Communications in
**Mathematical
Physics**

The Lower Tail of q-pushTASEP

Ivan Corwin, Milind Hegde

Columbia University, New York, USA. E-mail: milind@math.columbia.edu

Abstract: We study q-pushTASEP, a discrete time interacting particle system whose distribution is related to the q-Whittaker measure. We prove a uniform in N lower tail bound on the fluctuation scale for the location $x_N(N)$ of the right-most particle at time N when started from step initial condition. Our argument relies on a map from the q-Whittaker measure to a model of periodic last passage percolation (LPP) with geometric weights in an infinite strip that was recently established in Imamura et al. (Skew RSK dynamics: Greene invariants, affine crystals and applications to q-Whittaker polynomials. arXiv:2106.11922, 2021). By a path routing argument we bound the passage time in the periodic environment in terms of an infinite sum of independent passage times for standard LPP on $N \times N$ squares with geometric weights whose parameters decay geometrically. To prove our tail bound result we combine this reduction with a concentration inequality, and a crucial new technical result—lower tail bounds on $N \times N$ last passage times uniformly over all $N \in \mathbb{N}$ and all the geometric parameters in $(0, 1)$. This technical result uses Widom's trick (Widom in Int Math Res Notices 2002(9):455–464, 2002) and an adaptation of an idea of Ledoux introduced for the GUE (Ledoux, in: GAFA seminar notes, 2005) to reduce the uniform lower tail bound to uniform asymptotics for very high moments, up to order N, of the Meixner ensemble. This we accomplish by first obtaining sharp uniform estimates for factorial moments of the Meixner ensemble from an explicit combinatorial formula of Ledoux (Electron J Probab 10:1116–1146, 2005), and translating them to polynomial bounds via a further careful analysis and delicate cancellation.

Contents

1. Introduction, Main Results, and Proof Ideas

The Kardar–Parisi–Zhang (KPZ) universality class consists of a large variety of models, all of which are believed to exhibit certain universal behaviors; for example, common scaling limits. Most progress in this area has been in the setting of certain models that are known as *exactly solvable* or *integrable*, which possess certain algebraic structure that makes their analysis within reach, in comparison to non-integrable models.

In such models, it is often important for applications to have control on the upper and lower tails of the KPZ observable on the fluctuation scale. Of the two, it is more challenging to obtain this control on the lower tail (i.e., the one that typically has cubic tail exponent), though in what are known as zero-temperature models, such as TASEP and last passage percolation, a variety of techniques have been developed over the last two decades to do this (e.g., Riemann–Hilbert methods or analysis of determinantal representations; see Sect. 1.3 for a detailed discussion). In contrast, for *positive* temperature models such as the KPZ equation, stochastic six vertex model, ASEP, and polymer models, only a few techniques have recently been developed to approach this problem. Further, each technique only seems to be applicable in particular cases; due to fundamental limitations, there is no broad coverage.

In this paper we study the exactly solvable, discrete time interacting particle system model of *geometric q-pushTASEP*, a positive temperature model, but one for which the few methods available to obtain lower tails in positive temperature do not seem applicable. We develop a new technique for lower tail estimates on the position of the right-most particle, harnessing recently discovered connections between it and last passage percolation.

We start by introducing the model of study and our main results.

1.1. Principal objects and models of study.

1.1.1. Some notation and distributions The *q-Pochhammer symbol* $(z; q)_n$ is given by

$$(z; q)_n = \prod_{i=0}^{n-1} (1 - zq^i) \quad \text{for} \quad n = 0, 1, \ldots,$$

with $(z; q)_\infty$ defined by replacing $n - 1$ by ∞. The *q-binomial coefficient* is given by

$$\binom{n}{k}_q = \frac{(q; q)_n}{(q; q)_k (q; q)_{n-k}}. \tag{1}$$

The *q-deformed beta binomial distribution* is a distribution with parameters q, ξ, η, and m. Here $m \in \mathbb{Z}_{\geq 0}$ and the distribution is defined on $\{0, 1, \ldots, m\}$; the other parameters are non-negative real numbers, and are restricted to more specific domains in certain cases that we will describe. For $s \in \{0, \ldots, m\}$, the probability mass function at s is given by

$$\varphi_{q, \xi, \eta}(s \mid m) = \xi^s \frac{(\eta/\xi; q)_s (\xi; q)_{m-s}}{(\eta; q)_m} \cdot \binom{m}{s}_q.$$

We refer the reader to [MP16, Section 6.1] for more information regarding this distribution, including a discussion on why the above expression sums (over $s = 0, \ldots, m$) to 1 when the expression is well-defined and non-negative.

A special case is the q-Geometric distribution of parameter ξ (denoted q-Geo(ξ)), obtained by taking $m = \infty$, $\eta = 0$, and $q, \xi \in (0, 1]$, so that the probability mass function at $s \in \mathbb{Z}_{\geq 0}$ is given by

$$\varphi_{q,\xi,\eta}(s \mid \infty) = \xi^s \frac{(\xi; q)_\infty}{(q; q)_s}.$$

1.1.2. The model of q-pushTASEP The q-pushTASEP is a discrete time interacting particle system on $\mathbb{Z}$ first introduced in [MP16]. We have $N \in \mathbb{N}$ many particles which occupy distinct sites in $\mathbb{Z}$, and we label their position at time $T \in \mathbb{Z}_{\geq 0}$ in increasing order as $x_1(T) < x_2(T) < \cdots < x_N(T)$; we denote the collection of these random variables by $x(T)$. We also specify a collection of parameters $a_1, \ldots, a_N$ and $b_1, b_2, \ldots$, all lying in $(0, 1)$.

The evolution from time T to $T + 1$ is as follows (see Fig. 1). The particle positions are updated from left to right: for $k \in \{1, \ldots, N\}$,

$$x_k(T + 1) = x_k(T) + J_{k,T} + P_{k,T},$$

where $J_{k,T}$ and $P_{k,T}$ are independent random variables with $J_{k,T} \sim q$-Geo($a_k b_{T+1}$) (encoding a *jump* contribution) and

$$P_{k,T} \sim \varphi_{q^{-1}, \xi = q^{\mathrm{gap}_k(T)}, \eta=0}\big(\cdot \mid x_{k-1}(T+1) - x_{k-1}(T)\big),$$

(encoding a *push* contribution) where $\mathrm{gap}_k(T) = x_k(T) - x_{k-1}(T) - 1$, $x_0(T) = -\infty$ by convention and, by a slight abuse of notation, $\sim$ means the LHS is distributed according to the measure which has probability mass function given by the RHS. In other words, $P_{k,T}$ is a q-deformed beta binomial random variable with parameters q^{-1}, $\xi = q^{\mathrm{gap}_k(T)}$, $\eta = 0$, and $m = x_{k-1}(T+1) - x_{k-1}(T)$. Note in particular that x_1's motion does not depend on that of any other particle, i.e., marginally it follows a random walk. Further, the process is an exclusion process, i.e., particles always occupy distinct sites and also remain ordered.[1]

This model is integrable. More precisely, the distribution of $x_N(T)$, started from a special initial condition known as *step initial condition* where $x_k(0) = k$, can be related to a marginal of the q-Whittaker measure, a measure on partitions (equivalently, Young diagrams) defined in terms of q-Whittaker polynomials. This connection will be important for our arguments, and we will discuss it more in Sect. 1.5.

Apart from integrability, another reason q-pushTASEP is of interest is because it degenerates to other well-known models. Indeed, in the $q \to 1$ limit, when appropriately renormalized, $x_N(T)$ converges to the free energy of the log-gamma polymer model. While our results will not carry over to this limit, we will make some further remarks about this relationship between the models in Sect. 1.7. We also mention that after speeding up time and letting the jump rates go to zero, q-pushTASEP converges to the continuous time q-pushTASEP [BP16] which, when $q = 0$, is the well-known pushTASEP [BF08].

---

[1] Let $\Delta_{k-1}(T+1) = x_{k-1}(T+1) - x_{k-1}(T)$ and j be the value the probability mass function in the RHS of the previous display is evaluated at. When $\Delta_{k-1}(T+1) \geq \mathrm{gap}_k(T)$, the probability mass function is zero unless $j \geq \Delta_{k-1}(T+1) - \mathrm{gap}_k(T)$, as otherwise the factor $(q^{\mathrm{gap}_k(T)}; q^{-1})_{\Delta_{k-1}(T+1)-j}$ is zero. It is immediate from the definition that this implies $x_k(T) + j \geq x_{k-1}(T+1) + 1$, thus ordering and exclusion are maintained.
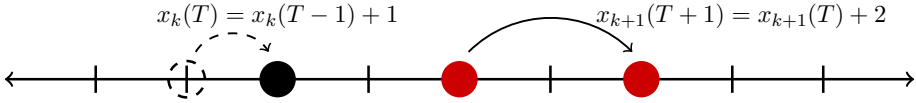
$$x_k(T) = x_k(T-1) + 1 \qquad\qquad x_{k+1}(T+1) = x_{k+1}(T) + 2$$

Fig. 1. A depiction of one step in the evolution of q-pushTASEP. The dotted circle is the position of the kth particle at time $T - 1$, and the solid black circle is it after it moves to its position at time T. The left red circle is the $(k + 1)$th particle at time T and the right one the same at time $T + 1$. The movement of the kth particle in the previous step effects the $P_{k,T}$ contribution to the total jump size of 2 of the $(k+1)$th particle at time T

1.1.3. Law of large numbers and asymptotic Tracy–Widom fluctuations of q-pushTASEP
In this work we will focus on q-pushTASEP when the parameters are equal, i.e., $u = a_i = b_j$ for all $i = 1, \ldots, N$ and $j = 1, 2, \ldots$ for some $u \in (0, 1)$, and when the initial condition is $x_k(0) = k$ for $k = 1, \ldots, N$. In this setting, and under some additional restrictions on the parameters, [Vet22] proved a law of large numbers for $x_N(T)$, which in the $T = N$ case states (with the convergence being in probability) that

$$\lim_{N \to \infty} \frac{x_N(N)}{N} = 2 \cdot \frac{\psi_q(\log_q u) + \log(1 - q)}{\log q} + 1 =: f_q; \tag{2}$$

here $\log_q u = \log u / \log q$ is the logarithm to the base q and ψ_q is the q-digamma function, given by

$$\psi_q(x) = \frac{1}{\Gamma_q(x)} \frac{\partial \Gamma_q(x)}{\partial x}, \tag{3}$$

where $\Gamma_q(x) = \frac{(q;q)_\infty}{(q^x;q)_\infty}(1 - q)^{1-x}$ is the q-gamma function.

Note that our definition of q-pushTASEP differs from that of [Vet22, MP16], in that particles move to the right for us rather than the left, thus introducing an extra negative sign in the law of large numbers. Our definition agrees with the one given in [IMS22].

[Vet22] also proves that the asymptotic fluctuation of x_N converges to the GUE Tracy–Widom distribution. To state this, let us consider the rescaled observable

$$X_N^{\text{sc}} = \frac{x_N(N) - f_q N}{(-\psi_q''(\log_q u))^{1/3}(\log q^{-1})^{-1} N^{1/3}}; \tag{4}$$

note that the denominator is a positive quantity, since $\psi_q''(x) < 0$ for all $x > 0$ (see e.g., [MS09]). Now for $q, u \in (0, 1)$, and under certain restrictions on those parameters that are used to simplify the analysis there, [Vet22, Theorem 2.2] asserts that $X_N^{\text{sc}} \Rightarrow F_{\text{GUE}}$, where F_{GUE} is the GUE Tracy–Widom distribution.

The proof given in [Vet22] relies on certain formulas for q-Laplace transforms of particle positions proved in [BCFV15]. The recent work [IMS22] gives different Fredholm determinant formulas for randomly shifted versions of $x_N(T)$ (see Corollary 5.1 there) from which it should also be possible to extract the above distributional convergence, with a perhaps simpler analysis; indeed, the analogous convergence is demonstrated for a half-space version of the model in [IMS22, Theorem 6.11].

It remains a question whether the conditions assumed in [Vet22] are necessary for this convergence to hold; our techniques suggest it should hold for any $q, u \in (0, 1)$. In particular, our results will hold for all $q, u \in (0, 1)$.

1.2. Main results. Our main theorem bounds the lower tail of the fluctuations of the centred and scaled position X_N^{sc} (as defined in (4)) of the Nth particle of q-pushTASEP, as introduced in Sect. 1.1.2.

Theorem 1.1. *Let $q, u \in (0, 1)$ and let $a_i = b_j = u$ for all i, j. There exist positive absolute constants c', C, and N_0 (independent of q and u) such that, with $\theta_0 = C(1 - u)^{-1/3}(1 \vee (\log q^{-1})^{-2/3})$ and $c = (1 - u)^{1/2}c'$, and for $N \geq N_0$ and $\theta > \theta_0$,*

$$\mathbb{P}\left(X_N^{sc} < -\theta\right) \leq \exp\left(-c\theta^{3/2}\right).$$

We believe the true lower tail behavior to be $\exp(-c\theta^3)$, at least for $\theta \ll N^{2/3}$, i.e., smaller than the large deviation regime, similar to other models in the KPZ class. We discuss ahead in Remark 1.7 why our arguments do not achieve this, and also how with additional different arguments it should be possible to attain the full exponent of 3.

It is an interesting question whether Theorem 1.1 can be extended to the case of general a_i and b_j. As we will see, a crucial connection to last passage percolation that we rely on continues to hold, so the general scheme is broadly applicable. However certain moment formulas available for the last passage percolation problem in the case of homogeneous parameters are not known for the general case, and this is where the strategy stops being directly applicable.

1.3. Lower tails of KPZ observables. In the past decade, integrable tools have been combined with other perspectives to slowly push out of strictly integrable settings. Examples include studies of geometric properties in models of last passage percolation (e.g., [Joh00b, BSS14, BSS19, BG21, BHS18, BGZ21, BGHH22, SSZ21, SS22]), process-level regularity properties (e.g., [CH14, CH16, Ham22, Ham19, CHH23, SV21, Dau23]) of processes whose finite-dimensional distributions are accessible via exactly solvable tools, recent progress on constructing the ASEP speed process [ACG23], edge and bulk scaling behavior of tiling or dimer models [Agg23, Hua21, AH21, HYZ23], as well as the construction of the directed landscape and convergence of LPP models to it [DV21a, DOV22, DV21b]. In these works, a crucial input from the integrable side has repeatedly been bounds on the tails of the relevant statistic. In the following, by zero temperature models we will mean models which can be embedded in determinantal point processes, e.g., integrable models of last passage percolation.

In the case of zero temperature models (e.g., geometric or exponential last passage percolation or the totally asymmetric simple exclusion process), these tail bounds had been studied two decades ago, with arguments relying in an essential way on determinantal structure possessed by these models. In positive temperature (e.g., the KPZ equation or the asymmetric simple exclusion process), where such structure is not directly available, progress on obtaining these important tail inputs has only been made in the last few years, but their availability promises to create the opportunity to bring the zero temperature successes to the positive temperature.

1.3.1. The relative difficulty of upper and lower tail bounds From a physical perspective, it is easy to see that the upper and lower tails should have different rates of decay, with the lower tail decaying faster. This is because the upper tail concerns making a single, "largest" object even larger—in q-pushTASEP, making the right-most particle lie even further to the right, which can be accomplished by demanding a single large jump of the right-most particle. So, in particular, the other particles are not a barrier. In contrast, in

the lower tail exactly the opposite happens: for the right-most particle to lie atypically to the left, *all* the other particles must also do so—in particular, many jumps, including those of other particles, must be suppressed. However, this intuition does not reveal the fact that, typically, it is technically much more challenging to obtain lower tail bounds than upper ones. Further, while this intuition turns out to be well-suited for arguments to understand large deviations behavior (e.g., [BGS21] for LPP), i.e., deviations on scale N, in applications one needs bounds on the fluctuation scale (i.e., deviations on scale $N^{1/3}$).

For solvable zero temperature models, which have determinantal descriptions, the difference in the difficulty of upper and lower tails can be seen from the fact that the upper tail bounds follow directly from bounds on the kernel of the associated determinantal point process, while this is not the case for lower tail bounds. Nevertheless, as mentioned above, in these models, a number of approaches have been developed over the last two decades. These include the Riemann–Hilbert approach (e.g., [BDM+01]), methods based on explicit formulas for moments (e.g., [Led05a]), connections to random matrix theory (e.g., [Joh00a,LR10,BGHK21,RRV11] as well as large deviation work e.g., [GH21,AGH21,CDG23]), and abstract concentration (see e.g., [ADH17, Section 3] for applications to first passage percolation).

The toolbox for the upper tail is already fairly well developed in positive temperature (i.e., non-determinantal but exactly solvable models). Here too one often has determinantal formulas for the distribution of the observable, and one can extract the upper tail by establishing decay of the kernel in these determinantal formulas. An instance where this is done is [BCD21, Theorem 1.4], in the context of the log-gamma polymer. Besides this approach, one can also try to extract upper tail estimates from the moments of the exponential of the random variable of interest (e.g., for the KPZ equation, this corresponds to moments of the stochastic heat equation, or, for our model, the analogue is q-moments). An example of this method is captured in [CG20a, Proposition 4.3 and Lemma 4.5], where tail estimates for the narrow-wedge KPZ equation are obtained through estimates on the kth moment of the stochastic heat equation. We also mention recent works [GH22] and [LS22a] which respectively make use of Gibbs properties and special structure of stationary versions of the relevant models (the KPZ equation and O'Connell–Yor polymer respectively) to prove upper tail estimates, but these methods by their nature are specific to models which have such probabilistic structure.

For our model of q-pushTASEP, a Fredholm determinant formula of the type the first approach relies on can be found in [BCFV15, Theorem 3.3] (via our model's connection to the q-Whittaker measure, see Sect. 1.5 ahead or the discussion in [MP16, Section 7.4]), and a different one in [IMS22, Corollary 5.1]. A formula for the q-moments in our model is also available, though there is a subtlety in that not all q-moments are finite; see Section 7.4 in [MP16] for the formula and a brief discussion of this point.

Having said this, while there are well-established approaches to obtain such upper tail estimates, it is certainly not a triviality to actually do so in any model, and we do not pursue them for q-pushTASEP in this work. However, we plan to revisit this question as part of subsequent work in which we will need both tail bounds to study further aspects of this model.

The toolbox for the lower tail in positive temperature is smaller but is being actively developed, and we briefly review some of the tools now. However, these do not seem applicable to our model, and so we are ultimately led to develop a new technique.

1.3.2. Work using determinantal representations of Laplace transforms The first class of techniques for lower tails in positive temperature models gives a determinantal representation for the Laplace transform (or q-Laplace transform) of the observable. This approach was initiated in [CG20b], which obtained fluctuation-scale lower tail bounds for the narrow wedge solution to the KPZ equation. [CG20b] used a formula from [BG16, ACQ11] which equates the Laplace transform of the fundamental solution of the stochastic heat equation (which is related to the KPZ equation via the Cole–Hopf transform) to an expectation of a multiplicative functional of the Airy point process, which is determinantal. That it is a multiplicative functional (as well as its precise form) is very useful as it allows the lower tail of the KPZ equation to be bounded in terms of the lower tail behavior of the particles at the edge of the Airy point process, which in turn can be controlled via determinantal techniques as outlined above.

The Laplace transform identity that this argument relies on can be seen as a special case of a general matching proved in [Bor18] between the stochastic six vertex model's height function and a multiplicative functional of the row lengths of a partition sampled according to the Schur measure. The stochastic six vertex model and the Schur measure are each known to specialize to a number of models also of interest; for example, one degeneration of the former is the asymmetric simple exclusion process (ASEP), and the analogous one for the latter is the discrete Laguerre ensemble, a determinantal process. This yields an identity between the q-Laplace transform of ASEP and a multiplicative functional of the discrete Laguerre ensemble [BO17], which was used to obtain a lower tail bound for the former in [ACG23], using the latter's connection to TASEP.

Unfortunately, not all degenerations to models of interest play nicely on both sides. For instance, for the O'Connell–Yor polymer, the Schur measure side of the stochastic six vertex model identity degenerates to an average of a multiplicative functional with respect to a point process whose measure is a *signed* measure instead of a probability measure [IS16]. Typical modes of analysis break down in the context of signed measures. For other models too, including ours, this issue of signed measures seems to arise.

A related recent approach brings in the machinery of Riemann–Hilbert problems and has been developed in [CC22]. There, in the setting of the KPZ equation, the mentioned expectation of the multiplicative functional of the Airy point process is expressed as a Fredholm determinant, and then the latter is written as a Riemann–Hilbert problem. So far this approach has only been developed at the level of the KPZ equation, and so it remains to be seen how broadly it can be applied.

1.3.3. Coupling and geometric methods in polymer models For the semi-discrete O'Connell–Yor and log-gamma polymer models, recent work [LS22b, LS22a] has obtained lower tail estimates via a mixture of exact formulas, coupling arguments, and geometric considerations. This builds on methods developed in the zero temperature model of exponential last passage percolation [EJS20, EJS21, EGO22]. The program has so far been implemented in full in the semi-discrete O'Connell–Yor model and in part for the log-gamma polymer. First, [LS22b] obtains the bounds for a stationary version of the model (where one can prove an explicit formula for the Laplace transform of the free energy), and then, for the O'Connell–Yor case, these are translated to the original model using geometric considerations of the polymer measure in [LS22a]. In fact, the Laplace transform bound obtains a lower tail exponent of 3/2, which is then upgraded to the sharp exponent of 3 by adapting geometric arguments from [GH23]. Since this method relies heavily on the polymer geometry, it is unclear how it could be extended to address the model of q-pushTASEP which only has a particle interpretation.

In summary, while there are a variety of methods in zero-temperature models to obtain lower tail bounds, so far only a handful of tools are available for positive temperature models. The ones available do not seem immediately applicable to our model. For this reason, we introduce a new method which does not rely on polymer structure or identities between q-Laplace transforms and multiplicative functionals of determinantal point processes, which are not directly available in q-pushTASEP. We rely instead on the recent work [IMS21] which relates the q-Whittaker measure on partitions to a model of periodic geometric last passage percolation. In this way, we are able to use both the polymer techniques and determinantal structure which *are* available in geometric last passage percolation to analyze q-pushTASEP. Our methods may also be useful in studying the lower tails of q-pushTASEP with certain other special initial data that also have connections to marginals of q-Whittaker measures, or q-pushTASEP with particle creation [BBC20,IMS22] which should have a description in terms of a similar last passage percolation problem via half-space q-Whittaker measures.

To explain our broad approach, we next describe this model of periodic last passage percolation.

1.4. Last passage percolation. We first describe the environment in which our last passage percolation (LPP) problem will exist. We consider a sequence of $N \times T$ "big rectangles" indexed by $k \in \mathbb{N} \cup \{0\}$, each of which contains NT "small squares" inside. These are arranged in a periodic strip as shown in Fig. 2. We use the coordinates $(i, j; k)$ (with $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$ and $k \in \{0, 1, \ldots, \}$) to denote the small square with coordinates (i, j) in the kth big square. The site $(i, j; k)$ is associated with an independent non-negative random variable $\xi_{(i,j;k)}$ which we call a *site weight*.

The distribution of the randomness of the site weights is as follows. We will say $X \sim \text{Geo}(z)$ if X is a random variable such that $\mathbb{P}(X \geq k) = z^k$ for $k = 0, 1, 2\ldots$; in other words, z is the *failure* probability in repeated independent trials and X is the number of failures before the first success. Then the site weights are specified as follows: $\xi_{(i,j;k)}$ are independent across all i, j, k, and distributed as $\text{Geo}(a_i b_j q^k)$ for $k = 0, 1, 2, \ldots$ and $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$.

Note that in our model with the specialization $u = a_i = b_j$, the site weights in the same big square, i.e., with the same value of k, are identically distributed as $\text{Geo}(u^2 q^k)$. For $s = (i, j; k)$ a site in the strip, we may also write ξ_s for $\xi_{(i,j;k)}$.

Remark 1.2. Observe that $\mathbb{P}(\xi_{(i,j;k)} \neq 0) = a_i b_j q^k$ for all i, j, k, which is summable over $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$ and $k = 0, 1, \ldots$. So by the Borel–Cantelli lemma, almost surely, for all large enough k and all $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$, $\xi_{(i,j;k)}$ will be zero.

We consider downward paths which are allowed to wrap around the strip, i.e., paths which are at $(i, T; k)$ may move to $(i, 1; k + 1)$ and paths which are at $(N, j; k)$ may move to $(1, j; k + 1)$ in the next step; again see Fig. 2. Each such path γ is assigned a weight $w(\gamma)$ given by $\sum_{v \in \gamma} \xi_v$. Note that while a priori the weight of γ could be infinite if γ is an infinite path, in our setting the environment will only have finitely many non-zero site weights almost surely (as noted above in Remark 1.2), and so this possibility will not arise.

Now, the last passage value $L_{v,w}$ between v and w small squares in the strip is defined as

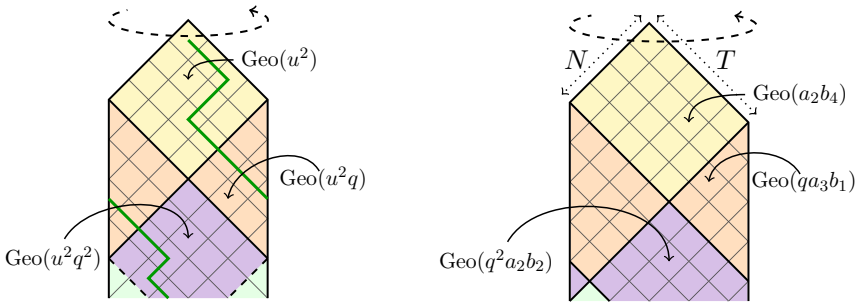$$L_{v,w} := \max_{\gamma: v \to w} w(\gamma),$$

Fig. 2. The environment in which the infinite last passage percolation occurs. The dashed arrow on top indicates the direction in which the squares wrap around, and the solid green line on the left is a downward path which wraps around the strip. In the left panel we have the model under consideration in this article with the specialization of $T = N$ and $u = a_i = b_j$ for all i, j, while on the right is the general model

where the maximum is over all downward paths from v to w, assuming at least one such path exists. If not, we define $L_{v,w}$ to be $-\infty$; we say this only to give a logically complete definition, but such cases will not actually arise in this paper.

This model of LPP is similar to other models of periodic LPP considered in the literature (e.g., [BL18, BL19, BL21, BO21, SS22] though perhaps with slightly different selections of parameters or distributions of the random variables), and also has connections to the periodic Schur measure [Bor07, BO21].

1.5. Relation between q-pushTASEP and LPP. We can now explain the exactly solvable connection between q-pushTASEP and the model of LPP in an infinite periodic strip just introduced that was recently discovered by Imamura–Mucciconi–Sasamoto, and on which our arguments crucially rely. We first state the equivalence precisely.

While this paper only considers the case $T = N$, we will state the LPP equivalence for general T. For this, we return to the LPP problem described in Sect. 1.4 where the infinite periodic environment has "fundamental domain" with $N \times T$. The parameters of the geometric random variables is as described there, i.e., site $(i, j; k)$ has parameter $a_i b_j q^k$.

Theorem 1.3. *Let L be the LPP value in the environment just described with $a_i, b_j \in (0, 1)$ for all $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$. Let $x_N(T)$ be the position of the Nth particle at time T in q-pushTASEP with the same parameters a_i, b_j for all $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$ and step initial condition. Then*

$$x_N(T) \overset{d}{=} L + N.$$

The connection between $x_N(T)$ and L runs through a measure on partitions known as the q-Whittaker measure, as mentioned above. More precisely, it was shown in [MP16] (and stated ahead as Theorem A.3) that $x_N(T)$, for any $N, T \in \mathbb{N}$ and with step initial condition (i.e., $x_k(0) = k$ for $k = 1, \ldots, N$), is distributed as the length of the top row in a partition (encoded as a Young diagram) distributed according to the q-Whittaker measure (after a deterministic shift by N). So it remains to establish a distributional equality between the q-Whittaker measure's top row and the LPP value L; the proof of this was explained to us by Matteo Mucciconi and relies on results from [IMS21] and we discuss it next.

[IMS21] proves a relation between the q-Whittaker measure and the periodic LPP model. From the perspective of the needs of this paper, the main consequence of [IMS21] is the development of a bijection which generalizes the Robinson–Schensted–Knuth correspondence. In traditional LPP on $\mathbb{Z}^2$, the RSK correspondence associates to the LPP environment (with non-negative integer site weights) a pair of Young tableaux of the same shape, with the property that LPP statistics are encoded in the row lengths of the tableaux; for instance, the LPP value is exactly the length of the top row of the tableaux. It is well-known that, under this correspondence, the measure on the environment given by i.i.d. geometric random variables gets pushed forward to give the Schur measure on partitions, i.e., Young diagrams. The generalization of RSK established in [IMS21], called skew RSK there, relates pairs of *skew* Young tableaux to pairs of *vertically strict tableaux* (tableaux where the ordering condition on the entries is imposed only along columns and not rows) along with some additional data.

In this bijection, LPP has not had a role to play. To involve LPP, we recall an earlier generalization of RSK known as the Sagan–Stanley correspondence [SS90], which can be interpreted as giving a bijection between the LPP environment in an infinite strip (again with non-negative integer site weights) and pairs of skew Young tableaux. [IMS21] also shows that, if one composes this bijection with the skew RSK bijection, then the LPP value is exactly the length of the top row of the vertically strict tableaux coming from the skew RSK.

It turns out that the generating function of vertically strict tableaux can be written in terms of the q-Whittaker polynomials. Using this fact, certain weight preservation properties of the bijection, and an argument similar to the well-known one that establishes the above mentioned relationship between geometric LPP and the Schur measure, one can show that the LPP value when the infinite strip has site weights given by independent geometric variables with parameter specified above has the same distribution as the top row of a random partition from the q-Whittaker measure. Since this statement is not recorded explicitly in [IMS21], we will give a proof using results from that paper in Appendix A. In fact, one can relate the lengths of all the rows of the partition to LPP values involving multiple disjoint paths, and we prove this stronger statement in Theorem A.4. Theorem 1.3 then follows by combining this theorem with Theorem A.3 (relating $x_N(T)$ and the top row of the q-Whittaker measure).

1.6. Proof ideas. We specialize to $N = T$ and $a_i = b_j = u \in (0, 1)$ for $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$.

To summarize, $x_N(N)$ is, up to a deterministic shift by N, the LPP value in an infinite periodic environment of inhomogeneous geometric random variables. Now, the weight of *any* path in this environment is a lower bound on the last passage percolation value. We consider a specific path which allows us to utilize the homogeneity of the geometric variable parameters inside a big square (as well as tail information of geometric LPP in such homogeneous squares) along with the independence across big squares.

More specifically, we consider the path formed by concatenating paths from the top to bottom of the big squares along the center, i.e., the squares in which the geometric parameter is $u^2 q^{2i}$ for some $i \in \mathbb{N} \cup \{0\}$. More precisely, we do not exactly concatenate the paths as they do not have a common site; we simply consider the sum of the weights of the paths, ignoring the positive weight of the extra site needed to actually join the paths. See Fig. 3.

Let us calculate the law of large numbers of this path, i.e., its weight up to first order, using the knowledge of the LLN for geometric LPP. Indeed, in an $N \times N$ square with
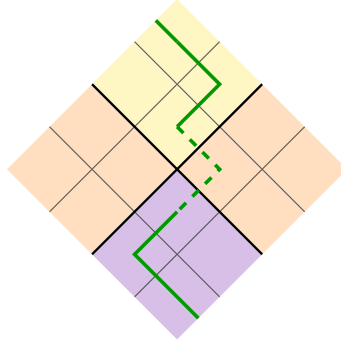
Fig. 3. A depiction of the paths we consider near the boundary between different big squares. The two solid green paths go from the topmost site to the bottommost site in their respective big squares, where the environment is homogeneous. We do not consider the dotted green path needed to connect them, which is valid for proving an upper bound on the lower tail since including its weight will only increase the overall weight

geometric parameter $u^2 q^{2i}$, to first order in N, the LPP value is $2N \cdot \frac{uq^i}{1-uq^i}$ (see for example [Joh00a] or Theorem 1.4 ahead), so that the overall LPP value of the path we have described is, again to first order,

$$2N \cdot \sum_{i=0}^{\infty} \frac{uq^i}{1 - uq^i} = 2N \cdot \sum_{i=0}^{\infty} \frac{q^{i+\log_q u}}{1 - q^{i+\log_q u}}. \tag{5}$$

To evaluate this sum we need the q-digamma function ψ_q, defined in (3). Now, the q-digamma function ψ_q is related to the sum in (5) by the formula

$$\psi_q(x) = -\log(1-q) + \log q \cdot \sum_{i=0}^{\infty} \frac{q^{i+x}}{1 - q^{i+x}}.$$

From this we see that (5) equals

$$2N \cdot \frac{\psi_q(\log_q(u)) + \log(1-q)}{\log q} = N(f_q - 1), \tag{6}$$

which is the first order term in the probability in Theorem 1.1 (remember that L and $x_N(N)$ differ by a constant term of N) and matches the LLN proved in [Vet22].

1.6.1. Uniform LPP control We have identified a concatenation of LPP problems which obtains the correct first order behaviour. Now, the order of fluctuations of geometric LPP of parameter $u^2 q^{2i}$ in an $N \times N$ square is $u^{1/3} q^{i/3} (1 - u^2 q^{2i})^{-1} N^{1/3}$ (again see for example [Joh00a] or Theorem 1.4 ahead). One can think of these fluctuations, once rescaled by this expression, as being approximately distributed according to the GUE Tracy–Widom distribution. The latter has a negative mean and, a calculation as in the previous subsection shows that the accumulated loss on the fluctuation scale across all the big squares is finite, in particular of order $(\log q^{-1})^{-1} |\log \log q^{-1}| N^{1/3}$ (ignoring the dependence on u). This means that if we can control the geometric LPP values across all the squares and use appropriate tools on concentration of sums of independent random variables, we will obtain a lower tail inequality for $x_N(N)$.

(In fact, the true behavior of $x_N(N)$ should be $f_q N - \Theta((\log q^{-1})^{-1} N^{1/3})$, i.e., the fluctuation term should not have the log log factor. Our approach does not seem able to achieve this, and we discuss this more ahead in Sect. 1.7, along with some of the consequences of the appearance of the extra log log factor.)

To apply concentration inequalities, we will need control over all the constituent geometric LPP problems. Observe that as the big squares get farther into the environment, the parameter $u^2 q^{2i}$ of the geometric random variables goes to zero. So, in fact, we need a tail bound on the geometric LPP problems which is *uniform* in the parameter q essentially in the entire range $(0, 1)$.

Now, the literature contains extremely sharp estimates on the upper and lower tails of geometric LPP for any fixed parameter q [BDM+01]. Unfortunately, these estimates are not stated uniformly in q in the required range, and the method of proof does not seem like it would yield such an estimate. Indeed, the arguments rely on steepest descent analysis of contour integrals, and the resulting contours implicitly depend on q, thus making it difficult to extract uniform-in-q estimates. Thus we need to prove new results. The following is our second main result and obtains a uniform lower tail in the entire parameter range of q. Here,

$$\mu_q = \frac{(1 + q^{1/2})^2}{1 - q}. \tag{7}$$

Theorem 1.4. *Let T_N be the LPP value from top to bottom of an $N \times N$ square in an environment given by i.i.d. $Geo(q)$ random variables. There exist positive constants c, x_0, and N_0 such that, for $q \in (0, 1)$, $N \geq N_0$, and $x > x_0$,*

$$\mathbb{P}\left(T_N \leq (\mu_q - 1)N - x \cdot \frac{q^{1/6}}{1 - q} N^{1/3}\right) \leq \exp(-cx^{3/2}).$$

We next make some remarks on aspects of this result before outlining how to use Theorem 1.4 to complete the proof of Theorem 1.1.

Remark 1.5 (Effective range of x). Observe that $\mu_q - 1 = \frac{2q^{1/2}(1 + q^{1/2})}{1 - q}$, so the first order term $(\mu_q - 1)N = O(q^{1/2}N/(1 - q))$. Thus, for $x > C(q^{1/3}N^{2/3})$ for some fixed constant C, the probability is actually zero (since $T_N \geq 0$ always). For this reason it will be enough to prove the theorem for $x < \delta(q^{1/2}N)^{2/3}$ for some small $\delta > 0$; then for $\delta(q^{1/2}N)^{2/3} \leq x \leq C(q^{1/2}N)^{2/3}$ one can obtain the claimed bound by modifying the constant c, and beyond that the bound holds trivially.

Remark 1.6 (Effective range of q). Though q is allowed to be arbitrarily close to zero, the statement is really only meaningful when q is lower bounded by a constant times N^{-2}. This is simply because when $q = o(N^{-2})$, then $(\mu_q - 1)N = O(q^{1/2}N) = o(1)$; similarly, the fluctuation scale is also $o(1)$. As a result the upper bound on x of $\delta(q^{1/2}N)^{2/3}$ under which we need to prove the theorem also becomes $o(1)$, where it is trivial. This effective lower bound on q reflects the fact that $q = \Theta(N^{-2})$ is the regime in which the number of points in $[1, N]^2 \cap \mathbb{Z}^2$ where the geometric random variable is non-zero is $O(1)$, and, more precisely, converges to a Poisson random variable; thus the geometric LPP problem converges to Poissonian LPP (see [Joh01]).

Remark 1.7 (Tail exponent of 3/2). While the tail bound we obtain is $\exp(-cx^{3/2})$, the true lower tail behavior is $\exp(-cx^3)$ as proven in [BDM+01] for fixed q. The same should be true uniformly in q as well, i.e., we expect the inequality in Theorem 1.4 with the RHS replaced by $\exp(-cx^3)$ to be true.

As we said earlier, the more usual method of obtaining lower tail bounds via steepest descent analysis of Riemann–Hilbert problems does not appear to be suited to obtain uniform estimates. Instead, we utilize a method, often referred to in the literature as "Widom's trick", which was first introduced by Widom [Wid02] to reduce the task to understanding the trace of the kernel operator of the Meixner ensemble, a determinantal point process associated to geometric LPP via the RSK correspondence. Widom's trick essentially treats the points of the Meixner ensemble as being independent, ignoring the repulsive behavior determinantal point processes exhibit. This simplifies the task of obtaining a lower tail bound, but at the cost of only yielding a tail bound with exponent 3/2.

This can likely be upgraded to the full cubic tail exponent uniformly in q using bootstrapping arguments developed in [GH23], but one would first have to obtain similar uniform lower tail estimates (i.e., with a non-optimal tail exponent like 3/2, though the framework in [GH23] requires only stretched exponential tails) to points displaced (on the $N^{2/3}$ scale) from (N, N). We will not pursue this here as it is not necessary for our bounds on q-pushTASEP, but extending our argument to other directions would involve considering more complicated formulas. Indeed, a formula for the factorial moments of the empirical distribution of the Meixner ensemble (see Sect. 1.6.2) simplifies in the (N, N) case (see (17)) and in the case of other directions one would need to perform asymptotic analysis on an additional layer of summation and handle an extra direction parameter in a uniform way in all of the estimates.

Remark 1.8 (Comparison to exponential LPP). The cubic lower tail exponent has also been obtained in the model of exponential LPP [LR10, BGHK21]; however, the method there is very different and crucially relies on the connection of that LPP value to the top eigenvalue of the Laguerre Unitary Ensemble random matrix theory model [Joh00a] and a certain tridiagonal representation of the same. Geometric LPP does not seem to have any analogous connection to random matrix theory so such techniques are not applicable. We also note that, since exponential random variables have a scale invariance property, one obtains tail estimates for exponential LPP for any rate of the random variables by considering only rate 1. Thus the delicacy of uniformity in the q-parameter for geometric LPP that we must deal with also has no analogue in the exponential case.

1.6.2. Meixner ensemble analysis Let us finally say a few words about what we need to know about the Meixner operator's trace. It is well-known (but proven here for completeness in Lemma 2.3, see also [Led05a]) that the trace can be expressed in terms of the upper tail of the expected empirical distribution $\nu_{q,N}$ of the Meixner ensemble. We then need to obtain a lower bound on the upper tail of $\nu_{q,N}$. An argument of Ledoux given in the context of the GUE in [Led05a] suggests that this can be accomplished by obtaining sharp asymptotics for the moments of $\nu_{q,N}$. In our context, this means that the estimates need to be sharp in both their q and N dependencies.

We obtain these estimates by first doing a careful analysis of formulas available for the factorial or Pochhammer moments of $\nu_{q,N}$ (i.e., $\mathbb{E}[X(X - 1) \cdots (X - k + 1)]$ for $X \sim \nu_{q,N}$) from [Led05b], performing a Laplace method type argument for sums instead of integrals. We then develop arguments to convert these estimates into ones for the polynomial moments. These two tasks comprise the bulk of the technical content of the paper. The difficulty of obtaining these estimates comes primarily from the high order of the moments needed to adapt Ledoux's idea to our setting: indeed, it turns out that we need moments up to order N, unlike in the GUE case where order $N^{2/3}$ sufficed, and with a particular q-dependence for the error term (see Remark 3.2).

Further, in Ledoux's exposition in [Led05a] for the GUE, estimates were obtainable for the polynomial moments directly, unlike here where we must first start with factorial moments. Since $X \sim \nu_{q,N}$ with extremely high probability takes values of order N, the factor by which X^k and $X(X-1)\cdots(X-k+1)$ typically differ when $k = O(N)$ becomes substantial, in fact, exponential in N. Thus the transfer between factorial and polynomial moments becomes delicate, and an exact cancellation needs to happen between factors which appear in the asymptotics for the factorial moments (coming from the value of the maximizer of the exponent in Laplace's method) and the exponential-in-N discrepancy factor just mentioned. It would be interesting to see if there is some other more direct method to obtain the polynomial moment estimates which avoids these cancellations or approaches them in a more systematic way.

However, even with this cancellation, it does not seem tractable to move directly from factorial to polynomial moments. The strategy we instead adopt also makes use of the "layer cake representation" of the polynomial moments (as well as of other quantities): $\mathbb{E}[X^k] = \int_0^\infty kt^{k-1}\mathbb{P}(X > t)\,dt$. This equality shows that, if we have sharp upper bounds for the upper tail of X (with the bounds being applicable in the entire tail), we can get sharp upper bounds for the polynomial moments of X; here by sharp we mean with the right dependencies on the various parameters in the exponent (and not the correct coefficient in the exponent, which we will not be concerned with).

We obtain such probability tail bounds using Markov's inequality combined with the factorial moment asymptotics and the cancellation; in fact, even ignoring the fact that we need moments of order N for our ultimate applications, we require estimates on the same order moments here too in order to obtain the sharp exponentially decaying upper tails (in particular, to get the right dependence on N in Proposition 4.5). Next turning to the lower bound on the polynomial moments, we develop a related but slightly more complicated argument for which also only the upper bounds on the tail of $X \sim \nu_{q,N}$ suffice.

With these precise upper bounds on the upper tail of $\nu_{q,N}$ (which recall is the Meixner ensemble's mean empirical distribution) in hand, it is essentially immediate to also obtain uniform-in-q upper tail estimates for geometric LPP with the correct $3/2$ tail exponent, and we record it below (though we do not need this estimate for any of our arguments concerning q-pushTASEP):

Theorem 1.9. *Let T_N be the LPP value from top to bottom of an $N \times N$ square in an environment given by i.i.d. Geo(q) random variables. There exist positive constants c, C, x_0, and N_0 such that, for $q \in (0,1)$, $N \geq N_0$, and $x_0 \leq x \leq (q^{1/2}N)^{2/3}$,*

$$\mathbb{P}\left(T_N \geq (\mu_q - 1)N + x \cdot \frac{q^{1/6}}{1-q}N^{1/3}\right) \leq C \exp(-cx^{3/2}).$$

For $x \geq (q^{1/2}N)^{2/3}$, the inequality holds with the RHS replaced by $Cq^{-1/4}N^{-1/2}(1+q^{1/6}xN^{-2/3})^{-cN}$.

1.6.3. Tying it together With Theorem 1.4, the last ingredient is a concentration inequality. The inequality must take into account the fact that the scale of the random variables is decreasing. Typical concentration inequalities are for sub-Gaussian tail decay (while here we only have tail exponent $3/2$) and are for deviations from the mean (while our estimates are from the law of large numbers centering). While there are results in the literature for different tail decays, e.g., [KC18], adapting these to address the second point directly to our setting results in a constant order loss for each term being summed,

independent of the scale of the summand. This is too lossy as we have an infinite number of terms. Instead we redo the arguments establishing these bounds, which ultimately rely on estimates on the moment generating function, in such a way to fit our applications. With this final step, we will obtain Theorem 1.1.

Remark 1.10 (An argument for the lower tail of $x_N(T)$). As we saw, the conceptual heart of the argument consisted of finding a concatenation of paths which, to first order, has the same weight as the law of large numbers (5) for the model. Now, if we were interested in $x_N(T)$ for general T, there is also a representation of it in terms of a periodic LPP problem, where the environment consists of periodic rectangles of dimension $N \times T$ instead of $N \times N$ squares as here. However, in such an environment, it is not clear what concatenation of paths would achieve the correct first order weight, and this is why we restrict to $T = N$ in this paper. We leave the general T case for future work.

1.7. A remark on convergence to the log-gamma free energy. Though not needed for the results in this paper, we also note that, as proven in [MP16], the $q \to 1$ limit of X_N^{sc}, when renormalized correctly, is the free energy of the log-gamma polymer introduced in [Sep12] (we refer to the reader to that paper for the precise definition of the model). Indeed for example, setting $q = \exp(-\varepsilon)$ and $u = \exp(-A\varepsilon)$ for a fixed $A > 0$, [MP16, Theorem 8.7] tell us that $\varepsilon(x_N(N) - (2N-1)\varepsilon^{-1}\log\varepsilon^{-1})$ converges in distribution to the log-gamma free energy where the parameters of the inverse gamma random variables are all $2A$. It can be checked that the appropriately normalized $q \to 1$ limit of f_q (as defined in (2)) is indeed the law of large numbers for the log-gamma polymer.

However, notice that the centering term for the convergence is $(2N-1)\varepsilon^{-1}\log\varepsilon^{-1}$, while the first order behavior (in N) we calculated in (6) via the connection to LPP, when written in terms of ε, was $2N\varepsilon^{-1}\log\varepsilon^{-1}$. In other words, there is a discrepancy of $\varepsilon^{-1}\log\varepsilon^{-1}$. This comes from the earlier noted point that the fluctuation scale we are able to prove (when written in terms of ε) is $\varepsilon^{-1}\log\varepsilon^{-1}$, unlike the true fluctuation scale of ε^{-1} suggested by [MP16]; equivalently, our lower tail bound (for $x_N(N)$ and not X_N^{sc}) only kicks in after $\varepsilon^{-1}\log\varepsilon^{-1}N^{1/3}$ into the tail. For this reason, unfortunately, our tail bounds do not survive in the limit to provide a tail bound on the log-gamma free energy.

The ultimate source of the discrepancy in the fluctuation scale that we are able to prove is that we are approximating the true LPP value in the infinite cylinder by a sum of LPP values in $N \times N$ big squares. In more detail, the portion of our path in the ith big square from the top suffers a loss of order $q^{i/6}(1 - q^{2i})^{-1}N^{1/3}$ (ignoring the u-dependence), essentially because this is the scale of fluctuations on which the LPP value in this box converges to the Tracy–Widom distribution, and the latter has a negative mean. Observing that $1 - q^{2i}$ is approximately εi up to constants when $q = \exp(-\varepsilon)$, we see that the sum of this loss from $i = 1$ to ∞ yields an overall loss of of order $N^{1/3}$ times $\varepsilon^{-1}\sum_{i=1}^{\infty}i^{-1}e^{-i\varepsilon/6} = \varepsilon^{-1}\log(1 - e^{-\varepsilon/6}) \approx \varepsilon^{-1}\log(\varepsilon^{-1})$. Thus to avoid the lossy factor of $\log\varepsilon^{-1}$ it seems one would need a different scheme of approximation.

As mentioned earlier in Sect. 1.3.3, very recent work [LS22b] has established a bound (with tail exponent 3/2) on the lower tail of the free energy of a *stationary* version of the log-gamma model (as well as other polymer models such as the O'Connell–Yor model) using a Burke property enjoyed by the model (proved in [OY01] which also introduced the model, and analogous properties in models such as exponential LPP were used to obtain exponentially decaying tail estimates earlier in [EJS20, EJS21, EGO22]), which gives access to formulas for the moment generating function of the free energy. For the O'Connell–Yor model, in [LS22a], these bounds were transferred to the

non-stationary version of the model using geometric arguments involving the polymer measure introduced in [FSV14], and the tail exponent was upgraded to the optimal 3 by adapting geometric methods from [GH23]. One expects that a similar program would deliver the corresponding bounds in the log-gamma case as well.

2. Widom's Trick Applied to the Lower Tail in Geometric LPP

In the next two sections we will prove Theorem 1.4, which provides an upper bound on the lower tail of the LPP value in an i.i.d. geometric environment, uniform in the parameter of the geometric random variables.

The argument relies on a trick introduced by Widom in [Wid02], which we explain next.

2.1. Widom's trick. We first need to introduce the Meixner ensemble, the determinantal point process associated to geometric LPP via the RSK correspondence. The fact that it is determinantal is the crucial property for Widom's argument.

Definition 2.1 (*Meixner ensemble*). First let μ_{Geo}^q denote the $\text{Geo}(q)$ distribution on $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, i.e., the distribution with discrete weights given by

$$\mu_{\text{Geo}}^q(\{x\}) = (1 - q)q^x.$$

For $q \in (0, 1)$ and $N \in \mathbb{N}$, the $N \times N$ *Meixner ensemble* is a determinantal point process on $\mathbb{N}_0$ with kernel given, for $x, y \in \mathbb{N}_0$ with $x \neq y$ and with respect to μ_{Geo}^q, by

$$K_N(x, y) = \frac{\kappa_{N-1}}{\kappa_N} \cdot \frac{M_N(x)M_{N-1}(y) - M_{N-1}(x)M_N(y)}{x - y}; \tag{8}$$

here $M_N = \kappa_N x^N + \kappa_{N-1}x^{N-1} + \cdots + \kappa_0$ are the orthonormal polynomials (which we call the Meixner polynomials, though it differs from the classical Meixner polynomials by a constant multiple due to the normalization) with respect to μ_{Geo}^q. The second factor on the right-hand side of (8) makes sense for $x, y \in \mathbb{R}$, and so the $x = y$ case can be defined by taking the appropriate limit.

Here is the relation between the Meixner ensemble and the geometric LPP value.

Proposition 2.2 (Proposition 1.3 of [Joh00a]). *Fix* $q \in (0, 1)$ *and* $N \in \mathbb{N}$. *Let* $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_N$ *be distributed according to the* $N \times N$ *Meixner ensemble and let* T_N *be the LPP value in the environment of i.i.d. Geo(q) random variables. Then* $T_N \overset{d}{=} \lambda_1 - N + 1$.

With this background we may explain Widom's trick. The fact that $(\lambda_1, \ldots, \lambda_N)$ is determinantal with kernel K_N given by (8) implies, using the Cauchy–Binet formula, that, for any $t \in \mathbb{R}$,

$$\mathbb{P}(\lambda_1 \leq t) = \det\left(I_N - K_N^t\right),$$

where K_N^t can be written as the Gram matrix of the Meixner polynomials, i.e.,

$$K_N^t = \left(\langle M_{\ell-1}, M_{k-1} \rangle_{\ell^2(\{t,t+1,\ldots\}, \mu_{\text{Geo}}^q)}\right)_{1 \leq k, \ell \leq N}. \tag{9}$$

The fact that Gram matrices are positive semi-definite implies that the eigenvalues of K_N^t are non-negative; also, we may write, for any unit vector $u \in \mathbb{R}^N$ and with $g(x) = \sum_{i=1}^{N} u_i M_{i-1}(x)$,

$$1 = \sum_{i=1}^{N} u_i^2 = \langle g, g \rangle_{\ell^2(\mathbb{N}_0, \mu_{\text{Geo}}^q)} = \langle g \mathbb{1}_{\cdot < t}, g \mathbb{1}_{\cdot < t} \rangle_{\ell^2(\mathbb{N}_0, \mu_{\text{Geo}}^q)} + \langle g \mathbb{1}_{\cdot \geq t}, g \mathbb{1}_{\cdot \geq t} \rangle_{\ell^2(\mathbb{N}_0, \mu_{\text{Geo}}^q)}$$

$$\geq \langle g \mathbb{1}_{\cdot \geq t}, g \mathbb{1}_{\cdot \geq t} \rangle_{\ell^2(\mathbb{N}_0, \mu_{\text{Geo}}^q)} = \langle g, g \rangle_{\ell^2(\{t, t+1, \dots\}, \mu_{\text{Geo}}^q)} = u^T K_N^t u,$$

which in turn implies that the eigenvalues of K_N^t are at most 1.

Let us label the eigenvalues of K_N^t as $\rho_1^t, \dots, \rho_N^t$. Since $1 - x \leq e^{-x}$ for $x \in [0, 1]$,

$$\mathbb{P}(\lambda_1 \leq t) = \det(I_N - K_N^t) = \prod_{i=1}^{N}(1 - \rho_i^t) \leq \exp\left(-\sum_{i=1}^{N} \rho_i^t\right) = \exp\left(-\text{Tr}(K_N^t)\right).$$

Thus Widom's trick reduces the problem of bounding the lower tail to understanding the trace of an associated operator. This in turn can be accomplished by lower bounding the upper tail of the expected empirical distribution $\nu_{q,N}$ of the Meixner ensemble, defined precisely by

$$\nu_{q,N} = \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}\right]. \tag{10}$$

We record the connection between the operator's trace and the tail of $\nu_{q,N}$ next.

Lemma 2.3. *For any $t \in \mathbb{N}$, $\text{Tr}(K_N^t) = N \nu_{q,N}([t, \infty))$.*

Proof. First we observe that, from (9),

$$\text{Tr}(K_N^t) = \sum_{\ell=0}^{N-1} \langle M_{\ell-1}, M_{\ell-1} \rangle_{\ell(\{t, t+1, \dots\}, \mu_{\text{Geo}}^q)} = \int_t^{\infty} \sum_{\ell=0}^{N-1} M_\ell^2 \, d\mu_{\text{Geo}}^q. \tag{11}$$

Now since $(\lambda_1, \dots, \lambda_N)$ is determinantal with kernel K_N with respect to μ_{Geo}^q, it is a standard fact of the theory of determinantal point processes (or see [Led05a, Proposition 1.2]) that, for any bounded measurable $f : \mathbb{N}_0 \to \mathbb{R}$,

$$\mathbb{E}\left[\prod_{i=1}^{N}[1 + f(\lambda_i)]\right] = \sum_{r=0}^{N} \frac{1}{r!} \int_{\mathbb{N}_0^r} \prod_{i=1}^{r} f(x_i) \det(K_N(x_i, x_j))_{1 \leq i, j \leq r} \, d\mu_{\text{Geo}}^q(x_1)$$
$$\cdots d\mu_{\text{Geo}}^q(x_r).$$

Replacing f by εf, taking the $\varepsilon \to 0$ limit, and thereby equating the order ε terms on both sides (since the constant-in-ε terms on both sides are easily seen to be 1), we obtain that

$$\mathbb{E}\left[\sum_{i=1}^{N} f(\lambda_i)\right] = \int_{\mathbb{N}_0} f(x) K_N(x, x) \, d\mu_{\text{Geo}}^q(x).$$

By the Christoffel–Darboux formula and (8), $K_N(x, x) = \sum_{\ell=0}^{N-1} M_\ell(x)^2$. With this, taking $f(x) = \mathbb{1}_{x \geq t}$ and using (11) yields the claim. $\square$

So the task is now to obtain a lower bound on the upper tail of $\nu_{q,N}$. The bound we prove is stated in the next theorem, and its proof will be the main goal of the remainder of this section as well as of the next two.

Theorem 2.4. *Let X be distributed as $\nu_{q,N}$ as defined in* (10) *and let μ_q be as in* (7). *There exist positive absolute constants c, C, and N_0 such that, for $N \geq N_0$, $\varepsilon \in [CN^{-2/3}, 1]$, and $q \in [\varepsilon^3, 1)$,*

$$\mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}\varepsilon) \right) \geq c\varepsilon^{3/2}.$$

In fact, the lower bound of order $\varepsilon^{3/2}$ is sharp and we prove a matching order upper bound in Proposition 4.4.

With Theorem 2.4 and Widom's trick, Theorem 1.4's proof is straightforward.

Proof of Theorem 1.4. Recall from Proposition 2.2 that if λ_1 is distributed as the largest particle of the Meixner ensemble, then $T_N \stackrel{d}{=} \lambda_1 - N + 1$. From Lemma 2.3, for any $t \in \mathbb{N}$,

$$\mathbb{P}\left(\lambda_1 \leq t \right) \leq \exp\left(-N\nu_{q,N}([t, \infty)) \right).$$

So we see that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(T_N \leq (\mu_q - 1)N - \mu_q N q^{1/6}\varepsilon \right) = \mathbb{P}\left(\lambda_1 \leq \mu_q N(1 - q^{1/6}\varepsilon) \right)$$
$$\leq \exp\left\{ -N\nu_{q,N}\left([\mu_q N(1 - q^{1/6}\varepsilon), \infty) \right) \right\}.$$

By Theorem 2.4, there exist positive constants c and c_0 such that, for all $q \in [\varepsilon^3, 1)$ and $CN^{-2/3} \leq \varepsilon \leq 1$,

$$\nu_{q,N}\left([\mu_q N(1 - q^{1/6}\varepsilon), \infty) \right) \geq c\varepsilon^{3/2}.$$

Putting the above together with $\varepsilon = xN^{-2/3}(1-q)^{-1}\mu_q^{-1} = xN^{-2/3}(1+q^{1/2})^2$ for $x > C$ and adjusting the constant in the exponent gives

$$\mathbb{P}\left(T_N \leq (\mu_q - 1)N - x\frac{q^{1/6}}{1-q}N^{1/3} \right) \leq \exp\left(-cx^{3/2} \right)$$

when $q \geq x^3 N^{-2}$ and $xN^{-2/3} \leq 1$ (since $q \geq \varepsilon^3$ and $\varepsilon \leq 1$ are required to apply Theorem 2.4), which are both implied by the hypothesis that $x \leq (q^{1/2}N)^{2/3}$. This completes the proof. □

To prove Theorem 2.4, we rely on a strategy of Ledoux explained in [Led05a, Section 5]. It relies on getting strong estimates on polynomial moments of the mean empirical distribution $\nu_{q,N}$. The bounds we prove are the following.

Theorem 2.5. *Let X be distributed according to $\nu_{q,N}$ as defined in* (10). *There exist positive C, α_0, and N_0 such that for any $N \geq N_0$ and (k, q) satisfying $k \leq (\alpha_0 N \wedge q^{-1/6}N^{2/3})$ and $q \in [k^{-2}, 1)$,*

$$C^{-1}(q^{1/6}k)^{-3/2}(\mu_q N)^k \leq \mathbb{E}[X^k] \leq C(q^{1/6}k)^{-3/2}(\mu_q N)^k.$$

Observe that the moments grow to first order like $(\mu_q N)^k$, which reflects that we expect $\nu_{q,N}$ to be supported on $[0, \mu_q N]$ (though more precisely there is a decaying-in-N amount of mass beyond this point, which we will in fact upper bound in Proposition 4.5). The polynomial dependence on k is what captures the behaviour of the tail of $\nu_{q,N}$ near this right edge, and this is the basic observation of Ledoux' argument. Indeed, the exponent of $-3/2$ for k is what gives the $3/2$ exponent of ε in Theorem 2.4.

Note also that we allow k to go up to a rather large value; essentially of order N, since, when $q = \Theta(N^{-2})$, then $q^{-1/6}N^{2/3} = \Theta(N)$. We will in fact need the estimate to allow such large values of k in the application, namely the proof of Theorem 2.4. We expand on this slightly ahead in Remark 2.7. Finally we remark that the expression $q^{-1/6}N^{2/3}$ in the upper bound on k is a technical feature, and may be an artifact, of the conversion we perform from factorial to polynomial moments (indeed, this expression does not appear in the bound on k in the upcoming Theorem 3.1 on factorial moment estimates).

We will prove Theorem 2.5 in Sects. 3 and 4; the upper and lower bounds are separated into Propositions 4.6 and 4.7 respectively.

We conclude this section by using Theorem 2.5 to implement Ledoux' argument to establish Theorem 2.4.

Proof of Theorem 2.4. First, by the Cauchy–Schwarz inequality, we have

$$\mathbb{E}[X^{2k}\mathbb{1}_{X\geq\mu_q N(1-q^{1/6}\varepsilon)}] \leq \mathbb{E}[X^{4k}]^{1/2}\mathbb{P}\left(X \geq \mu_q N(1-q^{1/6}\varepsilon)\right)^{1/2}. \qquad (12)$$

By Theorem 2.5, when $q \geq k^{-2}$, $4k \leq (\alpha_0 N \wedge q^{-1/6}N^{2/3})$ and $N \geq N_0$ (conditions we assume in the rest of the proof),

$$\mathbb{E}[X^{4k}] \leq C_1(q^{1/6}k)^{-3/2}(\mu_q N)^{4k}.$$

It is also easy to see that

$$\mathbb{E}[X^{2k}\mathbb{1}_{X\geq\mu_q N(1-q^{1/6}\varepsilon)}] = \mathbb{E}[X^{2k}] - \mathbb{E}[X^{2k}\mathbb{1}_{X<\mu_q N(1-q^{1/6}\varepsilon)}]$$
$$\geq \mathbb{E}[X^{2k}] - \mathbb{E}[X^k]\left(\mu_q N(1-q^{1/6}\varepsilon)\right)^k. \qquad (13)$$

Now, from Theorem 2.5, under the same conditions on (q, k) as above,

$$\mathbb{E}[X^{2k}] \geq C_2(q^{1/6}k)^{-3/2}(\mu_q N)^{2k},$$

while, again from Theorem 2.5 and the same conditions on (q, k),

$$\mathbb{E}[X^k] \leq C_3(q^{1/6}k)^{-3/2}(\mu_q N)^k.$$

Since overall it holds (under the condition that the expression in the parentheses is positive) from (12) and (13) that

$$\mathbb{P}\left(X \geq \mu_q N(1-q^{1/6}\varepsilon)\right) \geq \mathbb{E}[X^{4k}]^{-1}\left(\mathbb{E}[X^{2k}] - \mathbb{E}[X^k](\mu_q N)^k(1-q^{1/6}\varepsilon)^k\right)^2,$$

substituting the above recalled bounds on the moments of X yields (cancelling out all the common factors of $(\mu_q N)^{4k}$ on the right-hand side)

$$\mathbb{P}\left(X \geq \mu_q N(1-q^{1/6}\varepsilon)\right)$$

$$\geq C_1^{-1}(q^{1/6}k)^{3/2} \left[C_2(q^{1/6}k)^{-3/2} - C_3(q^{1/6}k)^{-3/2}(1 - q^{1/6}\varepsilon)^k \right]^2$$

$$\geq C_1^{-1}(q^{1/6}k)^{-3/2} \left[C_2 - C_3 \exp(-q^{1/6}\varepsilon k) \right]^2. \tag{14}$$

We need to set k such that the expression in the square brackets is uniformly positive; it is sufficient if

$$\frac{C_3}{C_2} \exp\left(-q^{1/6}\varepsilon k\right) < \frac{1}{2}. \tag{15}$$

We will verify that this condition is met with $k = Wq^{-1/6}\varepsilon^{-1}$ for a large absolute constant W to be chosen: the expression on the LHS of the previous display becomes

$$\frac{C_3}{C_2} \exp\left(-W\right). \tag{16}$$

We pick W such that $\frac{C_3}{C_2} \exp(-W) < \frac{1}{2}$. With this our choice of k has been made, and we next verify that the conditions that k, q, ε must satisfy to apply Theorem 2.5 indeed hold. After that we will check that (15) holds.

Recall that under the hypotheses of Theorem 2.4 that we are proving, we have a lower bound on ε of the form $N^{-2/3}$ times an absolute constant that we are free to choose and which we label c_0^{-1}, i.e., ε satisfies $\varepsilon > c_0^{-1}N^{-2/3}$. We also have $q \geq \varepsilon^3$. We need to verify two conditions: (i) $q \geq k^{-2}$ and (ii) $k \leq \frac{1}{4}\min(\alpha_0 N, q^{-1/6}N^{2/3})$.

For (i), it is easy to check that $q \geq k^{-2}$ is equivalent to $q \geq W^{-3}\varepsilon^3$ for our choice of k. If $W \geq 1$ (which we can ensure by raising W if needed), then clearly our hypothesis that $q \geq \varepsilon^3$ implies $q \geq k^{-2}$.

Next we set $c_0 > 0$ small enough (depending on W) that the condition $\varepsilon > c_0^{-1}N^{-2/3}$ implies $k \leq \frac{1}{4}\min(\alpha_0 N, q^{-1/6}N^{2/3})$ (i.e., that $\varepsilon > 4\max(W, \alpha_0^{-1})N^{-2/3}$, i.e., $c_0 < \frac{1}{4}\min(W^{-1}, \alpha_0)$). This verifies that the upper bound (ii) above on k holds, and thus that the hypotheses of Theorem 2.5 are satisfied.

Using (15) in (14) and substituting $k = Wq^{-1/6}\varepsilon^{-1}$ in the same, we obtain

$$\mathbb{P}\big(X \geq \mu_q N(1 - q^{1/6}\varepsilon)\big) \geq \tfrac{1}{4}C_1^{-1}C_2^2 W^{-3/2}\varepsilon^{3/2}.$$

Thus, overall, our choice of parameters yields that there exist positive absolute constants c and c_0, such that for all $\varepsilon > c_0^{-1}N^{-2/3}$ and $q \geq \varepsilon^3$,

$$\mathbb{P}\big(X \geq \mu_q N(1 - q^{1/6}\varepsilon)\big) \geq c\varepsilon^{3/2}.$$

This completes the proof.   □

Remark 2.6. Unfortunately, it does not seem possible to adapt the above argument to work with estimates on the factorial moments directly instead of first converting to polynomial moments, as such an adaptation would allow us to avoid the analysis to go from Theorem 3.1 to Theorem 2.5.

The reason is essentially that factorial moments do not work well with the Cauchy–Schwarz inequality: $\mathbb{E}[(X_k)] \leq \mathbb{E}[(X)_k^2]^{1/2}$, but we do not have any control on the RHS via estimates on factorial moments. This is in stark contrast to polynomial moments, where applying Cauchy–Schwarz results in another polynomial moment, which we do have estimates on.

Remark 2.7. We observe that the proof used the polynomial moment of $X \sim \nu_{q,N}$ of order $q^{-1/6}\varepsilon^{-1}$. When $\varepsilon = N^{-2/3}$ (which falls within the parameter ranges we are interested in since such a value of ε correspond to the KPZ fluctuation scale), the moments we are considering become of order $q^{-1/6}N^{2/3}$ (which further becomes N if $q = \Theta(N^{-2})$), and this is why we need Theorem 2.5 to allow k to go up to $q^{-1/6}N^{2/3}$.

3. Sharp Factorial Moment Bounds

Here we prove Theorem 2.5 on sharp bounds for the moments of the expected empirical distribution $\nu_{q,N}$ (as defined in (10)) of the Meixner ensemble. We adopt the notation

$$(x)_k := x(x-1)\cdots(x-k+1).$$

While there is no explicit formula available for the polynomial moments of $\nu_{k,N}$, there *is* one for the factorial moments, i.e., moments of the form $\mathbb{E}[(X)_k] = \mathbb{E}[X(X-1)\cdots(X-k+1)]$. Indeed letting X be distributed according to $\nu_{q,N}$, [Led05b, Lemma 5.2] states that

$$M^q(k, N) := \mathbb{E}[(X)_k] = \frac{q^k}{(1-q)^k} \sum_{i=0}^{k} q^{-i} \binom{k}{i}^2 \cdot \sum_{\ell=i}^{N-1} \frac{(\ell+k-i)!}{(\ell-i)!}.$$

In fact, this simplifies [CCO20, Eq. (3.5)] to

$$M^q(k, N) = \frac{q^k}{(1-q)^k} \frac{1}{N} \cdot \frac{1}{k+1} \sum_{i=0}^{k} q^{-i} \binom{k}{i}^2 \cdot \frac{(N+k-i)!}{(N-i-1)!}. \tag{17}$$

Our approach is to use this formula to obtain asymptotics on the factorial moments, and then later convert them into polynomial moments.

3.1. The factorial moment asymptotics.

Theorem 3.1. *Let X be distributed according to $\nu_{q,N}$ as defined in (10), and let $\mu_q = (1+q^{1/2})^2/(1-q)$ be as in (7). There exist positive constants C, N_0, and k_0 such that for all $N \geq N_0$, $k_0 \leq k \leq \frac{1}{2}N$, and $q \in [k^{-2}, 1)$, $M^q(k, N)$ is upper and lower bounded by (and letting $\alpha := k/N$)*

$$(q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(-N\left[(\mu_q - \alpha)\log\left(1 - \frac{\alpha}{\mu_q}\right) + \alpha\right]\right)$$

up to factors of $C \exp\left(Cq^{1/2}\frac{k^3}{N^2}\right)$ or its inverse. Further, the upper bound holds for all $q \in (0, 1)$.

As we said, $\nu_{N,q}$ has right edge of support roughly $\mu_q N = \frac{(1+q^{1/2})^2}{1-q}N$, and this is what gives that to first order, $M^q(k, N)N^{-k}$ grows like μ_q^k. The main task is to obtain the correct polynomial dependence on k and q, namely $q^{-1/4}k^{-3/2}$, of the same.

It turns out that the exponential factor in the theorem is the quantity to go between factorial and polynomial moments. In Sect. 4, when we use Theorem 3.1 to prove Theorem 2.5 on sharp upper and lower bounds on the polynomial moments, this factor will get

canceled. In fact, the expression in the exponent that one gets by directly doing Laplace's method on the sum (as is our broad strategy to prove Theorem 3.1 as we indicated in the introduction) is not the one recorded above. The recorded expression is instead obtained by identifying an additional cancellation as mentioned in the idea of proofs Sect. 1.6; this cancellation is isolated as Lemma B.2 in Appendix B, where most of the technical estimates required for Theorem 3.1 are done.

Remark 3.2. Let us also emphasize the fact that the coefficient of k^3/N^2 in the error term is of order $q^{1/2}$ and not unit order, which would have been easier to obtain (and would suffice if uniform-in-q bounds are not required). As we saw at the end of the previous section, the $q^{1/2}$ factor was important in obtaining the uniform lower bound in Theorem 2.4, as it ensures that $q^{1/2}k^3/N^2$ does not blow up as $q \to 0$ for the choice of k (which recall grows as $q^{-1/6}$).

Before turning to giving the proof of Theorem 3.1 in full (though as mentioned much of the technical estimates have been relegated to Appendix B), let us outline the strategy. First, we will use Stirling's approximation and a precise form of the fact that $\binom{k}{i} \approx \exp(kH(i/k))$ (with $H(x) = -x \log x - (1-x) \log(1-x)$ the entropy function) to write the sum in (17) as, approximately and up to absolute constants,

$$\frac{N^k q^k}{(1-q)^k} \cdot k^{-1} \cdot \frac{1}{k+1} \sum_{i=0}^{k} \frac{k^2}{i(k-i)} \exp\Big(i \log q^{-1} + 2k \cdot H(i/k)$$

$$+ (k+1) \log \Big(1 + \frac{k-i}{N} \Big) + (N-i-1) \log \Big(1 + \frac{k+1}{N-i-1} \Big) - (k+1) \Big). \quad (18)$$

Then, the idea is to obtain asymptotics for the sum using Laplace's method. Indeed, if we were to write k as αN and regard the sum (along with the $(k+1)^{-1}$ factor as being approximately the integral

$$\int_0^1 \frac{1}{x(1-x)} \exp\left(N f_\alpha(x) \right) \, dx,$$

where

$$f_\alpha(x) = \alpha x \log q^{-1} + 2\alpha H(x) + \alpha \log(1 + \alpha(1-x))$$

$$+ (1-\alpha x) \log \Big(1 + \frac{\alpha}{1-\alpha x} \Big) - \alpha, \quad (19)$$

then, letting $g(x) = (x(1-x))^{-1}$, one would expect from the Laplace method heuristic that the sum in (18) is approximately

$$(N|f_\alpha''(x_0)|)^{-1/2} \exp(N f_\alpha(x_0)) g(x_0),$$

up to constants, where $x_0 = x_0(\alpha)$ is the maximizer of f_α on $[0, 1]$. Evaluating $f_\alpha(x_0)$ and $g(x_0)$ will yield the claimed q and k dependencies in Theorem 3.1.

There are existing results in the literature, for example [Mas14], on Laplace's method for sums which also obtain the correct constant coefficient multiplying the previous display. However these are not directly useful to us: we need all our estimates to be uniform in the parameter q, which can be difficult to verify after applying black-box theorems. For this reason we perform the analysis explicitly ourselves; but we will not be

concerned with obtaining the correct constant dependencies as these are not necessary for our ultimate applications. However, since there is not much novel or probabilistic content in these computations, we defer to Appendix B the proofs of these bounds (which are stated as Propositions B.1 and B.5).

Proof of Theorem 3.1. In the following, we will include error terms of the form $\pm O(r(k, N))$ in front to emphasize that we allow the error to be positive or negative, as long as it is in magnitude bounded by $C r(k, N)$ for some constant C not depending on q, i, k, α, or N, in the range $(0 < \alpha \le \frac{1}{2})$. Similarly, a factor of $\Theta(1)$ indicates that the equality holds up to a factor of an absolute constant.

We will first obtain a bound on the summands in (17). We start with bounding $(N + k - i)!/(N - i - 1)!$ using Stirling's approximation (non-asymptotic form) to obtain

$$
\begin{aligned}
\frac{(N + k - i)!}{(N - i - 1)!} \\
= \frac{\sqrt{2\pi(N + k - i)}}{\sqrt{2\pi(N - i - 1)}} \exp\Big[(N + k - i)\log(N + k - i) \\
- (N + k - i) - (N - i - 1)\log(N - i - 1) + (N - i - 1) \pm O(N^{-1}) \Big] \\
= \Theta(1) \cdot \exp\Big[(N - i - 1)\log\frac{N + k - i}{N - i - 1} + (k + 1)\big(\log(N + k - i) - 1\big) \Big] \\
= \Theta(1) \cdot \exp\Big[(N - i - 1)\log\Big(1 + \frac{k + 1}{N - i - 1}\Big) + (k + 1)\big(\log(N + k - i) - 1\big) \Big].
\end{aligned}
$$

Substituting the above expression into that of $M^q(k, N)$ (17) and multiplying by N^{-k}, we obtain

$$
\begin{aligned}
M^q(k, N)N^{-k} &= \frac{\Theta(1)q^k}{(1 - q)^k(k + 1)} \sum_{i=0}^{k} q^{-i} \binom{k}{i}^2 \exp\Big[(N - i - 1)\log\Big(1 + \frac{k + 1}{N - i - 1}\Big) \\
&\quad + (k + 1)\log(N + k - i) - (k + 1)\log N - (k + 1) \Big] \\
&= \frac{\Theta(1)q^k}{(1 - q)^k(k + 1)} \sum_{i=0}^{k} q^{-i} \binom{k}{i}^2 \exp\Big[(N - i - 1)\log\Big(1 + \frac{k + 1}{N - i - 1}\Big) \\
&\quad + (k + 1)\log\Big(1 + \frac{k - i}{N}\Big) - (k + 1) \Big].
\end{aligned}
$$

Now since

$$
\binom{k}{i} = \Theta(1)\sqrt{\frac{k}{i(k - i)}} \exp(k H(i/k)),
$$

where $H(p) = -p \log p - (1 - p)\log(1 - p)$ is the entropy function, we obtain that $M^q(k, N)N^{-k}$ is equal to

$$
\frac{\Theta(1)q^k}{(1 - q)^k} \cdot k^{-1} \cdot \frac{1}{k + 1} \sum_{i=0}^{k} \frac{k^2}{i(k - i)}
$$

$$\exp\left[i \log q^{-1} + 2kH(i/k) + (N - i - 1) \log\left(1 + \frac{k+1}{N - i - 1}\right)\right.$$
$$\left. + (k + 1) \log\left(1 + \frac{k - i}{N}\right) - (k + 1)\right].$$

Rewriting the previous display a little, we obtain

$$\frac{\Theta(1)q^k}{(1 - q)^k} \cdot k^{-1} \cdot \frac{1}{k+1} \sum_{i=0}^{k} \frac{1}{\frac{i}{k}(1 - \frac{i}{k})} \exp\left[N\left\{\frac{i}{N} \log q^{-1} + \frac{2k}{N} \cdot H(i/k)\right.\right.$$
$$\left.\left. + \left(1 - \frac{i + 1}{N}\right) \log\left(1 + \frac{k+1}{N - i - 1}\right) + \frac{k+1}{N} \log\left(1 + \frac{k - i}{N}\right) - \frac{k+1}{N}\right\}\right]. \quad (20)$$

Note that the expression inside the curly brackets in the exponent is $f_\alpha(i/k)$ with $\alpha = k/N$ and f_α as defined in (19). The sum is upper and lower bounded using Propositions B.1 and B.5 (which perform the Laplace's method-type asymptotics mentioned above) in Appendix B, up to a factor of $C \exp(Cq^{1/2}k^3/N^2)$ or its inverse, and for $N \geq N_0$, $k_0 \leq k \leq N$, and $q \geq k^{-2}$, by

$$q^{-1/4}k^{1/2}\left[\frac{(1 + q^{1/2})^2}{q}\right]^k \exp\left(-N\left[(\mu_q - \alpha) \log\left(1 - \frac{\alpha}{\mu_q}\right)\right]\right).$$

Substituting this into (20) yields that $M^q(k, N)N^{-k}$ is equal to

$$\Theta(1) \cdot k^{-3/2}q^{-1/4}\left[\frac{(1 + q^{1/2})^2}{1 - q}\right]^k$$
$$\exp\left(-N\left[(\mu_q - \alpha) \log\left(1 - \frac{\alpha}{\mu_q}\right)\right] \pm O(1)q^{1/2}\frac{k^3}{N^2}\right), \quad (21)$$

which completes the proof after recalling that $\mu_q = (1 + q^{1/2})^2/(1 - q)$. □

4. Translating from Factorial to Polynomial Moments

Now we convert the factorial moment bounds of Theorem 3.1 to polynomial bounds and thus prove Theorem 2.5. The basic idea is the following. Notice that the difference between the bounds in the two theorems is essentially the factor of $\exp(-N[(\mu_q - \alpha) \log(1 - \frac{\mu_q}{\alpha}) + \alpha])$ in the bounds for the factorial moments, where $\alpha = k/N$. The presence of this factor is a problem because it is much smaller than $O(1)$. So our goal is to show that that factor goes away when we move to the kth polynomial moment.

Now, we can write the ratio $[X(X - 1) \cdots (X - k + 1)]/X^k$ as

$$\prod_{i=1}^{k-1}(1 - iX^{-1}) = \exp\left(-\sum_{i=1}^{k} \log\left(1 - \frac{i}{X}\right)\right).$$

If we believe that the support of X essentially has upper boundary $\mu_q N$, then, when considering high powers of X, heuristically one should get the correct behavior when replacing X by $\mu_q N$. Doing so, we can recognize the sum in the previous display as N

times a Riemann sum of the integral $\int_0^\alpha \log(1 - x/\mu_q)\, \mathrm{d}x = (\alpha - \mu_q) \log(1 - \frac{\alpha}{\mu_q}) - \alpha$. This exactly cancels the extra factor we noted above.

Because this factor which needs to be cancelled will come up a number of times in this section, let us adopt some notation for it. For $\alpha \in (0, 1)$, define $h_\alpha : [\alpha, \infty) \to \mathbb{R}$ by

$$h_\alpha(x) = (\alpha - x) \log\left(1 - \frac{\alpha}{x}\right) - \alpha. \tag{22}$$

The proof we provide does not exactly make the above heuristic precise. While the conversion between factorial and polynomial expressions outlined above essentially holds for deterministic x (see for example Lemma 4.1 ahead), due to the randomness of X, the procedure is not able to provide sharp estimates on the kth moment of X in the entire range of q and k that we will require. What we will do instead is obtain sharp upper bounds on the upper tail of X using the factorial moments, and then use the general formula (sometimes called the layer cake representation) $\mathbb{E}[X^k] = \int_0^\infty kx^{k-1}\mathbb{P}(X > x)\, \mathrm{d}x$ and variants to obtain the sharp estimates on $\mathbb{E}[X^k]$ in the complete parameter range we need (as was discussed briefly in the introduction in Sect. 1.6.2).

(In fact, for lower order moments it is possible to make the heuristic precise in a more direct way, without making use of the layer cake representation. For example, for moments of order $N^{2/3}$, the factor by which $\mathbb{E}[X^k]$ and $\mathbb{E}[(X)_k]$ differs becomes of unit order, and it is possible to approximate $h_\alpha(x)$ by $-\alpha^2/(2x) \pm O(\alpha^3)$ where recall $\alpha = k/N$, since $\alpha^3 N = k^3/N^2 = \Theta(1)$ when $k = \Theta(N^{2/3})$. But because we require moments up to order N, the errors in these approximations become much too large and so we must keep all the expressions, such as h_α, as they are and track all the cancellations.)

In the next Sect. 4.1 we collect some general statements connecting factorial and polynomial moments. In Sect. 4.2 we use the estimates on factorial moments to obtain sharp upper bounds on $\mathbb{P}(X > x)$. Finally in Sects. 4.3 and 4.4 we will use these estimates and the layer cake representation to obtain the upper and lower bounds respectively on $\mathbb{E}[X^k]$.

4.1. General lemmas connecting factorial and polynomial moments.

Lemma 4.1. Let $k, N, x \in \mathbb{N}$ with $k \leq x$ and let $\alpha = k/N$, $\beta = x/N$. Then, with h_α as in (22),

$$(x)_k \geq x^k \exp\left(Nh_\alpha(\beta)\right).$$

Proof. We see that

$$(x)_k = x(x-1)\cdots(x-k+1) = x^k \times \prod_{i=1}^{k-1}\left(1 - \frac{i}{x}\right) = x^k \cdot \exp\left(\sum_{i=1}^{k-1}\log\left(1 - \frac{i}{x}\right)\right).$$

Now, the RHS is lower bounded by

$$x^k \cdot \exp\left(\sum_{i=1}^{k-1}\log\left(1 - \frac{i}{\beta N}\right)\right) = x^k \cdot \exp\left(\sum_{i=1}^{\alpha N-1}\log\left(1 - \frac{i}{\beta N}\right)\right).$$

We recognize the sum in the exponent as N times the left Riemann sum associated to the integral $\int_0^\alpha \log(1 - x/\beta)\, \mathrm{d}x = (\alpha - \beta) \log(1 - \frac{\alpha}{\beta}) - \alpha$. Since $x \mapsto \log(1 - x/\beta)$

is a decreasing function, the integral is a lower bound for the left Riemann sum, i.e., the previous display is lower bounded by

$$x^k \cdot \exp\left(N\left[(\alpha - \beta)\log\left(1 - \frac{\alpha}{\beta} \right) - \alpha \right] \right).$$

Rearranging completes the proof. □

Lemma 4.2. *Let X be a non-negative integer-valued random variable. Then for any $k, N \in \mathbb{N}$ with $k \leq N/2$, $\alpha := k/N$, and $\beta > \alpha$, with h_α as in (22),*

$$\mathbb{E}[X^k] \geq \left(\mathbb{E}[(X)_k] - \mathbb{E}[(X)_k \mathbb{1}_{X > \beta N}] \right) \cdot \exp\left(-N h_\alpha(\beta) - \tfrac{1}{4}(\beta - \alpha)^{-1} \right).$$

Proof. As in the previous proof, we see that, under the condition that $x \leq \beta N$,

$$(x)_k = x^k \times \prod_{i=1}^{k-1}\left(1 - \frac{i}{x} \right) = x^k \times \exp\left(\sum_{i=1}^{k-1} \log\left(1 - \frac{i}{x} \right) \right) \leq x^k$$

$$\times \exp\left(\sum_{i=1}^{\alpha N - 1} \log\left(1 - \frac{i}{\beta N} \right) \right).$$

We again recognize the sum as N times the left Riemann sum of $\int_0^\alpha \log(1 - x/\beta)\,\mathrm{d}x = (\alpha - \beta)\log(1 - \frac{\alpha}{\beta}) - \alpha$. The absolute value of the deviation between the Riemann sum of f and the associated integral is bounded by $\sup_{x \in [0,\alpha]} |f'(x)|\alpha/N$ (e.g. by bounding $|f(x) - f(i/N)|$ on $[i/N, (i+1)/N]$ by $\sup_{[0,\alpha]} |f'|/N$, integrating the inequality on $[i/N, (i+1)/N]$, and summing over $i = 1, \ldots \alpha N$), and we can evaluate this expression to be $\alpha/((\beta - \alpha)N)$ for our particular f. So we see that, when $x \leq \beta N$,

$$(x)_k \leq x^k \exp\left(N\left[(\alpha - \beta)\log\left(1 - \frac{\alpha}{\beta} \right) - \alpha \right] + (\beta - \alpha)^{-1}\alpha \right). \tag{23}$$

So with this, and since X is non-negative, for any $\beta > \alpha$ and since $\alpha \leq \frac{1}{2}$,

$$\mathbb{E}[X^k] \geq \mathbb{E}[X^k \mathbb{1}_{X \leq \beta N}] \geq \mathbb{E}\left[(X)_k \mathbb{1}_{X \leq \beta N} \right]$$

$$\cdot \exp\left(N\left[(\beta - \alpha)\log\left(1 - \frac{\alpha}{\beta} \right) + \alpha \right] - \tfrac{1}{2}(\beta - \alpha)^{-1} \right).$$

Writing $(X)_k \mathbb{1}_{X \leq \beta N}$ as $(X)_k - (X)_k \mathbb{1}_{X > \beta N}$ completes the proof. □

In the next lemma we record some basic properties of h_α which will be useful ahead.

Lemma 4.3. *Recall $h_\alpha(x) = (\alpha - x)\log(1 - \frac{\alpha}{x}) - \alpha$ from (22). Then h_α is increasing and, for $\alpha \in (0, \frac{1}{2}]$ and $x \geq y$,*

$$h_\alpha(x) - h_\alpha(y) \leq \begin{cases} \alpha^2 y^{-2}(x - y) & y \geq 2\alpha \\ \alpha^2 & y \geq 1. \end{cases}$$

In particular, if $\alpha \in (0, \frac{1}{2}]$ and $x \geq y \geq 1$,

$$h_\alpha(x) - h_\alpha(y) \leq \alpha^2\left(y^{-2}(x - y) \wedge 1 \right).$$

Proof. That h_α is increasing is easily checked by differentiating.

Since $x \geq y$, $h_\alpha(x) \geq h_\alpha(y)$ as h_α is increasing. The lemma follows from the facts that (i) h_α is concave for $x > \alpha$, so that $h_\alpha(x) - h_\alpha(y) \leq (x - y)h'_\alpha(y)$, and noting by direct calculation that $h'_\alpha(y) = -\alpha/y - \log(1 - \alpha/y) \leq \alpha^2/y^2$ for all $y \geq 2\alpha$ and $\alpha \in [0, \frac{1}{2}]$; and (ii) since h is increasing and $y \geq 1$, $h_\alpha(x) - h_\alpha(y) \leq \lim_{z\to\infty} h_\alpha(z) - h_\alpha(1) = \alpha + (1 - \alpha)\log(1 - \alpha) \leq \alpha^2$ for all $\alpha \in [0, 1]$. $\qquad\square$

4.2. Sharp upper bounds on the upper tail of $\nu_{q,N}$. As we outlined earlier, for our ultimate goal of sharp asymptotics for $\mathbb{E}[X^k]$, we will need to obtain sharp upper bounds on expectations such as $\mathbb{E}[X^k]$ and $\mathbb{E}[(X)_k \mathbb{1}_{X \geq t}]$ (as appeared in Lemma 4.2). For this we will need upper bounds on $\mathbb{P}(X \geq t)$ for appropriate t, which will then be used to estimate the expectations using the "layer cake" representation (i.e., a general form of the identity $\mathbb{E}[X^k] = \int_0^\infty kt^{k-1}\mathbb{P}(X \geq t)\,\mathrm{d}t$).

In this section we obtain these sharp upper bounds on $\mathbb{P}(X \geq t)$, where $t = \mu_q N(1 + q^{1/6}\varepsilon)$. The nature of the tail differs between $\varepsilon \in (-1, 0)$ and $\varepsilon > 0$; in the former case, it decays polynomially in ε (indeed, Theorem 2.4 asserts a lower bound of order $\varepsilon^{3/2}$, and we will prove a matching upper bound), while in the latter case it decays exponentially. These are captured in the next two propositions. We recall the definitions of μ_q from (7) and the distribution $\nu_{q,N}$ as defined in (10).

Proposition 4.4. *Let $X \sim \nu_{q,N}$. There exist positive constants C and N_0 such that, for $N \geq N_0$, $\varepsilon \in (N^{-2/3}, 1)$ and $q \in [N^{-2}, 1)$,*

$$\mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}\varepsilon)\right) \leq C\varepsilon^{3/2}.$$

Proposition 4.5. *Let $X \sim \nu_{q,N}$. There exist positive constants c, C and N_0 such that, for $N \geq N_0$, $\varepsilon > 0$, and $q \in [N^{-2}, 1)$,*

$$\mathbb{P}\left(X \geq \mu_q N(1 + q^{1/6}\varepsilon)\right) \leq \begin{cases} C\varepsilon^{-3/4}N^{-3/2}\exp\left(-c\varepsilon^{3/2}N\right) & \varepsilon \in (0, q^{1/3}] \\ C(q^{1/6}N)^{-3/2}(1 + q^{1/6}\varepsilon)^{-cN} & \varepsilon \geq q^{1/3}. \end{cases}$$

When $\varepsilon \in [q^{1/3}, q^{-1/6}]$, we may replace $(1 + q^{1/6}\varepsilon)^{-cN}$ by $\exp(-cq^{1/6}\varepsilon N)$ in the second inequality.

Observe the change in behaviour of the bound at $\varepsilon = q^{1/3}$; in particular, the coefficients in the exponent depend on q when $\varepsilon > q^{1/3}$ and are no longer uniform. The reason for the change in the nature of the bound can be understood by recalling that the upper tail of X is closely related to the upper tail of T_N, the LPP value in geometric LPP: $\mathbb{P}(T_N \geq t - N + 1) \leq N \cdot \mathbb{P}(X \geq t)$ (we will prove this as well as use it in the proof of Theorem 1.9 on the upper tail of T_N at the end of Sect. 4.2 ahead). Setting $t = \mu_q N(1 + q^{1/6}\varepsilon)$, we see that

$$\mathbb{P}\left(T_N \geq (\mu_q - 1)N + \mu_q q^{1/6}\varepsilon N\right) \leq \mathbb{P}(X \geq \mu_q N(1 + q^{1/6}\varepsilon)).$$

Now $\mu_q - 1 = O(q^{1/2}/(1-q))$ and, when $\varepsilon = q^{1/3}$, the deviation $\mu_q q^{1/6}\varepsilon N$ equals $O(q^{1/2}N/(1 - q))$ as well, and so we are in the large deviation regime of geometric LPP. While in the moderate deviation regime we would expect universality (and hence the bounds are uniform in q), in the large deviation regime the bound should be expected

to depend on the last passage percolation vertex distribution; in this case it is $\text{Geo}(q)$, and so it is not surprising that the bound depends on q.

To prove Proposition 4.5, the basic idea will be to use Markov's inequality with the kth factorial moment of X for a k which will be optimized over.

Proof of Proposition 4.4. We may assume $\varepsilon < \frac{1}{2}$ by modifying the constant C. By Markov's inequality, Theorem 3.1, and Lemma 4.1 (to lower bound the denominator), for any $N \geq N_0$, $k_0 \leq k \leq \frac{1}{2}N$, and $q \in (0, 1)$ (and using that $k \leq N/2$, $\varepsilon < \frac{1}{2}$, and $q < 1$ guarantees $\mu_q N(1 - q^{1/6}\varepsilon) \geq k$ to apply Lemma 4.1),

$$
\mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}\varepsilon)\right) \leq \frac{\mathbb{E}[(X)_k]}{(\mu_q N(1 - q^{1/6}\varepsilon))_k}
$$

$$
\leq \frac{C(q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(Nh_\alpha(\mu_q) + Cq^{1/2}\frac{k^3}{N^2}\right)}{(\mu_q N)^k(1 - q^{1/6}\varepsilon)^k \exp\left(Nh_\alpha(\mu_q(1 - q^{1/6}\varepsilon))\right)},
$$

where $\alpha = k/N$ and $h_\alpha(x) = (\alpha - x)\log(1 - \frac{\alpha}{x}) - \alpha$ as in (22). We will ultimately pick $\varepsilon < \frac{1}{2}$. Since $(1 - x)^{-1} \leq \exp(2x)$ for $x \in [0, \frac{1}{2}]$, $\mu_q \geq 1$, and $|h_\alpha(x) - h_\alpha(y)| \leq \alpha^2 y^{-2}(x - y)$ for $x \geq y \geq 2\alpha$ from Lemma 4.3, we obtain

$$
\mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}\varepsilon)\right) \leq C(q^{1/6}k)^{-3/2} \exp\left(2q^{1/6}\varepsilon k + Cq^{1/6}\varepsilon\frac{k^2}{N} + Cq^{1/2}\frac{k^3}{N^2}\right).
$$

Setting $k = q^{-1/6}\varepsilon^{-1}$ (note that $k \leq N/2$ if $\varepsilon \geq 2^{2/3}N^{-2/3}$ since $q \geq N^{-2}$), and using that $k \leq N$, we obtain

$$
\mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}\varepsilon)\right) \leq C\varepsilon^{3/2} \exp\left(C + C\varepsilon^{-3}N^{-2}\right) \leq C\varepsilon^{3/2},
$$

the last inequality using that $\varepsilon \geq N^{-2/3}$. $\qquad\square$

Proof of Proposition 4.5. We address the case of $\varepsilon \in [0, q^{1/3}]$ first. As in the previous proof, by Markov's inequality and Theorem 3.1, for any $N \geq N_0$, $k_0 \leq k \leq \frac{1}{2}N$, and $q \in (0, 1)$,

$$
\mathbb{P}\left(X \geq \mu_q N(1 + q^{1/6}\varepsilon)\right) \leq \frac{\mathbb{E}[(X)_k]}{(\mu_q N(1 + q^{1/6}\varepsilon))_k}
$$

$$
\leq \frac{C(\mu_q N)^k(q^{1/6}k)^{-3/2} \exp\left(Nh_\alpha(\mu_q) + Cq^{1/2}\frac{k^3}{N^2}\right)}{(\mu_q N)^k(1 + q^{1/6}\varepsilon)^k \exp\left(Nh_\alpha(\mu_q(1 + q^{1/6}\varepsilon))\right)}.
$$

(24)

Since $(1 + x)^{-1} \leq \exp(-x/2)$ for $x \in [0, 1]$, and since h_α is increasing (as recorded in Lemma 4.3), we obtain

$$
\mathbb{P}\left(X \geq \mu_q N(1 + q^{1/6}\varepsilon)\right) \leq C(q^{1/6}k)^{-3/2} \exp\left(-cq^{1/6}k\varepsilon + Cq^{1/2}\frac{k^3}{N^2}\right),
$$

the last inequality for $k \leq \frac{1}{2}cC^{-1}N$. Setting $k = \delta\varepsilon^{1/2}q^{-1/6}N$ for a constant $\delta > 0$ to be set shortly, and using that $\varepsilon \leq q^{1/3}$ (so that $k \leq \frac{1}{2}cC^{-1}N$ if $\delta < \frac{1}{2}cC^{-1}$), we obtain

$$
\mathbb{P}\left(X \geq \mu_q N(1 + q^{1/6}\varepsilon)\right) \leq C_\delta\varepsilon^{-3/4}N^{-3/2} \exp\left(-c\delta\varepsilon^{3/2}N + C\delta^3\varepsilon^{3/2}N\right).
$$

Picking δ to be an appropriately small absolute constant completes the proof in the case $\varepsilon \leq q^{1/3}$.

Next we turn to the the case of $\varepsilon \geq q^{1/3}$. We look again at (24). Using that $h_\alpha(\mu_q(1+q^{1/6}\varepsilon)) > h_\alpha(\mu_q)$ as h_α is increasing (from Lemma 4.3) and taking $k = \alpha N$ in (24) for a small absolute constant α to be chosen, we conclude that

$$\mathbb{P}\left(X \geq \mu_q N(1 + q^{1/6}\varepsilon)\right) \leq C(q^{1/6}N)^{-3/2}(1 + q^{1/6}\varepsilon)^{-\alpha N} \exp\left(Cq^{1/2}\alpha^3 N\right)$$
$$\leq C(q^{1/6}N)^{-3/2}(1 + q^{1/6}\varepsilon)^{-\alpha N/2}$$

for all small enough α (using that $q^{1/6}\varepsilon \geq q^{1/2}$). In the case that $q^{1/6}\varepsilon \leq 1$, it holds that $(1 + q^{1/6}\varepsilon)^{-1} \leq \exp(-cq^{1/6}\varepsilon)$, which provides the claimed bound in the remaining case. This completes the proof of Proposition 4.5.

With these upper tail estimates of $X \sim \nu_{q,N}$ available, we can also quickly give the proof of Theorem 1.9 on the uniform upper tail of geometric LPP:

Proof of Theorem 1.9. As in the proof of Theorem 1.4, we have that $T_N \overset{d}{=} \lambda_1 - N + 1$, where $(\lambda_1, \ldots, \lambda_N)$ is the Meixner ensemble. So for any $t \in \mathbb{R}$,

$$\mathbb{P}(T_N \geq t - N + 1) = \mathbb{P}(\lambda_1 \geq t) \leq \mathbb{P}\left(N^{-1}\sum_{i=1}^{N}\delta_{\lambda_i/N}([t, \infty)) \geq \frac{1}{N}\right)$$
$$\leq N\nu_{q,N}([t, \infty)), \tag{25}$$

where that last inequality is Markov's inequality and recall $\nu_{q,N} = \mathbb{E}[N^{-1}\sum_{i=1}^{N}\delta_{\lambda_i/N}]$ is the expected empirical distribution of the Meixner ensemble defined in (10).

We set $t = \mu_q N(1 + q^{1/6}\varepsilon)$ for an $\varepsilon \in (0, 1)$ to be set later. Now if we let X be distributed as $\nu_{q,N}$, we see that we want to estimate

$$N\nu_{q,N}([\mu_q N(1 + q^{1/6}\varepsilon), \infty)) = N \cdot \mathbb{P}\left(X \geq \mu_q N(1 + q^{1/6}\varepsilon)\right).$$

By the first case of Proposition 4.5, the previous display is upper bounded by

$$C\varepsilon^{-3/4}N^{-1/2}\exp(-c\varepsilon^{3/2}N)$$

when $\varepsilon < q^{1/3}$. Taking $\varepsilon = xN^{-2/3}$ and noting that $x < q^{1/3}N^{2/3}$ implies $\varepsilon < q^{1/3}$ completes the proof in this range of x. For $x > q^{1/3}N^{2/3}$, doing the above using the bound from Proposition 4.5 for $\varepsilon > q^{1/3}$ completes the proof. $\qquad\blacksquare$

4.3. The upper bound on $\mathbb{E}[X^k]$. In this section we apply Lemmas 4.1 to obtain the upper bound on $\mathbb{E}[X^k]$, and we will turn to the lower bounds in Sect. 4.4. The precise upper bound is the following. (Recall from (7) that $\mu_q = (1 + q^{1/2})^2/(1 - q)$.)

Proposition 4.6. *Let X be distributed according to $\nu_{q,N}$ as defined in (10) and μ_q be as in (7). There exist positive constants C, k_0, α_0, and N_0 such that, for any $N \geq N_0$ and (k, q) satisfying $k_0 \leq k \leq \min(\alpha_0 N, q^{-1/6}N^{2/3})$ and $q \in [N^{-2}, 1)$, it holds that*

$$\mathbb{E}[X^k] \leq C(q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(Cq^{1/2}\frac{k^3}{N^2}\right).$$

We note that, for the range of k under consideration, we can absorb the factor $\exp(Cq^{1/2}k^3/N^2)$ into the constant factor C (and thus match the statement of Theorem 2.5). We keep this factor in this statement merely to match the form of Proposition 4.7 ahead where it cannot be similarly absorbed, and because it will appear naturally in the proof.

Proof of Proposition 4.6. We start with the breakup

$$\mathbb{E}[X^k] = \mathbb{E}[X^k \mathbb{1}_{X \leq \mu_q N(1-q^{1/6})}] + \mathbb{E}[X^k \mathbb{1}_{X > \mu_q N(1-q^{1/6})}]$$
$$\leq (\mu_q N)^k \exp(-q^{1/6}k) + \mathbb{E}[X^k \mathbb{1}_{X > \mu_q N(1-q^{1/6})}].$$

Next we estimate $\mathbb{E}[X^k \mathbb{1}_{X > \mu_q N(1-q^{1/6})}]$. Using the fundamental theorem of calculus and Fubini's theorem to write $\mathbb{E}[X^k \mathbb{1}_{X > s}] = \int_0^\infty kt^{k-1} \mathbb{P}(X > \max(s, t)) \, dt$,

$$\mathbb{E}[X^k \mathbb{1}_{X > \mu_q N(1-q^{1/6})}] = \int_0^\infty kt^{k-1} \mathbb{P}\left(X > \max(\mu_q N(1 - q^{1/6}), t)\right) dt$$
$$= (\mu_q N(1 - q^{1/6}))^k \mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6})\right)$$
$$+ \int_{\mu_q N(1-q^{1/6})}^\infty kt^{k-1} \mathbb{P}(X > t) \, dt$$
$$\leq (\mu_q N)^k \exp(-q^{1/6}k) + (q^{1/6}k)(\mu_q N)^k$$
$$\int_{-1}^\infty (1 + q^{1/6}s)^{k-1} \mathbb{P}\left(X \geq \mu_q N(1 + q^{1/6}s)\right) ds,$$

performing a change of variable $t \mapsto \mu_q N(1 + q^{1/6}s)$ in the last line and using $1 - x \leq \exp(-x)$.

We focus on the second term in the previous display. The tail probability in the integrand behaves differently for $s \in (-1, 0)$ and $s > 0$, as captured in Propositions 4.4 and 4.5, and so we break up the integral into two parts on this basis. Doing so, and doing a change of variable $s \mapsto -s$ in the resulting first integral, we obtain that the second term in the previous display equals

$$(q^{1/6}k)(\mu_q N)^k \int_0^1 (1 - q^{1/6}s)^{k-1} \mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}s)\right) ds$$
$$+ (q^{1/6}k)(\mu_q N)^k \int_0^\infty (1 + q^{1/6}s)^{k-1} \mathbb{P}\left(X \geq \mu_q N(1 + q^{1/6}s)\right) ds. \quad (26)$$

We start with the first term. Using that $1 - x \leq \exp(-x)$ and Proposition 4.4 (note that this proposition requires $\varepsilon > N^{-2/3}$, which necessitates the further break up below),

$$(q^{1/6}k)(\mu_q N)^k \int_0^1 (1 - q^{1/6}s)^{k-1} \mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}s)\right) ds$$
$$= (q^{1/6}k)(\mu_q N)^k \int_0^{N^{-2/3}} (1 - q^{1/6}s)^{k-1} \mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}s)\right) ds$$
$$+ (q^{1/6}k)(\mu_q N)^k \int_{N^{-2/3}}^1 (1 - q^{1/6}s)^{k-1} \mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}s)\right) ds$$

$$\leq (q^{1/6}k)(\mu_q N)^k N^{-2/3} \mathbb{P}\left(X \geq \mu_q N(1 - q^{1/6}N^{-2/3})\right)$$

$$+ (q^{1/6}k)(\mu_q N)^k \int_{N^{-2/3}}^1 s^{3/2} \exp(-q^{1/6}sk) \, ds$$

$$\leq C(q^{1/6}k)(\mu_q N)^k \left[N^{-5/3} + (q^{1/6}k)^{-5/2}\right].$$

Now we turn to the second term of (26). By applying the bound on the tail probability from Proposition 4.5 (in the two cases of $s < q^{1/3}$ and $s > q^{1/3}$) and using that $1 + x \leq \exp(x)$, we obtain that the second term of (26) is upper bounded by

$$C(q^{1/6}k)(\mu_q N)^k \left[N^{-3/2} \int_0^{q^{1/3}} s^{-3/4} \exp\left(q^{1/6}ks - cs^{3/2}N\right) \, ds \right.$$

$$\left. + q^{-1/4}N^{-3/2} \int_{q^{1/3}}^\infty (1 + q^{1/6}s)^{k-1-cN} \, ds\right]. \tag{27}$$

We write k in the exponent as αN and perform Laplace's method to bound the first integral, with $f(s) = q^{1/6}\alpha s - cs^{3/2}$, $s_0 = \text{argmax}_s f(s) = \frac{4}{9}c^{-2}q^{1/3}\alpha^2$ and $f''(s_0) = -\frac{3}{4}cs_0^{-1/2}$, to obtain that the first term (i.e., the first integral times its complete coefficient) of the previous display is upper bounded by (also using that $\alpha = k/N$)

$$C(\mu_q N)^k q^{1/6}k \cdot N^{-3/2}(N|f''(s_0)|)^{-1/2} \cdot s_0^{-3/4} \cdot \exp(Nf(s_0))$$

$$= C(\mu_q N)^k q^{1/6}k \cdot N^{-3/2}(Nq^{-1/6}\alpha^{-1})^{-1/2} \cdot q^{-1/4}\alpha^{-3/2} \cdot \exp\left(Nq^{1/2}\frac{k^3}{N^2}\right)$$

$$= CN^{-1}(\mu_q N)^k \exp\left(Cq^{1/2}\frac{k^3}{N^2}\right). \tag{28}$$

Under the condition that $k \leq cN/2$, the second term in (27) is upper bounded (using that $k \leq N$ and $(1 + q^{1/2})^{-1} \leq 1$ for the displayed inequality) by

$$C(q^{1/6}k)(\mu_q N)^k \cdot q^{-1/4}N^{-3/2} \cdot \frac{q^{-1/6}(1 + q^{1/2})^{-cN/2}}{cN} \leq q^{-1/4}k^{-3/2}(\mu_q N)^k.$$

Putting it all together, we have shown that

$$\mathbb{E}[X^k] \leq C(\mu_q N)^k \left[2\exp(-q^{1/6}k) + (q^{1/6}k)N^{-5/3} + 2(q^{1/6}k)^{-3/2}\right.$$

$$\left. + N^{-1}\exp\left(Cq^{1/2}\frac{k^3}{N^2}\right)\right].$$

Clearly $\exp(-q^{1/6}k) \leq C(q^{1/6}k)^{-3/2}$. The condition that $k \leq q^{-1/6}N^{2/3}$ implies that $(q^{1/6}k)N^{-5/3} \leq (q^{1/6}k)^{-3/2}$ and $N^{-1} \leq (q^{1/6}k)^{-3/2}$. Thus we obtain that the RHS in the previous display is upper bounded by

$$C(q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(Cq^{1/2}\frac{k^3}{N^2}\right),$$

completing the proof.   □

4.4. The lower bound on $\mathbb{E}[X^k]$. Here is the lower bound statement we prove:

Proposition 4.7. *Let X be distributed according to $\nu_{q,N}$ as defined in* (10) *and μ_q be as in* (7)*. Then there exist positive constants c, C, α_0, N_0, and k_0 such that, for any $N \geq N_0$, $k_0 \leq k \leq \alpha_0 N$, and $q \in [k^{-2}, 1)$,*

$$\mathbb{E}[X^k] \geq c(q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(-Cq^{1/2}\frac{k^3}{N^2}\right).$$

The basic idea is to combine Lemma 4.2 with Theorem 3.1. Recall that the lower bound from Lemma 4.2 has the term $\mathbb{E}[(X)_k] - \mathbb{E}[(X)_k \mathbb{1}_{X \geq \beta N}]$. While the first term can be lower bounded using Theorem 3.1, we do not currently have an estimate for the second term. To handle this we will again make use of the fundamental theorem of calculus and Fubini's theorem and make use of Proposition 4.5 (upper bound on the upper tail of X), as in the proof of Proposition 4.7.

Here too the fact that we need estimates on polynomial moments $\mathbb{E}[X^k]$ with k up to order N creates technical difficulties. For smaller values of k, e.g. up to order $N^{2/3}$, one can upper bound $\mathbb{E}[(X)_k \mathbb{1}_{X \geq \beta N}]$ by $\mathbb{E}[X^k \mathbb{1}_{X \geq \beta N}]$. The latter is less than $\mathbb{E}[X^{2k}]^{1/2}\mathbb{P}(X \geq \beta N)^{1/2}$ by Cauchy–Schwarz, and this can be bounded using the already proved upper bound on the polynomial moments and a crude upper bound on the tail probability. This strategy works for smaller k because, when $X = \Theta(N)$ as is typical, X^k and $(X)_k$ differ only by a constant factor; the same certainly does not hold when $k = \Theta(N)$. It is for this reason that we must follow the more delicate path outlined above, making use of a variant of the layer cake formula for $\mathbb{E}[(X)_k \mathbb{1}_{X \geq \beta N}]$ and essentially sharp upper tail bounds for X throughout the tail.

Proof of Proposition 4.7. We will take $\beta = \mu_q(1+q^{1/2})$ and ultimately apply Lemma 4.2. So we need to lower bound $\mathbb{E}[(X)_k] - \mathbb{E}[(X)_k \mathbb{1}_{X > \mu_q N(1+q^{1/2})}]$, in particular, to show that the second term is much smaller than the first. We start with lower bounding the first term using Theorem 3.1, which says that, for $N \geq N_0$, $k_0 \leq k \leq \frac{1}{2}N$, and $q \in [k^{-2}, 1)$,

$$\mathbb{E}[(X)_k] \geq C'(q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(N\left[h_\alpha(\mu_q) - C'\alpha^3 q^{1/2}\right]\right), \qquad (29)$$

where recall from (22) that $h_\alpha(x) = (\alpha - x)\log(1 - \frac{\alpha}{x}) - \alpha$ and $\alpha = k/N$.

We need to show that $\mathbb{E}[(X)_k \mathbb{1}_{X \geq \mu_q N(1+q^{1/2})}]$ is smaller than the previous display by at least a constant factor. The idea will be, as in the arguments in Sect. 4.3, to use the fundamental theorem of calculus and Fubini's theorem to write this expectation as an integral against the tail probability.

$$\begin{aligned}
\mathbb{E}[(X)_k \mathbb{1}_{X \geq \mu_q N(1+q^{1/2})}] &= \int_0^\infty \frac{d}{dx}(x)_k \cdot \mathbb{P}\left(X > \max(x, \mu_q N(1+q^{1/2}))\right) dx \\
&= \int_0^{\mu_q N(1+q^{1/2})} \frac{d}{dx}(x)_k \cdot \mathbb{P}\left(X > \mu_q N(1+q^{1/2})\right) dx \\
&\quad + \int_{\mu_q N(1+q^{1/2})}^\infty \frac{d}{dx}(x)_k \cdot \mathbb{P}(X > x) dx \\
&= \left(\mu_q N(1+q^{1/2})\right)_k \cdot \mathbb{P}\left(X > \mu_q N(1+q^{1/2})\right) \\
&\quad + \int_{\mu_q N(1+q^{1/2})}^\infty \frac{d}{dx}(x)_k \cdot \mathbb{P}(X > x) dx.
\end{aligned}$$

Now, we can write the derivative of $(x)_k$ as $(\psi(x+1) - \psi(x-k+1)) \cdot (x)_k$, where ψ is the digamma function (as can be verified by using the product rule to differentiate $(x)_k$ and the recursive relation $\psi(x+1) = \psi(x) + x^{-1}$ that ψ satisfies). The digamma function satisfies the inequalities, for $x > \frac{1}{2}$,

$$\psi(x) \in \left(\log(x - \tfrac{1}{2}), \log x\right),$$

so that, when $x > k - \frac{1}{2}$,

$$\psi(x+1) - \psi(x-k+1) \le \log(x+1) - \log(x-k+\tfrac{1}{2}) = -\log\left(1 - \frac{k+\frac{1}{2}}{x+1}\right)$$

$$\le \frac{k+\frac{1}{2}}{x+1}.$$

if $k < x/2$ (which holds in our situation since $k < \frac{1}{2}N$ and $x > \mu_q N$ with $\mu_q > 1$).

We recall that, by the second case of Proposition 4.5, $\mathbb{P}(X > \mu_q N(1+q^{1/2})) \le q^{-1/4}N^{-3/2}(1+q^{1/2})^{-cN}$.

Using these facts and performing the change of variable $x \mapsto \mu_q N(1+q^{1/2}y)$, we see that

$$\mathbb{E}[(X)_k \mathbb{1}_{X \ge \mu_q N(1+q^{1/2})}] \le C\left(\mu_q N(1+q^{1/2})\right)_k \cdot q^{-1/4}N^{-3/2}(1+q^{1/2})^{-cN}$$

$$+ C\mu_q N q^{1/2} \int_1^\infty \frac{k}{\mu_q N}(\mu_q N(1+q^{1/2}y))_k$$

$$\cdot \mathbb{P}\left(X > \mu_q N(1+q^{1/2}y)\right) dy. \tag{30}$$

We next estimate the falling factorial terms in the above expression. We know from (23) that, for any $\varepsilon > 0$ and $k = \alpha N$, (recalling $h_\alpha(x) = (\alpha - x)\log(1 - \frac{\alpha}{x}) - \alpha$ from (22))

$$\left(\mu_q N(1+\varepsilon)\right)_k \le C(\mu_q N(1+\varepsilon))^k \exp\left(Nh_\alpha(\mu_q(1+\varepsilon))\right)$$

$$\le C(\mu_q N)^k(1+\varepsilon)^k \exp\left(N\left[h_\alpha(\mu_q) + C\alpha^2(\mu_q^{-1}\varepsilon \wedge 1)\right]\right)$$

$$\le C(\mu_q N)^k(1+\varepsilon)^k \exp\left(N\left[h_\alpha(\mu_q) + C\alpha^2(\varepsilon \wedge 1)\right]\right),$$

the second inequality using Lemma 4.3 and third using $\mu_q \ge 1$.

Substituting this bound on the falling factorial into (30) and using the bounds on the tail probability from the second case of Proposition 4.5, we obtain that $\mathbb{E}[(X)_k \mathbb{1}_{X \ge \mu_q N(1+C_1 q^{1/2})}]$ is upper bounded by

$$\widetilde{C}(\mu_q N)^k \exp\left(N\left[(\alpha - \mu_q)\log\left(1 - \frac{\alpha}{\mu_q}\right) - \alpha\right]\right) \cdot q^{-1/4}N^{-3/2}$$

$$\times \left[\exp\left(-N\left[c\log(1+q^{1/2}) - C\alpha^2 q^{1/2}\right]\right)\right.$$

$$\left. + kq^{1/2}\int_1^\infty \exp\left(-N\left[c\log(1+q^{1/2}y) - C\alpha^2\left(1 \wedge (q^{1/2}y)\right)\right]\right) dy\right]. \tag{31}$$

To ensure that the coefficient of N in the exponential is negative in both the terms on the second and third lines of the previous display, we will need to restrict how big α can be. More precisely, since $\alpha < \alpha_0$ with α_0 an absolute constant which we are free to set, we set it such that, for all $x \in [0, 1]$

$$\frac{1}{2}c \log(1 + x) > C\alpha_0^2 x.$$

Then the second line of (31) is upper bounded by

$$(1 + q^{1/2})^{-\frac{1}{2}cN} \le \tfrac{1}{2},$$

using that $q \ge N^{-2}$.

Similarly the third line of (31) is upper bounded by

$$kq^{1/2} \int_1^\infty (1 + q^{1/2}y)^{-\frac{1}{2}cN} \, \mathrm{d}y = Ck\frac{(1 + q^{1/2})^{-cN}}{N} \le \tfrac{1}{2}.$$

Thus overall we see that there is a C such that

$$\mathbb{E}[(X)_k \mathbb{1}_{X \ge \mu_q N(1 + C_1 q^{1/2})}] \le Cq^{-1/4}N^{-3/2}(\mu_q N)^k$$
$$\exp\left(N\left[(\alpha - \mu_q)\log\left(1 - \frac{\alpha}{\mu_q}\right) - \alpha\right]\right).$$

Then, recalling the lower bound on $\mathbb{E}[(X)_k]$ from (29),

$$\frac{\mathbb{E}[(X)_k \mathbb{1}_{X \ge \mu_q N(1 + C_1 q^{1/2})}]}{\mathbb{E}[(X)_k]} \le C(C')^{-1}\left[\frac{(q^{1/6}k)^{3/2}}{(q^{1/6}N)^{3/2}}\right] \le C(C')^{-1}\alpha_0^{3/2} \le \frac{1}{2},$$

the last inequality by reducing α_0 further if necessary and using that $k \le \alpha_0 N$ and $q \ge k^{-2}$.

Now applying Lemma 4.2 with $\beta = \mu_q(1 + q^{1/2}) \ge 1$ and $\alpha \le \frac{1}{2}$ (and noting that thus $(\beta - \alpha)^{-1} \le 2$), using the previous display and (29), we get

$$\mathbb{E}[X^k] \ge \frac{1}{2}\mathbb{E}[(X)_k] \cdot \exp\left(N\left[(\beta - \alpha)\log\left(1 - \frac{\alpha}{\beta}\right) + \alpha\right] - \tfrac{1}{2}(\beta - \alpha)^{-1}\right)$$
$$\ge c(q^{1/6}k^{-3/2})(\mu_q N)^k \exp\left(-C\alpha^3 q^{1/2}N\right).$$

$\square$

5. Concentration Inequalities and the Proof of Theorem 1.1

In this section we combine Theorem 1.4 (on the uniform tail for geometric LPP) with the representation of the position $x_N(N)$ of the first particle in q-pushTASEP in terms of the LPP value in an infinite periodic strip of inhomogeneous geometric random variables, and so obtain an upper bound on the lower tail of $x_N(N)$. Recall from the proof outline given in Sect. 1.6 that the main idea is to lower bound the LPP value by a sum of independent LPP values, each one in an $N \times N$ square; the parameter of the geometric random variables is the same within each single such square, but varies across different ones.

For this argument we need one final ingredient: a concentration inequality for a sum of independent random variables that takes into account the possibly varying scales of the summands. Indeed, we will be considering a sum of geometric LPP values where the parameter of the geometric is q^i for varying i; the scale of fluctuation for fixed i is $q^{i/6}/(1 - q^i) \approx i^{-1}(|\log q|)^{-1} q^{i/6}$. Such a concentration inequality is recorded next, and, as its proof is fairly routine, will be proven in Appendix C.

Theorem 5.1. *Let $I \in \mathbb{N} \cup \{\infty\}$ and suppose $X_1, \dots, X_I$ are independent, and assume that there exists $C_1 < \infty$ and $\rho_1, \dots, \rho_I > 0$ such that each X_i satisfies*

$$\mathbb{P}(X_i \geq t) \leq C_1 \exp(-\rho_i t^{3/2})$$

for all $t > 0$. Let $\sigma_2 = \sum_{i=1}^I \rho_i^{-2}$ and $\sigma_{2/3} = \sum_{i=1}^I \rho_i^{-2/3}$. Then there exist positive absolute constants C and c such that, for $t > 0$,

$$\mathbb{P}\left(\sum_{i=1}^I X_i \geq t + C \cdot C_1 \sigma_{2/3}\right) \leq \exp\left(-c\sigma_2^{-1/2} t^{3/2}\right).$$

Using Theorem 5.1 we may give the proof of the main result, Theorem 1.1. We will need a simple lower bound on the coefficient rescaling the fluctuations in the definition (4) of X_N^{sc}.

Lemma 5.2. *For $q, u \in (0, 1)$,*

$$(-\psi_q''(\log_q u))^{1/3}(\log q^{-1})^{-1} \geq \frac{u^{1/3}}{(1-q)^{1/3}(1-u)^{2/3}}.$$

Proof. By [MS09, Eq. (1.6)], $\psi_q''(x) = (\log q)^3 \cdot \sum_{n=1}^\infty \frac{n^2 q^{nx}}{1-q^n}$. Taking $x = \log_q u = \log u / \log q$ yields that

$$(1-q)(-\psi_q''(\log_q u))(\log q)^{-3} = \sum_{n=1}^\infty \frac{n^2 u^n (1-q)}{1 - q^n} = \sum_{n=1}^\infty \frac{n^2 u^n}{1 + q + \cdots + q^{n-1}}.$$

Since the denominator is upper bounded by n and $\sum_{n=1}^\infty n u^n = u/(1-u)^2$, we obtain the lemma. $\square$

This bound is not sharp. Indeed, since $(-\psi_q''(\log_q u))(\log q)^{-3} = \sum_{n=1}^\infty n^2 u^n/(1-q)^n$, in the $q \to 0$ limit it equals $\sum_{n=1}^\infty n^2 u^n = u/(1-u)^3$ which is of larger order than $u/(1-u)^2$ in the $u \to 1$ regime.

Proof of Theorem 1.1. We have to upper bound, for $\theta > \theta_0 = \theta_0(q)$,

$$\mathbb{P}\left(x_N(N) \leq f_q N - (-\psi_q''(\log_q u))^{1/3}(\log q^{-1})^{-1}\theta N^{1/3}\right), \tag{32}$$

where we recall that f_q is defined in (2) as

$$f_q = 2 \cdot \frac{\psi_q(\log_q u) + \log(1-q)}{\log q} + 1.$$

By Lemma 5.2, with $\sigma_{u,q} = u^{1/3}(1-q)^{-1/3}(1-u)^{-2/3}$, (32) is upper bounded by

$$\mathbb{P}\left(x_N(N) \leq f_q N - \sigma_{u,q}\theta N^{1/3}\right). \tag{33}$$

We define $\tilde{f}_q = f_q - 1$. By Theorem 1.3, we know that $L + N \overset{d}{=} x_N(N)$, where L is the LPP value from the topmost site to ∞ in the infinite periodic environment defined in Sect. 1.4. Let $L_N^{(i)}$ be the last passage time from the top to the bottom of the ith large square on the vertical line from the top (which has i.i.d. $\mathrm{Geo}(u^2q^{2i})$ random variables associated to each small square). Then clearly $\sum_{i=0}^{\infty} L_N^{(i)} \leq L$, so

$$(33) = \mathbb{P}\left(L \leq \tilde{f}_q N - \sigma_{u,q}\theta N^{1/3}\right) \leq \mathbb{P}\left(\sum_{i=0}^{\infty} L_N^{(i)} \leq \tilde{f}_q N - \sigma_{u,q}\theta N^{1/3}\right).$$

We next add and subtract the law of large numbers term of $L_N^{(i)}$, which as we see from Theorem 1.4 is $2Nuq^i(1-uq^i)^{-1}$, as this is the term by which the random variables are centered to yield the tail bounds in the same theorem. So we can write the right-hand side of the previous display as

$$\mathbb{P}\left(\sum_{i=0}^{\infty}\left(L_N^{(i)} - 2N\cdot\frac{uq^i}{1-uq^i}\right) \leq \tilde{f}_q N - \sum_{i=0}^{\infty} 2N\cdot\frac{uq^i}{1-uq^i} - \sigma_{u,q}\theta N^{1/3}\right). \tag{34}$$

We have already evaluated the LLN sum in the proof ideas section. So we recall from (5) and (6) that

$$\sum_{i=0}^{\infty} 2N\cdot\frac{uq^i}{1-uq^i} = 2N\cdot\frac{\psi_q(\log_q(u))+\log(1-q)}{\log q} = \tilde{f}_q N.$$

Putting this back into (34), we see that

$$(34) = \mathbb{P}\left(\sum_{i=0}^{\infty}\left(L_N^{(i)} - 2N\cdot\frac{uq^i}{1-uq^i}\right) \leq -\sigma_{u,q}\theta N^{1/3}\right). \tag{35}$$

In the remainder of the proof we will invoke the concentration bound from Theorem 5.1 to upper bound the previous display.

Now, $L_N^{(i)}$ is the LPP value in an $N \times N$ square with i.i.d. geometric random variables of parameter u^2q^{2i}. We know from Theorem 1.4 that there exist positive constants c, t_0, and N_0 such that, when $u^2q^{2i} \in (0,1)$, $t > t_0$, and $N > N_0$,

$$\mathbb{P}\left(L_N^{(i)} - 2N\cdot\frac{uq^i}{1-uq^i} \leq -t\cdot\frac{u^{1/3}q^{i/3}}{1-u^2q^{2i}}N^{1/3}\right) \leq \exp(-ct^{3/2}).$$

Equivalently, there exist positive constants c, C_1, and N_0 such that, when $u^2q^{2i} \in (0,1)$, $t > 0$ (i.e., not t_0), and $N > N_0$,

$$\mathbb{P}\left(L_N^{(i)} - 2N\cdot\frac{uq^i}{1-uq^i} \leq -t\cdot\frac{u^{1/3}q^{i/3}}{1-u^2q^{2i}}N^{1/3}\right) \leq C_1\exp(-ct^{3/2}).$$

Letting $\sigma'_{i,q,u} = u^{1/3}q^{i/3}(1 - u^2q^{2i})^{-1}$, we see that, for $t > 0$,

$$\mathbb{P}\left(L_N^{(i)} - 2N \cdot \frac{uq^i}{1 - uq^i} \leq -tN^{1/3}\right) \leq C_1 \exp\left(-c(\sigma'_{i,q,u})^{-3/2}t^{3/2}\right).$$

We next want invoke Theorem 5.1 with $I = \infty$ and (from the previous display) $\rho_i = (\sigma'_{i,q,u})^{-3/2}$. Now, since $q \in (0, 1)$,

$$\sigma_2 = \sum_{i=0}^{\infty} \rho_i^{-2} = \sum_{i=0}^{\infty} (\sigma'_{i,q,u})^3 = \sum_{i=0}^{\infty} \frac{uq^i}{(1 - u^2q^{2i})^3} \leq \frac{u}{(1 - q)(1 - u^2)^3}$$

$$\text{and} \quad \sigma_{2/3} = \sum_{i=0}^{\infty} \rho_i^{-2/3} = \sum_{i=0}^{\infty} \sigma'_{i,q,u} = \sum_{i=0}^{\infty} \frac{u^{1/3}q^{i/3}}{1 - u^2q^{2i}} \leq \frac{u^{1/3}}{(1 - q^{1/3})(1 - u^2)}.$$

With these estimates, we obtain from Theorem 5.1 that there exist positive C and c such that, for all $t > 0$ (and using that $1 - u^2 = \Theta(1 - u)$),

$$\mathbb{P}\left(\sum_{i=0}^{\infty}\left(L_N^{(i)} - 2N \cdot \frac{uq^i}{1 - uq^i}\right) \leq -tN^{1/3} - C\frac{u^{1/3}}{(1 - q^{1/3})(1 - u)}\right)$$
$$\leq \exp\left(-cu^{-1/2}(1 - q)^{1/2}(1 - u)^{3/2}t^{3/2}\right).$$

So, putting in $t = 2\sigma_{u,q}\theta = u^{1/3}(1 - q)^{-1/3}(1 - u)^{-2/3}\theta$, we obtain that, for $\theta > C(1 - q)^{1/3}/((1 - q^{1/3})(1 - u)^{1/3})$ with C as in the last display, (35) is bounded as

$$\mathbb{P}\left(\sum_{i=0}^{\infty}\left(L_N^{(i)} - 2N \cdot \frac{uq^i}{1 - uq^i}\right) \leq -\sigma_{u,q}\theta N^{1/3}\right) \leq \exp\left(-c(1 - u)^{1/2}\theta^{3/2}\right).$$

Note that $(1 - q)^{1/3}/(1 - q^{1/3}) \leq C' \vee (\log q^{-1})^{-2/3}$ by writing $q = \exp(-\varepsilon)$ for some C'. Thus we obtain the desired bound (33) with $\theta_0(q, u) = C(1 - u)^{-1/3}(1 \vee (\log q^{-1})^{-2/3})$. $\qquad\square$

Appendix A: Proof of the LPP-q-Whittaker Connection

In this appendix we give the proof of Theorem 1.3 relating the observable $x_N(T)$ to an infinite last passage problem in a periodic and inhomogeneous environment. As mentioned, the proof goes through an equivalence to the q-Whittaker measure, and we start by introducing it.

A.1 q-Whittaker polynomials and measure.

Definition A.1 (q-Whittaker polynomial). For a skew partition μ/λ, the skew q-Whittaker polynomial in n variables $\mathcal{P}_{\mu/\lambda}(x_1, \ldots, x_n; q)$ is defined recursively by the branching rule

$$\mathcal{P}_{\mu/\lambda}(x_1, \ldots, x_n; q) = \sum_{\eta} \mathcal{P}_{\eta/\lambda}(x_1, \ldots, x_{n-1}; q)\mathcal{P}_{\mu/\eta}(x_n; q),$$

where, for a single variable $z \in \mathbb{C}$ (recalling the q-binomial coefficient defined in (1)),

$$\mathcal{P}_{\mu/\eta}(z; q) = \mathbb{1}_{\eta \prec \mu} \prod_{i \geq 1} z^{\mu_i - \eta_i} \binom{\mu_i - \mu_{i+1}}{\mu_i - \eta_i}_q.$$

For a partition μ, the q-Whittaker polynomial $\mathcal{P}_\mu$ is given by the skew q-Whittaker polynomial $\mathcal{P}_{\mu/\lambda}$ with λ taken to be the empty partition. The q-Whittaker polynomial is a special case ($t = 0$) of the Macdonald polynomials, for which a comprehensive reference is [Mac98, Section VI].

For a partition μ, we also define $b_\mu(q)$ by

$$b_\mu(q) = \prod_{i \geq 1} \frac{1}{(q; q)_{\mu_i - \mu_{i+1}}}.$$

Definition A.2 (q-Whittaker measure). The q-Whittaker measure $\mathbb{W}_{a;b}^{(q)}$, first introduced in [BC14], is the measure on the set of all partitions given by

$$\mathbb{W}_{a;b}^{(q)}(\mu) = \frac{1}{\Pi(a; b)} b_\mu(q)\mathcal{P}_\mu(a; q)\mathcal{P}_\mu(b; q),$$

where $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_t)$ satisfy $a_i, b_j \in (0, 1)$, and $\Pi(a; b)$ is a normalization constant given explicitly by

$$\Pi(a; b) = \prod_{i=1}^{n} \prod_{j=1}^{t} \frac{1}{(a_i b_j; q)_\infty}.$$

We may now record the important connection between $x_N(T)$ and the q-Whittaker measure which holds under general parameter choices and general times, when started from the narrow-wedge initial condition:

Theorem A.3 (Section 3.1 of [MP16]). *Let a, b be specializations of parameters respectively $(a_1, \ldots, a_N) \in (0, 1)^N$ and $(b_1, \ldots, b_T) \in (0, 1)^T$. Let $\mu \sim \mathbb{W}_{a;b}^{(q)}$ and let $x(T)$ be a q-pushTASEP under initial conditions $x_k(0) = k$ for $k = 1, \ldots, N$. Then,*

$$x_N(T) \stackrel{d}{=} \mu_1 + N.$$

q-Whittaker to LPP. Next we give the proof of Theorem 1.3, which was explained to us by Matteo Mucciconi. As indicated in Sect. 1.5, the proof we give relies heavily on the work [IMS21]. Before proceeding we introduce some terms that will be needed. First, a tableaux is a filling of a Young diagram with non-negative integers. It is called *semi-standard* if the entries in the rows and columns are non-decreasing, from left to right and top to bottom respectively. A *vertically strict tableaux* is a tableaux of non-negative integers in which the columns are strictly increasing from top to bottom, but there is no constraint on the row entries. A *skew tableaux* is a pair of partitions (λ, μ) such that the Young diagram of λ contains that of μ, and should be thought of as the boxes corresponding to $\lambda \setminus \mu$. A semi-standard skew tableaux is defined analogously to the semi-standard tableaux.

As we saw, Theorem A.3 on the relation between $x_N(T)$ and the q-Whittaker measure reduces the proof of Theorem 1.3 to proving the equality in distribution of the LPP value L and the length of the top row of a Young diagram λ sampled from the q-Whittaker measure. In fact, we will prove a stronger statement which relates all the row lengths of λ to appropriate last passage percolation observables. For this, we let $L^{(j)}$ be the maximum weight over all collections of j disjoint paths, one path starting from $(i, 1)$ for each $1 \leq i \leq j$ and all going to ∞ downwards, where the weight of a collection of paths is the sum of the weights of the individual paths.

Theorem A.4. *For $1 \leq j \leq \min(N, T)$, let $L^{(j)}$ be as defined above in the environment defined in Sect. 1.4 and let $\mu \sim \mathbb{W}_{a;b}^{(q)}$ with $a_i, b_j \in (0, 1)$ for all $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$. Then, jointly across $1 \leq j \leq \min(N, T)$,*

$$L^{(j)} \stackrel{d}{=} \mu_1 + \cdots + \mu_j.$$

Theorem 1.3 follows immediately from combining the $j = 1$ case of Theorem A.4 with Theorem A.3.

Proof of Theorem A.4. We prove this in the case $T = N$. It is easy to see that the same proof applies to the $T < N$ case by setting $m_{(i,j);k} = 0$ (in the same notation as below) for $j = T + 1, \ldots, N$, and similarly for $N < T$.

We will make use of two bijections. The first, known as the Sagan–Stanley correspondence and denoted by SS, is a bijection between the set of (M, ν) and the collection of pairs (P, Q) of semi-standard tableaux of general skew shape, where $M = (m_{(i,j);k})_{1 \leq i,j \leq N, k=0,1,\ldots}$ is a filling of the infinite strip by non-negative integers which are eventually all zero and ν is a partition. The second is a bijection Υ introduced in [IMS21] between the collection of such (P, Q) and tuples of the form $(V, W; \kappa, \nu)$, where V, W are vertically strict tableaux of shape μ (which is a function of P, Q) and $\kappa \in \mathcal{K}(\mu)$ is an ordered tuple of non-negative integers with certain constraints on the entries depending on μ, encoded by the set $\mathcal{K}(\mu)$. We will not need the definition of $\mathcal{K}(\mu)$ for our arguments, but the interested reader is referred to [IMS21, Eq. (1.22)] for it.

If one applies SS to (M, ν) and then Υ to the result, the ν in the resulting output $(V, W, \kappa; \nu)$ is the same as the starting one (see the line following [IMS21, Theorem 1.4]). Thus the ν can be factored out, yielding a bijection between the set of M and the set of (V, W, κ); we call this $\tilde{\Upsilon}$, as in [IMS21].

Now, [IMS21, Theorem 1.2] asserts that $L^{(j)}(M)$ is equal to $\mu_1 + \cdots + \mu_j$, the sum of the lengths of the first j rows in the partition μ from the previous paragraph, for all $1 \leq j \leq N$ (this is a deterministic statement for L defined with respect to any fixed

entries of the environment). So we need to understand the distribution of μ under the map $\tilde{\Upsilon}$ when the entries $m_{(i,j);k}$ of M are distributed as independent $\text{Geo}(q^k a_i b_j)$, in particular, show that it is $\mathbb{W}_{a,b}^{(q)}$.

For this task we will need certain weight preservation properties of $\tilde{\Upsilon}$ which we record in the next lemma.

Lemma A.5. *For an infinite matrix M and $(V, W; \kappa)$ as above, define weight functions*

$$W_1(M) = \left(\sum_k \sum_j m_{(i,j);k}\right)_{1 \leq i \leq N}$$

$$W_2(M) = \left(\sum_k \sum_i m_{(i,j);k}\right)_{1 \leq j \leq N}$$

$$W_3(M) = \sum_k \sum_{i,j} k m_{(i,j);k}.$$

and, with $\#(U, i)$ being the number of times the entry i appears in the vertically strict tableaux U,

$$\widetilde{W}_1(V, W, \kappa) = \big(\#(V, i)\big)_{1 \leq i \leq N}$$
$$\widetilde{W}_2(V, W, \kappa) = \big(\#(W, j)\big)_{1 \leq j \leq N}$$
$$\widetilde{W}_3(V, W, \kappa) = \mathcal{H}(V) + \mathcal{H}(W) + \sum_i \kappa_i,$$

where $\mathcal{H}$ is the "intrinsic energy function". Its definition is complicated and not strictly needed for our purposes, so the interested reader is referred to [IMS21, Definition 7.4] *for a precise definition.*

Then, if M and (V, W, κ) are in bijection via $\tilde{\Upsilon}$, it holds that, for $i = 1, 2, 3$,

$$W_i(M) = \widetilde{W}_i(V, W, \kappa).$$

We will prove this after completing the proof of Theorem 1.3. We wish to calculate $\mathbb{P}(L^{(k)}(M) = \sum_{i=1}^k \mu_i$ for all$1 \leq k \leq n)$ where M is distributed according to independent geometric random variables as above, and show that this is equal to the first k row lengths marginal of the q-Whittaker measure. We will instead show the stronger statement that the law of the shape of V (or W) obtained by applying $\tilde{\Upsilon}$ to M with $m_{(i,j);k} \sim \text{Geo}(q^k a_i b_j)$ is the q-Whittaker measure. Then marginalizing to the lengths of the first k rows will complete the proof. Denoting the V obtained by applying $\tilde{\Upsilon}$ to M by $V(M)$ and by $\text{vst}(\mu)$ the set of vertically-strict tableau of shape μ,

$$\mathbb{P}\left(V(M) \in \text{vst}(\mu)\right) \propto \sum_{M:V(M)\in\text{vst}(\mu)} \prod_{i,j,k} (q^k a_i b_j)^{m_{(i,j);k}}$$

$$= \sum_{M:V(M)\in\text{vst}(\mu)} q^{\sum_{i,j,k} k m_{(i,j);k}} \prod_i a_i^{\sum_j m_{(i,j);k}} \prod_j b_j^{\sum_i m_{(i,j);k}}$$

$$= \sum_{M:V(M)\in\text{vst}(\mu)} q^{W_3(M)} a^{W_1(M)} b^{W_2(M)}.$$

By applying the bijection $\widetilde{\Upsilon}$ and Lemma A.5, and recalling the definitions of $\widetilde{W}_i$, we see that the previous line equals (letting $x^V = \prod_i x_i^{\#(V,i)}$)

$$\sum_{\substack{V,W \in \mathrm{vst}(\mu) \\ \kappa \in \mathcal{K}(\mu)}} q^{\widetilde{W}_3(V,W,\kappa)} a^{\widetilde{W}_1(V,W,\kappa)} b^{\widetilde{W}_2(V,W,\kappa)}$$

$$= \sum_{\kappa \in \mathcal{K}(\mu)} q^{|\kappa|} \times \sum_{V \in \mathrm{vst}(\mu)} q^{\mathcal{H}(V)} a^V \times \sum_{W \in \mathrm{vst}(\mu)} q^{\mathcal{H}(W)} b^W.$$

Now it is known that the first factor is $\mathfrak{b}_\mu(q)$ (see for example [IMS21, Eq. (10.5)]), while the second and third factors are $\mathcal{P}_\mu(a; q)$ and $\mathcal{P}_\mu(b; q)$ (see [IMS21, Proposition 10.1]). Thus the RHS is the unnormalized probability mass function of $\mathbb{W}_{a,b}^{(q)}$ at μ, as desired. $\qquad\square$

Proof of Lemma A.5. That $W_3(M) = \widetilde{W}_3(V, W, \kappa)$ follows by combining Eq. (1.23) in [IMS21, Theorem 1.4] (on the conservation of the quantity under Υ) with Eq. (4.15) in [IMS21, Theorem 4.11] (on its conservation under SS).

For $W_i(M)$ for $i = 1, 2$ we will similarly quote separate statements for its conservation under SS and Υ. Under SS, this is a consequence of [SS90, Theorem 6.6] (which is also the source of [IMS21, Theorem 4.11] mentioned in the previous paragraph), i.e., it holds that $W_1(M) = (\#(P, i))_{1 \leq i \leq N}$ and $W_2(M) = (\#(Q, j))_{1 \leq j \leq N}$. For Υ this preservation property is not recorded explicitly in [IMS21], but it is easy to see it from its definition. Indeed, as described in [IMS21, Sections 1.2 and 3.3], the output (V, W) of Υ is obtained as the asymptotic result of iteratively applying a map known as the *skew RSK* map to (P, Q), and it is immediate from the definition of this map that it does not change the number of times any entry i appears in P or Q (only possibly the location of the entries and/or the shape of the tableaux). Thus this property carries over to Υ. $\square$

Appendix B: Asymptotics for the Sum in the Factorial Moments Formula

Here we obtain upper and lower bounds (with the correct dependencies on q and k) on the sum in (20), which we label S, i.e., (recall $H(x) = -x \log x - (1 - x) \log(1 - x)$)

$$S := \sum_{i=0}^{k} \frac{1}{\frac{i}{k}(1 - \frac{i}{k})} \exp\left[N \left\{ \frac{i}{N} \log q^{-1} + \frac{2k}{N} \cdot H(i/k) \right. \right.$$
$$\left. \left. + \left(1 - \frac{i+1}{N}\right) \log\left(1 + \frac{k+1}{N-i-1}\right) + \frac{k+1}{N} \log\left(1 + \frac{k-i}{N}\right) - \frac{k+1}{N} \right\} \right]. \quad (36)$$

As indicated, the idea behind the analysis is simply Laplace's method, but it must be done carefully and explicitly here since we need to obtain the estimates uniformly in q. We start with the upper bound, and turn to the lower bound in Sect. B.2.

B.1 The upper bound.

Proposition B.1. *There exist positive constants C, k_0, and N_0 such that for all $N \geq N_0$, $k_0 \leq k \leq N$, and $q \in (0, 1)$, there exists $x_0 = 1 - \Theta(q^{1/2})$ such that*

$$S \leq Cq^{-1/4}k^{1/2} \left[\frac{(1+q^{1/2})^2}{q} \right]^k \exp\left(N \left[(\alpha - \mu_q) \log\left(1 - \frac{\alpha}{\mu_q}\right) - \alpha + Cq^{1/2}\alpha^3 \right] \right).$$

Proof. Recall that α is defined by $k = \alpha N$. Also recall the definition of $f_\alpha, g : [0, 1] \to \mathbb{R}$ from (19) by

$$f_\alpha(x) = \alpha x \log q^{-1} + 2\alpha H(x) + \alpha \log(1 + \alpha(1 - x))$$
$$+ (1 - \alpha x) \log\left(1 + \frac{\alpha}{1 - \alpha x}\right) - \alpha g(x) = (x(1 - x))^{-1}. \tag{37}$$

Then the sum S is

$$S = \sum_{i=1}^{k-1} g(i/k) \exp\left(N f_\alpha(i/k) +)(1)\right);$$

the $O(1)$ term is to account for the stray ± 1 we have ignored when going from the definition of S to its form in terms of f_α.

The first step is to identify the location x_0 where f_α is maximized.

$$f_\alpha'(x) = \alpha \log q^{-1} + 2\alpha \log\left(\frac{1-x}{x}\right) - \alpha \log\left(1 + \frac{\alpha}{1 - \alpha x}\right)$$
$$+ (1 - \alpha x) \cdot \frac{1}{1 + \frac{\alpha}{1-\alpha x}} \cdot \frac{\alpha}{(1 - \alpha x)^2} \cdot \alpha$$
$$+ \alpha \cdot \frac{1}{1 + \alpha(1 - x)} \cdot (-\alpha)$$
$$= \alpha \log q^{-1} + 2\alpha \log\left(\frac{1-x}{x}\right) - \alpha \log\left(1 + \frac{\alpha}{1 - \alpha x}\right).$$

Therefore we see that $f_\alpha'(x) = 0$ is equivalent to

$$x^{-1} - 1 = q^{1/2}\left(1 + \frac{\alpha}{1 - \alpha x}\right)^{1/2}. \tag{38}$$

Observe that the LHS tends to ∞ as $x \to 0$ and equals 0 when $x = 1$. Further the LHS is decreasing while the RHS is increasing in x, and both are continuous in x. Thus there is a unique $x_0 \in (0, 1)$ satisfying (38).

Further, the same observations yield (by evaluating the RHS of (38) at $x = 0$ and $x = 1$ and solving for x on the LHS) that

$$x_0 \in \left[\frac{1}{1 + q^{1/2}(1 + \alpha)^{1/2}}, \frac{1}{1 + q^{1/2}(1 - \alpha)^{-1/2}}\right] =: I_{q,\alpha}. \tag{39}$$

We further see that the size of $I_{q,\alpha}$ is

$$O\left(q^{1/2}\left[(1 - \alpha)^{-1/2} - (1 + \alpha)^{1/2}\right]\right) = O\left(q^{1/2}\alpha^2\right). \tag{40}$$

As a result, if $x \in I_{q,\alpha}$, and since $|f_\alpha'(x)| = O(\alpha)$ for such x, the mean value theorem implies that $|f_\alpha(x) - f_\alpha(x_0)| = O(q^{1/2}\alpha^3)$.

Evaluating $f_\alpha(x_0)$. Using these estimates we may obtain an estimate of $f_\alpha(x_0)$. Let us first evaluate the first three terms from (37) at

$$x = x^* := \left(1 + q^{1/2} \exp(\tfrac{1}{2}\alpha)\right)^{-1}.$$

(Note that since $1 + x \leq \exp(x) \leq (1-x)^{-1}$, $x^* \in I_{q,\alpha}$.) We see that the three terms evaluated at x^* equal

$$\alpha x^* \log q^{-1} - 2\alpha (x^* \log x^* + (1 - x^*) \log(1 - x^*)) + \alpha \log \left(1 + \alpha(1 - x^*)\right)$$

$$= \alpha x^* \log q^{-1} - 2\alpha \left[x^* \log x^* + (1 - x^*) \log \frac{q^{1/2} \exp(\tfrac{1}{2}\alpha)}{1 + q^{1/2} \exp(\tfrac{1}{2}\alpha)} \right]$$

$$+ \alpha \log \left(1 + \alpha(1 - x^*)\right)$$

$$= \alpha x^* \log q^{-1} - 2\alpha$$

$$\left[x^* \log x^* + \tfrac{1}{2}(1 - x^*) \log (q \exp(\alpha)) + (1 - x^*) \log \frac{1}{1 + q^{1/2} \exp(\tfrac{1}{2}\alpha)} \right]$$

$$+ \alpha \log \left(1 + \alpha(1 - x^*)\right).$$

We recognize $(1 + q^{1/2} \exp(\alpha/2))^{-1}$ to be x^*, which results in a cancellation with the $x^* \log x^*$ term. With this, and writing $\log(1 + \alpha(1 - x^*)) = \alpha(1 - x^*) \pm O(\alpha^2(1 - x^*)^2)$, we see that the previous display equals

$$\alpha x^* \log q^{-1} - 2\alpha \left[\log x^* + \tfrac{1}{2}(1 - x^*) \log q + \tfrac{1}{2}(1 - x^*)\alpha \right]$$

$$+ \alpha \cdot (\alpha(1 - x^*)) \pm O(\alpha^3(1 - x^*)^2)$$

$$= \alpha x^* \log q^{-1} - 2\alpha \log x^* - \alpha(1 - x^*) \log q \pm O(\alpha^3(1 - x^*)^2)$$

$$= \alpha \log \left(q^{-1}(1 + q^{1/2} \exp(\tfrac{1}{2}\alpha))^2\right) \pm O\left(\alpha^3 q\right)$$

$$= \alpha \log \left(q^{-1}(1 + q^{1/2})^2\right) + \alpha^2 \frac{q^{1/2}}{1 + q^{1/2}} \pm O(\alpha^3 q^{1/2}),$$

where we performed a Taylor expansion of $\exp(\alpha/2)$ as well as of the logarithm in the last step.

Next we must move from this calculation done at x^* to x_0. Observe that the derivative with respect to x of both the first three terms in (37) of f_α (which we will refer to as $f_\alpha^{(1)}$ below) gives an expression with a common factor of α. By the mean value theorem, we therefore have that, for some $y \in [x_0, x^*]$ (or $y \in [x^*, x_0]$ depending on which is greater),

$$|f_\alpha^{(1)}(x_0) - f_\alpha^{(1)}(x^*)| = |x_0 - x^*| \cdot |(f_\alpha^{(1)})'(y)| = O(1)\alpha^3 q^{1/2},$$

the last equality using that $|(f_\alpha^{(1)})'(y)| = O(\alpha)$ and $|x_0 - x^*| = O(\alpha^2 q^{1/2})$ from (40) and since $x_0, x^* \in I_{q,\alpha}$.

Putting back in the the remaining two terms of f_α not included in $f_\alpha^{(1)}$, overall we have shown that

$$f_\alpha(x_0) = \alpha \log \left(q^{-1}(1 + q^{1/2})^2\right) + (1 - \alpha x_0) \log \left(1 + \frac{\alpha}{1 - \alpha x_0}\right)$$

$$- \alpha + \alpha^2 \frac{q^{1/2}}{1 + q^{1/2}} \pm O\left(\alpha^3 q^{1/2}\right).$$

This can be simplified further using the following cancellation, which we prove in Sect. B.3.

Lemma B.2. *Let $x_0 = x_0(\alpha)$ be the maximizer of f_α over $[0, 1]$. For $\alpha \in (0, \frac{1}{2}]$ (and adopting the convention that $O(1)$ refers to a quantity whose value is upper bounded by an absolute constant)*

$$(1 - \alpha x_0) \log \left(1 + \frac{\alpha}{1 - \alpha x_0}\right) + \alpha^2 \cdot \frac{q^{1/2}}{1 + q^{1/2}} + (\mu_q - \alpha) \log \left(1 - \frac{\alpha}{\mu_q}\right)$$
$$= \pm O(1) q^{1/2} \alpha^3. \tag{41}$$

This yields that

$$f_\alpha(x_0) = \alpha \log \left(q^{-1}(1 + q^{1/2})^2\right) + (\alpha - \mu_q) \log \left(1 - \frac{\alpha}{\mu_q}\right) - \alpha \pm O\left(\alpha^3 q^{1/2}\right). \tag{42}$$

Performing Laplace's method. For future reference we also record that

$$f_\alpha''(x) = -\frac{2\alpha}{x(1 - x)} - \frac{\alpha^3}{(1 - \alpha x)^2 + \alpha(1 - \alpha x)} \implies f_\alpha''(x_0) = -\Theta(\alpha q^{-1/2}) \tag{43}$$

since $x_0 = 1 - \Theta(q^{1/2})$. To apply Laplace's method, we need to have bounds on f_α over its domain, which we will obtain by Taylor approximations. We will expand to third order as we need to include the just calculated second order term precisely. So next we bound the third derivative of f_α.

Fix $c > 0$. We observe that there exists C such that for $x \in [\frac{1}{4}, 1 - cq^{1/2}]$ and $\alpha \in (0, \frac{1}{2}]$,

$$f_\alpha'''(x) = -2\alpha \left(\frac{1}{(1 - x)^2} - \frac{1}{x^2}\right) - \frac{\alpha^4 (2(1 - \alpha x) + \alpha)}{\left((1 - \alpha x)^2 + \alpha(1 - \alpha x)\right)^2} \geq -C\alpha q^{-1}. \tag{44}$$

So by Taylor's theorem, we see that, if $\varepsilon > 0$ is such that $x_0 - \varepsilon q^{1/2} \in [\frac{1}{4}, 1 - cq^{1/2}]$, then for some $y \in [x_0 - \varepsilon q^{1/2}, x_0]$,

$$f_\alpha(x_0 - \varepsilon q^{1/2}) = f_\alpha(x_0) - \varepsilon q^{1/2} f_\alpha'(x_0) + \frac{1}{2}\varepsilon^2 q f_\alpha''(x_0) - \frac{1}{6}\varepsilon^3 q^{3/2} f'''(y)$$
$$\leq f_\alpha(x_0) - \varepsilon q^{1/2} f_\alpha'(x_0) + \frac{1}{2}\varepsilon^2 q f_\alpha''(x_0) + \frac{1}{6}C\alpha q^{1/2}\varepsilon^3$$
$$= f_\alpha(x_0) + \frac{1}{2}\varepsilon^2 q f_\alpha''(x_0) + \frac{1}{6}C\alpha q^{1/2}\varepsilon^3$$

since $f_\alpha'(x_0) = 0$. We may pick ε_0 a constant depending only on C such that, if $0 < \varepsilon < \varepsilon_0$, then (recalling that $f_\alpha''(x_0) < 0$)

$$\frac{1}{6}C\alpha q^{1/2}\varepsilon < \frac{1}{4}q|f_\alpha''(x_0)|;$$

that ε_0 can be taken to not depend on q or α follows from the fact that $|f_\alpha''(x_0)|$ can be lower bounded by an absolute constant times $\alpha q^{-1/2}$, due to (43). Similarly, since $f'''(x) < C\alpha$ for $x > \frac{1}{4}$, it holds for $\varepsilon > 0$ that (since $x_0 > \frac{1}{4}$ always)

$$f_\alpha(x_0 + \varepsilon q^{1/2}) \le f_\alpha(x_0) + \tfrac{1}{2}\varepsilon^2 q f_\alpha''(x_0) + \tfrac{1}{6}C\alpha\varepsilon^3 q^{3/2} \le f_\alpha(x_0) + \tfrac{1}{4}\varepsilon^2 q f_\alpha''(x_0),$$

the last inequality for $\varepsilon < \varepsilon_0$ for some absolute constant ε_0 by similar reasoning as above.

So, for $-\varepsilon_0 < \varepsilon < \varepsilon_0$,

$$f_\alpha(x_0 + \varepsilon q^{1/2}) \le f_\alpha(x_0) - \tfrac{1}{4}\varepsilon^2 q |f_\alpha''(x_0)|. \tag{45}$$

The above controls f_α inside $[x_0 - \varepsilon q^{1/2}, x_0 + \varepsilon q^{1/2}]$, where $0 < \varepsilon < \varepsilon_0$. We will also need control outside this interval, which we turn to next. We observe that, since f_α is concave on $(0, 1)$ and $f_\alpha'(x_0) = 0$, it holds for $0 < \varepsilon < \varepsilon_0$ and $x \in (0, x_0 - \varepsilon q^{1/2}]$ that

$$\begin{aligned}
f_\alpha(x) &\le f_\alpha(x_0 - \varepsilon q^{1/2}) + (x - x_0 - \varepsilon q^{1/2}) f'(x_0 - \varepsilon q^{1/2}) \\
&\le f_\alpha(x_0) - \tfrac{1}{4}\varepsilon^2 q |f''(x_0)| - (x_0 - x) f'(x_0 - \varepsilon q^{1/2}) \\
&\le f_\alpha(x_0) - (x_0 - x)C\varepsilon\alpha. \tag{46}
\end{aligned}$$

the last inequality using again that $|f_\alpha''(x_0)| = \Theta(\alpha q^{-1/2})$ and using Taylor's theorem for f_α' around x_0 to obtain $f_\alpha'(x_0 - \varepsilon q^{1/2}) = f_\alpha'(x_0) - \varepsilon q^{1/2} f_\alpha''(y)$ for some $y \in [x_0 - \varepsilon q^{1/2}, x_0]$ and using (43) to then obtain that $f_\alpha'(x_0 - \varepsilon q^{1/2}) \ge C\varepsilon\alpha$.

Similarly, since f_α is concave, $f'(x_0) = 0$, and (45), it follows for $x \in [x_0 + \varepsilon q^{1/2}, 1]$ that

$$f_\alpha(x) \le f_\alpha(x_0 + \varepsilon q^{1/2}) \le f_\alpha(x_0) - \tfrac{1}{4}q\varepsilon^2 |f_\alpha''(x_0)| \le f_\alpha(x_0) - C\alpha q^{1/2}\varepsilon^2. \tag{47}$$

We next analyze the actual sum S. We break up S into three subsums S_1, S_2, and S_3. Let $0 < \varepsilon < \varepsilon_0$ be fixed. S_1 corresponds to $i = 1$ to $i = \lceil k(x_0 - \varepsilon q^{1/2})\rceil$, S_2 to $i = \lceil k(x_0 - \varepsilon q^{1/2})\rceil + 1$ to $\lceil k(x_0 + \varepsilon q^{1/2})\rceil$, and S_3 to $\lceil k(x_0 + \varepsilon q^{1/2})\rceil + 1$ to $k - 1$.

We start by bounding S_1 using the above groundwork. Recall that $g(x) = (x(1 - x))^{-1}$. Recall that $g(i/k) = k^2/(i(k-i))$. We drop the $\lceil\rceil$ in the notation for convenience. Using (46), that $q \le 1$, and that $\alpha N = k$,

$$\begin{aligned}
S_1 &= \sum_{i=1}^{k(x_0 - \varepsilon q^{1/2})+1} g(i/k)\exp\left(Nf_\alpha(i/k)\right) \\
&\le e^{Nf_\alpha(x_0)} \sum_{i=1}^{k(x_0 - \varepsilon q^{1/2})+1} \frac{k^2}{i(k-i)} \exp\left(-C\varepsilon(kx_0 - i)\right) \\
&= e^{Nf_\alpha(x_0)} \sum_{j=k\varepsilon q^{1/2}-1}^{kx_0-1} \frac{k^2}{(kx_0 - j)(k(1 - x_0) + j)} \exp\left(-C\varepsilon j\right).
\end{aligned}$$

It is easy to see that the sum is bounded by $Cq^{-1/2}$ for an absolute constant C depending on ε, as this is the behaviour near $j = k\varepsilon q^{1/2}$ (using that $x_0 = 1 - \Theta(q^{1/2})$).

At the same time, it is easy to check that for $k \geq 2$ and $1 \leq j \leq k - 1$, $g(i/k) \leq 2k$, so that S_1 is also upper bounded by $Ck \exp(Nf_\alpha(x_0))$. Thus

$$S_1 \leq C \min(q^{-1/2}, k) e^{Nf_\alpha x_0} \leq Cq^{-1/4} k^{1/2} e^{Nf_\alpha(x_0)}.$$

Next we bound S_3. We apply (47) and use again that $g(i/k) \leq 2k$ and $x_0 = 1 - \Theta(q^{1/2})$ to see that

$$
\begin{aligned}
S_3 &= \sum_{i=k(x_0+\varepsilon q^{1/2})+1}^{k-1} g(i/k) \exp\left(Nf_\alpha(i/k)\right) \\
&\leq e^{Nf_\alpha(x_0)} \cdot 2k \cdot k(1 - x_0 - \varepsilon q^{1/2}) \exp(-C\varepsilon^2 N\alpha q^{1/2}) \\
&\leq Ce^{Nf_\alpha(x_0)} \cdot k^2 \cdot q^{1/2} \exp(-Ckq^{1/2}).
\end{aligned}
$$

Now $k^2 q^{1/2} = q^{-1/4} k^{1/2} \cdot (kq^{1/2})^{3/2}$ and, since $x \mapsto x^{3/2} \exp(-cx)$ is uniformly bounded over $x \geq 0$, this implies from the previous display that

$$S_3 \leq Cq^{-1/4} k^{1/2} e^{Nf_\alpha(x_0)}.$$

Finally we turn to the main sum, S_2, which consists of the range $\frac{i}{k} \in [x_0 - \varepsilon q^{1/2}, x_0 + \varepsilon q^{1/2}]$. We first want to say that, for x in the same range, $g(x) \leq Cg(x_0)$ for some absolute constant C. Observe that g blows up near 1 (and x_0 can be arbitrarily close to 1), and it is to avoid this and thereby be able to control g on the mentioned interval that its upper boundary is of order $q^{1/2}$ above x_0.

Lemma B.3. *There exists an absolute constant C such that for $x \in [x_0 - \varepsilon q^{1/2}, x_0 + \varepsilon q^{1/2}]$, $g(x) \leq Cg(x_0)$.*

Proof. We have to upper bound $g(x)/g(x_0) = \frac{x_0(1-x_0)}{x(1-x)}$. When $x \in [x_0 - \varepsilon q^{1/2}, x_0]$, this ratio is upper bounded by x_0/x; since $x \geq x_0 - \varepsilon q^{1/2}$, and $x_0 \geq \frac{1}{4}$ always, if $\varepsilon < \frac{1}{8}$ say, the ratio is uniformly upper bounded. When $x \in [x_0, x_0 + \varepsilon q^{1/2}]$, this ratio is upper bounded by $(1 - x_0)/(1 - x)$; since $1 - x \geq 1 - x_0 - \varepsilon q^{1/2} \geq \frac{1}{2}(1 - x_0)$ (using that $\varepsilon q^{1/2} \leq \frac{1}{2}(1 - x_0) = cq^{1/2}$ whenever ε is small enough), the ratio is again upper bounded by a constant. $\square$

So we see, from Lemma B.3 and (45) that S_2 equals

$$\sum_{i=k(x_0-\varepsilon q^{1/2})+1}^{k(x_0+\varepsilon q^{1/2})} g(i/k) \exp\left(Nf_\alpha(i/k)\right)$$

$$\leq Cg(x_0) \sum_{i=k(x_0-\varepsilon q^{1/2})+1}^{k(x_0+\varepsilon q^{1/2})} \exp\left[N\left(f_\alpha(x_0) - c\left(\frac{i}{k} - x_0\right)^2 |f_\alpha''(x_0)|\right)\right]$$

$$\leq Cg(x_0)e^{Nf_\alpha(x_0)} \sum_{i=-\infty}^{\infty} \exp\left[-cNk^{-2}(i - kx_0)^2 |f_\alpha''(x_0)|\right]$$

$$= Cg(x_0)e^{Nf_\alpha(x_0)} \sum_{i=-\infty}^{\infty} \exp\left[-c\alpha^{-1}k^{-1}i^2 |f_\alpha''(x_0)|\right].$$

Recall that $f''_\alpha(x_0) = -\Theta(\alpha q^{-1/2})$, and so the coefficient of i^2 is of order $c(q^{1/2}k)^{-1}$. We bound the above series using Proposition B.4 ahead, which says, with $\gamma = c\alpha^{-1}k^{-1}$ $|f''_\alpha(x_0)| = \Theta((q^{1/2}k)^{-1})$, and using that $g(x_0) = \Theta(q^{-1/2})$,

$$S_2 \leq Cg(x_0)(\alpha^{-1}k^{-1}|f''_\alpha(x_0)|)^{-1/2}e^{Nf_\alpha(x_0)} \leq Cq^{-1/2} \cdot k^{1/2}q^{1/4} \cdot e^{Nf_\alpha(x_0)}$$
$$= Cq^{-1/4}k^{1/2}e^{Nf_\alpha(x_0)}.$$

Note that this estimate holds for all $q > 0$ and dos not require $q \geq k^{-2}$. Thus overall we have shown that

$$S = S_1 + S_2 + S_3 \leq Cq^{-1/4}k^{1/2}e^{Nf_\alpha(x_0)}.$$

Using the expression for $f_\alpha(x_0)$ from (42) and recalling $\alpha N = k$ completes the proof.
□

The following is the bound on the discrete Gaussian sum, more precisely a Jacobi theta function, which we used in the proof. It can be proved using the Poisson summation formula and straightforward bounds.

Proposition B.4 (page 157 of [SS11]). *There exists C and, for any $M > 0$, a constant $c_M > 0$ such that for $0 < \gamma \leq M$ (for the first inequality) and $\gamma > 0$ (for the second),*

$$c_M\gamma^{-1/2} \leq \sum_{i\in\mathbb{Z}} e^{-\gamma i^2} \leq C\gamma^{-1/2}.$$

B.2 The lower bound. Recall the definition of S from (36).

Proposition B.5. *There exist positive constants C and k_0 such that for all $q \in [k^{-2}, 1)$ and $k_0 \leq k \leq N$,*

$$S \geq C^{-1}q^{-1/4}k^{1/2}\left[\frac{(1+q^{1/2})^2}{q}\right]^k$$
$$\exp\left(N\left[(\alpha - \mu_q)\log\left(1 - \frac{\alpha}{\mu_q}\right) - \alpha - C\alpha^3 q^{1/2}\right]\right).$$

Proof. As in the proof of Proposition B.1, we focus around the point Nx_0, where $x_0 = (1 + q^{1/2}\exp(\frac{1}{2}\alpha))^{-1} \pm O(q^{1/2}\alpha^2)$ satisfies (38). So, for $\varepsilon, \varepsilon' > 0$ to be chosen,

$$S \geq \sum_{i=k(x_0-\varepsilon)}^{k(x_0+\varepsilon')} \frac{1}{\frac{i}{k}(1-\frac{i}{k})} \exp\left[N\left\{\frac{i}{N}\log q^{-1} + \frac{2k}{N}\cdot H(i/k)\right.\right.$$
$$\left.\left. + \left(1 - \frac{i+1}{N}\right)\log\left(1 + \frac{k+1}{N-i-1}\right) + \frac{k+1}{N}\log\left(1 + \frac{k-i}{N}\right) - \frac{k+1}{N}\right\}\right]$$
$$= \sum_{i=k(x_0-\varepsilon)}^{k(x_0+\varepsilon')} g(i/k)\exp\left[Nf_\alpha(i/k)\right],$$

where $g(x) = (x(1-x))^{-1}$ and f_α is as defined in (37). We break into two cases depending on whether $x_0 > \frac{1}{2}$ or $x_0 \leq \frac{1}{2}$. We start with the first case.

In this case the basic issue is that x_0 can be arbitrarily close to 1. Thus, since the contribution of the sum from $i = kx_0$ to $i = k - 1$ will be small anyway, we will ignore it and set $\varepsilon' = 0$. We will also set ε to be an absolute constant. With these values, we want to show that, for some $c > 0$ and all $x \in [x_0 - \varepsilon, x_0]$, it holds that $g(x) \geq cg(x_0)$. This is easy to verify by upper bounding $g(x)/g(x_0)$ for x in the same range using the expression for $g(x)$, the value of ε, and that $x_0 \in [\frac{1}{4}, 1]$. So, for some absolute constant $c > 0$,

$$S \geq cg(x_0) \sum_{i=k(x_0-\varepsilon)}^{kx_0} \exp\left[Nf_\alpha(i/k)\right].$$

Now as in the proof of Proposition B.1, we Taylor expand f_α around x_0 to obtain a lower bound on $f_\alpha(x)$ for $x \in [x_0 - \varepsilon, x_0]$:

$$f_\alpha(x) \geq f_\alpha(x_0) - \frac{(x - x_0)^2}{2}|f_\alpha''(x_0)| + \frac{1}{6} \inf_{y \in [x, x_0]} (x - x_0)^3 f_\alpha'''(y)$$

$$= f_\alpha(x_0) - \frac{(x - x_0)^2}{2}|f_\alpha''(x_0)| + \frac{1}{6}(x - x_0)^3 \sup_{y \in [x, x_0]} f_\alpha'''(y);$$

the second equality by noting that, since $x - x_0 < 0$, the infimum is equivalent to maximizing $f_\alpha'''(y)$ over $y \in [x, x_0]$. Now, if $x \geq \frac{1}{2}$, then $y \geq \frac{1}{2}$ in the previous display and $f_\alpha'''(y) < 0$ by the explicit formula (44), and so the last term in the previous display is non-negative, i.e., can be lower bounded by zero. If $x < \frac{1}{2}$, we can ensure that $\varepsilon < \frac{1}{4}$, so that $x_0 < \frac{3}{4}$. Then we see that each of $|f_\alpha''(x_0)|$ and $|f_\alpha'''(x)|$ are uniformly bounded by $C\alpha$ where C is independent of q, so, by picking ε small enough also independent of q, we can ensure that $|x - x_0|f_\alpha''(x) < Cf_\alpha''(x_0)$ for some absolute constant C.

Thus over all, we have shown that, in the case that $x_0 \geq \frac{1}{2}$, there exists $\varepsilon > 0$ and $C > 0$ independent of q such that, for $x \in [x_0 - \varepsilon, x_0]$,

$$f_\alpha(x) \geq f_\alpha(x_0) - C(x - x_0)^2|f_\alpha''(x_0)|.$$

Using this we see that

$$S \geq cg(x_0)e^{Nf_\alpha(x_0)} \sum_{i=k(x_0-\varepsilon)}^{kx_0} \exp\left(-CN|f_\alpha''(x_0)|\left(\frac{i}{k} - x_0\right)^2\right)$$

$$= \frac{1}{2}cg(x_0)e^{Nf_\alpha(x_0)} \sum_{i=k(x_0-\varepsilon)}^{k(x_0+\varepsilon)} \exp\left(-CN|f_\alpha''(x_0)|\left(\frac{i}{k} - x_0\right)^2\right).$$

Now it is easy to see that $\sum_{i:|i-kx_0|>k\varepsilon} \exp\left(-CNk^{-2}|f_\alpha''(x_0)|(i - kx_0)^2\right) \leq C \exp(-cN|f_\alpha''(x_0)|\varepsilon^2)$, so the previous display is lower bounded by (using Proposition B.4 in the third line)

$$cg(x_0)e^{Nf_\alpha(x_0)}\left[\sum_{i=-\infty}^{\infty} \exp\left(-CNk^{-2}|f_\alpha''(x_0)|(i - kx_0)^2\right) - C\exp(-cN|f_\alpha''(x_0)|\varepsilon^2)\right]$$

$$= cg(x_0)e^{Nf_\alpha(x_0)}\left[\sum_{i=-\infty}^{\infty} \exp\left(-C\alpha^{-1}k^{-1}|f_\alpha''(x_0)|i^2\right) - C\exp(-cN|f_\alpha''(x_0)|\varepsilon^2)\right]$$

$$\geq cg(x_0)e^{Nf_\alpha(x_0)}\left[\alpha^{1/2}k^{1/2}|f''_\alpha(x_0)|^{-1/2} - C\exp(-c\varepsilon^2 N|f''_\alpha(x_0)|)\right]$$

$$\geq cg(x_0)e^{Nf_\alpha(x_0)}\left[q^{1/4}k^{1/2} - C\exp(-c\varepsilon^2 kq^{-1/2})\right],$$

using that $|f''_\alpha(x_0)| = \Theta(\alpha q^{-1/2})$ and $\alpha N = k$. This is in turn lower bounded by $cq^{-1/4}k^{1/2}e^{Nf_\alpha(x_0)}$ since $g(x_0) = \Theta(q^{-1/2})$ since $x_0 = 1 - \Theta(q^{1/2})$.

In the case that $x_0 \leq \frac{1}{2}$, the same proof works after noting that, since also $x_0 \geq \frac{1}{4}$ for all q, α (as can be observed from (39)), quantities like $|f''(x_0)|$, $f'''(x)$, and $g(x)$ are all bounded above and below by absolute constants for $x \in [x_0 - \varepsilon, x_0 + \varepsilon]$ (assuming $\varepsilon < \frac{1}{8}$ say) and $\alpha \in (0, \frac{1}{2}]$. 　　　　　　　　　　　　□

B.3 The factorial to polynomial moment cancellation. Here we give the proof of Lemma B.2. The proof goes by doing a full series expansion of the expression in α around 0. Performing a Taylor expansion to second order in α of the expression under consideration for fixed x, and uniformly (over $\alpha \in (0, \frac{1}{2})$ and $x = x_0 = 1 - \Theta(q^{1/2})$) bounding the third derivative error term by $Cq^{1/2}$ is also possible but somewhat messier.

Proof of Lemma B.2. To establish this we will utilize the series expansion for the logarithm as well as for $(1 - x)^{-k}$. We start with the first term: for any x,

$$(1 - \alpha x)\log\left(1 + \frac{\alpha}{1 - \alpha x}\right) = \sum_{j=1}^{\infty}(-1)^{j-1}\frac{\alpha^j}{j(1 - \alpha x)^{j-1}}$$

$$= \alpha + \sum_{j=2}^{\infty}(-1)^{j-1}\frac{\alpha^j}{j}\sum_{i=0}^{\infty}(\alpha x)^i\binom{j+i-2}{i}.$$

Collecting the α^ℓ terms together, the previous line equals

$$\alpha + \sum_{\ell=2}^{\infty}\alpha^\ell\sum_{i=0}^{\ell-2}\frac{(-1)^{\ell-i-1}}{\ell-i}x^i\binom{\ell-2}{i} = \alpha - \tfrac{1}{2}\alpha^2 + \sum_{\ell=3}^{\infty}\alpha^\ell\sum_{i=0}^{\ell-2}\frac{(-1)^{\ell-i-1}}{\ell-i}x^i\binom{\ell-2}{i}.$$

Invoking Lemma B.6 ahead, the sums in the previous display equals

$$\sum_{\ell=3}^{\infty}\frac{\alpha^\ell}{\ell(\ell-1)}\left[(-1+x)^\ell + \ell(-1+x)^{\ell-1} - x^\ell\right]$$

$$= \sum_{\ell=3}^{\infty}\frac{\alpha^\ell}{\ell(\ell-1)}\left[(-1+x)^{\ell-1}(-1+x+\ell) - x^\ell\right].$$

Now turning to the third term in (41),

$$(\mu_q - \alpha)\log\left(1 - \frac{\alpha}{\mu_q}\right) = -(\mu_q - \alpha)\sum_{j=1}^{\infty}\frac{\alpha^j}{j\mu_q^j} = -\alpha + \frac{\alpha^2}{2\mu_q} + \sum_{\ell=3}^{\infty}\frac{\alpha^\ell}{\ell(\ell-1)\mu_q^{\ell-1}}.$$

Thus we see overall that the LHS of (41) equals (where x_0 is the solution of (38))

$$\alpha^2 \left(-\frac{1}{2} + \frac{1}{2\mu_q} + \frac{q^{1/2}}{1 + q^{1/2}} \right)$$

$$+ \sum_{\ell=3}^{\infty} \frac{\alpha^\ell}{\ell(\ell - 1)} \left[(-1 + x_0)^{\ell-1}(-1 + x_0 + \ell) - x_0^\ell + \mu_q^{-(\ell-1)} \right]. \tag{48}$$

We focus on the coefficient of α^2 first, which, using that $\mu_q = (1+q^{1/2})/(1-q^{1/2})$, simplifies to 0.

Now we turn to the sum in (48). We observe that $\mu_q = 1 + \Theta(q^{1/2})$ and recall that $x_0 = 1 - \Theta(q^{1/2})$ (from (39)), so that the expression in the square brackets is equal to

$$\Theta(1)(-1)^{\ell-1}\ell q^{(\ell-1)/2} - 1 + \Theta(\ell q^{1/2}) + 1 - \Theta(\ell q^{1/2}) = \pm O(\ell q^{1/2}),$$

implying that the sum in (48) is $\pm O(1) \sum_{\ell=3}^{\infty} \frac{q^{1/2}\alpha^\ell}{(\ell-1)} = \pm O(1)q^{1/2}\alpha^3$. This completes the proof. $\qquad\square$

Lemma B.6. *It holds for any $x \in \mathbb{R}$ and $\ell \geq 2$ that*

$$\ell(\ell - 1) \sum_{i=0}^{\ell-2} \frac{(-1)^{i+1}}{\ell - i} x^i \binom{\ell - 2}{i} = (-1 + x)^\ell + \ell(-1 + x)^{\ell-1} - x^\ell.$$

Proof. We do the change of variable $i \mapsto \ell - i - 2$ and use $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$ twice to write the LHS as

$$\ell(\ell - 1) \sum_{i=0}^{\ell-2} \frac{(-1)^{\ell-i-1}}{i + 2} x^{\ell-i-2} \binom{\ell - 2}{i} = \sum_{i=0}^{\ell-2} (-1)^{i+1}(i + 1)x^{\ell-i-2} \binom{\ell}{i + 2}.$$

Now we do a change of variable $(i \mapsto i - 2)$ and add and subtract the terms corresponding to $i = 0$ and $i = 1$ in the new indexing to obtain

$$\sum_{i=0}^{\ell} (-1)^{i+1} x^{\ell-i}(i - 1) \binom{\ell}{i} - x^\ell.$$

Multiplying out over the $(i - 1)$ factor and using the binomial theorem on each of the resulting terms yields the claim. $\qquad\square$

Appendix C: Concentration Inequality Proofs

Here we prove the concentration inequality Theorem 5.1. As is typical for such inequalities, the main step is to obtain a bound on the moment generating function.

Proposition C.1. *Suppose X is such that*

$$\mathbb{P}(X \geq t) \leq C_1 \exp(-\rho t^{3/2}) \tag{49}$$

for some $\rho > 0$, and all $t \geq 0$. Then, there exist positive absolute constants C and $c > 0$ such that, for $\lambda > 0$,

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \exp\left\{C \cdot C_1\left[\lambda\rho^{-2/3} + \lambda^3\rho^{-2}\right]\right\}.$$

Proof. By rescaling X as $\rho^{2/3}X$ it is enough to prove the proposition with $\rho = 1$. First, we see that

$$\mathbb{E}[e^{\lambda X}] = \mathbb{E}\left[\lambda\int_{-\infty}^{X} e^{\lambda x}\, dx\right] = \mathbb{E}\left[\lambda\int_{-\infty}^{\infty} e^{\lambda x}\mathbb{1}_{X \geq x}\, dx\right] = \lambda\int_{-\infty}^{\infty} e^{\lambda x} \cdot \mathbb{P}(X \geq x)\, dx.$$

Next, we break up the integral into two at 0 and use the hypothesis (49) for the second term:

$$\lambda\int_{-\infty}^{\infty} e^{\lambda x} \cdot \mathbb{P}(X \geq x)\, dx = \lambda\int_{-\infty}^{0} e^{\lambda x} \cdot \mathbb{P}(X \geq x)\, dx + \int_{0}^{\infty} e^{\lambda x} \cdot \mathbb{P}(X \geq x)\, dx$$

$$\leq \lambda\int_{-\infty}^{0} e^{\lambda x}\, dx + C_1\lambda\int_{0}^{\infty} e^{\lambda x - x^{3/2}}\, dx$$

$$= 1 + C_1\lambda\int_{0}^{\infty} e^{\lambda x - x^{3/2}}\, dx.$$

We focus on the second integral now, and make the change of variables $y = \lambda^{-2}x \iff x = \lambda^2 y$, to obtain

$$C_1\lambda\int_{0}^{\infty} e^{\lambda x - x^{3/2}}\, dx = C_1\lambda^3\int_{0}^{\infty} e^{\lambda^3(y - y^{3/2})}\, dy.$$

We upper bound this essentially using Laplace's method. We first note that, for all $y > 0$, it holds that $y - y^{3/2} \leq 1 - \frac{1}{2}y^{3/2}$. So

$$C_1\lambda^3\int_{0}^{\infty} e^{\lambda^3(y - y^{3/2})}\, dy \leq C_1\lambda^3 \cdot \int_{0}^{\infty} e^{\lambda^3(1 - \frac{1}{2}y^{3/2})}\, dy = C_1\lambda^3 \cdot 2^{2/3}\Gamma(5/3)\lambda^{-2}e^{\lambda^3}$$

$$= C \cdot C_1\lambda e^{\lambda^3}.$$

for positive absolute constants C and c. Putting all the above together yields that

$$\mathbb{E}[e^{\lambda X}] \leq 1 + C \cdot C_1\lambda e^{c\lambda^3}.$$

Now we break into two cases depending on whether λ is less than or greater than 1: if $\lambda \leq 1$, then, since $1 + x \leq \exp(x)$ and $C \cdot C_1\exp(c\lambda) \leq C \cdot C_1\exp(c)$, we obtain, for an absolute constant $\tilde{C}$,

$$\mathbb{E}[e^{\lambda X}] \leq 1 + C \cdot C_1\exp(c)\lambda \leq \exp\left(\tilde{C} \cdot C_1\lambda\right).$$

On the other hand if $\lambda \geq 1$, we observe that, since $x \leq \exp(x)$, and by increasing the coefficient in the exponent,

$$1 + C \cdot C_1\lambda^3 e^{c\lambda^3} \leq e^{c'\lambda^3}.$$

It is easy to see that we may take c' to depend linearly on C_1. Thus overall, since $\max(a, b) \leq a + b$ when $a, b \geq 0$, we obtain, for some universal constant C,

$$\mathbb{E}\left[e^{\lambda X}\right] \le \exp\left(C \cdot C_1(\lambda + \lambda^3)\right).$$

$\square$

Proof of Theorem 5.1. Following the proof of the Chernoff bound, we exponentiate inside the probability (with $\lambda > 0$ to be chosen shortly) and apply Markov's inequality:

$$\mathbb{P}\left(\sum_{i=1}^{I} X_i \ge t + C \cdot C_1\sigma_{2/3}\right) = \mathbb{P}\left(e^{\lambda \sum_{i=1}^{I} X_i} \ge e^{\lambda t + \lambda C \cdot C_1 \cdot \sigma_{2/3}}\right)$$

$$\le e^{-\lambda t - \lambda C \cdot C_1 \cdot \sigma_{2/3}} \prod_{i=1}^{I} \mathbb{E}\left[e^{\lambda X_i}\right].$$

Using Proposition C.1, this is bounded by

$$\exp\left(-\lambda t - \lambda C \cdot C_1\sigma_{2/3} + C \cdot C_1\left[\lambda \sum_{i=1}^{I} \rho_i^{-2/3} + \lambda^3 \sum_{i=1}^{I} \rho_i^{-2}\right]\right)$$

$$= \exp\left(-\lambda t + C \cdot C_1 \cdot \lambda^3 \sigma_2\right),$$

the penultimate line using Proposition C.1 for $\lambda > 0$ to be chosen soon. Optimizing over λ and setting it to $c't^{1/2}\sigma_2^{-1/2}$ for some small constant $c' > 0$ now yields that,

$$\mathbb{P}\left(\sum_{i=1}^{I} X_i \ge t + C \cdot C_1 \cdot \sigma_{2/3}\right) \le \exp\left(-ct^{3/2}\sigma_2^{-1/2}\right).$$

$\square$

References

[ACG23] Aggarwal, A., Corwin, I., Ghosal, P.: The ASEP speed process. Advances in Mathematics **422**, 109004 (2023)

[ACQ11] Amir, G., Corwin, I., Quastel, J.: Probability distribution of the free energy of the continuum directed random polymer in 1+ 1 dimensions. Communications on pure and applied mathematics **64**(4), 466–537 (2011)

[ADH17] Auffinger, A., Damron, M., Hanson, J.: 50 Years of First-Passage Percolation, vol. 68. American Mathematical Society (2017)

[Agg23] Aggarwal, A.: Universality for lozenge tiling local statistics. Ann. Math. (2023) (To appear)

[AGH21] Augeri, F., Guionnet, A., Husson, J.: Large deviations for the largest eigenvalue of sub-Gaussian matrices. Commun. Math. Phys. **383**, 997–1050 (2021)

[AH21] Aggarwal, A., Huang, J.: Edge statistics for lozenge Tilings of polygons, II: Airy line ensemble. arXiv:2108.12874 (2021)

[BBC20] Barraquand, G., Borodin, A., Corwin, I.: Half-space Macdonald processes. In: Forum of Mathematics, Pi, vol. 8, p. e11. Cambridge University Press (2020)

[BC14] Borodin, A., Corwin, I.: Macdonald processes. Probab. Theory Relat. Fields **158**(1), 225–400 (2014)

[BCD21] Barraquand, G., Corwin, I., Dimitrov, E.: Fluctuations of the log-gamma polymer free energy with general parameters and slopes. Probab. Theory Relat. Fields **181**(1), 113–195 (2021)

[BCFV15] Borodin, A., Corwin, I., Ferrari, P., Vető, B.: Height fluctuations for the stationary KPZ equation. Math. Phys. Anal. Geom. **18**(1), 1–95 (2015)

[BDM+01] Baik, J., Deift, P., McLaughlin, K.D.T., Miller, P., Zhou, X.: Optimal tail estimates for directed last passage site percolation with geometric random variables. Adv. Theor. Math. Phys. **5**(6), 1–41 (2001)

[BF08] Borodin, A., Ferrari, P.: Large time asymptotics of growth models on space-like paths I: PushASEP. Electron. J. Probab. **13**, 1380–1418 (2008)

[BG16] Borodin, A., Gorin, V.: Moments match between the KPZ equation and the Airy point process. SIGMA. Symmetry Integrability Geom. Methods Appl. **12**, 102 (2016)

[BG21] Basu, R., Ganguly, S.: Time correlation exponents in last passage percolation. In: In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius, pp. 101–123 (2021)

[BGHH22] Basu, R., Ganguly, S., Hammond, A., Hegde, M.: Interlacing and scaling exponents for the geodesic watermelon in last passage percolation. Commun. Math. Phys. **393**(3), 1241–1309 (2022)

[BGHK21] Basu, R., Ganguly, S., Hegde, M., Krishnapur, M.: Lower deviations in β-ensembles and law of iterated logarithm in last passage percolation. Isr. J. Math. **242**(1), 291–324 (2021)

[BGS21] Basu, R., Ganguly, S., Sly, A.: Upper tail large deviations in first passage percolation. Commun. Pure Appl. Math. **74**(8), 1577–1640 (2021)

[BGZ21] Basu, R., Ganguly, S., Zhang, L.: Temporal correlation in last passage percolation with flat initial condition via Brownian comparison. Commun. Math. Phys. **383**, 1805–1888 (2021)

[BHS18] Basu, R., Hoffman, C., Sly, A.: Nonexistence of bigeodesics in integrable models of last passage percolation. arXiv:1811.04908 (2018)

[BL18] Baik, J., Liu, Z.: Fluctuations of TASEP on a ring in relaxation time scale. Commun. Pure Appl. Math. **71**(4), 747–813 (2018)

[BL19] Baik, J., Liu, Z.: Multipoint distribution of periodic TASEP. J. Am. Math. Soc. **32**(3), 609–674 (2019)

[BL21] Baik, J., Liu, Z.: Periodic TASEP with general initial conditions. Probab. Theory Relat. Fields **179**(3), 1047–1144 (2021)

[BO17] Borodin, A., Olshanski, G.: The ASEP and determinantal point processes. Commun. Math. Phys. **353**(2), 853–903 (2017)

[BO21] Betea, D., Occelli, A.: Peaks of cylindric plane partitions. arXiv:2111.15538 (2021)

[Bor07] Borodin, A.: Periodic Schur process and cylindric partitions. Duke Math. J. **140**(3), 391–468 (2007)

[Bor18] Borodin, A.: Stochastic higher spin six vertex model and Macdonald measures. J. Math. Phys. **59**(2), 023301 (2018)

[BP16] Borodin, A., Petrov, L.: Nearest neighbor Markov dynamics on Macdonald processes. Adv. Math. **300**, 71–155 (2016)

[BSS14] Basu, R., Sidoravicius, V., Sly, A.: Last passage percolation with a defect line and the solution of the slow bond problem. arXiv:1408.3464 (2014)

[BSS19] Basu, R., Sarkar, S., Sly, A.: Coalescence of geodesics in exactly solvable models of last passage percolation. J. Math. Phys. **60**(9), 093301 (2019)

[CC22] Cafasso, M., Claeys, T.: A Riemann–Hilbert approach to the lower tail of the Kardar–Parisi–Zhang equation. Commun. Pure Appl. Math. **75**(3), 493–540 (2022)

[CCO20] Cohen, P., Cunden, F.D., O'Connell, N.: Moments of discrete orthogonal polynomial ensembles. Electron. J. Probab. **25**, 1–19 (2020)

[CDG23] Cook, N.A., Ducatez, R., Guionnet, A.: Full large deviation principles for the largest eigenvalue of sub-Gaussian Wigner matrices. arXiv:2302.14823 (2023)

[CG20a] Corwin, I., Ghosal, P.: KPZ equation tails for general initial data. Electron. J. Probab. **25** (2020)

[CG20b] Corwin, I., Ghosal, P.: Lower tail of the KPZ equation. Duke Math. J. **169**(7), 1329–1395 (2020)

[CH14] Corwin, I., Hammond, A.: Brownian Gibbs property for Airy line ensembles. Invent. Math. **195**(2), 441–508 (2014)

[CH16] Corwin, I., Hammond, A.: KPZ line ensemble. Probab. Theory Relat. Fields **166**(1), 67–185 (2016)

[CHH23] Calvert, J., Hammond, A., Hegde, M.: Brownian structure in the KPZ fixed point. Astérisque (2023) (To appear)

[Dau23] Dauvergne, D.: Wiener densities for the Airy line ensemble. arXiv:2302.00097 (2023)

[DOV22] Dauvergne, D., Ortmann, J., Virág, B.: The directed landscape. Acta Math. (2022) (To appear)

[DV21a] Dauvergne, D., Virág, B.: Bulk properties of the Airy line ensemble. Ann. Probab. **49**(4), 1738–1777 (2021)

[DV21b] Dauvergne, D., Virág, B.: The scaling limit of the longest increasing subsequence. arXiv:2104.08210 (2021)

[EGO22] Emrah, E., Georgiou, N., Ortmann, J.: Coupling derivation of optimal-order central moment bounds in exponential last-passage percolation. arXiv:2204.06613 (2022)

[EJS20] Emrah, E., Janjigian, C., Seppäläinen, T.: Right-tail moderate deviations in the exponential last-passage percolation. arXiv:2004.04285 (2020)

[EJS21] Emrah, E., Janjigian, C., Seppäläinen, T.: Optimal-order exit point bounds in exponential last-passage percolation via the coupling technique. arXiv:2105.09402 (2021)

[FSV14]   Flores, G., Seppäläinen, T., Valkó, B.: Fluctuation exponents for directed polymers in the inter-mediate disorder regime. Electron. J. Probab. **19**, 1–28 (2014)

[GH21]   Guionnet, A., Husson, J.: Asymptotics of *k* dimensional spherical integrals and applications. arXiv:2101.01983 (2021)

[GH22]   Ganguly, S., Hegde, M.: Sharp upper tail estimates and limit shapes for the KPZ equation via the tangent method. arXiv:2208.08922 (2022)

[GH23]   Ganguly, S., Hegde, M.: Optimal tail exponents in general last passage percolation via bootstrap-ping and geodesic geometry. Probab. Theory Relat. Fields **186**(1), 221–284 (2023)

[Ham19]   Hammond, A.: A patchwork quilt sewn from Brownian fabric: regularity of polymer weight profiles in Brownian last passage percolation. In: Forum of Mathematics, Pi, vol. 7. Cambridge University Press (2019)

[Ham22]   Hammond, A.: Brownian regularity for the Airy line ensemble, and multi-polymer watermelons in Brownian last passage percolation. Mem. Am. Math. Soc. **277**(1363) (2022)

[Hua21]   Huang, J.: Edge statistics for lozenge Tilings of polygons, I: concentration of height function on strip domains. arXiv:2108.12872 (2021)

[HYZ23]   Huang, J., Yang, F., Zhang, L.: Pearcey universality at cusps of polygonal lozenge tiling. arXiv:2306.01178 (2023)

[IMS21]   Imamura, T., Mucciconi, M., Sasamoto, T.: Skew RSK dynamics: Greene invariants, affine crys-tals and applications to *q*-Whittaker polynomials. arXiv:2106.11922 (2021)

[IMS22]   Imamura, T., Mucciconi, M., Sasamoto, T.: Solvable models in the KPZ class: approach through periodic and free boundary Schur measures. arXiv:2204.08420 (2022)

[IS16]   Imamura, T., Sasamoto, T.: Determinantal structures in the O'Connell–Yor directed random polymer model. J. Stat. Phys. **163**(4), 675–713 (2016)

[Joh00a] Johansson, K.: Shape fluctuations and random matrices. Commun. Math. Phys. **209**(2), 437–476 (2000)

[Joh00b] Johansson, K.: Transversal fluctuations for increasing subsequences on the plane. Probab. Theory Relat. Fields **116**(4), 445–456 (2000)

[Joh01]   Johansson, K.: Discrete orthogonal polynomial ensembles and the Plancherel measure. Ann. Math. **153**(1), 259–296 (2001)

[KC18]   Kuchibhotla, A.K., Chakrabortty, A.: Moving beyond sub-Gaussianity in high-dimensional statis-tics: applications in covariance estimation and linear regression. arXiv:1804.02605 (2018)

[Led05a] Ledoux, M.: Deviation inequalities on largest eigenvalues. In: GAFA Seminar Notes (2005)

[Led05b] Ledoux, M.: Distributions of invariant ensembles from the classical orthogonal polynimials: the discrete case. Electron. J. Probab. **10**, 1116–1146 (2005)

[LR10]   Ledoux, M., Rider, B.: Small deviations for beta ensembles. Electron. J. Probab. **15**, 1319–1343 (2010)

[LS22a]   Landon, B., Sosoe, P.: Tail bounds for the O'Connell–Yor polymer. arXiv:2209.12704 (2022)

[LS22b]   Landon, B., Sosoe, P.: Upper tail bounds for stationary KPZ models. arXiv:2208.01507 (2022)

[Mac98]   Macdonald, I.G.: Symmetric Functions and Hall Polynomials. Oxford University Press, Oxford (1998)

[Mas14]   Masoero, D.: A Laplace's method for series and the semiclassical analysis of epidemiological models. arXiv:1403.5532 (2014)

[MP16]   Matveev, K., Petrov, L.: *q*-randomized Robinson–Schensted–Knuth correspondences and random polymers. Ann. Inst. Henri Poincaré D **4**(1), 1–123 (2016)

[MS09]   Mansour, T., Shabani, A.S.: Some inequalities for the *q*-digamma function. J. Inequal. Pure Appl. Math. **10**(1) (2009)

[OY01]   O'Connell, N., Yor, M.: Brownian analogues of Burke's theorem. Stoch. Process. Appl. **96**(2), 285–304 (2001)

[RRV11]   Ramirez, J., Rider, B., Virág, B.: Beta ensembles, stochastic Airy spectrum, and a diffusion. J. Am. Math. Soc. **24**(4), 919–944 (2011)

[Sep12]   Seppäläinen, T.: Scaling for a one-dimensional directed polymer with boundary conditions. Ann. Probab. **40**(1), 19–73 (2012)

[SS90]   Sagan, B.E., Stanley, R.P.: Robinson–Schensted algorithms for skew tableaux. J. Comb. Theory Ser. A **55**(2), 161–193 (1990)

[SS11]   Stein, E.M., Shakarchi, R.: Fourier Analysis: An Introduction, vol. 1. Princeton University Press, Princeton (2011)

[SS22]   Schmid, D., Sly, A.: Mixing times for the TASEP on the circle. arXiv:2203.11896 (2022)

[SSZ21]   Sarkar, S., Sly, A., Zhang, L.: Infinite order phase transition in the slow bond TASEP. arXiv:2109.04563 (2021)

[SV21]   Sarkar, S., Virág, B.: Brownian absolute continuity of the KPZ fixed point with arbitrary initial condition. Ann. Probab. (2021) (To appear)

[Vet22] Vető, B.: Asymptotic fluctuations of geometric q-TASEP, geometric q-PushTASEP and q-PushASEP. Stoch. Process. Their Appl. **148**, 227–266 (2022)

[Wid02] Widom, H.: On convergence of moments for random young tableaux and a random growth model. Int. Math. Res. Notices **2002**(9), 455–464 (2002)

Communicated by K. Johansson