THE LOWER TAIL OF q-PUSHTASEP

IVAN CORWIN AND MILIND HEGDE

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 [IMS21]. 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 as well as asymptotics extracted herein for high moments of the Meixner ensemble.

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, 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. 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, interacting particle system model of 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

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

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 supported 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 $x_1(T) < x_2(T) < \ldots < x_N(T)$; 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. 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 $\mathrm{gap}_k(T) = x_k(T) - x_{k-1}(T)$, $x_0(T) = -\infty$ 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

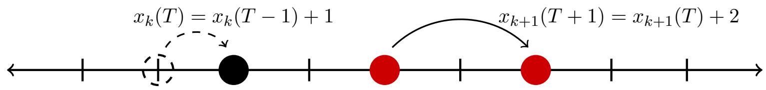

FIGURE 1. A depiction of one step in the evolution of q-pushTASEP. The dotted circle is the position of the k^{th} particle at time $T-1$, and the solid black circle is it after it moves to its position at time T. The left red circle is the $(k+1)^{\text{th}}$ particle at time T and the right one the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

particular that x_1's motion does not depend on that of any other particle, i.e., marginally it follows a random walk.

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 specify it more precisely, along with the definition of q-Whittaker measures, in Section 2.

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 Section 1.7.

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

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

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] and [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 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 N^{th} particle of q-pushTASEP, as introduced in Section 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 $\theta_0 = C|\log(\log q^{-1})|$and for $N \geq N_0$ and $\theta > \theta_0$,*

$$\mathbb{P}\left(X_N^{\mathrm{sc}} < -\theta N^{1/3}\right) \leq \exp\bigl(-c\theta^{3/2}\bigr).$$

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 1.6 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.

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. [BSS14, BGHH20, SS22]), process-level regularity properties (e.g. [CH14, CH16, CHH19, Ham22, Ham19]) of processes whose finite-dimensional distributions are accessible via exactly solvable tools, recent progress on constructing the ASEP speed process [ACG22], edge scaling behavior of tiling or dimer models [Hua21, AH21], as well as the construction of the directed landscape [DV21, DOV22]. In these works, a crucial input from the integrable side has repeatedly been bounds on the tails of the relevant statistic.

In the case of zero temperature models, these inputs had been studied two decades ago, with arguments relying in an essential way on determinantal structure possessed by these models. In positive temperature, 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. [BGS17] 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]), and connections to random matrix theory (e.g. [LR10, BGHK21]).

The toolbox for the upper tail is already fairly well developed in positive temperature. 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 k^{th} 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 Section 2 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 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 [ACG22], 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 [GH20]. 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. To explain this, 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 $N \times N$ "big squares" indexed by $k \in \mathbb{N} \cup \{0\}$, each of which contains N^2 "small squares" inside. These are arranged in a periodic strip as shown in Figure 2. To each small square with coordinates (i, j) (with $i, j \in \{1, \ldots, N\}$) in the k^{th} big square is associated an independent non-negative random variable $\xi_{(i,j);k}$ which we call a *site weight*; in our model, the site weights in the same big square, i.e., with the same value of k, will additionally be identically distributed. For s a site in the strip (say, with coordinates (i, j) in the k^{th} big box), we may also write ξ_s for $\xi_{(i,j);k}$.

We consider downward paths which are allowed to wrap around the strip; again see Figure 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 (see 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

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

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.

We now specify the distribution of the randomness of the site weights. We will say $X \sim \mathrm{Geo}(z)$ 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 $\mathrm{Geo}(u^2q^k)$ for $k = 0, 1, 2, \ldots$ and $i, j \in \{1, \ldots, N\}$.

Remark 1.2. Observe that $\mathbb{P}(\xi_{(i,j);k} \neq 0) = u^2q^k$ for all i, j, k, which is summable over $i, j \in \{1, \ldots, N\}$ and $k = 0, 1, \ldots$. So by the Borel-Cantelli lemma, almost surely, for all large enough k and all $i, j \in \{1, \ldots, N\}$, $\xi_{(i,j);k}$ will be zero.

This model of LPP is similar to other models of periodic LPP considered in the literature [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].

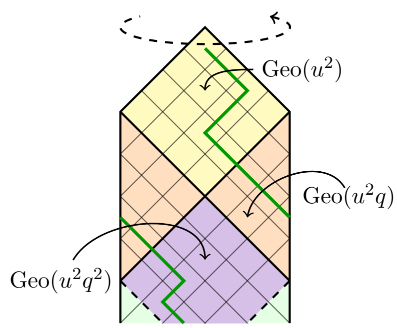

FIGURE 2. The environment in which the infinite last passage percolation occurs. The dashed arrow on top indicates the direction in which the squares wrap around, and the solid green line is a downward path which wraps around the strip.

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.

The connection 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 2.3) 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). See Theorem 2.4.

[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

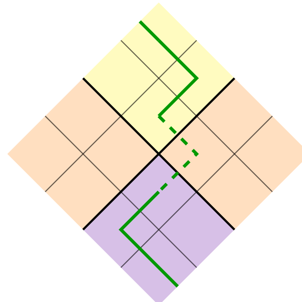

FIGURE 3. A depiction of the paths we consider near the boundary between different big squares. The two solid green paths go from the topmost site to the bottommost site in their respective big squares, where the environment is homogeneous. We do not consider the dotted green path needed to connect them, which is valid for proving an upper bound on the lower tail since including its weight will only increase the overall weight.

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.

1.6. **Proof ideas.** 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^2q^{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 Figure 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 geometric parameter u^2q^{2i}, to first order in N, the LPP value is $2N \times \frac{uq^i}{1-uq^i}$ (see for example [Joh00] or Theorem 1.3 ahead), so that the overall LPP value of the path we have described is, again to first order,

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

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

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

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^2q^{2i} in an $N \times N$ square is $u^{1/3}q^{i/3}(1-u^2q^{2i})^{-1}N^{1/3}$ (again see for example [Joh00] or Theorem 1.3 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 $f_qN - \Theta((\log q^{-1})^{-1}N^{1/3})$, 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 Section 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^2q^{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 1.3. *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 and N_0 such that, for $q \in (0,1)$, $N \geq N_0$, and $x > 0$,*

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

We next make some remarks on aspects of this result before outlining how to use Theorem 1.3 to complete the proof of Theorem 1.1.

Remark 1.4 (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 1.5 (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 $o(1)$. 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 1.6 (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. As we said earlier, the 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 [GH20], but one would first have to obtain similar uniform lower tail estimates to points displaced (on the $N^{2/3}$ scale) from (N,N). This should be doable, but we have not pursued it in this work as it is not necessary for our bounds on q-pushTASEP.

Let us finally say a few words about what we need to know about the Meixner operator's trace. It is well-known 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 [Led05a] shows 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 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] and then converting these into ones for the polynomial moments.

1.6.2. *Tying it together.* With Theorem 1.3, 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, eg. [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 1.7 (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 i^{th} 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 Section 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, 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 [GH20]. One expects that a similar program would deliver the corresponding bounds in the log-gamma case as well.

Acknowledgements. The authors thank Matteo Mucciconi for explaining the proof of Theorem 2.4, as well as Philippe Sosoe and Benjamin Landon for sharing their preprint [LS22a] with us in advance. I.C. was partially supported by the NSF through grants DMS:1937254, DMS:1811143, DMS:1664650, as well as through a Packard Fellowship in Science and Engineering, a Simons Fellowship, and a W.M. Keck Foundation Science and Engineering Grant. I.C. also thanks the Pacific Institute for the Mathematical Sciences (PIMS) and Centre de Recherches Mathematique (CRM), where some materials were developed in conjunction with the lectures he gave at the PIMS-CERM summer school in probability (which is partially supported by NSF grant DMS:1952466). M.H. was partially supported by NSF grant DMS:1937254.

2. The q-Whittaker measure and last passage percolation

As mentioned earlier, our strategy will rely on first relating the observable $x_N(N)$ to an infinite last passage problem in a periodic and inhomogeneous environment. The first step in establishing this relation requires us to introduce the q-Whittaker measure, and we start there.

2.1. q-Whittaker polynomials and measure.

Definition 2.1 (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 $\mathscr{b}_\mu(q)$ by

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

Definition 2.2 (q-Whittaker measure)**.** The q-Whittaker measure $\mathbb{W}^{(q)}_{a;b}$, first introduced in [BC14], is the measure on the set of all partitions given by

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

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

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

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 2.3 (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 $\mu \sim \mathbb{W}^{(q)}_{a;b}$ and let $x(T)$ be a q-PushTASEP under initial conditions $x_k(0) = k$ for $k = 1, \ldots, N$. Then,*

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

Finally, we can give the equivalence between q-pushTASEP and the LPP value. As indicated, this is a straightforward consequence of the results in [IMS21], and we will provide a proof, explained to us by Matteo Mucciconi, in Appendix A.

While this paper only considers the case $T = N$, we will state the LPP equivalence for general T. For this, the LPP problem we consider is in an infinite periodic environment where the "fundamental domain" has dimension $N \times T$ instead of $N \times N$ as described in Section 1.4. The distribution of the geometric random variables remains the same, i.e., it is $u^2 q^k$ in the k^{th} copy of the fundamental domain.

Theorem 2.4. *Let L be the LPP value in the environment just described with $u, q \in (0,1)$. Let $x_N(T)$ be the position of the N^{th} particle at time T in q-pushTASEP with $a_i = b_j = u$ for all $(i,j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$ and step initial condition. Then*

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

The proof actually combines Theorem 2.3 with a statement giving a distributional equality between L and the first row of a partition sampled from the q-Whittaker measure. 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.1.

3. Widom's trick applied to the lower tail in geometric LPP

In the next two sections we will prove Theorem 1.3, 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.

3.1. **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 3.1 (Meixner ensemble)**.** First let μ^q_{Geo} denote the $\mathrm{Geo}(q)$ distribution on $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, i.e., the distribution with discrete weights given by

$$\mu^q_{\mathrm{Geo}}(\{x\}) = (1-q)q^x.$$

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^q_{Geo}, 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 $M_N = \kappa_N x^N + \kappa_{N-1}x^{N-1} + \ldots + \kappa_0$ 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^q_{Geo}. 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 3.2 (Proposition 1.3 of [Joh00])**.** *Fix $q \in (0,1)$ and $N \in \mathbb{N}$. Let $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_N$ 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 $T_N \stackrel{d}{=} \lambda_1 - N + 1$.*

With this background we may now 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 $t \in \mathbb{R}$,

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

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

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

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 $g(x) = \sum_{i=1}^N u_i M_{i-1}(x)$,

$$\begin{aligned}1 = \sum_{i=1}^N u_i^2 = \langle g, g\rangle_{\ell^2(\mathbb{N}_0,\mu^q_{\mathrm{Geo}})} &= \langle g\mathbb{1}_{\cdot<t}, g\mathbb{1}_{\cdot<t}\rangle_{\ell^2(\mathbb{N}_0,\mu^q_{\mathrm{Geo}})} + \langle g\mathbb{1}_{\cdot\geq t}, g\mathbb{1}_{\cdot\geq t}\rangle_{\ell^2(\mathbb{N}_0,\mu^q_{\mathrm{Geo}})}\\ &\geq \langle g\mathbb{1}_{\cdot\geq t}, g\mathbb{1}_{\cdot\geq t}\rangle_{\ell^2(\mathbb{N}_0,\mu^q_{\mathrm{Geo}})} = \langle g, g\rangle_{\ell^2(\{t,t+1,\ldots\},\mu^q_{\mathrm{Geo}})} = u^T K_N^t u,\end{aligned}$$

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 $\rho_1^t, \ldots, \rho_N^t$. Since $1 - x \leq e^{-x}$ for $x \in [0,1]$,

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

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

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

$$\mathrm{Tr}(K_N^t) = \sum_{\ell=0}^{N-1} \langle M_{\ell-1}, M_{\ell-1}\rangle_{\ell(\{t,t+1,\ldots\},\mu^q_{\mathrm{Geo}})} = \int_t^\infty \sum_{\ell=0}^{N-1} M_\ell^2 \, \mathrm{d}\mu^q_{\mathrm{Geo}}. \tag{11}$$

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

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

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

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

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. ☐

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.

Theorem 3.4. *Let X be distributed as $\nu_{q,N}$ as defined in (10). There exist positive absolute constants c and N_0 such that, for $N \geq N_0$, $\varepsilon \in (0, \frac{1}{2})$, and $q \in [\varepsilon^3, 1)$,*

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

With this statement and Widom's trick, Theorem 1.3's proof is straightforward.

Proof of Theorem 1.3. Recall from Proposition 3.2 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 3.3, for any $t \in \mathbb{N}$,

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

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

$$\begin{aligned}\mathbb{P}\Big(T_N \leq (\mu_q - 1)N - \mu_q N q^{1/6}\varepsilon)\Big) &= \mathbb{P}\Big(\lambda_1 \leq \mu_q N(1 - q^{1/6}\varepsilon)\Big)\\ &\leq \exp\Big\{-c\nu_{q,N}\big([\mu_q N(1 - q^{1/6}\varepsilon), \infty)\big)\Big\}.\end{aligned}$$

By Theorem 3.4, there exists $c > 0$ such that, for all $q \in [\varepsilon^3, 1)$ and $0 < \varepsilon < \frac{1}{2}$,

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

Putting the above together with $\varepsilon = xN^{-2/3}$ 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 $q \geq x^3 N^{-2}$, which translates to $x \leq (q^{1/2}N)^{2/3}$. This completes the proof. ☐

To prove Theorem 3.4, 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 3.5. *Let X be distributed according to $\nu_{q,N}$ as defined in (10). For any $W > 0$, there exist positive C_W and N_0 such that for any $N \geq N_0$, $k \leq WN^{2/3}$ and $q \in [k^{-2}, 1)$,*

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

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). 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 3.4.

We will prove Theorem 3.5 in Sections 4 and 5; the upper and lower bounds are separated into Propositions 5.3 and 5.5 respectively.

We conclude this section by using Theorem 3.5 to implement Ledoux' argument to establish Theorem 3.4.

Proof of Theorem 3.4. First, by the Cauchy-Schwarz inequality, we have

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

By Theorem 3.5, when $q \geq k^{-2}$ 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

$$\begin{aligned}\mathbb{E}[X^{2k}\mathbb{1}_{X\geq\mu_q N(1-\varepsilon)}] &= \mathbb{E}[X^{2k}] - \mathbb{E}[X^{2k}\mathbb{1}_{X<\mu_q N(1-\varepsilon)}]\\ &\geq \mathbb{E}[X^{2k}] - \mathbb{E}[X^k]\left(\mu_q N(1-\varepsilon)\right)^k.\end{aligned}$$

Now, from Theorem 3.5,

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

while, again from Theorem 3.5,

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

Since overall it holds that

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

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

$$\begin{aligned}\mathbb{P}\big(X \geq \mu_q N(1-\varepsilon)\big) &\geq C_1^{-1}(q^{1/6}k)^{3/2}\left[C_2(q^{1/6}k)^{-3/2} - C_3(q^{1/6}k)^{-3/2}(1-\varepsilon)^k\right]^2\\ &= C_1^{-1}(q^{1/6}k)^{-3/2}\left[C_2 - C_3(1-\varepsilon)^k\right]^2\end{aligned}$$

We pick the absolute constant W such that, with $k = W\varepsilon^{-1}$, $C_3(1-\varepsilon)^k \leq \frac{1}{2}C_2$. This yields that for some absolute constant $C > 0$,

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

The earlier assumed condition that $q \geq k^{-2}$ translates into $q \geq \varepsilon^2$. Replacing ε by $q^{1/6}\varepsilon$ completes the proof, after noting that the resulting condition on q is that $q \geq \varepsilon^3$. ☐

4. Sharp factorial moment bounds

Here we prove Theorem 3.5 on sharp bounds for the moments of the expected empirical distribution $\nu_{q,N}$ (as defined in (10)) of the Meixner ensemble. While there is no explicit formula available for these moments, there *is* one for the factorial moments, i.e., moments of the form $\mathbb{E}[X(X-1)\cdots(X-k+1)]$. Indeed letting X be distributed according to $\nu_{q,N}$, [Led05b, Lemma 5.2] states that

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

In fact, this simplifies [CCO20, eq. (39)] to

$$M^q(k,N) = \frac{q^k}{(1-q)^k}\frac{1}{N}\cdot\frac{1}{k+1}\sum_{i=0}^{k} q^{-i}\binom{k}{i}^2\cdot\frac{(N+k-i)!}{(N-i-1)!}. \tag{12}$$

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

4.1. The factorial moment asymptotics.

Theorem 4.1. *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). For any $W > 0$, there exist positive constants C_W, N_0, and k_0 such that for all $N \geq N_0$, $k_0 \leq k \leq WN^{2/3}$, and $q \in [k^{-2}, 1)$,*

$$C_W^{-1}(q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(-\frac{k^2}{2\mu_q N}\right) \leq M^q(k,N) \leq C_W(q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(-\frac{k^2}{2\mu_q N}\right).$$

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.

Before turning to giving the proof in full, 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 (12) as, approximately and up to absolute constants,

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

Then, the idea is to obtain asymptotics for the sum using Laplace's method. Indeed, if we were to regard the sum (along with the $(k+1)^{-1}$ factor and ignoring the $-k^2/(2N)$ term in the exponent) as being approximately the integral

$$\int_0^1 \frac{1}{x(1-x)} \exp\left[k\left(x\log q^{-1} + 2H(x) + \frac{k}{N}(1-x)\right)\right] \mathrm{d}x$$

then, writing the integrand as $g(x)\exp(kf(x))$, where

$$f(x) = x\log q^{-1} + 2H(x) + \frac{k(1-x)}{N} \qquad \text{and} \qquad g(x) = (x(1-x))^{-1}$$

one would expect from the Laplace method intuition that

$$\frac{1}{k+1}\sum_{i=0}^{k} \frac{1}{\frac{i}{k}(1-\frac{i}{k})} \exp\left[k\left(\tfrac{i}{k}\log q^{-1} + 2H(i/k) + \frac{k(1-\frac{i}{k})}{N}\right)\right] \approx (k|f''(x_0)|)^{-1/2}\exp(kf(x_0))g(x_0),$$

up to constants, where x_0 is the maximizer of f on $[0,1]$. Evaluating $f(x_0)$ and $g(x_0)$ will yield the claimed q and k dependencies in Theorem 4.1.

There are existing results in the literature, for example [Mas14], on Laplace's method for sums which also obtain the correct constant. However these are not directly useful to us: our function f depends on k, which is not typical in Laplace's method, and we need all our estimates to be uniform in the parameter q, which can be difficult to verify after applying black-box theorems. For these reasons 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.4).

Proof of Theorem 4.1. In the following, an equality with a factor of $\Theta(f(k,N))$ on the right-hand side means that the left-hand side is upper and lower bounded by the rest of the right-hand side up to a constant times $f(k,N)$ replacing $\Theta(f(k,N))$, where the constant may depend on W but not q, i, k, or N (at least in the range $1 \leq i \leq k \leq WN^{2/3}$).

We will first obtain a bound on the summands in (12). We start with bounding $(N+k-i)!/(N-i-1)!$ using Stirling's approximation (non-asymptotic form) to obtain

$$\begin{aligned}\frac{(N+k-i)!}{(N-i-1)!} &= \frac{\sqrt{2\pi(N+k-i)}}{\sqrt{2\pi(N-i-1)}}\exp\Big[(N+k-i)\log(N+k-i)\\ &\qquad -(N+k-i)-(N-i-1)\log(N-i-1)+(N-i-1)+\Theta(N^{-1})\Big]\\ &=\Theta(1)\cdot\exp\left[(N-i-1)\log\frac{N+k-i}{N-i-1}+(k+1)\big(\log(N+k-i)-1\big)\right]\\ &=\Theta(1)\cdot\exp\left[(N-i-1)\log\left(1+\frac{k+1}{N-i-1}\right)+(k+1)\big(\log(N+k-i)-1\big)\right]\\ &=\Theta(1)\cdot\exp\left[k+1-\frac{(k+1)^2}{2(N-i-1)}+(k+1)\big(\log(N+k-i)-1\big)+\Theta\left(\frac{k^3}{N^2}\right)\right]\\ &=\Theta(1)\cdot\exp\left[(k+1)\log(N+k-i)-\frac{k^2}{2N}+\Theta\left(\frac{k^3}{N^2}+\frac{k}{N}\right)\right]\end{aligned}$$

Now since $k\le WN^{2/3}$, the k^3/N^2 and k/N terms are both bounded by a power of W, and so may be absorbed into the $\Theta(1)$ factor outside. So we see from (12) that

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

Now since

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

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

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

Rewriting the previous display a little, we obtain

$$\frac{\Theta(1)q^k}{(1-q)^k}\cdot k^{-1}\cdot\frac{1}{k+1}\sum_{i=0}^{k}\frac{1}{\frac{i}{k}(1-\frac{i}{k})}\exp\left[k\left(\tfrac{i}{k}\log q^{-1}+2H(i/k)+\frac{k(1-\frac{i}{k})}{N}\right)-\frac{k^2}{2N}\right].\tag{13}$$

The sum is upper and lower bounded Propositions B.1 and B.4 in Appendix B, up to a constant factor and for $N\ge N_0$, $k_0\le k\le N$, and $q\ge k^{-2}$, by

$$q^{-1/4}k^{1/2}\left[\frac{\left(1+q^{1/2}\exp\left(\frac12 kN^{-1}\right)\right)^2}{q}\right]^k.$$

Substituting this into (13) yields that $M^q(k,N)N^{-k}$ is equal to

$$\Theta(1)\cdot k^{-3/2}q^{-1/4}\exp\left[k\log\left(\frac{(1+q^{1/2}\exp(\frac{1}{2}kN^{-1}))^2}{1-q}\right)-\frac{k^2}{2N}\right]. \tag{14}$$

Now since $\exp\left(\frac{1}{2}kN^{-1}\right)=1+\frac{1}{2}kN^{-1}+\Theta(k^2/N^2)$, we can expand the logarithm of the numerator term in the previous display to obtain

$$\begin{aligned}\log\left(1+q^{1/2}e^{k/2N}\right)&=\log\left(1+q^{1/2}+\frac{1}{2}q^{1/2}kN^{-1}+\Theta(1)qk^2N^{-2}\right)\\&=\log(1+q^{1/2})+\frac{k}{2N}\cdot\frac{q^{1/2}}{1+q^{1/2}}+\Theta(1)\frac{qk^2}{N^2}\end{aligned}$$

Substituting this into (14) yields that $M^q(k,N)N^{-k}$ is equal to

$$\Theta(1)k^{-3/2}q^{-1/4}\exp\left[k\log\left(\frac{(1+q^{1/2})^2}{1-q}\right)+\frac{k^2}{2N}\left(\frac{2q^{1/2}}{1+q^{1/2}}-1\right)+\Theta(1)\frac{qk^3}{N^2}\right],$$

which, after simplifying and using that $1\le k\le WN^{2/3}$, equals

$$\Theta(1)(q^{1/6}k)^{-3/2}\mu_q^k\exp\left[-\frac{k^2}{2N}\cdot\frac{1-q^{1/2}}{1+q^{1/2}}\right]=\Theta(1)(q^{1/6}k)^{-3/2}\mu_q^k\exp\left[-\frac{k^2}{2\mu_qN}\right],$$

which completes the proof. □

5. Translating from factorial to polynomial moments

Now we convert the factorial moment bounds of Theorem 4.1 to polynomial bounds and thus prove Theorem 3.5. The basic idea is the following. Notice that the difference between the bounds in the two theorems is essentially the factor of $\exp(-k^2/(2\mu_qN))$ in the bounds for the factorial moments. The presence of this factor is a problem because, when $k=\Theta(N^{2/3})$ (a value of k we will need to take), this factor equals $\exp(-\mu_q^{-1}N^{1/3})$, i.e., much smaller than $O(1)$. So our goal is to show that the $\exp(-k^2/(2\mu_qN))$ goes away when we move to the k^{th} polynomial moment.

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

$$\prod_{i=1}^{k-1}(1-iX^{-1})\approx\exp\left(-X^{-1}\sum_{i=1}^{k}i\right)\approx\exp\left(-\frac{k^2}{2X}\right).$$

If we believe that the support of X essentially has upper boundary μ_qN, then, when considering high powers of X, heuristically one should get the correct behavior when replacing X by μ_qN. Doing so, the above then suggests that $\mathbb{E}[X^k]\approx\mathbb{E}[X(X-1)\cdots(X-k+1)]\exp(k^2/(2\mu_qN))$, which exactly cancels the extra factor we noted above for the latter.

The proof essentially comes down to making this heuristic precise. In many of the statements we will have terms of the form k^3/N^2 in the exponent, which comes from the next-order term after k^2/N; note that this term is $O(1)$ when $k=O(N^{2/3})$, the regime we care about, and so is controlled.

5.1. **General lemmas connecting factorial and polynomial moments.**

We start with two statements capturing, in each direction, the correct (to first order) conversion factor between factorial and polynomial moments for general $\mathbb{N}$-valued random variables. We adopt the notation

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

so that, for example, $M^q(k,N)=\mathbb{E}[(X)_k]$ when $X\sim\nu_{q,N}$ (as defined in (10)).

Lemma 5.1. *Let X be a non-negative integer-valued random variable. Then for any $k \in \mathbb{N}$ and $R \geq 2k$,*

$$\mathbb{E}[X^k \mathbb{1}_{X\geq R}] \leq \mathbb{E}[(X)_k] \exp\left(\frac{k^2}{2R} + \frac{k^3}{3R^2}\right).$$

Proof. As above, for any non-negative integer x, we see that

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

Next we use that $\log(1-y) \geq -y-y^2$ for $y \in (0, \frac{1}{2})$, so that, for $k-1 \leq x/2$, we obtain

$$\begin{aligned}(x)_k \geq x^k \exp\left(-\sum_{i=1}^{k-1}\left(\frac{i}{x}+\frac{i^2}{x^2}\right)\right) &= x^k \exp\left(-\frac{k(k-1)}{2x} - \frac{k(k-1)(2k-1)}{6x^2}\right)\\ &\geq x^k \exp\left(-\frac{k^2}{2x} - \frac{k^3}{3x^2}\right).\end{aligned}$$

Substituting X in place of x and taking expectations yields (since we have assumed $R \geq 2k$)

$$\mathbb{E}[(X)_k] \geq \mathbb{E}[(X)_k \mathbb{1}_{X\geq R}] \geq \mathbb{E}\left[X^k \exp\left(-\frac{k^2}{2X} - \frac{k^3}{3X^2}\right)\mathbb{1}_{X\geq R}\right].$$

Now since $x \mapsto \exp(-k^2/(2x) - k^3/(3x^2))$ is increasing, we see that

$$\mathbb{E}\left[X^k \exp\left(-\frac{k^2}{2X} - \frac{k^3}{3X^2}\right)\mathbb{1}_{X\geq R}\right] \geq \mathbb{E}\left[X^k \mathbb{1}_{X\geq R}\right]\exp\left(-\frac{k^2}{2R} - \frac{k^3}{3R^2}\right),$$

which, after rearranging, completes the proof. □

Lemma 5.2. *Let X be a non-negative integer-valued random variable. Then for any $k \in \mathbb{N}$ and $R > 0$,*

$$\mathbb{E}[X^k] \geq \left(\mathbb{E}[(X)_k] - \mathbb{E}[X^{2k}]^{1/2}\mathbb{P}(X > R)^{1/2}\right)\exp\left(\frac{k(k-1)}{2R}\right).$$

Proof. As in the previous proof, we see that (since $\log(1-x) \leq -x$)

$$(x)_k = x^k \times \prod_{i=1}^{k-1}\left(1-\frac{i}{x}\right) \leq x^k \times \exp\left(-\sum_{i=1}^{k-1}\frac{i}{x}\right) = x^k \times \exp\left(-\frac{k(k-1)}{2x}\right).$$

So with this, and since X is non-negative, for any $R > 0$,

$$\mathbb{E}[X^k] \geq \mathbb{E}[X^k \mathbb{1}_{X\leq R}] \geq \mathbb{E}\left[(X)_k \exp\left(\frac{k(k-1)}{2X}\right)\mathbb{1}_{X\leq R}\right] \geq \mathbb{E}\left[(X)_k \mathbb{1}_{X\leq R}\right]\cdot\exp\left(\frac{k(k-1)}{2R}\right),$$

the last inequality since $\exp(k(k-1)/(2x))$ is decreasing in x. To lower bound $\mathbb{E}[(X)_k \mathbb{1}_{X\leq R}]$ we use that $(X)_k \leq X^k$ and the Cauchy-Schwarz inequality to obtain

$$\begin{aligned}\mathbb{E}[(X)_k \mathbb{1}_{X\leq R}] &= \mathbb{E}[(X)_k] - \mathbb{E}[(X)_k \mathbb{1}_{X>R}]\\ &\geq \mathbb{E}[(X)_k] - \mathbb{E}[X^{2k}]^{1/2}\mathbb{P}(X > R)^{1/2}.\end{aligned}$$

□

5.2. **Applying the general lemmas to $\nu_{q,N}$: the upper bound.** In this section we apply Lemmas 5.1 to obtain the upper bound on $\mathbb{E}[X^k]$, and we will turn to the lower bounds in Section 5.3. The precise upper bound is the following. (Recall from (7) that $\mu_q = (1+q^{1/2})^2/(1-q)$.)

Proposition 5.3. *Let X be distributed according to $\nu_{q,N}$ as defined in (10) and μ_q be as in (7). For any $W > 0$, there exist positive constants C_W, k_0, and N_0 such that, for any $N \geq N_0$, $k_0 \leq k \leq WN^{2/3}$, and $q \in [k^{-2}, 1)$,*

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

To prove this, in the next Lemma 5.4, we first obtain a bound on $\mathbb{E}[X^k \mathbb{1}_{X \geq \mu_q N(1-\varepsilon)}]$ by directly applying Lemma 5.1. We will then convert this into a bound on $\mathbb{E}[X^k]$ to prove Proposition 5.3, thus proving the upper bound half of Theorem 3.5.

Lemma 5.4. *Let X be distributed according to $\nu_{q,N}$ as defined in (10) and μ_q be as in (7). Then for any $W > 0$ there exist positive constants C_W, k_0, and N_0 such that, for any $N \geq N_0$, $k_0 \leq k \leq WN^{2/3}$, $q \in [k^{-2}, 1)$, and $0 < \varepsilon < \frac{1}{2}$,*

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

Proof. By Lemma 5.1, and since $(1-\varepsilon)^{-1} \leq 1 + 2\varepsilon$ for $0 < \varepsilon < \frac{1}{2}$,

$$\begin{aligned}\mathbb{E}[X^k \mathbb{1}_{X \geq \mu_q N(1-\varepsilon)}] &\leq \mathbb{E}[(X)_k] \exp\left(\frac{k^2}{2\mu_q N(1-\varepsilon)} + \frac{k^3}{3\mu_q^2 N^2(1-\varepsilon)^2}\right) \\ &\leq \mathbb{E}[(X)_k] \exp\left(\frac{k^2}{2\mu_q N}(1+2\varepsilon) + \frac{Ck^3}{\mu_q^2 N^2}\right).\end{aligned}$$

By Theorem 4.1, for fixed W and $N \geq N_0$, $k_0 \leq k \leq WN^{2/3}$, and $q \geq k^{-2}$,

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

Thus we see that

$$\mathbb{E}[X^k \mathbb{1}_{X \geq \mu_q N(1-\varepsilon)}] \leq C_W (q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(\frac{k^2\varepsilon}{\mu_q N} + \frac{Ck^3}{\mu_q^2 N^2}\right),$$

which yields the inequality in the statement (using also that $\mu_q^{-1} \leq 1$ and relabeling C_W). ☐

Proof of Proposition 5.3. First we write, for any $\varepsilon > 0$,

$$\mathbb{E}[X^k] = \mathbb{E}[X^k \mathbb{1}_{X > \mu_q N(1-\varepsilon^{1/2})}] + \mathbb{E}[X^k \mathbb{1}_{X \leq \mu_q N(1-\varepsilon^{1/2})}].$$

By Lemma 5.4, the first term is upper bounded, for $N \geq N_0$, $k_0 \leq k \leq WN^{2/3}$, and $q \in [k^{-2}, 1)$, by

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

while, trivially, the second term is upper bounded by

$$(\mu_q N)^k (1-\varepsilon^{1/2})^k \leq (\mu_q N)^k \exp(-\varepsilon^{1/2}k).$$

Setting $\varepsilon = k^{-1}$, overall, we obtain that

$$\begin{aligned}\mathbb{E}[X^k] &\leq C_W (q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(\frac{k^{3/2}}{N}\right) + (\mu_q N)^k \exp(-k^{1/2}) \\ &\leq C_W (q^{1/6}k)^{-3/2}(\mu_q N)^k + (\mu_q N)^k \exp(-k^{1/2}),\end{aligned}$$

the last inequality since $k \leq WN^{2/3}$. Since $q < 1$ and $k^{-3/2} > \exp(-k^{1/2})$, the second term is bounded by (a constant times) the first, which completes the proof. □

5.3. **Applying the general lemmas to $\nu_{q,N}$: the lower bound.** Here is the lower bound statement we prove:

Proposition 5.5. *Let X be distributed according to $\nu_{q,N}$ as defined in (10) and μ_q be as in (7). Then for any $W > 0$, there exist positive constants $c_W > 0$, N_0, and k_0 such that, for any $N \geq N_0$, $k_0 \leq k \leq WN^{2/3}$, and $q \in [k^{-2}, 1)$,*

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

The idea is to combine Lemma 5.2 with Theorem 4.1. Recall that the lower bound from Lemma 5.2 has the term $\mathbb{E}[(X)_k] - \mathbb{E}[X^{2k}]^{1/2}\mathbb{P}(X > R)^{1/2}$; in our argument, we will show that this is lower bounded by $\frac{1}{2}\mathbb{E}[(X)_k]$.

Now, we can lower bound the first term by Theorem 4.1 and upper bound $\mathbb{E}[X^{2k}]^{1/2}$ by the just established Proposition 5.3. However the former is smaller by a factor of $\exp(-k^2/(2\mu_q N))$ which must be compensated for by $\mathbb{P}(X > R)^{1/2}$ to achieve the overall lower bound of $\frac{1}{2}\mathbb{E}[(X)_k]$; here we will take $R = \mu_q N(1 + N^{-1/3})$, though the choice of $N^{-1/3}$ is not significant. The next statement gives a strong enough bound on this probability, using Markov inequality and the just established upper bounds on $\mathbb{E}[X^k]$.

Lemma 5.6. *Let $L > 0$ and X be distributed according to $\nu_{q,N}$ as defined in (10). There exists $N_0 = N_0(L)$ such that, for $q \in [N^{-2}, 1)$ and $N \geq N_0$,*

$$\mathbb{P}\left(X \geq \mu_q N(1 + N^{-1/3})\right) \leq \exp\left(-LN^{1/3}\right)$$

In fact, one expects that this probability should decay something like $\exp(-(N^{1/3})^{3/2}) = \exp(-N^{1/2})$, since $N \times N^{-1/3} = N^{2/3} = N^{1/3} \times N^{1/3}$, and one expects the upper tail of the expected empirical distribution to be similar to that of the top particle, which indeed fluctuates on scale $N^{1/3}$ with an upper tail exponent of 3/2. Proving this would require our estimates on the polynomial moment bounds from Proposition 5.3 to go up to $O(N)$ instead of $O(N^{2/3})$, which we omit doing since we have no other need for it.

We also observe that, when $k \leq WN^{2/3}$ and for a correspondingly large choice of L, $\exp(-LN^{1/3}) \ll \exp(-k^2/(2\mu_q N))$ as the latter is at least $\exp(-\frac{1}{2}W^2N^{1/3})$. Thus the requirement for this probability estimate outlined above is met.

Proof of Lemma 5.6. By Markov's inequality and Proposition 5.3, for any $N \geq N_0$, $k_0 \leq k \leq 2LN^{2/3}$, and $q \in [k^{-2}, 1)$,

$$\mathbb{P}\left(X \geq \mu_q N(1 + \varepsilon^{1/2})\right) \leq \frac{\mathbb{E}[X^k]}{(\mu_q N)^k (1 + \varepsilon^{1/2})^j} \leq C_L (q^{1/6}k)^{-3/2}(1 + \varepsilon^{1/2})^{-k}.$$

We will ultimately pick $\varepsilon < \frac{1}{2}$. Since $(1 + x)^{-1} \leq 1 - \frac{x}{2}$ for $x \in [0, \frac{1}{2}]$, we obtain

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

Using that $q \geq N^{-2}$ and setting $k = 2LN^{2/3}$, $\varepsilon = N^{-2/3}$, we obtain

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

which is clearly smaller than $\exp(-LN^{1/3})$ for all large N (depending on L). □

Proof of Proposition 5.5. We will take $R = \mu_q N(1+\varepsilon^{1/2})$ for $\varepsilon = N^{-2/3}$ and apply Lemma 5.2. So we need to lower bound $\mathbb{E}[(X)_k] - \mathbb{E}[X^{2k}]^{1/2}\mathbb{P}(X > \mu_q N(1+N^{-1/3}))^{1/2}$, in particular, to show that the second term is much smaller than the first. We start with upper bounding the second term using Proposition 5.3 and 5.6, for a L to be chosen soon and for N large enough (depending on L):

$$\mathbb{E}[X^{2k}]^{1/2}\mathbb{P}\left(X > \mu_q N(1+N^{-1/3})\right)^{1/2} \le C_W (q^{1/6}k)^{-3/4}(\mu_q N)^k \times \exp\left(-\tfrac{1}{4}LN^{1/3}\right).$$

On the other hand, by Theorem 4.1, for $N \ge N_0$, $k_0 \le k \le WN^{2/3}$, and $q \in [k^{-2}, 1)$ and using that $\mu_q \ge 1$ in the second inequality,

$$\begin{aligned}\mathbb{E}[(X)_k] &\ge C'_W (q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(-\frac{k^2}{2\mu_q N}\right)\\ &\ge C'_W (q^{1/6}k)^{-3/2}(\mu_q N)^k \exp\left(-W^2 N^{1/3}\right). \end{aligned} \tag{15}$$

So we see that

$$\frac{\mathbb{E}[X^{2k}]^{1/2}\mathbb{P}\left(X > \mu_q N(1+N^{-1/3})\right)^{1/2}}{\mathbb{E}[(X)_k]} \le \frac{C_W}{C'_W}(q^{1/6}k)^{3/4}\exp\left(-\left[\tfrac{1}{4}L + W^2\right]N^{1/3}\right).$$

Next we use that $k \le WN^{2/3}$ and set $L = 4W^2 + 1$. Since $q < 1$, the previous right-hand side is upper bounded by

$$\frac{C_W}{C'_W}N^{1/2}\exp\left(-N^{1/3}\right),$$

which is less than $\frac{1}{2}$ for all large enough N depending only on W; also note that L was chosen as a function of W, so we may take the overall lower bound on N to be purely a function of W.

Now applying Lemma 5.2 with $R = \mu_q N(1+\varepsilon^{1/2})$ with $\varepsilon = N^{-2/3}$, Theorem 4.1, and using the above, we get

$$\begin{aligned}\mathbb{E}[X^k] &\ge \frac{1}{2}\mathbb{E}[(X)_k]\exp\left(\frac{k(k-1)}{2\mu_q N(1+\varepsilon^{1/2})}\right)\\ &\ge c_W(q^{1/6}k^{-3/2})(\mu_q N)^k \exp\left(-\frac{k^2}{2\mu_q N} + \frac{k(k-1)}{2\mu_q N(1+\varepsilon^{1/2})}\right).\end{aligned}$$

Since $(1+x)^{-1} \ge 1 - x$,

$$\begin{aligned}\exp\left(-\frac{k^2}{2\mu_q N} + \frac{k(k-1)}{2\mu_q N(1+\varepsilon^{1/2})}\right) &\ge \exp\left(-\frac{k^2}{2\mu_q N} + \frac{k(k-1)}{2\mu_q N}(1-\varepsilon^{1/2})\right)\\ &\ge \exp\left(-\frac{k^2\varepsilon^{1/2}}{2\mu_q N} - \frac{k}{2\mu_q N}\right)\\ &\ge \exp(-cW^2),\end{aligned}$$

since $\varepsilon^{1/2} = N^{-1/3}$, $\mu_q \ge 1$ and $k \le WN^{2/3}$. So overall we see that

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

□

6. Concentration inequalities and the proof of Theorem 1.1

In this section we combine Theorem 1.3 (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 Section 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 will be proven in Section 6.1.

Theorem 6.1. *Let $I \in \mathbb{N} \cup \{\infty\}$ and suppose $X_1, \ldots, X_I$ are independent, and assume that there exists $C_1 < \infty$ and $\rho_1, \ldots, \rho_I > 0$ such that each X_i satisfies*

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

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 6.1 we may give the proof of the main result, Theorem 1.1.

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

$$\mathbb{P}\left(x_N(N) \leq f_q N - (\log q^{-1})^{-1}\theta N^{1/3}\right), \tag{16}$$

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

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

(In going from the statement of Theorem 1.1 to (16) we implicitly use that $\psi_q''(\log_q u)$ is bounded over $q \in (0,1)$ for fixed $u \in (0,1)$, see for example [MS09, eq. (1.6)], so that its effect can be absorbed into the constant c which will show up in the exponent of the final probability bound.)

We let ε be such that $q = e^{-\varepsilon}$, and we will sometimes, when convenient, write things in terms of ε in the proof, so that, for example, $(\log q^{-1})^{-1}$ becomes ε^{-1}. We also define $\tilde{f}_q = f_q - 1$. By Theorem 2.4, we know that $L + N \stackrel{d}{=} x_N(N)$, where L is the LPP value from the topmost site to ∞ in the infinite periodic environment defined in Section 1.4. Let $L_N^{(i)}$ be the last passage time from the top to the bottom of the i^{th} 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

$$(16) = \mathbb{P}\left(L \leq \tilde{f}_q N - \varepsilon^{-1}\theta N^{1/3}\right) \leq \mathbb{P}\left(\sum_{i=0}^\infty L_N^{(i)} \leq \tilde{f}_q N - \varepsilon^{-1}\theta N^{1/3}\right).$$

We next add and subtract the law of large numbers term of $L_N^{(i)}$, which as we see from Theorem 1.3 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

$$\mathbb{P}\left(\sum_{i=0}^\infty \left(L_N^{(i)} - 2N\frac{uq^i}{1-uq^i}\right) \leq \tilde{f}_q N - \sum_{i=0}^\infty 2N\frac{uq^i}{1-uq^i} - \varepsilon^{-1}\theta N^{1/3}\right) \tag{17}$$

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

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

Putting this back into (17), we see that

$$(17) = \mathbb{P}\left(\sum_{i=0}^{\infty}\left(L_N^{(i)} - 2N\frac{uq^i}{1-uq^i}\right) \leq -\varepsilon^{-1}\theta N^{1/3}\right). \tag{18}$$

In the remainder of the proof we will invoke the concentration bound from Theorem 6.1 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 1.3 that there exist positive constants c and N_0 such that, when $u^2q^{2i} \in (0,1)$, $t > 0$, and $N > N_0$,

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

Rewriting $u^{1/3}q^{i/3}(1-u^2q^{2i})^{-1}$ as $\Theta(q^{i/3}(\log q)^{-1}i^{-1}) = \Theta(q^{i/3}\varepsilon^{-1}i^{-1})$ (where we absorb u dependencies into the Θ notation), we see that, for $t > 0$ and possibly a different c,

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

which implies that, again for $t > 0$,

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

We next want invoke Theorem 6.1 with $I = \infty$ and (from the previous display) $\rho_i = \varepsilon^{3/2}i^{3/2}q^{-i/2}$. Now,

$$\sigma_2 = \sum_{i=1}^{\infty}\rho_i^{-2} = \sum_{i=1}^{\infty}\varepsilon^{-3}i^{-3}q^i = \Theta(\varepsilon^{-3}) \quad \text{and} \quad \sigma_{2/3} = \sum_{i=1}^{\infty}\rho_i^{-2/3} = \sum_{i=1}^{\infty}\varepsilon^{-1}i^{-1}q^{i/3} = \Theta(\varepsilon^{-1}|\log\varepsilon^{-1}|);$$

where the bounds on $\sigma_{2/3}$ hold because, for some $c' > 0$, $c'\sum_{i=1}^{\varepsilon^{-1}} i^{-1} \leq \sigma_{2/3} \leq \sum_{i=1}^{\infty} i^{-1}e^{-\varepsilon i/3}$, and both lower and upper bounds are of order $\log(\varepsilon^{-1})$ for small ε, and bounded for ε bounded away from zero.

With these estimates, we obtain from Theorem 6.1 that, for all $t > 0$,

$$\mathbb{P}\left(\sum_{i=0}^{\infty}\left(L_N^{(i)} - 2N\frac{uq^i}{1-uq^i}\right) \leq -tN^{1/3} - C\varepsilon^{-1}|\log\varepsilon^{-1}|\right) \leq \exp\left(-c\varepsilon^{3/2}t^{3/2}\right).$$

So, putting in $t = \theta\varepsilon^{-1}$, we obtain that, for $\theta > 2C|\log\varepsilon^{-1}|$ with C as in the last display, (18) is bounded as

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

Thus we obtain the desired bound (16) with $\theta_0(q) = C|\log\varepsilon^{-1}| = 2C|\log(\log q^{-1})|$. $\square$

6.1. **Proving the concentration inequality.** The only remaining task is to prove Theorem 6.1. As is typical for such inequalities, the main step is to obtain a bound on the moment generating function.

Proposition 6.2. *Suppose X is such that*

$$\mathbb{P}(X \geq t) \leq C_1\exp(-\rho t^{3/2}) \tag{19}$$

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. First, we see that

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

Next, we break up the integral into two at 0 and use the hypothesis (19) for the second term:

$$\begin{aligned}
\lambda \int_{-\infty}^{\infty} e^{\lambda x}\cdot \mathbb{P}(X \geq x)\,\mathrm{d}x &= \lambda \int_{-\infty}^{0} e^{\lambda x}\cdot \mathbb{P}(X \geq x)\,\mathrm{d}x + \int_{0}^{\infty} e^{\lambda x}\cdot \mathbb{P}(X \geq x)\,\mathrm{d}x \\
&\leq \lambda \int_{-\infty}^{0} e^{\lambda x}\,\mathrm{d}x + C_1\lambda \int_{0}^{\infty} e^{\lambda x - \rho x^{3/2}}\,\mathrm{d}x \\
&= 1 + C_1\lambda \int_{0}^{\infty} e^{\lambda x - \rho x^{3/2}}\,\mathrm{d}x.
\end{aligned}$$

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

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

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

$$\begin{aligned}
C_1\lambda^3\rho^{-2} \int_{0}^{\infty} e^{\lambda^3\rho^{-2}(y - y^{3/2})}\,\mathrm{d}y &\leq C_1\lambda^3\rho^{-2} \cdot \int_{0}^{\infty} e^{\lambda^3\rho^{-2}(1 - \frac{1}{2}y^{3/2})}\,\mathrm{d}y \\
&= C_1\lambda^3\rho^{-2} \cdot e^{\lambda^3\rho^{-2}} \cdot c(\lambda^3\rho^{-2})^{-2/3}\Gamma(5/3) \\
&= C \cdot C_1(\lambda^3\rho^{-2})^{1/3} e^{\lambda^3\rho^{-2}}.
\end{aligned}$$

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

$$\mathbb{E}[e^{\lambda X}] \leq 1 + C \cdot C_1(\lambda^3\rho^{-2})^{1/3} e^{c\lambda^3\rho^{-2}}$$

Now we break into two cases depending on whether $\lambda^3\rho^{-2}$ is less than or greater than 1: if $\lambda^3\rho^{-2} \leq 1$, then, since $1+x \leq \exp(x)$ and $C\cdot C_1 \exp(c\lambda^3\rho^{-2}) \leq C\cdot C_1\exp(c)$, we obtain, for an absolute constant $\tilde{C}$,

$$\mathbb{E}[e^{\lambda X}] \leq 1 + C\cdot C_1 \exp(c)(\lambda^3\rho^{-2})^{1/3} \leq \exp\left(\tilde{C}\cdot C_1\lambda\rho^{-2/3}\right).$$

On the other hand if $\lambda^3\rho^{-2} \geq 1$, we observe that, since $x \leq \exp(x)$, and by increasing the coefficient in the exponent,

$$1 + C\cdot C_1\lambda^3\rho^{-2}e^{c\lambda^3\rho^{-2}} \leq e^{c'\lambda^3\rho^{-2}}.$$

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,

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

□

Proof of Theorem 6.1. Following the proof of the Chernoff bound, we exponentiate inside the probability (with $\lambda > 0$ to be chosen shortly) and apply Markov's inequality:

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{I} X_i \geq t + C \cdot C_1\sigma_{2/3}\right) &= \mathbb{P}\left(\exp\left(\lambda\sum_{i=1}^{I} X_i\right) \geq \exp\Big(\lambda t + \lambda C \cdot C_1 \cdot \sigma_{2/3}\Big)\right)\\ &\leq e^{-\lambda t - \lambda C\cdot C_1\cdot\sigma_{2/3}}\prod_{i=1}^{I}\mathbb{E}\left[e^{\lambda X_i}\right]\\ &\leq \exp\left(-\lambda t - \lambda C\cdot C_1\sigma_{2/3} + C\cdot C_1\left[\lambda\sum_{i=1}^{I}\rho_i^{-2/3} + \lambda^3\sum_{i=1}^{I}\rho_i^{-2}\right]\right),\\ &= \exp\left(-\lambda t + C\cdot C_1\cdot\lambda^3\sigma_2\right)\end{aligned}$$

the penultimate line using Proposition 6.2 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,

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

Appendix A. Proof of the LPP–q-Whittaker connection

Here we give the proof of Theorem 2.4, which was explained to us by Matteo Mucciconi. As indicated earlier, 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 2.3 on the relation between $x_N(T)$ and the q-Whittaker measure reduces the proof of Theorem 2.4 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.1. *For $1 \leq j \leq \min(N, T)$, let $L^{(j)}$ be as defined above in the environment defined before Theorem 2.4 and let $\mu \sim \mathbb{W}^{(q)}_{a;b}$ with $a_i = b_j = u$ for all $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$. Then, jointly across $1 \leq j \leq \min(N, T)$,*

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

Theorem 2.4 follows immediately from combining the $j = 1$ case of Theorem A.1 with Theorem 2.3.

Proof of Theorem A.1. 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 $m_{(i,j);k} = 0$ 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. 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. Thus one has a bijection between the set of M and the set of (V, W, κ); we call this $\overline{\Upsilon}$.

Now, [IMS21, Theorem 1.2] asserts that $L^{(j)}(M)$ is equal to $\mu_1 + \ldots + \mu_j$, 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 $\overline{\Upsilon}$ when the entries $m_{(i,j);k}$ of M are distributed as independent $\mathrm{Geo}(q^k a_i b_j)$, in particular, show that it is $\mathbb{W}^{(q)}_{a,b}$.

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

Lemma A.2. *For a infinite matrix M and $(V, W; \kappa)$ as above, define weight functions*

$$W_1(M) = \Big(\sum_k \sum_j m_{(i,j);k}\Big)_{1\leq i\leq N}$$
$$W_2(M) = \Big(\sum_k \sum_i m_{(i,j);k}\Big)_{1\leq j\leq N}$$
$$W_3(M) = \sum_k \sum_{i,j} k m_{(i,j);k}.$$

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 $\overline{\Upsilon}$, it holds that, for $i = 1, 2, 3$,*

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

We will prove this after completing the proof of Theorem 2.4. 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-row-length 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 $\overline{\Upsilon}$ to M with $m_{(i,j);k} \sim \mathrm{Geo}(q^k a_i b_j)$ is the q-Whittaker measure. Then marginalizing to the length of the first row will complete the proof. Denoting the V obtained by applying $\overline{\Upsilon}$ to M by $V(M)$,

$$\begin{aligned}
\mathbb{P}\left(V(M) \in \mathrm{vst}(\mu)\right) &\propto \sum_{M : V(M) \in \mathrm{vst}(\mu)} \prod_{i,j,k} (q^k a_i b_j)^{m_{(i,j);k}} \\
&= \sum_{M : V(M) \in \mathrm{vst}(\mu)} q^{\sum_{i,j,k} k m_{(i,j);k}} \prod_i a_i^{\sum_j m_{(i,j);k}} \prod_j b_j^{\sum_i m_{(i,j);k}}
\end{aligned}$$

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

By applying the bijection $\overline{\Upsilon}$ and Lemma A.2, and recalling the definitions of $\widetilde{W}_i$, we see that the previous line equals

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

where $x^V = \prod_i x_i^{\#(V,i)}$.

Now it is known that the first factor is $b_\mu(q)$ (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 $\mathbb{W}^{(q)}_{a,b}$ at μ, as desired. □

Proof of Lemma A.2. 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 $W_1(M) = (\#(P,i))_{1\le i\le N}$ and $W_2(M) = (\#(Q,j))_{1\le j\le N}$. 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. □

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 (13), which we label S, i.e.,

$$S := \sum_{i=1}^{k-1} \frac{1}{\frac{i}{k}(1-\frac{i}{k})} \exp\left[k\left(\tfrac{i}{k}\log q^{-1} + 2H(i/k) + \frac{k(1-\frac{i}{k})}{N}\right)\right]. \tag{20}$$

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 Section B.2.

B.1. The upper bound.

Proposition B.1. *There exist positive constants C, k_0, and N_0 such that for all $N \ge N_0$, $k_0 \le k \le N$, and $q \in [k^{-2}, 1)$,*

$$S \le Cq^{-1/4}k^{1/2}\left[\frac{\left(1+q^{1/2}\exp\left(\frac{1}{2}kN^{-1}\right)\right)^2}{q}\right]^k.$$

Proof. Define $f, g : [0,1] \to \mathbb{R}$ by

$$f(x) = x\log q^{-1} + 2H(x) + \frac{k}{N}(1-x) \qquad \text{and} \qquad g(x) = (x(1-x))^{-1}.$$

Then the sum S is

$$S = \sum_{i=1}^{k-1} g(i/k) \exp\{kf(i/k)\}.$$

The first step is to identify the location x_0 where f is maximized. Calculating $f'(x)$ and equating to zero,

$$f'(x) = \log q^{-1} + 2\log\left(\frac{1-x}{x}\right) - \frac{k}{N} = 0$$

yields that

$$x_0 = \left(1 + q^{1/2}\exp(\tfrac{1}{2}kN^{-1})\right)^{-1}. \tag{21}$$

For future reference we also record that

$$f''(x) = -\frac{1}{x} - \frac{1}{1-x} = -\frac{1}{x(1-x)} \implies f''(x_0) = -\frac{(1+q^{1/2}\exp(\frac{1}{2}kN^{-1}))^2}{q^{1/2}\exp(\frac{1}{2}kN^{-1})} = -\Theta(q^{-1/2}). \tag{22}$$

To apply Laplace's method, we need to have bounds on f over its domain, which we will obtain by Taylor approximations. We will expand to third order as we need to include the second order term precisely, which we just calculated. So next we obtain bounds on the third derivative of f.

We observe that there exists C such that for $x \in [\frac{1}{4}, 1)$,

$$f'''(x) = \frac{1}{x^2} - \frac{1}{(1-x)^2} \le C.$$

So by Taylor's theorem, we see that, if $\varepsilon > 0$ is such that $x_0 - \varepsilon \in [\frac{1}{4}, 1]$

$$f(x_0 - \varepsilon) \le f(x_0) - \varepsilon f'(x_0) + \frac{\varepsilon^2}{2}f''(x_0) + \frac{C}{6}\varepsilon^3 = f(x_0) + \frac{\varepsilon^2}{2}f''(x_0) + \frac{C}{6}\varepsilon^3$$

since $f'(x_0) = 0$. We may pick ε_0 an absolute constant such that, if $0 < \varepsilon < \varepsilon_0$, then (recalling that $f''(x_0) < 0$)

$$\frac{C}{6}\varepsilon < \frac{1}{4}|f''(x_0)|;$$

that ε_0 can be taken to not depend on q follows from the fact that $|f''(x_0)|$ can be lower bounded by an absolute constant, due to (22). So, for $-\varepsilon_0 < \varepsilon < \varepsilon_0$,

$$f(x_0 - \varepsilon) \le f(x_0) - \frac{\varepsilon^2}{4}|f''(x_0)|. \tag{23}$$

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

$$f(x) \le f(x_0) - \frac{\varepsilon^2}{4}|f''(x_0)| \le f(x_0) - Cq^{-1/2}\varepsilon^2, \tag{24}$$

the second inequality using again that $|f''(x_0)| = \Theta(q^{-1/2})$. We will apply this for both ε and ε' shortly.

We next break up the sum S into three subsums S_1, S_2, and S_3. Let $0 < \varepsilon' < \varepsilon_0$ be a constant whose value will be set shortly and $0 < \varepsilon < \varepsilon_0$ be fixed. S_1 corresponds to $i = 1$ to $i = \lceil k(x_0 - \varepsilon)\rceil$, S_2 to $i = \lceil k(x_0 - \varepsilon)\rceil + 1$ to $\lceil k(x_0 + \varepsilon')\rceil$, and S_3 to $\lceil k(x_0 + \varepsilon')\rceil + 1$ to $k - 1$.

We start by bounding S_1 using the above groundwork. Recall that $g(x) = (x(1-x))^{-1}$. Note that, for $1 \leq i \leq k-1$ and $k \geq 2$, $g(i/k) \leq 2k$. We drop the $\lceil\rceil$ in the notation for convenience. Using (24) and that $q \leq 1$,

$$S_1 = \sum_{i=1}^{k(x_0-\varepsilon)+1} g(i/k)\exp\left(kf(i/k)\right) \leq e^{kf(x_0)} \cdot 2k^2 \exp(-Ck\varepsilon^2) \leq Ce^{kf(x_0)}.$$

Next we bound S_3. Because x_0 can be arbitrarily close to 1, we cannot set ε' to be a constant, as then $x_0 + \varepsilon' > 1$ for some range of q. So we set $\varepsilon' = \min(\varepsilon_0, \frac{1}{2}(1-x_0)) = \Theta(q^{1/2})$ and apply (24) to see that

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

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

$$S_3 \leq Ck^{1/2}q^{-1/4}e^{kf(x_0)}.$$

Finally we turn to the main sum, S_2, which consists of the range $\frac{i}{k} \in [x_0-\varepsilon, x_0+\varepsilon']$. 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 we have taken ε' to depend on x_0.

Lemma B.2. *There exists an absolute constant C such that for $x \in [x_0-\varepsilon, x_0+\varepsilon']$, $g(x) \leq Cg(x_0)$.*

Proof. We have to upper bound $g(x)/g(x_0) = \frac{x_0(1-x_0)}{x(1-x)}$. When $x \in [x_0-\varepsilon, x_0]$, this ratio is upper bounded by x_0/x; since $x \geq x_0 - \varepsilon$, and $x_0 \geq \frac{1}{4}$ always, if $\varepsilon < \frac{1}{8}$ say, the ratio is uniformly upper bounded. When $x \in [x_0, x_0+\varepsilon]$, this ratio is upper bounded by $(1-x_0)/(1-x)$; since $1-x \geq 1-x_0-\varepsilon' = \frac{3}{2}(1-x_0)$, the ratio is again upper bounded by a constant. □

So we see, from Lemma B.2 and (23)

$$\begin{aligned} S_2 &= \sum_{i=k(x_0-\varepsilon)+1}^{k(x_0+\varepsilon')} g(i/k)\exp\left(kf(i/k)\right) \\ &\leq Cg(x_0) \sum_{i=k(x_0-\varepsilon)+1}^{k(x_0+\varepsilon')} \exp\left[k\left(f(x_0) - c\left(\frac{i}{k}-x_0\right)^2 |f''(x_0)|\right)\right] \\ &\leq Cg(x_0)e^{kf(x_0)} \sum_{i=-\infty}^{\infty} \exp\left[-ck^{-1}\left(i-kx_0\right)^2 |f''(x_0)|\right] \\ &= Cg(x_0)e^{kf(x_0)} \sum_{i=-\infty}^{\infty} \exp\left[-ck^{-1}i^2|f''(x_0)|\right] \end{aligned}$$

Recall that $f''(x_0) = \Theta(q^{-1/2})$, and so the coefficient of i^2 is at least $c(q^{1/2}k)^{-1}$. We want to bound the above series using Proposition B.3 ahead, which requires that this coefficient of i^2 is positive and bounded. This is verified using the lower bound assumption on q, i.e., that $q \geq k^{-2}$, as this immediately yields that $q^{1/2}k \geq 1$, so that the coefficient of i^2 is at most c. Then by Proposition B.3 with $\gamma = c(q^{1/2}k)^{-1}$, and using that $g(x_0) = \Theta(q^{-1/2})$,

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

Thus overall we have showed that

$$S = S_1 + S_2 + S_3 \leq Ce^{kf(x_0)}\left[1 + q^{-1/4}k^{1/2} + q^{-1/4}k^{1/2}\right] \leq Cq^{-1/4}k^{1/2}e^{kf(x_0)}.$$

Now it can be computed that $f(x_0)$ simplifies to

$$\log\left[\frac{(1+q^{1/2}\exp(\frac{1}{2}kN^{-1}))^2}{q}\right],$$

which 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.3 (page 157 of [SS11])**.** *For any $M > 0$, there exists a constant $C > 0$ such that for $0 < \gamma \leq M$,*

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

B.2. **The lower bound.** Recall the definition of S from (20).

Proposition B.4. *There exist positive constants C and k_0 such that for all $q \in [k^{-2}, 1)$ and $k_0 \leq k \leq N$,*

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

Proof. As in the proof of Proposition B.1, we focus around the point kx_0, where $x_0 = (1 + q^{1/2}\exp(\frac{1}{2}kN^{-1}))^{-1}$ from (21). So, for $\varepsilon, \varepsilon' > 0$ to be chosen,

$$\begin{aligned} S &\geq \sum_{i=k(x_0-\varepsilon)}^{k(x_0+\varepsilon')} \frac{1}{\frac{i}{k}(1-\frac{i}{k})}\exp\left[k\left(\tfrac{i}{k}\log q^{-1} + 2H(i/k) + \frac{k(1-\frac{i}{k})}{N}\right)\right] \\ &= \sum_{i=k(x_0-\varepsilon)}^{k(x_0+\varepsilon')} g(i/k)\exp\left[kf(i/k)\right]. \end{aligned}$$

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 $c > 0$,

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

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

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

$$= f(x_0) - \frac{(x-x_0)^2}{2}|f''(x_0)| + \frac{1}{6}(x-x_0)^3 f'''(x);$$

the second equality by noting that, since $x - x_0 < 0$, the infimum is equivalent to maximizing $f'''(y)$ over $y \in [x, x_0]$ and then noting that, since $f'''(y)$ is a decreasing function, the maximum is achieved at x. Now, if $x \geq \frac{1}{2}$, then $f'''(x) < 0$, 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 $f''(x_0)$ and $f'''(x)$ are uniformly bounded independent of q, so, by picking ε small enough also independent of q, we can ensure that $|x - x_0| f'''(x) < C f''(x_0)$ for some absolute constant C.

Thus over all, we have shown that there exists $\varepsilon > 0$ and $C > 0$ independent of q such that, for $x \in [x_0 - \varepsilon, x_0]$,

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

Using this we see that

$$\begin{aligned}
S &\geq c g(x_0) e^{k f(x_0)} \sum_{i=k(x_0-\varepsilon)}^{k x_0} \exp\left(-Ck|f''(x_0)|\left(\frac{i}{k} - x_0\right)^2\right) \\
&= \frac{1}{2} c g(x_0) e^{k f(x_0)} \sum_{i=k(x_0-\varepsilon)}^{k(x_0+\varepsilon)} \exp\left(-Ck|f''(x_0)|\left(\frac{i}{k} - x_0\right)^2\right)
\end{aligned}$$

Now it is easy to see that $\sum_{i:|i-kx_0|>k\varepsilon} \exp\left(-Ck^{-1}|f''(x_0)|(i-kx_0)^2\right) \leq C\exp(-ck|f''(x_0)|\varepsilon^2)$, so the previous display is lower bounded by (using Proposition B.3 in the third line)

$$\begin{aligned}
& c g(x_0) e^{k f(x_0)} \left[\sum_{i=-\infty}^{\infty} \exp\left(-Ck^{-1}\left(i - kx_0\right)^2\right) - C\exp(-ck\varepsilon^2)\right] \\
&= c g(x_0) e^{k f(x_0)} \left[\sum_{i=-\infty}^{\infty} \exp\left(-Ck^{-1}|f''(x_0)| i^2\right) - C\exp(-ck|f''(x_0)|\varepsilon^2)\right] \\
&\geq c g(x_0) e^{k f(x_0)} \left[k^{1/2}|f''(x_0)|^{-1/2} - C\exp(-c\varepsilon^2 k|f''(x_0)|)\right] \\
&\geq c g(x_0) e^{k f(x_0)} \left[q^{1/4} k^{1/2} - C\exp(-c\varepsilon^2 k q^{-1/2})\right] \\
&\geq c q^{-1/4} k^{1/2} e^{k f(x_0)}.
\end{aligned}$$

In the case that $x_0 \leq \frac{1}{2}$, the same proof works after noting that, since also $x_0 \geq \frac{1}{4}$, 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). ☐

References

[ACG22] Amol Aggarwal, Ivan Corwin, and Promit Ghosal. The ASEP speed process. *arXiv preprint arXiv:2204.05395*, 2022.

[ACQ11] Gideon Amir, Ivan Corwin, and Jeremy Quastel. 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.

[AH21] Amol Aggarwal and Jiaoyang Huang. Edge statistics for lozenge tilings of polygons, II: Airy line ensemble. *arXiv preprint arXiv:2108.12874*, 2021.

[BC14] Alexei Borodin and Ivan Corwin. Macdonald processes. *Probability Theory and Related Fields*, 158(1):225–400, 2014.

[BCD21] Guillaume Barraquand, Ivan Corwin, and Evgeni Dimitrov. Fluctuations of the log-gamma polymer free energy with general parameters and slopes. *Probability Theory and Related Fields*, 181(1):113–195, 2021.

[BCFV15] Alexei Borodin, Ivan Corwin, Patrik Ferrari, and Bálint Vető. Height fluctuations for the stationary KPZ equation. *Mathematical Physics, Analysis and Geometry*, 18(1):1–95, 2015.

[BDM+01] Jinho Baik, Percy Deift, Kenneth DT McLaughlin, Peter Miller, and Xin Zhou. Optimal tail estimates for directed last passage site percolation with geometric random variables. *Advances in Theoretical and Mathematical Physics*, 5(6):1–41, 2001.

[BG16] Alexei Borodin and Vadim Gorin. Moments match between the KPZ equation and the Airy point process. *SIGMA. Symmetry, Integrability and Geometry: Methods and Applications*, 12:102, 2016.

[BGHH20] Riddhipratim Basu, Shirshendu Ganguly, Alan Hammond, and Milind Hegde. Interlacing and scaling exponents for the geodesic watermelon in last passage percolation. *arXiv preprint arXiv:2006.11448*, 2020.

[BGHK21] Riddhipratim Basu, Shirshendu Ganguly, Milind Hegde, and Manjunath Krishnapur. Lower deviations in β-ensembles and law of iterated logarithm in last passage percolation. *Israel Journal of Mathematics*, 2021+. To appear.

[BGS17] Riddhipratim Basu, Shirshendu Ganguly, and Allan Sly. Upper tail large deviations in first passage percolation. *arXiv preprint arXiv:1712.01255*, 2017.

[BL21] Jinho Baik and Zhipeng Liu. Periodic TASEP with general initial conditions. *Probability Theory and Related Fields*, 179(3):1047–1144, 2021.

[BO17] Alexei Borodin and Grigori Olshanski. The ASEP and determinantal point processes. *Communications in Mathematical Physics*, 353(2):853–903, 2017.

[BO21] Dan Betea and Alessandra Occelli. Peaks of cylindric plane partitions. *arXiv preprint arXiv:2111.15538*, 2021.

[Bor07] Alexei Borodin. Periodic Schur process and cylindric partitions. *Duke Mathematical Journal*, 140(3):391–468, 2007.

[Bor18] Alexei Borodin. Stochastic higher spin six vertex model and Macdonald measures. *Journal of Mathematical Physics*, 59(2):023301, 2018.

[BSS14] Riddhipratim Basu, Vladas Sidoravicius, and Allan Sly. Last passage percolation with a defect line and the solution of the slow bond problem. *arXiv preprint arXiv:1408.3464*, 2014.

[CC22] Mattia Cafasso and Tom Claeys. A Riemann-Hilbert approach to the lower tail of the Kardar-Parisi-Zhang equation. *Communications on Pure and Applied Mathematics*, 75(3):493–540, 2022.

[CCO20] Philip Cohen, Fabio Deelan Cunden, and Neil O'Connell. Moments of discrete orthogonal polynomial ensembles. *Electronic Journal of Probability*, 25:1–19, 2020.

[CG20a] Ivan Corwin and Promit Ghosal. KPZ equation tails for general initial data. *Electronic Journal of Probability*, 25, 2020.

[CG20b] Ivan Corwin and Promit Ghosal. Lower tail of the KPZ equation. *Duke Mathematical Journal*, 169(7):1329–1395, 2020.

[CH14] Ivan Corwin and Alan Hammond. Brownian Gibbs property for Airy line ensembles. *Inventiones mathematicae*, 195(2):441–508, 2014.

[CH16] Ivan Corwin and Alan Hammond. KPZ line ensemble. *Probability Theory and Related Fields*, 166(1):67–185, 2016.

[CHH19] Jacob Calvert, Alan Hammond, and Milind Hegde. Brownian structure in the KPZ fixed point. *arXiv preprint arXiv:1912.00992*, 2019.

[DOV22] Duncan Dauvergne, Janosch Ortmann, and Bálint Virág. The directed landscape. *Acta Mathematica*, 2022+. To appear.

[DV21] Duncan Dauvergne and Bálint Virág. Bulk properties of the Airy line ensemble. *The Annals of Probability*, 49(4):1738–1777, 2021.

[EGO22] Elnur Emrah, Nicos Georgiou, and Janosch Ortmann. Coupling derivation of optimal-order central moment bounds in exponential last-passage percolation. *arXiv preprint arXiv:2204.06613*, 2022.

[EJS20] Elnur Emrah, Chris Janjigian, and Timo Seppäläinen. Right-tail moderate deviations in the exponential last-passage percolation. *arXiv preprint arXiv:2004.04285*, 2020.

[EJS21] Elnur Emrah, Christopher Janjigian, and Timo Seppäläinen. Optimal-order exit point bounds in exponential last-passage percolation via the coupling technique. *arXiv preprint arXiv:2105.09402*, 2021.

[FSV14] Gregorio Moreno Flores, Timo Seppäläinen, and Benedek Valkó. Fluctuation exponents for directed polymers in the intermediate disorder regime. *Electronic Journal of Probability*, 19:1–28, 2014.

[GH20] Shirshendu Ganguly and Milind Hegde. Optimal tail exponents in general last passage percolation via bootstrapping & geodesic geometry. *arXiv preprint arXiv:2007.03594*, 2020.

[GH22] Shirshendu Ganguly and Milind Hegde. Sharp upper tail estimates and limit shapes for the KPZ equation via the tangent method. *arXiv preprint arXiv:2208.08922*, 2022.

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

[Ham22] Alan Hammond. Brownian regularity for the Airy line ensemble, and multi-polymer watermelons in Brownian last passage percolation. *Memoirs of the American Mathematical Society*, 277(1363), 2022.

[Hua21] Jiaoyang Huang. Edge statistics for lozenge tilings of polygons, I: Concentration of height function on strip domains. *arXiv preprint arXiv:2108.12872*, 2021.

[IMS21] Takashi Imamura, Matteo Mucciconi, and Tomohiro Sasamoto. Skew RSK dynamics: Greene invariants, affine crystals and applications to q-Whittaker polynomials. *arXiv preprint arXiv:2106.11922*, 2021.

[IMS22] Takashi Imamura, Matteo Mucciconi, and Tomohiro Sasamoto. Solvable models in the KPZ class: approach through periodic and free boundary Schur measures. *arXiv preprint arXiv:2204.08420*, 2022.

[IS16] Takashi Imamura and Tomohiro Sasamoto. Determinantal structures in the O'Connell-Yor directed random polymer model. *Journal of Statistical Physics*, 163(4):675–713, 2016.

[Joh00] Kurt Johansson. Shape fluctuations and random matrices. *Communications in mathematical physics*, 209(2):437–476, 2000.

[Joh01] Kurt Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Annals of Mathematics*, 153(1):259–296, 2001.

[KC18] Arun Kumar Kuchibhotla and Abhishek Chakrabortty. Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *arXiv preprint arXiv:1804.02605*, 2018.

[Led05a] Michel Ledoux. Deviation inequalities on largest eigenvalues. In *GAFA seminar notes*, 2005.

[Led05b] Michel Ledoux. Distributions of invariant ensembles from the classical orthogonal polynimials: the discrete case. *Electronic Journal of Probability*, 10:1116–1146, 2005.

[LR10] Michel Ledoux and Brian Rider. Small deviations for beta ensembles. *Electron. J. Probab.*, 15:1319–1343, 2010.

[LS22a] Benjamin Landon and Philippe Sosoe. Tail bounds for the O'Connell-Yor polymer. *arXiv preprint arXiv:2209.12704*, 2022.

[LS22b] Benjamin Landon and Philippe Sosoe. Upper tail bounds for stationary KPZ models. *arXiv preprint arXiv:2208.01507*, 2022.

[Mac98] Ian Grant Macdonald. *Symmetric functions and Hall polynomials*. Oxford university press, 1998.

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

[MP16] Konstantin Matveev and Leonid Petrov. q-randomized Robinson-Schensted-Knuth correspondences and random polymers. *Annales de l'Institut Henri Poincaré D*, 4(1):1–123, 2016.

[MS09] Toufik Mansour and Armend Sh Shabani. Some inequalities for the q-digamma function. *J. Inequal. Pure Appl. Math*, 10(1), 2009.

[Sep12] Timo Seppäläinen. Scaling for a one-dimensional directed polymer with boundary conditions. *The Annals of Probability*, 40(1):19–73, 2012.

[SS90] Bruce E Sagan and Richard P Stanley. Robinson-Schensted algorithms for skew tableaux. *Journal of Combinatorial Theory, Series A*, 55(2):161–193, 1990.

[SS11] Elias M Stein and Rami Shakarchi. *Fourier analysis: an introduction*, volume 1. Princeton University Press, 2011.

[SS22] Dominik Schmid and Allan Sly. Mixing times for the TASEP on the circle. *arXiv preprint arXiv:2203.11896*, 2022.

[Vet22] Bálint Vető. Asymptotic fluctuations of geometric q-TASEP, geometric q-PushTASEP and q-PushASEP. *Stochastic Processes and their Applications*, 148:227–266, 2022.

[Wid02] Harold Widom. On convergence of moments for random young tableaux and a random growth model. *International Mathematics Research Notices*, 2002(9):455–464, 2002.