

6.3.7. *Conclusion of the proof of Proposition 6.2.* Let $\epsilon' > 0$ be a positive constant which will be fixed later. For this value of $\epsilon := \epsilon'$ we apply Lemmas 6.8 and 6.9. We define G to be the open neighborhood of the measure μ defined as the intersection of the neighborhoods provided by these lemmas.

By summing the contributions of all of the above cases it follows that there exists a sequence C_n which converges to zero and n_0 such that for any $n \geq n_0$ and any measure $\nu \in G$ which fulfills the local refinement condition (6.6) the following bound is true:

$$\mathbb{E}[\mathcal{E}_\nu(\mathbf{u}_n)] \leq \mathcal{E}_0 + 2\epsilon' + f_{\max}^2 |J| + C_n.$$

Since $\epsilon' > 0$ as well as the length of the interval J can be chosen to be arbitrarily small, Proposition 6.2 follows.

6.4. *Proposition 6.2 implies the upper bound in Assumption (B2).* For an integer $n \geq 1$, define the following random variables:

$$\Delta_n = \mathcal{E}_{\frac{1}{\sqrt{n}}; \mu_{\omega_n}}(\mathbf{u}_n) - \mathcal{E}_0,$$

$$\Delta_n^+ = \max(\Delta_n, 0),$$

$$\Delta_n^- = \max(-\Delta_n, 0),$$

such that $\Delta_n = \Delta_n^+ - \Delta_n^-$ and $\Delta_n^- \geq 0$, as well as $\Delta_n^+ \geq 0$.

Fix $\epsilon > 0$ and $\eta > 0$. Our goal is to show that

$$(6.19) \quad \limsup_{n \rightarrow \infty} \mathbb{P}[\Delta_n > \epsilon] = \limsup_{n \rightarrow \infty} \mathbb{P}[\Delta_n^+ > \epsilon] < \eta.$$

In the pursuit of this goal we may condition on events whose probabilities converge to 1, preserving the validity of our proof while offering greater analytical flexibility. We will perform such a conditioning twice. We start by setting $\delta > 0$ and G to be the parameters provided by Proposition 6.2 for $\epsilon' := \frac{1}{2}\epsilon\eta$.

Firstly, recall that the sequence of random measures (μ_{ω_n}) converges to μ_Ω in the weak topology of probability measures, in probability. (For a detailed discussion of this convergence, refer to the paragraph preceding Equation (6.4).) It follows that we can condition on the event that the global proximity condition (6.5) is satisfied for $\nu := \omega_n$:

$$\mu_{\omega_n} \in G.$$

Secondly, by (6.1), we can additionally condition on the event that the local refinement condition (6.6) holds true for $\nu := \mu_{\omega_n}$.

Under these two conditionings, Proposition 6.2 is applicable for $\nu := \omega_n$ and all sufficiently large values of n , thus by applying the expected value over ω_n it follows that

(6.20)
$$\limsup_{n \rightarrow \infty} \mathbb{E} \Delta_n \leq \limsup_{n \rightarrow \infty} \mathbb{E} [\mathcal{E}_{\mu_{\omega_n}}(\mathbf{u}_n)] - \mathcal{E}_0 < \epsilon' = \frac{1}{2}\epsilon\eta.$$

The random variables $\Delta_n^- \geq 0$ are uniformly bounded from above by $\mathcal{E}_0$. Thus,

$$\Delta_n^- \leq \frac{1}{2}\epsilon\eta + \mathcal{E}_0 \, \mathbb{1}_{[\Delta_n^- > \frac{1}{2}\epsilon\eta]},$$

and

$$\mathbb{E} \Delta_n^- \leq \frac{1}{2}\epsilon\eta + \mathbb{P}\left[\Delta_n^- > \frac{1}{2}\epsilon\eta\right] \mathcal{E}_0.$$

By applying (6.2) with $\epsilon' = \frac{1}{2}\epsilon\eta$ it follows that the second term on the right-hand side converges to zero. Consequently,

(6.21)
$$\limsup_{n \rightarrow \infty} \mathbb{E} \Delta_n^- \leq \frac{1}{2}\epsilon\eta.$$

By Markov’s inequality,

$$\mathbb{P}[\Delta_n > \epsilon] = \mathbb{P}[\Delta_n^+ > \epsilon] \leq \frac{\mathbb{E} \Delta_n^+}{\epsilon} = \frac{\mathbb{E} \Delta_n + \mathbb{E} \Delta_n^-}{\epsilon}.$$

Combining (6.20) and (6.21), we find that the right-hand side is asymptotically bounded by η , thus establishing (6.19) and completing the proof of Assumption (B2).

6.5. *Towards Assumption (B3). Cauchy transform in a random point is not too singular.* Recall that the modified Cauchy transform is defined in (5.1). This quantity $\mathbf{G}_{\rho_n}^+(z)$ is particularly effective in scenarios where ρ_n is part of a sequence of probability measures supported on an interval of the form

$$[-\Theta(\sqrt{n}), \Theta(\sqrt{n})],$$

which expands proportionally to $\sqrt{n}$, and where the argument $z = \Theta(\sqrt{n})$ scales in a similar manner.

In some cases, it is advantageous to dilate the n th element of the sequence (ρ_n) by a factor of $\frac{1}{\sqrt{n}}$, resulting in a new sequence of probability measures supported on a common compact interval. The modified Cauchy transform $\mathbf{G}_{\rho_n}^+(z)$ for the original measure is related to the following expression for the dilated measure:

(6.22)
$$\mathbf{G}_{\rho_n}^+(\sqrt{n}u) = \frac{1}{\sqrt{n}} \int \frac{1}{|x - u| + n^{-\frac{1}{2}}} \, \mathrm{d}D_{\frac{1}{\sqrt{n}}, \rho_n}(x),$$

where $D_{\frac{1}{\sqrt{n}}, \rho_n}$ denotes the dilation of the measure ρ_n by the factor $\frac{1}{\sqrt{n}}$.

The following result concerns the integral on the right-hand side of (6.22) and is crucial for establishing Assumption (B3). Heuristically, even if the modified Cauchy transform exhibits locally very large values, its behavior becomes significantly more regular when evaluated at a point perturbed by random noise.

LEMMA 6.10. *Let (ν_n) be a sequence of random probability measures on the real line, and let μ be a probability measure on the same line. Assume there exist real numbers $a < b$ such that the measure μ restricted to the interval $[a, b]$ is absolutely continuous, with a bounded density.*

Furthermore, assume that the CDFs of the measures (ν_n) converge to the CDF of μ at the prescribed speed:

$$(6.23) \quad \sup_{u \in [a, b]} \sqrt[4]{n} |F_{\nu_n}(u) - F_{\mu}(u)| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Let $\mathbf{Q}$ and $\mathbf{u}_n$ be as in the statement of Theorem 5.1. Then, for any $u_0 \in (a, b)$, we have that

$$(6.24) \quad \frac{1}{(\log n)^2} \int \frac{1}{|x - \mathbf{u}_n| + n^{-\frac{1}{2}}} d\nu_n(x) \xrightarrow[n \rightarrow \infty]{P} 0.$$

The remaining part of this section is devoted to the proof.

6.5.1. Conditioning. Our general strategy is to prove that the probability of the left-hand side of (6.24) exceeding some constant $\epsilon > 0$ converges to zero. For any constant $A > 0$, this probability is bounded above by the sum:

$$\mathbb{P}[|\mathbf{Q}| > A] + \mathbb{P}\left[\int \frac{1}{|x - \mathbf{u}_n| + n^{-\frac{1}{2}}} d\nu_n(x) > \epsilon (\log n)^2 \mid |\mathbf{Q}| \leq A\right].$$

We can choose $A > 0$ sufficiently large such that the first summand becomes arbitrarily small. In the remaining part of the proof we will show that the conditional probability in the second summand converges to zero as $n \rightarrow \infty$. Given our conditional setup, we can assume in the following that $\mathbf{Q}$ is a random variable supported on the interval $[-A, A]$, obtained by conditioning the standard normal random variable.

By employing a similar conditioning argument, we can invoke (6.23) to assume that

$$(6.25) \quad \sup_{u \in [a, b]} \sqrt[4]{n} |F_{\nu_n}(u) - F_{\mu}(u)| < 1$$

holds with probability 1 for all sufficiently large values of n .

6.5.2. Split the domain of integration. We will split the domain of integration in

$$(6.26) \quad \int_{\mathbb{R}} \frac{1}{|x - \mathbf{u}_n| + n^{-\frac{1}{2}}} d\nu_n(x)$$

into three parts, and investigate the asymptotic behavior of each part separately.

6.5.2.1. Easy deterministic estimate, very far from u_0 . We start with the first integral which contributes to (6.26), namely

$$\int_{\mathbb{R} \setminus [a, b]} \frac{1}{|x - \mathbf{u}_n| + n^{-\frac{1}{2}}} d\nu_n(x).$$

Let $c > 0$ be a positive constant satisfying

$$[u_0 - c, u_0 + c] \subset (a, b).$$

Given that the random perturbation $\mathbf{Q}$ is bounded, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, the following inequality holds uniformly for any $x \in \mathbb{R} \setminus [a, b]$ contributing to the integral:

$$|x - \mathbf{u}_n| \geq c.$$

Consequently, we can establish an almost sure upper bound for the integral:

$$(6.27) \quad \int_{\mathbb{R} \setminus [a, b]} \frac{1}{|x - \mathbf{u}_n| + n^{-\frac{1}{2}}} dv_n(x) \leq \frac{1}{c}.$$

This bound holds with probability 1 for all $n \geq n_0$.

6.5.2.2. Deterministic estimate, intermediate distance to u_0 . We consider now the second integral which contributes to (6.26). In this section, we will show that:

$$(6.28) \quad \int_{x \in [a, b] \text{ and } |x - u_0| > 2An^{-\frac{1}{4}}} \frac{1}{|x - \mathbf{u}_n| + n^{-\frac{1}{2}}} dv_n(x) = O(\log n).$$

For any x contributing to the integral (6.28) we have $|x - u_0| > 2An^{-\frac{1}{4}}$; on the other hand $|\mathbf{u}_n - u_0| \leq An^{-\frac{1}{4}}$; it follows that

$$|x - \mathbf{u}_n| \geq (|x - u_0|) - (|\mathbf{u}_n - u_0|) \geq \frac{|x - u_0|}{2}.$$

It follows that (6.28) is bounded from above by

$$(6.29) \quad 2 \int_{x \in [a, b] \text{ and } |x - u_0| > 2An^{-\frac{1}{4}}} \frac{1}{|x - u_0|} dv_n(x).$$

The integration area encompasses two distinct segments: one to the left and one to the right of u_0 . In the following discussion, we will focus on the segment to the right. The integral over the left segment can be derived using a similar approach.

We will utilize integration by parts, as outlined below:

$$\begin{aligned} & \int_{u_0 + 2An^{-\frac{1}{4}}}^b \frac{1}{x - u_0} dv_n(x) \\ &= \left[\frac{1}{x - u_0} F_{v_n}(x) \right]_{u_0 + 2An^{-\frac{1}{4}}}^b + \int_{u_0 + 2An^{-\frac{1}{4}}}^b \frac{1}{(x - u_0)^2} F_{v_n}(x) dx. \end{aligned}$$

An equivalent formula can be established for the corresponding integral over the same interval with respect to the measure μ . By subtracting these two formulas, and using the bound (6.25), we can derive:

$$\begin{aligned} & \left| \int_{u_0 + 2An^{-\frac{1}{4}}}^b \frac{1}{x - u_0} dv_n(x) - \int_{u_0 + 2An^{-\frac{1}{4}}}^b \frac{1}{x - u_0} d\mu(x) \right| \\ (6.30) \quad & \leq \frac{|F_{v_n}(b) - F_{\mu}(b)|}{b - u_0} + \frac{|F_{v_n}(u_0 + 2An^{-\frac{1}{4}}) - F_{\mu}(u_0 + 2An^{-\frac{1}{4}})|}{2An^{-\frac{1}{4}}} \\ & \quad + \int_{u_0 + 2An^{-\frac{1}{4}}}^b \frac{|F_{v_n}(x) - F_{\mu}(x)|}{(x - u_0)^2} dx = O(1) \end{aligned}$$

in the sense that there is a universal, deterministic constant which is an upper bound for the right-hand side.

We denote by C any upper bound on the density of the measure μ on the interval $[a, b]$. The integral which is the subtrahend on the left-hand side of (6.30) can be bounded as follows:

$$\int_{u_0+2An^{-\frac{1}{4}}}^b \frac{1}{x-u_0} d\mu(x) \leq C \int_{u_0+2An^{-\frac{1}{4}}}^b \frac{1}{x-u_0} dx = O(\log n).$$

The preceding considerations establish that the random integral (6.29) possesses a deterministic upper bound of $O(\log n)$. As a direct consequence, the claimed upper bound in (6.28) is also valid.

6.5.2.3. *Probabilistic estimate, very close to u_0 .* Consider the interval

$$I_n = \left[u_0 - \frac{2A}{\sqrt[4]{n}}, u_0 + \frac{2A}{\sqrt[4]{n}} \right].$$

Our goal in this section is to show that

$$(6.31) \quad \frac{1}{(\log n)^2} \int_{I_n} \frac{1}{|x - \mathbf{u}_n| + n^{-\frac{1}{2}}} dv_n(x) \xrightarrow[n \rightarrow \infty]{P} 0.$$

Let X_n be a random variable generated by the following two-step sampling procedure. First, we sample the random variables v_n and $\mathbf{Q}$. In the second step we sample X_n whose distribution is derived from the conditioning of the measure v_n to the interval I_n . In this way, the conditional expectation of the integral on the left-hand side of (6.31):

$$(6.32) \quad \mathbb{E} \left[\int_{I_n} \frac{1}{|x - \mathbf{u}_n| + n^{-\frac{1}{2}}} dv_n(x) | v_n \right] = v_n(I_n) \quad \mathbb{E} \left[\frac{1}{|X_n - \mathbf{u}_n| + n^{-\frac{1}{2}}} | v_n \right]$$

can be expressed in terms of the conditional expected value of a random variable involving X_n . In the following, we shall investigate each of the two factors on the right-hand side. Without loss of generality, we may assume that n is large enough so that $I_n \subset (a, b)$.

By Equation (6.25):

$$|v_n(I_n) - \mu(I_n)| \leq \frac{2}{\sqrt[4]{n}}.$$

As before, let C denote any upper bound on the density of μ on the interval (a, b) . It follows that

$$\mu(I_n) \leq \frac{4AC}{\sqrt[4]{n}}.$$

In this way, we found a deterministic upper bound

$$v_n(I_n) = O\left(\frac{1}{\sqrt[4]{n}}\right)$$

for the first factor on the right-hand side of (6.32).

In order to bound the second factor on the right-hand side of (6.32) we use conditioning over the measure v_n . With this conditioning, the second factor on the right-hand side of (6.32) can be seen as the integral

$$(6.33) \quad \mathbb{E} \left[\frac{1}{|X_n - \mathbf{u}_n| + n^{-\frac{1}{2}}} \right] = \int \frac{1}{|x| + n^{-\frac{1}{2}}} d\mu_{Z_n}(x)$$

over the (conditional) distribution of the random variable

$$Z_n = X_n - \mathbf{u}_n = (X_n - u_0) - \frac{1}{\sqrt[4]{n}}\mathbf{Q}.$$

First, notice that the absolute value of this random variable is almost surely bounded by

$$\frac{3A}{\sqrt[4]{n}}.$$

Second, since the probability distribution of Z_n is a convolution of:

- the distribution of $(X_n - u_0)$, and
- the Gaussian distribution with the standard deviation equal to $\frac{1}{\sqrt[4]{n}}$,

it follows that the distribution of Z_n is absolutely continuous, with the density bounded from above by that of a Gaussian measure, which is $\frac{\sqrt[4]{n}}{\sqrt{2\pi}}$.

By combining these observations it follows that the integral (6.33) is bounded from above by

$$\frac{2\sqrt[4]{n}}{\sqrt{2\pi}} \int_{n^{-\frac{1}{2}}}^{3An^{-\frac{1}{4}}} \frac{1}{x} \, dx = O(\log n \sqrt[4]{n}).$$

Our analysis thus far shows that the conditional expectation given in (6.32) is almost surely bounded above by a deterministic sequence that grows at most like $O(\log n)$. Consequently, the left-hand side of (6.31) converges to 0 in the L^1 norm. This also implies that the convergence in probability stated in (6.31) holds true.

6.5.3. *Conclusion of the proof of Lemma 6.10.* Our investigation of the integral (6.26) is complete. Equations (6.27), (6.28), and (6.31) establish the proof of Lemma 6.10.

6.6. *Assumptions of Theorem 4.3 imply Assumption (B3).* Our strategy is to apply Lemma 6.10 for the sequence of random measures (ν_n) , where

$$\nu_n := \mu_{\omega_n}$$

is the transition measure of the random continual diagram ω_n , and for $\mu := \mu_\Omega$ equal to the transition measure of the limit continual diagram Ω .

We begin by verifying that the assumptions of Lemma 6.10 are satisfied. Recall that we have already established (6.1), which demonstrates that the CDFs converge at the required rate. Moreover, Assumption (A3) ensures that the limit measure possesses a locally bounded density, as required. Thus, we conclude that Lemma 6.10 is applicable.

Upon applying this lemma and utilizing (6.22), we obtain the following result:

$$\frac{n^{\frac{4}{8}}}{(\log n)^2} \mathbf{G}_{\lambda^{(n)}}^+(\sqrt{n} \mathbf{u}_n) \xrightarrow[n \rightarrow \infty]{P} 0$$

which provides a significantly stronger bound than what is stipulated by Assumption (B3).

7. The first tool: The cumulative function of a tableau.

7.1. *Cumulative function of a tableau.* In our recent paper [18], we introduced the *cumulative function* for a Poissonized tableau T . This function, denoted as $F_T : \mathbb{R} \rightarrow [0, 1]$, is defined as follows:

$$F_T(u) = \inf\{z \in [0, 1] : u\text{-Ins}(T; z) > u\} \quad \text{for } u \in \mathbb{R};$$

where $u\text{-Ins}(T; z)$ represents the u -coordinate of the box $\text{Ins}(T; z)$, as defined in Section 1.3. If the infimum is taken over an empty set, we define $F_T(u) = 1$. For an example, see Figure 11.

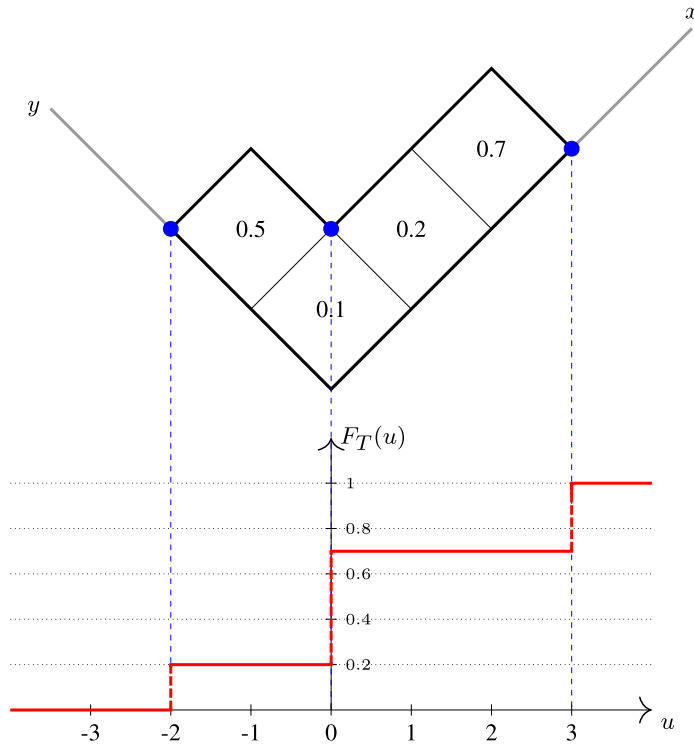


FIG. 11. A Poissonized tableau T shown in Russian coordinates. The red line depicts its cumulative function F_T . This zigzag line splits the infinite rectangle $\mathbb{R} \times [0, 1]$ into two regions: the northwest region and the southeast region.

Due to the monotonicity of Schensted row insertion, the following equivalence holds for any $u \in \mathbb{R}$ and $z \in [0, 1]$:

$$(7.1) \quad F_T(u) \leq z \iff \text{u-Ins}(T; z) \geq u.$$

This relationship implies that the cumulative function can be regarded as the inverse of the insertion function $[0, 1] \ni z \mapsto \text{u-Ins}(T; z) \in \mathbb{R}$. Specifically, the relationship between F_T and $\text{u-Ins}(T; \cdot)$ is analogous to that between the CDF of a measure and its quantile function. Consequently, the cumulative function is an ideal tool for proving the main results of this paper, as they all pertain to the function u-Ins .

7.2. Wider perspective: Cumulative function of a random tableau. Consider a Young diagram λ (chosen either randomly or deterministically) and let T be a random Poissonized tableau of shape λ . An intriguing problem in this context is the study of the asymptotics of the random function $u \mapsto F_T(u)$ as the number of boxes in λ approaches infinity.

7.2.1. Plancherel-distributed Poissonized tableaux. We consider a specific instance of this general problem by examining the random Plancherel-distributed Poissonized tableau T with n boxes (see Section 2.1.1). Figures 12 and 13 illustrate the results of a single Monte Carlo simulation, comparing three key functions:

- (i) the cumulative function F_T of the tableau (thick red line);
- (ii) the CDF K_λ of the transition measure of λ , where λ is the shape of T (thin blue line);
- (iii) the CDF of the rescaled arcsine law (smooth black line).

As $n \rightarrow \infty$, both curves (i) and (ii) converge to curve (iii). This convergence is well established:

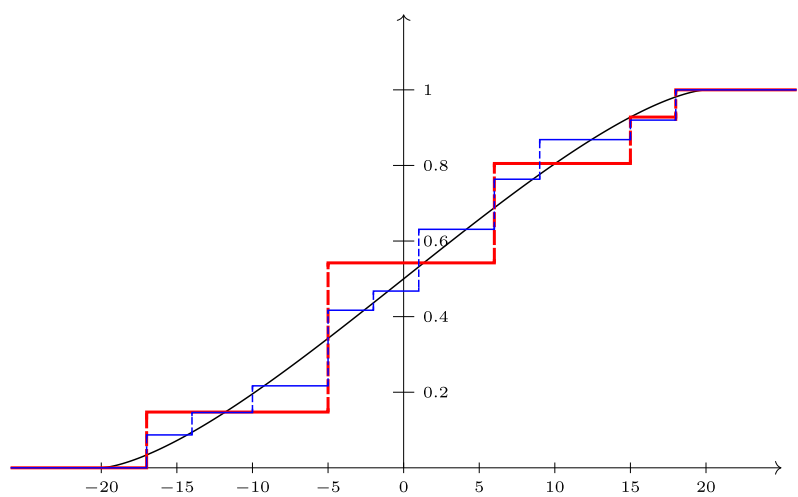


FIG. 12. Results of a single Monte Carlo simulation. The insertion tableau $T = P(w_1, \dots, w_n)$ with $n = 100$ boxes was generated using the RSK algorithm applied to a sequence of i.i.d. $U(0, 1)$ random variables. This figure displays two functions related to this tableau. The thin blue line represents the CDF K_λ of the transition measure for the shape λ of T . The thick red line shows the cumulative function F_T of the tableau T . For reference, the smooth black line depicts the CDF of the semicircle distribution, supported on the interval $[-2\sqrt{n}, 2\sqrt{n}]$.

- For curve (i), it is a reformulation of the result by Romik and the second named author (Theorem 1.4) about the asymptotic determinism of the Schensted insertion, as well as a special case of Theorem 4.1.
- For curve (ii), it is a classical result by Kerov on the transition measure of Plancherel-distributed diagrams [13, 14].

Intriguingly, the apparent rates of convergence differ between the two curves:

- (i) The observed rate appears to be of order $\Theta(n^{-\frac{1}{4}})$. While we lack a rigorous proof for this assertion, it aligns well with the findings presented in Section 10.
- (ii) The convergence seems to be much faster, with the rate appearing to be of a significantly smaller order, $O(n^{-\frac{1}{2}+\epsilon})$, as conjectured in Conjecture 5.4.

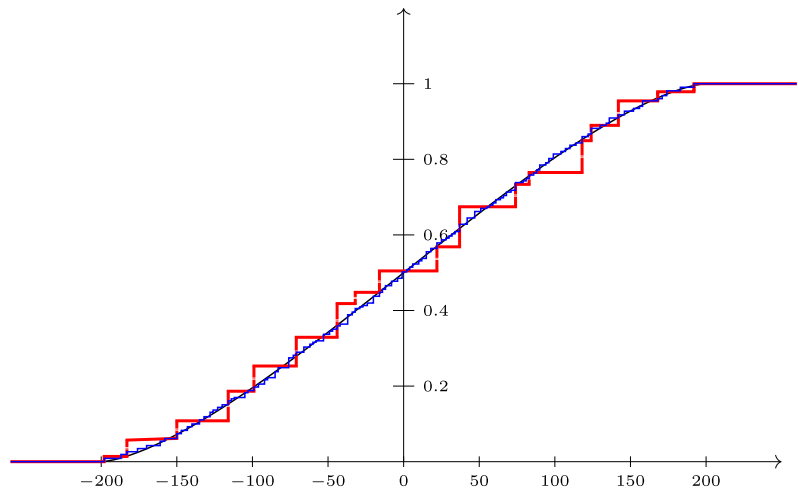


FIG. 13. An analogue of Figure 12 for $n = 10^4$ boxes.

7.2.2. Future research directions. An intriguing avenue for future research is the investigation of the expected number of discontinuities in the cumulative function F_T and their spatial distribution as the number of boxes tends to infinity. This study is closely related to the density of infinite geodesics in the percolation tree associated with Plancherel-distributed infinite standard tableaux on one side, and the geometry of bumping routes and bumping forests on the other side [17, 24]. Understanding these discontinuities could provide deeper insights into the asymptotic behaviors of random tableaux. We aim to address this problem in forthcoming work.

7.3. Single-point evaluation of the cumulative function. While a comprehensive analysis of the statistical properties of the cumulative function F_T is beyond the scope of this paper, we focus on a more targeted approach. Our primary interest lies in understanding the probability distribution of $F_T(u)$, where the cumulative function is evaluated at a single point u . This approach provides valuable insights while remaining computationally manageable.

In our recent work [18], we derived explicit combinatorial formulas for the cumulants of this probability distribution in terms of Kerov's transition measure μ_λ of the tableau's shape λ . For the purposes of this paper, we will utilize a few key results from that work, which are particularly useful for the asymptotic analysis, as detailed in Lemma 7.1 below.

7.3.1. Cumulants of a probability distribution. Recall that cumulants [11, 36] are quantities that provide an alternative to moments for characterizing probability distributions. For a random variable X , its n th cumulant $\kappa_n = \kappa_n(X)$ is defined as the coefficient in the Taylor expansion of the logarithm of the moment-generating function:

$$\log \mathbb{E}[e^{tX}] = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!}.$$

The first cumulant κ_1 is the mean, the second cumulant κ_2 is the variance, and higher-order cumulants provide information about the shape of the distribution. The standard normal distribution is uniquely characterized by its cumulants: its first cumulant (mean) is 0, its second cumulant (variance) is 1, and all higher-order cumulants (κ_n for $n \geq 3$) are zero.

7.3.2. Cumulants of the cumulative function.

LEMMA 7.1 ([18], Equations (3.2) and (3.3) and Corollary 3.4). *Let λ be a fixed Young diagram and T be a uniformly random Poissonized tableau of shape λ . Let $u \in \mathbb{R}$ be fixed. The expected value and the variance of $F_T(u)$ are given by:*

$$(7.2) \quad \mathbb{E}F_T(u) = \sum_{x_1 \leq u} \mu_\lambda(x_1),$$

$$(7.3) \quad \text{Var } F_T(u) = \sum_{\substack{x_1 \leq u \\ x_2 > u}} \frac{1}{x_2 - x_1 + 1} \mu_\lambda(x_1) \mu_\lambda(x_2).$$

For any $k \geq 1$, the k th cumulant of $F_T(u)$ satisfies the following bound:

$$|\kappa_k(F_T(u))| \leq (k-1)! [\mathbf{G}_\lambda^+(u)]^{k-1},$$

where $\mathbf{G}_\lambda^+(u)$ was defined in Section 5.2.

8. The second tool: The double cumulative function of a Young diagram. For a Young diagram λ , we introduce the *double cumulative function* $\mathcal{F}_\lambda(u, z)$, which serves as a two-dimensional analogue of the usual CDF from probability theory. This function is a tool for analyzing the insertion of a number into a uniformly random Poissonized tableau of shape λ . When a number is inserted into such a tableau, the double cumulative function $\mathcal{F}_\lambda(u, z)$ combines information about the new box’s position, the inserted number, and the associated probability. This provides a comprehensive analytical framework for studying Schensted insertion into Poissonized tableaux.

8.1. *Definition.* Let λ be a fixed Young diagram, and let T be a uniformly random Poissonized tableau of shape λ . We define the *double cumulative function* of λ as:

$$\mathcal{F}_\lambda : \mathbb{R} \times [0, 1] \rightarrow [0, 1].$$

This function is given by:

$$\mathcal{F}_\lambda(u, z) = \mathbb{P}(\text{u-Ins}(T; z) \geq u) = \mathbb{P}(F_T(u) \leq z) \quad \text{for } u \in \mathbb{R}, z \in [0, 1].$$

It represents the probability that one of the two equivalent conditions (7.1) holds true, or equivalently, that the point (u, z) belongs to the northwest region with respect to the plot of the cumulative function $u \mapsto F_T(u)$ (see Figure 11). For an example corresponding to $\lambda = (3, 1)$, see Figure 14.

8.2. *Partitioning of the domain.* The double cumulative function $\mathcal{F}_\lambda(u, z)$ exhibits discontinuities along vertical lines where the first coordinate equals the u -coordinates of the concave corners of the Young diagram λ . These vertical lines partition the infinite rectangle

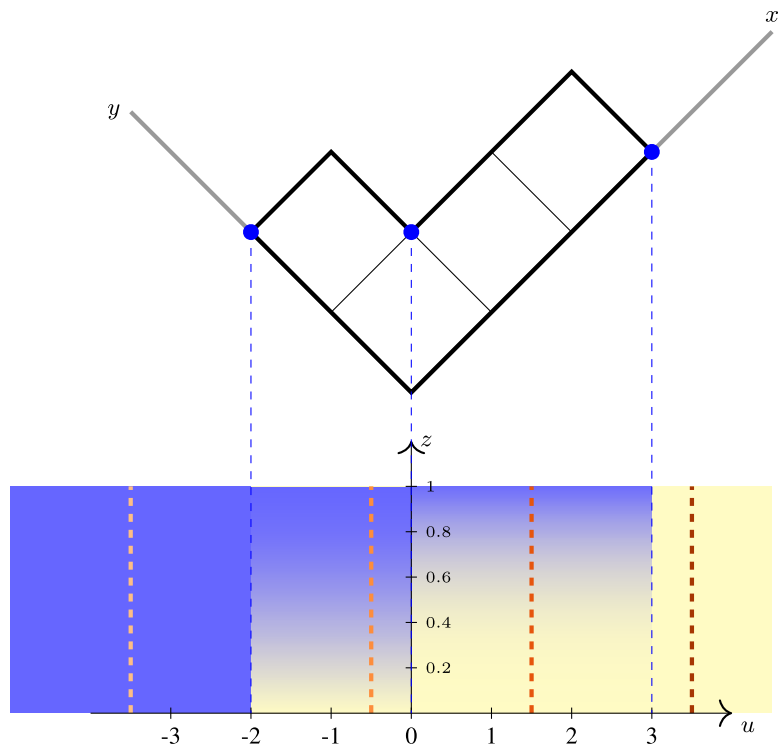


FIG. 14. Density plot of the double cumulative function $\mathcal{F}_\lambda(u, z)$ for the Young diagram $\lambda = (3, 1)$. Bright yellow indicates areas where the insertion function takes values close to 0, while dark blue indicates values close to 1. The plots of the double cumulative function along the four vertical dashed lines are shown in Figure 15.

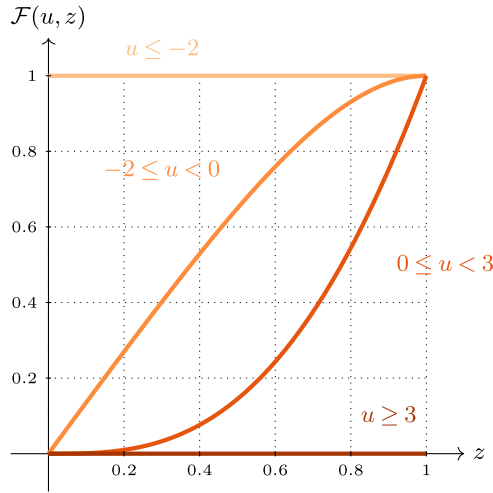


FIG. 15. Double cumulative function plots for Young diagram $\lambda = (3, 1)$. This figure extends the example from Figure 14, showing $z \mapsto \mathcal{F}_\lambda(u, z)$ for various values of u . These plots illustrate how the function behaves across different u values, providing insight into the insertion probabilities for Poissonized tableaux of this shape.

$\mathbb{R} \times [0, 1]$ into a series of rectangles, which may be finite or semi-infinite. Within each rectangle, $\mathcal{F}_\lambda(u, z)$ depends solely on the second variable z . This behavior will be further described in Section 8.3. See Figure 15 for an illustration.

8.3. *Properties for fixed u .* For each fixed u , the function

$$(8.1) \quad [0, 1] \ni z \mapsto \mathcal{F}_\lambda(u, z)$$

serves as the CDF for the random variable $F_T(u)$. Consequently, $\mathcal{F}_\lambda$ is weakly increasing with respect to the second variable z . Results from our recent paper [18] provide information about this random variable $F_T(u)$, which we will use later in Proposition 9.1 and Theorem 10.1 to analyze the behavior of $\mathcal{F}_\lambda$ for fixed values of the first variable.

REMARK 8.1. Since the distribution of the random variable $F_T(u)$ is a mixture of some beta distributions $\text{Beta}(\alpha, \beta)$ over integers $\alpha, \beta \geq 1$ with $\alpha + \beta = |\lambda| + 2$; it follows that the function (8.1) is a polynomial, with degree bounded from above by $|\lambda|$.

8.4. *Properties for fixed z .* For a fixed $z \in [0, 1]$, the tail function:

$$(8.2) \quad \mathbb{R} \ni u \mapsto 1 - \mathcal{F}_\lambda(u, z) = \mathbb{P}(\text{u-Ins}(T; z) < u)$$

is *almost* equal to the CDF of the random variable $\text{u-Ins}(T; z)$. This property will be crucial for proving Theorems 4.1 and 5.1 in Section 11. (Strictly speaking, the standard definition of CDF involves a *strict* inequality on the right-hand side of (8.2) while we employ a *weak* inequality. This minor discrepancy will not affect correctness of our arguments presented in Section 11.) A simple consequence of this observation is that the double cumulative function is a weakly decreasing function of the first variable.

The probability distribution of $\text{u-Ins}(T; z)$ can be directly inferred from plots such as those in Figure 15. For a fixed z , the vertical unit interval is partitioned by the plotted curves into sub-intervals. The lengths of these sub-intervals correspond to the probabilities associated with specific concave corners of the Young diagram. This visual representation allows for an intuitive understanding of how the insertion process behaves for different values of z .

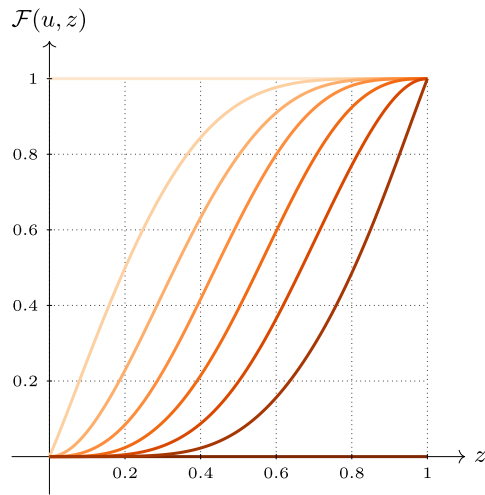


FIG. 16. Double cumulative function plots for the staircase Young diagram $\lambda = (6, 5, 4, 3, 2, 1)$. This diagram features 7 concave corners, whose u -coordinates partition the real line into 8 intervals (finite or infinite). The topmost bright curve (constant at 1) corresponds to the leftmost (infinite) interval, while the bottommost dark curve (constant at 0) represents the rightmost (infinite) interval. This figure is analogous to Figure 15.

8.5. *Navigating the two viewpoints.* Our analytical strategy employs a dual approach to comprehensively examine the double cumulative function and its implications for the insertion process. We begin by fixing u and use the results from [18]. The goal is to derive ‘two-dimensional’ results concerning $\mathcal{F}_\lambda(u, z)$ for an arbitrary choice of u and z . Subsequently, we shift our focus to fixed values of z , allowing us to deduce vital information about the CDF of the insertion function $\text{u-Ins}(T; z)$.

8.6. *Examples.*

8.6.1. *The example for $\lambda = (3, 1)$.* The double cumulative function for $\lambda = (3, 1)$ is visualized in Figure 14 as a density plot. The cumulative function plots along the four vertical lines are shown in Figure 15.

8.6.2. *Staircase tableaux.* Expanding on our earlier discussion in Section 4.4.1, we present visualizations of the double cumulative function for two distinct staircase diagrams in Figures 16 and 17.

9. First-order asymptotics of the double cumulative function. The phenomenon we aim to elucidate is vividly exemplified by large staircase diagrams, as depicted in Figure 17. Let us examine the density plot of the double cumulative function $\mathcal{F}_\lambda : \mathbb{R} \times [0, 1] \rightarrow [0, 1]$ for a Young diagram λ . The domain, an infinite rectangle $\mathbb{R} \times [0, 1]$, is bisected by the graph of the cumulative function $K_\lambda : \mathbb{R} \rightarrow [0, 1]$, delineating two distinct regions:

- the southeast region, characterized by small values of $\mathcal{F}_\lambda$ (approaching 0),
- the northwest region, distinguished by large values of $\mathcal{F}_\lambda$ (nearing 1).

The following proposition formalizes this observation:

PROPOSITION 9.1. *Consider a sequence $(\lambda^{(n)})$ of random Young diagrams. We posit the existence of a probability measure ν on $\mathbb{R}$ such that for each $u \in \mathbb{R}$ where the cumulative distribution function F_ν is continuous, the following convergence in probability holds:*

$$K_{\lambda^{(n)}}(\sqrt{n}u) \xrightarrow[n \rightarrow \infty]{P} F_\nu(u).$$

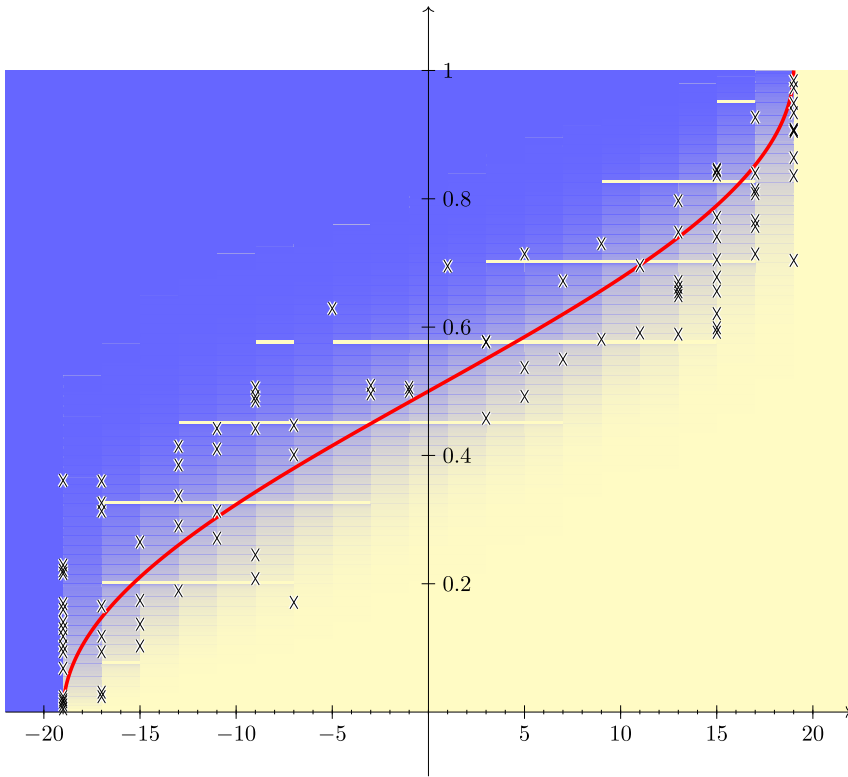


FIG. 17. Double cumulative function visualization for the staircase Young diagram $\lambda = (19, 18, \dots, 1)$. This figure extends the analysis presented in Figure 14, illustrating the behavior for a larger staircase shape. The crosses represent points $(u\text{-Ins}(T; z), z)$, where z is sampled from the uniform distribution $U(0, 1)$, and T is an independently sampled random Poissonized tableau of shape λ . The thick red curve depicts the CDF of the (dilated) arcsine law, which is the limit measure of transition measures for large staircase tableaux.

For $u \in \mathbb{R}$ and $z \in [0, 1]$:

If $z < F_\mu(u^-) = \lim_{v \rightarrow u^-} F_\mu(v)$, then:

$$\mathcal{F}_{\lambda(n)}(\sqrt{n}u, z) \xrightarrow[n \rightarrow \infty]{P} 0.$$

Conversely, if $z > F_\mu(u)$, then

$$\mathcal{F}_{\lambda(n)}(\sqrt{n}u, z) \xrightarrow[n \rightarrow \infty]{P} 1.$$

We begin with a lemma concerning the double cumulative function $\mathcal{F}_\lambda$ in a deterministic setting. The key idea is that when a point (u, z) lies below the graph of K_λ and is sufficiently separated from it both vertically and horizontally, there exists an upper bound on the value of $\mathcal{F}_\lambda(u, z)$.

LEMMA 9.2. Let λ be a Young diagram, and consider real numbers $u' < u$ and z satisfying:

$$0 \leq z < K_\lambda(u').$$

Then the following inequality holds:

$$0 \leq \mathcal{F}_\lambda(u, z) \leq \frac{1}{(u - u')(K_\lambda(u') - z)^2}.$$

PROOF. Let T be a uniformly random Poissonized tableau of shape λ . The explicit formula (7.3) for the variance of $F_T(v)$ implies:

$$\int_{u'}^u \text{Var } F_T(v) \, dv \leq 1.$$

By the mean value theorem for integrals, there exists $u'' \in [u', u]$ such that:

$$\text{Var } F_T(u'') \leq \frac{1}{u - u'}.$$

From (7.2), we know the mean value satisfies:

$$\mathbb{E} F_T(u'') = K_\lambda(u'') > z.$$

This allows us to apply the Bienaymé–Chebyshev inequality:

$$\begin{aligned} 0 \leq \mathcal{F}_\lambda(u, z) &\leq \mathcal{F}_\lambda(u'', z) = \mathbb{P}(F_T(u'') \leq z) \\ &\leq \mathbb{P}(|F_T(u'') - \mathbb{E} F_T(u'')| \geq \mathbb{E} F_T(u'') - z) \\ &\leq \frac{\text{Var } F_T(u'')}{(\mathbb{E} F_T(u'') - z)^2} \leq \frac{1}{(u - u')(K_\lambda(u'') - z)^2}, \end{aligned}$$

as required. $\square$

PROOF OF PROPOSITION 9.1. We begin by addressing the first part of the claim, assuming $z < F_\mu(u^-)$. There exists $u' < u$ such that $F_v(u') > z$. Without loss of generality, we may assume that u' is a point of continuity for F_v , ensuring:

$$K_{\lambda(n)}(\sqrt{n}u') \xrightarrow[n \rightarrow \infty]{P} F_v(u').$$

Applying Lemma 9.2, we obtain:

$$0 \leq \mathcal{F}_{\lambda(n)}(\sqrt{n}u, z) \leq \begin{cases} \frac{1}{(u - u')\sqrt{n}(K_{\lambda(n)}(\sqrt{n}u') - z)^2} & \text{if } K_{\lambda(n)}(\sqrt{n}u') > z, \\ 1 & \text{otherwise.} \end{cases}$$

The random variable on the right-hand side converges in probability to zero, completing the proof of the first part.

The second part of the claim is established through a parallel line of reasoning, albeit with an adjustment of certain inequality directions. This modification is applied to the function:

$$(u, z) \mapsto 1 - \mathcal{F}_{\lambda(n)}(\sqrt{n}u, z)$$

which pertains to the complementary events. $\square$

10. Fine asymptotics of the double cumulative function. The phenomenon we aim to describe is exemplified by large staircase diagrams, as vividly depicted in Figure 17. As previously discussed in Section 9, the domain of the density plot for the double cumulative function $\mathcal{F}_\lambda : \mathbb{R} \times [0, 1] \rightarrow [0, 1]$ is partitioned by the graph of the cumulative function $K_\lambda : \mathbb{R} \rightarrow [0, 1]$. This partition delineates two distinct regions: the southeast region, characterized by values approaching 0, and the northwest region, distinguished by values nearing 1.

Of particular intrigue is the narrow band immediately adjacent to the plot of K_λ , where the transition between these regions unfolds. As the magnitude of λ increases, the relative width of this transition zone converges to zero, creating a fascinating boundary layer phenomenon. The crux of this section, encapsulated in Theorem 10.1, reveals a remarkable property: within this transition area, the double cumulative function can be approximated by the CDF of the standard normal distribution.

10.1. *New coordinate system.* We employ the notations from Theorem 5.1. To provide a formal, quantitative description of the behavior of $\mathcal{F}_{\lambda^{(n)}}$, we introduce a new coordinate system on its domain $\mathbb{R} \times [0, 1]$. The new coordinates (q, ξ) correspond to a point $(U_{n,q}, Z_{n,q}) \in \mathbb{R} \times [0, 1]$, where:

$$\begin{aligned} U_{n,q} &= \sqrt{n}u_0 + \sqrt[4]{n}q, \\ Z_{n,q} &= z_0 + \frac{\xi}{\sqrt[4]{n}}. \end{aligned}$$

This coordinate system is centered around the point $(\sqrt{n}u_0, z_0)$ that represents a first-order approximation for the insertion $T^{(n)} \leftarrow z_0$. Notably, the units of this system are scaled in proportion to $\sqrt[4]{n}$ or its inverse, which aligns with the scale of problem.

10.2. *The Gaussian profile of the double cumulative function.* The following theorem encapsulates the behavior of the double cumulative function in the transition area:

THEOREM 10.1. *Under the assumptions of Theorem 5.1*

$$(10.1) \quad \mathcal{F}_{\lambda^{(n)}}(U_{n,q}, Z_{n,\xi}) \xrightarrow[n \rightarrow \infty]{P} \Phi\left(\frac{\xi - f_0q}{\sqrt{\mathcal{E}}}\right)$$

for each $q, \xi \in \mathbb{R}$, where $\Phi(x)$ denotes the CDF of the standard normal distribution, given by

$$(10.2) \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

The remaining part of this section is devoted to the proof.

REMARK 10.2. We shall now examine the geometric properties of the surface defined by the mapping

$$(q, \xi) \mapsto \Phi\left(\frac{\xi - f_0q}{\sqrt{\mathcal{E}}}\right),$$

which emerges as the limiting function on the right-hand side of equation (10.1). A notable characteristic of this surface is that its level curves are straight lines, expressible in the form

$$\xi = f_0q + c,$$

where c denotes a constant. These lines are parallel to the tangent line to the local approximation of the graph of $K_{\lambda^{(n)}}$ given by the Assumption (B1) in Theorem 5.1. The aforementioned geometric property is visually discernible in the illustrative example presented in Figure 17.

10.3. *Case study: Large staircase tableaux.* We revisit the staircase tableaux example introduced in Section 4.4.1, now examining it through the lens of Theorem 10.1.

A consequence of Theorem 10.1 is that for a large Young diagram λ and fixed u , the function $z \mapsto \mathcal{F}_{\lambda}(u, z)$ can be approximated by the CDF of a normal distribution. The parameters of this approximating distribution—mean and variance—are determined by equations (7.2) and (7.3), respectively.

A key insight emerges for staircase tableaux: the variance, up to a geometric scaling factor, closely approximates the interaction energy (4.7) of the arcsine distribution. This distribution arises as the limit transition measure for large staircase diagrams (see Section 4.4.1). Notably, for this measure, the interaction energy remains constant across the entire spectrum of valid u values.

Figure 16 provides visual corroboration of this concept. The depicted family of curves resembles a series of shifted plots of a single scaled CDF of the standard normal distribution. This aligns with our theoretical expectations of the normal distribution approximation with a fixed variance. A similar phenomenon is observed in Figure 17: for a wide range of u values, the color profile along a fixed u coordinate appears similar, up to a vertical shift.

Moreover, for large staircase diagrams, the mean value of the approximating normal distribution, given by (7.2), is approximately the CDF of the aforementioned arcsine measure. This relationship is visually evident in Figure 17, where the region of significant color change closely follows the plot of the arcsine measure’s CDF.

10.4. *Normalized cumulative function.* For an integer $n \geq 1$ and $q \in \mathbb{R}$ we consider the random variable

$$X_{n,q} = \frac{\sqrt[4]{n}}{\sqrt{\mathcal{E}}} \left[F_{T^{(n)}}(U_{n,q}) - \left(z_0 + \frac{f_0 q}{\sqrt[4]{n}} \right) \right] = \frac{\sqrt[4]{n}}{\sqrt{\mathcal{E}}} [F_{T^{(n)}}(U_{n,q}) - z_0] - \frac{f_0 q}{\sqrt{\mathcal{E}}}.$$

This random variable is the result of an affine transformation applied to $F_{T^{(n)}}(U_{n,q})$, which is the cumulative function of $T^{(n)}$ evaluated at the point with the new coordinate q . As we shall see, the constants in this affine transformation were chosen in such a way that $X_{n,q}$ behaves asymptotically like a standard normal random variable.

The double cumulative function $\mathcal{F}_{\lambda^{(n)}}$ is intimately linked to $X_{n,q}$ through the following lemma:

LEMMA 10.3. *For any $q, \xi \in \mathbb{R}$*

$$\mathcal{F}_{\lambda^{(n)}}(U_{n,q}, Z_{n,\xi}) = \mathbb{P} \left[X_{n,q} \leq \frac{\xi - f_0 q}{\sqrt{\mathcal{E}}} \right].$$

PROOF. The value of the double cumulative function concerns the insertion $T^{(n)} \leftarrow Z_{n,\xi}$ and represents the probability that one of the following equivalent conditions occurs:

- The insertion position in the transformed coordinates is greater than q .
- The original coordinate of the insertion exceeds $U_{n,q}$.
- The cumulative function $F_{T^{(n)}}$ evaluated at $U_{n,q}$ is less than or equal to $Z_{n,\xi}$.
- The random variable $X_{n,q}$ is at most $\frac{\xi - f_0 q}{\sqrt{\mathcal{E}}}$.

Formally, this can be expressed as:

(10.3)

$$\begin{aligned} \frac{\text{u-Ins}(T^{(n)}; Z_{n,\xi}) - \sqrt{nu}u_0}{\sqrt[4]{n}} \geq q &\iff \text{u-Ins}(T^{(n)}; Z_{n,\xi}) \geq U_{n,q} \iff \\ &F_{T^{(n)}}(U_{n,q}) \leq Z_{n,\xi} \iff X_{n,q} \leq \frac{\xi - f_0 q}{\sqrt{\mathcal{E}}}, \end{aligned}$$

see (7.1). $\square$

10.5. *Randomized probability distribution of $X_{n,q}$.* Our objective is to analyze the probability distribution of $X_{n,q}$. However, the assumptions in Theorem 5.1 prove insufficient to draw conclusions about its behavior for an arbitrary constant $q \in \mathbb{R}$ (see Remark 5.2). To overcome this limitation, we introduce randomness to the spatial parameter q .

10.5.1. *Randomized approach.* Instead of examining $X_{n,q}$ for a fixed transformed coordinate q , we study $X_{n,\mathbf{Q}}$, where $\mathbf{Q}$ is a standard Gaussian random variable independent from $T^{(n)}$. Without loss of generality, we assume that this random variable coincides with the Gaussian random variable from Theorem 5.1.

This probabilistic approach allows us to leverage the power of probabilistic methods by considering average behavior over a distribution of spatial parameters, and to shift our goal from proving a result that holds for each value of q to demonstrating that the probability of our claimed result converges to 1 for a random choice of q .

10.5.2. *Formal framework.* Let $\mathfrak{F}_n$ denote the σ -field generated by the random variables $\mathbf{Q}$ and $\lambda^{(n)}$. Our target result is encapsulated in the following result.

PROPOSITION 10.4. *The conditional probability distribution of the random variable $X_{n,\mathbf{Q}}$, given $\mathfrak{F}_n$, converges in probability to the standard Gaussian measure. More precisely, the supremum distance between the corresponding CDFs converges in probability to 0:*

$$(10.4) \quad \sup_t |\mathbb{P}(X_{n,\mathbf{Q}} \leq t | \mathfrak{F}_n) - \Phi(t)| \xrightarrow[n \rightarrow \infty]{P} 0,$$

where $\Phi(t)$ is the CDF of the standard normal distribution.

This result formalizes the convergence of the conditional probability distribution, which is a random probability measure on the real line, to the standard Gaussian measure. The use of the supremum distance between cumulative distribution functions provides a rigorous metric for this convergence.

The proof of this result is postponed to Section 10.5.6.

10.5.3. *Two spacial coordinate systems.* In the proof of Proposition 10.4, it is necessary to employ the assumptions from Theorem 5.1, which utilize slightly different notation. Recall that the random variable $\mathbf{u}_n$, defined in (5.2), was used in Theorem 5.1 to refer to the rescaled diagram $\omega_n = \omega_{\frac{1}{\sqrt{n}}\lambda^{(n)}}$. This variable is clearly related to the random variable $U_{n,\mathbf{Q}}$ by a simple scaling:

$$(10.5) \quad \mathbf{u}_n = \frac{1}{\sqrt{n}} U_{n,\mathbf{Q}}.$$

Since $U_{n,\mathbf{Q}}$ will be used to refer to the original, nonrescaled diagram $\lambda^{(n)}$, it follows that $\mathbf{u}_n$ and $U_{n,\mathbf{Q}}$ represent the same point in the diagram but in two different coordinate systems.

10.5.4. *Conditional moments and conditional cumulants.* The proof strategy for Proposition 10.4 employs an adaptation of the method of moments to a conditional framework. Rather than directly examining the conditional moments $\mathbb{E}[X_{n,\mathbf{Q}}^k | \mathfrak{F}_n]$, we shall pursue a more nuanced approach by focusing on the *conditional cumulants* of $X_{n,\mathbf{Q}}$. This conditional setup can be intuitively understood as fixing the values of the spacial variable $\mathbf{Q}$ and the Young diagram $\lambda^{(n)}$. Consequently, Lemma 7.1 becomes applicable, the implications of which we shall elucidate below.

Formula (7.2) from Lemma 7.1 and the relationship (10.5) imply that the conditional expected value is given by

$$(10.6) \quad \begin{aligned} \mathbb{E}[X_{n,\mathbf{Q}} | \mathfrak{F}_n] &= \frac{\sqrt[4]{n}}{\sqrt{\mathcal{E}}} [\mathbb{E}[F_{T^{(n)}}(U_{n,\mathbf{Q}}) | \mathfrak{F}_n] - z_0] - \frac{f_0 \mathbf{Q}}{\sqrt{\mathcal{E}}} \\ &= \frac{\sqrt[4]{n}}{\sqrt{\mathcal{E}}} [K_{\omega_n}(\mathbf{u}_n) - z_0] - \frac{f_0 \sqrt[4]{n}(\mathbf{u}_n - u_0)}{\sqrt{\mathcal{E}}} \xrightarrow[n \rightarrow \infty]{P} 0 \end{aligned}$$

which converges in probability to zero by Assumption (B1).

Formula (7.3) and the relationship (10.5) imply that the conditional variance is given by

(10.7)
$$\text{Var}[X_{n,\mathbf{Q}}|\mathfrak{F}_n] = \frac{\sqrt[2]{n}}{\mathcal{E}} \text{Var } F_{T^{(n)}}(U_{n,\mathbf{Q}}) \xrightarrow[n \rightarrow \infty]{P} 1,$$

where the convergence is a consequence of the Assumption (B2).

For $k \geq 3$ the absolute value of the corresponding cumulant can be bounded thanks to Lemma 7.1 as follows:

(10.8)
$$\begin{aligned} |\kappa_k[X_n|\mathfrak{F}_n]| &= \left(\frac{\sqrt[4]{n}}{\sqrt{\mathcal{E}}}\right)^k |\kappa_k[F_{T^{(n)}}(U_{n,\mathbf{Q}})|\mathfrak{F}_n]| \\ &\leq \left(\frac{\sqrt[4]{n}}{\sqrt{\mathcal{E}}}\right)^k (k-1)! [\mathbf{G}_\lambda^+(U_{n,\mathbf{Q}})]^{k-1} \xrightarrow[n \rightarrow \infty]{P} 0 \end{aligned}$$

which converges in probability to 0 by Assumption (B3).

The fundamental implication of equations (10.6)–(10.8) is that the sequence of conditional cumulants of $X_{n,\mathbf{Q}}$ converges in probability to that of the standard normal distribution as $n \rightarrow \infty$.

10.5.5. *Finite number of moments is enough.*

LEMMA 10.5. *For any $\epsilon > 0$, there exist an integer $n_0 \geq 1$ and $\delta > 0$ such that the following holds:*

Let μ be an arbitrary probability measure on the real line whose first n_0 moments (or cumulants) exist and are δ -close to the corresponding moments (or cumulants) of the standard normal distribution $N(0, 1)$. Then, the supremum distance between the CDF of μ and that of the standard normal distribution is less than ϵ . Formally,

(10.9)
$$\sup_{t \in \mathbb{R}} |F_\mu(t) - \Phi(t)| < \epsilon.$$

PROOF. We proceed by contradiction. Suppose the statement were false. Then there would exist a sequence of probability measures that converges in moments to the standard normal distribution but does not converge weakly to it. However, this is impossible, as the normal distribution is uniquely determined by its moments (a property known as moment determinacy). The contradiction establishes the lemma. $\square$

10.5.6. *Proof of Proposition 10.4.* PROOF OF PROPOSITION 10.4. Let $\epsilon > 0$ be arbitrary, let n_0 and δ be the values provided by Lemma 10.5. Equations (10.6)–(10.8) imply that the probability that the first n_0 conditional cumulants of X_n are δ -close to their counterparts for the standard normal distribution converges to 1. Consequently, by Lemma 10.5, we have that the probability of the event

$$\sup_t |\mathbb{P}(X_{n,\mathbf{Q}} \leq t | \mathfrak{F}_n) - \Phi(t)| < \epsilon$$

converges to 1 as $n \rightarrow \infty$, as required. $\square$

10.6. *Conditioning on a bounded interval.* The conclusion of Proposition 10.4 remains valid under additional conditioning. Specifically, for any constants $q_1 < q_2$, we may condition on the event $\{q_1 < \mathbf{Q} < q_2\}$ without affecting the lemma’s result. Consequently, we can proceed with our analysis under the assumption that the random variable $\mathbf{Q}$ is confined to the interval (q_1, q_2) without loss of generality.

10.7. *Conditional CDF of $X_{n,\mathbf{Q}}$.* The following result bears a strong resemblance to Proposition 10.4. However, a key distinction lies in the nature of the conditional probabilities considered:

- In Proposition 10.4, the probability is conditioned on both the choice of $\lambda^{(n)}$ and $\mathbf{Q}$.
- In contrast, Proposition 10.6 considers a probability conditioned solely on the choice of $\lambda^{(n)}$.

PROPOSITION 10.6. *For each $t \in \mathbb{R}$*

$$(10.10) \quad \mathbb{P}[X_{n,\mathbf{Q}} \leq t | \lambda^{(n)}] \xrightarrow[n \rightarrow \infty]{P} \Phi(t).$$

PROOF. The difference of the random variable on the left-hand side of (10.10), and the hypothetical limit:

$$(10.11) \quad \mathbb{P}[X_{n,\mathbf{Q}} \leq t | \lambda^{(n)}] - \Phi(t) = \mathbb{E}[\mathbb{P}[X_{n,\mathbf{Q}} \leq t | \mathfrak{F}_n] - \Phi(t) | \lambda^{(n)}]$$

can be obtained by taking the conditional expected value in two steps. Its L^1 norm is bounded from above by the L^1 norm of the random variable within the most external conditional expected value on the right-hand side, which is

$$\mathbb{E}[|\mathbb{P}[X_{n,\mathbf{Q}} \leq t | \mathfrak{F}_n] - \Phi(t)|]$$

which converges to zero by Proposition 10.4 $\square$

10.8. *The consequences of monotonicity of the double cumulative function.* The double cumulative function $\mathcal{F}_{\lambda^{(n)}}$ exhibits weak monotonicity with respect to its first coordinate. Consequently, we can establish a lower bound for (10.1) by averaging over the random variable $\mathbf{Q} \in (q_1, q_2)$. This bound can be expressed as follows:

$$(10.12) \quad \begin{aligned} \mathcal{F}_{\lambda^{(n)}}(U_{n,q_1}, Z_{n,\xi}) &\geq \mathbb{E}[\mathcal{F}_{\lambda^{(n)}}(U_{n,\mathbf{Q}}, Z_{n,\xi}) | \lambda^{(n)}] = \mathbb{P}\left[X_{n,\mathbf{Q}} \leq \frac{\xi - f_0 \mathbf{Q}}{\sqrt{\mathcal{E}}} \middle| \lambda^{(n)}\right] \\ &\geq \mathbb{P}\left[X_{n,\mathbf{Q}} \leq \frac{\xi - f_0 q_2}{\sqrt{\mathcal{E}}} \middle| \lambda^{(n)}\right] \xrightarrow[n \rightarrow \infty]{P} \Phi\left(\frac{\xi - f_0 q_2}{\sqrt{\mathcal{E}}}\right), \end{aligned}$$

where the equality follows from Lemma 10.3, and the convergence in probability is a consequence of Proposition 10.6.

Let $\epsilon > 0$. We set $q_1 := q$ and choose $q_2 > q$ such that the right-hand side of (10.12) is ϵ -close to $\Phi\left(\frac{\xi - f_0 q}{\sqrt{\mathcal{E}}}\right)$. This choice ensures that the probability of the event

$$\mathcal{F}_{\lambda^{(n)}}(U_{n,q}, Z_{n,\xi}) \geq \Phi\left(\frac{\xi - f_0 q}{\sqrt{\mathcal{E}}}\right) - 2\epsilon$$

converges to 1 as $n \rightarrow \infty$. This demonstrates that the lower bound for the double cumulative function $\mathcal{F}_{\lambda^{(n)}}$ is asymptotically close to the standard normal CDF, within an error margin of 2ϵ .

Proof of the opposite inequality follows similarly and is omitted for brevity. This completes the proof of Theorem 10.1.

11. Proofs of the main results. In this short section, we present the long-awaited proofs of Theorems 4.1 and 5.1. For Theorem 4.1, which addresses the asymptotic determinism of Schensted insertion, we will take a detour through Theorem 11.1. Although both theorems reach the same conclusion, the assumptions of Theorem 11.1 are framed in terms of Kerov's transition measure.

11.1. *The main idea.* Let λ be a random Young diagram, and let T be a uniformly random Poissonized tableau of shape λ . Consider a constant $z \in [0, 1]$. The central concept that underlies all proofs in this section is the relationship between the tail distribution function of the u -coordinate of the newly inserted box during the insertion $T \leftarrow z$ and the expected value of the double cumulative function. Specifically, we have:

(11.1)
$$\mathbb{P}[\text{u-Ins}(T; z) \geq u] = \mathbb{E}\mathcal{F}_\lambda(u, z)$$

for any $u \in \mathbb{R}$.

It is important to note that when λ is deterministic, the above equality represents the definition of the double cumulative function. In the general case, this relationship can be established by taking the appropriate average of both sides of the equation.

11.2. *First-order asymptotics for Schensted insertion.*

11.2.1. *Kerov transition measure formulation.* We begin our analysis with a refined version of Theorem 4.1, formulated in the language of Kerov transition measure. This formulation not only operates under weaker assumptions but also serves as a crucial first step in proving Theorem 4.1.

THEOREM 11.1. *Let the following conditions hold:*

- *For each $n \geq 1$ we are given a random Young diagram $\lambda^{(n)}$.*
- *There is a probability measure ν on the real line such that for each $u \in \mathbb{R}$ which is a continuity point of its CDF, the limit*

$$K_{\lambda^{(n)}}(\sqrt{n}u) \xrightarrow[n \rightarrow \infty]{P} F_\nu(u)$$

holds true in probability.

- *For a given real number $0 < z < 1$, the quantile function F_ν^{-1} is continuous at z .*

Let $T^{(n)}$ be a uniformly random Poissonized tableau of shape $\lambda^{(n)}$. Then

(11.2)
$$\frac{1}{\sqrt{n}} \text{u-Ins}(T^{(n)}; z) \xrightarrow[n \rightarrow \infty]{P} F_\nu^{-1}(z).$$

PROOF. Our strategy hinges on demonstrating that the limit of the CDF for the random variable on the left-hand side of (11.2) converges to a step function. Specifically, we aim to prove that for any $u \neq u_0$:

$$\lim_{n \rightarrow \infty} \mathbb{P}[\text{u-Ins}(T^{(n)}; z) < u] = \begin{cases} 0 & \text{if } u < u_0, \\ 1 & \text{if } u > u_0. \end{cases}$$

We begin by considering the case where $u < u_0$. By monotonicity, we can assume without loss of generality that u is a continuity point of F_ν . Applying (11.1), we obtain

$$\mathbb{P}[\text{u-Ins}(T^{(n)}; z) < u] = 1 - \mathbb{E}\mathcal{F}_{\lambda^{(n)}}(u, z) \xrightarrow[n \rightarrow \infty]{} 0.$$

This convergence to zero is a direct consequence of the second part of Proposition 9.1.

The case where $u > u_0$ follows an analogous reasoning process, completing our proof of convergence to the step function. $\square$

11.2.2. *Proof of Theorem 4.1.* PROOF OF THEOREM 4.1. The first three assumptions of Theorem 4.1 imply that the random sequence of continuous diagrams ω_n converges to Ω in probability, as defined by the topology introduced by Kerov (see Proposition 3.1 and the associated references). Consequently, Proposition 3.1 becomes applicable, ensuring that the conditions of Theorem 11.1 are satisfied with $\nu := \mu_\Omega$. $\square$

11.3. *Gaussianity of insertion fluctuations: The proof.* PROOF OF THEOREM 5.1. By applying Theorem 10.1 for the special choice of the vertical coordinate $\xi = 0$, we can demonstrate that the CDF of the random variable in question converges to that of a centered normal distribution with variance $\frac{\mathcal{E}}{f_0^2}$ as $n \rightarrow \infty$. Specifically, for any $q \in \mathbb{R}$ by Equation (11.1):

$$\begin{aligned} & \mathbb{P}\left[\sqrt[4]{n}\left[\frac{\mathbf{u}\text{-Ins}(T^{(n)}; z_0)}{\sqrt{n}} - u_0\right] < q\right] \\ &= 1 - \mathbb{E}[\mathcal{F}_{\lambda^{(n)}}(U_{n,q}, Z_{n,0})] \xrightarrow{n \rightarrow \infty} 1 - \Phi\left(\frac{-f_0 q}{\sqrt{\mathcal{E}}}\right) = \Phi\left(\frac{f_0 q}{\sqrt{\mathcal{E}}}\right). \quad \square \end{aligned}$$

Acknowledgments. We express our gratitude to Márton Balázs, Gaetan Borot, Marek Bożejko, Maciej Dołęga, Valentin Féray, Pablo Ferrari, Patrik Ferrari, Mustazee Rahman, and Dan Romik for their valuable discussions and suggestions regarding the bibliography.

Additionally, we thank Maciej Hendzel for conducting the Monte Carlo experiments related to Conjecture 1.5.

Funding. Research was supported by Narodowe Centrum Nauki, grant number 2017/26/A/ST1/00189.

Additionally, the first named author was supported by Narodowe Centrum Badań i Rozwoju, grant number POWR.03.05.00-00-Z302/17-00.

REFERENCES

- [1] BAIK, J., DEIFT, P. and JOHANSSON, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12** 1119–1178. MR1682248 <https://doi.org/10.1090/S0894-0347-99-00307-0>
- [2] BEN AROUS, G. and GUIONNET, A. (1997). Large deviations for Wigner’s law and Voiculescu’s non-commutative entropy. *Probab. Theory Related Fields* **108** 517–542. MR1465640 <https://doi.org/10.1007/s004400050119>
- [3] BIANE, P. (1998). Representations of symmetric groups and free probability. *Adv. Math.* **138** 126–181. MR1644993 <https://doi.org/10.1006/aima.1998.1745>
- [4] BIANE, P. (2001). Approximate factorization and concentration for characters of symmetric groups. *Int. Math. Res. Not.* **4** 179–192. MR1813797 <https://doi.org/10.1155/S1073792801000113>
- [5] BOGACHEV, L. V. and SU, Z. (2007). Gaussian fluctuations of Young diagrams under the Plancherel measure. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **463** 1069–1080. MR2310137 <https://doi.org/10.1098/rspa.2006.1808>
- [6] DOŁĘGA, M., FÉRAY, V. and ŚNIADY, P. (2010). Explicit combinatorial interpretation of Kerov character polynomials as numbers of permutation factorizations. *Adv. Math.* **225** 81–120. MR2669350 <https://doi.org/10.1016/j.aim.2010.02.011>
- [7] FERRARI, P. A., MARTIN, J. B. and PIMENTEL, L. P. R. (2009). A phase transition for competition interfaces. *Ann. Appl. Probab.* **19** 281–317. MR2498679 <https://doi.org/10.1214/08-AAP542>
- [8] FULTON, W. (1997). *Young Tableaux: With Applications to Representation Theory and Geometry*. London Mathematical Society Student Texts **35**. Cambridge Univ. Press, Cambridge. MR1464693
- [9] GORIN, V. and RAHMAN, M. (2019). Random sorting networks: Local statistics via random matrix laws. *Probab. Theory Related Fields* **175** 45–96. MR4009705 <https://doi.org/10.1007/s00440-018-0886-1>
- [10] GUSTAVSSON, J. (2005). Gaussian fluctuations of eigenvalues in the GUE. *Ann. Inst. Henri Poincaré Probab. Stat.* **41** 151–178. MR2124079 <https://doi.org/10.1016/j.anihpb.2004.04.002>
- [11] KENDALL, M. G. and STUART, A. (1979). *The Advanced Theory of Statistics. (In Three Volumes) Vol. 2: Inference and Relationship*. 4th ed. Charles Griffin & Co., Ltd., London, High Wycombe. 748 p. £ 23.50.
- [12] KEROV, S. (1993). Gaussian limit for the Plancherel measure of the symmetric group. *C. R. Acad. Sci. Paris Sér. I Math.* **316** 303–308. MR1204294
- [13] KEROV, S. V. (1993). Transition probabilities of continual Young diagrams and the Markov moment problem. *Funct. Anal. Appl.* **27** 104–117. MR1251166 <https://doi.org/10.1007/BF01085981>

- [14] KEROV, S. V. (2003). *Asymptotic Representation Theory of the Symmetric Group and Its Applications in Analysis. Translations of Mathematical Monographs* **219**. Amer. Math. Soc., Providence, RI. Translated from the Russian manuscript by N. V. Tsilevich. With a foreword by A. Vershik and comments by G. Olshanski. [MR1984868](#) <https://doi.org/10.1090/mmono/219>
- [15] KEROV, S. V. and VERSHIK, A. M. (1986). The characters of the infinite symmetric group and probability properties of the Robinson–Schensted–Knuth algorithm. *SIAM J. Algebr. Discrete Methods* **7** 116–124. [MR0819713](#) <https://doi.org/10.1137/0607014>
- [16] LOGAN, B. F. and SHEPP, L. A. (1977). A variational problem for random Young tableaux. *Adv. Math.* **26** 206–222. [MR1417317](#) [https://doi.org/10.1016/0001-8708\(77\)90030-5](https://doi.org/10.1016/0001-8708(77)90030-5)
- [17] MARCINIAK, M., MAŚLANKA, Ł. and ŚNIADY, P. (2021). Poisson limit of bumping routes in the Robinson–Schensted correspondence. *Probab. Theory Related Fields* **181** 1053–1103. [MR4344138](#) <https://doi.org/10.1007/s00440-021-01084-y>
- [18] MARCINIAK, M. and ŚNIADY, P. (2024). Cumulants of threshold for Schensted row insertion into random tableaux. Preprint. Available at [arXiv:2407.06213](#). <https://doi.org/10.48550/arXiv>
- [19] MINGO, J. A. and SPEICHER, R. (2017). *Free Probability and Random Matrices. Fields Institute Monographs* **35**. Springer, New York. [MR3585560](#) <https://doi.org/10.1007/978-1-4939-6942-5>
- [20] OKOUNKOV, A. (2000). Random matrices and random permutations. *Int. Math. Res. Not.* **20** 1043–1095. [MR1802530](#) <https://doi.org/10.1155/S1073792800000532>
- [21] PITTEL, B. and ROMIK, D. (2007). Limit shapes for random square Young tableaux. *Adv. in Appl. Math.* **38** 164–209. [MR2290809](#) <https://doi.org/10.1016/j.aam.2005.12.005>
- [22] RAHMAN, M. and VIRÁG, B. (2025). Infinite geodesics, competition interfaces and the second class particle in the scaling limit. *Ann. Inst. Henri Poincaré Probab. Stat.* **61** 1075–1126. [MR4901636](#) <https://doi.org/10.1214/23-aihp1454>
- [23] ROMIK, D. (2015). *The Surprising Mathematics of Longest Increasing Subsequences. Institute of Mathematical Statistics Textbooks* **4**. Cambridge Univ. Press, New York. [MR3468738](#)
- [24] ROMIK, D. and ŚNIADY, P. (2015). Jeu de taquin dynamics on infinite Young tableaux and second class particles. *Ann. Probab.* **43** 682–737. [MR3306003](#) <https://doi.org/10.1214/13-AOP873>
- [25] ROST, H. (1981). Nonequilibrium behaviour of a many particle process: Density profile and local equilibria. *Z. Wahrsch. Verw. Gebiete* **58** 41–53. [MR0635270](#) <https://doi.org/10.1007/BF00536194>
- [26] ŚNIADY, P. (2006). Gaussian fluctuations of characters of symmetric groups and of Young diagrams. *Probab. Theory Related Fields* **136** 263–297. [MR2240789](#) <https://doi.org/10.1007/s00440-005-0483-y>
- [27] ŚNIADY, P. (2025). Modulus of continuity of Kerov transition measure for continual Young diagrams. *Electron. J. Probab.* **30** Paper No. 108, 26. [MR4926018](#) <https://doi.org/10.1214/25-ejp1369>
- [28] STANLEY, R. P. (1999). *Enumerative Combinatorics, Vol. 2. Cambridge Studies in Advanced Mathematics* **62**. Cambridge Univ. Press, Cambridge. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin. [MR1676282](#) <https://doi.org/10.1017/CBO9780511609589>
- [29] THOMA, E. (1964). Die unzerlegbaren, positiv-definiten Klassenfunktionen der abzählbar unendlichen, symmetrischen Gruppe. *Math. Z.* **85** 40–61. [MR0173169](#) <https://doi.org/10.1007/BF01114877>
- [30] TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. [MR1257246](#)
- [31] VASSILIEV, N. N., DUZHIN, V. S. and KUZMIN, A. D. (2019). Investigation of properties of equivalence classes of permutations by inverse Robinson–Schensted–Knuth transformation. *Inf. Control Syst.* **98** 11–22. [https://doi.org/10.31799/1684-8853-2019-1-11-22](#)
- [32] VERSHIK, A. M. and KEROV, S. V. (1981). Asymptotic theory of the characters of a symmetric group. *Funktsional. Anal. i Prilozhen.* **15** 15–27, 96. [MR0639197](#)
- [33] VERSHIK, A. M. and KEROV, S. V. (1977). Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux. *Sov. Math., Dokl.* **18** 527–531. [MR0480398](#)
- [34] VOICULESCU, D. (1994). The analogues of entropy and of Fisher’s information measure in free probability theory. II. *Invent. Math.* **118** 411–440. [MR1296352](#) <https://doi.org/10.1007/BF01231539>
- [35] WOJTYNIAK, K. (2019). Asymptotyka losowych trajektorii jeu de taquin, Bachelor’s Thesis. Nicolaus Copernicus Univ. in Toruń.
- [36] LAURITZEN, S. L. and HALD, A. (2002). *Thiele: Pioneer in Statistics*. Oxford Univ. Press, New York. Thiele’s papers translated from the Danish by Steffen L. Lauritzen. With appreciations of Thiele’s work by Lauritzen and A. Hald. [MR2055773](#) <https://doi.org/10.1093/acprof:oso/9780198509721.001.0001>